Computational studies of Lignocellulose deconstruction

Gnana S. Gnanakaran

Los Alamos National Labs
I. Thermal decomposition of biomass
II. Cellulose Properties responsible for recalcitrance to digestion
III. Alter cellulose substrate for efficient digestion
IV. Mechanistic kinetic model for Cellulases
U.S. economy is heavily dependent on oil imports.

U.S. has 4% of the world's population but consumes 25% world's oil production.

Domestic production of crude oil has declined but the dependence on imported oil has increased and accounts for 65% of crude oil supplies.

Biofuels are an alternative to conventional energy sources that increase our Nation's energy security by dramatically reducing our dependence on imported oil.
Currently, our Nation's growing bioethanol industry is largely based on the use of starch in grains such as corn, that are also needed for food.

Corn for ethanol production has increased by nearly 5-fold from 2000 to 2008.
More than 130 ethanol plants are now in operation. Process is similar to making “moonshine” which initially became popular during prohibition.
Where will the food come from?


36 billion gallons of annual renewable fuel use by 2022 and required that 60% to be met by advanced biofuels, including cellulosic ethanol
Biofuels need to be made from a wide range of hardy and fast-growing plants, such as switchgrass—which is a perennial native to American prairies. It requires about a quarter of the irrigation and fertilization of row crops.

Lignocellulosic biomass, the inedible fibrous material from wood and plant stems, is an abundant alternative source.

corn stalks are more challenging!
Industry has been slow to explore the potential of lignocellulosic biomass because so far it is more difficult and costly than starch to convert to ethanol.
Biomass Conversion

Main problem: complex structure of plant Cell wall, impedes efficient Conversion into sugars that can be fermented to ethanol

Source: DOE EERE Office of the Biomass Program, Multi-year Program Plan, Appendix C.
Biomass Production

Solids: cellulose, hemicellulose, lignin

Pretreatment

Chemicals

Cellulose

Enzymatic Hydrolysis

Dissolved sugars

Fermentation of Sugars

Biomass recalcitrance

Lignocellulose

Hemicellulose

Cellulose

Lignin

Pretreatment

Effect of Pretreatment
Biomass

Plant cell walls contain cellulose microfibrils, hemicelluloses & lignins

Fragment of a cellulose molecule

Alternating glucose residues are in an inverted orientation so the cellubiose (a disaccharide) is the repeating structural unit.

Crystalline cellulose

The glucan chains contain thousands of glucose residues.

Cellulose microfibrils are composed of linear chains of glucose molecules that hydrogen bond to form the microfibrils.
Lignin
- Complex aromatic structure
- High Energy Content
- Resists biochemical conversion

Hemicellulose
- Xylose
- Polymer of 5- and 6- carbon sugars
Pretreatment

- Lignin
- Cellulose
- Hemicellulose

The diagram illustrates the process of pretreatment, denoted by the arrow labeled "Pretreatment," showing the transformation from a dense structure to a more dispersed one.
LIGNIN DEGRADATION

Microorganisms:

Woodrotting fungi – depolymerization of lignins by C-C bond cleavage

Part of the few microorganisms capable of a complete lignin degradation: lignin $\rightarrow$ $\text{CO}_2 + \text{H}_2\text{O}$

Polymeric lignin

Lignin Degrading Enzymes
- Manganese peroxidase
- Lignin peroxidase

H$_2$O$_2$ producing enzymes
- Glyoxal oxidase
- Glucose-2-oxidase

Small metabolites
- oxalate, veratryl alcohol, unsaturated lipids

Phanerochaete chrysosporium
QM Studies on Lignin Models

- Radical cationic state playing a essential part in Lignin fragmentation
- Evaluate relative chemical reactivity trends from Ionization potentials
- Identify weak C—C bond fragmentation in lignin compounds
- Recognize the potential reactive sites

Study of Diverse Linkages in Lignin

Monomeric precursors of lignin

- syringyl (S)
- guaiacyl (G)
- \(p\)-hydroxyphenylpropane (H)

Common linkages in lignin

Density Functional Calculations of bond fragmentation in 65 distinct Lignin Model Compounds
Bond dissociation energies (kcal/mol)

Ether bond linkages

- β-O-4 – L1-L21
- α-O-4 – L22-L31
- 4-O-5 – L32 & L33

C-C bond linkages

- β-1 – L34-L46
- α-1 – L47-L53
- β-5 – L54-L57
- 5-5 – L58-L65
- β-O-4 – L1-L21
Weakest lignin linkage

