Recent Advances in Computer-Aided Interpretation of 2D Representations of 3D Molecules

Karl Nedwed, Gregory M. Banik, Ph.D., Ty Abshear, Keith Kunitsky & Michelle D'Souza, Ph.D.

Bio-Rad Laboratories, Inc. Informatics Division

August 16, 2017 – InChI Workshop @ NIH, Bethesda, MD

BIO-RAD

Many Sources, Many Chemists, Many Structures

BIO-RAD

SADTLER RESEARCH LABORATORIES INC

WILEY

NIST

universität wien

JASCO

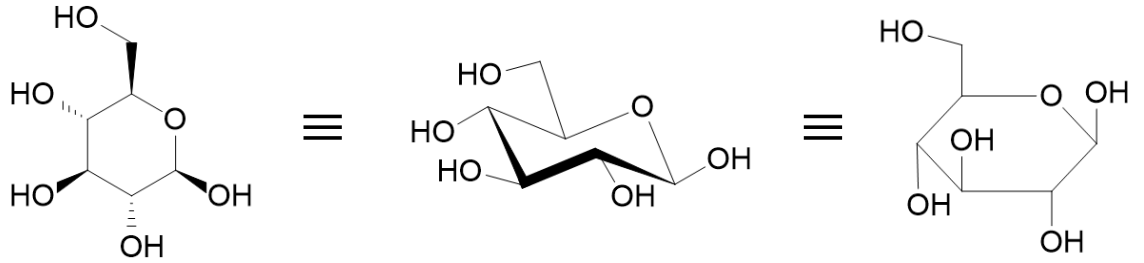
HORIBA

Health Canada

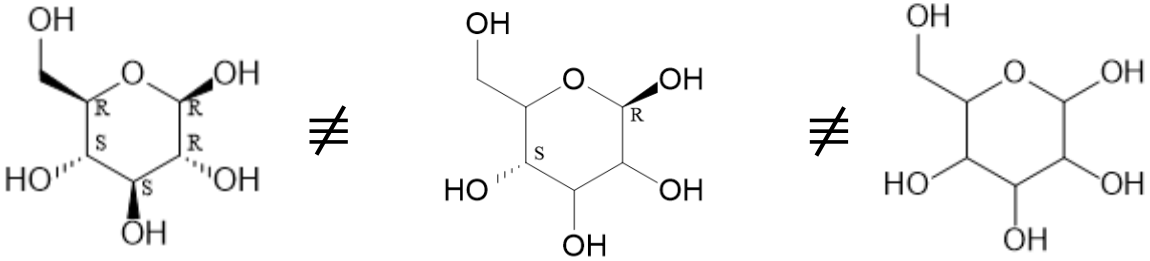
GBI STATE OF GEORGIA BUREAU OF INVESTIGATION 1937

