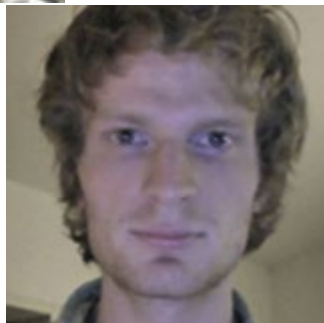




Norbert Sas

**Consultancy**

# Consultancy at ChemAxon





taking care of:

- architect solutions
- integration
- migration
- customization / extension

# Customizing Compound Registration workflows

## Workflow steps

The registration system has a certain number of workflows. On this page you can configure which steps should be executed for each workflow.

Selected workflow: **Auto registration** ▾

### Workflow steps

SourceValidator ✕ LotIdValidator ✕ LnbRefValidator ✕ StructureTypeValidator ✕ +

RejectMarkushValidator ✕ FieldValidator ✕ FieldGenerator ✕ MarkushStructureValidator ✕

StandardizerErrorValidator ✕ AmfValidator ✕ PolymerValidator ✕ SaltSolvateValidator ✕

ParentMultiplicityValidator ✕ SaltFragmentsValidator ✕ ChargeValidator ✕ StructureCheckerValidator ✕

StructureQualityCheckerValidator ✕ SaltFragmentsValidator ✕ StereoCommentsValidator ✕

Save steps

# Custom ID generator

Single compound

CXN-1234

# Custom ID generator

Single compound

CXN-1234

Single compound + salt/solvent

CXN-**S**-1234

# Custom ID generator

Single compound

CXN-1234

Single compound + salt/solvent

CXN-**S**-1234

Polymer

CXN-**P**-1234

# Custom ID generator

Single compound

CXN-1234

Single compound + salt/solvent

CXN-**S**-1234

Polymer

CXN-**P**-1234

Mixture

CXN-**M**-1234



# Custom ID generator

Single compound	CXN-1234
Single compound + salt/solvent	CXN- <b>S</b> -1234
Polymer	CXN- <b>P</b> -1234
Mixture	CXN- <b>M</b> -1234
Formulation	CXN- <b>F</b> -1234

# Custom ID generator

Single compound	CXN-1234
Single compound + salt/solvent	CXN- <b>S</b> -1234
Polymer	CXN- <b>P</b> -1234
Mixture	CXN- <b>M</b> -1234
Formulation	CXN- <b>F</b> -1234
No structure	CXN- <b>NS</b> -1234

# Custom Field Validator



Registration Upload Staging Search

Quick search: type here any ID...

nsas@chemaxon.com

## Register new compound

Register

Using source REGISTRAR  Created by nsas@chemaxon.com  Advanced mode  Virtual Compound

Single Structure

View Find

No salts or solvates added.

+ Add new

Chemically Significant Text

Enter any text value

Actions



Click to draw structure

### Submission Details

LnbRef

E.g. N1111-1-1

Specified molweight

Submitter

nsas@chemaxon.com

### Additional Data

Project

Lot Molecule Formula

Test 3

IUPAC name

logP

logD (pH:7.4)

Shipped

OFF

test2 (positive int)

+

PSA

Testing

Data (valid JSON)

Stereochemistry

# Custom Field Validator



Registration Upload Staging Search

Quick search: type here any ID...

nsas@chemaxon.com

## Register new compound

Register

Using source REGISTRAR · Created by nsas@chemaxon.com · Advanced mode OFF · Virtual Compound OFF

Single Structure

View

Find

No salts or solvates added.

+ Add new

Chemically Significant Text

Enter any text value

Actions



Click to draw structure

### Submission Details

LnbRef

E.g. N1111-1-1

This field is required.

Specified molweight

Submitter

nsas@chemaxon.com

### Additional Data

Project

Lot Molecule Formula

Test 3

IUPAC name

logP

logD (pH:7.4)

Shipped

OFF

test2 (positive int)

PSA

Testing

Data (valid JSON)

Stereochemistry

# Custom Field Validator



Registration Upload Staging Search

Quick search: type here any ID...

nsas@chemaxon.com

## Register new compound

Register

Using source REGISTRAR  Created by nsas@chemaxon.com  Advanced mode  Virtual Compound

Single Structure

View Find

No salts or solvates added.

