CHEMINFORMATICS

KnowItAll® Stereochemistry Toolkit

A wide range of applications

We recomend this tool for the following applications to interpret structures:

- Chemistry Publishing
- Chemical Information
- Pharma Discovery
- Any chemist who builds large structure databases

Stereochemical Interpretation Solved!

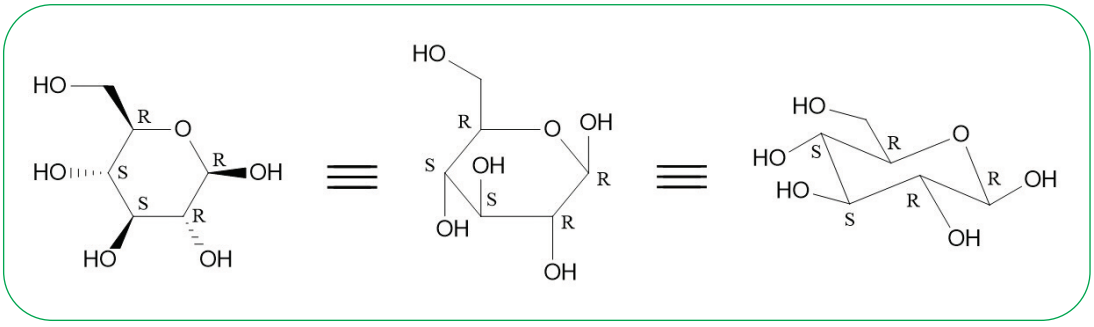
Introduction

Chemical structures are the primary means of communication between chemists and researchers. In today's ever more interconnected world, innovations such as the IUPAC International Chemical Identifier (InChI) have created an explosion of interconnectedness for the multitude of chemical structure databases. Unfortunately, for over a century, chemists have drawn chemical structures using implicit 2D information to infer 3D information (for example, chair/boat representations of cyclohexane, Haworth Projections, Fischer Projections, pseudo-3D projections, etc.). The IUPAC InChI code and most cheminformatics software on the market today do not accurately interpret this implicit 2D information, leading to incorrectly assigned InChIs and therefore incorrectly matched structures.

Bio-Rad's KnowItAll® Stereochemistry Toolkit (TK) provides a fast, efficient, and highly reliable platform for the interpretation of implicit stereochemical information in traditionally-drawn structures that allows them to be used for indexing, searching, and other forms of processing. In addition, the KnowItAll Stereochemistry TK generates accurate Cahn-Ingold-Prelog (CIP) stereodescriptors defined in the *IUPAC Nomenclature of Organic Chemistry (Blue Book)* for all descriptor classes (R/S, E/Z, M/P) of organic molecules.

Interpretation of Stereochemical Information

Most cheminformatics software interprets stereochemical information only if clearly and explicitly drawn with wedged or hashed bonds. The KnowItAll Stereochemistry TK provides an interpretation of a variety of additional drawing styles (*see list*) that allows the implicit stereochemical information of such structures to be elucidated. A frequently-used approach in the industry is to convert structures into a normalized representation with additional wedged or hashed bonds. The KnowItAll



BIO-RAD

Stereochemistry TK does not use this method for two reasons:

1. Most chemists prefer to see the structure in the drawing style as originally used by the person who drew the structure, because a normalized view may be harder to interpret with the human eye.

2. The normalization process itself is prone to introducing interpretation errors that lead to incorrectly assigned stereocenters. This is particularly true for pseudo-3D projections of complex polycyclic organic molecules.

The KnowItAll Stereochemistry TK interprets the stereochemical information in the original structure without any conversion and calculates parity information for every found stereocenter. These parity values (odd/even) can then be used as input information to other processing steps such as the calculation of an InChI or InChIKey. The toolkit provides two ways to add stereochemical information to a molecule:

1. **The IXA_MOL_HANDLE format as defined by the InChI source code.** The IXA suite of functions and objects is a convenient way to define an input molecule for the calculation of an InChI or InChIKey. By providing an IXA molecule object as input to the KnowItAll Stereochemistry TK, adding implicit stereochemical information to InChIs requires only one additional function call. The toolkit interprets the molecule defined by the IXA_MOL_HANDLE object and adds parity information to it that will be used by the InChI code to generate InChIs with proper /t, /b, /m and /s information layers.

2. **A molfile passed in as a character string.** The KnowItAll Stereochemistry TK interprets the structure defined by the molfile and inserts parity values into the corresponding atom stereo parity column within the atom block of the connection table (Ctab) section. This information can be interpreted by other software to deduce the correct stereochemical orientation of every stereocenter of the molecule.

As opposed to normalization, the process of interpretation allows chemists to keep the original chemical structures as-is without any loss of information while fully extracting all implicit and explicit stereochemical information. The resulting gain of information allows structures to be matched correctly regardless of their drawing style, which is critical for database building, searching, indexing, web publishing, etc.

