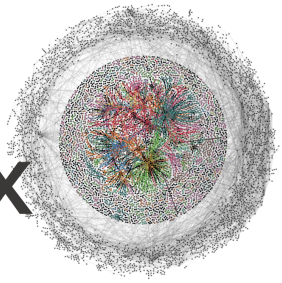
EPFL

Laboratory of
Computational
Systems
Biotechnology

ATLASx



User guide

Navigating known and hypothetical biochemical space

Content

Getting started

Create a user account

Please fill out the online request form:

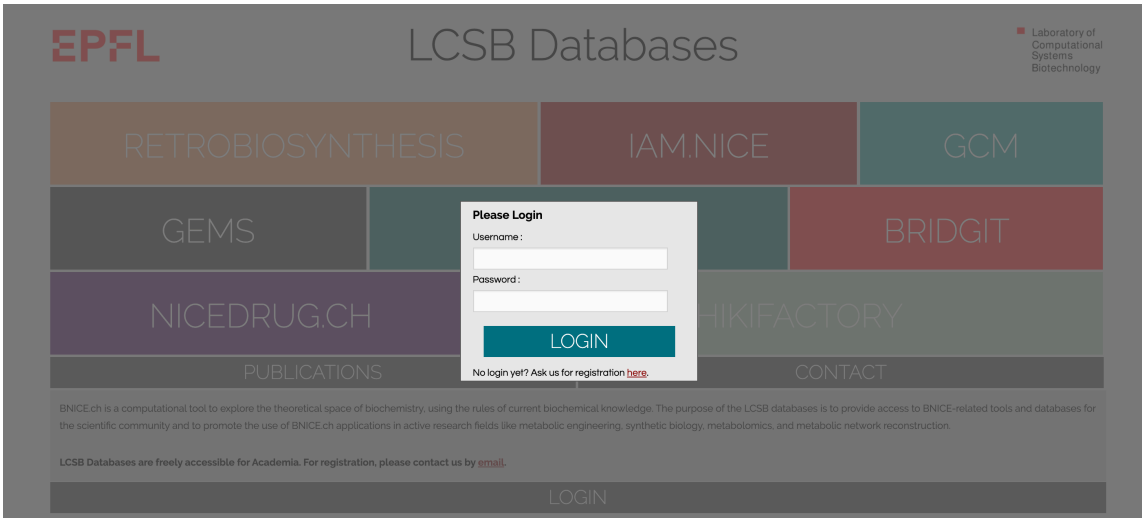
https://lcsb-databases.epfl.ch/pathways/database_requests

To apply for an academic license, please use your academic email address when registering.

We will send you a license agreement to be signed, and provide your username & password

Log in

Click on the LOGIN tab at the bottom of the LCSB webpage and insert your login info:



Alternatively, click on any tab within ATLASx and it will request your login credentials.

ATLASx database and website content

Database scopes

ATLASx databases organizes known and predicted biochemical reactions in *3 different scopes:*

- **bioDB** – unification of publicly available biochemical databases (e.g., KEGG, SEED, MetaCyc, BiGG, Brenda), also includes bioactive compounds (e.g., from ChEBI, CHEMBL, KEGG Drug)
- **bioATLAS** – predicted biochemical reactions within the biological and bioactive compound space
- **chemATLAS** – predicted biochemical reactions involving compounds from the chemical compound space (PubChem)

Data organization

In LCSB databases, we have unique identifiers (LCSB IDs) for each compound, reaction, and biotransformation (i.e., a substrate-product pair, or reactant pair. One biotransformation can be performed by more than one reaction.)

Database entries can be accessed directly via their LCSB ID:

- Compounds: https://lcsb-databases.epfl.ch/Graph2/**loadCompound**/1467886019
- Reactions: https://lcsb-databases.epfl.ch/Graph2/**loadReaction**/2592908327
- Biotransformations: https://lcsb-databases.epfl.ch/Graph2/**loadReactantPair**/2677718711