Option (A) for Merging Structures: Standardization

Original



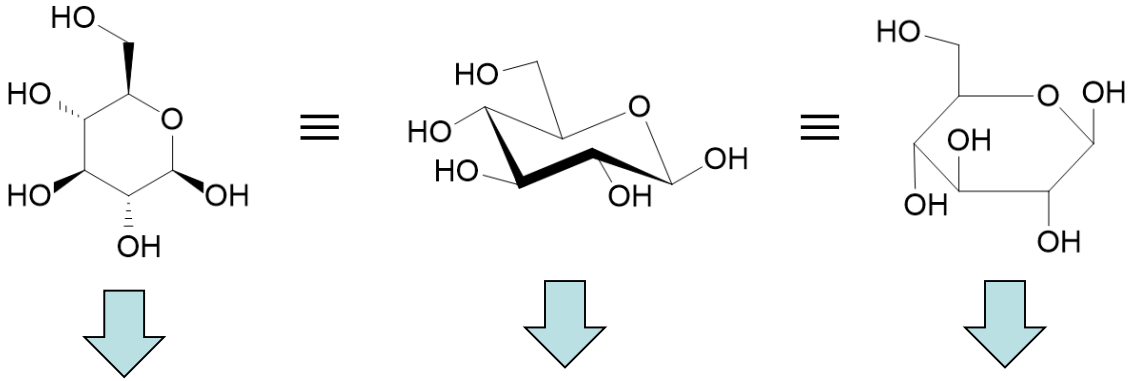
After



BIO-RAD

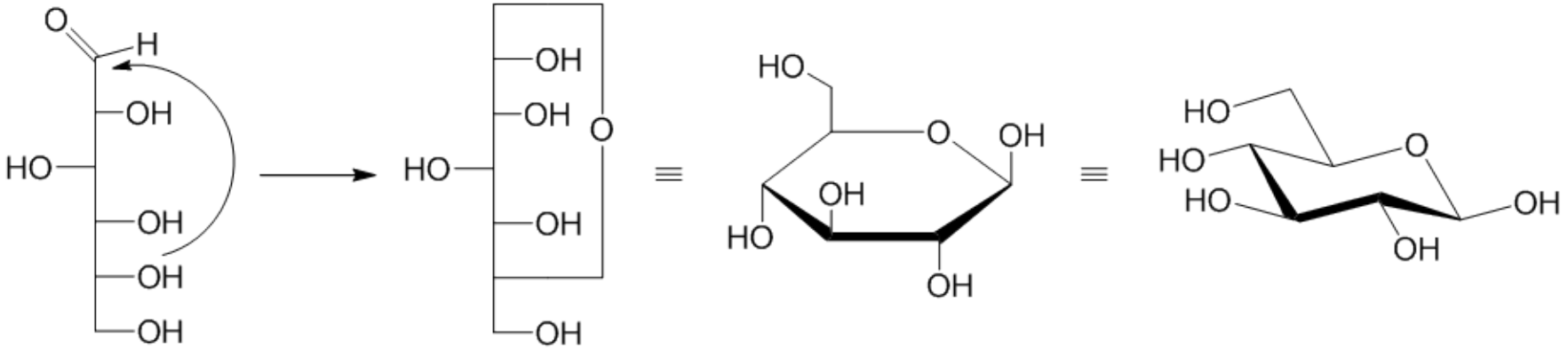
Option (B) for Merging Structures: Interpretation

Original



There Is No After

Wikipedia Page for “Pyranose”



Fischer Projections

Haworth Projection

Chair Representation

Source of Image: https://https://en.wikipedia.org/wiki/Pyranose - Accessed July 10, 2017.

BIO-RAD

Volume 95 Issue 21 | p. 7 | News of The Week
Issue Date: May 22, 2017 | Web Date: May 18, 2017

New role in cells suggested for ATP

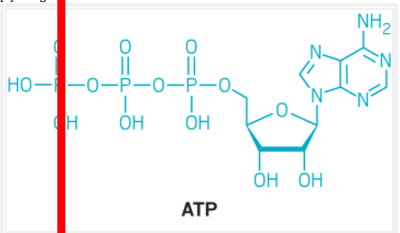
Known as an energy carrier, molecule can also solubilize proteins

By ***Jyllian Kemsley***

Adenosine triphosphate (ATP) performs many jobs in a cell. It carries energy, serves as a signaling molecule, and is the source of adenosine in DNA and RNA.

But cells contain far more ATP—as much as 5 mM in the cytoplasm—than these known uses seem to require. That might be because ATP also can solubilize proteins, suggests a new study (*Science* 2017, DOI: **10.1126/science.aaf6846**).

