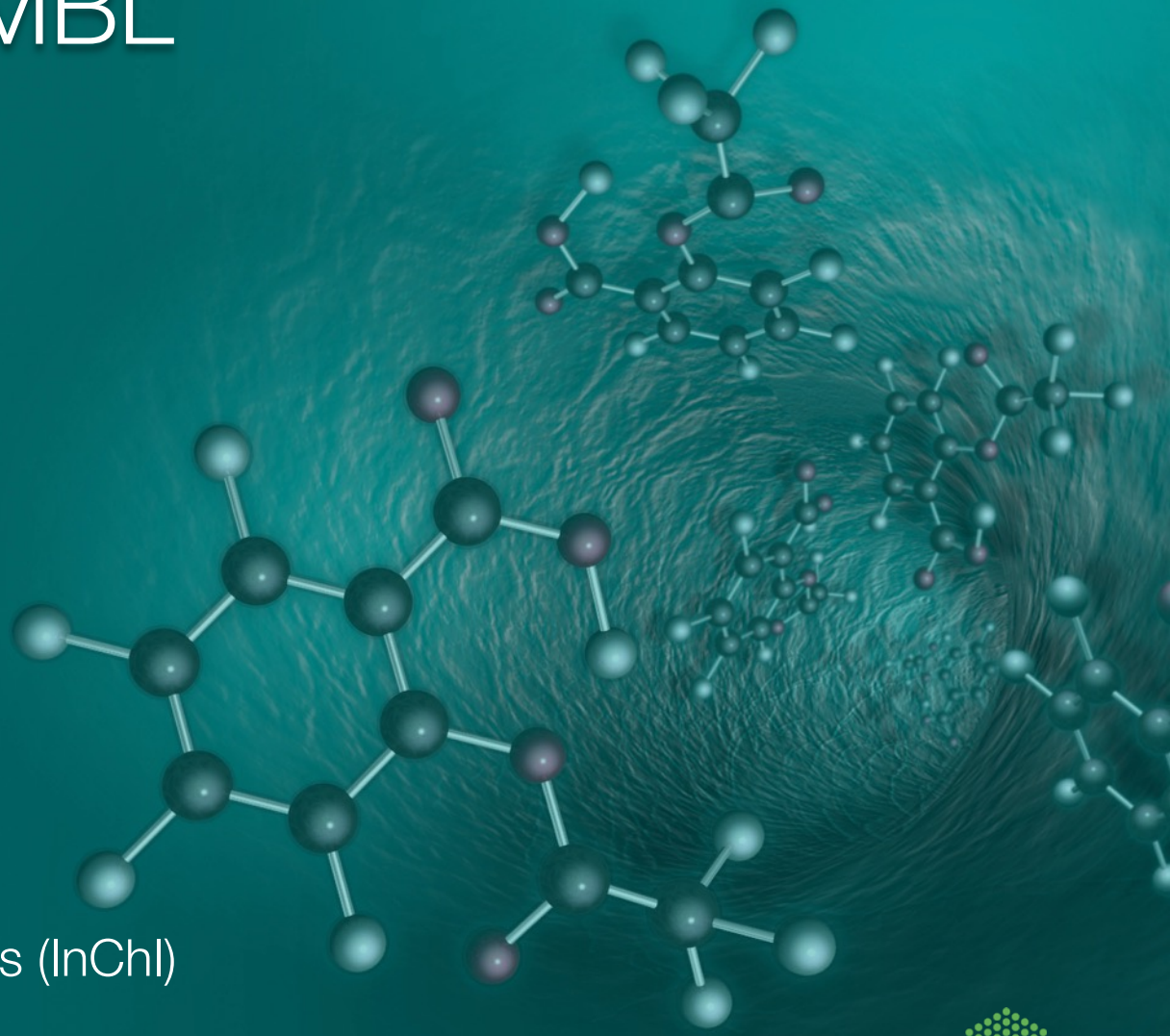
InChI Applications in ChEMBL



Patrícia Bento
Data Integration & Development Officer
Chemogenomics Group – EMBL-EBI

Virtual Workshop on International Chemical Identifiers (InChI)

24-03-2021

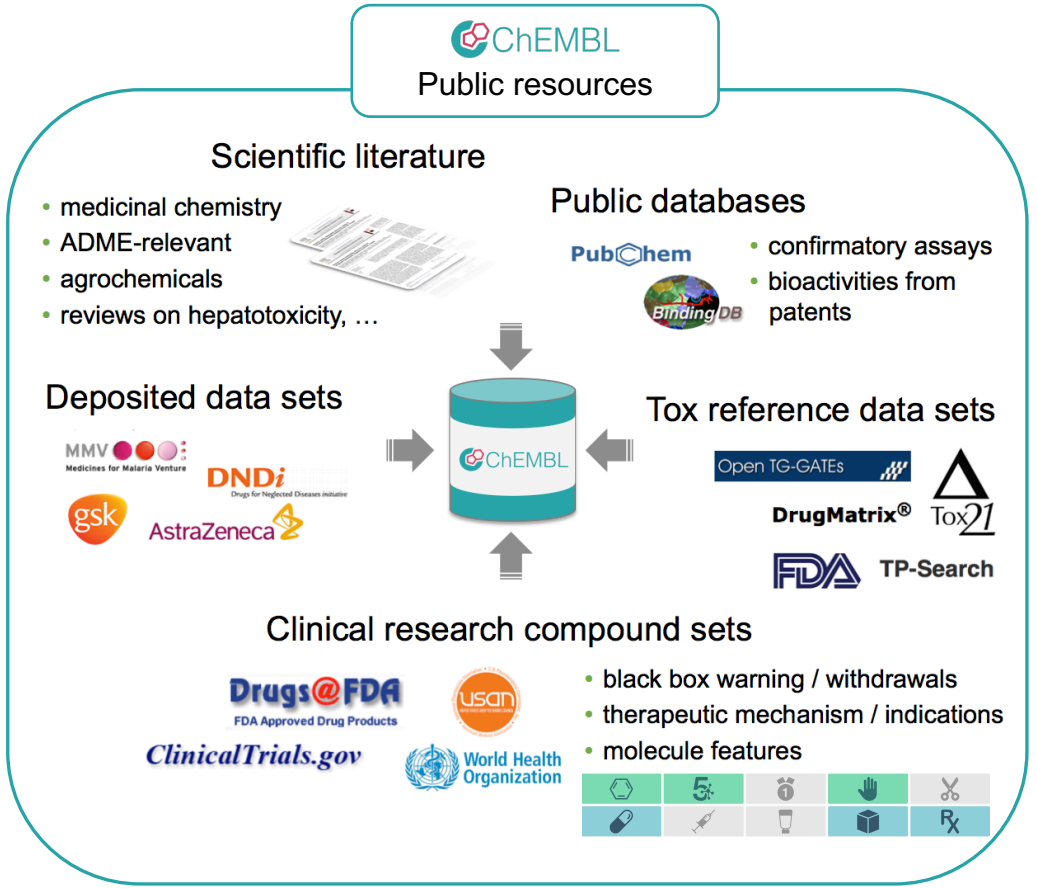
EMBL-EBI

Outline

- ChEMBL Database overview and content
- InChI uses in ChEMBL and UniChem
  - Compound Registration
  - ChEMBL Structure Pipeline
  - Cross-References

ChEMBL Database

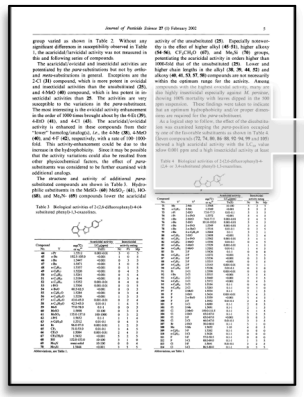
- Open access database for drug discovery
- Core bioactivity data from:
  - Key MedChem journals
  - Public databases
  - Review articles, book chapters, etc
  - Contributed datasets
- Integrated with data on:
  - Clinical development and marketed drugs (from ClinicalTrials.gov, FDA, USANs, INNs)
- Latest release (version 28) statistics:



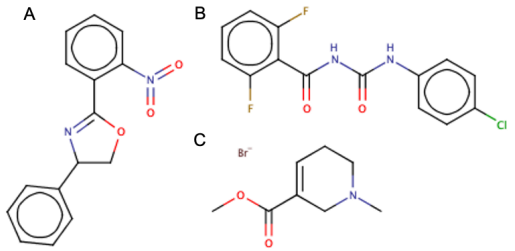
14,347	2,086,898	17,276,334	80,480	170
Targets	Distinct compounds	Activities	Publications	Deposited Datasets

InChI Applications in ChEMBL

Documents



Compounds



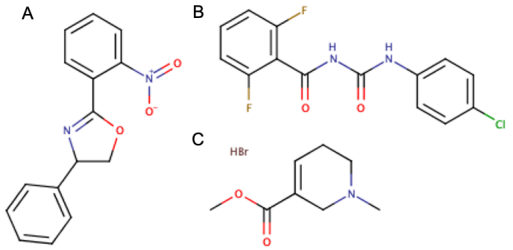
ChEMBL Structure Pipeline

Check

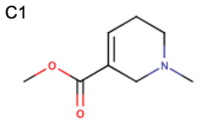
```
A   [ [ 5, "molecule has a bond with an illegal type" ],
      [ 2, "InChI: Omitted undefined stereo" ] ]

B   [ [ 5, "molecule has a bond with an illegal type" ],
      [ 2, "InChI: Omitted undefined stereo" ] ]
```

Standardizer



GetParent



Cross References

ChEBI

MetaboLights

eMolecules

PHARMGKB

mcule

EPA

FDA

DRUGBANK

PDBe Protein Data Bank in Europe

IUPHAR

PubChem

ZINC15

SureChEMBL

UniChem

ChEMBL

Hierarchy

`compound C1 → compound C`

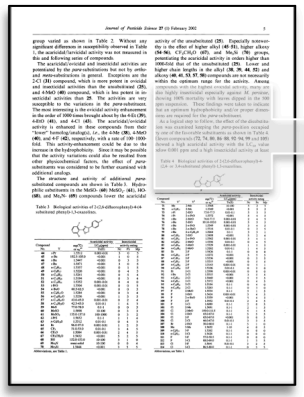
Registration

```
InChI=1S/C14H9ClF2N2O2/c15-8-4-6-9…
InChI=1S/C15H12N2O3/c18-17(19)14-9…
InChI=1S/C8H13NO2/c1-9-5-3-4-7(6-9)8…
InChI=1S/C8H13NO2.BrH/c1-9-5-3-4-7(6…
```

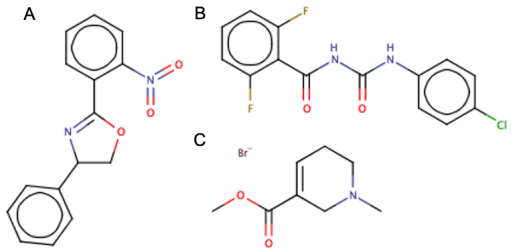
EMBL-EBI

InChI Applications in ChEMBL – Registration

Documents



Compounds



ChEMBL Structure Pipeline

Check

A [[5, "molecule has a bond with an illegal type"], [2, "InChI: Omitted undefined stereo"]]

B [[5, "molecule has a bond with an illegal type"], [2, "InChI: Omitted undefined stereo"]]

Standardizer

GetParent

Cross References

ChEBI
MetaboLights
eMolecules
PHARMGKB
mcule
EPA
FDA
DRUGBANK
PDBe
Protein Data Bank in Europe
IUPHAR
PubChem
ZINC15
SureChEMBL
UniChem

Hierarchy

compound C1 → compound C

Registration

InChI=1S/C14H9ClF2N2O2/c15-8-4-6-9…
InChI=1S/C15H12N2O3/c18-17(19)14-9…
InChI=1S/C8H13NO2/c1-9-5-3-4-7(6-9)8…
InChI=1S/C8H13NO2.BrH/c1-9-5-3-4-7(6…

EMBL-EBI

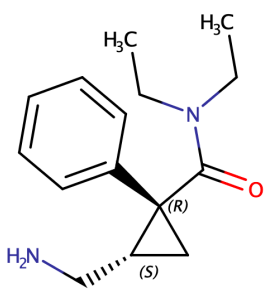
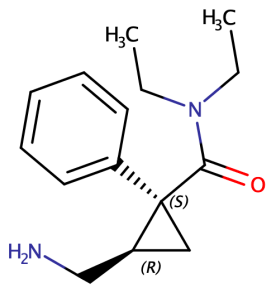
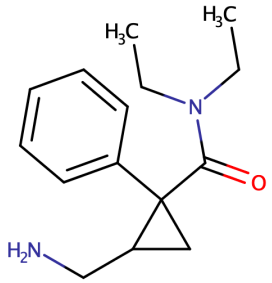
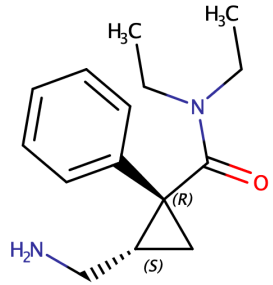
ChEMBL Compound Registration

- Standardised V2000 molfile as the primary chemical structure representation
- Standard InChI and corresponding hashed InChIKey are calculated from the molfile using software provided by InChI Trust (current version 1.05)
- Standard InChI and InChIKey used as a measure of uniqueness for a chemical structure in ChEMBL
- Compounds from different sources with the same Standard InChI and InChIKey are considered to be the same compound and are assigned the same ChEMBL identifier (CHEMBL_ID)
- However, V2000 molfile and Standard InChI have a few limitations, *e.g.*,
  - Do not support the use of relative stereochemistry, only absolute or no stereochemistry

ChEMBL Compound Registration

- Several drugs are formulated as racemic mixtures (e.g., Orphenadrine, Milnacipran) or as unequal mixture of both enantiomers (e.g., Amphetamine)
- Chemical representation of these mixtures is a challenge and different data providers may have different registration rules

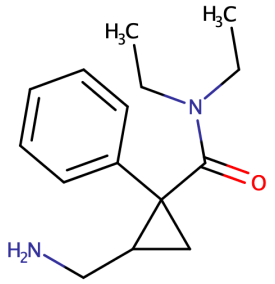
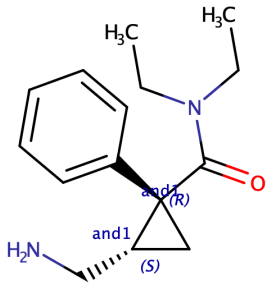
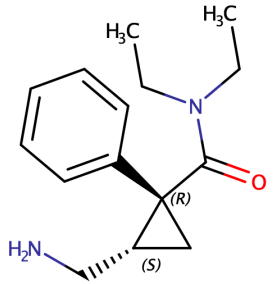
Example of structure representations from various sources for Milnacipran

Source	USAN	ChEMBL/DailyMed/DrugBank	PubChem/Kegg
Structure Representation	 		
Standard InChIKey	GCWZHWKRAHFLED-NATOKKMCSA-N	GJJFMKBJSRMPLA-UHFFFAOYSA-N	GJJFMKBJSRMPLA-HIFRSBDPSA-N

ChEMBL Compound Registration

- Relative stereochemistry can be annotated in V3000 molfiles
- ChEMBL has started to store it internally for manually annotated compounds

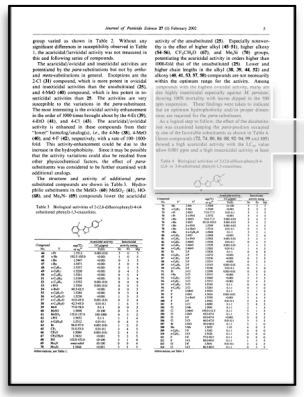
Structure representation of different molfile versions
(and corresponding Standard InChIKey) for Milnacipran