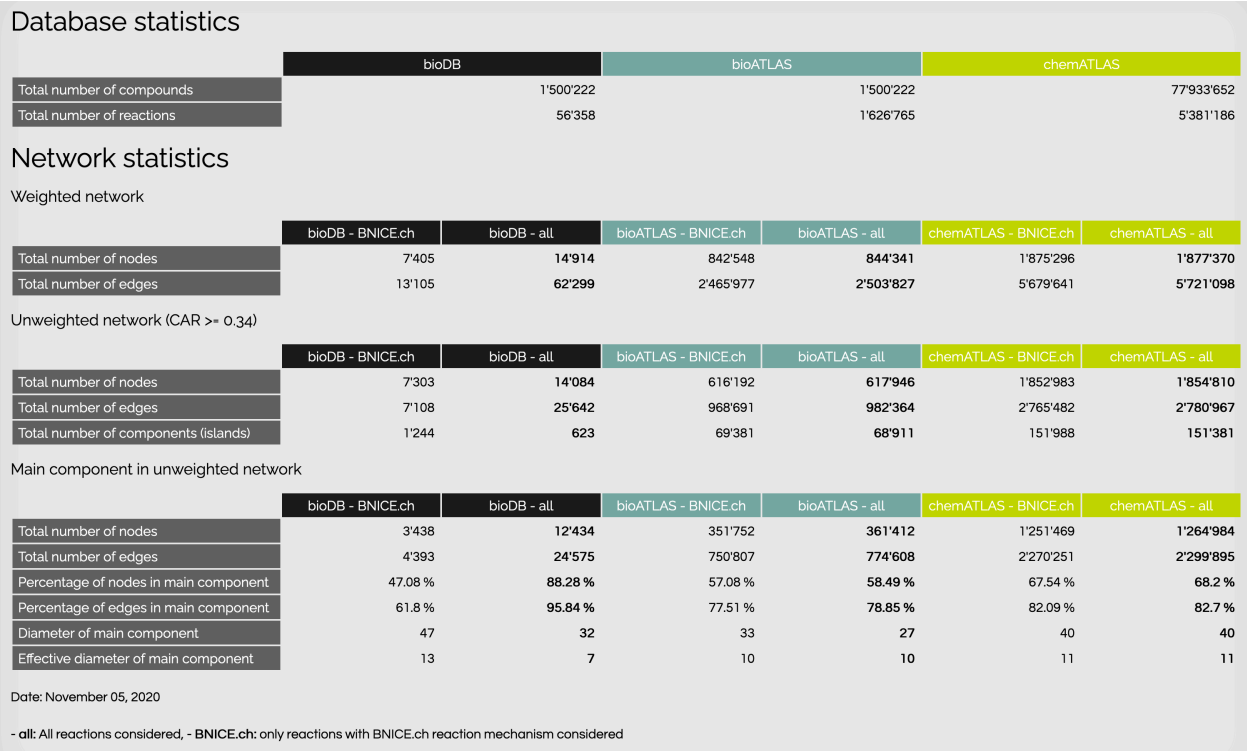
Website organization

The ATLASx website is organized in 3 parts:

- Statistics overview**:** Current statistics on compounds and reactions in ATLASx
- Search for compounds, reactions, pathways**:** Search interface to find compounds, reactions, and pathways
- Analysis platform**:** Additional analysis tools to query ATLASx

Statistics overview

To get updated compound and reaction counts of ATLASx, go to the STATISTICS tab.



Database statistics

	bioDB	bioATLAS	chemATLAS
Total number of compounds	1'500'222	1'500'222	77'933'652
Total number of reactions	56'358	1'626'765	5'381'186

Network statistics

Weighted network

	bioDB - BNICE.ch	bioDB - all	bioATLAS - BNICE.ch	bioATLAS - all	chemATLAS - BNICE.ch	chemATLAS - all
Total number of nodes	7'405	**14'914**	842'548	**844'341**	1'875'296	**1'877'370**
Total number of edges	13'105	**62'299**	2'465'977	**2'503'827**	5'679'641	**5'721'098**

Unweighted network (CAR >= 0.34)

	bioDB - BNICE.ch	bioDB - all	bioATLAS - BNICE.ch	bioATLAS - all	chemATLAS - BNICE.ch	chemATLAS - all
Total number of nodes	7'303	**14'084**	616'192	**617'946**	1'852'983	**1'854'810**
Total number of edges	7'108	**25'642**	968'691	**982'364**	2'765'482	**2'780'967**
Total number of components (islands)	1'244	**623**	69'381	**68'911**	151'988	**151'381**

Main component in unweighted network

	bioDB - BNICE.ch	bioDB - all	bioATLAS - BNICE.ch	bioATLAS - all	chemATLAS - BNICE.ch	chemATLAS - all
Total number of nodes	3'438	**12'434**	351'752	**361'412**	1'251'469	**1'264'984**
Total number of edges	4'393	**24'575**	750'807	**774'608**	2'270'251	**2'299'895**
Percentage of nodes in main component	47.08 %	**88.28 %**	57.08 %	**58.49 %**	67.54 %	**68.2 %**
Percentage of edges in main component	61.8 %	**95.84 %**	77.51 %	**78.85 %**	82.09 %	**82.7 %**
Diameter of main component	47	**32**	33	**27**	40	**40**
Effective diameter of main component	13	**7**	10	**10**	11	**11**

Date: November 05, 2020

- **all:** All reactions considered, - **BNICE.ch:** only reactions with BNICE.ch reaction mechanism considered

The table is updated once a day to give the current statistics of ATLASx.

