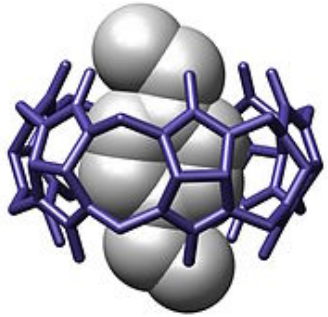
A Linked Data based Approach to Car bohydrate Complexation



Yinling An,
Karpjoo Jeong
Department of Advanced Technology Fusion

Daham Jeong,
Seunho Jung
Department of Bioscience and Biotechnology
Konkuk University

CiR Cyber-Infrastructure Research

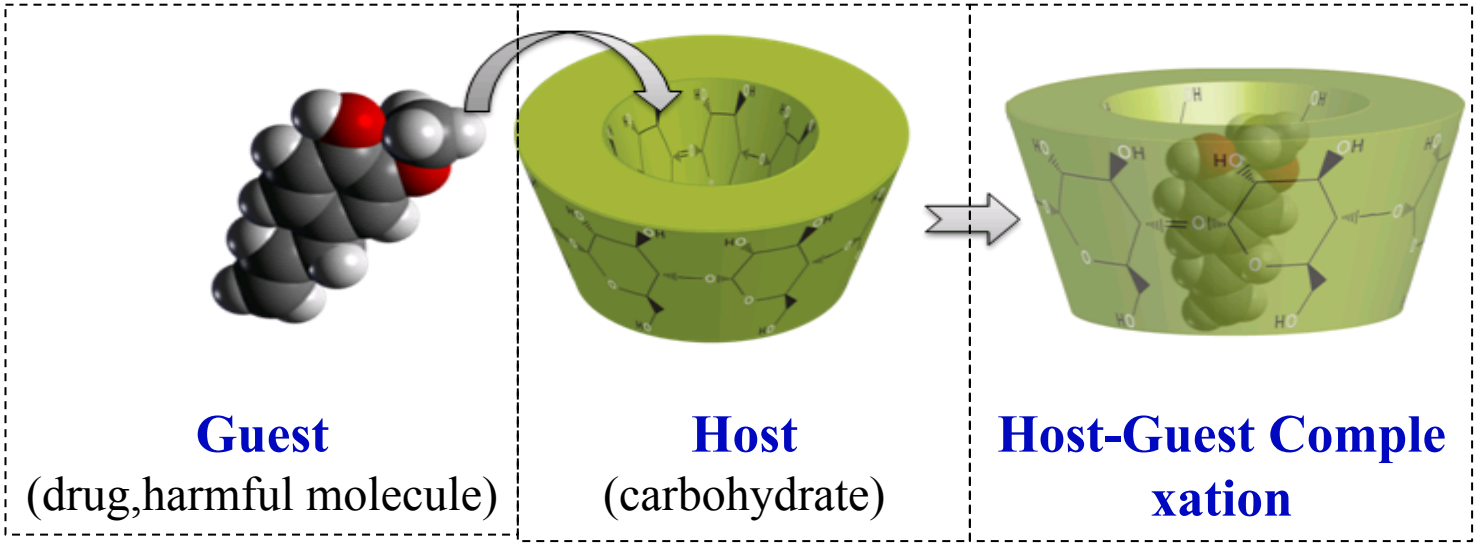
KU KONKUK UNIVERSITY

Contents

- Background
- Our Goal
- Existing Carbohydrate & Chemical Databases
- System Design
- System Implementation

Carbohydrate Complexation

- Complexation
  - A chemical process where a host molecule and a guest molecule recognize each other, interact with each other, and form a complex by non-covalent bonding



Major Applications

- **Improve solubility of substances**
- **Remove hazardous molecules**
- Stabilize of light or oxygen-sensitive substances
- Modify of the chemical reactivity of guest molecules
- Modify of liquid substances
- Mask of ill smell and taste

