



# NCI/CADD Chemical Identifier Resolver: Indexing and Analysis of Available Chemistry Space

Markus Sitzmann<sup>1</sup>, Wolf-Dietrich Ihlenfeldt<sup>2</sup>, and  
Marc C. Nicklaus<sup>1</sup>

[1] Computer-Aided Drug Design Group, Chemical Biology Laboratory,  
NCI-Frederick, NIH, DHHS

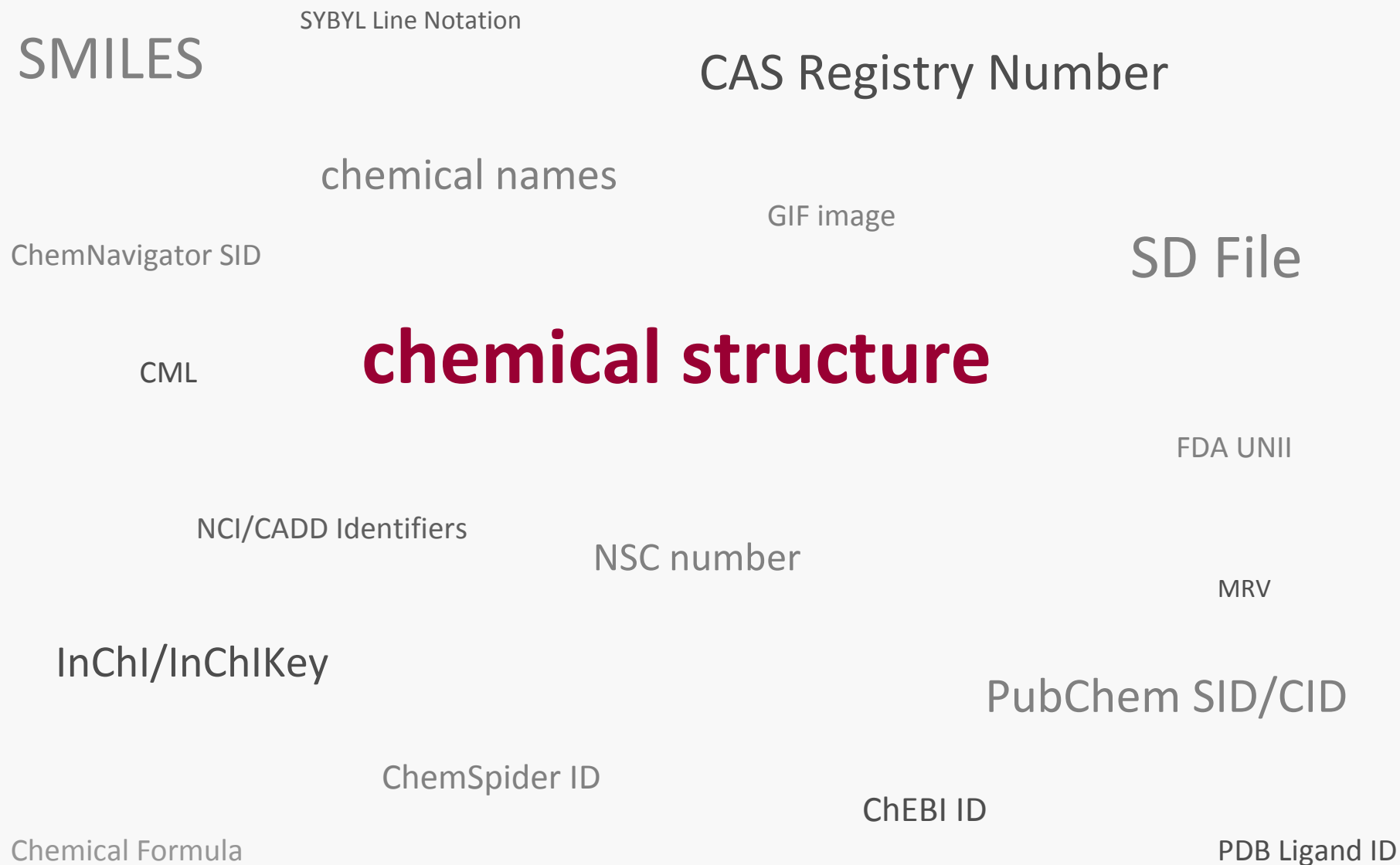
[2] Xemistry GmbH, Auf den Stieden 8, D-35094 Lahntal, Germany

---

# Chemistry Space Analysis

- how many small-molecules are there currently?
- since the early 2000s: enormous increase of the number of databases containing small molecules, e.g. PubChem, ChemSpider, ChEMBL, DrugBank – what is the overlap?
- many ambiguities in the representation of small molecules (e.g. tautomerism, salts, ionic resonance forms)
- growing number of chemical structure identifiers (InChI/InChIKey, PubChem SID/CID, ChemSpider ID, ChEBI ID, ...)

# Chemical Identifier Resolver



# Chemical Identifier Resolver

The screenshot shows the web interface for the Chemical Identifier Resolver. At the top, there is a red header with the URL 'cactus.nci.nih.gov' and the title 'Chemical Identifier Resolver beta 2'. Below the header, there is a form with a 'Structure Identifier' input field, a 'Representation' dropdown menu set to 'Standard InChIKey', and a 'Submit' button. The main content area is titled 'Getting started ...' and contains the following text: 'This service works as a resolver for different chemical structure identifiers and allows one to convert a given structure identifier into another representation or structure identifier. You can either use the resolver web form above or use the following simple URL API scheme:   
`http://cactus.nci.nih.gov/chemical/structure/"structure identifier"/"representation"`   
  
Example: Chemical name to Standard InChIKey:   
`http://cactus.nci.nih.gov/chemical/structure/aspirin/stdinchikey`   
  
The service returns the requested new structure representation with a corresponding MIME-Type specification (in most cases *MIME-Type: text/plain*). If a requested URL is not resolvable for the service an *HTML 404 status* message is returned. In the (unlikely) case of an error, an *HTML 500 status* message is generated. [Read more.](#)

At the bottom of the page, there is a footer with the following text: 'Getting started ... | [Documentation](#) | [Blog](#) | [Contact](#) | [Disclaimer](#) | [Privacy Statement](#)   
Markus Sitzmann (sitzmann+++helix.nih.gov)   
NCI/CADD Group 2009   
15th March 2010 16:38

*Works as a resolver for different chemical structure identifiers. Allows one to convert a given structure identifier into another representation or structure identifier.*

**first beta release: July 2009**  
**current release (beta 4): April 2011**

<http://cactus.nci.nih.gov/chemical/structure>

## Chemical Identifier Resolver

- it is usable by a simple URL API:

**`http://cactus.nci.nih.gov/chemical/structure/"identifier"/"representation"`**

**XML format:** `http://cactus.nci.nih.gov/chemical/structure/"identifier"/"representation"/xml`

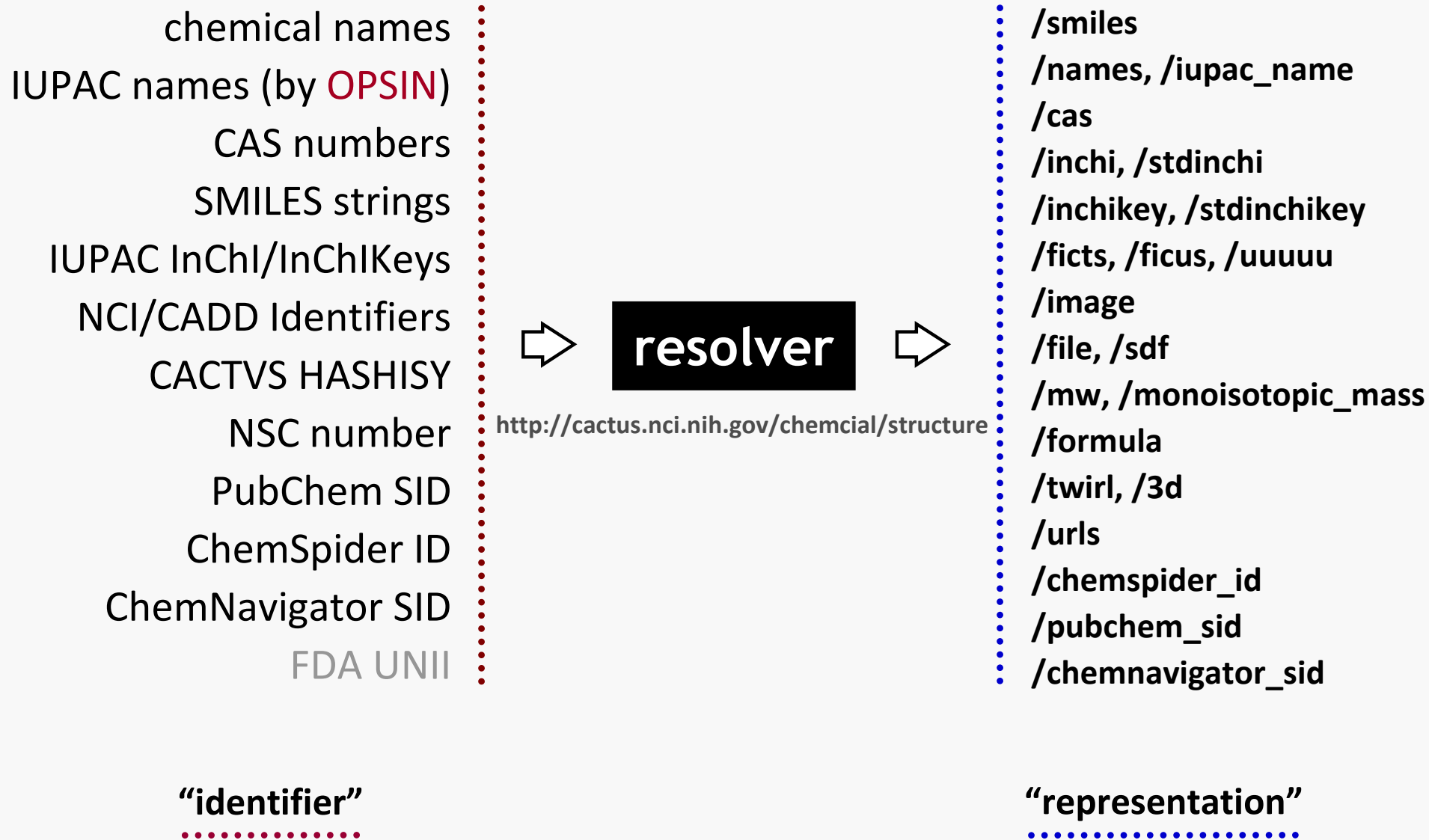
**example:** `http://cactus.nci.nih.gov/chemical/structure/Tamiflu/cas`



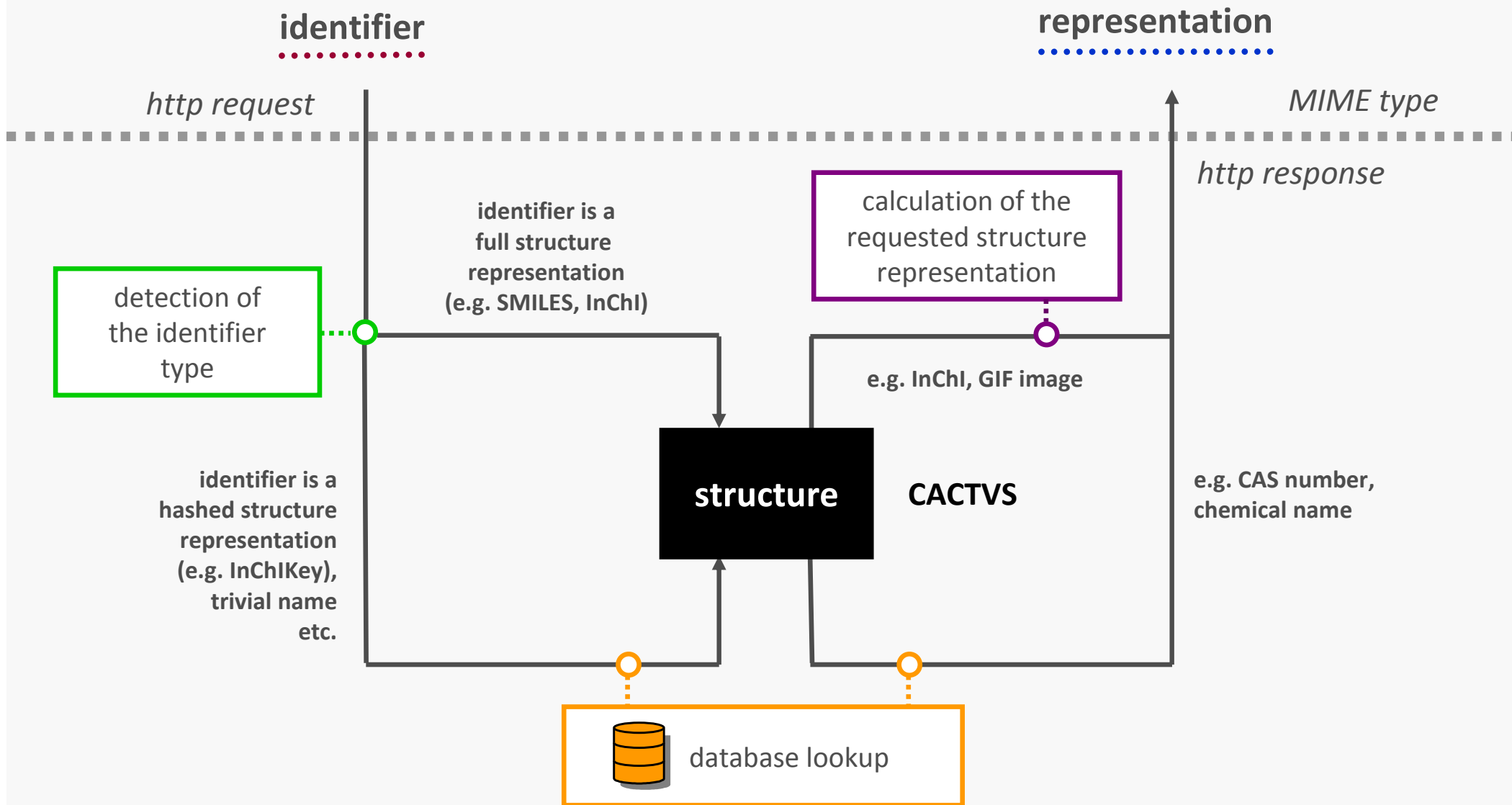
**204255-11-8** MIME type: text/plain

- if a request is not resolvable: HTTP404 status message

# Chemical Identifier Resolver



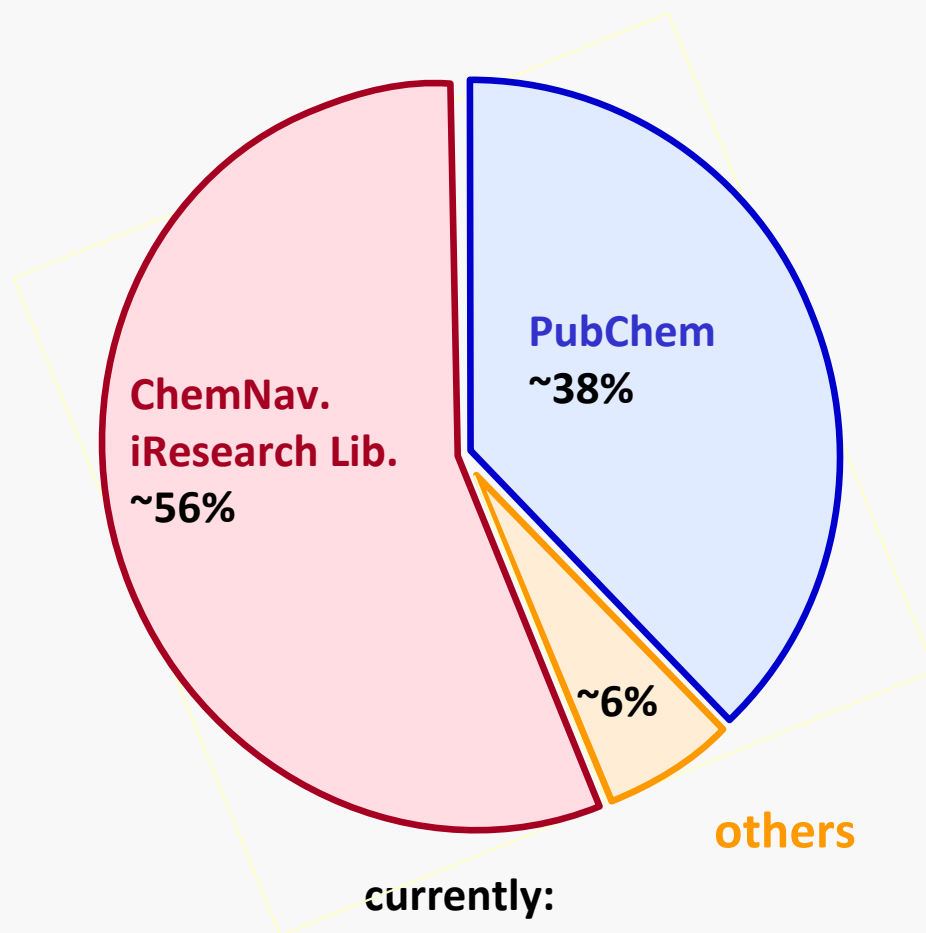
# Chemical Identifier Resolver



NCI/CADD Chemical Structure Database (CSDB)