Ether linkage - $\alpha$-O-4  
BDE = 48.31 kcal/mol

C-C linkage - $\beta$-1  
BDE = 64.7 kcal/mol
Strongest lignin linkage

Ether linkage - 4-O-5

\[
\text{L32}
\]

\[82.54 \text{kcal/mol}\]

C-C linkage - \(\beta\)-5

\[
\text{L56}
\]

\[127.6 \text{kcal/mol}\]

\(\beta\)-1 double bond

\[
\text{L43}
\]

\[165.8 \text{kcal/mol}\]
Influence of *ortho* methoxy group substitutions

Bond Dissociation Energies of $\alpha$-O-4 Ether linkages

<table>
<thead>
<tr>
<th>Compound</th>
<th>Bond Dissociation Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L22</td>
<td>50.24</td>
</tr>
<tr>
<td>L25</td>
<td>56.24</td>
</tr>
<tr>
<td>L27</td>
<td>48.31</td>
</tr>
</tbody>
</table>

These substitutions tend to lower the strength of $\beta$-O-4 linkages as well.
Metal oxides as catalysts for fast pyrolysis

Collaboration with Tom Elder (USDA)
Catalytic Pyrolysis of Bibenzyl

Dehydrogenation – stilbene
Favorable – lower temperature

CH2-CH2 cleavage – facilitated by CaO

Typical pyrolysis products
Exploiting quantum and classical computational methods to study the atomistic details of the interactions of NaCl with β-glucose, oligomeric and crystalline fibrillar cellulose.
Simulation Details

• System
  – Aqueous α-glucose (α-D-Glc)
  – Aqueous β-glucose (β-D-Glc)

• Replica exchange molecular dynamics (REMD) simulations
  – Control (H2O, 0% NaCl, Sol)
  – 5.0% wt. NaCl (0.9 M)
  – 10.5% wt. NaCl (2 M)

• NVT REMD simulations
  – Temperature range 275 ~ 505 K
  – Used Glycam06 for glucose
  – Ion force field from Joung & Cheatham, JPCB 2008, 112, 9020
Radial distribution function – Solvation by Water

No Significance difference between $\alpha$-glucose and $\beta$-glucose
Influence of Alkali Metal on stereoisomers of glucose

Alkali metal effects that depend on stereochemistry
- Coordination with Alkali Metals

Alkali metal effects that didn’t depend on stereochemistry
- Conformations of Glucose
Affinity of Na to hydroxyl or hydroxymethyl oxygen: O2>O6>O3

Affinity for β-glucose is lower, and the order is slightly different: O2≈O6>O3
• Several bridging coordination positions with Na+ possible

• However, probabilities are lower for β-glucose
Summary – Part I

• Reactivity trends of diverse linkages and substitutions of lignin model compounds

Ongoing Work

• Interactions of bi-benzyl with (CaO)$_2$ metal oxide have been studied using DFT approach to understand mechanistic insights involved in the catalytic pyrolysis reactions

• Acid Catalyzed reactions of beta5 model compounds

• Effect of Na+ on cellulose pyrolysis
Enzymatic Degradation

endoglucanases randomly attack the cellulose chain, creating free ends for exoglucanases to process along the chain and create dimers of glucose (cellobiose). β-glucosidases, then, hydrolyze these cellobiose units.
PART II.

Cellulose Properties responsible for Recalcitrance to digestion
Native plant-derived cellulose I microfibrils

Crystalline core

Fibril periphery

Inter-Sheet

Intra-Sheet

Cellulose I\(_\beta\)

References:

Chundawat et al., 2011, J Am Chem Soc, 133, 11163
Ding et al., 2006, J Agric Food Chem, 54, 597
Cellulose – A Dynamical System


Movie: Al French (USDA)
Multiscale modeling of cellulose assembly/disassembly

- **intra-strand H-bonding:**
  - All-atom MD Simulation
  - Enhanced Sampling

- **2D sheet formation:**
  - Inter vs intra H-bonding
  - Coarse-grained Model

- **3D fiber formation:**
  - Replica Exchange MD Simulations

- **Enzymatic degradation:**
  - All-atom MD simulations
  - Agent Based Modeling
Initial Studies

Coarse-grained Hydrogen Bond Network Model for Crystalline Cellulose

The multiplicity in H-bond network and switching between different H-bond networks with increasing temperature make the cellulose assembly very stable.

