

2.3: Molecular Graph Issues

Learning Objectives:

- Identify issues with representing molecules as molecular graphs
- Review chemical principles that can provide issues with structural data

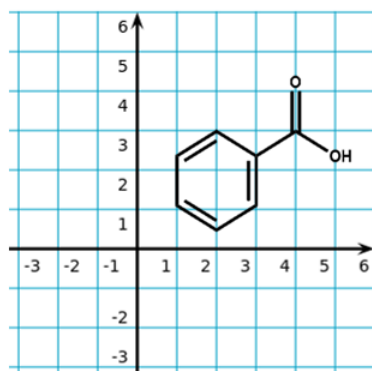
Introduction

This section is intended to be a review for chemistry students, while also assisting non-chemistry majors in identifying some of the complexities of chemistry that come into play when describing molecular structures. Even chemistry students are advised to quickly skim through this material, as we are parsing common topics from the perspective of a Simple Connection Table (SCT), which may give them a deeper understanding of the nuances of chemical structure representation.

As many of these issues require features beyond those of the simple connection table, many of these issues will be revisited when we look at real data files in section [2.5 Structural Data Files](#).

Atom Coordinates

Note that the SCT atom table does not tell you anything about the relative position of atoms. (As we have seen, you often have to go to the bond table just to figure out which atom is which.) Many connection table formats contain two- or three-dimensional spatial coordinates for each atom entry. These coordinates may simply record the relative position of atoms in a structural formula sketched in a chemical drawing program (*SCT XII*).



SCT XII			
Atom table			
Atom #	Atom type	X	Y
1	C	3	2.5
2	C	2	3
3	C	1	2.5
4	C	1	1
5	C	2	0.5
6	C	3	1
7	C	4	3
8	O	5	2.5
9	O	4	4
Bond table			
Atom 1	Atom 2	Bond order	
1	2	1	
2	3	2	
3	4	1	
4	5	2	
5	6	1	
6	1	2	
1	7	1	
7	8	1	
7	9	2	

Figure 2.3.1: The addition of two dimensional coordinates added to a connection table based on an external coordinate system, a 3D system would have an additional column for Z values.

We will address atom coordinates in section 2.5 when we look at real structural data files.

Stereochemistry

Isomers are different molecules with the same atomic constituency, that is they have the same number of atoms for each element and the atom tables are essentially identical, (the numbering of the atoms may be different, but the two atom tables are isomorphic). There are two basic types of isomers, constitutional isomers and stereoisomers. Constitutional isomers are also called structural isomers and have different bond connectivity for the same atoms. This means they have different (non-isomorphic) bond tables,

and so the Simple Connection Table (SCT) has no problem distinguishing constitutional isomers. Stereoisomers have the same (isomorphic) SCTs, that is, both the atom and the bond table are essentially the same, (the atom numbering may differ, but this is reflected in the bond connections and so the SCTs are essentially the same (isomorphic). What distinguishes the atoms of stereochemical isomers is the atomic arrangement in space, not the connections.

You may ask why is this important? One example often used in textbooks is the biological significance of two stereoisomers of thalidomide, a chemical used as an antidepressant for pregnant mothers in the 1960s. In synthesizing the chemical the "drug" was actually a mixture of both isomers, one of which was an effective medication and the other of which caused horrific birth defects. This was clearly an "unintended consequence" and one of the most important functions of cheminformatics is to help scientists identify unintended effects of potential drugs by looking at a multiplicity of bioassays, including toxicological screening assays.

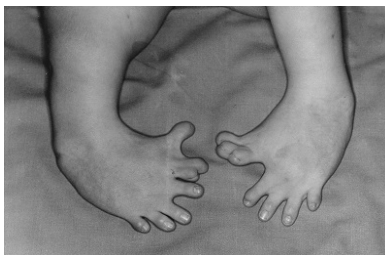


Figure 2.3.2: Birth defect caused by the mutagenic isotope of thalidomide, which was prescribed by the mother's doctor.

Isomer Review

First it is prudent to review isomers. Figure 2.3.3 gives an flowchart of the major type of isomers chemists deal with. This section is dealing with the issues of representing stereochemistry in simplified connection tables and so will focus on that. This section will also introduce conformational isomers, which are different pseudo-stable orientations around rotatable bonds and so are not really separate molecules but can be of interest to cheminformatics because the different conformations can be favored in different environments, and exhibit different chemical behaviors (the most probable conformation in a protein environment may be different than in free space).

