TAUTOMERS AND MESOMERS: REGISTRATION AND SEARCHING

ROGER SAYLE

NEXTMOVE SOFTWARE LTD

CAMBRIDGE UK

Chemical Registry Systems, EBI Industry Workshop, UK, 11th October 2011

OVERVIEW

- Context, Definitions and Examples
- Searching for Tautomers
- Registration of Tautomers
- Conclusions

INTRODUCTION

- The existence of multiple tautomeric forms of small molecules creates a problems for registration in chemical information systems, physical property calculation and compound clustering/similarity searching.
- In 1999, Sayle and Delany described novel algorithms for enumerating and for generating a canonical form of a tautomeric molecule.
- This talk describes advances in the field over the past 12 years, and suggestions for the state-of-the-art.

A LITTLE HISTORY

- Roger Sayle and Jack Delany, "**Canonicalization and Enumeration of Tautomers**", Innovation Computational Applications, San Francisco, October 25-27, 1999.
- http://www.daylight.com/meetings/emug99/Delany/taut_html/slide01.html
- SciTegic Pipeline Pilot, "**Chemistry Collection: Advanced Chemistry**", page 20, "Enumerate Tautomers Component".
- Paul Labute, "**On the Perception of Molecules from 3D Atomic Coordinates**", JCIM, Vol. 45, No. 2, pp. 215-221, 2005.
- Frank Oellien et al., "**The Impact of Tautomer Forms on Pharmacophore-based Virtual Screening**", JCIM, Vol. 46, No. 6, pp. 2342-2354, 2006.
- Roger Sayle, "**So You Think You Understand Tautomerism?**", Journal of Computer-Aided Molecular Design (JCAMD), Vol. 24, Nos. 6-7, pp. 485-496, 2010.

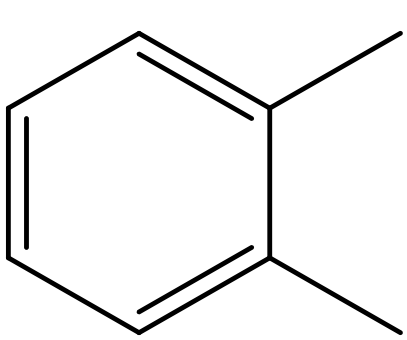
A LITTLE OF THE FUTURE

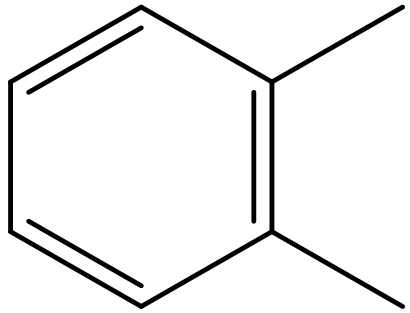
- The IUPAC International Chemical Identifier: InChI v1.03 and later [mixtures, reactions, polymers, organometalics, etc.]
- U.S. Food and Drug Administration (FDA), Substance Registration System – Unique Ingredient Identifier (UNII).
- European Union IMI OpenPhacts.

DEFINITIONS

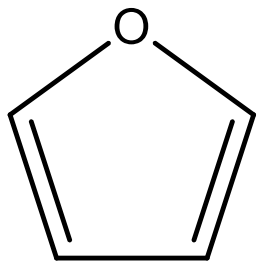
- Tautomers are special kinds of structural isomers that are rapidly inter-convertible through a rapid equilibration.
- The movement of a hydrogen atom between discrete sites in the “same” molecule.