+ Add new

Chemically Significant Text

Enter any text value

Actions



Click to draw structure

### Submission Details

LnbRef

E.g. N1111-1-1

Specified molweight

Submitter

nsas@chemaxon.com

### Additional Data

Project

Lot Molecule Formula

Test 3

IUPAC name

logP

logD (pH:7.4)

Shipped

OFF

test2 (positive int)

+

PSA

Testing

Data (valid JSON)

Stereochemistry

# Custom Field Validator

The screenshot shows the 'Dictionary Manager' interface. At the top, there is a blue navigation bar with 'Registration', 'Upload', 'Staging', and 'Search' options. A search bar on the right contains the text 'Quick search: type here any ID...'. The user's email 'nsas@chemaxon.com' is displayed in the top right corner. On the left, a dark sidebar menu lists various application areas, with 'Dictionary Manager' highlighted. The main content area is titled 'Dictionary Manager' and is split into two panels. The left panel shows a list of dictionaries: 'Chem. Sig. Text', 'Double Bond Panel', 'Geometric Isomerism', 'scientist\_lmbref' (which is selected and highlighted), 'Stereocenter Panel', and 'Stereochemistry'. The right panel, titled '"scientist\_lmbref" items', contains a search bar with the text 'Search dictionary items ...' and a table with two columns. The table has one row with the name 'Alice' and the identifier 'CXN-12345'. At the bottom of the right panel, it states '2 items total.'

Registration Upload Staging Search

Quick search: type here any ID...

nsas@chemaxon.com

MENU

- Access Control
- Forms & Fields
- Form Editor
- Dictionary Manager**
- Chemistry
- Chemical Structures
- Pages
- General Settings
- ADVANCED
- Workflow
- Notifications
- Integration
- Tools
- Navigation

## Dictionary Manager

+ Add dictionary

- Chem. Sig. Text
- Double Bond Panel
- Geometric Isomerism
- scientist\_lmbref**
- Stereocenter Panel
- Stereochemistry

### "scientist\_lmbref" items

+ Add item

Search dictionary items ...

Alice	CXN-12345
-------	-----------

2 items total.

# Structure Based Field Validation



Registration Upload Staging Search

Quick search: type here any ID...

nsas@chemaxon.com

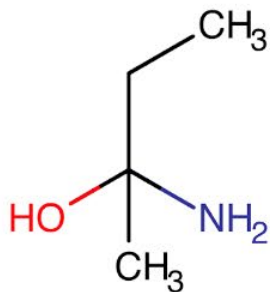
## Register new compound

Register

Using source REGISTRAR · Created by nsas@chemaxon.com · Advanced mode OFF · Virtual Compound OFF

Single Structure

View Find



No salts or solvates added.

+ Add new

Chemically Significant Text

Enter any text value

Actions

### Submission Details

LnbRef

E.g. N1111-1-1

Specified molweight

Submitter

nsas@chemaxon.com

### Additional Data

Project

Lot Molecule Formula

Test 3

IUPAC name

logP

logD (pH:7.4)

Shipped

OFF

test2 (positive int)

PSA

Testing

Data (valid JSON)

Stereochemistry

# Salt name manner

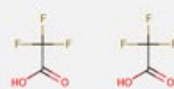
Registration Upload Staging Search Quick search: type here any ID... admin

## Salts & Solvates

SHOW SETTINGS

Salts Solvates

+ Add salt Filter by name or ID...

$\text{Ba}^{2+}$ $\text{Ba}^{2+}$	$\text{Ca}^{2+}$ $\text{Ca}^{2+}$	HBr Br—H	HCl Cl—H	HF F—H	HI I—H
Name 2Ba ID 28	Name 2Ca ID 24	Name 2HBr ID 34	Name 2HCl ID 32	Name 2HF ID 30	Name 2HI ID 36
$\text{K}^+$ $\text{K}^+$	$\text{Li}^+$ $\text{Li}^+$	$\text{Mg}^{2+}$ $\text{Mg}^{2+}$	$\text{Na}^+$ $\text{Na}^+$		$\text{Ba}^{2+}$ $\text{Ba}^{2+}$ $\text{Ba}^{2+}$
Name 2K ID 20	Name 2Li ID 22	Name 2Mg ID 26	Name 2Na ID 18	Name 2TFA ID 39	Name 3Ba ID 29
$\text{Ca}^{2+}$ $\text{Ca}^{2+}$	HBr HBr	HCl HCl	HF HF	HI HI	$\text{K}^+$ $\text{K}^+$



# Salt name manner

## Register new compound

Using source REGISTRAR ▾ • Created by nsas@chemaxon.com • Advanced mode  OFF • Virtual Compound  OFF

Single Structure ▾

View ▾

Find

Search for salts/solvates

chl

hydrochloride (Salt, #4)

sodium chloride (Salt, #6)

Add

Cancel

Chemically Significant Text

Enter any text value ▾

⋮ Actions

  
Click to draw  
structure

Submission Details

LnbRef

E.g. N1111-1-1

Additional Data ▾

Project

+

IUPAC name

Shipped

OFF

Testing

# Salt name mapper

Registration Upload Staging Search

Quick search: type here any ID... nsas@chemax

## Compound\_00000001\_000025000.sdf

Browse or edit content

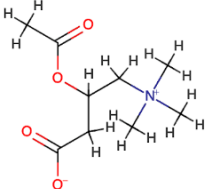
Record nr. 1 < >

Field mapping:

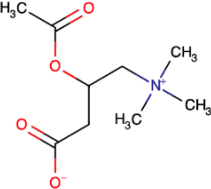
- Structure → Structure  
1 -OEChem-03021912542D 31 30 0 1 0 0 0 0 0999 V200 0 ...
- PUBCHEM\_COMPOUND\_CID 1
- PUBCHEM\_COMPOUND\_CANONICALIZED 1
- PUBCHEM\_CACTVS\_COMPLEXITY 214
- PUBCHEM\_CACTVS\_HBOND\_ACCEPTOR 4
- PUBCHEM\_CACTVS\_HBOND\_DONOR 0
- PUBCHEM\_CACTVS\_ROTATABLE\_BOND 5
- PUBCHEM\_CACTVS\_SUBSKEYS  
AAADceBy0AAA  
AAHgAAAAAACBThgAYCCA...
- PUBCHEM\_IUPAC\_OPENEYE\_NAME  
3-acetoxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_CAS\_NAME  
3-acetyloxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_NAME

Compounds Salt/Solvate

**Original**



**Normalized**



Using source REGISTRAR

Checkers

- Wiggly Bond Checker
- Crossed Double-Bond Checker
- Straight Double Bond Checker

Registration Options

- Analyze Salt Solvate Fragments
- Register New Lots
- Register New Versions
- Register 2D Parent Matches
- Register Tautomer Parent Matches
- Register 2D Tautomer Parent Matches
- Register Restricted Matches
- Perform Quality Checks
- Register With Unknown CST
- Reject Duplicate IDs
- Calculate Stereo Comments

Library identifier: Auto-generated

Use two-step registration

Upload



# Salt name mapper

Registration Upload Staging Search  nsas@chemax

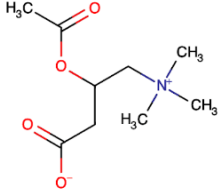
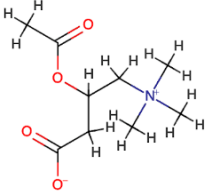
  

**Compound\_00000001\_000025000.sdf**  
Browse or edit content

**Field mapping:** Record nr. 1 < >

- Structure → Structure  
1 -OEChem-03021912542D 31 30 0 1 0 0 0 0 0999 V200 0 ...
- PUBCHEM\_COMPOUND\_CID 1
- PUBCHEM\_COMPOUND\_CANONICALIZED 1
- PUBCHEM\_CACTVS\_COMPLEXITY 214
- PUBCHEM\_CACTVS\_HBOND\_ACCEPTOR 4
- PUBCHEM\_CACTVS\_HBOND\_DONOR 0
- PUBCHEM\_CACTVS\_ROTATABLE\_BOND 5
- PUBCHEM\_CACTVS\_SUBSKEYS  
AAADceBy0AAA  
AAHgAAAAAACBThgAYCCA...
- PUBCHEM\_IUPAC\_OPENEYE\_NAME  
3-acetoxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_CAS\_NAME  
3-acetyloxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_NAME

**Original** **Normalized**



Library identifier: Auto-generated

Use two-step registration

Using source REGISTRAR

**Checkers**  
 Wiggly Bond Checker  
 Crossed Double-Bond Checker  
 Straight Double Bond Checker

**Registration Options**  
 Analyze Salt Solvate Fragments  
 Register New Lots  
 Register New Versions  
 Register 2D Parent Matches  
 Register Tautomer Parent Matches  
 Register 2D Tautomer Parent Matches  
 Register Restricted Matches  
 Perform Quality Checks  
 Register With Unknown CST  
 Reject Duplicate IDs  
 Calculate Stereo Comments



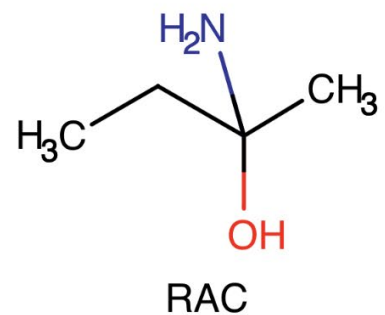
><SaltName>  
*HCl*

# StereoCommentCSTGenerator

## Register new compound

Using source REGISTRAR ▾ · Created by nsas@chemaxon.com · Advanced mode  · Virtual Compound

Single Structure ▾ View ▾ Find



RAC

*No salts or solvates added.*  
[+ Add new](#)

Chemically Significant Text

⋮ Actions

### Submission Details

LnbRef

---

### Additional Data ▾

Project

IUPAC name

Shipped

Testing

# StereoCommentCSTGenerator

Registration Upload Staging Search  nsas@chemax

Compound\_00000001\_000025000.sdf

[Browse](#) or [edit content](#)