## Chemical Structure Database (CSDB)

- **ChemNavigator iResearch Library**  
compilation of commercially available screening compounds from ~300 international chemistry suppliers
- **PubChem database**  
including Open NCI database, EPA DSSTox databases, NIAID HIV databases, NIST Webbook, NLM ChemIDplus, ChemSpider ...
- **Commercial Sources / others**  
Asinex, Comgenex, eMolecules, ChEMBL, ...



currently:  
~150 chemical structure databases  
~120 million structure records  
~81.6 million unique structures by  
NCI/CADD FICuS Identifier  
~84 million unique structures by Std. InChIKey

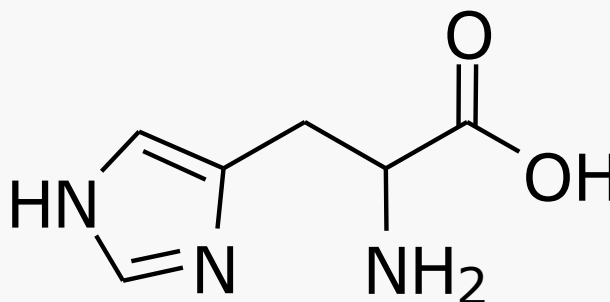
# NCI/CADD Structure Identifiers

**FICTS, FICuS, uuuuu**

Unique Representation of Chemical Structures

## NCI/CADD Structure Identifiers

- based on hashcodes calculated by the chemoinformatics toolkit CACTVS

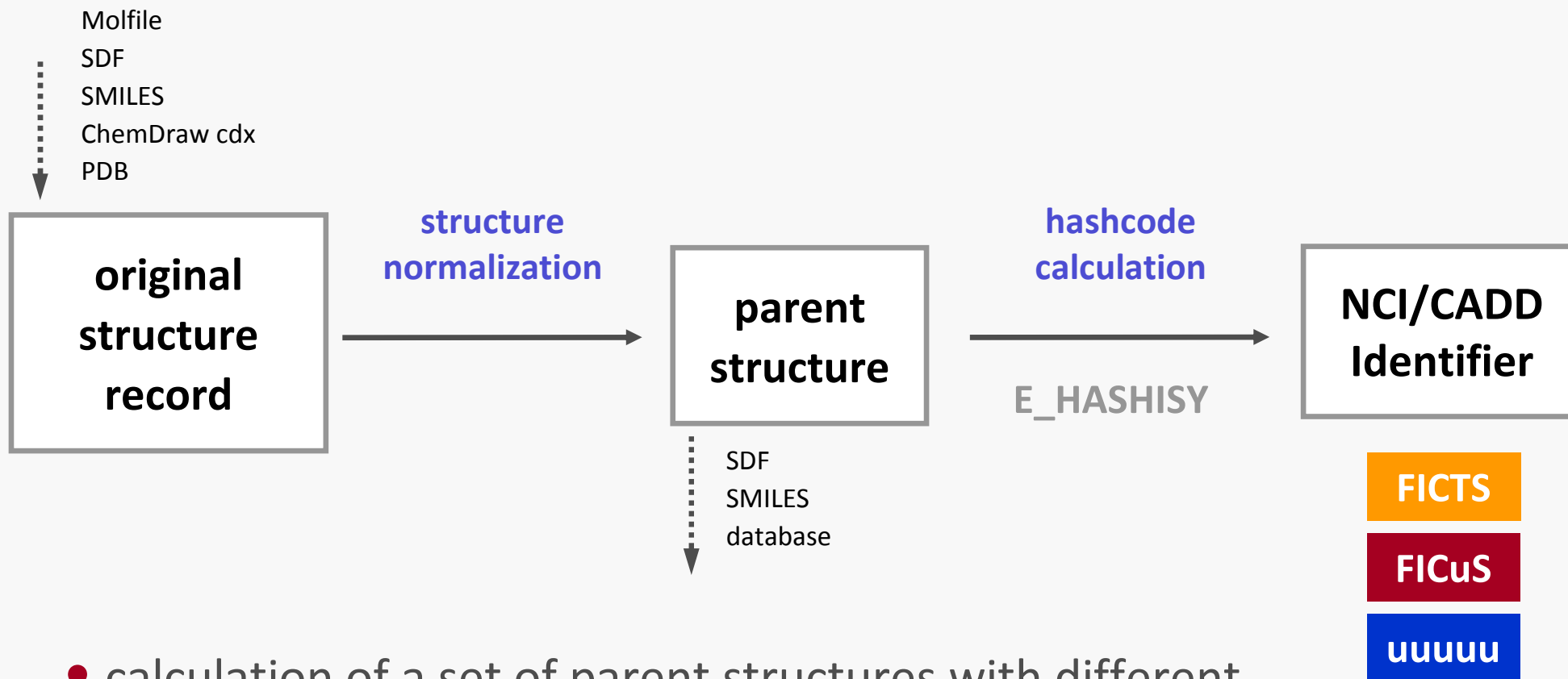


**9850FD9F9E2B4E25**

- **CACTVS hashcodes:**
  - represent a chemical structure uniquely as 16-digit hexadecimal number (64-bit unsigned)
  - high sensitivity to structural features of a compound
  - change if connectivity changes

# Unique Representation of Chemical Structures

## NCI/CADD Structure Identifiers

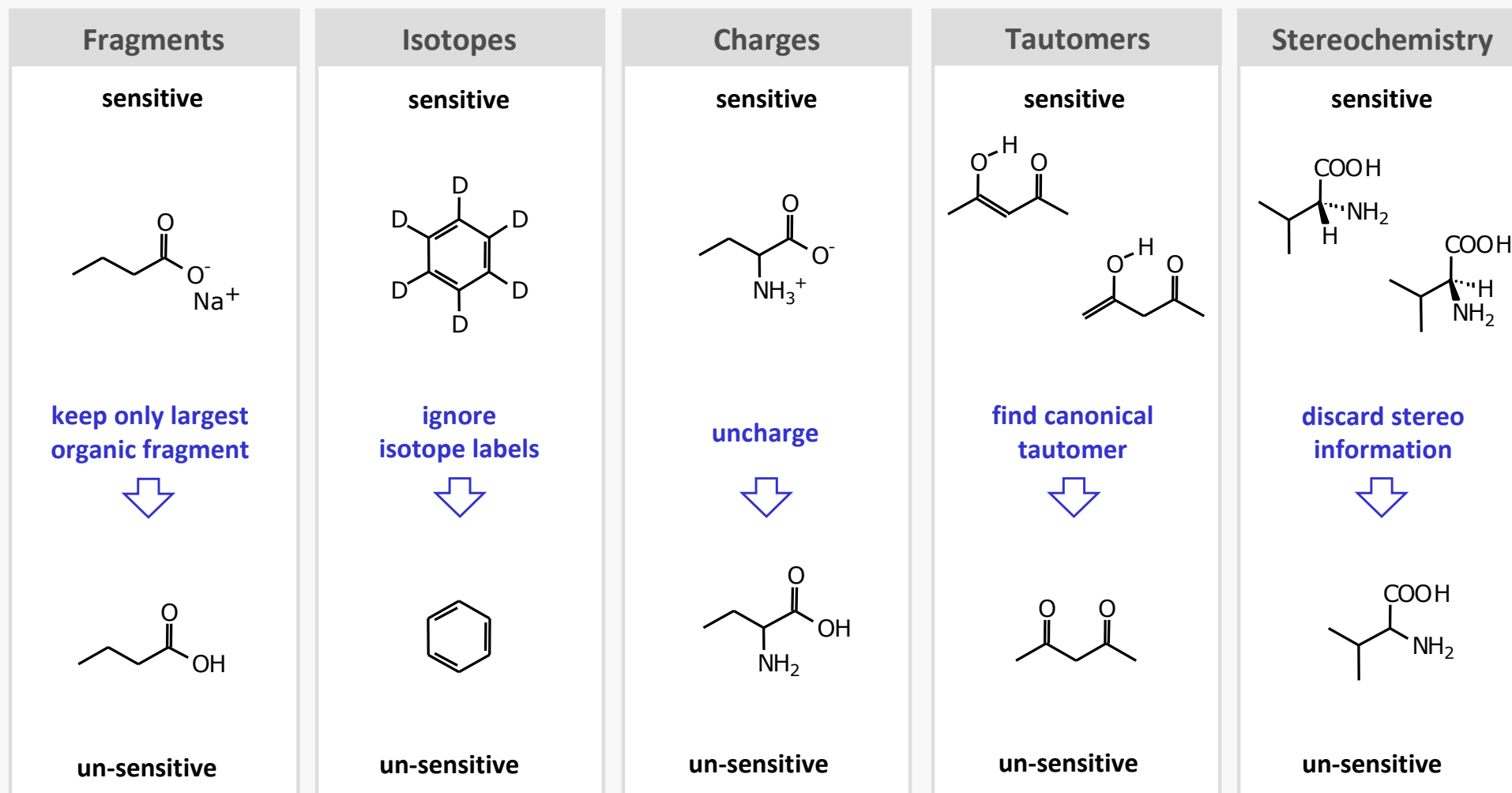


- calculation of a set of parent structures with different sensitivity to chemical features
- representation of chemical structures on different levels

# Unique Representation of Chemical Structures

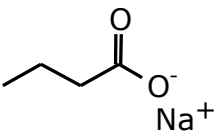
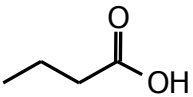
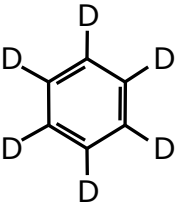

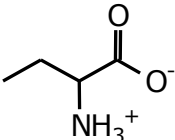
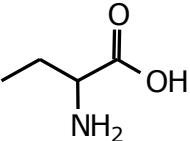
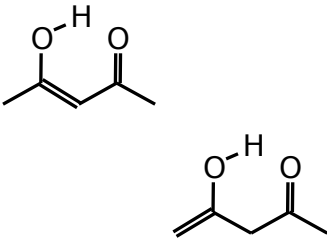
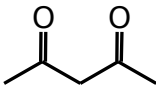
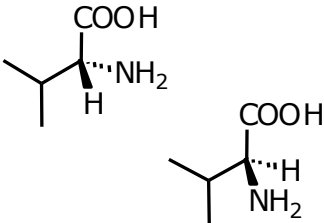
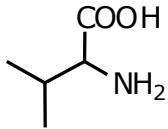
## NCI/CADD Structure Identifiers

- adjustable levels of sensitivity:



# Unique Representation of Chemical Structures

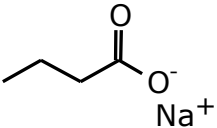
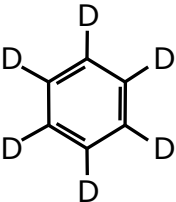
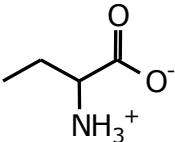
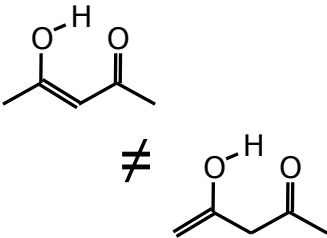
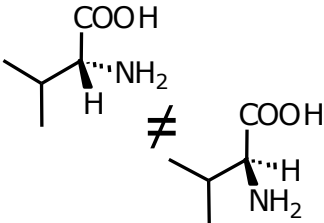
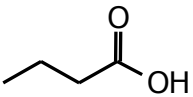

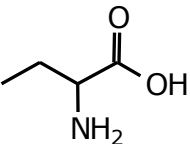
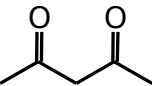
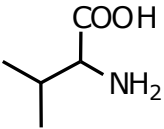
## NCI/CADD Structure Identifiers

Fragments	Isotopes	Charges	Tautomers	Stereochemistry
<p data-bbox="275 555 409 587">sensitive</p>  	<p data-bbox="660 555 795 587">sensitive</p>   <p data-bbox="629 1406 808 1437">un-sensitive</p>	<p data-bbox="1046 555 1180 587">sensitive</p>   <p data-bbox="1016 1406 1196 1437">un-sensitive</p>	<p data-bbox="1431 555 1565 587">sensitive</p>   <p data-bbox="1408 1406 1588 1437">un-sensitive</p>	<p data-bbox="1816 555 1951 587">sensitive</p>   <p data-bbox="1794 1406 1973 1437">un-sensitive</p>

# Unique Representation of Chemical Structures

## NCI/CADD Structure Identifiers

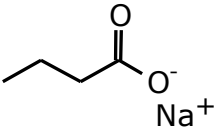
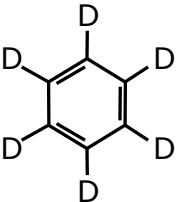
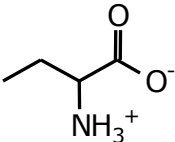
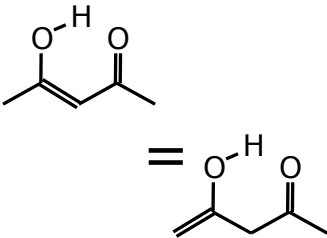
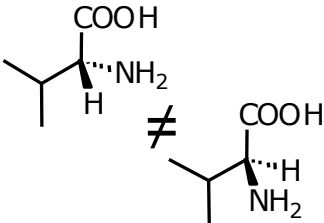
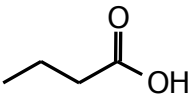

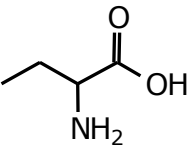
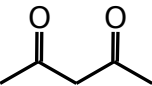
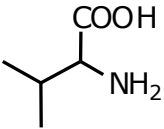
**FICTS** representation of the exact drawing

Fragments	Isotopes	Charges	Tautomers	Stereochemistry
<b>sensitive</b>	<b>sensitive</b>	<b>sensitive</b>	<b>sensitive</b>	<b>sensitive</b>
				
<b>F</b>	<b>I</b>	<b>C</b>	<b>T</b>	<b>S</b>
$\neq$	$\neq$	$\neq$	$\neq$	$\neq$
				
<b>un-sensitive</b>	<b>un-sensitive</b>	<b>un-sensitive</b>	<b>un-sensitive</b>	<b>un-sensitive</b>