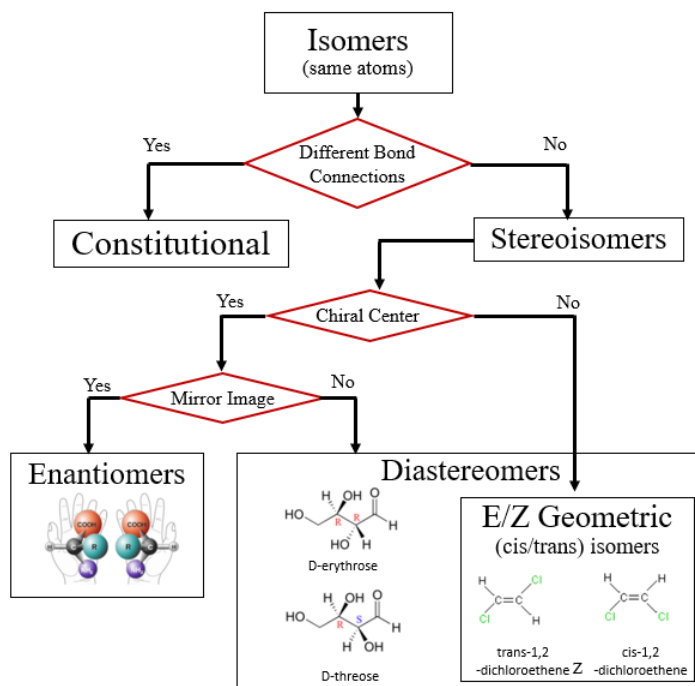
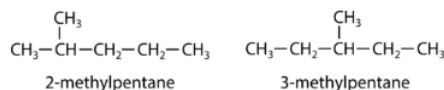


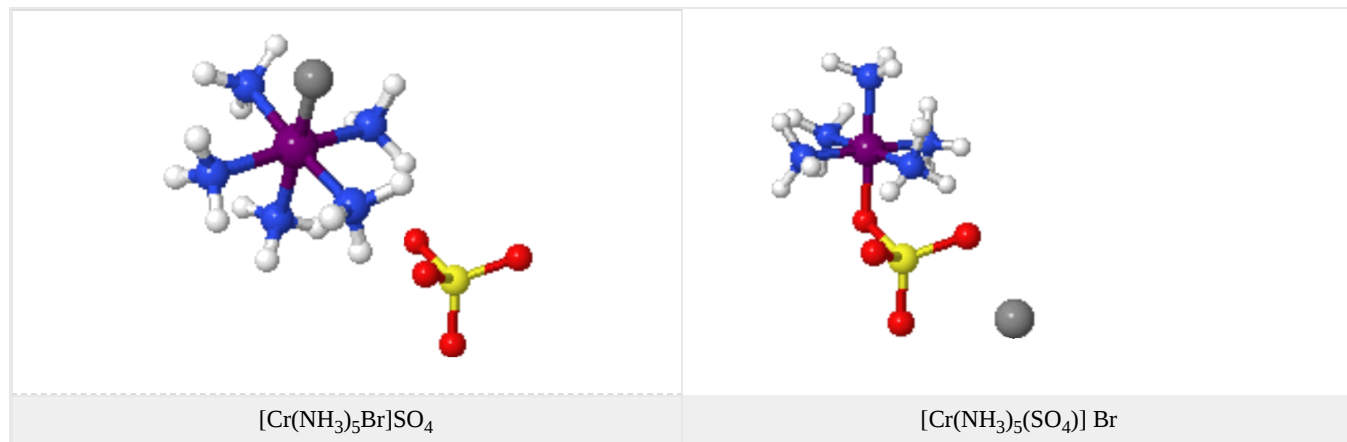
Figure 2.3.3: Flowchart of the common types of isomers that

Constitutional Isomers

Constitutional or structural isomers are molecules composed of the same atoms but with different bonds (connectivity). There are two subsets to constitutional isomers, linkage isomers and ionization sphere isomers. Linkage isomers are simple ones within a molecule where the bond links are different, and so they can not have the same bond table.



Ionization sphere isomers are multi-covalent unit coordination complexes (salts) where the neutral salt that is composed of a positive cation and a negative anion have the same formula, but in the two isomers the anions exchange with a [ligand](#). These types of species present challenges for connection tables and these issues will be approached in the multicovalent unit and coordination complex sections of this chapter.