[+]Enlarge



ATP has the general characteristics of a hydrotrope, an amphiphilic molecule that has both a hydrophilic and a hydrophobic component but does not assemble into structures such as micelles. Hydrotropes are used industrially to solubilize hydrophobic species in aqueous solution. The hydrophobic portion of hydrotropes—such as ATP's adenosine—likely

Haworth Projection

Volume 95 Issue 26 | News of The Week
Issue Date: June 26, 2017 | Web Date: June 21, 2017

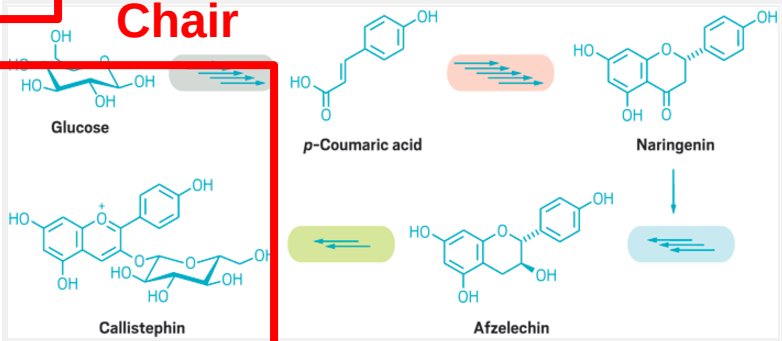
Engineered microbes make 'natural' colorants

Four engineered bacteria cooperate to synthesize possible replacements for artificial colorings

By *Stu Borman*

[+]Enlarge

Chair



Relay

Four bacteria (ovals) work together to biosynthesize the anthocyanin callistephin from glucose. Arrows inside bacteria are biosynthetic steps.

MOST POPULAR

Viewed | Commented | Shared

Cancer, redefined

Periodic graphics: The chemistry of frozen desserts

Global Top 50

Plants inspire exceptionally strong and elastic graphene aerogels

Arylamines made easy

*Most Viewed in the last 7 days

RELATED ARTICLES

Sisyphean Task: Changing Chemists' Drawing Habits

- Structure drawing conventions such as boat/chair cyclohexane, Fischer projections, Haworth projections, and 2.5D representations are deeply ingrained in all chemists today
- Every newly minted chemist in the world is being trained to draw structures with these conventions
- Changing the drawing habits of all current and future chemists is the definition of a Sisyphean task

Source of Image: https://commons.wikimedia.org/wiki/File:Punishment_sisyph.jpg - Accessed July 16, 2017.

BIO-RAD

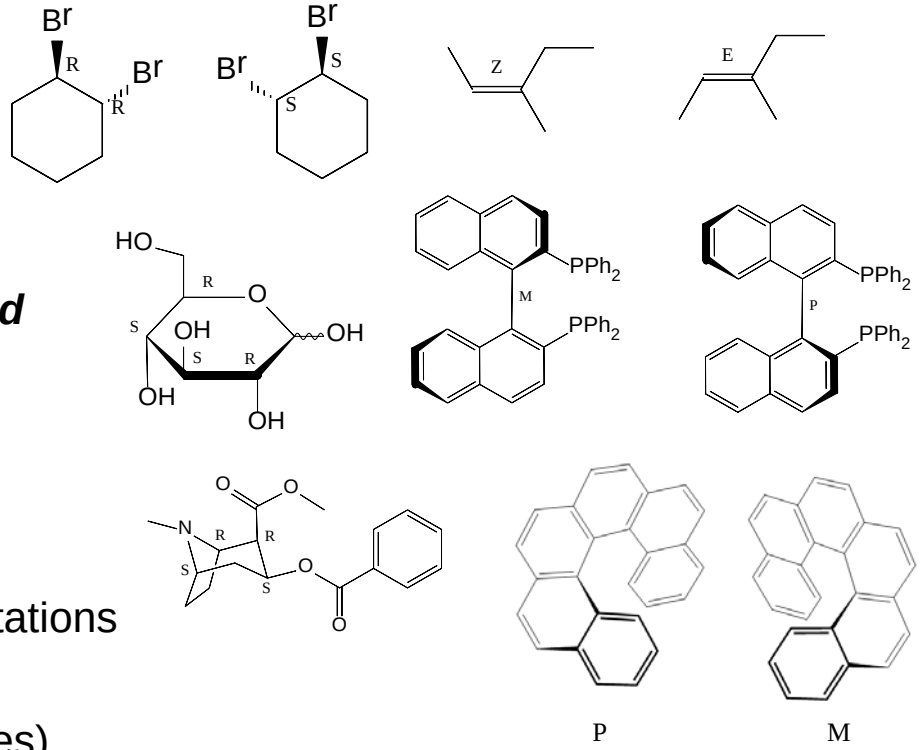
Computer Interpretation of 2D Structures