# Unique Representation of Chemical Structures

## NCI/CADD Structure Identifiers

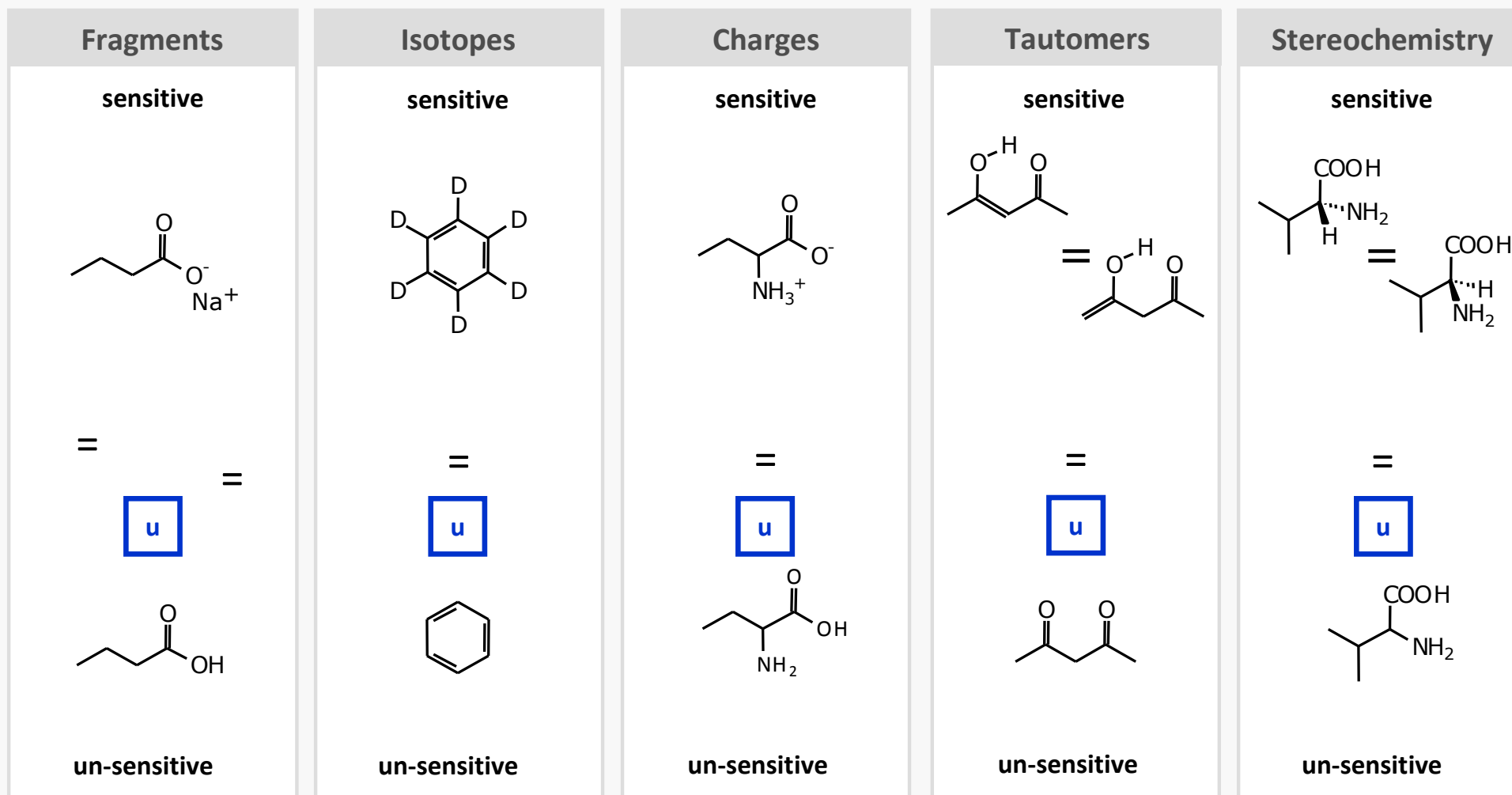
**FICuS** comes closest to how a chemist perceives a compound

Fragments	Isotopes	Charges	Tautomers	Stereochemistry
<b>sensitive</b>	<b>sensitive</b>	<b>sensitive</b>	<b>sensitive</b>	<b>sensitive</b>
				
<b>F</b>	<b>I</b>	<b>C</b>	<b>u</b>	<b>S</b>
$\neq$	$\neq$	$\neq$	$=$	$\neq$
				
<b>un-sensitive</b>	<b>un-sensitive</b>	<b>un-sensitive</b>	<b>un-sensitive</b>	<b>un-sensitive</b>

# Unique Representation of Chemical Structures

## NCI/CADD Structure Identifiers

**uuuuu** closely related forms of the same compound

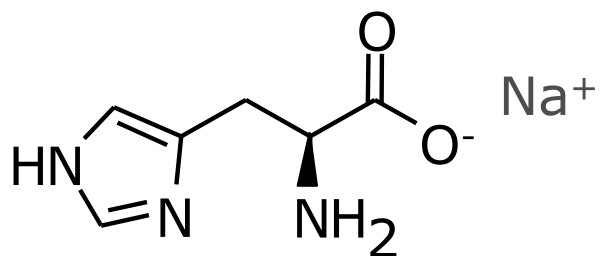


# Unique Representation of Chemical Structures

## NCI/CADD Structure Identifiers

● sensitive / ● not sensitive

	Fragments	Isotopes	Charges	Tautomers	Stereo
FICTS	●	●	●	●	●
FICuS	●	●	●	●	●
uuuuu	●	●	●	●	●

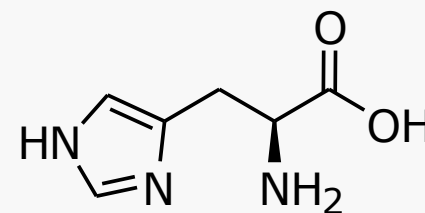
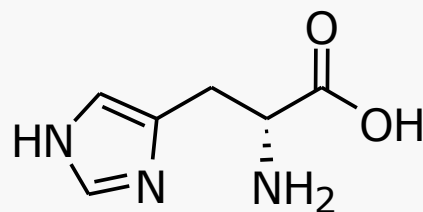
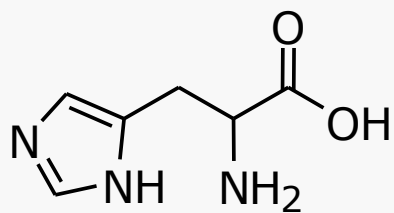


4A122D094098B50D-FICTS-01-1D

0E26B623DF7FAD30-FICuS-01-70

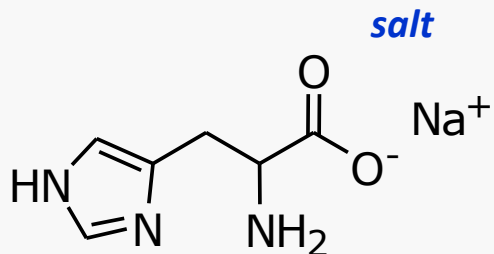
9850FD9F9E2B4E25-uuuuu-01-27

<CACTVS hashcode (E\_HASHISY)>-<tag>-<version>-<checksum>

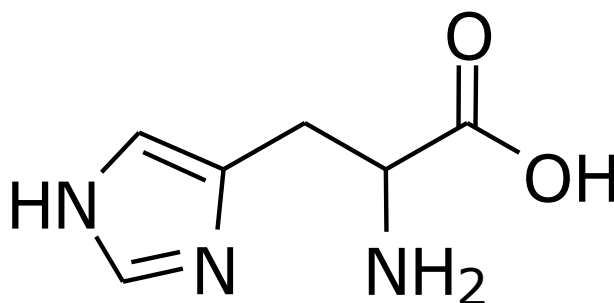


*tautomer*

*stereoisomers*

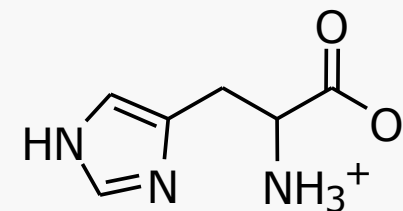


*salt*

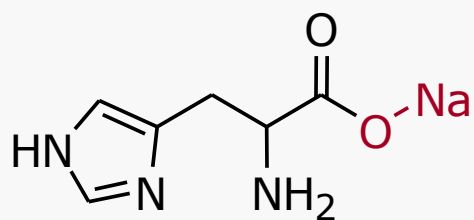


*histidine*

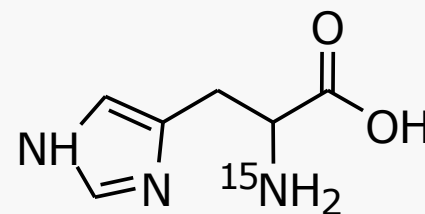
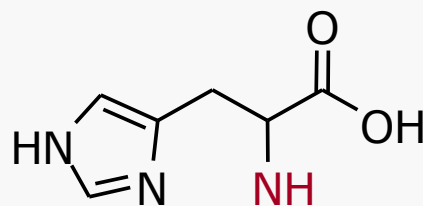
*charged form*

