Glyco – MGrid

CyberInfraStructure for e-Glycomics
Biomacromolecules & Carbohydrates - Structural Motif

DNA (102D)  
CGCAATTTCGCGCGC  
AAATTTCG

Protein (1W4Q)  
LYS GLU THR ALA ALA ALA LYS PHE GLU ARG GLN HIS MET ASP SER SER THR SER ALA ALA SER SER SER ASN TYR CYS ASN

Carbohydrate (1CAP)  
b-D-Manp-(1-6)+
|  
a-D-Manp-(1-6)+
|  
b-D-Manp-(1-3)+
b-D-Manp-(1-4)-b-D-GlcNAc
Gene – Genomics
Protein – Proteomics
Glycoconjugates – Glycomics

e-Glycomics

DNA Protein

Sweet/Sugar

Carbohydrate
Grid computing

Grid protocols link everything else - databases, simulation tools, and underutilized computing power. Computer grids function analogously to the electric grid. Widely dispersed on-line resources become available to focus on the problem at hand.

Glycomics

Glycomics is the effort to understand and harness sugars that are naturally made by the human body, in order to improve health by bolstering the immune system or by arresting disease processes. Sounds easy, but there is, as of yet, no code that determines the structure of the sugars.
## Databases for Carbohydrates and Glyco-compounds in WWW

<table>
<thead>
<tr>
<th>Database</th>
<th>1D Sequence</th>
<th>3D Structure</th>
<th>User-Interface</th>
<th>Conformation Computing</th>
<th>Solvent Effect</th>
<th>Carbo-Parameter</th>
<th>Grid System</th>
<th>Sharing System</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarbBank</td>
<td>O</td>
<td>×</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>SUGABASE</td>
<td>O</td>
<td>×</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>SWEET</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>GlycoMaps Database</td>
<td>×</td>
<td>×</td>
<td>O</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Glydict</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>CAZy</td>
<td>O</td>
<td>×</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Glycoconjugate DataBank</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>GlycoSuite</td>
<td>O</td>
<td>×</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Functional Glycomics</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>MD-Simulator</td>
<td>×</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>BioSimGrid</td>
<td>×</td>
<td>×</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>×</td>
<td>O</td>
<td>×</td>
</tr>
</tbody>
</table>

Glyco-MGrid System O O O O O O O O O

CarbBank [http://bssy01.lancs.ac.uk/rgp/pages/gag/carbbank.htm](http://bssy01.lancs.ac.uk/rgp/pages/gag/carbbank.htm)
SUGABASE [http://www.boc.chem.uu.nl/sugabase/sugabase.html](http://www.boc.chem.uu.nl/sugabase/sugabase.html)
GlycoMaps [http://www.dkfz.de/spec/glycosciences.de/glycomaps/)](http://www.dkfz.de/spec/glycosciences.de/glycomaps/]
Glycoconjugates DB [http://akashia.sci.hokudai.ac.jp/](http://akashia.sci.hokudai.ac.jp/)
GlycoSuite [https://tmat.proteomesystems.com/glycosuite/](https://tmat.proteomesystems.com/glycosuite/)
Glydict [http://www.dkfz.de/spec/glydict](http://www.dkfz.de/spec/glydict)
Functional Glycomics [http://www.functionalglycomics.org/static/consortium/](http://www.functionalglycomics.org/static/consortium/)
Parameters are derived principally from experimental data, and some calculations. Sources include:

- IR spectra; bond stretching, angle bending, ...
- X-ray crystal structures; geometries and energetics
- solvent simulation to match physical properties (TIP3P)
- ab initio calculations; partial charges
- free energy perturbation with a known experimental result

\[ E_{\text{total}} = \sum_r K_b (r - r_0)^2 + \sum_\theta K_\theta (\theta - \theta_0)^2 + \sum_\phi |K_\phi - K_\phi \cos n\phi + K_\sigma (\omega - \omega_0)^2 \]

\[ + \sum \left[ (\frac{A}{r_{ij}^{12}}) - (\frac{B}{r_{ij}^6}) \right] s\lambda (r_{ij}^2, r_{on}^2, r_{off}^2) + \sum \left( \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} \right) [1 - 2 \left( \frac{r_{ij}}{r_{cut}} \right)^2 + \left( \frac{r_{ij}}{r_{cut}} \right)^4] \]
"Publish or perish" refers to the pressure to publish work constantly in order to further or sustain one's career in academia. The competition for tenure-track faculty positions in academia puts increasing pressure on scholars to publish new work frequently.
CHARMM