AROMATICITY





AROMATICITY AND QSAR



Name: Furan
Test: #96
Exptl: 1.34

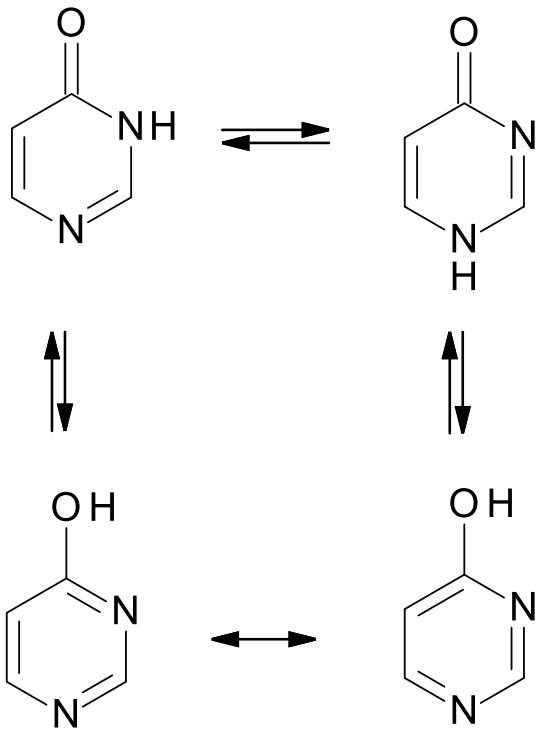
Original XLogP

R-O-R:	0.327	0.327
R=CHX	-0.166	-0.332
R=CHR	0.236	0.472
H	0.046	0.184
Total:		0.651

Aromatic Furan XLogP

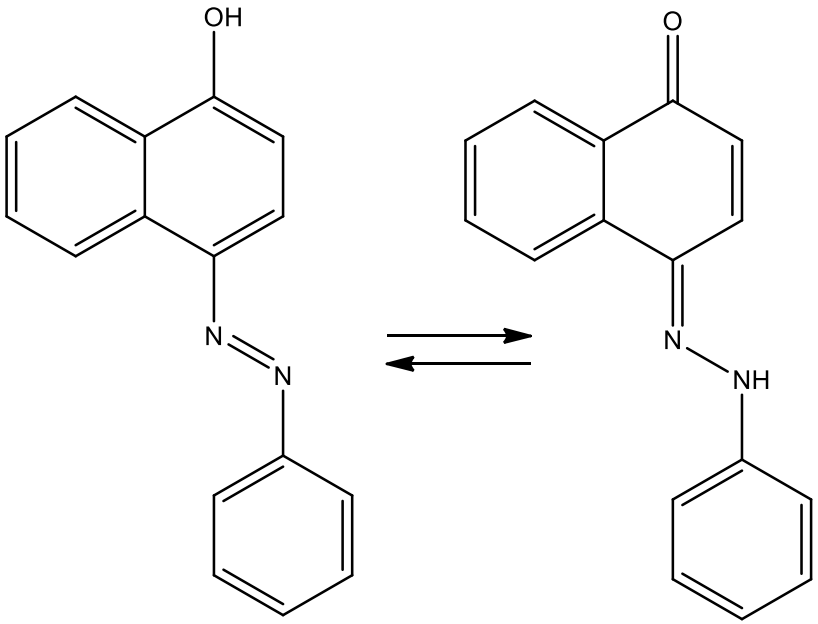
R-O-R	0.327	0.327
R-CH-X	0.142	0.284
R-CH-R	0.281	0.562
H	0.046	0.184
Total:		1.357

CLASSIC TAUTOMERISM: 4-PYRIMIDONE



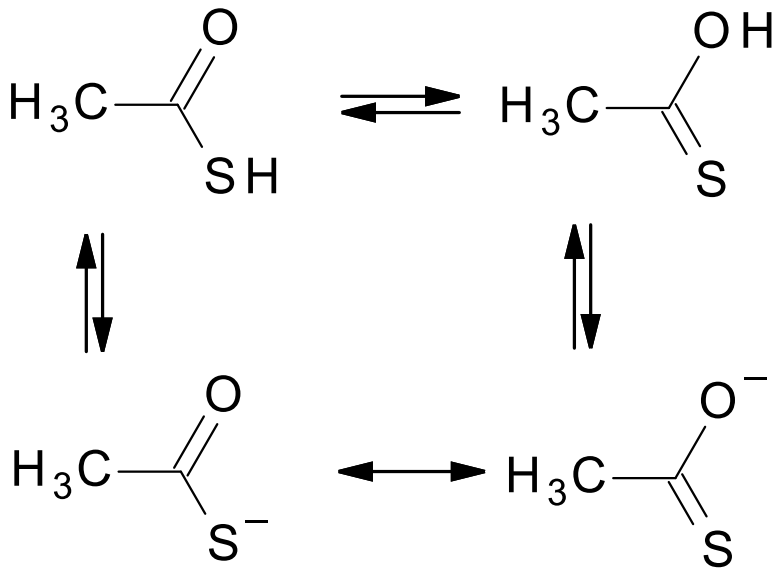
InChI=1S/C4H4N2O/c7-4-1-2-5-3-6-4/h1-3H,(H,5,6,7)