V2000 format	V3000 format		V2000 format
			
GJJFMKBJSRMPLA-UHFFFAOYSA-N	GJJFMKBJSRMPLA-HIFRSBDPSA-N	=	GJJFMKBJSRMPLA-HIFRSBDPSA-N

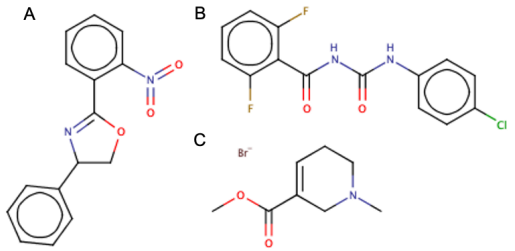
- V3000 molfile not fully represented by Standard InChI

InChI Applications in ChEMBL – Structure Pipeline

Documents



Compounds



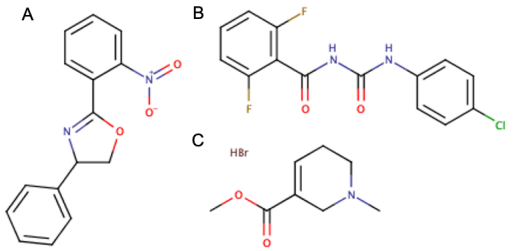
ChEMBL Structure Pipeline

Check

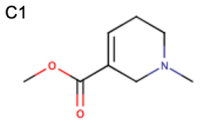
A `[ [ 5, "molecule has a bond with an illegal type" ], [ 2, "InChI: Omitted undefined stereo" ] ]`

B `[ [ 5, "molecule has a bond with an illegal type" ], [ 2, "InChI: Omitted undefined stereo" ] ]`

Standardizer



GetParent



Cross References

Hierarchy

compound C1 → compound C

Registration

InChI=1S/C14H9ClF2N2O2/c15-8-4-6-9…
InChI=1S/C15H12N2O3/c18-17(19)14-9…
InChI=1S/C8H13NO2/c1-9-5-3-4-7(6-9)8…
InChI=1S/C8H13NO2.BrH/c1-9-5-3-4-7(6…

EMBL-EBI

ChEMBL Structure Pipeline

- ChEMBL Structure Pipeline is comprised of three processes:
  - *Checker*: identifies and validates structures and identifies problems before structures are loaded into the database
  - *Standardizer*: processes (standardises) chemical structures according to a set of predefined rules
  - *GetParent*: generates parent structures based a set of rules and defined list of salts and solvents
- *Standardizer and GetParent* have been rewritten and adapted from rules originally implemented using a commercial software toolkit
- *Checker* was developed more recently in an attempt to identify problem structures

ChEMBL Structure Pipeline – *Checker*

- *Checker* process validates structures prior to loading compounds into ChEMBL
- If an error or problem is detected in the structure, a score is assigned
- The higher the score, more serious is the problem

Penalty Score	Penalty Explanation	Database Loader Actions
7	Error -9986 (Cannot process aromatic bonds) Illegal input InChI: Unknown element(s)	a fatal error and no data is loaded into ChEMBL
6	all atoms have zero coordinates InChI: Accepted unusual valence(s) InChI: Empty structure molecule has 3D coordinates molecule has a radical that is not found in the known list molecule has six (or more) atoms with exactly the same coordinates number of atoms less than 1 polymer information in mol file V3000 mol file	data is loaded into the database but without a molfile
5	InChI_RDKit/Mol stereo mismatch Mol/Inchi/RDKit stereo mismatch RDKit_Mol/InChI stereo mismatch molecule has a bond with an illegal stereo flag molecule has a bond with an illegal type molecule has a crossed bond in a ring molecule has two (or more) atoms with exactly the same coordinates	all data is loaded
2	InChI_Mol/RDKit stereo mismatch molecule has a stereo bond in a ring molecule has an atom with multiple stereo bonds molecule has a stereo bond to a stereocenter molecule has the 3D flag set for a 2D conformer Other InChI Warnings	

- InChI warnings are a component of the *Checker*

ChEMBL Structure Pipeline Publication

Bento *et al. J Cheminform* (2020) 12:51
https://doi.org/10.1186/s13321-020-00456-1

Journal of Cheminformatics

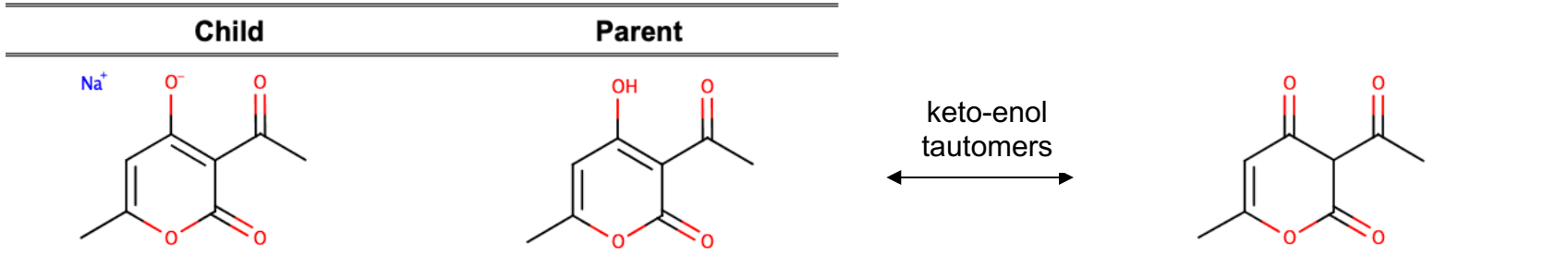
METHODOLOGY Open Access

An open source chemical structure curation pipeline using RDKit

A. Patrícia Bento[1], Anne Hersey[1], Eloy Félix[1], Greg Landrum[2], Anna Gaulton[1], Francis Atkinson[1,3], Louisa J. Bellis[1,4], Marleen De Veij[1] and Andrew R. Leach[1*]

ChEMBL Structure Pipeline Updates – *Standardizer*

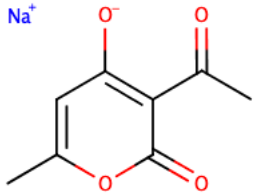
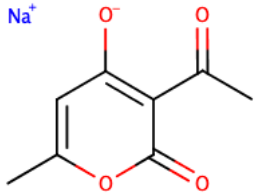
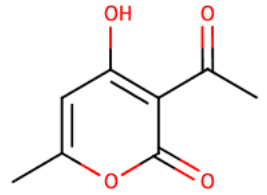
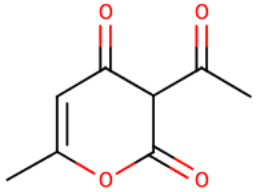
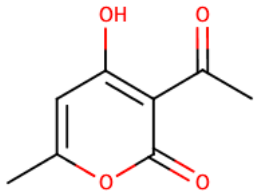
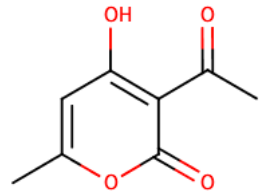
- Generate canonical tautomer
  - Standard InChI is independent of the tautomeric form of a compound and hence different tautomers of a compound will have the same Standard InChI. Therefore, in ChEMBL they are considered as being the same compound
  - However, Standard InChI has a few limitations, *e.g.*,
    - unable to recognise some keto-enol tautomers as being the same compound