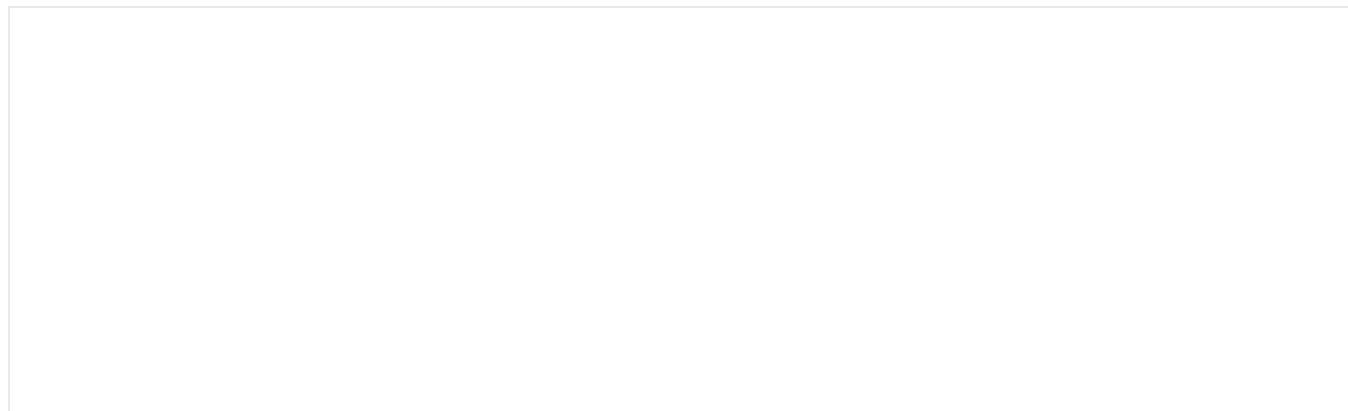


2.3.4: On the left the bromine ligand (grey) is bonded to the Chromium of the pentaamminebromoChromium(III) cation and the sulfate is the anion, while on the right the bromine is the anion and the "sulfate" is the ligand of the pentaamminebromosulfato(III) cation.

Conformers

These are often called conformational isomers but they do not represent true isomers, but are often of great importance in cheminformatic investigations. If you have a single bond a molecule can freely rotate, and as it rotates the groups attached to the rotating atom change their geometric positions and thus change their interaction energies both with respect to other atoms in the molecule, but also with respect to atoms in their environment. There are approaching an infinity of conformations a molecule posses and these can be understood by looking at a simple hydrocarbon, ethane (CH_3CH_3) where there is free rotation around the C-C bond. Ethane has two extreme conformations, the eclipsed and the staggered which can be visualized by the [Newman projection](#) (right of figure) where you are looking down the C-C bond axis.

These are not really isomers in that it is the same molecule and the following animated gif shows the potential energy changing for an isolated molecule of ethane as it goes through these rotations.



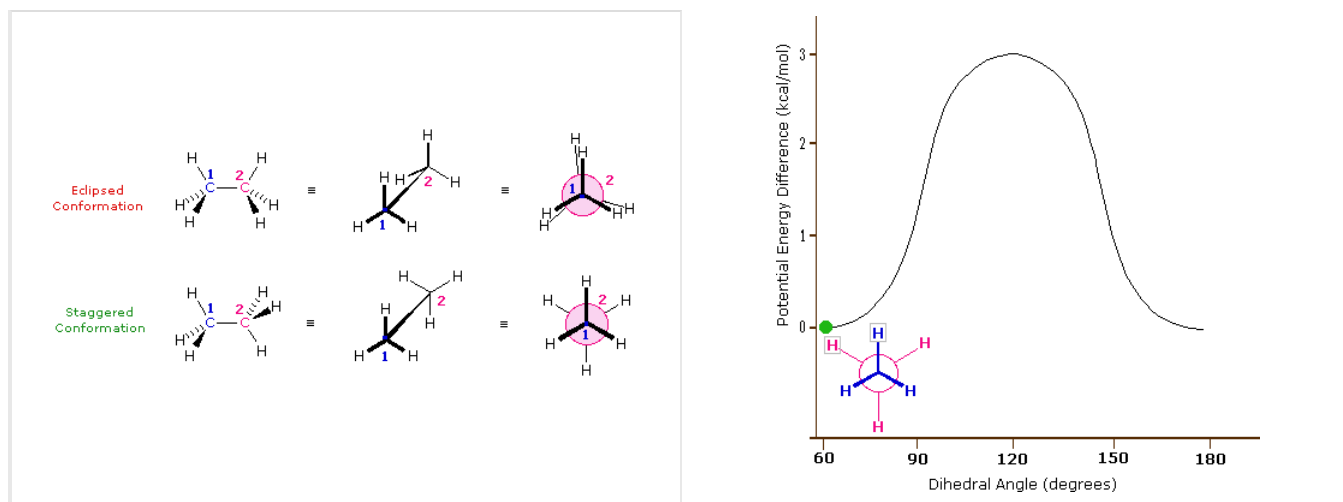


Figure 2.3.5: Perspective and Newman drawing of the two extreme conformations of ethane (left) and an animated gif showing the potential energy change as a result of rotation around the C-C bond (right).