CLASSIC TAUTOMERISM: LAAR 1886



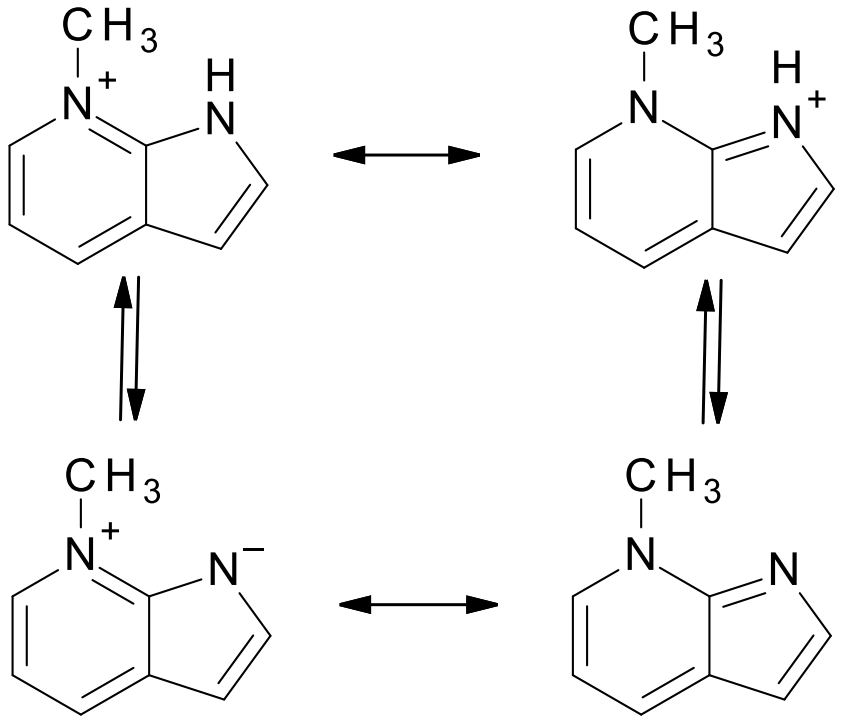
InChI=1S/C16H12N20/c19-16-11-10-15(13-8-4-5-9-14(13)16)18-17-12-6-2-1-3-7-12/h1-11,19H
InChI=1S/C16H12N20/c19-16-11-10-15(13-8-4-5-9-14(13)16)18-17-12-6-2-1-3-7-12/h1-11,17H