Field mapping: Record nr. 1 < >

- Structure → Structure  
1 -OEChem-03021912542D 31 30 0 1 0 0 0 0 0999 V200  
0 ...
- PUBCHEM\_COMPOUND\_CID 1
- PUBCHEM\_COMPOUND\_CANONICALIZED 1
- PUBCHEM\_CACTVS\_COMPLEXITY 214
- PUBCHEM\_CACTVS\_HBOND\_ACCEPTOR 4
- PUBCHEM\_CACTVS\_HBOND\_DONOR 0
- PUBCHEM\_CACTVS\_ROTATABLE\_BOND 5
- PUBCHEM\_CACTVS\_SUBSKEYS  
AAADceBy0AAA  
AAHgAAAAAACBThgAYCCA...
- PUBCHEM\_IUPAC\_OPENEYE\_NAME  
3-acetoxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_CAS\_NAME  
3-acetyloxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_NAME

Comounds Salt/Solvate

Original

Normalized

Using source REGISTRAR ▾

Checkers

- Wiggly Bond Checker
- Crossed Double-Bond Checker
- Straight Double Bond Checker

Registration Options

- Analyze Salt Solvate Fragments
- Register New Lots
- Register New Versions
- Register 2D Parent Matches
- Register Tautomer Parent Matches
- Register 2D Tautomer Parent Matches
- Register Restricted Matches
- Perform Quality Checks
- Register With Unknown CST
- Reject Duplicate IDs
- Calculate Stereo Comments

Library identifier:

Use two-step registration

# StereoCommentCSTGenerator

Registration Upload Staging Search  nsas@chemax

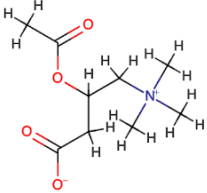
**Compound\_00000001\_000025000.sdf**  
[Browse](#) or [edit content](#)

Field mapping: Record nr. 1 < >

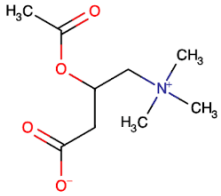
- Structure → Structure  
1 -OEChem-03021912542D 31 30 0 1 0 0 0 0 0999 V200  
0 ...
- PUBCHEM\_COMPOUND\_CID 1
- PUBCHEM\_COMPOUND\_CANONICALIZED 1
- PUBCHEM\_CACTVS\_COMPLEXITY 214
- PUBCHEM\_CACTVS\_HBOND\_ACCEPTOR 4
- PUBCHEM\_CACTVS\_HBOND\_DONOR 0
- PUBCHEM\_CACTVS\_ROTATABLE\_BOND 5
- PUBCHEM\_CACTVS\_SUBSKEYS  
AAADceBy0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAHgAAAAAACBThgAYCCA...
- PUBCHEM\_IUPAC\_OPENEYE\_NAME  
3-acetoxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_CAS\_NAME  
3-acetyloxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_NAME

Comounds Salt/Solvate

Original



Normalized



Using source REGISTRAR

Checkers

- Wiggly Bond Checker
- Crossed Double-Bond Checker
- Straight Double Bond Checker

Registration Options

- Analyze Salt Solvate Fragments
- Register New Lots
- Register New Versions
- Register 2D Parent Matches
- Register Tautomer Parent Matches
- Register 2D Tautomer Parent Matches
- Register Restricted Matches
- Perform Quality Checks
- Register With Unknown CST
- Reject Duplicate IDs
- Calculate Stereo Comments

Library identifier: Auto-generated

Use two-step registration

Upload



# StereoCommentCSTGenerator

Registration Upload Staging Search  nsas@chemax

**Compound\_00000001\_000025000.sdf**  
[Browse](#) or [edit content](#)

Field mapping: Record nr. 1 < >

- Structure → Structure  
1 -OEChem-03021912542D 31 30 0 1 0 0 0 0 0999 V200  
0 ...
- PUBCHEM\_COMPOUND\_CID 1
- PUBCHEM\_COMPOUND\_CANONICALIZED 1
- PUBCHEM\_CACTVS\_COMPLEXITY 214
- PUBCHEM\_CACTVS\_HBOND\_ACCEPTOR 4
- PUBCHEM\_CACTVS\_HBOND\_DONOR 0
- PUBCHEM\_CACTVS\_ROTATABLE\_BOND 5
- PUBCHEM\_CACTVS\_SUBSKEYS  
AAADceBy0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAHgAAAAAACBThgAYCCA...
- PUBCHEM\_IUPAC\_OPENEYE\_NAME  
3-acetoxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_CAS\_NAME  
3-acetyloxy-4-(trimethylammonio)butanoate
- PUBCHEM\_IUPAC\_NAME

Compounds Salt/Solvate

Original Normalized

Library identifier: Auto-generated

Use two-step registration

Using source REGISTRAR

Checkers

- Wiggly Bond Checker
- Crossed Double-Bond Checker
- Straight Double Bond Checker

Registration Options

- Analyze Salt Solvate Fragments
- Register New Lots
- Register New Versions
- Register 2D Parent Matches
- Register Tautomer Parent Matches
- Register 2D Tautomer Parent Matches
- Register Restricted Matches
- Perform Quality Checks
- Register With Unknown CST
- Reject Duplicate IDs
- Calculate Stereo Comments