ChEMBL Structure Pipeline Updates – *Standardizer*

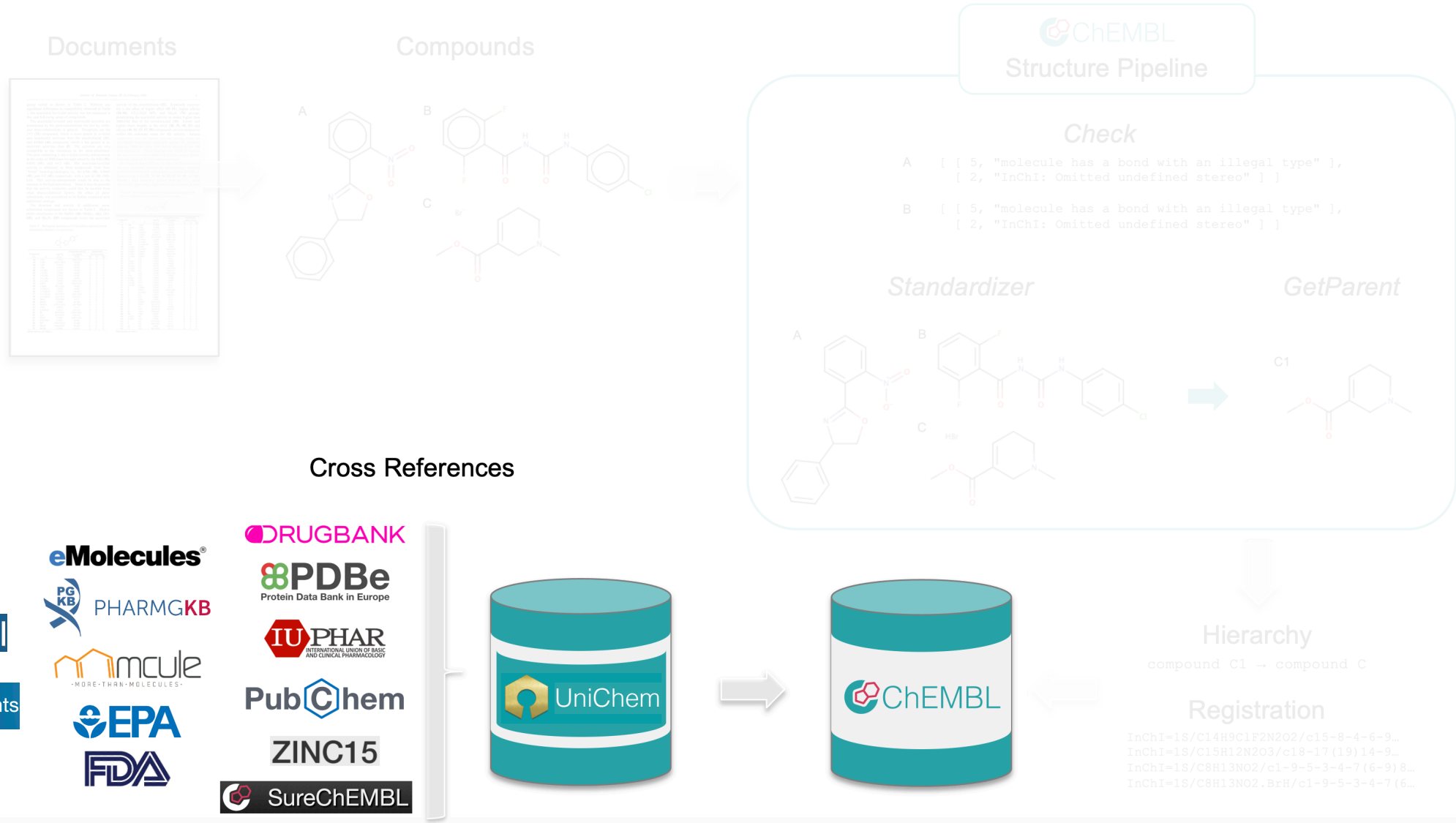
- Generate canonical tautomer

- Update

	Child	Canonical Tautomer	Parent Canonical Tautomer
SODIUM DEHYDROACETATE			
DEHYDROACETIC ACID			

Bioactivity data linked

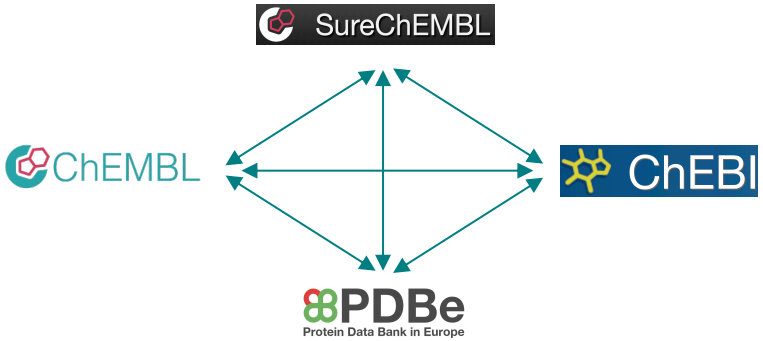
InChI Applications in ChEMBL – Cross References



UniChem

- Cross-references to compounds in other sources provided by UniChem
- UniChem is a compound identifier mapping service developed at EMBL-EBI that provides structure-based hyperlinks to chemical structures in other databases

- Historically, each EMBL-EBI resource would individually manage its own links to all other resources



- In UniChem, mappings are maintained centrally, allowing for the simple inclusion of additional resources

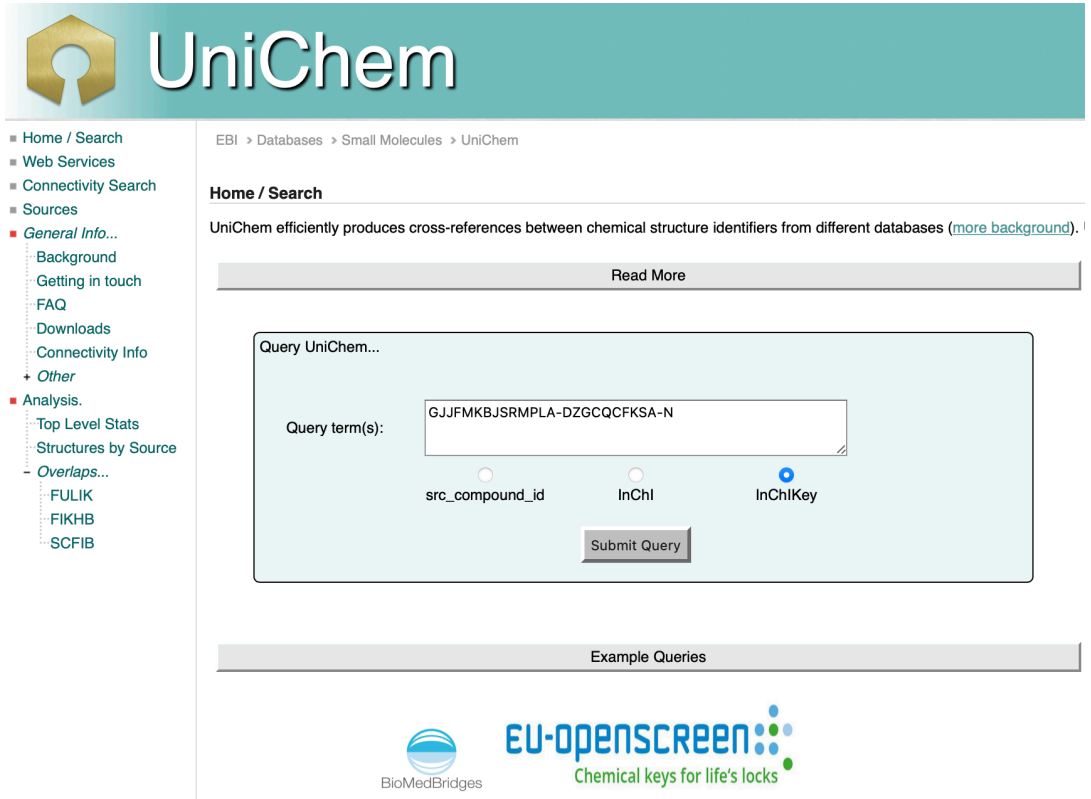
UniChem

- Standard InChI and InChIKey are the chemical structure normalizing key for UniChem
- For UniChem, each new structure (as defined by the Standard InChI) is assigned a UniChem identifier (UCI) within the system and this association between the structure and its UCI is never updated or deleted. The only updates permitted are to indicate whether the assignment is ‘current’ or ‘obsolete’.
- All structures and assignments of whatever historical status are captured
- UniChem Stats

Release Date	04-MAR-2021
Total number of Structures	176,340,296
Total number Assignments*	285,463,117
Number Current Assignments	214,367,308
Number Obsolete Assignments	71,095,809
Number of Sources	40