Chemistry at HARvard
Macromolecular Mechanics

Supporting Area
- Molecular Mechanics
- Molecular Dynamics
- Monte Carlo
- QM/MM
- Free Energy Calculation
- Coordinate Analysis

http://www.charmm.org
Personal Job

Shell Terminal

Molecular Simulations & e-Science

Web-Interface

Portal

Collaboration

Multidisciplinary Work

e-Science CyberInfrastructure
Vision of MGrid

Cooperative Simulation
- Protein Scientists
- Structural Biologists
- Glycan Scientists

Integrated Environment
- MGrid-Portal
- MGrid-PSE
- MGrid-CG
- MGrid-SDG
- MGrid-AG
- Glyco-MGrid
- Proteo-MGrid
- Chiral-MGrid

Comparative Study
- Simulation Parameter Sweeping
- MGrid Structure Database
- Researcher & User Group
- Results Sharing

Research Community
Architecture of Glyco-MGrid
Components of Glyco-MGrid

Computation Facility

Simulation Program
- Molecular Coordinate File
- Simulation Input File
- Molecular Parameter File
- Molecular Topology File

Computing Resources

DataBase Facility

Simulation Input
- Job Title
- Job Description
- Molecular Name
- Force Field
- Program
- Target System
- Temperature
- Pressure
- Frame Number
- Temp. Bath
- Pressure Bath
- Solvation
- PBC
- Crystal Type
- Ensemble
- Dielectrics
- NonBond Option
- Simulation Time
- Time Step
- Total Step
- Update Number
- Save Frequency
- Restart Saving

Simulation Output
- Trajectory File
- Structure File
- Coordinate File
- Restart File
- Velocity File
- Output Log File
- Float Number
- Number List
- Molecular Figure
- Data Plotting
- 2-D Scatter Plot
- Probability Plot
Research Example

Thermal Stability of Unglycosylated RNaseA and N-acetylglycosylated Glc_RNase

Protein $+$ Carbohydrates $\Rightarrow$ Stability Increase
The radial distribution function (RDF) of water molecules around disaccharides. (A) RDF around all oxygen atoms of each disaccharide (B) RDF around glycosidic oxygen of 1, 2, and 3 (C) RDF around glycosidic oxygen of 1, 4, 5, 6, and 7 (D) RDF around glycosidic oxygen of 1, 8, 9, 10, and 11. The RDF of water around whole trehalose is not dissimilar to other sugars. But, the RDF of water around glycosidic oxygen of trehalose is quite dissimilar to other linkage typed disaccharides. Only 1,1-linked trehalose showed anisotropic water distribution among the other 1,2-, 1,3-, 1,4- or 1,6-linked sugars.

Schematic diagrams of the glycosidic linkage model of the sugars studied in this work.

Cooperative Simulation

Research Community

RNaseA Protein

glycosylation

Glycosylation with different position

Researcher B

RNaseA Protein

Glycosylation with different sugar

Researcher C

Integrated Environment

MGrid

MGrid -PSE
MGrid -CG
MGrid -SDG
MGrid -AG

Researcher A

Trajectory File
Structure File
Coordinate File
Restart File
Velocity File
Output Log File

MGrid Portal

Browsing & Searching
Similar Research

Data Retrieval & Sharing

Researcher B and C perform Re-simulation from the results of Researcher A

Comparative Study & Results Sharing

DataBase

Molecular Model & Simulation Output

Statistical & Data Mining

DataBase

MGrid Portal

Data Retrieval & Sharing

MGrid -AG

MGrid -CG

MGrid -SDG

MGrid -PSE

Researcher A

Researcher B and C

RNaseA Protein

Institute
Glyco-Structure Motif Searching Project Scenario

Researcher A
Glyco-Coordinates

Researcher B
Glyco-Coordinates

Researcher C
Fragment Coordinates

Cooperative Simulation

Glyco-MGrid
- MGrid-PSE
- MGrid-CG
- Glyco-ATK
- Glyco-SimDB

Molecular Simulations with Parameter Sweeping

Motif A
Motif B
Motif C
Motif D

Combination

Pattern Clustering & Motif Discovering

Research Community

Integrated Environment

Comparative & Statistical Study

Results Sharing
Disease-Related Glycan Modeling Project Scenario

- **Researcher A**
  - Glyco-Coordinates

- **Researcher B**
  - Glyco-Coordinates

- **Researcher C**
  - Fragment Coordinates

**Cooperative Simulation**

**Glyco-MGrid**
- MGrid - PSE
- MGrid - CG
- Glyco - ATK
- Glyco-SimDB

**Molecular Simulations with Parameter Sweeping**

- **3D-Structure A**
- **3D-Structure B**
- **3D-Structure C**
- **3D-Structure D**