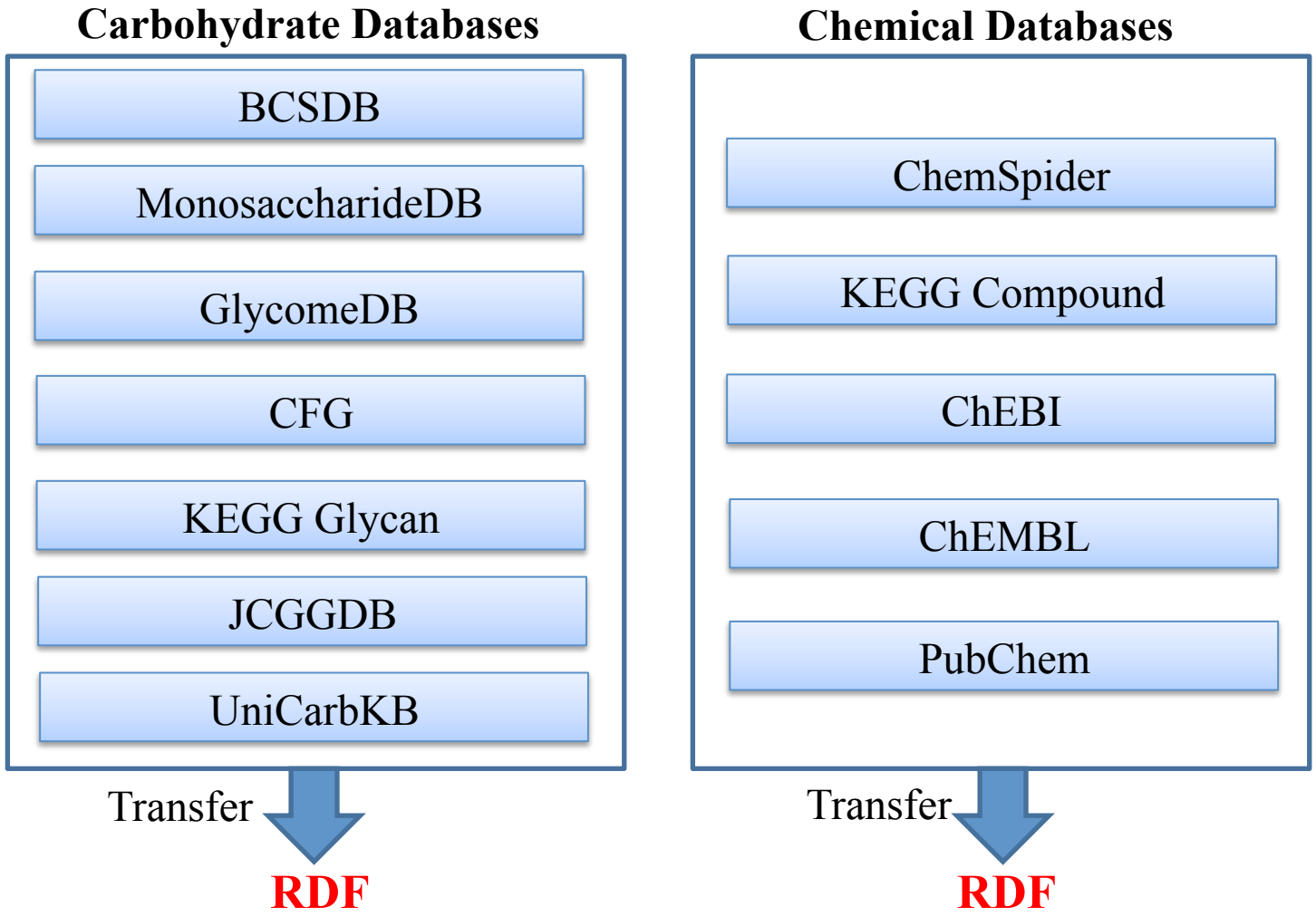
Challenging Issues

- **Guest:** a large number of molecules
  - Ex: PubChem (https://pubchem.ncbi.nlm.nih.gov/). Contains 54 million compo und entries
- **Host(carbohydrate):** complicated, dynamic and modifiable structures
  - Potentially, a number of new variants
- **A single complexation experiment takes a few days**, but finding effective hosts requires the comparative evaluation of many complexation experime nts

Our Goal

- Build carbohydrate based host-guest complexation database
- Recommend candidate hosts for a specific guest
  - Study common patterns in previous complexation results
    - Structure based
    - Property based