UniChem – Use cases

- UniChem can be used to:
  - Search for identical small molecule data across a wide variety of chemically aware resources, using either src_compound_ids, InChI or InChIKeys



Show 50 entries Apply filter:

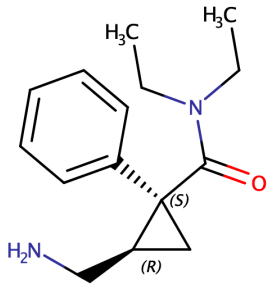
src_id	Source Name	src_compound_id	Currently Assigned	LR *	UCI **	Standard InChIKey
1	chembl	CHEMBL99946	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
7	chebi	136040	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
15	surechembl	SCHEMBL1414867	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
2	drugbank	DB08918	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
4	gtopdb	7435	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
9	zinc	ZINC000000000506	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
11	ibm	BCAB701D6C6F434F6E3261DFA5BDCE4A	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
14	fdasrs	UGM0326TXX	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
21	pubchem_tpharma	16010452	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
22	pubchem	6917779	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
29	nikkaji	J259.860D	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
31	bindingdb	50032379	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
34	drugcentral	4864	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
39	chemicalbook	CB13133687	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
40	dailymed_old	levomilnacipran	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
45	dailymed	LEVOMILNACIPRAN	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
46	clinicaltrials	F 2695	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
46	clinicaltrials	F-2695	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
46	clinicaltrials	F2695	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
46	clinicaltrials	FETZIMA	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
46	clinicaltrials	LEVOMILNACIPRAN	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
47	rxnorm	FETZIMA	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
47	rxnorm	LEVOMILNACIPRAN	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
47	rxnorm	LEVOMILNACIPRAN HYDROCHLORIDE	Yes		10620	GJJFMKBJSRMPLA-DZGCQCFKSA-N
src_id		src_compound_id				

Showing 1 to 24 of 24 entries

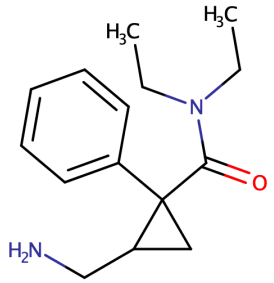
UniChem – Use cases

- UniChem can be used to:
  - Search for molecules with the same atom connectivity but different stereochemistry or isotopic composition or existing in a different salt form, using either src_compound_ids or InChIKeys

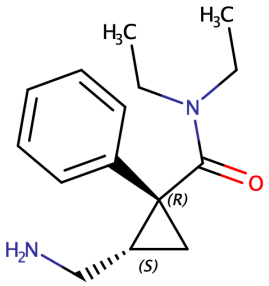
Different stereochemistry and isotopes



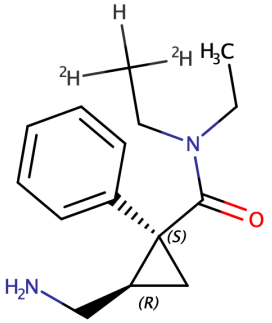
GJJFMKBJSRMPLA-DZGCQCFKSA-N



GJJFMKBJSRMPLA-UHFFFAOYSA-N



GJJFMKBJSRMPLA-HIFRSBDPSA-N

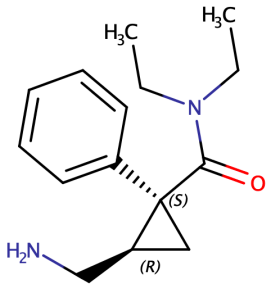


GJJFMKBJSRMPLA-PSBOALLFSA-N

UniChem – Use cases

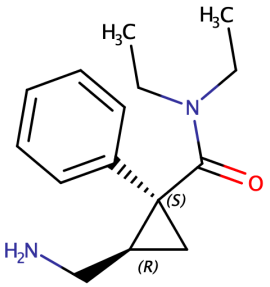
- UniChem can be used to:
  - Search for molecules with the same atom connectivity but different stereochemistry or isotopic composition or existing in a different salt form, using either src_compound_ids or InChIKeys

Different salt forms



GJJFMKBJSRMPLA-
DZGCQCFKSA-N

InChI=1S/C15H22N2O/
c1-3-17(4-2)14(18)15(10-13(15)11-16)12-8-6-5-7-9-12/
h5-9,13H,3-4,10-11,16H2,1-2H3/t13-,15+/m0/s1



HCl

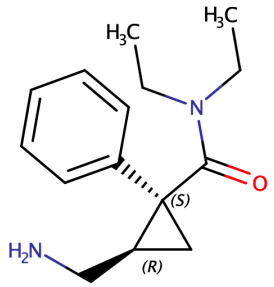
XNCDYJFPRPDERF-
NQQJLSKUSA-N

InChI=1S/C15H22N2O.ClH/
c1-3-17(4-2)14(18)15(10-13(15)11-16)12-8-6-5-7-9-12;/
h5-9,13H,3-4,10-11,16H2,1-2H3;1H/t13-,15+;/m0./s1

UniChem – Use cases

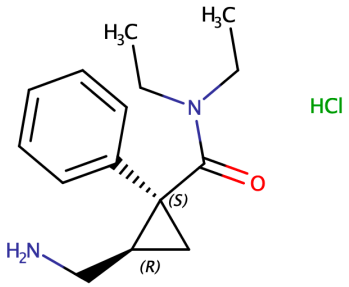
- UniChem can be used to:
  - Search for molecules with the same atom connectivity but different stereochemistry or isotopic composition or existing in a different salt form, using either src_compound_ids or InChIKeys

Different salt forms



GJJFMKBJSRMPLA-
DZGCQCFKSA-N

InChI=1S/**C15H22N2O**/
c1-3-17(4-2)14(18)15(10-13(15)11-16)12-8-6-5-7-9-12/
h5-9,13H,3-4,10-11,16H2,1-2H3/**t13-,15+**/**m0**/**s1**