**Combination**

**Disease Diagnosis & Cure Design**

**Research Community**

**Integrated Environment**

**Comparative & Statistical Study**

**Results Sharing**
## Current-Status of Glyco-MGrid Project

<table>
<thead>
<tr>
<th>Glyco-compounds</th>
<th>Research Project</th>
<th>Job Details</th>
<th>Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycan</td>
<td>Branching</td>
<td>Branching effect on the glycan conformation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glycan Core</td>
<td>Calculation of biological core glycan structure</td>
<td></td>
</tr>
<tr>
<td>Sugar compounds</td>
<td>Glycosidic Linkage</td>
<td>Glycosidic angle map of each glycosidic linkage</td>
<td></td>
</tr>
<tr>
<td>Poly-saccharide</td>
<td>Microbial Carbohydrates</td>
<td>MC and MD conformational searching of bacterial carbohydrates</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cyclic Carbohydrates</td>
<td>Conformational characteristics of cyclic carbohydrates</td>
<td></td>
</tr>
<tr>
<td>Glycoprotein</td>
<td>Unfolding</td>
<td>Unfolding simulations for the stability and conformations of glycosylated proteins</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Folding</td>
<td>Folding simulations for the conformational effects of glycosylation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Function</td>
<td>MD simulations for the biological function of glycoproteins</td>
<td></td>
</tr>
</tbody>
</table>
Use on the MGrid

1. Project & CHARMM Job
2. CHARMM Input Script & File Upload
3. CHARMM Job Running
4. Job Monitoring

Chemistry at Harvard Macromolecular Mechanics

Copyright(c) 1984-2001 President and Fellows of Harvard College
All Rights Reserved
Current operating system: Linux-2.6.3-1.687kbase (EBB)@nodall.konuk.ac.
Created on 10/17/6 at 9:19:36 by user: kgrid1001

Maximum number of ATOMS: 50120, and RESIDUES: 18030
Current HEAP size: 1024000, and STACK size: 2000000

RDFITL> *******************************************
RDFITL> B-D-GAL-ALAC-(1-2)-A-D-GAL-ALAC CON TOPOLOGY FILE
RDFITL> BY YOUNG JIN CHOI (KONUK UNIVERSITY, SEOUL, KOREA), 2005
RDFITL> *******************************************
RDFITL> *******************************************

CHARMM
CHARMM: bomb -1
CHARMM: open read unit 10 card name "con.rtf"
YODPS: Attempting to open: con.rtf:
OPEN: Unit 10 opened for READONLY access to con.rtf
CHARMM: read rdf file unit 10
MAINL: Residue topology file being read from unit 10.
TITLE: COMBINED CHARMM ALL-HETEROGEN TOPOLOGY FILE FOR CHARMM2 PROTEINS AND CHARMM7 LIPIDS
TITLE: ALL-HETEROGEN TOPOLOGY FILE FOR PROTEINS
TITLE: ALL-HETEROGEN TOPOLOGY FILE FOR LIPIDS
TITLE: ALL-HETEROGEN TOPOLOGY FILE FOR LIPIDS
TITLE: DIRECT COMMENTS TO ALEXANDER D. MACKERELL JR.
TITLE: EMAIL: ALEX.MACKERELL@ABML.FUDU.UCB.US
TITLE: 410-706-1442 OR EMAIL: ALEX.MACKERELL@ABML.FUDU.UCB.US
TITLE: DIRECT COMMMENTS TO ALEXANDER D. MACKERELL JR.
TITLE: EMAIL: ALEX.MACKERELL@ABML.FUDU.UCB.US
TITLE: 410-706-1442 OR EMAIL: ALEX.MACKERELL@ABML.FUDU.UCB.US

410-706-1442 OR EMAIL: ALEX.MACKERELL@ABML.FUDU.UCB.US

There were 1 warning(s) from RFFDR.

(CHARMM) close unit 10.
5. Post-Processing of the Trajectory