- *Database statistics:* numbers of compounds and reactions in the different database scopes
- *Network statistics:* Compounds (nodes) connected by biotransformations (or reactant pairs, edges)
    - *Weighted network:* networks underlying the pathway search algorithm
    - *Unweighted network:* All reactant pairs (edges) with a CAR < 0.34 are removed, and the remaining edges are assigned a weight = 1.
    - *Main component:* From the unweighted network, the main component is extracted to calculate network properties (i.e., size, diameter)

NOTE: Since the database is updated continuously, the numbers presented here might not match the numbers given in the paper.

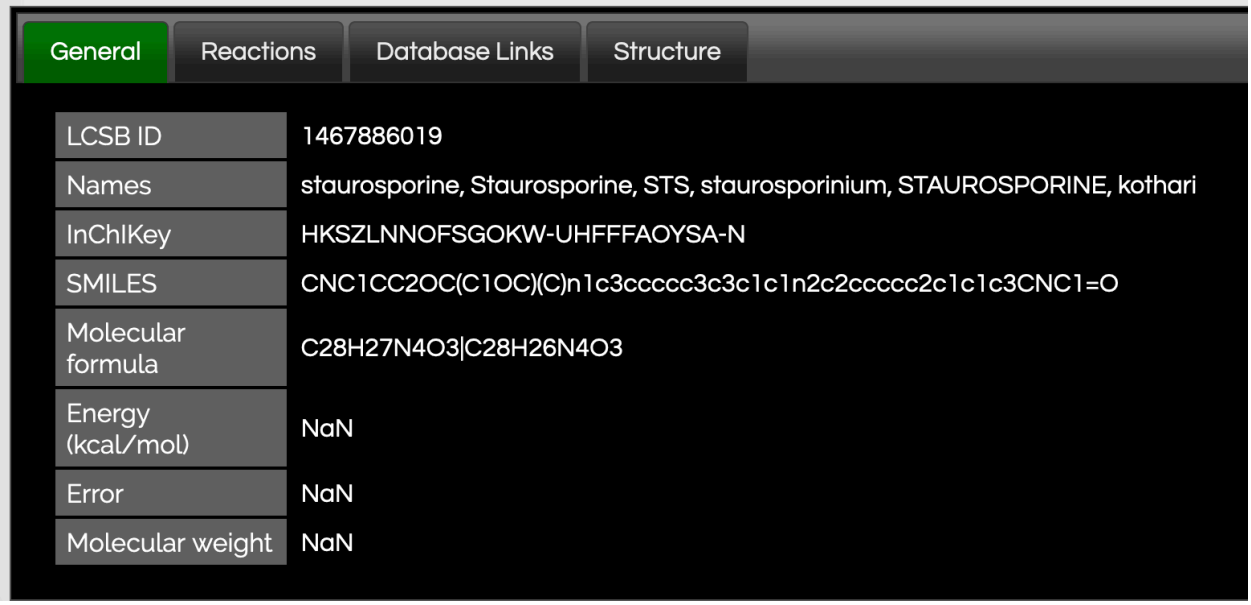
Search for compounds, reactions, pathways

To query ATLASx, go to the SEARCH tab.