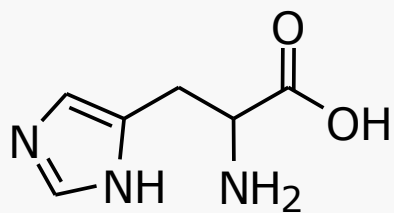


*isotope*



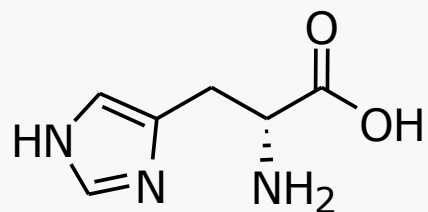
*"errors"*





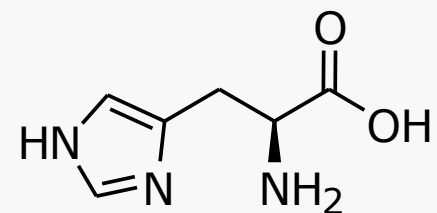
6C16DE2351F9FF50-FICTS

*tautomer*

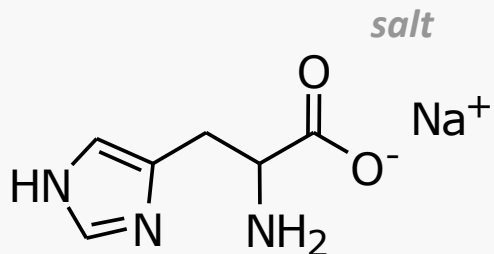


E92E4BA2869F3611-FICTS

*stereoisomers*

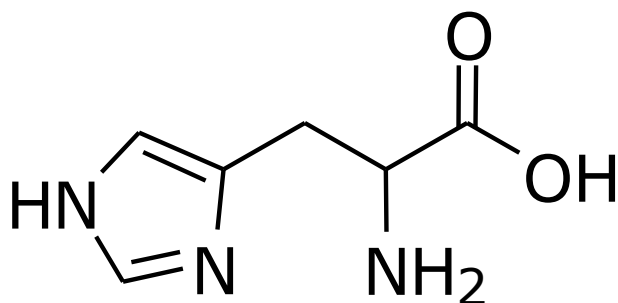


8A7AD1EB498CC76A-FICTS



E5F83F10C5DB080A-FICTS

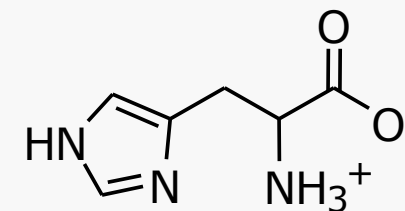
*salt*



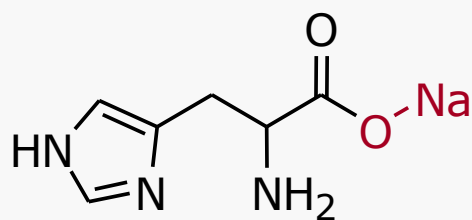
9850FD9F9E2B4E25-FICTS

FICTS

*charged form*

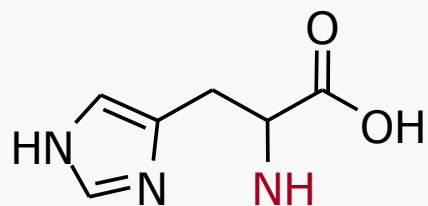


A3DAE0788050DDE4-FICTS



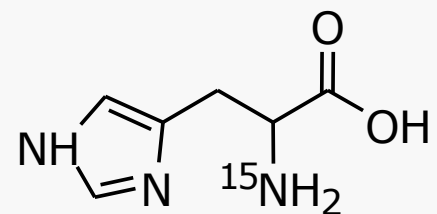
E5F83F10C5DB080A-FICTS

*"errors"*

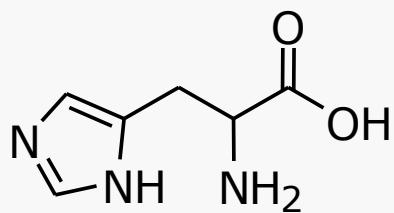


9850FD9F9E2B4E25-FICTS

*isotope*

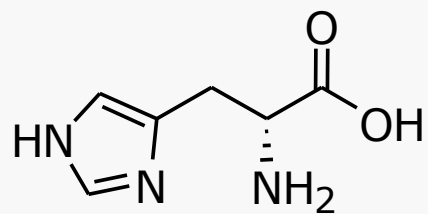


B2FDA68AEDA06DB9-FICTS



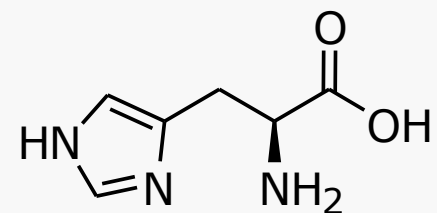
9850FD9F9E2B4E25-FICuS

*tautomer*

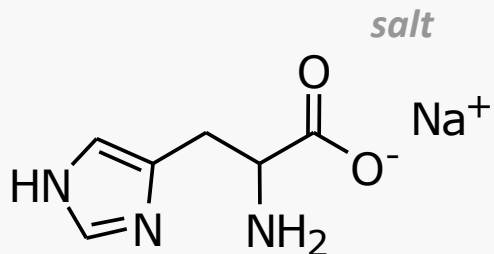


E92E4BA2869F3611-FICuS

*stereoisomers*

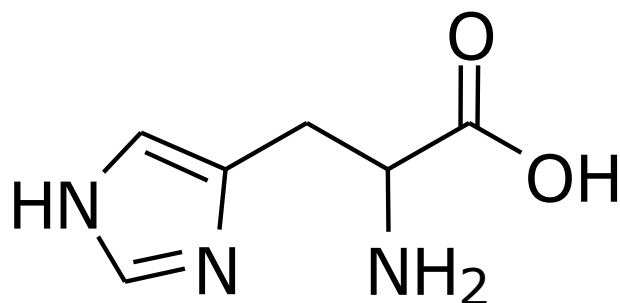


8A7AD1EB498CC76A-FICuS



E5F83F10C5DB080A-FICuS

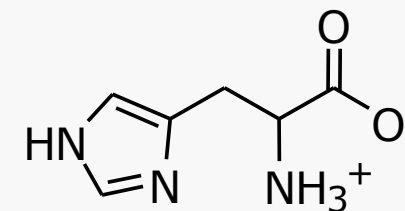
*salt*



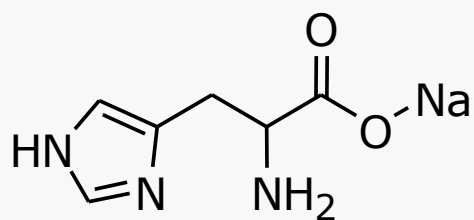
9850FD9F9E2B4E25-FICuS

**FICuS**

*charged form*

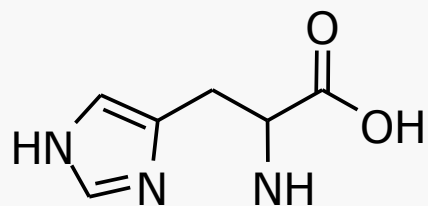


A3DAE0788050DDE4-FICuS



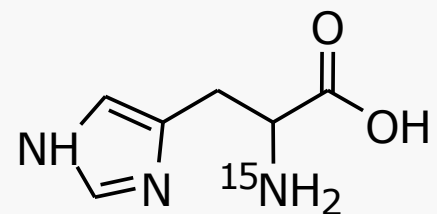
E5F83F10C5DB080A-FICuS

*"errors"*

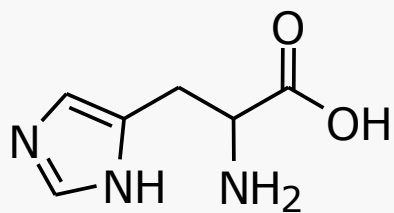


9850FD9F9E2B4E25-FICuS

*isotope*

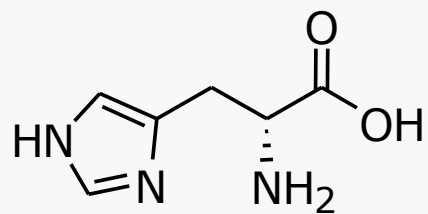


B2FDA68AEDA06DB9-FICuS



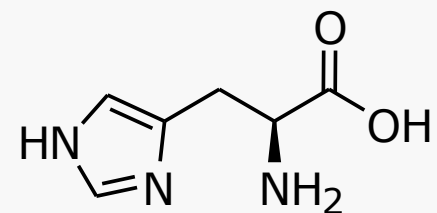
9850FD9F9E2B4E25-uuuuu

*tautomer*

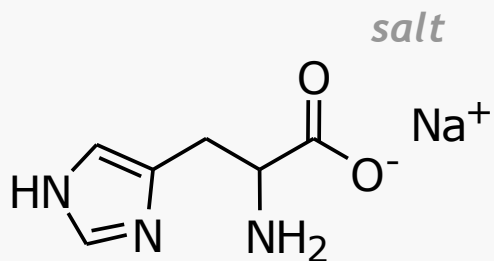


9850FD9F9E2B4E25-uuuuu

*stereoisomers*

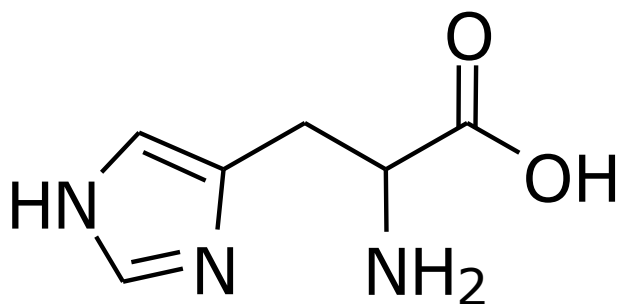


9850FD9F9E2B4E25-uuuuu



9850FD9F9E2B4E25-uuuuu

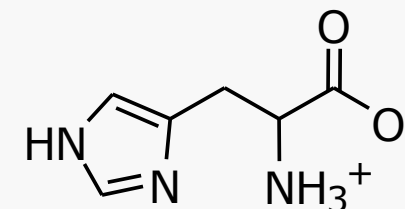
*salt*



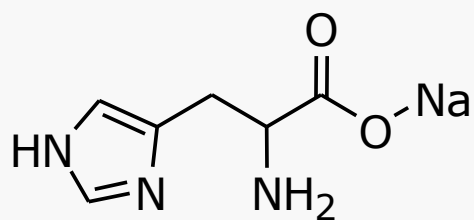
9850FD9F9E2B4E25-uuuuu

uuuuu

*charged form*

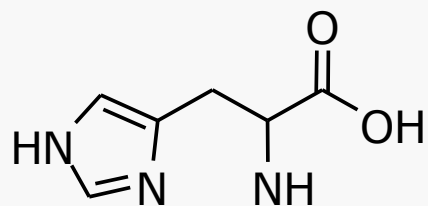


9850FD9F9E2B4E25-uuuuu



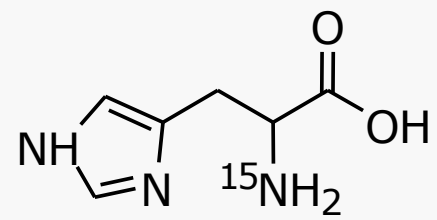
9850FD9F9E2B4E25-uuuuu

*“errors”*

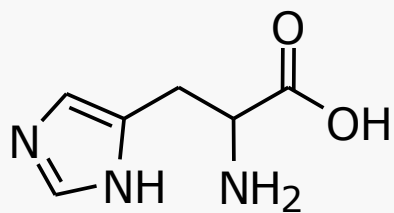


9850FD9F9E2B4E25-FICuS

*isotope*

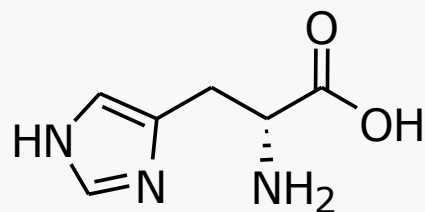


9850FD9F9E2B4E25-uuuuu



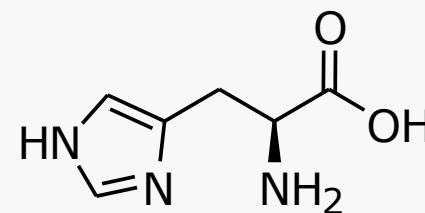
HNDVDQJICGZPNO-UHFFFAOYSA-N

*tautomer*

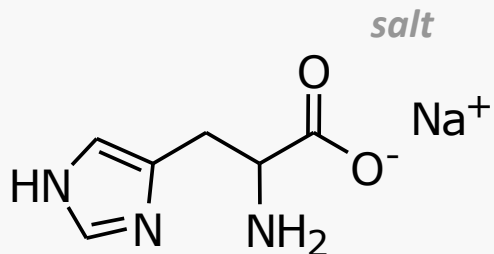


HNDVDQJICGZPNO-RXMQYKEDSA-N

*stereoisomers*

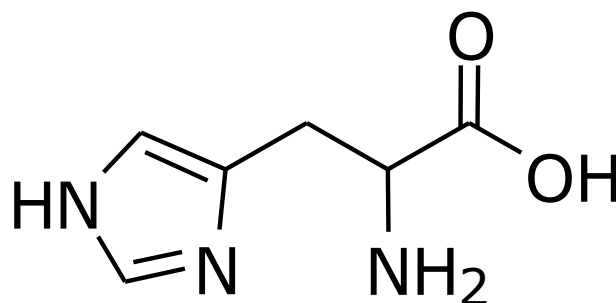


HNDVDQJICGZPNO-YFKPBYRVSA-N



UHPNKBYGGMJTIM-UHFFFAOYSA-M

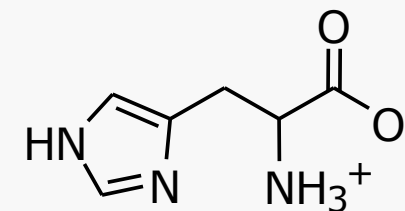
*salt*



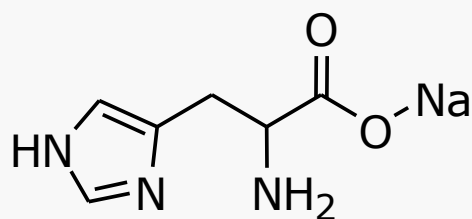
HNDVDQJICGZPNO-UHFFFAOYSA-N

Std. InChIKey

*charged form*

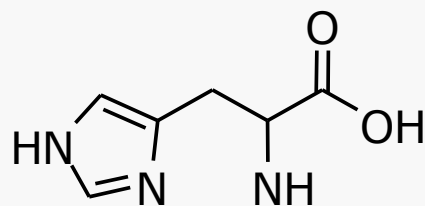


HNDVDQJICGZPNO-UHFFFAOYSA-N



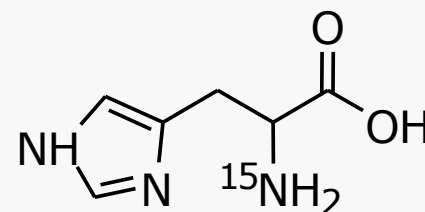
UHPNKBYGGMJTIM-UHFFFAOYSA-M

*“errors”*



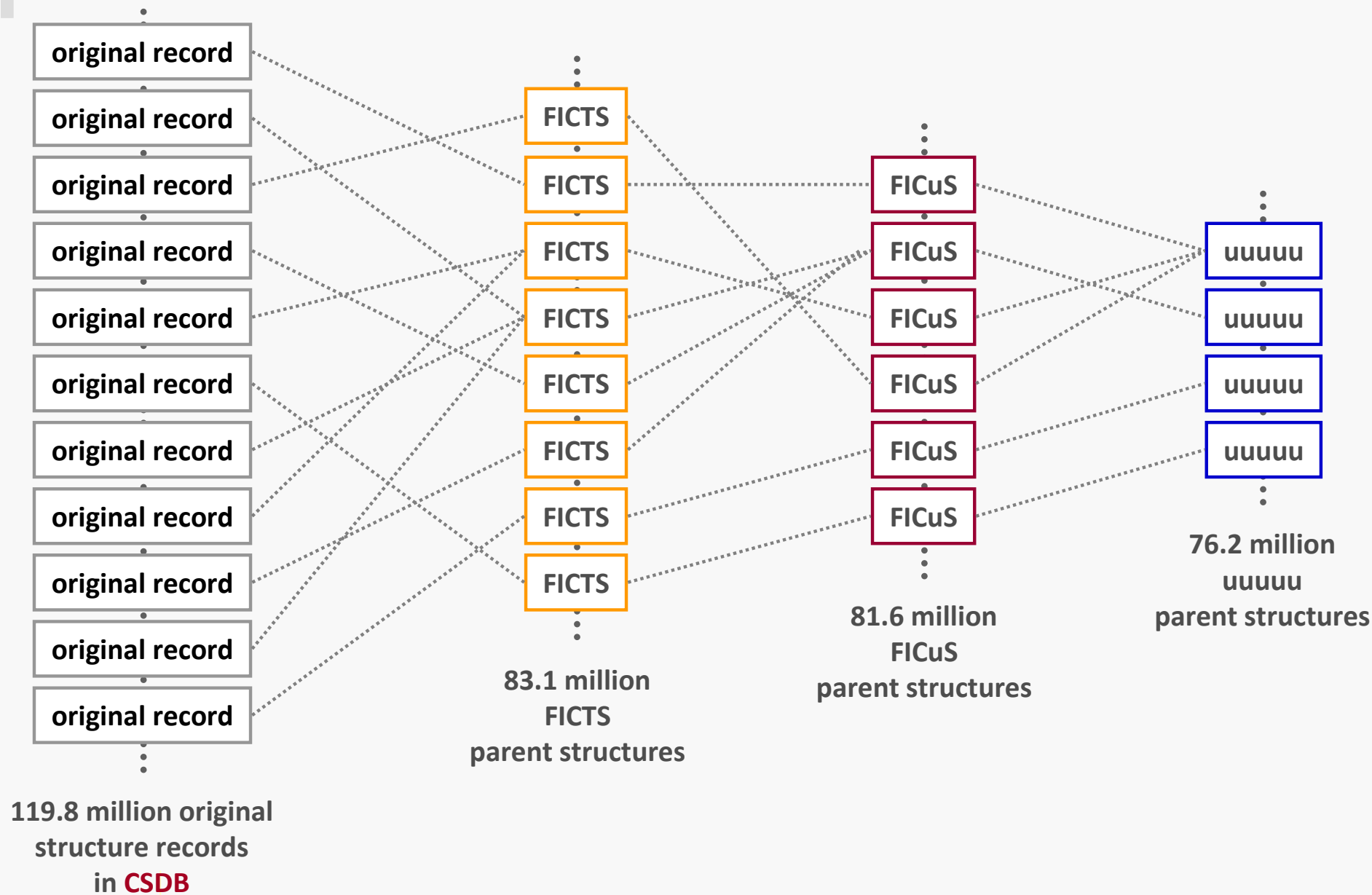
HNDVDQJICGZPNO-UHFFFAOYSA-N

*isotope*

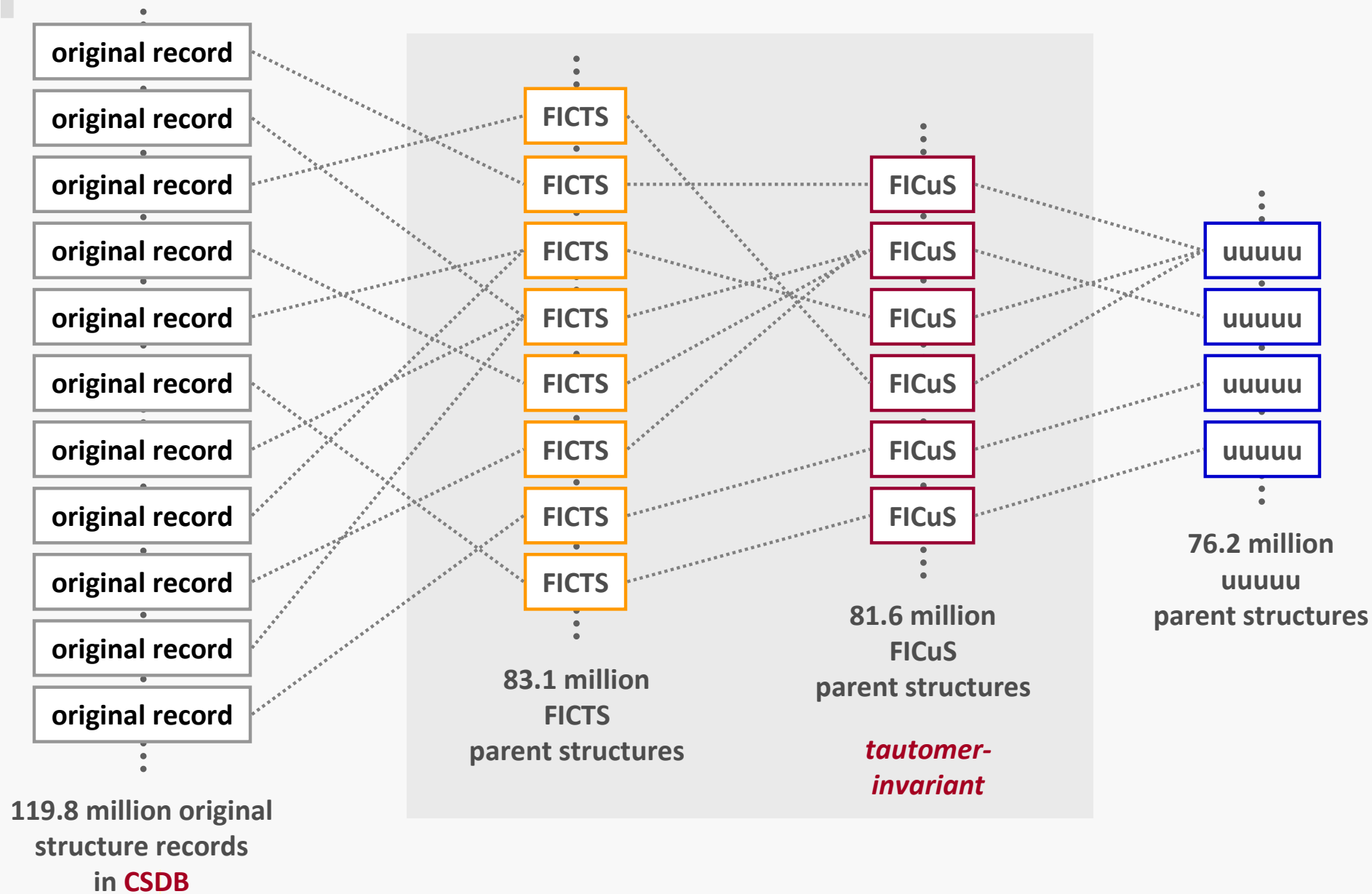


HNDVDQJICGZPNO-CDZYAPPSA-N

# Structure Normalization



# Structure Normalization



# Tautomer Analysis

How much “chemical space” is “just generated” by drawing tautomers?