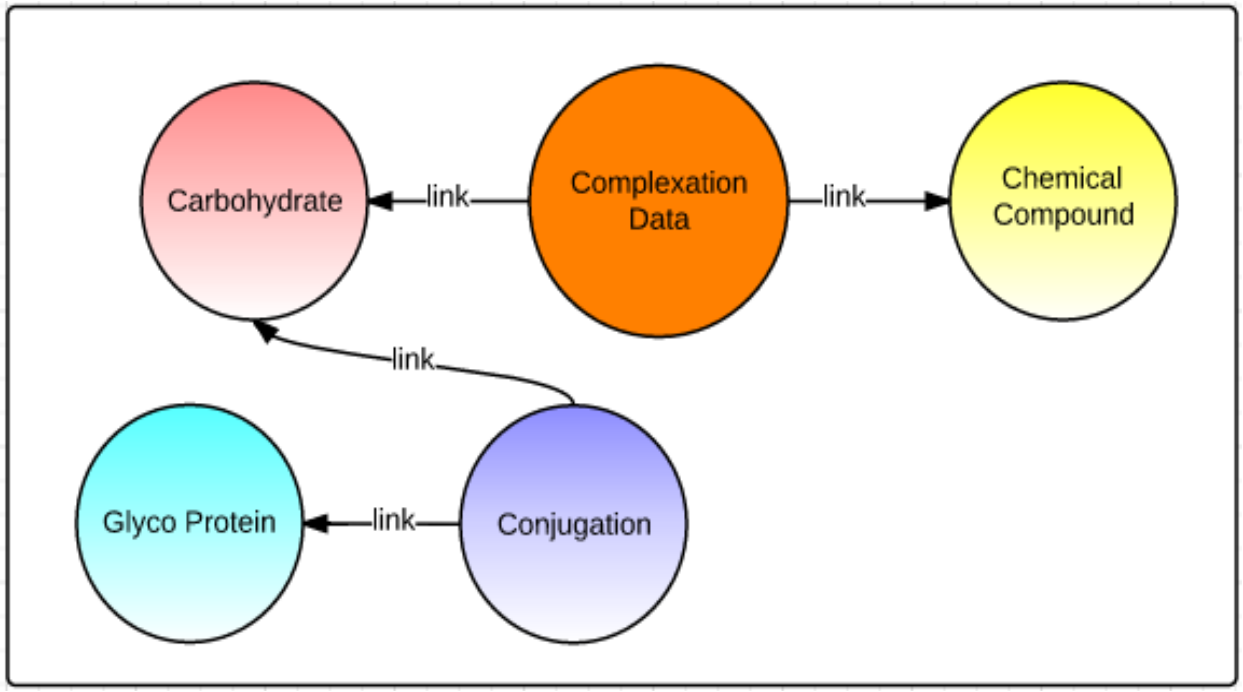
Current Carbohydrate & Chemical Data bases



Developed standard ontology: *GlycoRDF*

Approach – Linked Data

- Linked Data
  - Connect distributed data across the web
  - Machine readable data



Carbohydrate Complexation Database

Host DB

- Data from PubChem
  - Property info
  - Structure info
- Links to Monosaccharide DB
- GlycoRDF Data Model