M SED 1 Rac

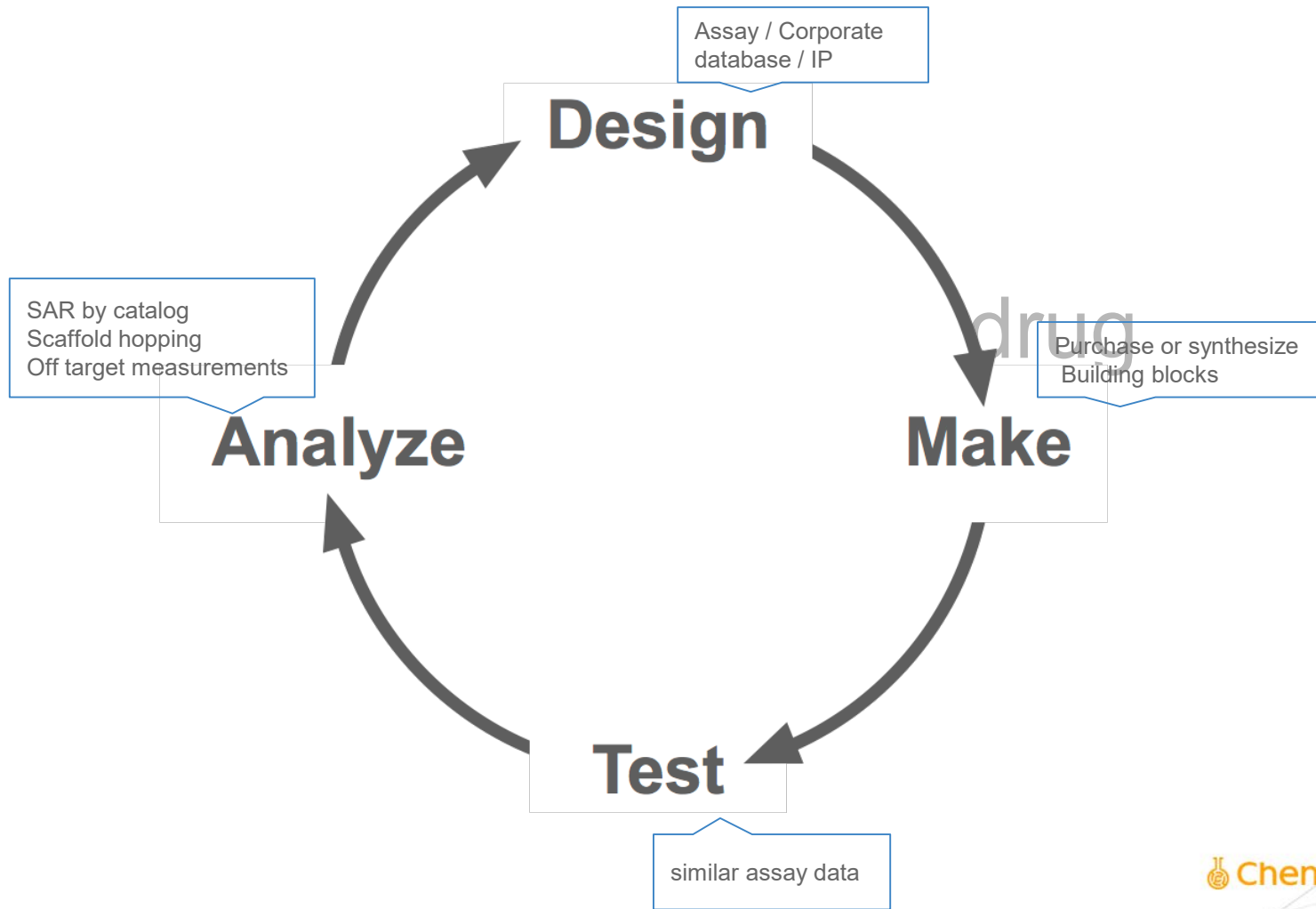






A large stack of hay bales, arranged in a somewhat irregular, stepped fashion. The bales are golden-brown and appear to be made of straw or hay. The stack is the central focus of the image, with a few more bales scattered in the foreground. The background is a bright, overcast sky, and a small portion of a green field is visible on the right side.

# Haystack Project



# Goal

## Quickly explore your chemical space

Indexing arbitrary amount of data and running substructure and similarity search.