HCl

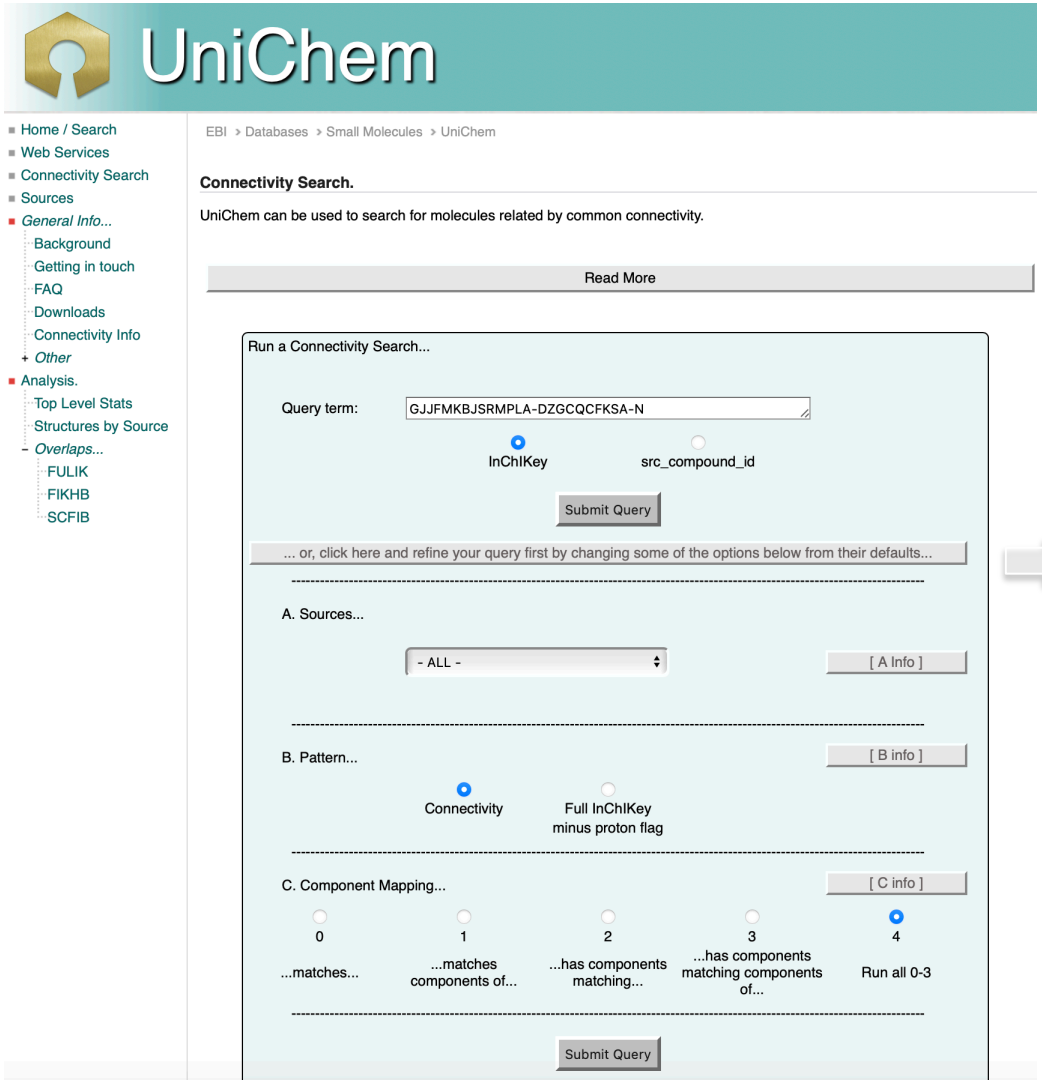
XNCDYJFPRPDERF-
NQQJLSKUSA-N

InChI=1S/**C15H22N2O**.ClH/
c1-3-17(4-2)14(18)15(10-13(15)11-16)12-8-6-5-7-9-12;/
h5-9,13H,3-4,10-11,16H2,1-2H3;1H/**t13-,15+**;/**m0**./**s1**

↕

GJJFMKBJSRMPLA-
DZGCQCFKSA-N

UniChem – Use cases



Layer	Description
/p	... *Protons. The number of protons that must be added to or removed from the formula to give the input composition.* (see * below)
/b	... *Double bond stereo.*
/t	... *Tetrahedral stereo.*
/m	... *Inversion flag for tetrahedral stereo. Note that the enantiomers of a chiral molecule will have the same /t layer but different /m flags; either /m0 or /m1.*
/s	... *Stereo type. Refers only to stereo that changes upon spatial inversion. Since inversion cannot change the double bond stereo, /s modifies the /t layer only, not the /b layer.*
/i	... *Isotope. Defines the presence of isotopic atoms, and includes changes in stereochemistry created by their presence.*

CMR. Query InChI...	src_id	Source	src_compound_id	Asn	label	p	b	t	m	s	i	En	src_compound_id InChIKey	#
...matches cpts of...	1	ChEMBL	CHEMBL1237129	1	HCl		0	0	1	0	0	1	XNCDYJFPRPDERF-PBCQUBLHSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL2094056	1	HCl		0	1	0	0	0		XNCDYJFPRPDERF-SLHAJLBXSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL2105732	1	HCl		0	0	0	0	0		XNCDYJFPRPDERF-NQQJLSKUSA-N	1
...matches...	1	ChEMBL	CHEMBL252923	1		0	0	0	1	0	0	1	GJJFMKBJSRMPLA-HIFRSBDPSA-N	1
...matches...	1	ChEMBL	CHEMBL259209	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL3527085	1	HCl		0	1	1	1	1		XNCDYJFPRPDERF-ZEVXLFMTSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL4297064	1	HCl		0	1	1	1	0		XNCDYJFPRPDERF-UHFFFAOYSA-N	1
...matches...	1	ChEMBL	CHEMBL471001	1		0	0	1	0	0	0		GJJFMKBJSRMPLA-ZFWWWQNUSA-N	1
...matches...	1	ChEMBL	CHEMBL99946	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	2	DrugBank	DB04896	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches...	2	DrugBank	DB08918	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	4	Guide to Pharmacology	7435	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	4	Guide to Pharmacology	7436	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches cpts of...	5	PubChem: Drugs of the Future	12013109	1	HCl		0	0	0	0	0		XNCDYJFPRPDERF-NQQJLSKUSA-N	1
...matches...	7	ChEBI	135005	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches...	7	ChEBI	136040	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches cpts of...	7	ChEBI	31854	1	HCl		0	0	1	0	0	1	XNCDYJFPRPDERF-PBCQUBLHSA-N	1
...matches...	7	ChEBI	94468	1		0	0	0	1	0	0	1	GJJFMKBJSRMPLA-HIFRSBDPSA-N	1

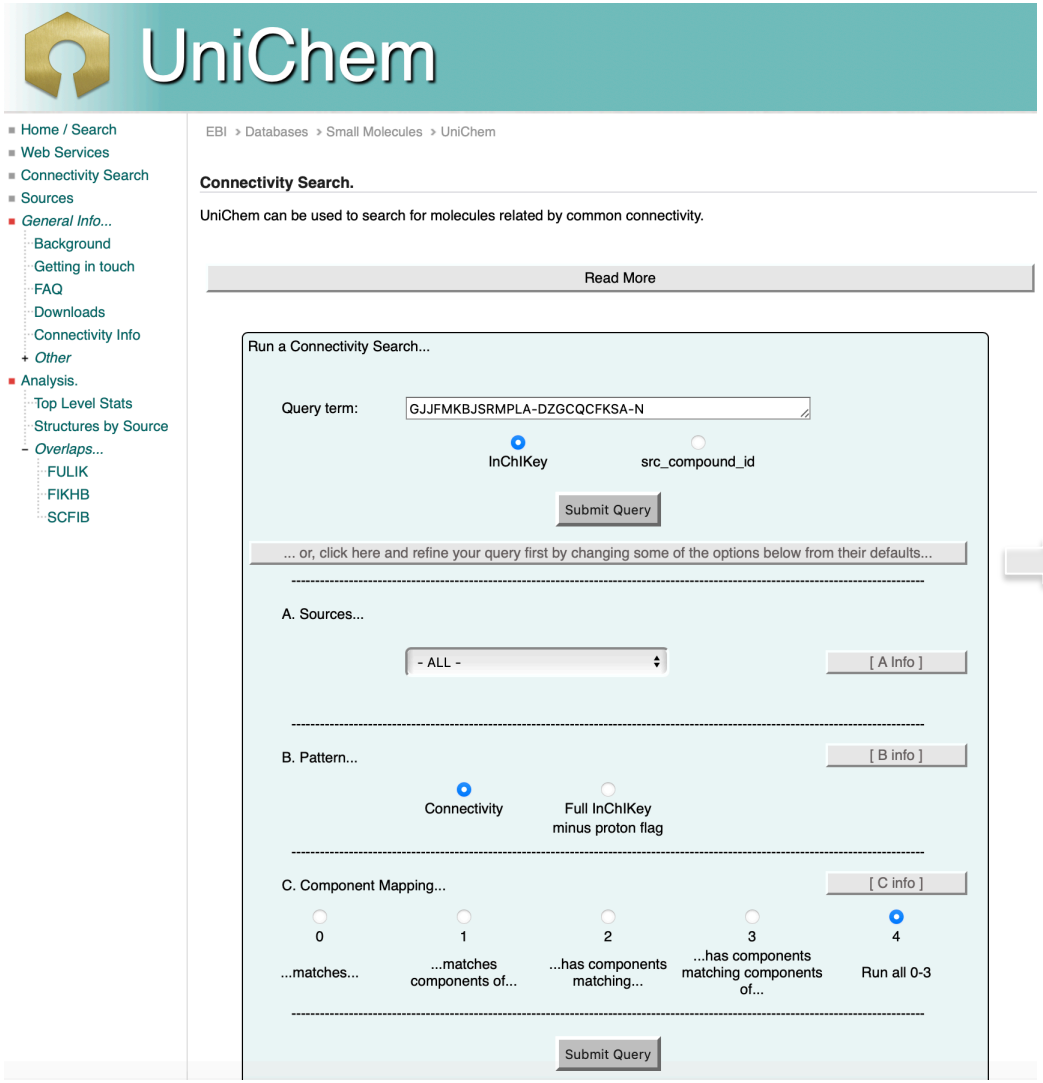
Showing 1 to 15 of 238 entries

First Previous 1 2 3 4 5 Next Last

www.ebi.ac.uk/unichem/

EMBL-EBI

UniChem – Use cases



Layer	Description
/p	... *Protons. The number of protons that must be added to or removed from the formula to give the input composition.* (see * below)
/b	... *Double bond stereo.*
/t	... *Tetrahedral stereo.*
/m	... *Inversion flag for tetrahedral stereo. Note that the enantiomers of a chiral molecule will have the same /t layer but different /m flags; either /m0 or /m1.*
/s	... *Stereo type. Refers only to stereo that changes upon spatial inversion. Since inversion cannot change the double bond stereo, /s modifies the /t layer only, not the /b layer.*
/i	... *Isotope. Defines the presence of isotopic atoms, and includes changes in stereochemistry created by their presence.*