TAUTOMERISM, MESOMERISM & PKA

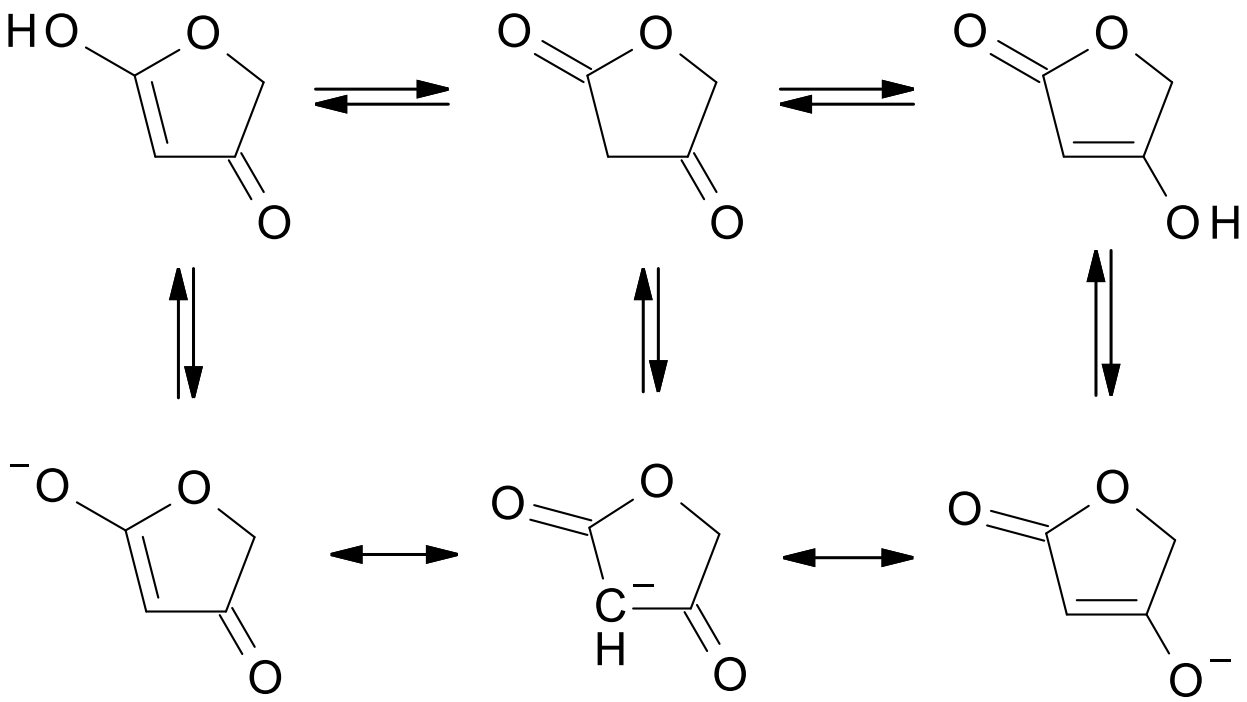


InChI=1S/C2H4OS/c1-2(3)4/h1H3,(H,3,4)
InChI=1S/C2H4OS/c1-2(3)4/h1H3,(H,3,4)/p-1

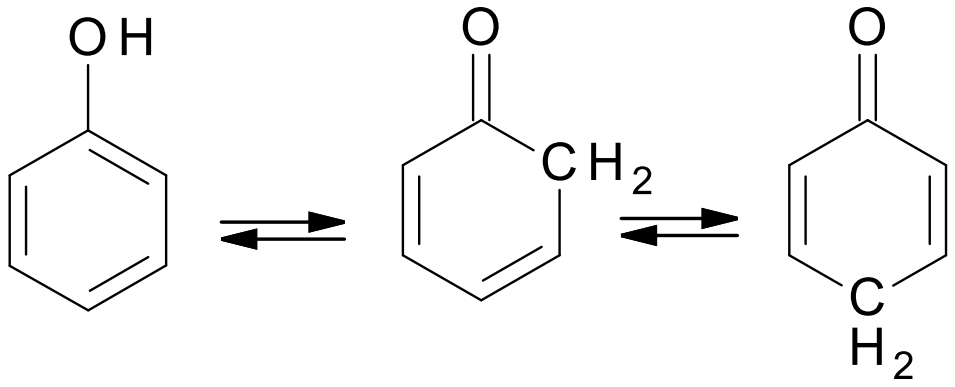
CRYPTIC NEUTRALIZATION



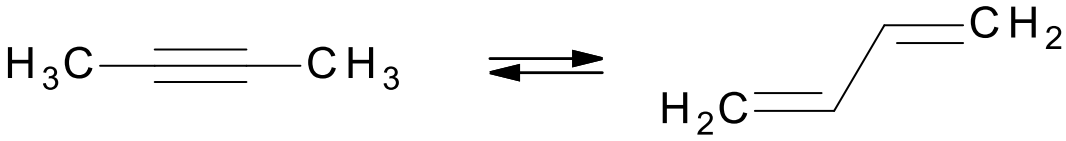
MYTHICAL ACIDIC CARBON ACIDS

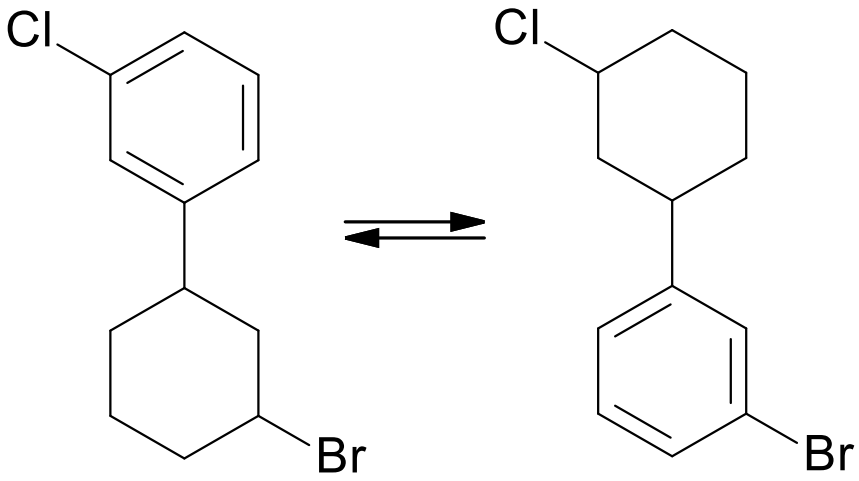


TAYLOR C-TYPE TAUTOMERISM

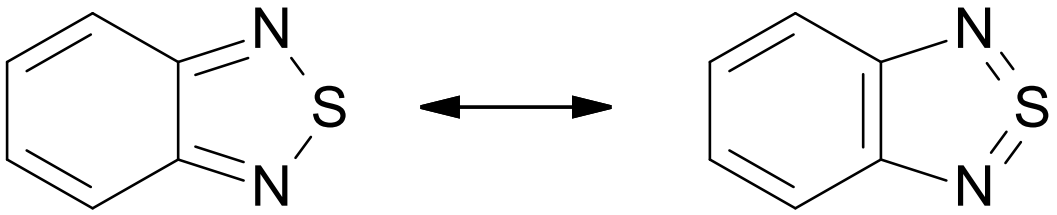