The toolkit gives feedback on stereocenters that could not be interpreted because of ambiguities or errors in the drawing style. This error information can then be used to clean up structures with apparent drawing problems.

The KnowItAll Stereochemistry TK interprets the following drawing styles:

1. **Wedged, hashed, and hashed wedged bonds.**
2. **Chair/boat projections in any orientation in space.**

   Note that vertically oriented chair or boat projections are inherently ambiguous unless drawn with additional visual clues such as wedged or bold bonds.
3. **Haworth projections in any orientation in space.**

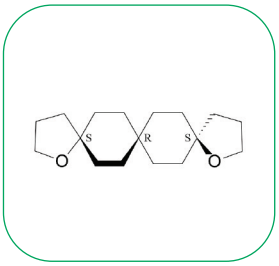
   Note that vertically oriented Haworth projections are inherently ambiguous unless drawn with additional visual clues such as wedged or bold bonds.
4. **Fischer projections in any orientation in space.**
5. **Pseudo-3D projections (2.5D projections).**

   Such projections are widely used for ring systems such as norbornanes, adamantanes, or fullerenes. Bond overlap information is used to infer the depth (implicit Z coordinate) information from these structures. Note that the correct stereochemical information can only be deduced if the bonds in the input structure are defined in the correct order.

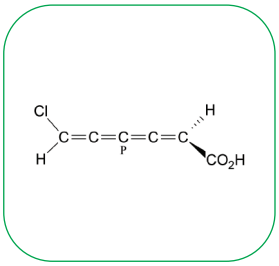
Stereochemical CIP Descriptors

In addition to the interpretation of stereochemical information, which is not explicitly visible to users looking at structures in a browser, database, etc., the KnowItAll Stereochemistry TK also provides a function to assign Cahn-Ingold-Prelog stereochemical descriptors to organic molecules. These descriptors can be used as feedback and helpful additional information to users.

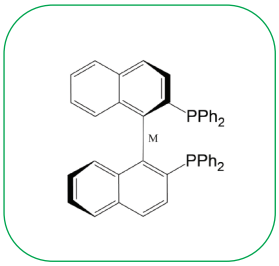
The toolkit supports all stereochemical descriptors of organic structures as defined by the *IUPAC Nomenclature of Organic Chemistry (Blue Book)*, 2013 Edition. These descriptors include:



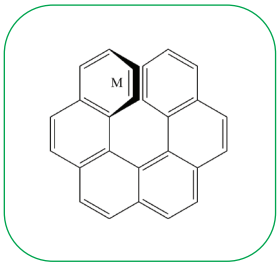
Tetrahedral chirality centers (R/S), including spiro compounds



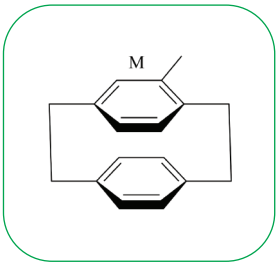
Chirality axes such as those found in allenes and cumulenes with an even number of double bonds (M/P)



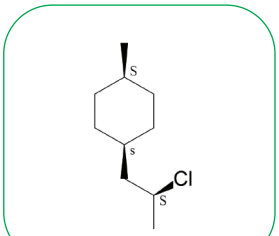
Stereogenic axes such as found in some o-substituted biphenyls (M/P)



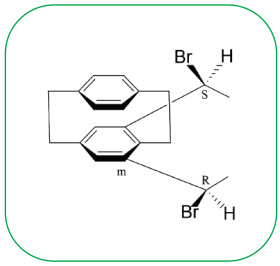
Helical stereogenic axes (M/P)



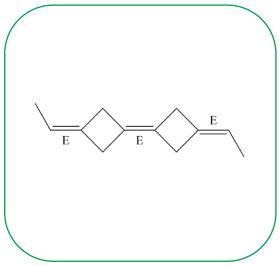
Chirality planes such as found in substituted cyclophanes (M/P)



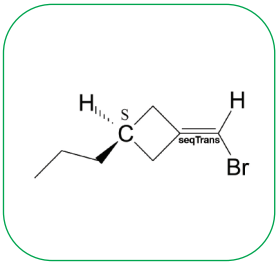
Pseudochiral (pseudoasymmetric) centers (r/s)



Pseudochiral (pseudoasymmetric) centers (m/p)



Cis/trans isomers as found for double bonds and cumulenes with an odd number of double bonds (Z/E or seqcis/seqtrans)



Enantiomorphic *cis/trans* isomers as found for double bonds and cumulenes with an odd number of double bonds (seqCis/seqTrans)

BIO-RAD

Bio-Rad Laboratories, Inc.

Informatics Division **Website:** www.knowitall.com/stereochem **Contact Us:** www.knowitall.com/contactus