Enhanced Stereochemistry and its use in ASAP

Overview

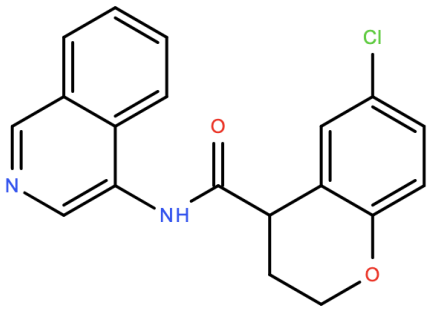
This document is intended as a very brief introduction to enhanced stereochemistry, and its use on ASAP projects. Enhanced stereochemistry does take some getting used to, but it is a powerful tool for representing chemical data once learned. Please note that our conventions may differ from other projects using similar representations -- the key is consistency among our projects.

As canonical references to this material, I highly suggest the blog post V3000 Molfile Enhanced Stereochemistry Representation and the references therein. These documents will provide much more detail than this doc, and thus are worth reading. Also, feel free reach out to manifold@postera.ai with any questions.

A first example

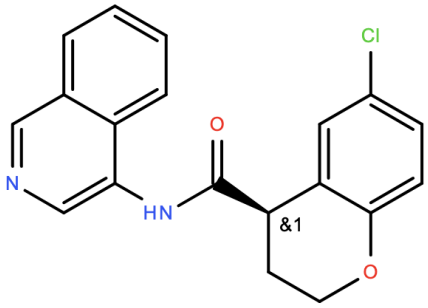
Racemates

Starting with a practical example, let's imagine a project with a chiral center in the main series. An example lead structure is shown below:



To most people, they may recognize the lack of stereochemistry at the chiral center as representing a racemate. However, in ASAP, we are going to only use "flat chiral centers" when the stereochemical configuration at the chiral center is truly unknown at that stereocenter.

Our preferred representation of the racemate is the following:



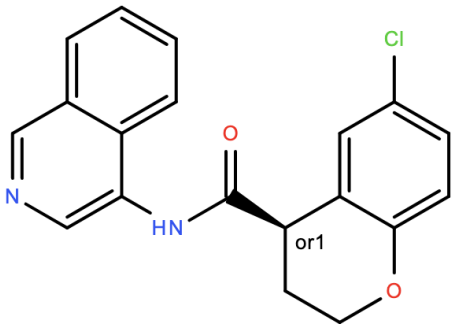
Note that the wedge can be drawn as up or down. It doesn't matter, since the `&1` symbol indicates that the stereocenter is a member of an `AND group`, and more specifically `AND group #1`. I.e. the molecule is a mixture of both the (R) AND (S) configurations at the specified stereocenter.

Some may be thinking that this is a very long way of saying the compound is "racemic." However, this does add precision to our representation, and you'll learn to like it.

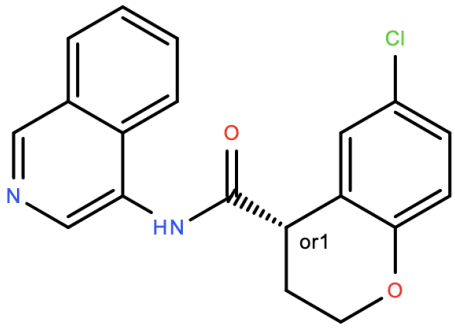
Dealing with Chiral Separations

Now let's pretend we want to separate the racemate shown above into its pure enantiomers, in order to measure the properties of the stereochemically pure compounds.

If a route to chiral pure materials cannot be developed, we have to resort to separation on a chiral column. The two compounds that come off the column can be represented as follows:



First eluting off column.
Single enantiomer of unknown absolute configuration.

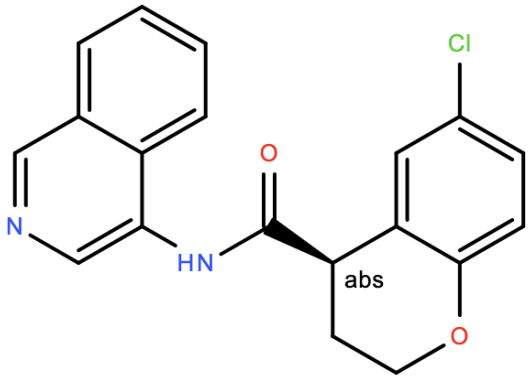


Second eluting off column.
Single enantiomer of unknown absolute configuration.