## Tautomer Analysis

- CACTVS: generation of all formal tautomers for a given organic compound (prototropic tautomerism)
- rule set of 21 transforms encoded as (CACTVS-extended) SMIRKS
- rule set is systematically applied to the original structure (and all tautomers that have been generated in previous steps)
- tautomer generation is limited to 1000 SMIRKS transform operations/structure
- all tautomers are ranked by a scoring function
- the highest ranked tautomer is defined as the **canonical tautomer**

## Tautomer Analysis

- 21 SMIRKS transform rules:

**rule 1:** 1.3 (thio)keto/(thio)enol

**rule 2:** 1.5 (thio)keto/(thio)enol

**rule 3:** simple (aliphatic) imine

**rule 4:** special imine

**rule 5:** 1.3 aromatic heteroatom H shift

**rule 6:** 1.3 heteroatom H shift

**rule 7:** 1.5 (aromatic) heteroatom H shift (1)

**rule 8:** 1.5 aromatic heteroatom H shift (2)

**rule 9:** 1.7 (aromatic) heteroatom H shift

**rule 10:** 1.9 (aromatic) heteroatom H shift

**rule 11:** 1.11 (aromatic) heteroatom H shift

**rule 12:** furanones

**rule 13:** keten/ynol exchange

**rule 14:** ionic nitro/aci-nitro

**rule 15:** pentavalent nitro/aci-nitro

**rule 16:** oxim/nitroso

**rule 17:** oxim/nitroso via phenol

**rule 18:** cyanic/iso-cyanic acids

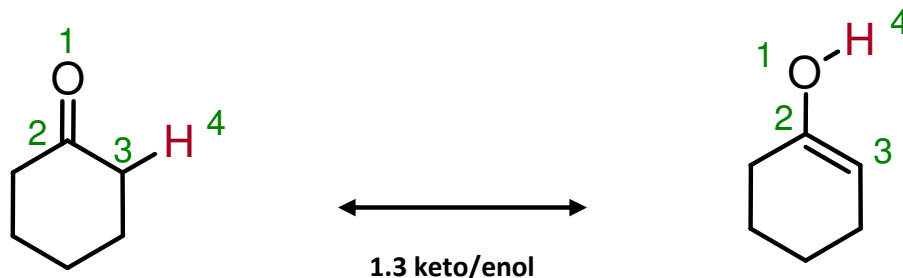
**rule 19:** formamidinesulfinic acids

**rule 20:** isocyanides

**rule 21:** phosphonic acids

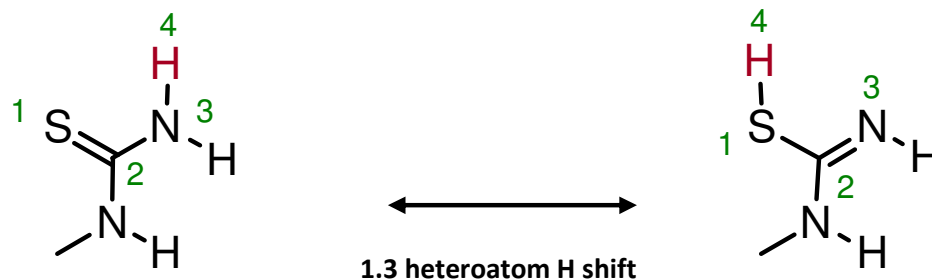
# Tautomer Analysis

**rule 1:** 1.3 (thio)keto/(thio)enol



$[O, S, Se, Te; X1:1] = [C; z\{1-2\}:2] [CX4R\{0-2\}:3] [\#1:4] \gg$   
 $[\#1:4] [O, S, Se, Te; X2:1] [\#6; z\{1-2\}:2] = [C, cz\{0-1\}R\{0-1\}:3]$

**rule 6:** 1.3 heteroatom H shift

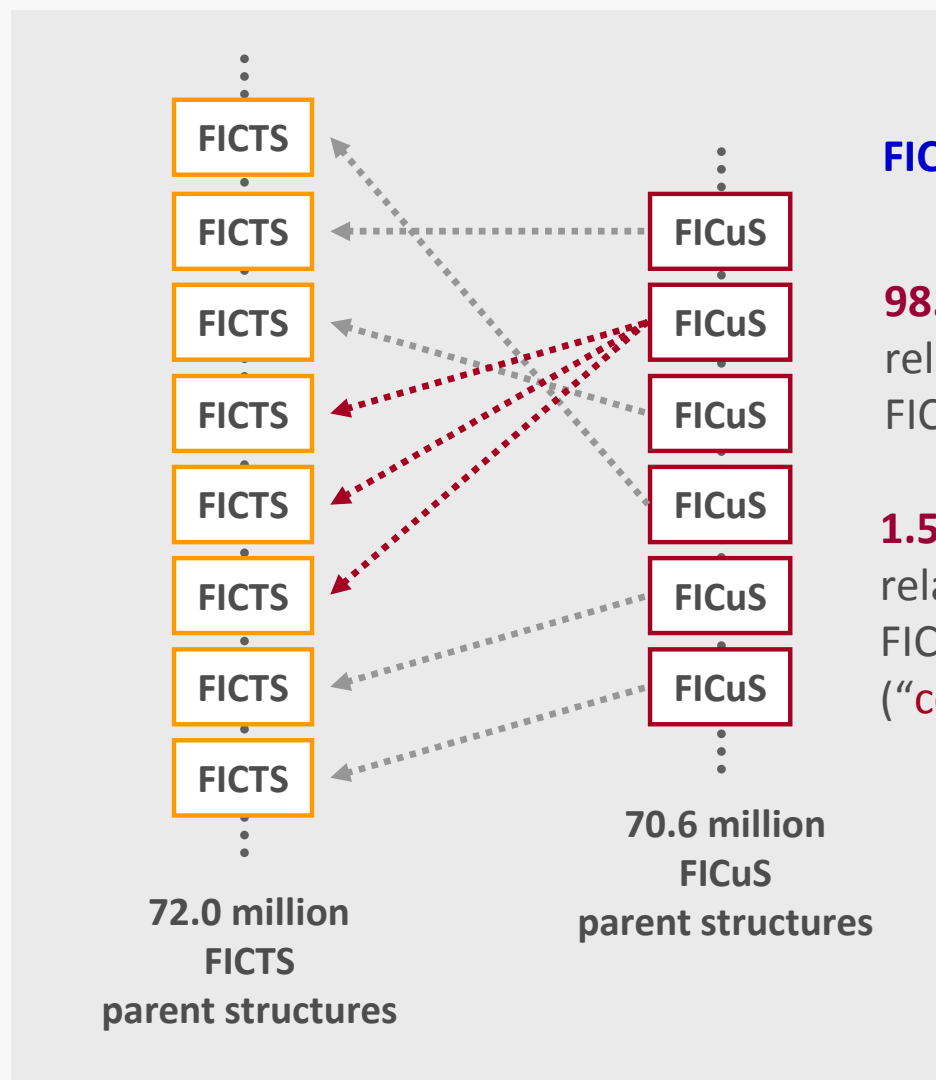


$[N, n, S, s, O, o, Se, Te:1] = [NX2, nX2, C, c, P, p:2] [N, n, S, O, Se, Te:3] [\#1:4] \gg$   
 $[\#1:4] [N, n, S, O, Se, Te:1] [NX2, nX2, C, c, P, p:2] = [N, n, S, s, O, o, Se, Te:3]$

# Tautomer Analysis

## FICTS parent structures

**8.6%** change tautomeric form during FICuS normalization



## FICuS parent structures

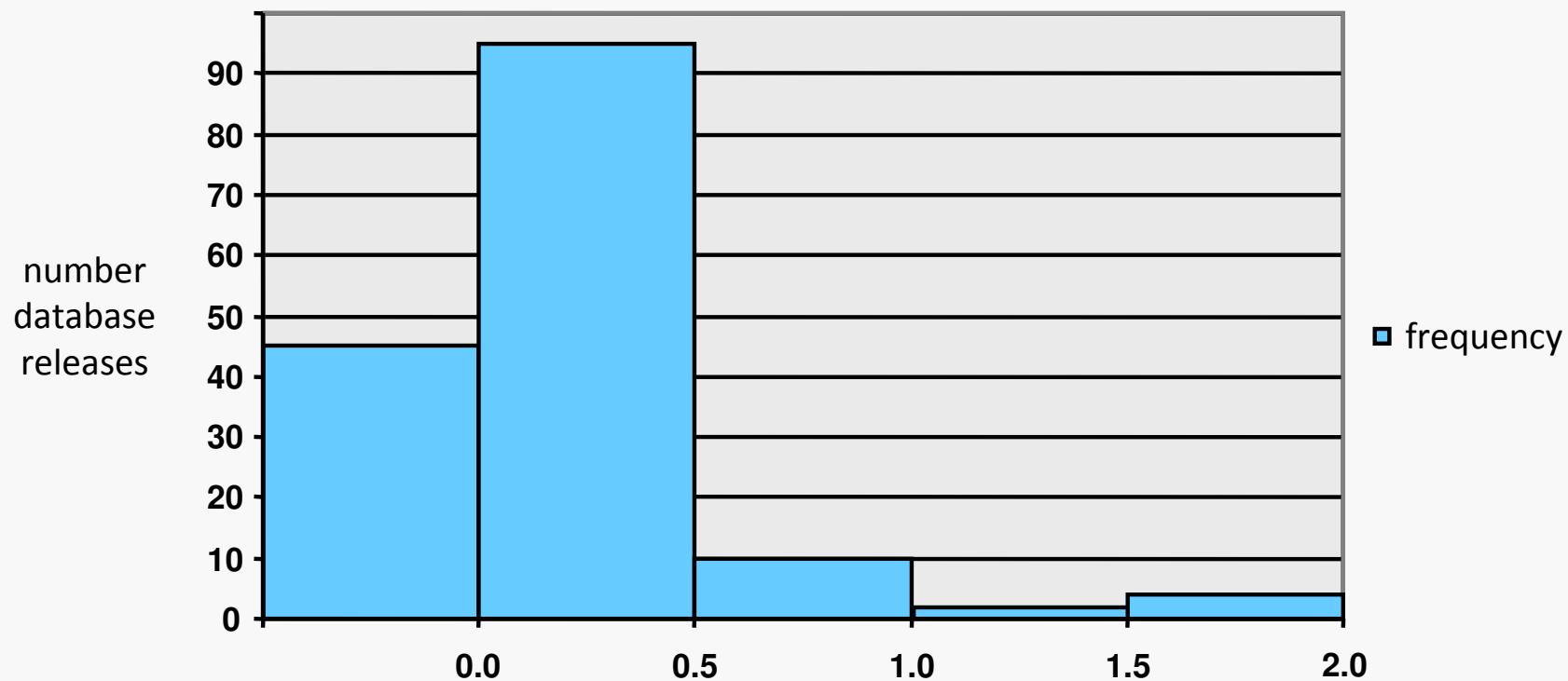
**98.5%** have an one-to-one relationship to a single FICTS parent structure

**1.5%** have an one-to-many relationship to several FICTS parent structures (“**conflict**”)

**structure counts are on basis of the 2009 version of CSDB (103.9 million structure records)**

# Tautomer Analysis

tautomeric overlap within each individual database release (%)

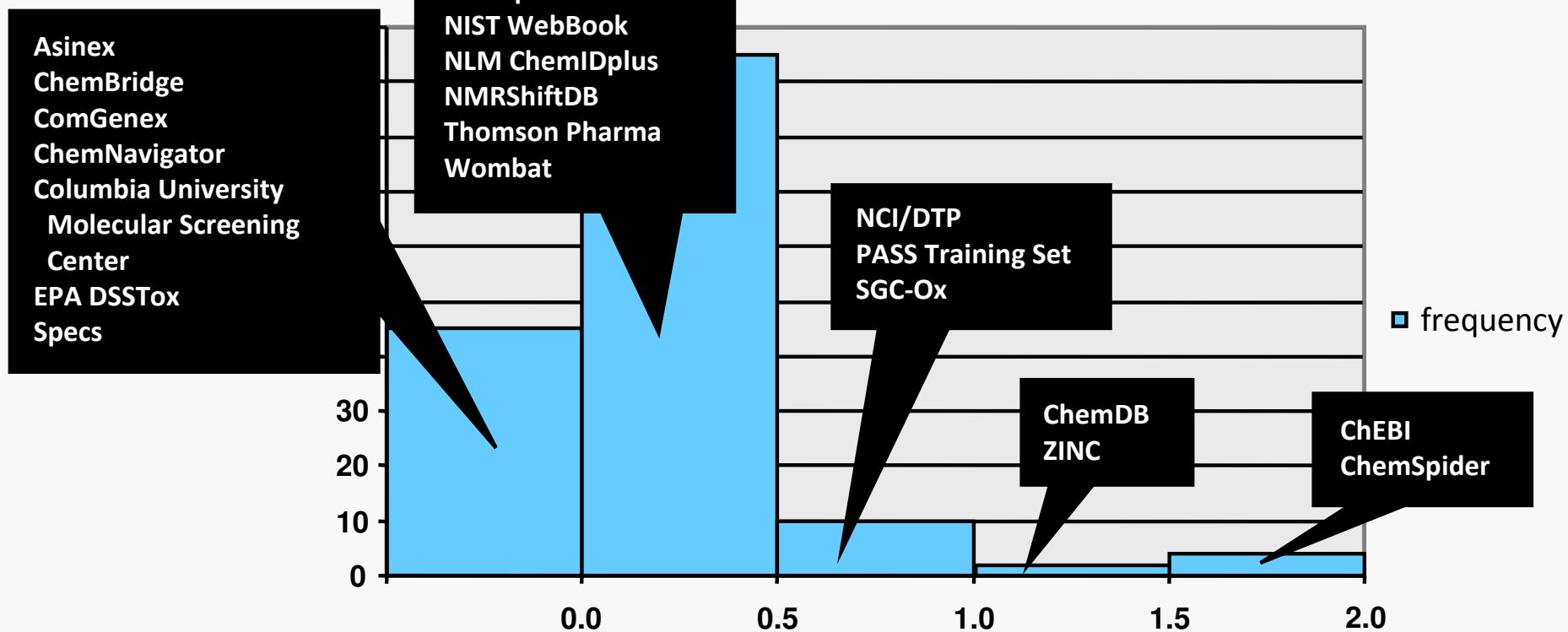


average: **~0.3% of original structure records**

# NCI/CADD Chemical Structure Database

## Tautomer A

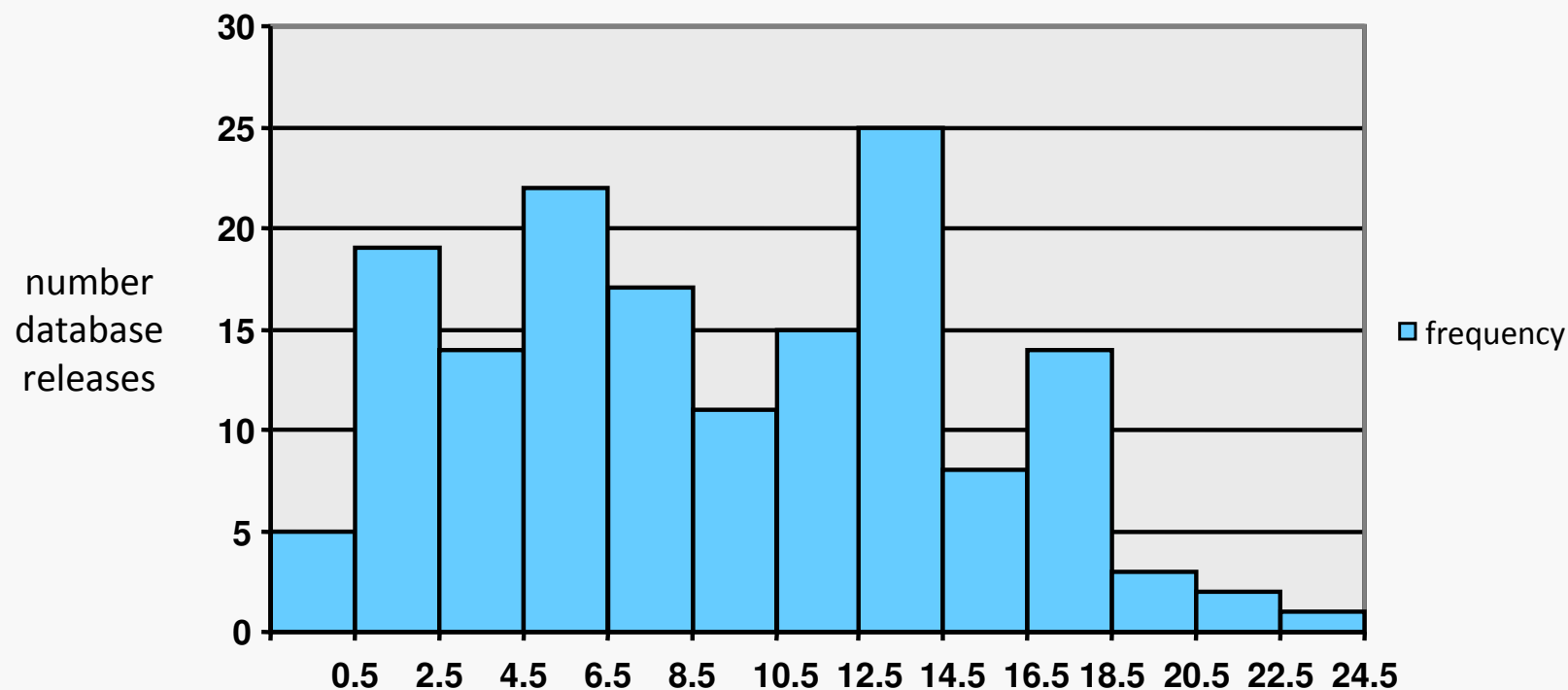
tautomer A in each individual database release (%)



average: ~0.3% of original structure records

## Tautomer Analysis

occurrence of “tautomerism-critical” molecules within each individual database release (%)

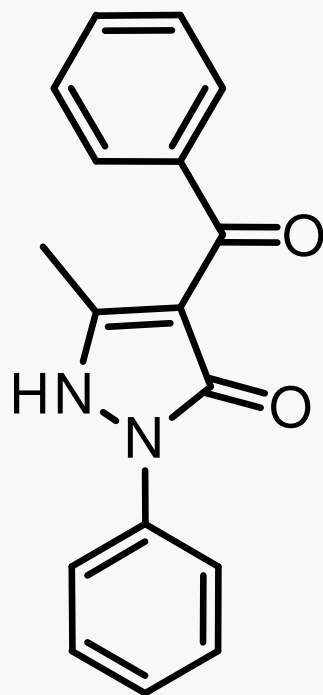


**average: ~9.5% of FICuS parent structures**

percentage of FICuS parent structure in each database release occurring somewhere in CSDB with a conflict