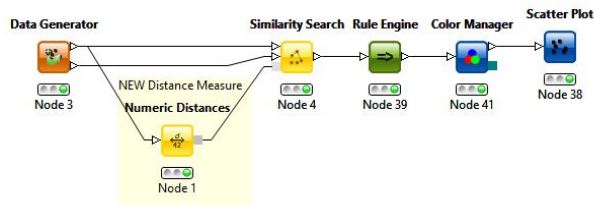
# Typical workflows

*Interactive search*



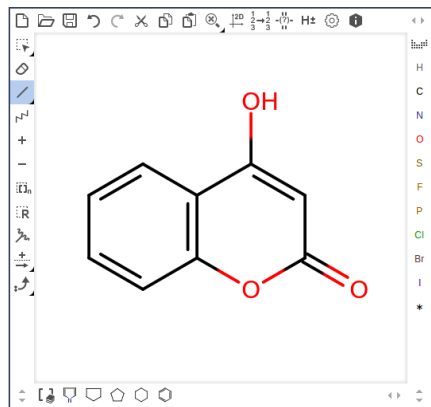
**VS**

*WorkFlow usage*

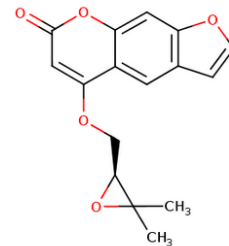
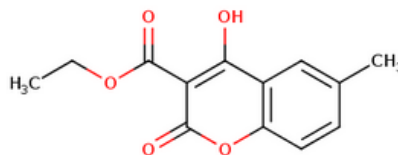
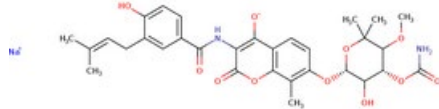
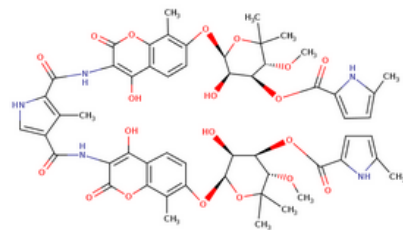
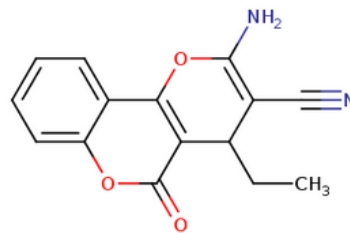
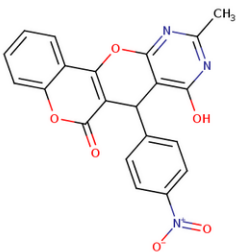


# Relevance search

Unordered result

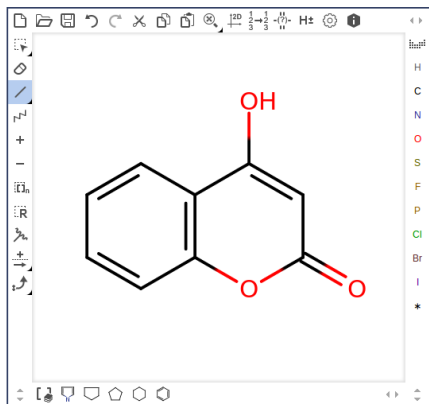


Hits :

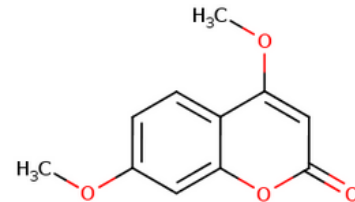
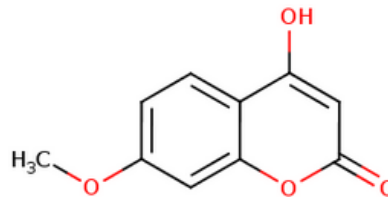
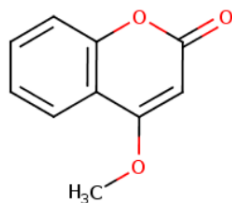
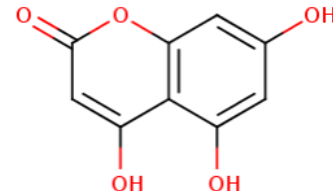
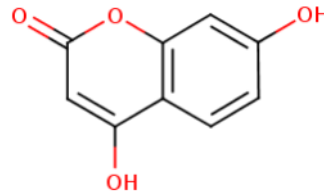
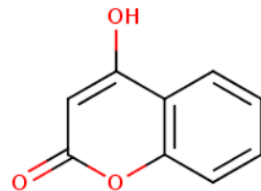