What is important to recognize is that a connection table will often give the coordinates of the atoms in a molecule, and in reality the coordinates change as the molecule transitions from one conformer to another. But also, the above diagram represents a simple isolated molecule, and in real systems, molecules are often in protein environments that will define the most stable state, and also the reactivity of a molecule will depend on its conformational state. So to truly represent a molecule on computer one needs to take into account a multiplicity of conformational states, all of which can be represented in a connection table.

Stereoisomers

Stereoisomers have isomorphic simplified connection tables (they have the same connectivity) but differ in the arrangement in space, and so need additional information to distinguish them. There are two basic types of stereoisomers, enantiomers and diastereoisomers. If two nonsuperimposable stereoisomers are mirror images of each other they are enantiomers, and if two stereoisomers are not mirror images of each other they are diastereoisomers.

Enantiomers

Enantiomers are chiral molecules, that is if you invert the molecule through a mirror plan it is not superimposable on its mirror image. If it is superimposable, then it is the same molecule as its mirror image and they are not isomers. The term chirality comes from the Greek word for handedness, and your hand is a chiral structure. If a carbon atom in an organic molecule is attached to four different substituents it is a chiral atom. Bromochlorofluoromethane has a nonsuperimposable mirror image across the chiral carbon and so is a chiral molecule, while dichlorofluoromethane is superimposable on its mirror image, and therefore is not an isomer with its mirror image.

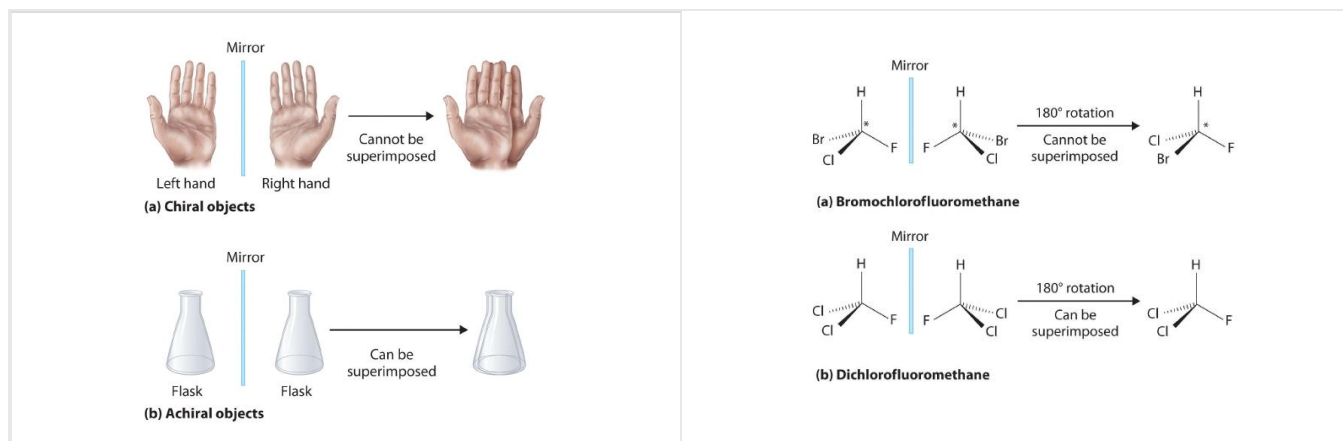


Figure 2.3.6: The top objects are chiral and the bottom are not. These means that there are two different types of bromochlorofluoromethane (the "left" and the "right" handed ones), while there is only one type of dichlorofluoromethane (because it is superimposable on its mirror image, and so it is the same molecule as its mirror image).

Chirality is a function of structure and so there are chiral centers (the plane going through the central carbon, hydrogen and fluorine in the above example was one such center of inversion). These can get very complicated very quickly, but one thing to note is that in an organic molecule a carbon that is bonded to four other atoms is chiral if none of the groups bonded to it are identical (note in the bottom part of figure 2.3.6 two of the groups are identical (Cl), and so it was not chiral).