- ***The Problem:***
  - Heretofore, most chemical software packages have been unable to accurately interpret “traditional” 2D representations of 3D molecules
- ***Recent Advance:***
  - Bio-Rad’s KnowItAll® software has been reengineered to accurately interpret “traditional” 2D representations of 3D molecules by adding the ability to understand the sometimes subtle 3D intentions of 2D drawing conventions including:
    - Interposition
    - Foreshortening
    - Perspective Scaling
  - Ambiguous stereocenters are detected

BIO-RAD

KnowItAll ChemWindow – CIP Stereodescriptors

- ***Cahn-Ingold-Prelog Assignments***
  - R/S
  - E/Z
  - M/P
- ***Stereochemical Depictions Recognized***
  - Boat / Chair Representations
  - Fischer Projections
  - 2.5D (Pseudo 3D) Projections
  - Haworth Projections
  - e-Chemist's Hash/Wedge Representations
  - Helical stereochemistry
  - Stereogenic planes (e.g., cyclophanes)
  - Stereogenic axes (e.g., sterically-hindered o-substituted biphenyls)



PubChem Substances Considered for Study

200,000,000	Ending PubChem Substance ID analyzed
92,207,320	Unused PubChem Substance IDs
107,792,680	Used PubChem Substance IDs
665,837	PubChem Substances with no structure
15,559,322	PubChem Substances with connection tables but no coordinates
27,781,844	PubChem Substances with structures already normalized
194,249	PubChem Substances that failed to generate an InChI
63,591,428	PubChem Substances with valid, un-normalized structures
57,556,031	PubChem Substances without defined stereocenters
6,035,397	PubChem Substances with defined stereocenters

BIO-RAD

PubChem Substances Considered for Study

662,971	PubChem Substances with 2.5D perspective drawings	(10.98%)
51,038	PubChem Substances with chair/boat rings	(0.85%)
27,639	PubChem Substances with Haworth projections	(0.46%)
2,015	PubChem Substances with Fischer projections	(0.03%)

6,035,397 PubChem Substances with defined stereocenters

BIO-RAD

Interpretation Technology Validation Sets

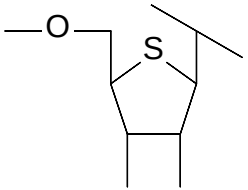
- Using a maximum diversity algorithm, 150 matched pairs of PubChem Substances and PubChem Compounds were selected for each of the following categories:
  - Haworth Projections
  - Fischer Projections
  - Chair/Boat Representations
  - 2.5D Representations
- Because the selections were diverse, they also tended to include some difficult examples…

BIO-RAD

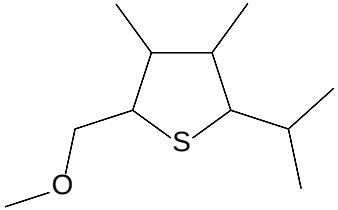
Validation Protocol

Compare InChIKeys for both representations

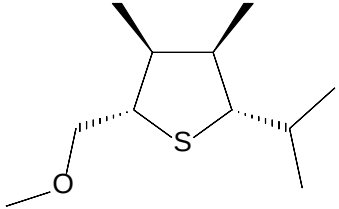
Investigate discrepancies



PubChem Substance

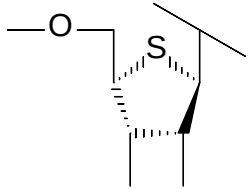


PubChem Compound (Standardized)