LIMITS OF C-TYPE TAUTOMERISM

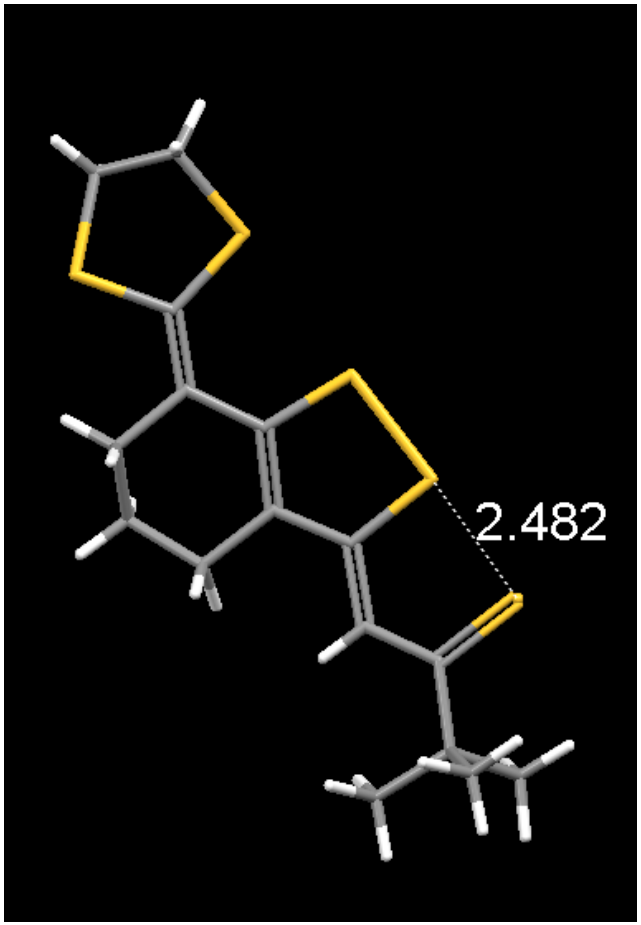


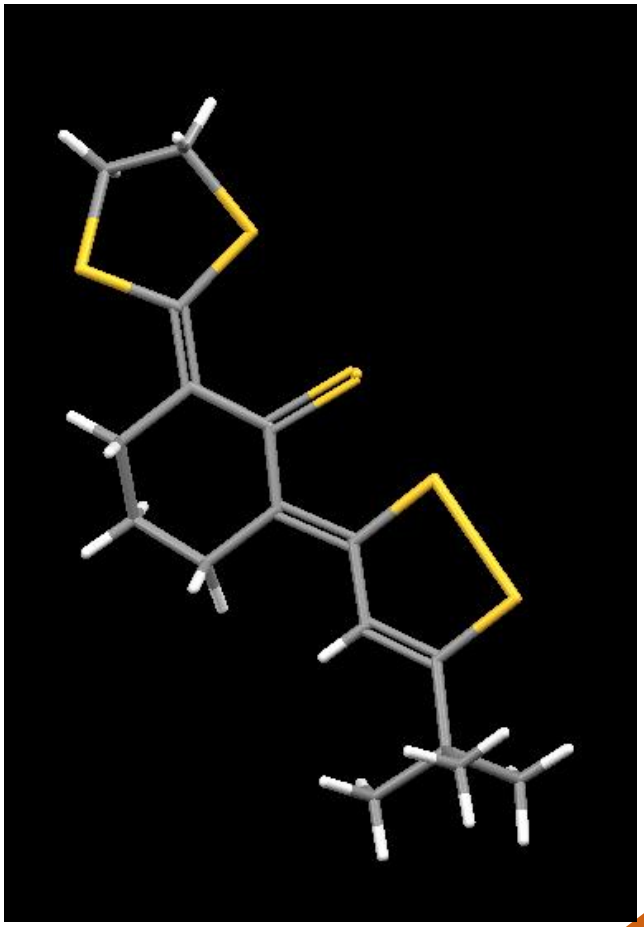


MESOMERISM



RING-RING MESOMERISM





SITZMANN "FICTS" TERMINOLOGY

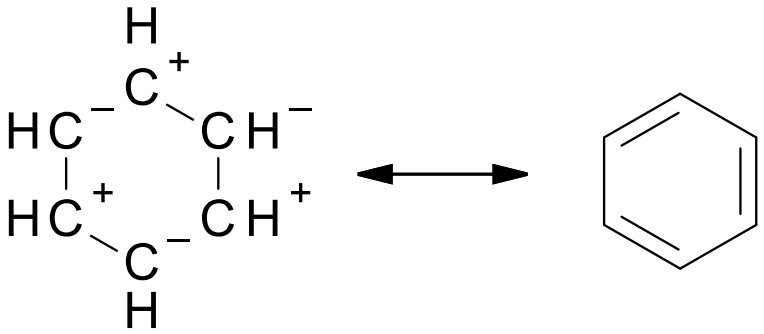
- Markus Sitzmann, Wolf-Deitrich Ihlenfeldt and Marck C. Nicklaus, "Tautomerism in Large Databases", JCAMD 24(6-7):521-551, 2010.
- InChI/InChIKey vs. NCI/CADD Structure Identifiers: A Comparison