Lets look at thalidomide, the molecule responsible for the birth defects in figure 2.3.2. Figure 2.3.7 shows four ways of drawing this structure.

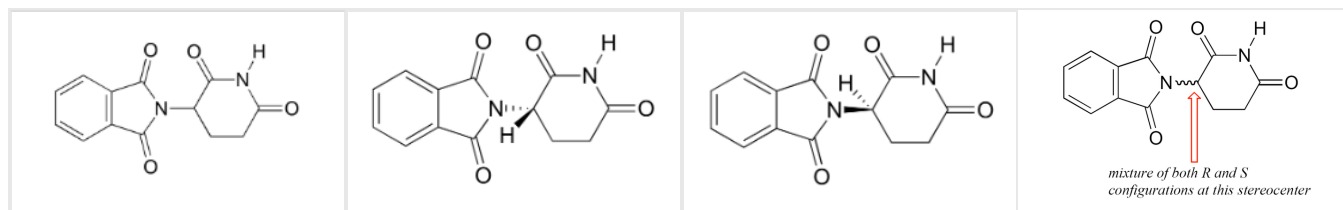


Figure 2.3.7: Four ways of drawing thalidomide, each means something different. Sequentially (left to right) these are; undefined, left-hand (S), right-hand(R) and a mixture. See ICP rules (below) to understand R and S notation.

Figure 2.3.7 represents a major challenge in cheminformatics in that authors often do not define the stereochemistry of molecules when they submit structures in their publications (left most image) and this can lead to issues when data is abstracted from the literature.

Before going into how we identify chirality we should first take a look at the "Simplified Connection Table" (SCT) to understand the issue with respect to representing chiral molecules on computer.

Chirality and Connection Tables

VERY IMPORTANT: In this class we will be using chemical compound databases to retrieve and store chemical information and in the following two figures you see three connection tables for theses stereoisomers. As pointed out in figure 2.3.7, when someone measures and uploads data concerning a chiral compound there are actually four possible identities, the stereo chemistry is not defined (left most image), it is defined (middle and right images) or it is a mixture (right image).

Lets look at the connection table of the chiral compound 2-butanol (figure 2.3.8. If you look at the SCT, they are all the same. That is, we need to add additional information, and when we get to actual chemical structure files we will look deeper into this.

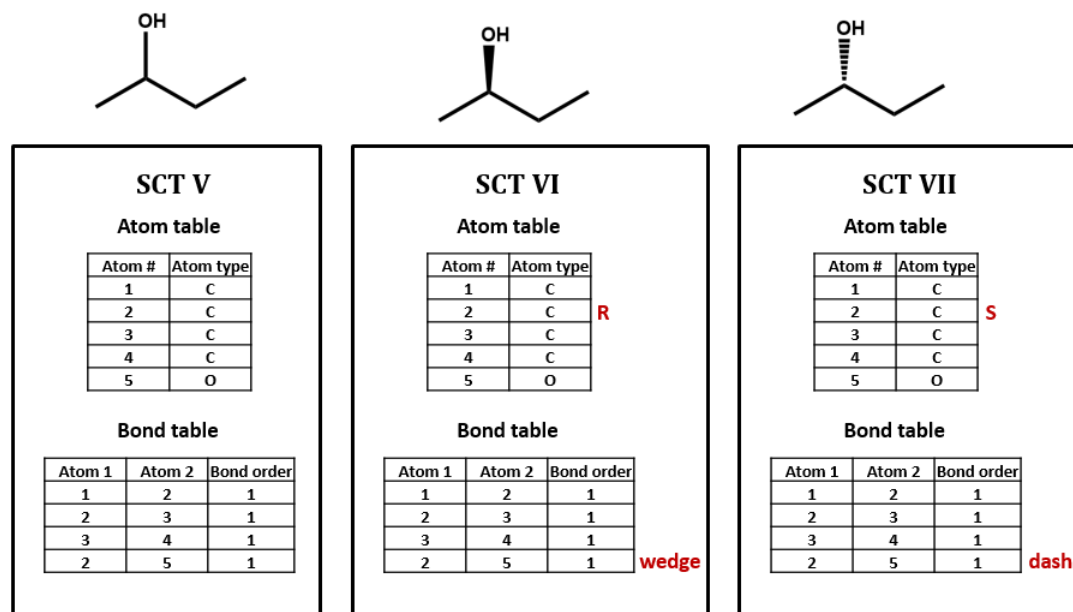


Figure 2.3.8: Three ways of representing 2-butanol. Beyond the information of the SCT, we need to add the material in red, identifying which atom is chiral, and which bond goes into the plane (dash) and out of the plane (wedge) of the drawing.