UYNPITADNKUEIY-ZDCRXTMVSA-N

Chemist manually adds stereobonds to standardized PubChem Compound to match the PubChem Substance and generate InChI and InChIKey



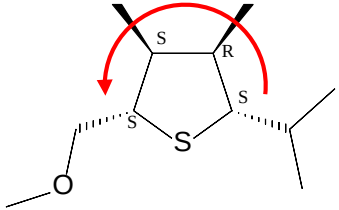
UYNPITADNKUEIY-ZDCRXTMVSA-N

KnowItAll perceives implicit stereochemistry from 2D representation and automatically adds stereobonds internally to original PubChem Substance to generate InChI and InChIKey

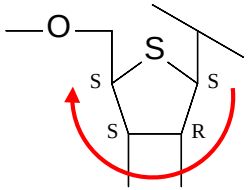
BIO-RAD

Related Topic: CIP Stereodescriptors Match

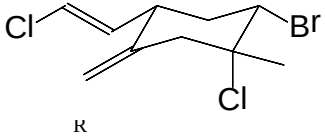
S,R,S,S



S,R,S,S

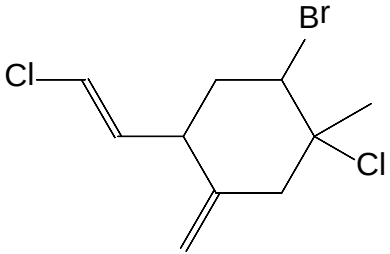


Validation Example – Cyclohexane Chair



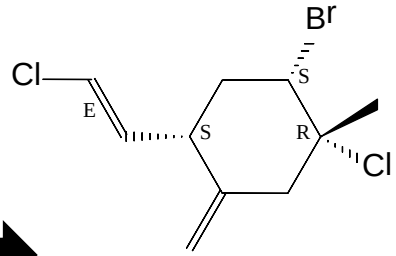
PubChem Substance (Original)

S,S,R



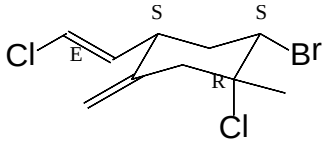
PubChem Compound (Standardized)

?,?,?



PubChem Compound (Edited)

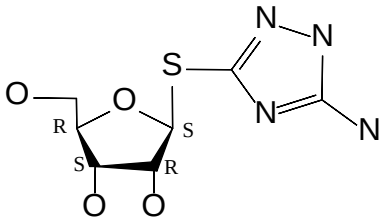
S,S,R



PubChem Substance (Interpreted)

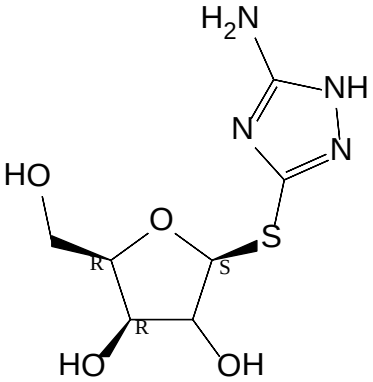
S,S,R

Validation Example – Haworth Projection with Misinformation



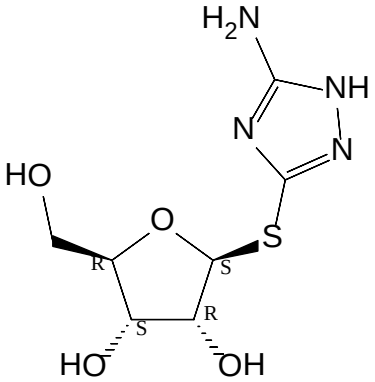
PubChem
Substance

S,R,S,R



PubChem
Compound
(Standardized)