"The stability of cellulose: a statistical perspective from a coarse grain model of hydrogen bonding"

All-atom molecular dynamics of Cellulose Oligomers and their assembly

- Effective sampling of structural properties cellulose including ring flips
- Mechanical properties- chains are getting less flexible with increasing # of monomers

Stability of cellulose fibers is due to

- *Intra*-molecular hydrogen bonding

- *Inter*-molecular hydrogen bonding

Why is cellulose so stable?

M. Wada, JPSB (2002)
Cross-strand H-bonding elements

Total 5 cross-strand bonding states: φ, e, f, g, e+f

At low T, either O3….HO6 or [O3….HO6, O2….HO6]
As T increases, a switch to O3….HO6 & O6….HO2

Single chains (2mer, 4mer, 6mer):
- Torsions
- Flexibility
- Ring flips
Chains (2mer, 4mer, 6mer): End-to-end distances and Flexibilities

\[
\cos \langle \theta \rangle (T) = \frac{k_B T}{c}
\]

\( c = 47.8 \text{ kJ/mol} \)

Persistence length, \( l_0 \propto c / k_B T \approx 10.5 \text{ nm} \)

Rigidity: 6mer > 4mer > 2mer
Ring Flipping Tendencies in Cellulose Oligomers

Rings at the non-reducing end is more stable
"Fiber Formation"

- At high $T$ (>460K), the system is in a random configuration with little order; while at lower $T$, the system is in a much more ordered state.

- At low $T$ (<310K), all factors indicate the system is going to a different state.
Rigidity increases as chain length increases & Ring Flipping Tendencies

Ability to form Multiple Hydrogen Bond Networks

Capturing synergy of enzyme cocktails with Rule-based Model

Summary

PART I I

Multiscale models of cellulose

Self assembly of chains
PART III.

Alter cellulose substrate for efficient digestion
Cellulose can exist in various crystalline forms

Perez & Samain, 2010
How to efficiently hydrolyze cellulose?

- Get a better hydrolyzing agent
- Altering substrate cellulose
  - Decrease crystallinity
  - Alter hydrogen bond network
- Enzyme (cellulases) engineering

Liquid Ammonia
Rearrangement of H-bonding during treatment of cellulose with NH₃

Liquid Ammonia Pretreatment

Nishiyama et al., Biomacromolecules 9(11), 3133-3140 (2008)
Why is it easier to digest cellulose III?

**Restructuring Cellulose Hydrogen Bond Network**

**Chemical Properties:**
Quantum Mechanical Calculations
(JPC A, 2011, 115 pp. 14191-14202)

**Dynamical Properties:**
All-atom Molecular Dynamics
(JACS, 2011, 133 pp. 11163-11174)
(JPC B, 2011, 115 pp. 9782-9788)

**Bulk Properties:**
Coarse-grained Description
(JPC B, 2012, PMID: 22712833)
Initial Molecular Events Associated with Liquid NH$_3$ Treatment

Cellulose $I_\beta$ and Cellulose $III_\parallel$ Model conformations

All calculations were carried out using hybrid DFT m062X/6-31+G(d,p) using G09
Interplay between H-bonding and Stacking

**Cellulose I$_{\beta}$**

- Strong Hydrogen bonding interactions within a sheet
- Highly cooperative stacking between sheets

**Cellulose III$_{1}$**

- Slightly weaker hydrogen bonding but cooperative
- Non-cooperative stacking interactions

The color scale indicates the charges on the atoms: red = most negative, green = neutral, blue = most positive charge: ±0.04 au isosurface.

Improved exposure of the hydroxyl groups and glycosidic oxygens in cellulose III$_1$
O6 side chain based Conformational Variability

**gg** is preferred for solvated oligomers

No **tg** of Cellulose Ib & Ia crystals!

No **gt** of dimer, trimer & tetramer x-ray structures
Overall, cellulose III$_{\text{I}}$ surface is more “similar” to amorphous cellulose.
COARSE-GRAINED MODEL (II)

\[
H = \sum_{i \neq j} 4\varepsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \lambda_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \\
\sum_{bonds} \frac{K_b}{2} (r_{ij} - r_0)^2 + \\
\sum_{angles} \frac{K_\theta}{2} (\theta_{ij} - \theta_0)^2 + \\
\sum_{dihe} C \cos(\gamma - \delta)
\]

Cellulose I\(_{\beta}\) \(\delta = 180^\circ\) \(tg\)