- http://acscinf.org/docs/meetings/237nm/presentations/237nm17.pdf
- www.slideshare.net/sitzmann/iccs9-2011-talk

- Defines a nomenclature and (5-level) hierarchy of 32 equivalences. For example, Daylight's "absolute" SMILES is equivalent to FuCTu.

THE FIVE (OR SIX) COMPUTATIONS

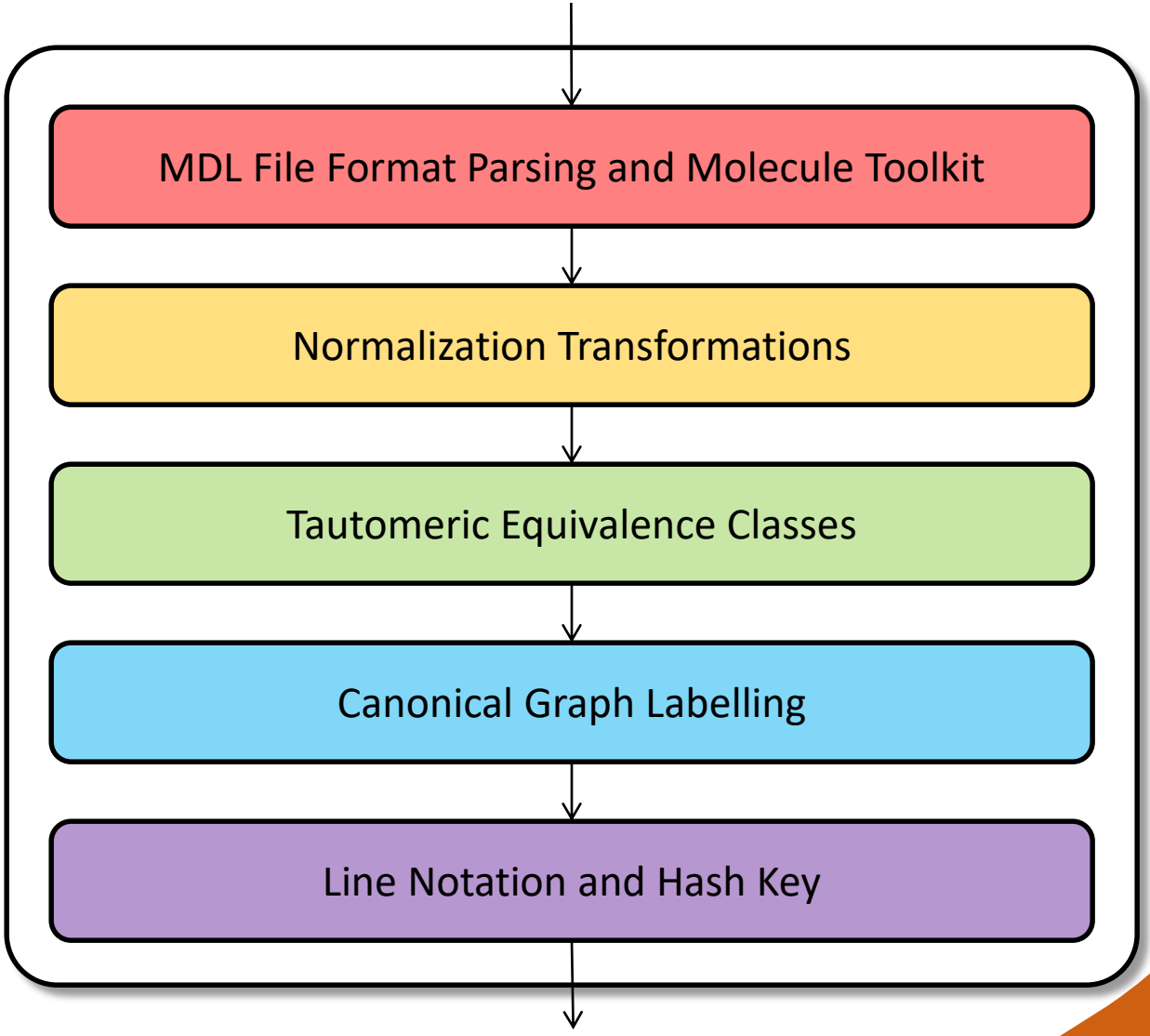
1. Comparison.
2. Canonicalization.
3. Enumeration.
4. Visualization.
5. Selection.
6. Prediction.

