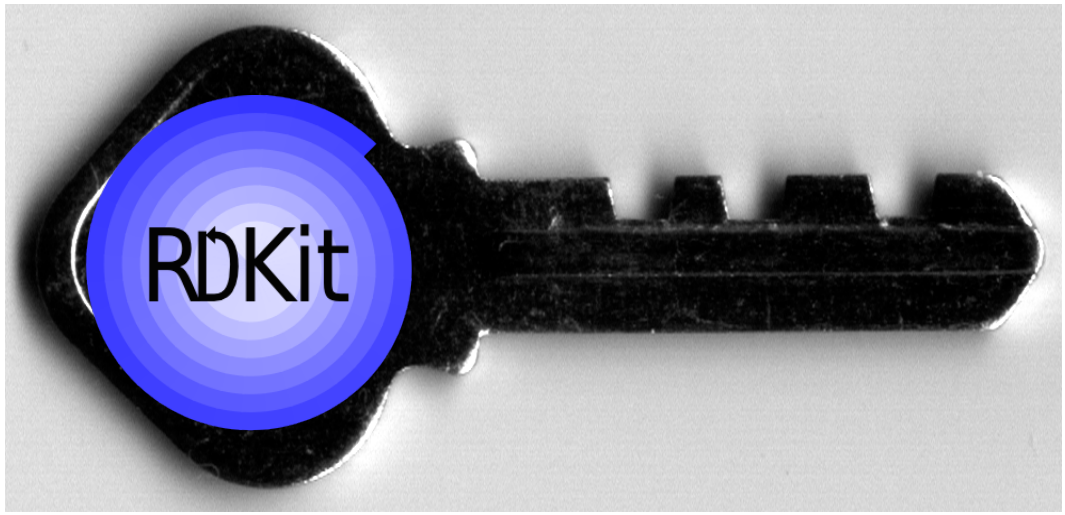


A Key ~~to~~for Registration

Gregory Landrum

NIBR Informatics

Novartis Institutes for BioMedical Research

RDKit UGM 2015, Zurich

NOVARTIS

Image source: https://commons.wikimedia.org/wiki/File:Skeletonkey.png

A key for structure registration

Requirements

- Generate a structure-derived key for every compound that can be used for uniqueness checking and as a primary key in a database
- Register tautomers as-drawn
- Handle ambiguous/partially known stereochemistry
- Provide a mechanism so that a new key can *always* be generated for a structure

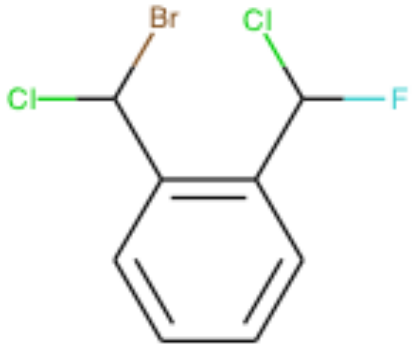
A minimalist solution

- Start from a CTAB (mol block), a stereo code (from a controlled vocabulary), and a stereo comment (free text)
- Do minimal standardization (using STRUCHK from RDKit AvalonTools integration)
- Generate a *non-standard* InChI with a fixed H layer and unknown/unspecified stereo always indicated (options /FixedH /SUU)
- Generate an MD5 hash of the InChI + the stereo code + the stereo comment
- Include the version
- An example: `1|5H9R3LvclagMXHp3Clrc/g==`

Possible stereo codes

- S_ACHIR: Single, achiral
- S_ABS: Single, absolute stereochemistry known
- S_REL: Single, relative stereochemistry known
- S_PART: Single, partial stereochemistry known
- S_UNKN: Single, stereochemistry unknown
- S_ABS_ACHIR: Single, absolute stereochemistry known, achiral
- R_ONE: Racemate, one chiral center
- R_REL: Racemate, relative stereochemistry known
- R_OTHER: Racemate, other
- MX_ENANT: Mixture of enantiomers
- MX_DIAST: Mixture of diastereomers
- MX_SP2: Mixture of SP2 isomers
- MX_DIAST_ABS: Mixture of diastereomers with absolute stereochemistry known
- MX_DIAST_REL: Mixture of diastereomers with relative stereochemistry known
- OTHER
- UNDEFINED

Example



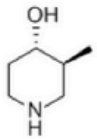
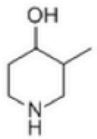
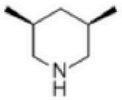
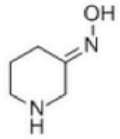
Two undefined stereocenters
InChI: InChI=1/C8H6BrCl2F/c9-7(10)5-3-1-2-4-6(5)8(11)12/h1-4,7-8H/t7?,8?

Stereo code = S_UNKN (single stereoisomer, unknown):
1|5H9R3LvclagMXHp3Clrc/g==

Stereo code = S_REL (single stereoisomer, relative stereo known)
1|cqKWVsUEY6QNpGCbDaDTYA==

Dealing with structures from external SDFs

Specific Examples for Structures Coming from SDFs

Drawing	Notes	Stereochem tag if chiral flag set	Comment required?	Stereochem tag if chiral flag not set	Comment required?
		Needs inspection		Single compound - achiral	no
	all stereocenters specified	Single compound - absolute stereo	no	Racemate - >1 chiral centre. Relative stereo known	yes, indicating why the relative stereo was set Should be drawn flat?
	one stereocenter, unspecified	Needs inspection		Racemate - 1 chiral centre	
	multiple stereocenters, unspecified	Needs inspection		S_UNKNOWN	
	stereocenters specified, but plane makes molecule achiral	Single compound – absolute stereo but achiral		Single compound – absolute stereo but achiral	
		Chiral flag irrelevant if drawn with squiggle or crossed bond, then SP2 mixture			

Dealing with structures from external SDFs

Need to come up with a stereo code and, possibly, a comment

Nspecified	Nunspecified	Meso?	Chiral Flag?	flag	comment	warning?
	0	0 F	T	S_ACHIRAL		Chiral flag on molecule with no chiral centers
	0	0 F	F	S_ACHIRAL		
	1	0 F	T	S_ABS		
	1	0 F	F	R_ONE		
	0	1 F	T	S_UNKN		Chiral flag on molecule with no chiral centers
	0	1 F	F	R_ONE		
>1		0 F	T	S_ABS		
>1		0 F	F	R_REL		
	0 >1	F	T	S_UNKN		Chiral flag on molecule with no chiral centers
	0 >1	F	F	MX_DIAST		
	1	1 F	T	S_PART		
	1	1 F	F	MX_DIAST		
>1		0 T	T	S_ABS_ACHIR		
>1		0 T	F	S_ABS_ACHIR		
	1 >1	F	T	S_PART		
	1 >1	F	F	MX_DIAST		
>1		1 F	T	S_PART		
>1		1 F	F	MX_DIAST_REL		
>1	>1	F	T	S_PART		
>1	>1	F	F	MX_DIAST_REL		

Acknowledgements

- Thomas Mueller
- Matthias Huebscher
- Willi Sieber
- Bernhard Rohde
- Matthias Wrobel
- Werner Breitenstein