S,?,R,R



PubChem
Compound
(Corrected)

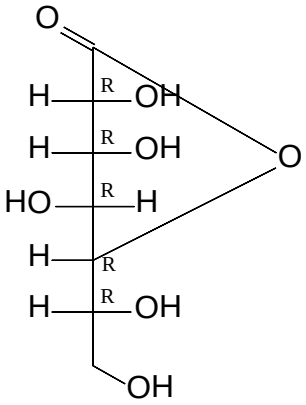
S,R,S,R

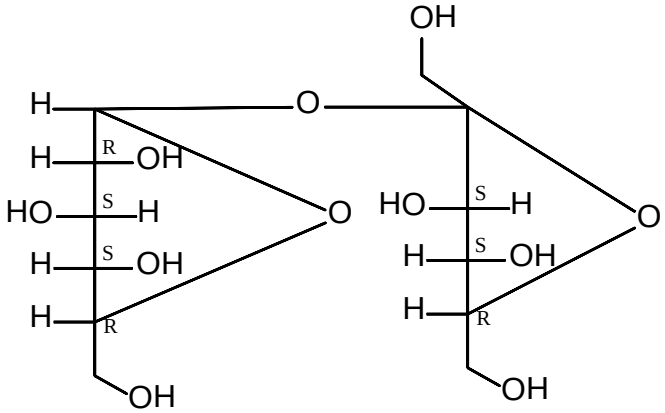
BIO-RAD

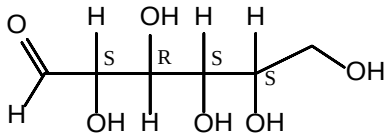
Validation Results

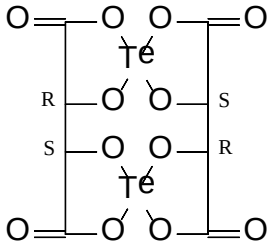
Number of records:	595	
Number of stereocenters:	1,573	
Number of matching stereocenters overall:	1,522	(96.76%)
Number of matching R/S assignments:	**1,512**	**(98.63%)**
Number of missed stereocenters:	20	(1.27%)
Number of added stereocenters:	10	(0.64%)
Number of stereocenters with reversed R/S assignments:	**21**	**(1.34%)**