# Relevance search

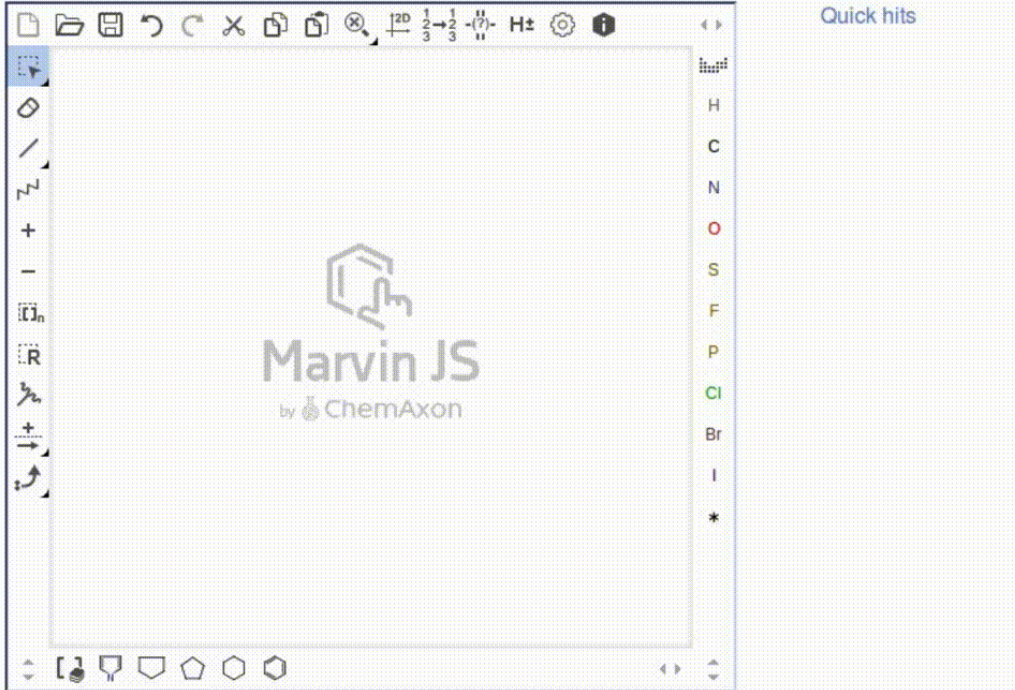
Order substructure search hits according to the relevance to the query structure.



**Hits :**



# Hits as you draw



The image shows a screenshot of the Marvin JS chemical drawing software interface. The main drawing area is a grid with the Marvin JS logo and 'by ChemAxon' text in the center. The interface includes a top toolbar with various drawing tools, a left sidebar with additional drawing tools, and a right sidebar titled 'Quick hits' containing a list of chemical elements: H, C, N, O, S, F, P, Cl, Br, I, and \*. The 'Quick hits' panel is currently empty.



# Chemical space:

Current, partial solutions

Some searchable implementations for specific databases:

- eMolecules
- ChEMBL
- SureChEMBL
- ...



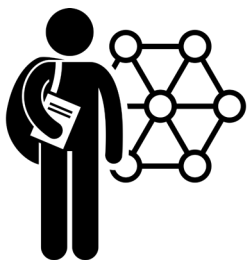
Corporate  
Compound  
Repository  
1.5M – 3M