Guest DB

- Data from PubChem
  - Property info
  - Structure info

Monosaccharide DB

- Structure info
  - GlycoRDF Monosaccharide Data Model

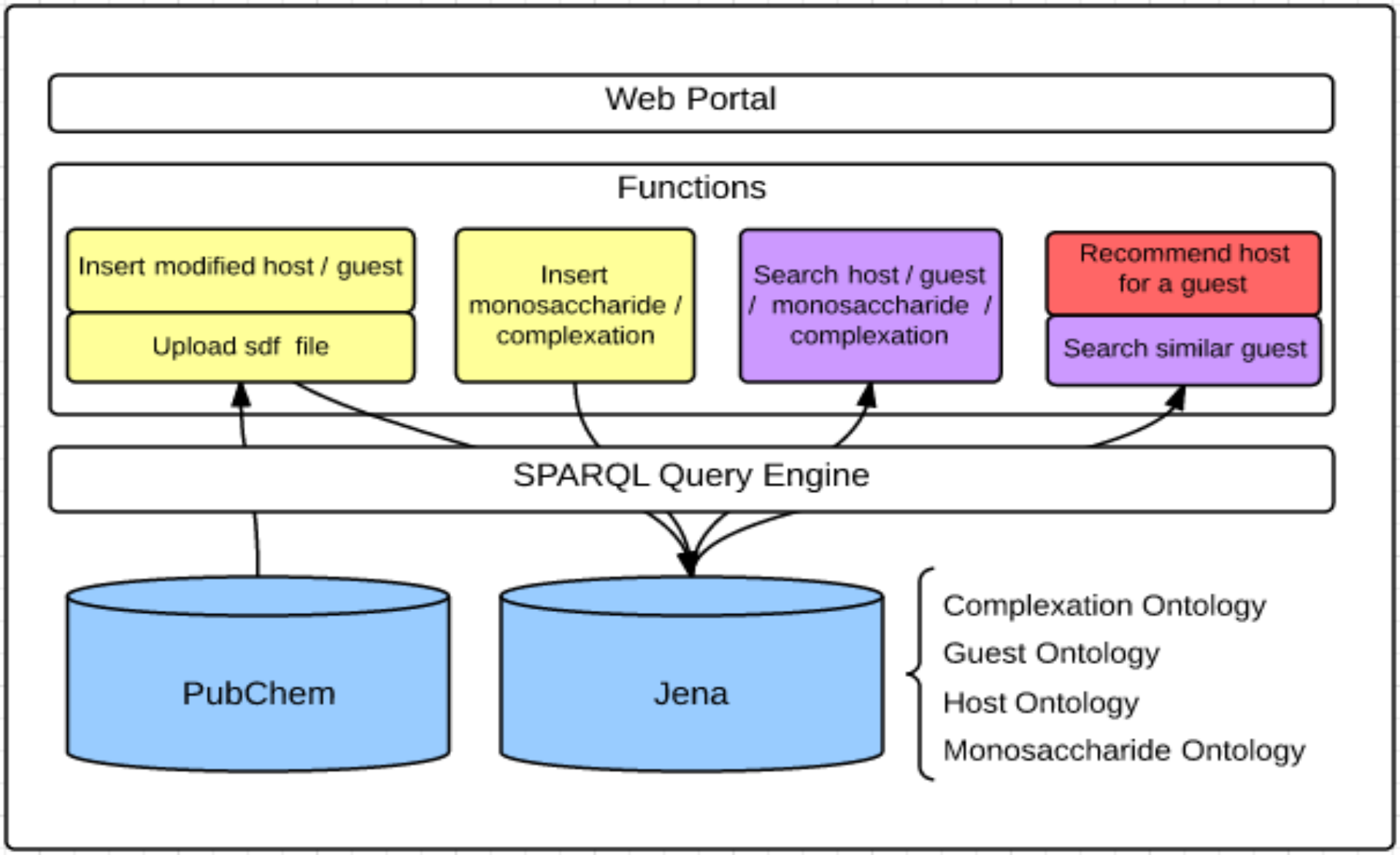
Complexation DB

- Data from experiments and publication
  - Experimental info
  - Binding result info