Example for a Tautomer “Conflict”

## HPMBP (1-Phenyl-3-methyl-4-benzoyl-pyrazolone-5)



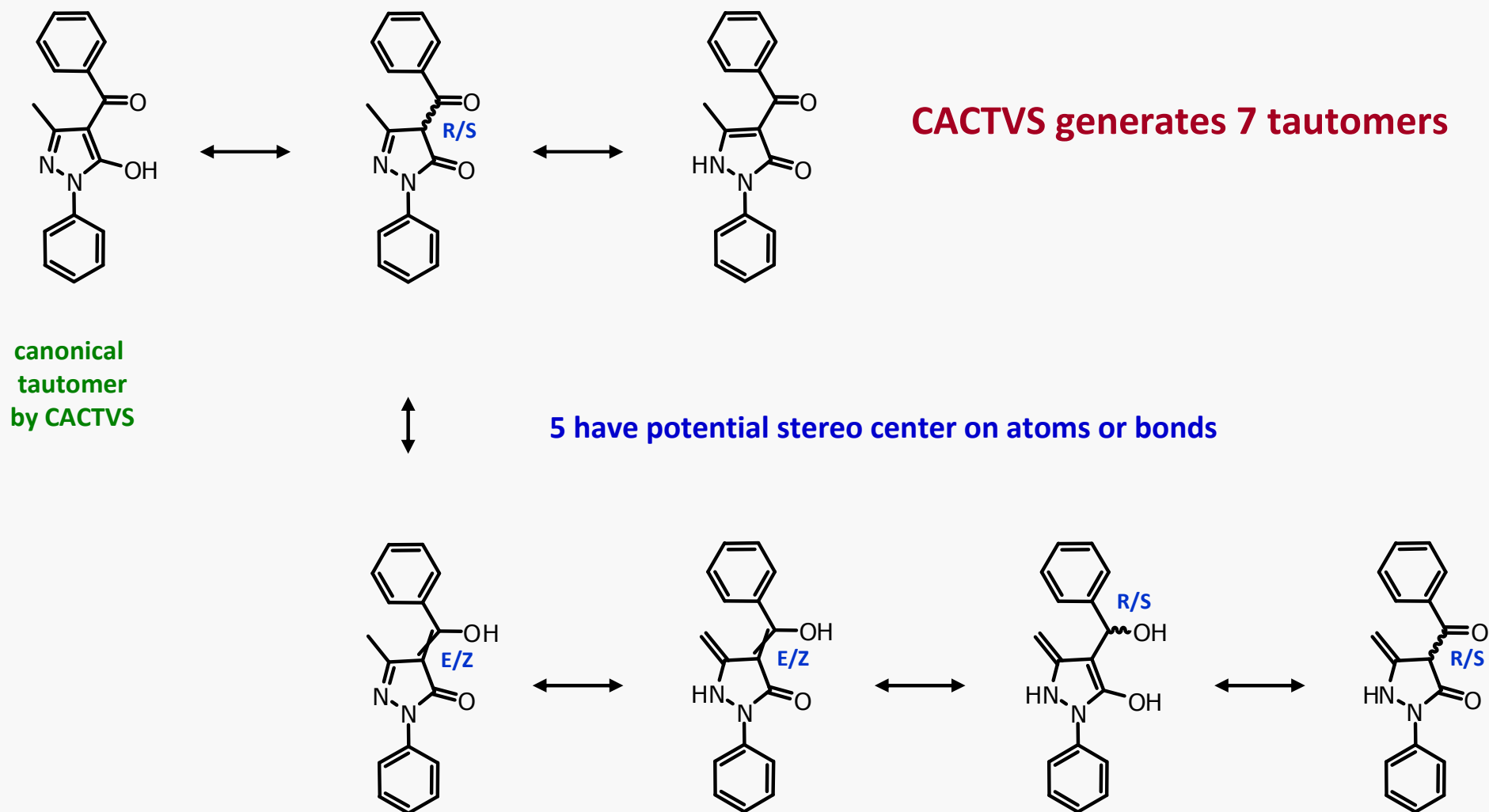
- HPMBP is used in liquid membranes (selective removal of metal ions)
- selectivity and efficiency depends on the tautomeric form of HPMBP
- the tautomeric form depends on solvent and concentration of HPMBP

He, D.; Li Z.; Ma M.; Huang J.; Yang Y. Study of extraction characteristics of HPMBP.

1. Tautomer and extraction characteristics. J. Chem. Eng. Data **2009**, 54(10), 2944-2947

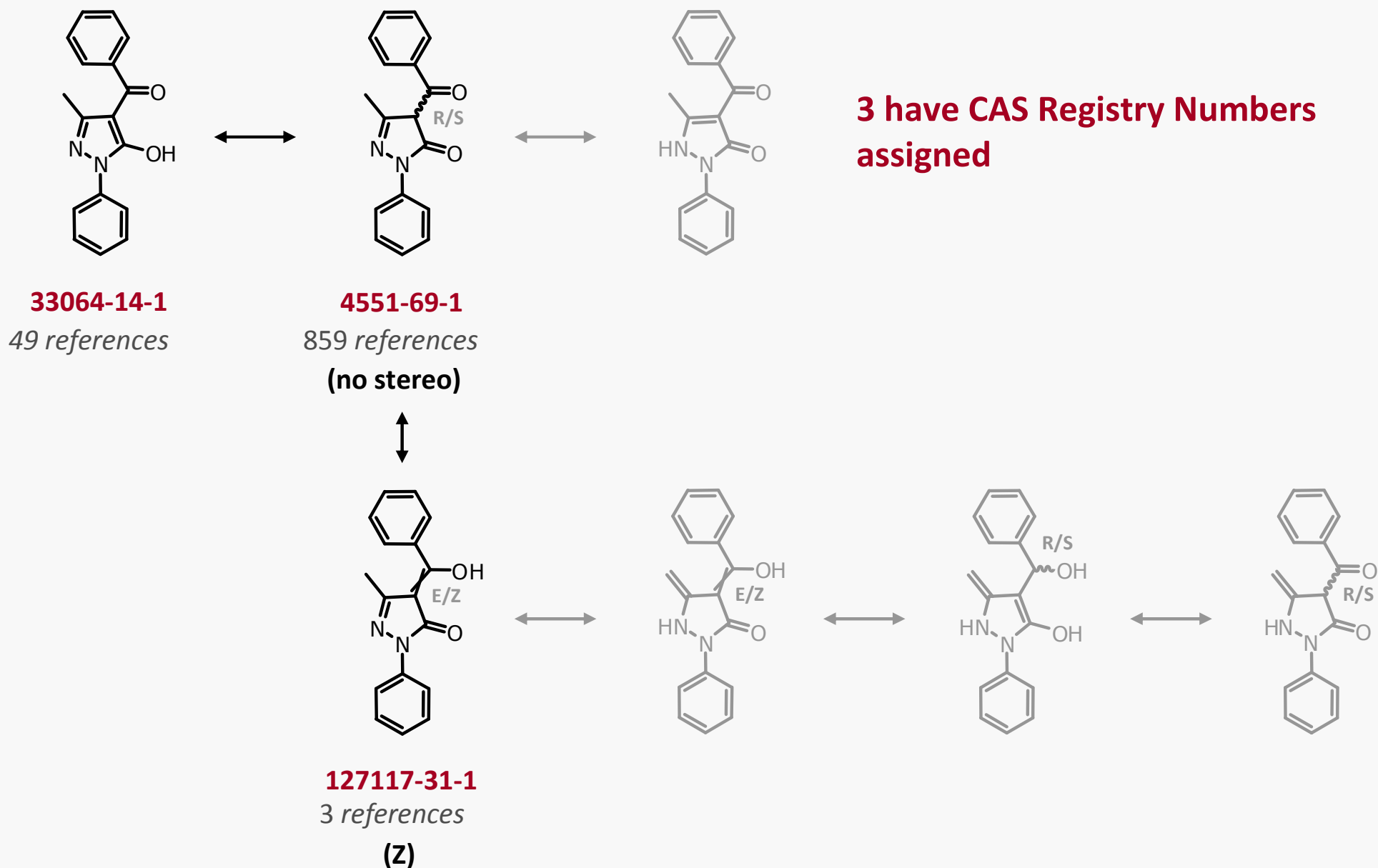
Example for a Tautomer "Conflict"

## HPMBP (1-Phenyl-3-methyl-4-benzoyl-pyrazolone-5)



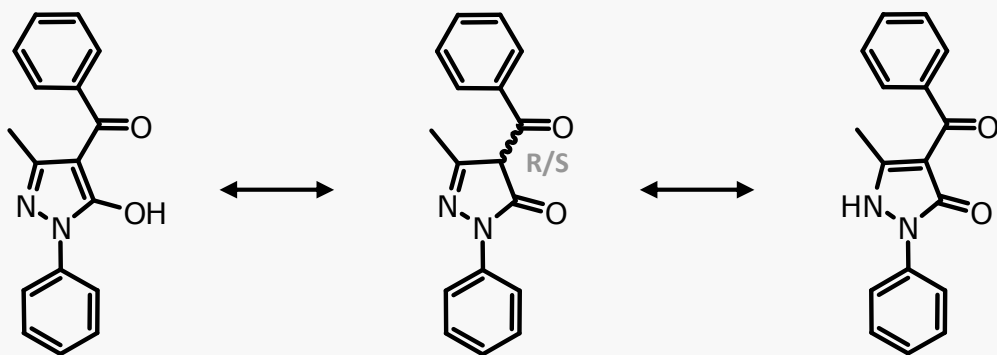
Example for a Tautomer “Conflict”

## HPMBP (1-Phenyl-3-methyl-4-benzoyl-pyrazolone-5)



Example for a Tautomer "Conflict"

## HPMBP (1-Phenyl-3-methyl-4-benzoyl-pyrazolone-5)



**occurrences in databases  
indexed in CSDB**

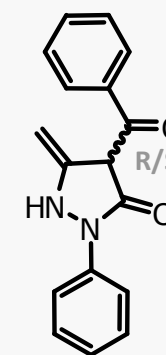
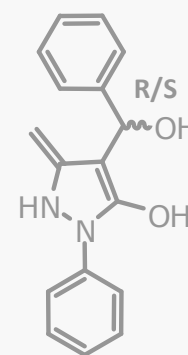
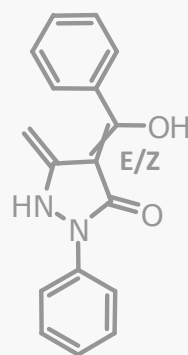
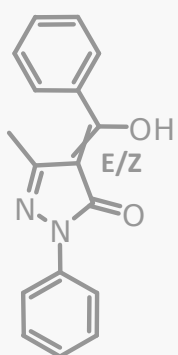
**12 databases**

**16 databases** (no stereo)

**3 databases** (R)

**2 databases** (S)

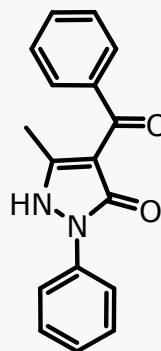
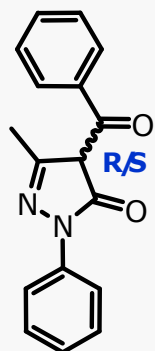
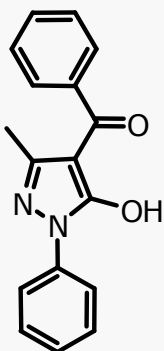
**6 databases**



**1 database**  
(no stereo)

Example for a Tautomer "Conflict"

## HPMBP (1-Phenyl-3-methyl-4-benzoyl-pyrazolone-5)



occurs in **10 databases**

Ambinter  
ChemDB  
ChemSpider  
DiscoveryGate  
ChemNavigator  
Thomson Pharma

**12 databases**

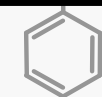
ACD 3D  
Ambinter  
BindingDB  
ChemBank  
ChemDB  
ChemSpider  
ChemNavigator  
MLSMR  
NIAID  
Scripps Screening Center  
Thomson Pharma  
ZINC

**16 databases (no stereo)**

**3 databases (R)**  
**2 databases (S)**

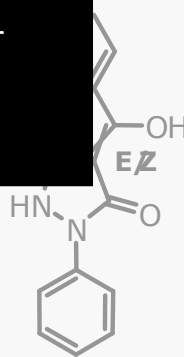


ChemSpider  
ZINC



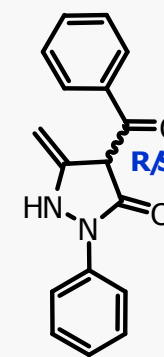
**6 databases**

ChemSpider  
ECOTOX  
ZINC



ACD 3D  
ACX  
Ambinter  
BioByte QSAR  
ChemBank  
ChemBridge  
ChemDB  
ChemSpider  
DiscoveryGate  
EPA GCES  
MLSMR  
NCI Open Database  
NIST MS-Lib  
NLM ChemIDplus  
Sigma-Aldrich  
Thomson Pharma

ChemDB

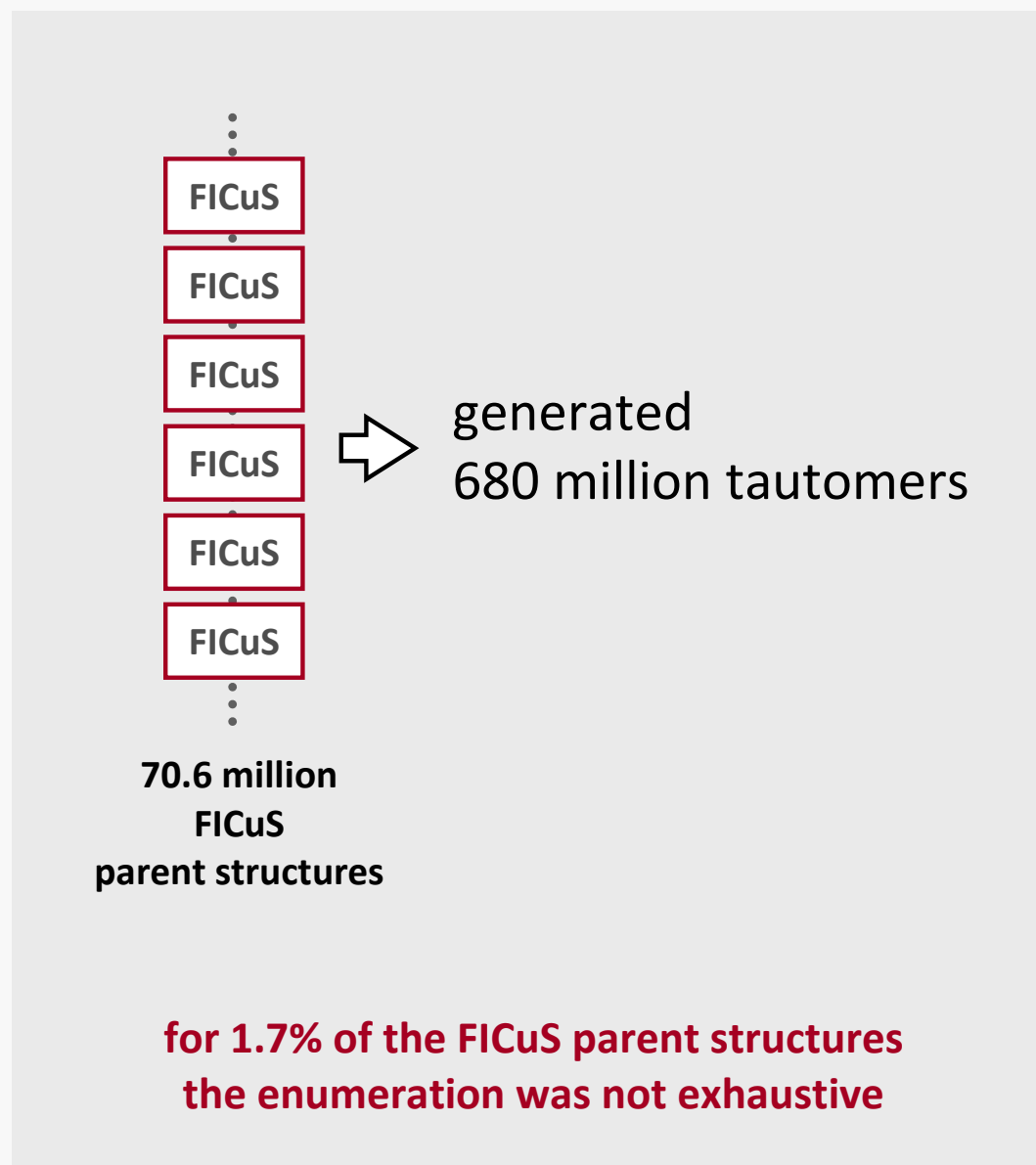


**1 database**  
(no stereo)

## Tautomer Analysis

starting from the set of FICuS parent structures we systematically generated all tautomers based on the 21 SMIRKS rule set available in CACTVS

- how many tautomers are generated?
- how often is each rule applied (type of tautomerism)?
- how many tautomers per structure?