ENUMERATION IS FUTILE

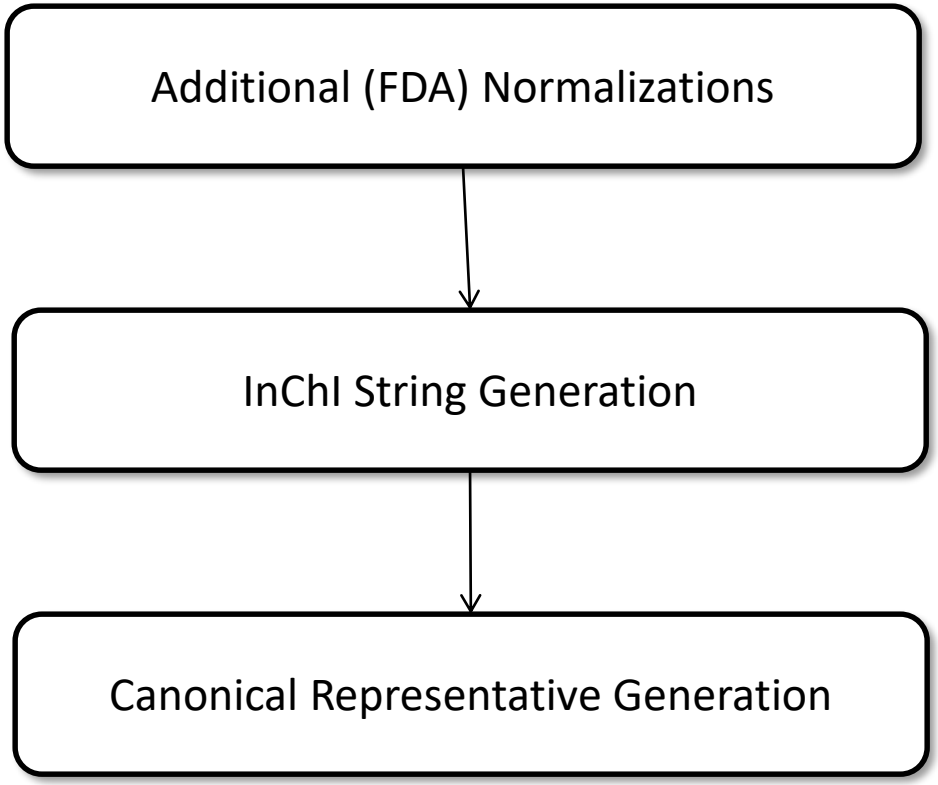


InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H

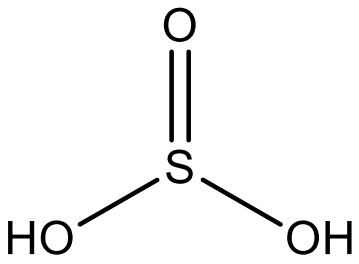
INCHI TECHNOLOGY COMPONENTS



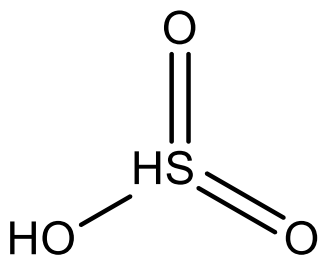
PROPOSED STRATEGY



EXAMPLE CHEBI DUPLICATES



Sulfurous Acid [Hydrogen Sulfite]
CHEBI:48854
InChI=1S/H2O3S/c1-4(2)3/h(H2,1,2,3)
CHEMBL:1161699