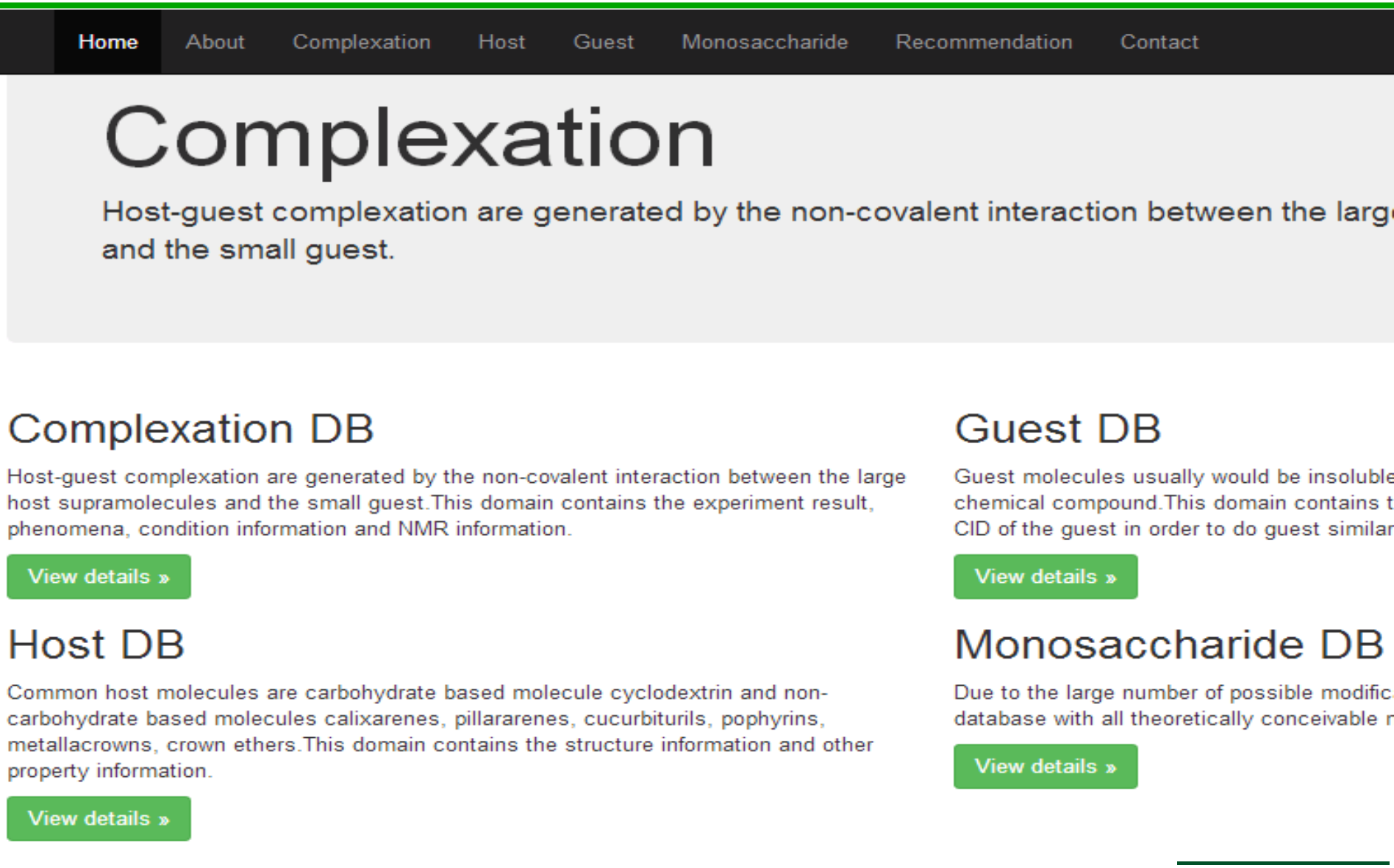
Host Recommendation

- Guest structure-based similarity search
  - Steps
    1. Compare the structure between the guests and the specific molecule
    2. Get hosts complex with all similar structure guests
  - Structure comparation using Tanimoto Coefficient
    - The most popular similarity measure for comparing chemical structures represented by means of fingerprints

System Design



Implementation: Web Portal



Search Complexation

Home | About | Complexation | Host | Guest | Monosaccharide | Recommendation | Contact

Complexation DB

Search | Insert

Search

Binding Result:

Host:	beta-cyclodextrin	Guest:	flurbiprofen
Binding Constant:	2483.8	Stoichiometry Binding	1:1