## Tautomer Analysis

- usage of SMIRKS rules (1/2):

tautomer rule	generated tautomers	
	count	%
<b>rule 1:</b> 1.3 (thio)keto/(thio)enol	173,002,712	<b>25.4</b>
<b>rule 2:</b> 1.5 (thio)keto/(thio)enol	11,541,452	1.7
<b>rule 3:</b> simple (aliphatic) imine	35,917,415	5.3
<b>rule 4:</b> special imine	4,306,155	0.6
<b>rule 5:</b> 1.3 aromatic heteroatom H shift	25,678,446	3.8
<b>rule 6:</b> 1.3 heteroatom H shift	250,453,882	<b>36.8</b>
<b>rule 7:</b> 1.5 (aromatic) heteroatom H shift (1)	27,542,770	4.0
<b>rule 8:</b> 1.5 aromatic heteroatom H shift (2)	26,819	<0.1
<b>rule 9:</b> 1.7 (aromatic) heteroatom H shift	57,242,472	<b>8.4</b>
<b>rule 10:</b> 1.9 (aromatic) heteroatom H shift	5,061,731	0.7
<b>rule 11:</b> 1.11 (aromatic) heteroatom H shift	1,374,235	0.2
<b>rule 12:</b> furanones	17,860,604	2.6

## Tautomer Analysis

- usage of SMIRKS rules (2/2):

tautomer rule	generated tautomers	
	count	%
<b>rule 13:</b> keten/ynol exchange	57,989	<0.1
<b>rule 14:</b> ionic nitro/aci-nitro	428,266	<0.1
<b>rule 15:</b> pentavalent nitro/aci-nitro	129	<0.1
<b>rule 16:</b> oxim/nitroso	505,695	<0.1
<b>rule 17:</b> oxim/nitroso via phenol	131,502	<0.1
<b>rule 18:</b> cyanic/iso-cyanic acids	181	<0.1
<b>rule 19:</b> formamidinesulfinic acids	1392	<0.1
<b>rule 20:</b> isocyanides	229	<0.1
<b>rule 21:</b> phosphonic acids	54,926	<0.1

# Tautomer Analysis

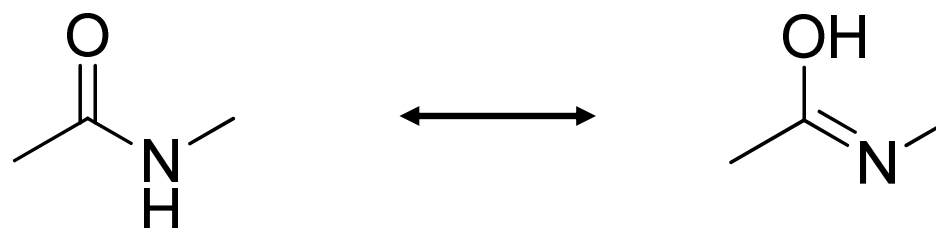
- number of tautomers per structure:

<b>FICuS structures with</b>	<b>count</b>	<b>%</b>
no tautomers	9,756,186	13.8
one tautomer	10,721,845	15.2
2-10 tautomers	33,532,284	47.5
11-25 tautomers	10,870,312	15.4
25-50 tautomers	2,622,587	3.7
51-100 tautomers	1,136,066	1.6
101-200 tautomers	565,199	0.8
201-300 tautomers	104,875	<0.1
301-400 tautomers	35,144	<0.1
401-500 tautomers	17,241	<0.1
501-600 tautomers	4,323	<0.1
601-700 tautomers	1,400	<0.1
701-800 tautomers	362	<0.1
801-832 tautomers	3	<0.1

# Tautomer Analysis

- number of tautomers per structure:

FICuS structures with	count	%
no tautomers	9,756,186	13.8
one tautomer	10,721,845	15.2
2-10 tautomers	33,532,284	47.5
11-25 tautomers	10,870,312	15.4
26-50 tautomers	1,587	3.7
51-70 tautomers	1066	1.6
71-100 tautomers	199	0.8
101-200 tautomers	875	0.1
201-300 tautomers	144	<0.1
301-400 tautomers	241	<0.1
401-500 tautomers	323	<0.1
501-600 tautomers	400	<0.1
701-800 tautomers	362	<0.1
801-832 tautomers	3	<0.1



many minor tautomeric forms  
(but you find them in databases)

# Tanimoto Similarities of Tautomers

- **canonical tautomer** vs. generated tautomers (680 million tautomer set)

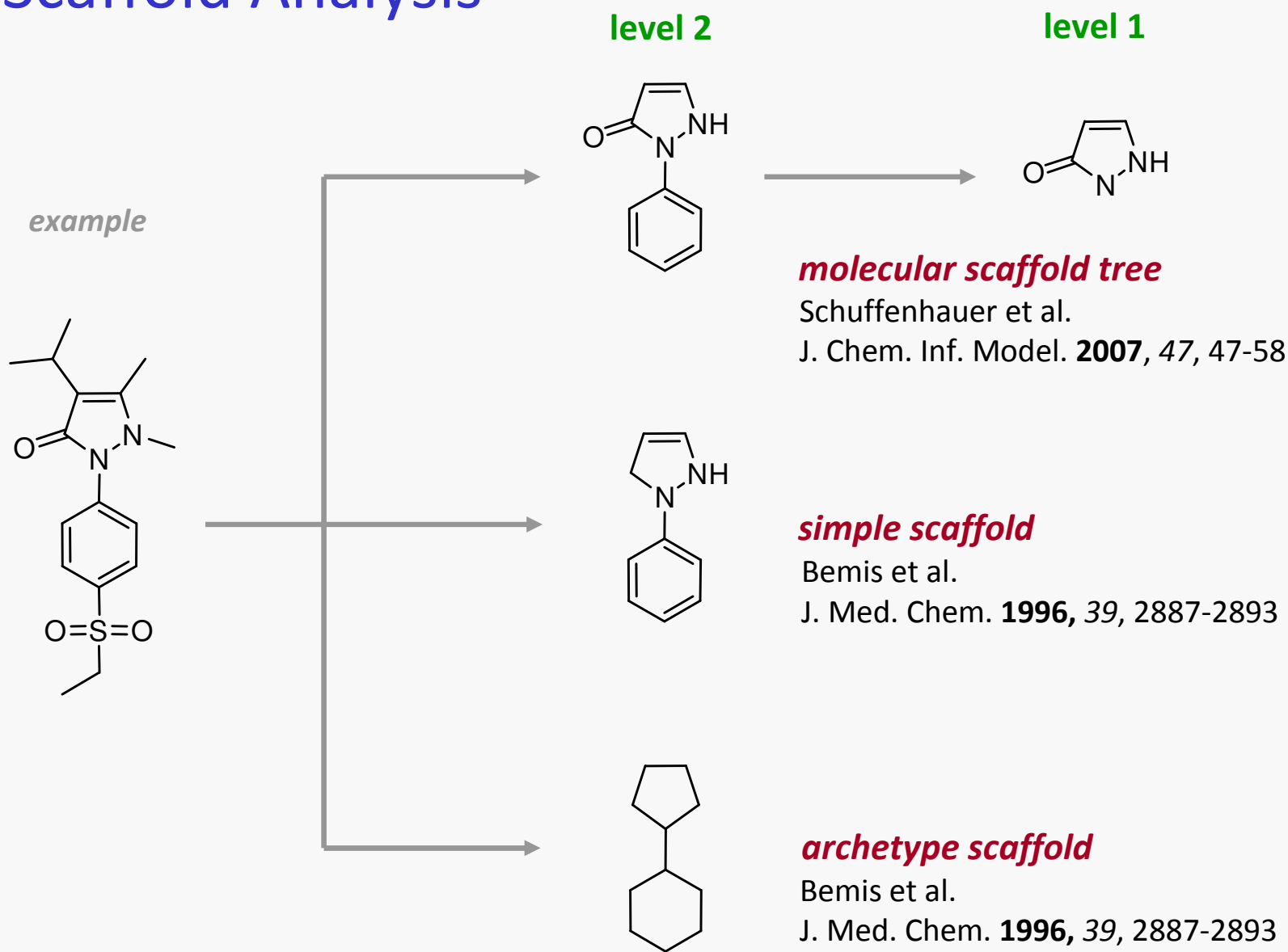
Tanimoto index range	Count	%
>0.0-0.2	0	0.0
>0.2-0.3	6	<0.1
>0.3-0.4	6,580	<0.1
>0.4-0.5	369,331	<0.1
>0.5-0.6	6,304,436	0.9
>0.6-0.7	36,448,651	5.3
>0.7-0.8	111,954,384	16.4
>0.8-0.9	214,747,976	31.5
>0.9-1.0	310,725,465	45.6

~ 23% below 0.8 Tanimoto similarity (although the same molecule)

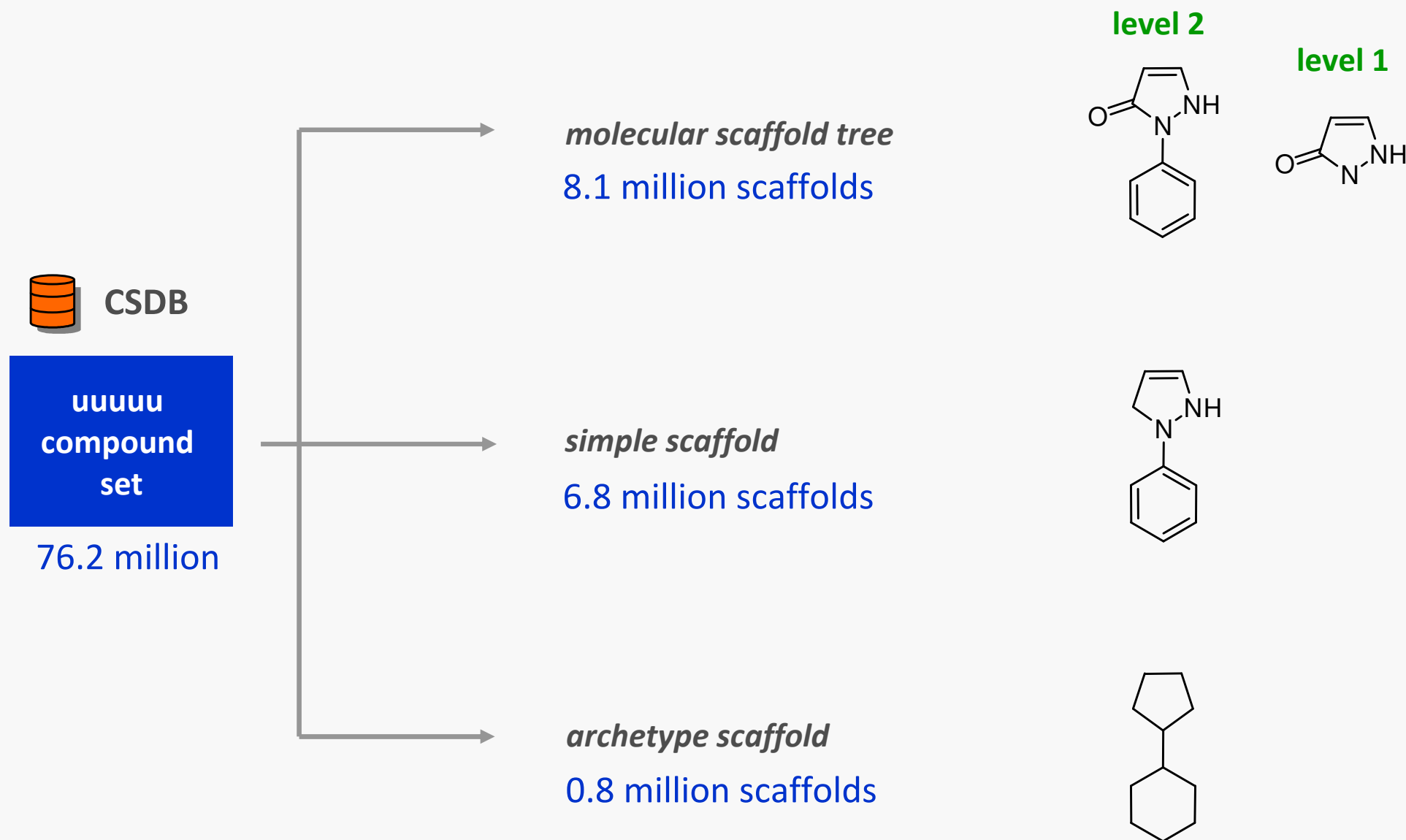
PubChem/CACTVS E\_SCREEN bitvector (881 bits)

# Scaffold Analysis

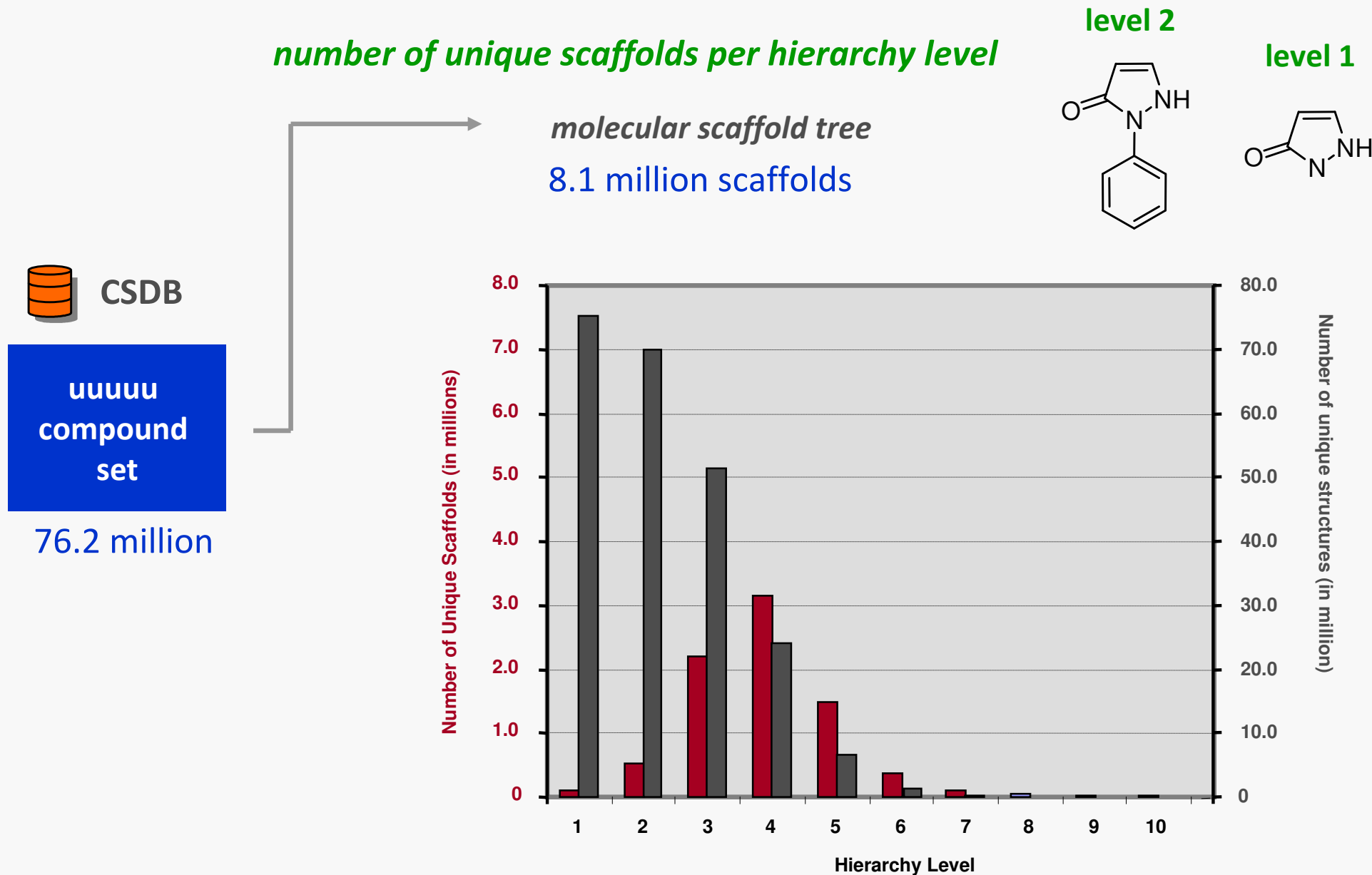
# Scaffold Analysis



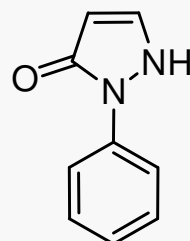
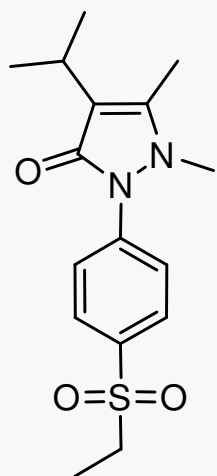
# Scaffold Analysis