Sulfonic Acid
CHEBI:29214
InChI=1S/H2O3S/c1-4(2)3/h4H,(H,1,2,3)

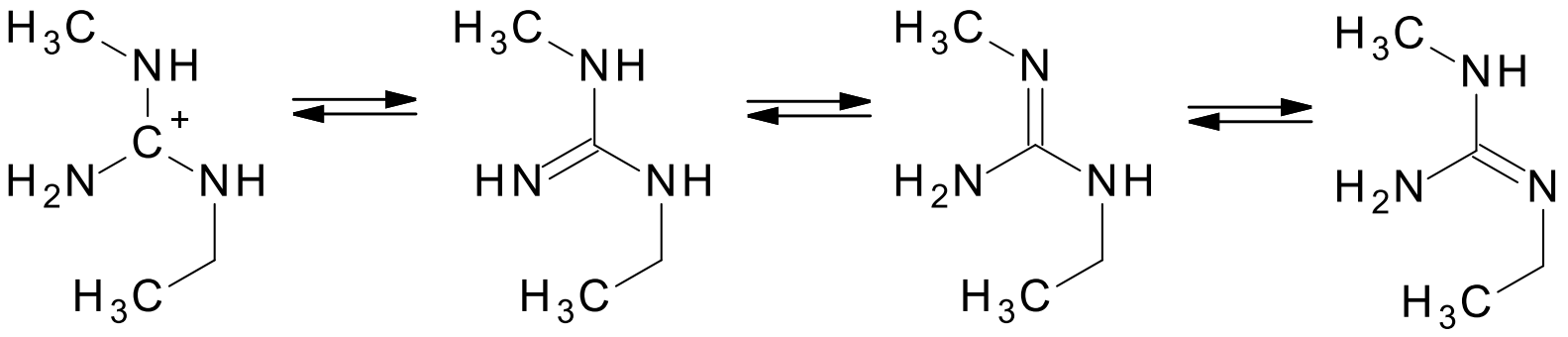
ENUMERATION FROM INCHI

- Torsten Thalheim, Armin Vollmer, Ralf-Uwe Ebert, Ralph Kühne and Gerrit Schüürman, “Tautomer Identification and Tautomer Structure Generation Based on the InChI Code”, JCIM 50():1223-1232, 2010.
- As implemented in EBI’s OrChem.
- Mark L. Rijnbeek and Christoph Steinbeck, “OrChem: An Open Source Chemistry Search Engine for Oracle”, Journal of Cheminformatics, 1:17, 2009.

TAUTOMERIC PREFERENCE

- Aromaticity is unimportant.
- Extending conjugated systems is significant.
- Amide formation is important.
- Geometry and hybridization.
- Lone pair repulsion/electrostatics.
- Electron Withdrawing/donating (σ*).

GUANIDINE


H3C
NH
C+
H2N
NH
H3C
H3C
NH
HN
NH
H3C
H3C
N
H2N
NH
H3C
H3C
NH
H2N
N
H3C

SUMMARY

- There are many (competing, complementary and conflicting) notions of molecular equivalence.
- InChI is the most significant development in the field in the past decade.
- Best practices record/track deposited form.
- Suggestion is that future standardization be built on top of existing infrastructure.

ACKNOWLEDGEMENTS

- Peter Taylor, ICI/Zeneca, UK (retired).
- John Bradshaw, GSK/Daylight, UK (retired).
- Peter Kenny, AstraZeneca, UK (retired).
- Dave Weininger, Daylight CIS, USA (retired).
- Yvonne Martin, Abbott, USA (retired).
- Robin Taylor, Taylor Cheminformatics Software, UK.
- Andrew Grant, AstraZeneca, UK.
- Sorel Muresan, AstraZeneca, Sweden.
- Evan Bolton, NCBI, USA.

THANK YOU FOR YOUR TIME

- Thanks to AstraZeneca for funding.

- And thank you for your questions.