MolPort  
All Stock  
~7M

Enamine REAL  
~680M

ChEMBL  
1.7M

SureChEMBL  
~18M



BindingDB  
651K

Namiki  
Shoji  
~6M

Screening  
library  
250k – 1M

eMolecules  
~18M

PubChem  
Compounds  
~95M



Confidential

# Haystack Prototype

Corporate Compound Repository  
1.5M – 3M

Corporate Compound Repository  
1.5M – 3M

MolPort Stock

ChEMBL  
1.7M

eMolecules  
~22M

BindingDB  
651K

SureChEMBL  
~18M

ChEMBL  
1.7M

SureChEMBL  
~18M

PubChem Compounds

En

PubChem Compounds  
~95M

Stuff

Design Application Layer

Screening library  
250k – 1M

MolPort Stock  
~7M

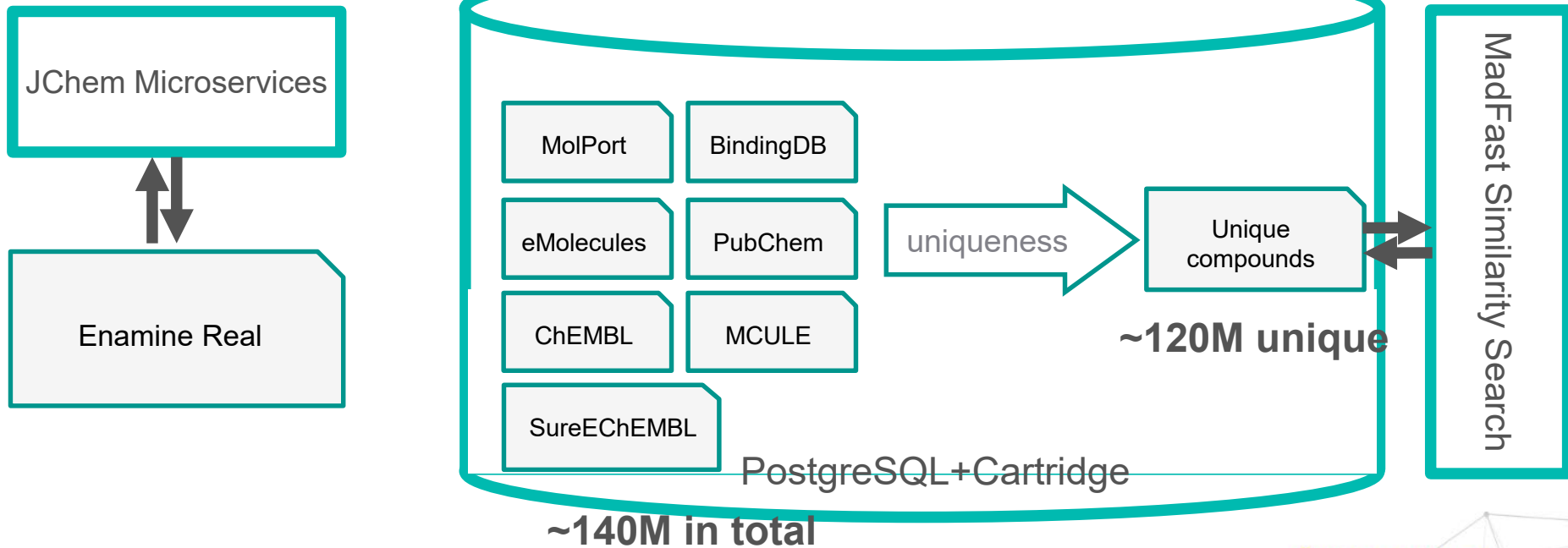
Enamine REAL  
~680M

Namiki Shoji  
~6M

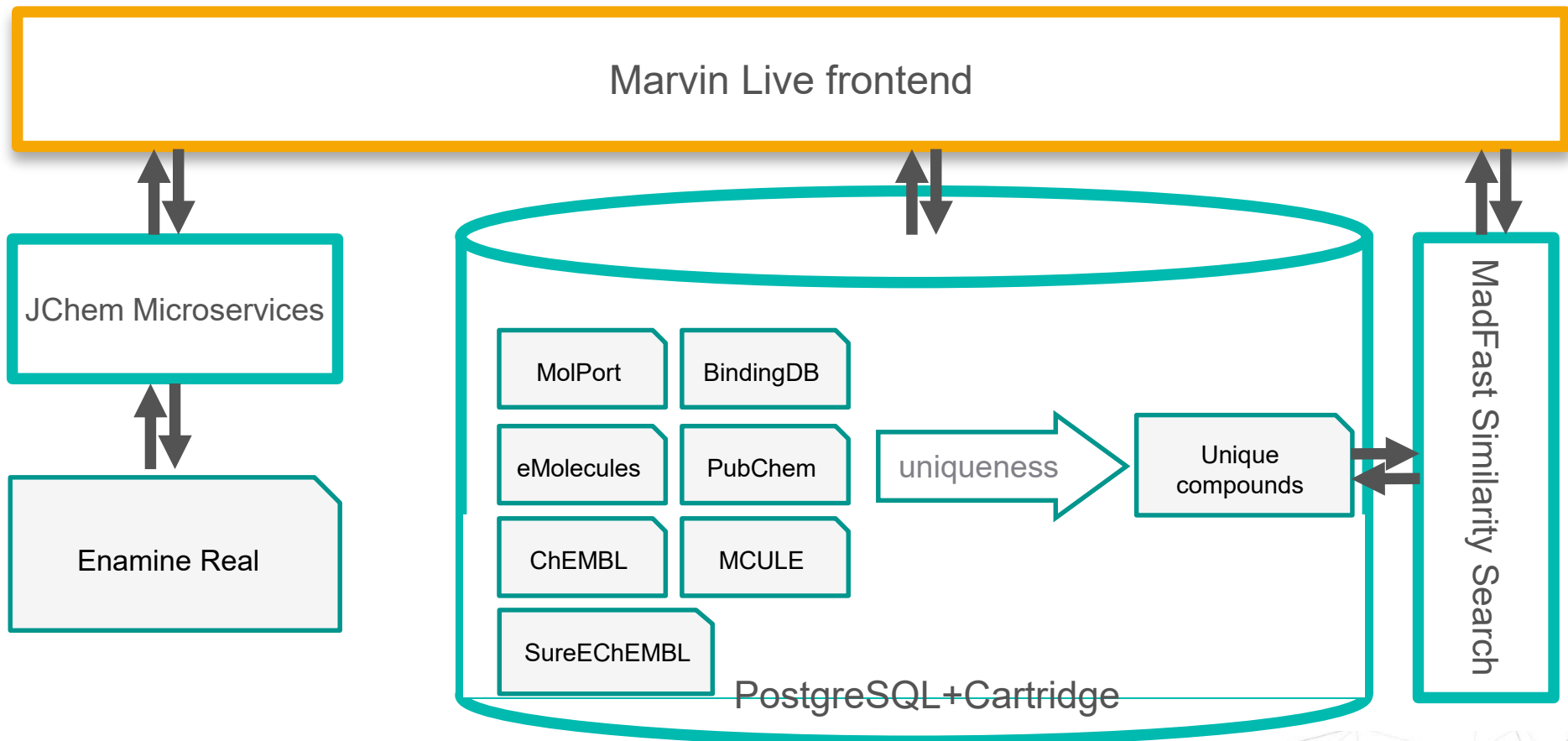
eMolecules  
~18M

ChemAxon

# The Haystack Database - prototype



# The Haystack Database - prototype



DEM  
O



THANK YOU!