CMR. Query InChI...	src_id	Source	src_compound_id	Asn	label	p	b	t	m	s	i	En	src_compound_id InChIKey	#
...matches cpts of...	1	ChEMBL	CHEMBL1237129	1	HCl		0	0	1	0	0	1	XNCDYJFPRPDERF-PBCQUBLHSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL2094056	1	HCl		0	1	0	0	0		XNCDYJFPRPDERF-SLHAJLBXSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL2105732	1	HCl		0	0	0	0	0		XNCDYJFPRPDERF-NQQJLSKUSA-N	1
...matches...	1	ChEMBL	CHEMBL252923	1		0	0	0	1	0	0	1	GJJFMKBJSRMPLA-HIFRSBDPSA-N	1
...matches...	1	ChEMBL	CHEMBL259209	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL3527085	1	HCl		0	1	1	1	1		XNCDYJFPRPDERF-ZEVXLFMTSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL4297064	1	HCl		0	1	1	1	0		XNCDYJFPRPDERF-UHFFFAOYSA-N	1
...matches...	1	ChEMBL	CHEMBL471001	1		0	0	1	0	0	0		GJJFMKBJSRMPLA-ZFWWWQNUSA-N	1
...matches...	1	ChEMBL	CHEMBL99946	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	2	DrugBank	DB04896	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches...	2	DrugBank	DB08918	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	4	Guide to Pharmacology	7435	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	4	Guide to Pharmacology	7436	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches cpts of...	5	PubChem: Drugs of the Future	12013109	1	HCl		0	0	0	0	0		XNCDYJFPRPDERF-NQQJLSKUSA-N	1
...matches...	7	ChEBI	135005	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches...	7	ChEBI	136040	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches cpts of...	7	ChEBI	31854	1	HCl		0	0	1	0	0	1	XNCDYJFPRPDERF-PBCQUBLHSA-N	1
...matches...	7	ChEBI	94468	1		0	0	0	1	0	0	1	GJJFMKBJSRMPLA-HIFRSBDPSA-N	1

Showing 1 to 15 of 238 entries

First Previous 1 2 3 4 5 Next Last

www.ebi.ac.uk/unichem/

EMBL-EBI

UniChem – Use cases

UniChem

- Home / Search
- Web Services
- Connectivity Search
- Sources
- *General Info...*
  - Background
  - Getting in touch
  - FAQ
  - Downloads
  - Connectivity Info
- *Other*
- Analysis.
  - Top Level Stats
  - Structures by Source
- *Overlaps...*
  - FULIK
  - FIKHB
  - SCFIB

EBI › Databases › Small Molecules › UniChem

Connectivity Search.

UniChem can be used to search for molecules related by common connectivity.

Read More

Run a Connectivity Search...

Query term: GJJFMKBJSRMPLA-DZGCQCFKSA-N

(•) InChIKey ( ) src_compound_id

Submit Query

... or, click here and refine your query first by changing some of the options below from their defaults...

A. Sources...

- ALL - [A Info]

B. Pattern... [B info]

(•) Connectivity ( ) Full InChIKey minus proton flag

C. Component Mapping... [C info]

0	1	2	3	4
...matches...	...matches components of...	...has components matching...	...has components matching components of...	Run all 0-3

Submit Query

Layer	Description
/p	... *Protons. The number of protons that must be added to or removed from the formula to give the input composition.* (see * below)
/b	... *Double bond stereo.*
/t	... *Tetrahedral stereo.*
/m	... *Inversion flag for tetrahedral stereo. Note that the enantiomers of a chiral molecule will have the same /t layer but different /m flags; either /m0 or /m1.*
/s	... *Stereo type. Refers only to stereo that changes upon spatial inversion. Since inversion cannot change the double bond stereo, /s modifies the /t layer only, not the /b layer.*
/i	... *Isotope. Defines the presence of isotopic atoms, and includes changes in stereochemistry created by their presence.*

CMR. Query InChI...	src_id	Source	src_compound_id	Asn	label	p	b	t	m	s	i	En	src_compound_id InChIKey	#
...matches cpts of...	1	ChEMBL	CHEMBL1237129	1	HCl		0	0	1	0	0	1	XNCDYJFPRPDERF-PBCQUBLHSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL2094056	1	HCl		0	1	0	0	0		XNCDYJFPRPDERF-SLHAJLBXSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL2105732	1	HCl		0	0	0	0	0		XNCDYJFPRPDERF-NQQJLSKUSA-N	1
...matches...	1	ChEMBL	CHEMBL252923	1		0	0	0	1	0	0	1	GJJFMKBJSRMPLA-HIFRSBDPSA-N	1
...matches...	1	ChEMBL	CHEMBL259209	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL3527085	1	HCl		0	1	1	1	1		XNCDYJFPRPDERF-ZEVXLFMTSA-N	1
...matches cpts of...	1	ChEMBL	CHEMBL4297064	1	HCl		0	1	1	1	0		XNCDYJFPRPDERF-UHFFFAOYSA-N	1
...matches...	1	ChEMBL	CHEMBL471001	1		0	0	1	0	0	0		GJJFMKBJSRMPLA-ZFWWWQNUSA-N	1
...matches...	1	ChEMBL	CHEMBL99946	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	2	DrugBank	DB04896	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches...	2	DrugBank	DB08918	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	4	Guide to Pharmacology	7435	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches...	4	Guide to Pharmacology	7436	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches cpts of...	5	PubChem: Drugs of the Future	12013109	1	HCl		0	0	0	0	0		XNCDYJFPRPDERF-NQQJLSKUSA-N	1
...matches...	7	ChEBI	135005	1		0	0	1	1	1	0		GJJFMKBJSRMPLA-UHFFFAOYSA-N	1
...matches...	7	ChEBI	136040	1		0	0	0	0	0	0		GJJFMKBJSRMPLA-DZGCQCFKSA-N	1
...matches cpts of...	7	ChEBI	31854	1	HCl		0	0	1	0	0	1	XNCDYJFPRPDERF-PBCQUBLHSA-N	1
...matches...	7	ChEBI	94468	1		0	0	0	1	0	0	1	GJJFMKBJSRMPLA-HIFRSBDPSA-N	1

Showing 1 to 15 of 238 entries First Previous 1 2 3 4 5 Next Last

www.ebi.ac.uk/unichem/

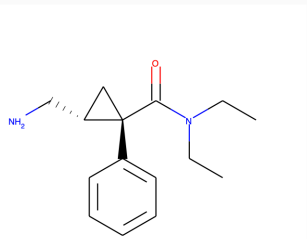
EMBL-EBI

UniChem – Use cases

- UniChem can be used to:
  - Automatically generate hyperlinks from within a resource to other resources