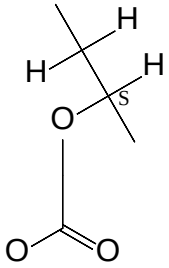
Examples: Fischer Projections

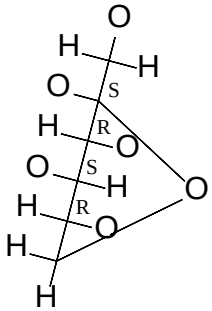


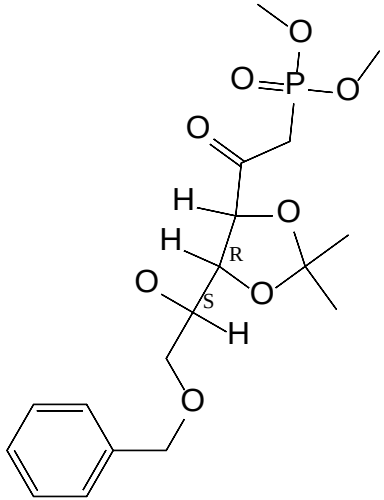




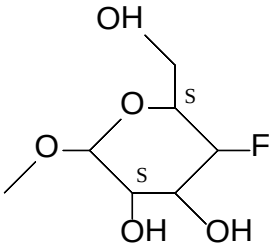


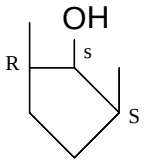


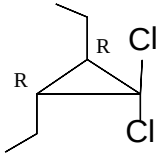


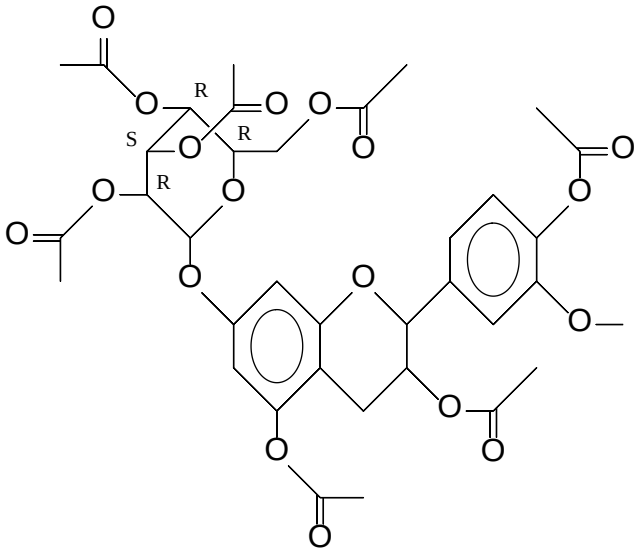


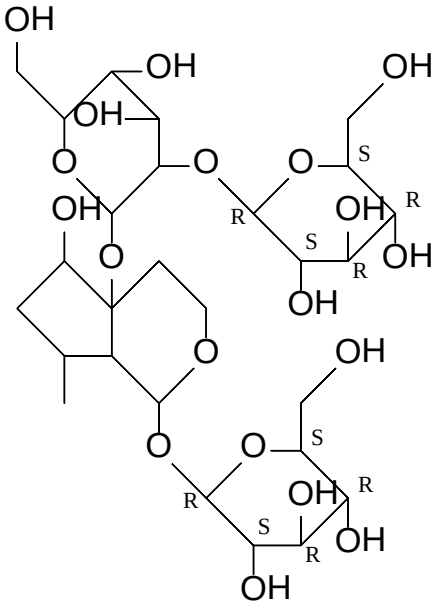
Examples: Haworth Projections



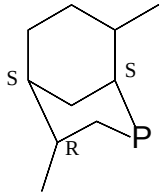


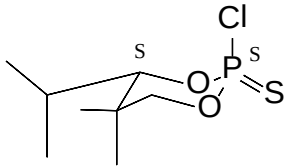


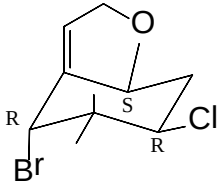


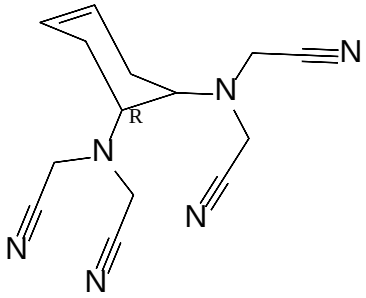


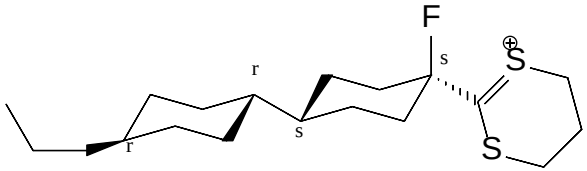
Examples: Chair Projections

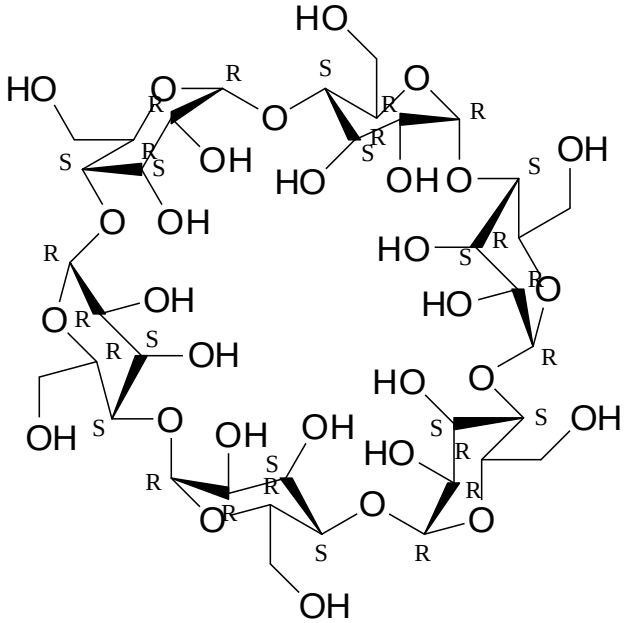




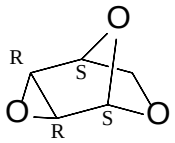


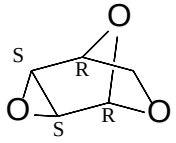