Cahn-Ingold-Prelog (CIP) Nomenclature

Chiral molecules are distinguished by the letters R- (rectus, Latin for Right handed) and S- (sinister, Latin for left handed). The rules used for identifying chiral centers are the [Cahn-Ingold-Prelog \(CIP\) rules](#). A review of organic chemistry may be required as these can get a bit complicated, but here is the jist using the bromochlorofluoromethane molecule of figure 2.3.6 as an example.

Step 1: Identify chiral atom(s) and rank others by order of priority of atomic mass (number 1 is largest, 3 is smallest) (Br=1, Cl=2, F=3 H=4)

Step 2: Place your thumb of either the right of left hand along the axis of the chiral carbon towards the atom of smallest priority (H here

Step 3, Starting with the atom of highest priority, the fingers of one of your hands will point in the direction of next highest (sequentially decreasing priority), and that hand tells you if it is R or S. So the image on the left of the mirror is R-Bromochlorofluoromethane and the one on the right is the S-Bromochlorofluoromethane.

Now applying this to the middle right image of figure 2.3.7 shows that image is the R isomer.

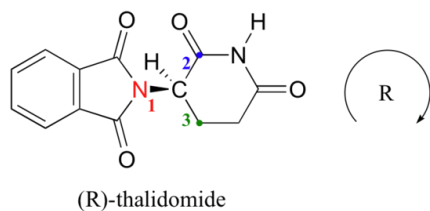


Figure 2.3.8: Applying CIP rules to the middle right thalidomide structure in figure 2.3.7.

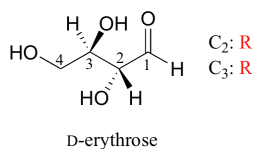
NOTE: If an atom has one chiral center it must be chiral, if it has more than one, it may or may not be chiral. Compounds with more than one stereocenter may be either an enantiomer or a diastereomer.

Diastereomers

These are stereoisomers that are not mirror images of each other. There are two types, geometric and those with chiral centers.

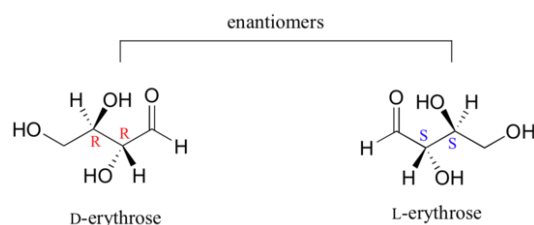
Compounds with Multiple Chiral Centers

We turn our attention next to molecules which have more than one stereocenter. We will start with a common four-carbon sugar called D-erythrose.



A note on sugar nomenclature: biochemists use a special system to refer to the stereochemistry of sugar molecules, employing names of historical origin in addition to the designators 'D' and 'L'. You will learn about this system if you take a biochemistry class. We will use the *D/L* designations here to refer to different sugars, but we won't worry about learning the system.

As you can see, *D*-erythrose is a chiral molecule: C₂ and C₃ are stereocenters, both of which have the *R* configuration. In addition, you should make a model to convince yourself that it is impossible to find a plane of symmetry through the molecule, regardless of the conformation. Does *D*-erythrose have an enantiomer? Of course it does – if it is a chiral molecule, it must. The enantiomer of erythrose is its mirror image, and is named *L*-erythrose (once again, you should use models to convince yourself that these mirror images of erythrose are not superimposable).



Notice that both chiral centers in *L*-erythrose both have the *S* configuration. *In a pair of enantiomers, all of the chiral centers are of the opposite configuration.*

What happens if we draw a stereoisomer of erythrose in which the configuration is *S* at C₂ and *R* at C₃? This stereoisomer, which is a sugar called *D*-threose, is *not* a mirror image of erythrose. *D*-threose is a **diastereomer** of both *D*-erythrose and *L*-erythrose.

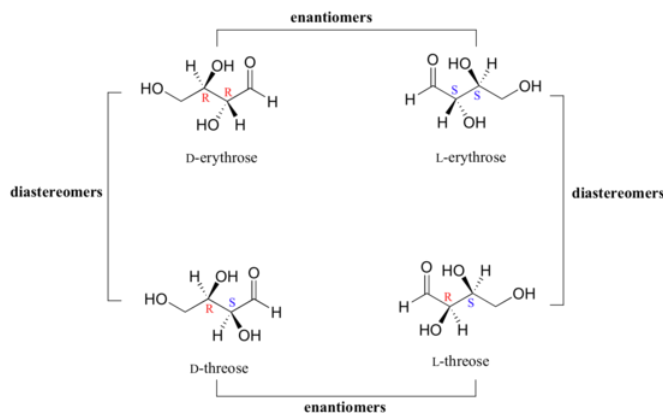


Figure 2.3.10 Looking at the diastereomers of erythrose. Note that each of the threose is a diastereomer to both of the erythroses. That is, there is at least one stereocenter that is not of opposite configuration.