Cellulose III\(_{\beta}\) \(\delta = 130^\circ\) \(gt\)

\(\sigma = 2.0, \lambda = 1, \varepsilon = 2.0\)

\(\sigma = 3.5, \lambda = 1, \varepsilon = 2.0\)

\(\sigma = 2.0, \lambda = 0, \varepsilon = 1.0\)

\(\sigma = 4.0, \lambda = 1, \varepsilon = 2.0\)

\(\sigma = 3.0, \lambda = 1, \varepsilon = 2.0\)
Summary II.
Differences between cellulose I and III

**Cellulose I\(_{\beta}\)**
- Strong Hydrogen bonding interactions within sheet
- Cooperative stacking
- Less structural flexibility
- Low hydration
- More stable
- Ideal packing

**Cellulose III\(_{1}\)**
- Slightly weaker hydrogen bonding but cooperative
- Non-cooperative stacking
- More structural flexibility
- High hydration
- Less stable
- Loosely packed
I. Rapid formation of an extended H-bond network between surface chains and NH$_3$.

II. Relative shifting of the layers that in turn leads to the formation of channels orthogonal to the (100) and (-100)

III. These channels allow NH$_3$ to penetrate into the cellulose fibril
PART IV.

Mechanistic kinetic model for Cellulases
But something odd was noticed about cellulase binding

4 Degrees C, 2 hrs equilibration time

Shishir Chundawat, GLBRC
How does lower enzyme binding in Cellulose III lead to increase in catalytic activity?

"Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis,”
D. Gao, S.P.S. Chundawat, A. Sethi, V. Balan, S. Gnanakaran and Bruce E Dale
(under review)
Overview of kinetic model for CBH1

This model takes into account adsorption of carbohydrate binding module (CBM) and catalytic domain (CD) individually to the cellulose surface.

In addition to chain ends, CBM can bind to other regions of the cellulose as well (CD cannot do this).

Once CD is adsorbed to the cellulose surface, the cellulose chain can slide into the active site a single glucose unit at a time.

Once a productive enzyme (10 glucose units in active site) is formed, the protein can hydrolyze the cellulose to form product-bound enzyme (P10).

Cellobiose can then undock from the enzyme to form the enzyme bound substrate with 8 glucose units remaining within CD (NP8-CB).
Key Steps in the Model

Binding

Sliding

Hydrolysis
Binding of enzyme

Off-pathway nonproductive complex OP

 apo Enzyme E

NP \(_1\) (1 glucose unit in CD active site)

Both domains bound to cellulose

\[
\frac{d[E]}{dt} = -k^{CBM}_{on}[E][C_{fe}] + k^{CBM}_{off}[NP_0] - k^{CD}_{on}[E][C_{fe}] + k^{CD}_{off}[NP_1] - k^{CBM}_{on}[E][C_{inner}] + k^{CBM}_{off}[OP]
\]

\[
\frac{d[OP]}{dt} = k^{CBM}_{on}[E][C_{inner}] - k^{CBM}_{off}[OP]
\]

\[
\frac{d[NP_0]}{dt} = k^{CBM}_{on}[E][C_{fe}] - k^{CBM}_{off}[NP_0] + k^{CD}_{off}[NP_1] - k^{CD}_{on}[NP_0]
\]
Intramolecular binding constants can be up to 100 times greater.

Sliding of enzyme

\[ k_{rev} \leftrightarrow k_{slide} \]

\[ \frac{k_{CBM}}{k_{on}} \]

\[ k_{off} \]

\[ k_{rev} \leftrightarrow k_{slide} \]

\[ NP_i \leftrightarrow \overline{NP}_i \]

\[ NP_i \leftrightarrow \overline{NP}_i \]

\[ NP_{i-1} \leftrightarrow \overline{NP}_{i-1} \]

\[ NP_{i+1} \leftrightarrow \overline{NP}_{i+1} \]

\[ i \text{ refers to number of glucose units inside the catalytic domain of the protein.} \]

The cellulose chain can slide in one glucose unit at a time into the CD.