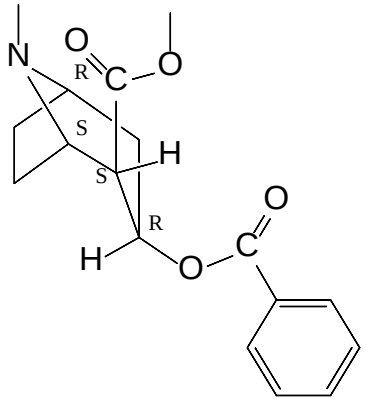


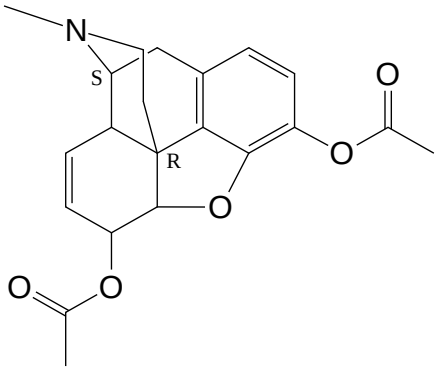


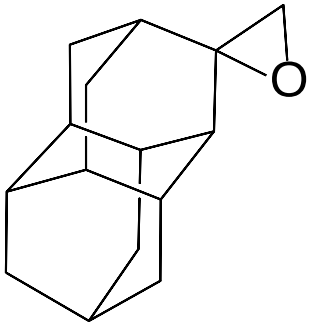
Examples: 2.5D Projections

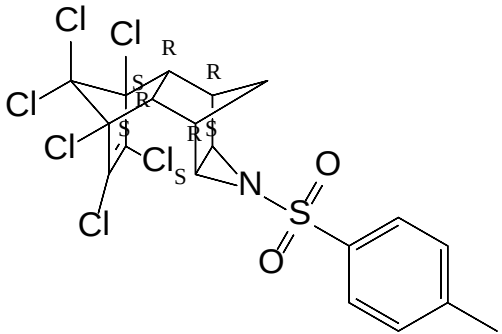


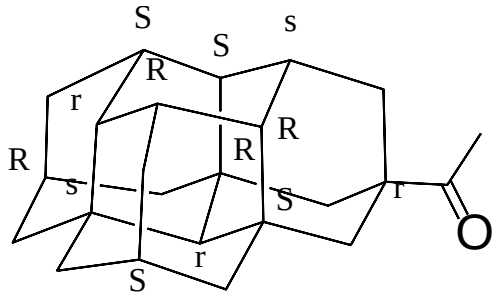












Summary

- Part art form and part language, 2D representations of chemical structures have a very long history.
- Traditional structure drawing styles like cyclohexane chairs and boats, Fischer projections, Haworth projections, and 2.5D perspective projections can be accurately interpreted by computer software, so…
- e-Chemists can end their battle against 127 years of scientific tradition, embrace the past, and live in peace and harmony with the rest of the chemical community.

BIO-RAD