The stereocenters are now members of "OR" groups, which indicate they are either the (R) OR (S) configuration, but not both! That is, we know the enantiomers are pure and separated, we just don't know how to assign the absolute configuration.

This is really useful because we no longer have to arbitrarily assign each stereocenter to (R) or (S) and write a note that the "stereochemistry is arbitrarily defined," thus confusing the reader who misses the note and assumes the absolute configuration is known.

On the other hand, if we do develop a synthetic route to obtain chirality pure material, we might be able to obtain only the desired enantiomer,



where the "ABS" group indicates that the stereocenter is absolutely defined and as shown. Note that we do not require the "ABS" flag on compounds, though it can

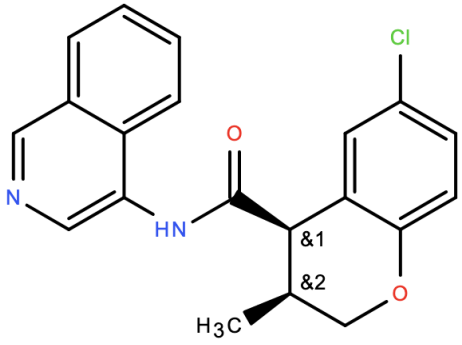
sometimes be useful. Our convention is that a compound drawn as (R) or (S) is understood to be that configuration unless an OR or AND flag is present.

Putting it all together

To sum up, each stereocenter should be considered part of a "stereochemical group," and the group should be designated as either:

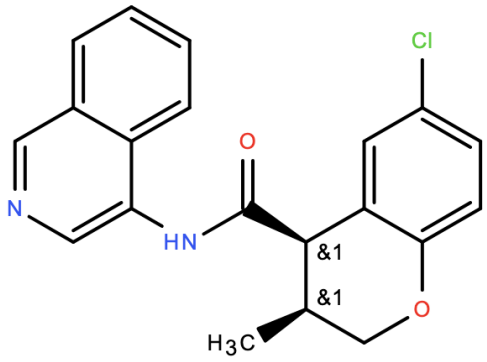
- `absolute`: with either no label by default or `abs`
- `racemic` (AND): with the label `&<group #>`
- `relative` (OR): with the label `or<group #>`

Of note is that multiple stereocenters can belong to the same group. This allows us to represent disastereoisomers in useful ways. For example, if we pretend we start pursuing a diastereomeric series, me might expect to obtain a true equal mixture of all 4 stereoisomers upon first synthesis, which can be represented as follows:



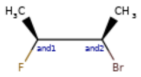
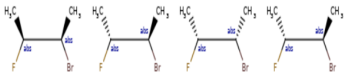
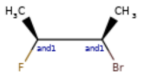
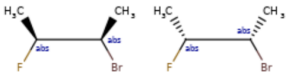
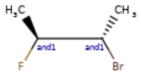
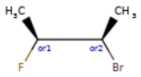
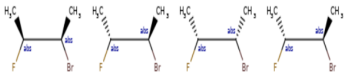
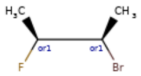
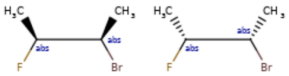
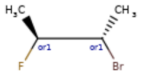
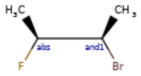
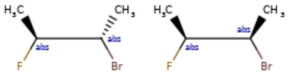
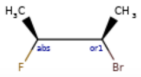
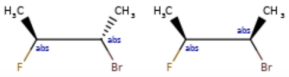
In this case, each stereocenter is can be both (R) AND (S), so we get a mixture of all 4 stereoisomers (R,R), (R,S), (S,R), (S,S).

If however, we only get the cis-stereoisomers for some reason, that can be represented as follows:



Note that now the two stereocenters are "tied together." I.e., when one of the stereocenters is up, the other is up, and vice versa. Thus, we are representing a mixture of the two cis stereoisomers.

This chart from RDKit's documentation nicely summarizes what the many combinations of stereocenter designations mean.

What's drawn	Mixture?	What it means
	mixture	
	mixture	
	mixture	
	single	
	single	
	single	
	mixture	
	single	

Getting the MOL V3000 files or CXSMILES needed to represent these molecules

Note that in order to transfer these "enhanced stereochemistry" representations of molecules between programs, you will either need to use "V3000 MOL files" or "enhanced SMILES" (often known as CXSMILES).

The following video gives a way to do this using the drawing program at postera.ai/manifold.

Please take a look at the above video, or email manifold@postera.ai if you have troubles getting the representation you need.