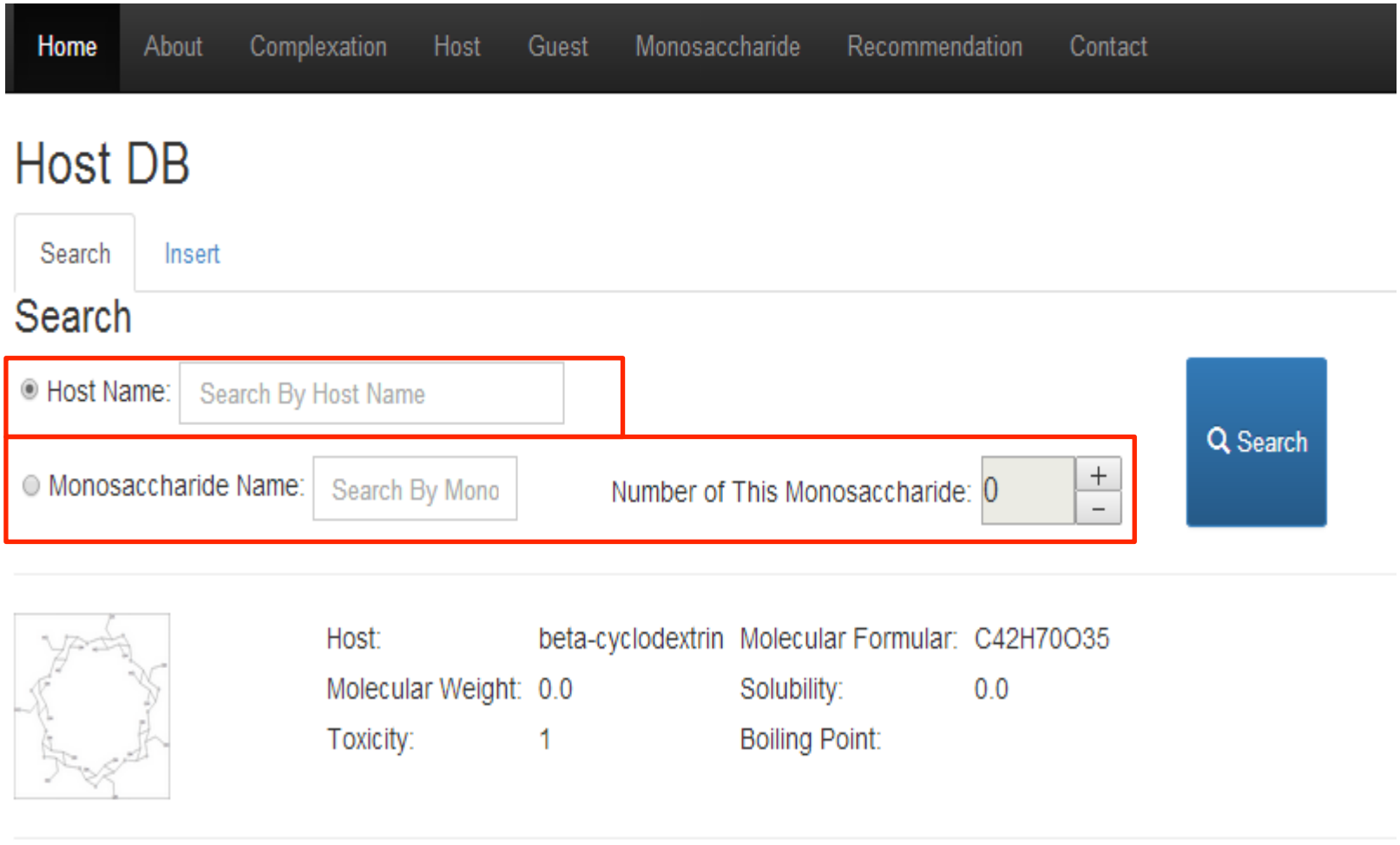
Experiment Condition:

Temperature	25.0	Solvent:	water
PH:	7.5	Mixing Time:	24.0

NMR evidence:

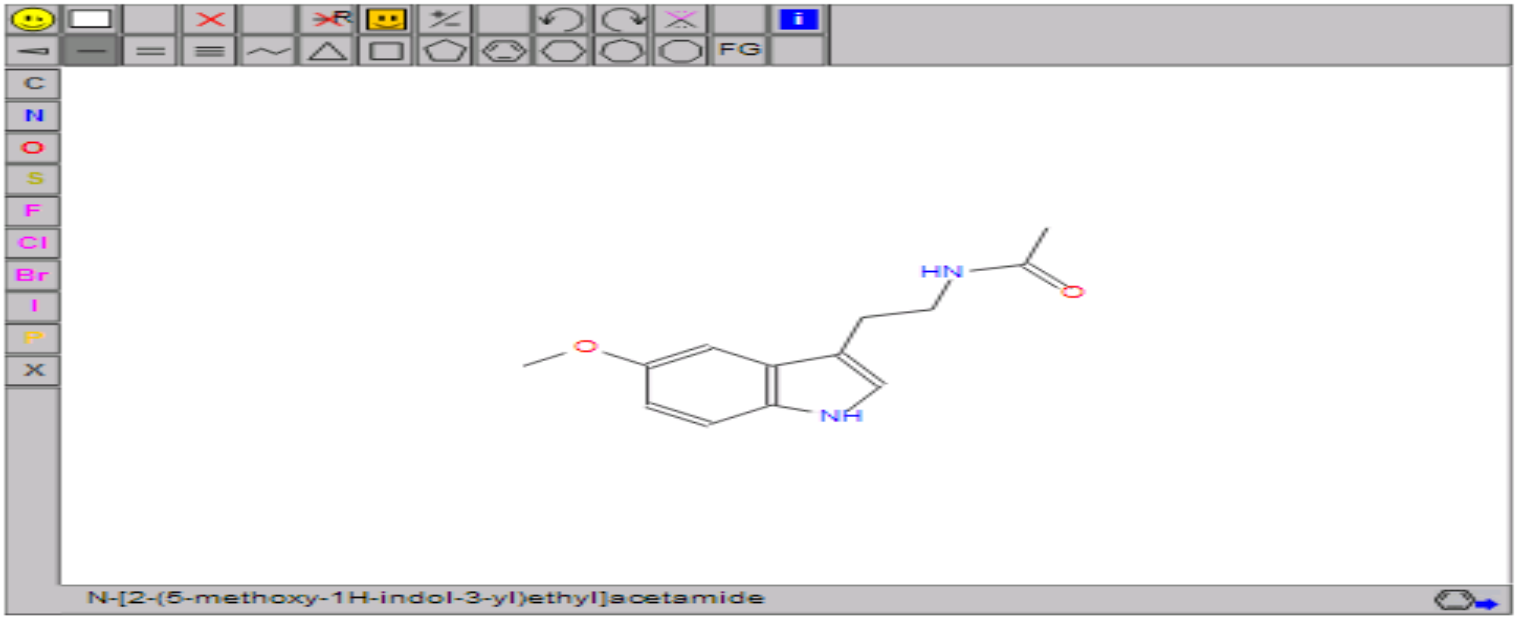
Host Concentration:	2.5	Guest Concentration:	2.5
Buffer:	d2o	Frequency:	4000.0
Temperature:	25.0	Mixing Time:	0.0
PH:	7.5		
Spectrum:			

Search Host


Home
About
Complexation
Host
Guest
Monosaccharide
Recommendation
Contact
Host DB
Search
Insert
Search
Host Name:
Search By Host Name
Monosaccharide Name:
Search By Mono
Number of This Monosaccharide:
0
+
−
Search
Host: beta-cyclodextrin
Molecular Formular: C42H70O35
Molecular Weight: 0.0
Solubility: 0.0
Toxicity: 1
Boiling Point:

Recommendation By Guest Structure-based S imilarity Search

1. Search Similar Structure By JSME Applet



JME Help

2. Search Similar Structure By Mol File

Similar Structure Guest Search

Found similar compound, and the complexation information:

Tanimoto: 100.0%

Guest:	melatonin	PubChem CID:	896
Molecular Weight:	232.27834	Molecular Formular:	C13H16N2O2
IUPAC:	N-[2-(5-methoxy-1H-indol-3-yl)ethyl]acetamide		

Development Environment

- Database
  - Jena Triple Storage
- Required Tool
  - Checkmol/Matchmol
  - Mol2ps
  - JSME
- Development Language
  - JSP
- Development Tool
  - Eclipse
- OS
  - Windows

Conclusion

- Carbohydrate complexation
  - Applicable for industrial applications
- Challenges
  - Finding effective hosts by experiments requires lots of time and efforts
- Carbohydrate complexation database
  - Informatics based approach to finding effective hosts
  - Use the linked data technology

Current & Future Work

- Collect data from previous experiments and publications
- Study data mining techniques for common patterns in complex ation data