\[ i = 2 \text{ to } 7 \]

\[
\frac{d[NP_i]}{dt} = k_{slide}[NP_{i-1}] - k_{rev}[NP_i] + k_{rev}[NP_{i+1}] - k_{slide}[NP_i] + k_{off}^{CBM}[\overline{NP}_i] - \frac{k_{on}^{CBM}}{k_{off}^{CBM}}[NP_i]
\]

\[
\frac{d[NP_i]}{dt} = k_{slide}[NP_{i-1}] - k_{rev}[NP_i] + k_{rev}[NP_{i+1}] - k_{slide}[NP_i] + \frac{k_{on}^{CBM}}{k_{off}^{CBM}}[NP_i] - k_{off}^{CBM}[\overline{NP}_i]
\]
Catalytic activity by CBH`

\[ \frac{d[NP_8]}{dt} = k_{slide}[NP_7] - k_{rev}[NP_8] + k_{rev}[NP_9] - k_{slide}[NP_8] + k_{off}^{CBM}[NP_8] - k_{on}^{CBM}[NP_8] + k_u^{CB}[NP_8 - CB] - k_b^{CB}[CB][NP_8] \]

\[ \frac{d[P_{10}]}{dt} = k_{slide}[NP_9] - k_{rev}[P_{10}] + k_{off}^{CBM}[P_{10}] - k_{on}^{CBM}[P_{10}] + k_b^{CB}[NP_8 - CB] - k_h[P_{10}] \]

\[ \frac{d[NP_8 - CB]}{dt} = k_b^{CB}[NP_8][CB] - k_u^{CB}[NP_8 - CB] + k_h[P_{10}] - k_b^{CB}[NP_8 - CB] + k_{off}^{CBM}[NP_8 - CB] - k_{on}^{CBM}[NP_8 - CB] \]

Binding and undocking rate of cellobiose

Forward and backward catalytic rates by enzyme.

These four parameters are independent of cellulose crystal and depend on enzyme.
Consider different scenarios for combination of factors that could lead to increase of hydrolysis rate of enzyme simultaneously with decrease in amount of bound enzyme as seen in experiments.
only one scenario explains observations
Higher Sliding & Lower Binding Rates in Cellulose III

Only increase in sliding rate with decrease in binding rates of both domains to the surface of cellulose reproduces the observations.
Can we implement a more complex kinetic model?

1) How does the substrate change during hydrolysis?
2) What cellulase cocktails would result in an efficient cellulose degradation?
Samples from Agent-based Simulations

Time=0.00  Crystalline cellulose surface

Time=0.513  Enzyme binding

Time=0.60  Disrupts the hydrogen bonds

Time=10.662  Cleaving the glycosidic bond

Oligomers entering into solution

Current Studies – Enzymatic Degradation
Enzymatic Hydrolysis of Cellulose:

Results: Endo-enzymes

Time course of hydrolysis

Distribution of oligomers

Enzyme loading

Enzyme adsorption
Enzymatic Hydrolysis of Cellulose:

Results: Cellulase Cocktails

- **Endo(Num)**
- **Exo(Num)**
- **Endo+Exo(Num)**
- **Endo+Exo(Expected,Num)**
Summary

Simple mechanistic kinetic model finds
- Sliding rate of CBH1 is faster in Cellulose III
- Binding rate of CBH1 is lower in Cellulose III

Stochastic model that captures substrate properties during hydrolysis and effect of enzyme cocktail has been developed
- possible 3d extension of the model
Increasing Spatial & Temporal Scales

**Biophys. J. 2009, 96:3032**

**Research in Progress**

**Biophys. J. 2009, 96:3032**

**Optimal Cellulase Cocktails for Efficient Degradation of Cellulose**

**Biophs. J. 2009, 96:3032**

**Plasticity of Hydrogen Bond Network in Cellulose**

**Amorphous Cellulose**

**J. Am. Chem. Soc. 2009 131: 14786**

**Transformation of Cellulose I to III**


**Stacking & Hydrogen Bonding in Cellulose**

**Cellulose. 2011. 18, 191**

**Stereochemistry of glucose**

**Manuscript in Preparation**

**Acta Crystal. D. 2010. 66:1184**

**Multi-resolution Theoretical Approaches**

- Quantum Mechanical
- Classical MD Simulations
- Coarse-grained Simulations
- Rule-based Models

**Manuscript in Preparation**

J. Phys. Chem. A. 2011 133, 11163


J. Am. Chem. Soc. 2011 133, 11163

J. Am. Chem. Soc. 2009 131: 14786


Cellulose. 2011. 18, 191

Manuscript in Preparation


Optimal Cellulase Cocktails for Efficient Degradation of Cellulose

Quantum Mechanical

Classical MD Simulations

Coarse-grained Simulations


Rule-based Models
Enzymatic Digestion Game

Bullet Physics – heavily used by the game developers