Compound Report Card

Name And Classification



Structure Search

ID:	CHEMBL99946
Name:	LEVOMILNACIPRAN
Max Phase:	4 Approved
Molecular Formula:	C15H22N2O
Molecular Weight:	246.35
ChEMBL Synonyms:	F2695 F 2695 F-2695 LEVOMILNACIPRAN
Synonyms From Alternate Forms:	F 2695 HYDROCHLORIDE F-2695 HYDROCHLORIDE F2695 HYDROCHLORIDE LEVOMILNACIPRAN HYDROCHLORIDE
Trade Names:	FETZIMA
Molecule Type:	Small molecule

UniChem Cross References

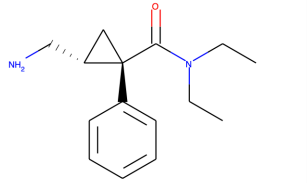
BindingDB	50032379
clinicaltrials	LEVOMILNACIPRAN, F 2695, F2695, FETZIMA, F-2695
ChemicalBook	CB13133687
ChEBI	136040
DailyMed	levomilnacipran, LEVOMILNACIPRAN
DrugBank	DB08918
DrugCentral	4864
FDA SRS	UGM0326TXX
GSRS	f319b645-15ca-418f-bdf8-89c1027b29d8
Guide to Pharmacology	7435
IBM Patent System	BCAB701D6C6F434F6E3261DFA5BDCE4A
IBM Patents	US7309799, US20100197797, US20040259953, US20040106681, US7074833, WO2004075886A1, EP2110129A1, US20050282898, US20090048233, US7038085, US20040162334, US20100016636, EP1601349B1, EP1601349A1, US7005452, US20060014837, US20100274050, US20050282859, US20040142904, US20050096395, US20040034101

UniChem – Use cases

- UniChem can be used to:
  - Automatically generate hyperlinks from within a resource to other resources

Compound Report Card

Name And Classification



Structure Search

ID:	CHEMBL99946
Name:	LEVOMILNACIPRAN
Max Phase:	4 Approved
Molecular Formula:	C15H22N2O
Molecular Weight:	246.35
ChEMBL Synonyms:	F2695 F 2695 F-2695 LEVOMILNACIPRAN
Synonyms From Alternate Forms:	F 2695 HYDROCHLORIDE F-2695 HYDROCHLORIDE F2695 HYDROCHLORIDE LEVOMILNACIPRAN HYDROCHLORIDE
Trade Names:	FETZIMA
Molecule Type:	Small molecule

UniChem Connectivity Layer Cross References

UniChem Connectivity Layer Cross References for CHEMBL99946

GJJFMKBJSRMPLA-DZGCQCFKSA-N

Full InChI:

InChI=1S/C15H22N2O/c1-3-17(4-2)14(18)15(10-13(15)11-16)12-8-6-5-7-9-12/h5-9,13H,3-4,10-11,16H2,1-2H3/t13-,15+/m0/s1

Source	Identical Component	S	I	P	SI	IP	SP	SIP
ChEMBL	CHEMBL99946 CHEMBL2105732	CHEMBL252923 CHEMBL4297064 CHEMBL471001 CHEMBL2094056 CHEMBL259209 CHEMBL1237129			CHEMBL3527085			
DrugBank	DB08918	DB04896						
Guide to Pharmacology	7435	7436						
PubChem: Drugs of the Future	12013109							
ChEBI	136040	94468 135005 31854						
NIH Clinical Collection		SAM001247035						
ZINC	ZINC000000000506	ZINC000000014037 ZINC000005649654 ZINC000001846944						
eMolecules	31706396 30100405	36754037 36500460 36500460 30069726 29543180			36754038			
IBM Patent System	BCAB701D6C6F434F6E3261DFA5BDCE4A	7B5D61094EBF88BDDBD5CAC963214F7A AACD9B490ECD4EA5F9FD06A92ED88D94 98ED58873FC70F5561D8B5B6D53849CB 0E9EA3DAADBB667A092086134ABFAEF8 DCFA8846FBE580E408339BE0C13975EC 0EC7C98A888839109AE7F8FC4759DE34						
FDA SRS	UGM0326TXX 371U2ZK31U	ES1O38J3C4						

UniChem Publications

Chambers *et al. Journal of Cheminformatics* 2013, **5**:3
http://www.jcheminf.com/content/5/1/3

Journal of Cheminformatics

DATABASE **Open Access**

UniChem: a unified chemical structure cross-referencing and identifier tracking system

Jon Chambers[1*], Mark Davies[1], Anna Gaulton[1], Anne Hersey[1], Sameer Velankar[2], Robert Petryszak[3], Janna Hastings[4], Louisa Bellis[1], Shaun McGlinchey[1] and John P Overington[1]

Chambers *et al. Journal of Cheminformatics* 2014, **6**:43
http://www.jcheminf.com/content/6/1/43

Journal of Cheminformatics

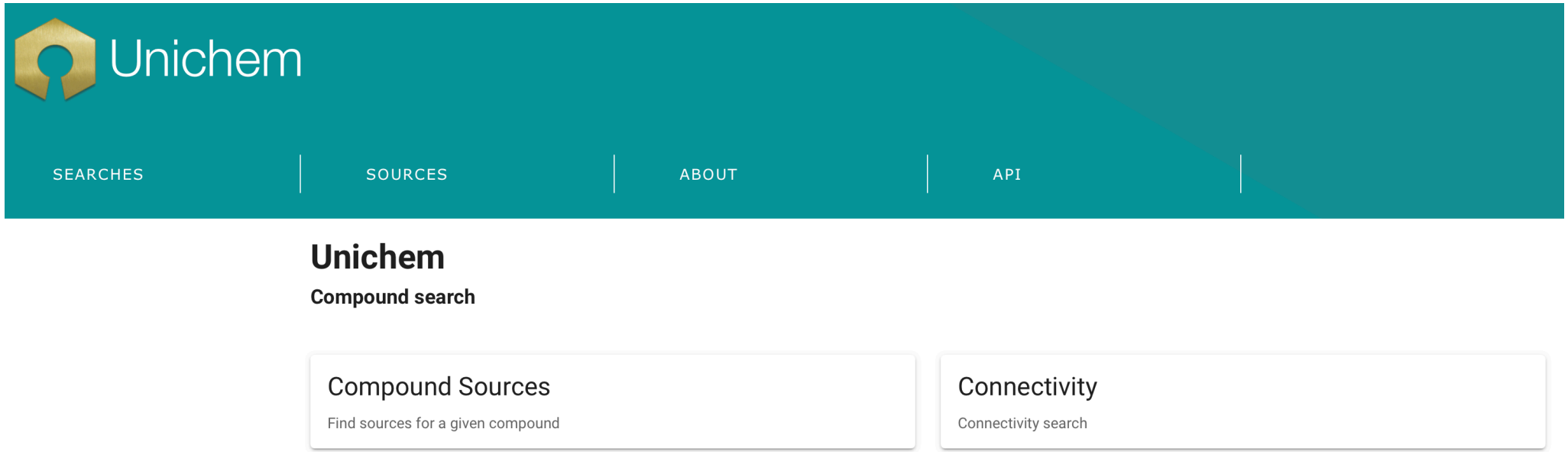
DATABASE **Open Access**

UniChem: extension of InChI-based compound mapping to salt, connectivity and stereochemistry layers

Jon Chambers*, Mark Davies, Anna Gaulton, George Papadatos, Anne Hersey and John P Overington

UniChem Update Coming Soon

- New interface
- New functionalities



Acknowledgements

- ChEMBL Structure Pipeline

ChEMBL

Anne Hersey
Patrícia Bento
Eloy Félix
Anna Gaulton
Francis Atkinson
Louisa Bellis
Marleen de Veij
Andrew Leach

RDKit Support and Chemical Data Science - T5 Informatics

Greg Landrum

- UniChem

ChEMBL

Jon Chambers
Anne Hersey
Anna Gaulton
Mark Davies
George Papadatos
Shaun McGlinchey
Louisa Bellis
James Blackshaw
John Overington

ChEBI

Chris Steinbeck
Janna Hastings

PDBe
Protein Data Bank in Europe

Sameer Velankar

Expression Atlas

Robert Petryszak

Funding:

wellcome trust
Strategic Award

EMBL-EBI

BioMedBridges

EU-OPENSCREEN
Chemical keys for life's locks