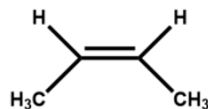
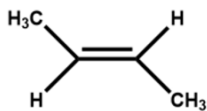
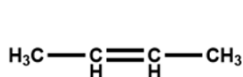
The definition of diastereomers is simple: if two molecules are stereoisomers (same molecular formula, same connectivity, different arrangement of atoms in space) but are *not* enantiomers, then they are diastereomers by default. *In practical terms, this means that at least one - but not all - of the chiral centers are opposite in a pair of diastereomers.* By definition, two molecules that are diastereomers are *not* mirror images of each other.

L-threose, the enantiomer of *D*-threose, has the *R* configuration at C₂ and the *S* configuration at C₃. *L*-threose is a diastereomer of both erythrose enantiomers.

In general, a structure with *n* stereocenters will have 2^{*n*} different stereoisomers. (We are not considering, for the time being, the stereochemistry of double bonds – that will come later).

Geometric Isomers

Geometric isomers are a type of non-chiral diastereomers that have the same connectivity but differ in orientation. Figure 2.2.4 shows 2-butene, which is a planar molecule having a 120° bond angle for the carbons attached to the double bond. Since the double bond can not rotate, the orientations are fixed, meaning the hydrogens are on the same side or opposite side. The image on the left does not define the stereoisomerism, while the two on the right do. The middle image has the hydrogens opposite each other, which is classified as the trans or "E" configuration, while the one on the right has them on the same side of the double bond, which is the cis or "Z" configuration.



SCT VIII

Atom table

Atom #	Atom type
1	C
2	C
3	C
4	C

Bond table

Atom 1	Atom 2	Bond order
1	2	1
2	3	2
3	4	1

SCT IX

Atom table

Atom #	Atom type
1	C
2	C
3	C
4	C

Bond table

Atom 1	Atom 2	Bond order
1	2	1
2	3	2
3	4	1

E

SCT X

Atom table

Atom #	Atom type
1	C
2	C
3	C
4	C

Bond table

Atom 1	Atom 2	Bond order
1	2	1
2	3	2
3	4	1

Z

Figure 2.3.11: Three Lewis dot structures and their connection tables for 2-butene. Not, all three tables have the same SCT, and so a real file would have three options, to define the stereochemistry as E or Z, or not to define it at all.

In the case of geometric isomers physical data can be different, for example, the z isomer may have a slightly different boiling point, and so when someone reports a value and uploads it to the database, it needs to be determined which isomer they had, if they know, or if it is a mixture. So often times databases may have different properties for the same substance because the scientist who made the measurement did not know or report the structure correctly.

Once again, the SCT can not distinguish these isomers, and you need more information than what the atoms are, and what is bonded to what.

Resonance Structures

Lewis dot structures and connection tables consider a covalent bond to consist of two electrons shared between two nuclei, thus forming a bonding orbital. But many times pi bonds of adjacent atoms can overlap to produce an orbital that involves electrons being shared between three or more nuclei. In this case you can't draw one Lewis dot structure (or connection table), and have to draw two (or more) with each of these structures being a **resonance structure**, and the real molecule being a sort of average of all resonance structures. In the case of nonaromatic resonance structure, the current protocol is to draw each resonance structure as its own connection table. Connection tables have a special way of representing aromatic compounds that have rings of delocalized electrons.

Aromatic Structures

Aromatic structures are common in organic chemistry and involve conjugated ring systems where electrons in p orbitals combine into pi-orbital rings systems forming delocalized orbitals over multiple nuclei. Benzene is the simplest aromatic compound, and because these are so ubiquitous, they are typically given a bond order of 4.