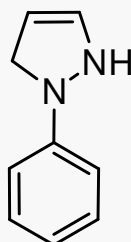
# Scaffold Analysis



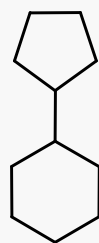
# Scaffold Analysis



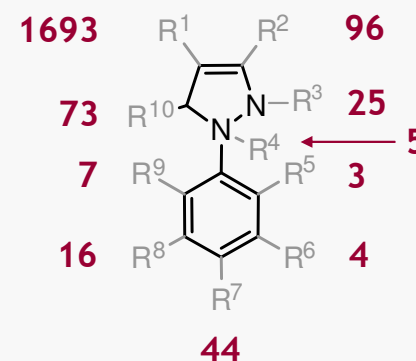
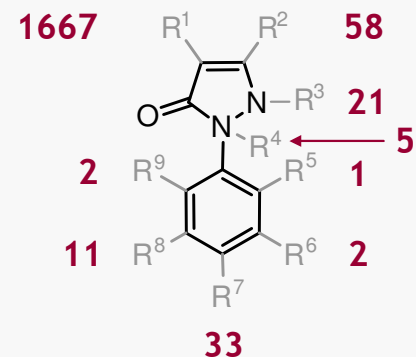
2,281 uuuuu  
parent structures  
5334 structure records  
in 64 databases



2,726 uuuuu  
parent structures  
6007 structure records  
in 66 databases

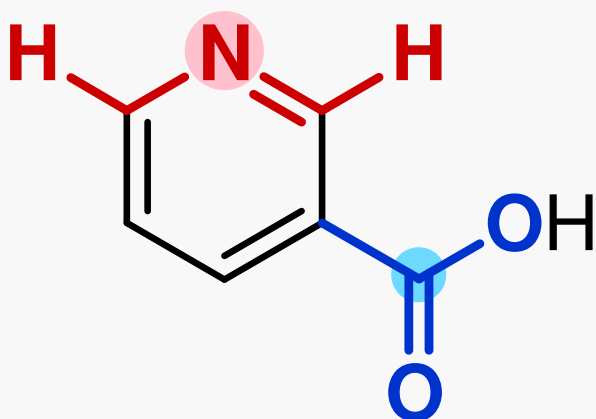


744,469 uuuuu  
parent structures  
1,069,046 structure records  
in 66 databases



# Atom Neighborhoods

# Multilevel Neighborhoods of Atoms (MNA)



## MNA level 1

HC  
HO  
CHCC  
CHCN  
CCCC  
**CCOO**  
NCC  
OHC  
OC

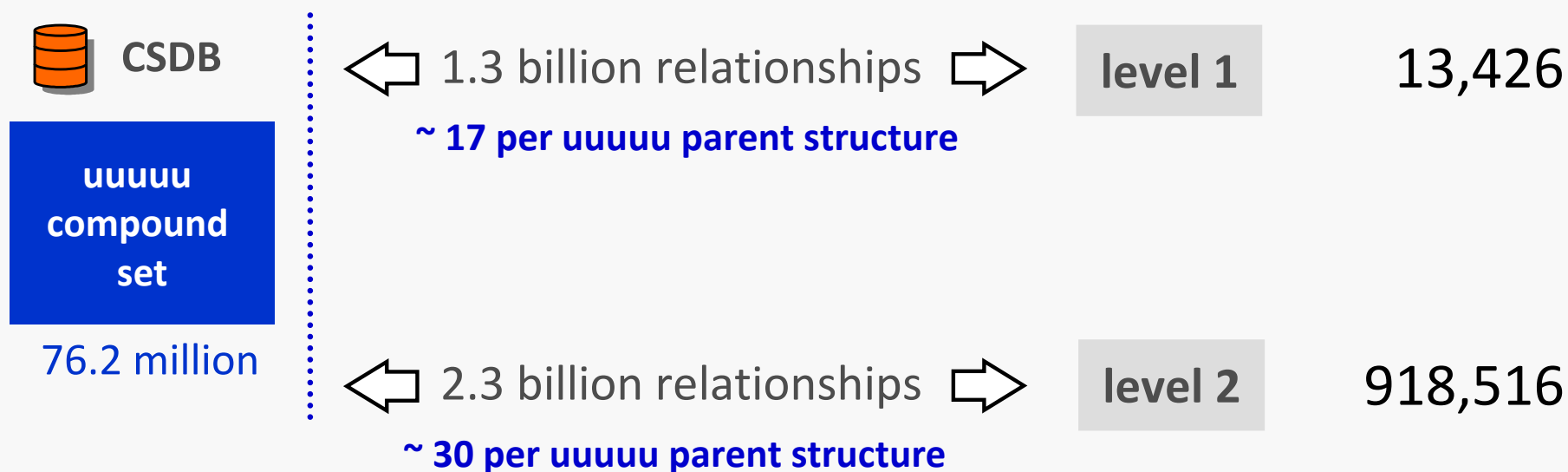
## MNA level 2

C(C(CC-H)C(CC-C)-H(C))  
C(C(CC-H)C(CN-H)-H(C))  
C(C(CC-H)C(CN-H)-C(C-O-O))  
C(C(CC-H)N(CC)-H(C))  
C(C(CC-C)N(CC)-H(C))  
**N(C(CN-H)C(CN-H))**  
-H(C(CC-H))  
-H(C(CN-H))  
-H(-O(-H-C))  
-C(C(CC-C)-O(-H-C)-O(-C))  
-O(-H(-O)-C(C-O-O))  
-O(-C(C-O-O))

Filimonov D., Poroikov V., Borodina Yu., Glorizova T. J.  
Chem. Inf. Comput. Sci., **1999**, 39 (4), 666-670.

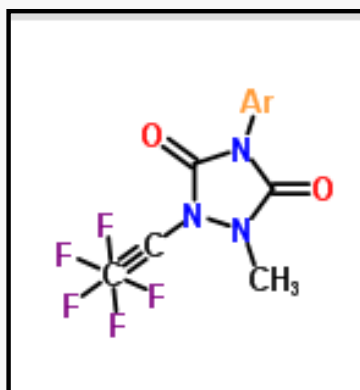
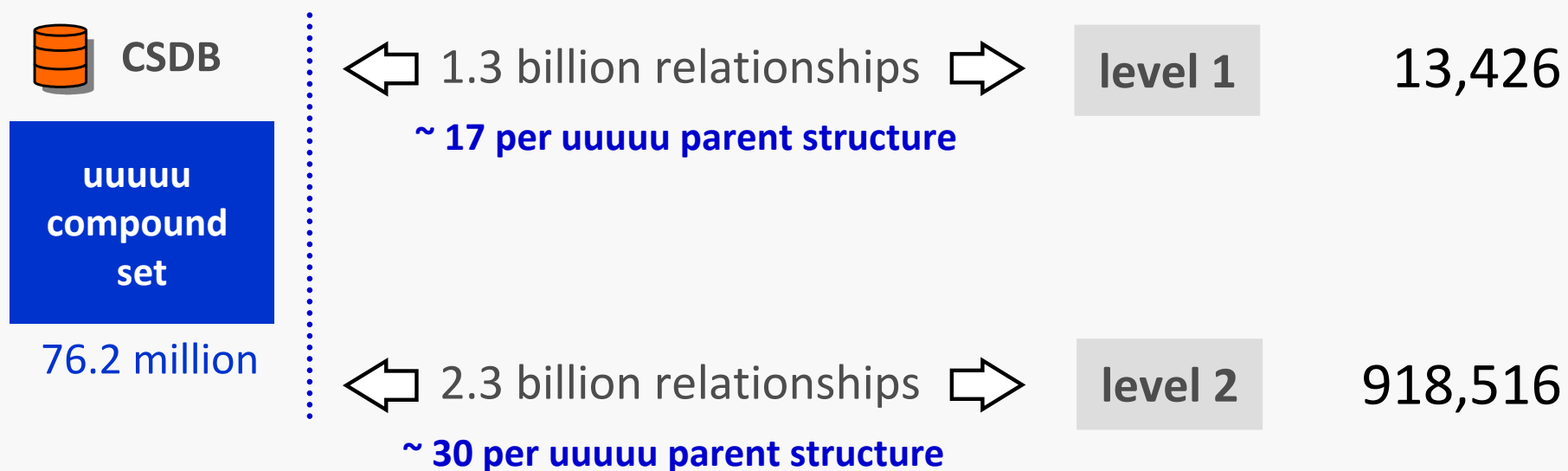
# Multilevel Neighborhoods of Atoms (MNA)

## Unique MNAs



# Multilevel Neighborhoods of Atoms (MNA)

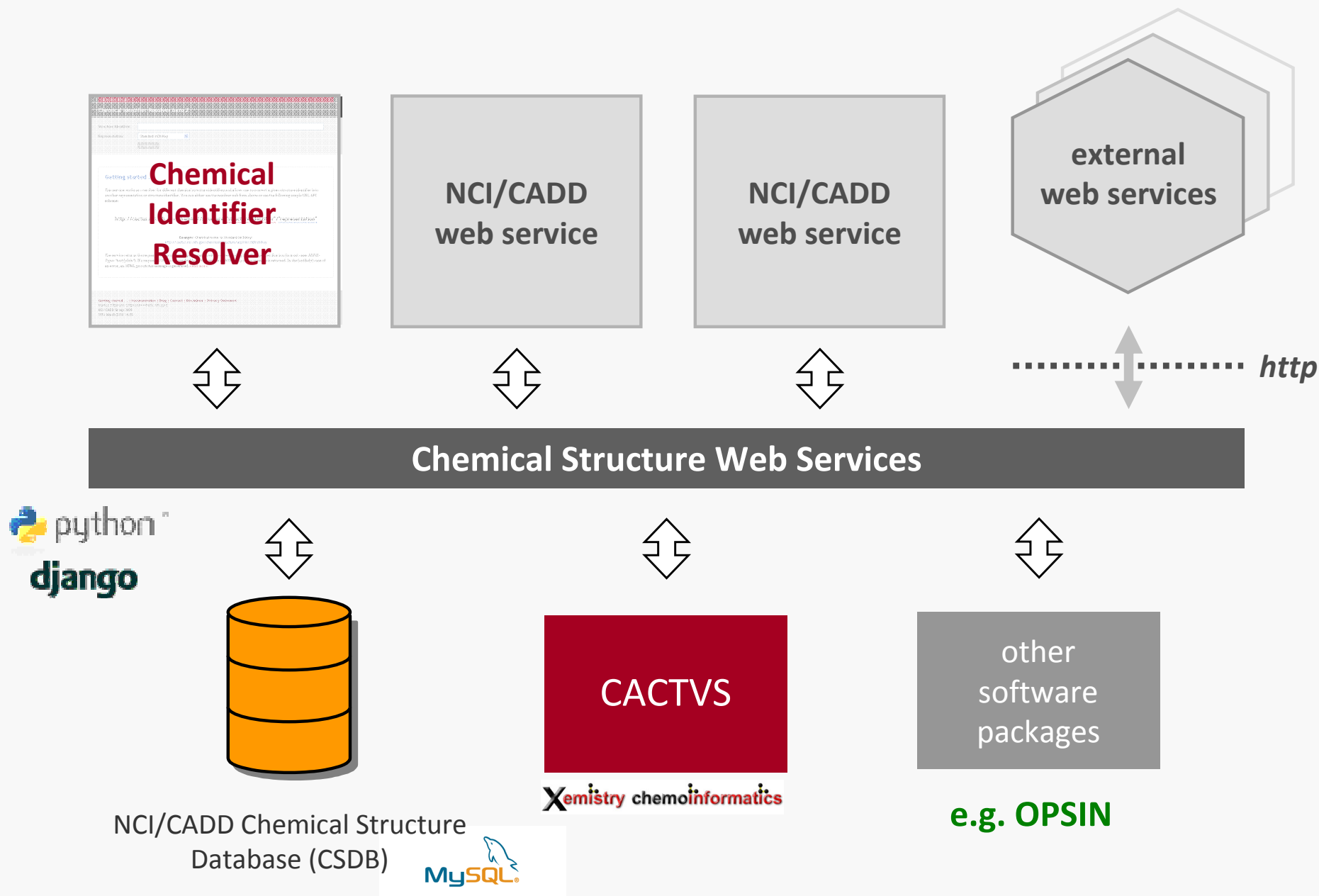
Unique MNAs



424,784 MNAs (level 2) are exclusive to a set of 1,3 million structures in ChemSpider

# Indexing Chemical Space

## Chemical Structure Web Services



# NCI/CADD Web Resources

## Chemical Identifier Resolver

cactus.nci.nih.gov

### Chemical Identifier Resolver *beta 2*

Structure Identifier:

Representation:

#### Getting started ...

This service works as a resolver for different chemical structure identifiers and allows one to convert a given structure identifier into another representation or structure identifier. You can either use the resolver web form above or use the following simple URL API scheme:

[http://cactus.nci.nih.gov/chemical/structure/"structure identifier"/"representation"](http://cactus.nci.nih.gov/chemical/structure/)

**Example:** Chemical name to Standard InChIKey:  
<http://cactus.nci.nih.gov/chemical/structure/aspirin/stdinchikey>

The service returns the requested new structure representation with a corresponding MIME-Type specification (in most cases *MIME-Type: text/plain*). If a requested URL is not resolvable for the service an *HTML 404* status message is returned. In the (unlikely) case of an error, an *HTML 500* status message is generated. [Read more.](#)

[Getting started ...](#) | [Documentation](#) | [Blog](#) | [Contact](#) | [Disclaimer](#) | [Privacy Statement](#)  
Markus Sitzmann (sitzmann+++helix.nih.gov)  
NCI/CADD Group 2009  
15th March 2010 16:38

<http://cactus.nci.nih.gov/chemical/structure>

### /chemical/structure Blog

About new web services at <http://cactus.nci.nih.gov>

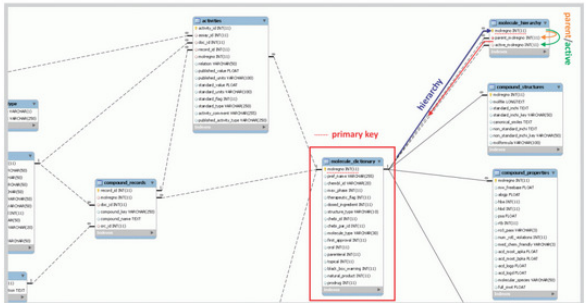
Home | About | Contact | Disclaimer | Privacy

## CADD Group Chemoinformatics Tools and User Services

Home About

### Using pychembl (3) – Active & Parent Molecules

Posted on May 23, 2011 by Markus



A quite interesting table in ChEMBLdb, also linked to table *molecule\_dictionary* by the mutual primary key *molregno*, is table *molecule\_hierarchy*. As the name suggests, it stores hierarchical relationships between row entries in table *molecule\_dictionary* and provides a linkage to the parent and active form of a molecule if available in ChEMBLdb.

But first of all, let us load an example molecule from the database again:

```
> molecule = chembl_db.query(MoleculeDictionary).filter(MoleculeDictionary.molregno==47940).one()
```

<http://cactus.nci.nih.gov/blog>

# Acknowledgments

Thanks to all database providers!

**CADD Group, CBL, NCI**

Igor Filippov

**ChemNavigator**

Scott Hutton

Tad Hurst

**University of Cambridge**

Daniel Lowe

Peter Murray-Rust

Noel' O Boyle (University College Cork, Ireland)

Richard Apodaca (Metamolecular)

Hans-Juergen Himmler

**Our web site:**

<http://cactus.nci.nih.gov>

# Acknowledgments - Software



ChemWriter



Python Web Framework

JME Molecular Editor 😊

Peter Ertl (Novartis)



Python SQL Library



Javascript library