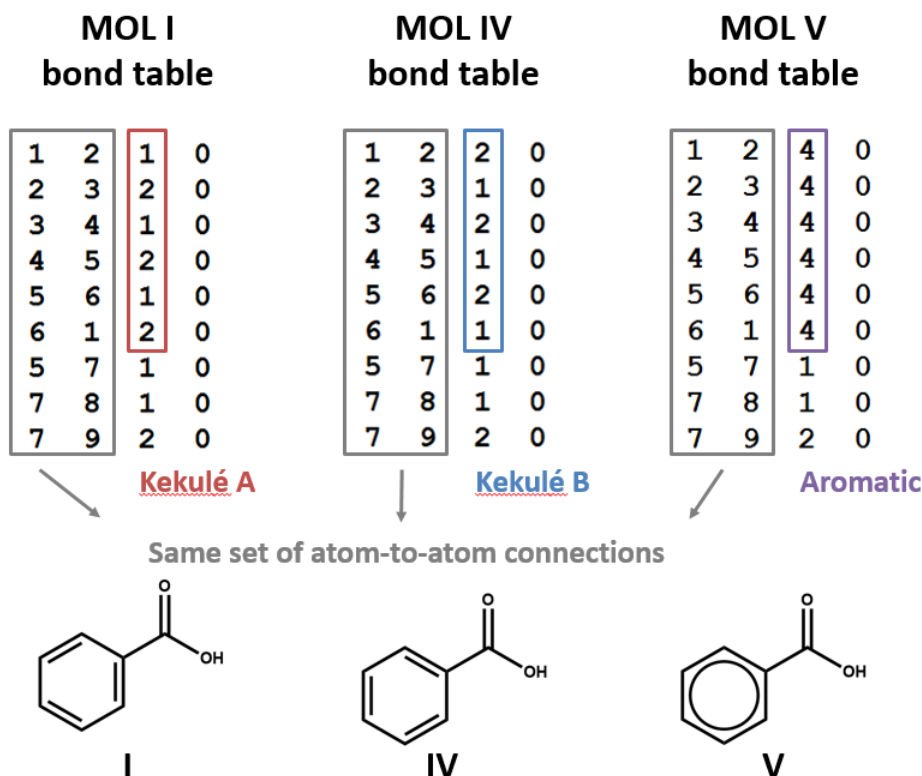


Figure 2.3.12 The two Kekulé structures represent the resonance structures of the benzene ring, which are often "combined" to form the ring structure on the right, which is given a bond order of 4.

Tautomers

Hydrogens are often labile and can easily jump from one atom to another and this can occur very rapidly. So a molecule may be jumping back and forth between two Lewis dot structures/connection tables.

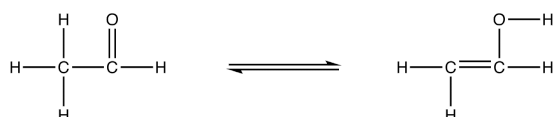


Figure 2.3.3: keto-enol tautomerism showing the two structures associated with acetaldehyde, where the hydrogen is jumping between the two carbons as the electron pair of the double bond switches between C=O and C=C.

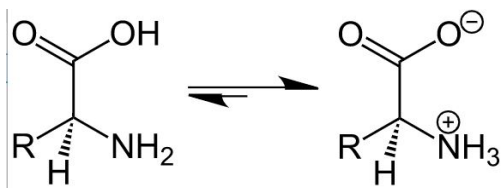


Figure 2.3.4 Amino acid undergoing tautomerism as it transforms between neutral and bi-charged (zwitter ionic) forms.

Zwitterions are typically dealt with as two files, and then connected by the database.

Multicovalent Units

A covalent unit is a single chemical entity held together by covalent bonds. Many chemical substances consists of multiple covalent units, like salts and mixtures. In the case of salts, each covalent unit has a charge being positive (cation) or negative (anion) and the sum of the charges must be neutral or the salt will not form.

A salt can be represented by a Simple Connection Table through the bond table where two groups of atoms in the atom table are simply not connected. We will take a closer look at this in section 2.5 when we look at actual chemical structural data files. But in essence, a file for a chemical multivalent unit substance contains several disconnected bond groups within the bond table, and shows every atom of the salt. In the case of crystal structures these can include the 3D coordinates.

Mixtures are more complicated because neither the coordinates or the ratio of the substances are typically defined, and so they are typically not represented as structural data.

Contributors

[Robert E. Belford](#) (University of Arkansas Little Rock; Department of Chemistry). The breadth, depth and veracity of this work is the responsibility of Robert E. Belford, rebelford@ualr.edu. You should contact him if you have any concerns. This material has both original contributions, and content built upon prior contributions of the LibreTexts Community and other resources, including but not limited to:

- Evan Hepler-Smith
- Leah R. McEwen
- [Material Adopted from 2017 Cheminformatics OLCC](#)

2.3: Molecular Graph Issues is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by LibreTexts.