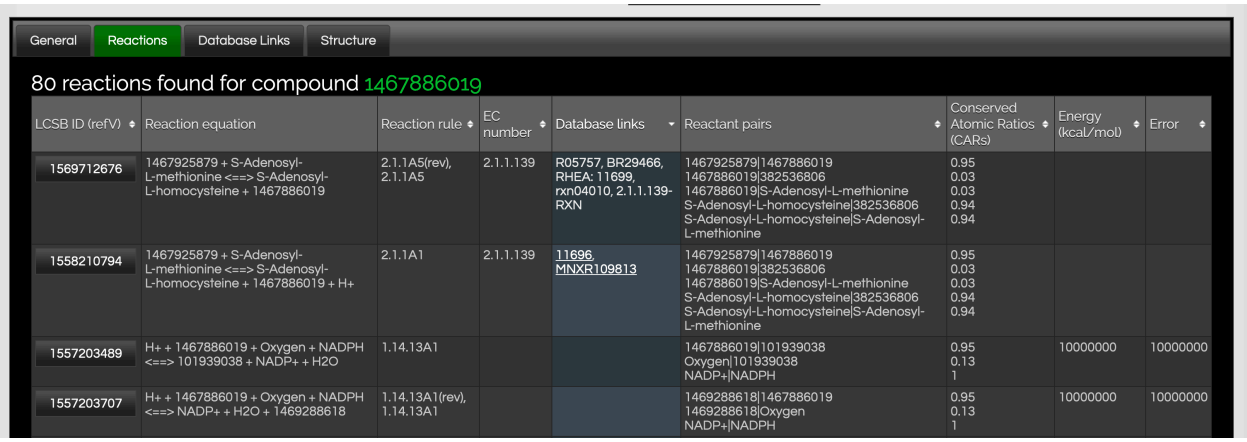
Find compounds

1. Search for compound by name, InChIKey, or LCSB ID

2. Output
    a. General overview



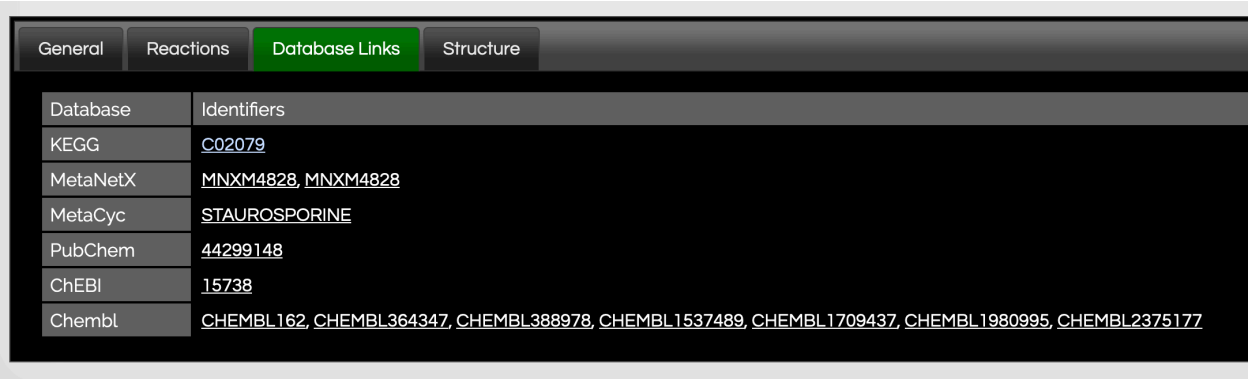
b. All reactions involving the query compound



LCSB ID (refV)	Reaction equation	Reaction rule	EC number	Database links	Reactant pairs	Conserved Atomic Ratios (CARs)	Energy (kcal/mol)	Error
1569712676	1467925879 + S-Adenosyl-L-methionine <==> S-Adenosyl-L-homocysteine + 1467886019	2.1.1A5(rev), 2.1.1A5	2.1.1.139	R05757, BR29466, RHEA: 11699, rxn04010, 2.1.1.139-RXN	1467925879|1467886019 1467886019|382536806 1467886019|S-Adenosyl-L-methionine S-Adenosyl-L-homocysteine|382536806 S-Adenosyl-L-homocysteine|S-Adenosyl-L-methionine	0.95 0.03 0.03 0.94 0.94		
1558210794	1467925879 + S-Adenosyl-L-methionine <==> S-Adenosyl-L-homocysteine + 1467886019 + H+	2.1.1A1	2.1.1.139	11696, MNXR109813	1467925879|1467886019 1467886019|382536806 1467886019|S-Adenosyl-L-methionine S-Adenosyl-L-homocysteine|382536806 S-Adenosyl-L-homocysteine|S-Adenosyl-L-methionine	0.95 0.03 0.03 0.94 0.94		
1557203489	H+ + 1467886019 + Oxygen + NADPH <==> 101939038 + NADP+ + H2O	1.14.13A1			1467886019|101939038 Oxygen|101939038 NADP+|NADPH	0.95 0.13 1	10000000	10000000
1557203707	H+ + 1467886019 + Oxygen + NADPH <==> NADP+ + H2O + 1469288618	1.14.13A1(rev), 1.14.13A1			1469288618|1467886019 1469288618|Oxygen NADP+|NADPH	0.95 0.13 1	10000000	10000000

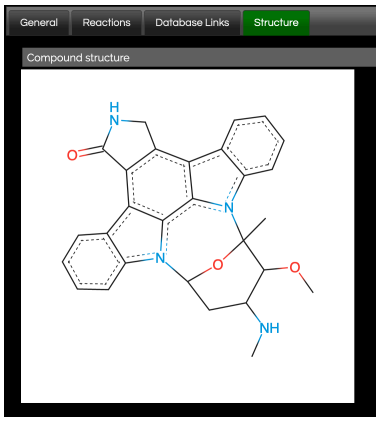
Note: The reactant pairs column shows all possible substrate-product pairs and their associated CAR (Conserved Atom Ratio, a metric between 0 and 1 indicating the ratio of conserved atoms between the substrate and the product)

c. Links to external databases



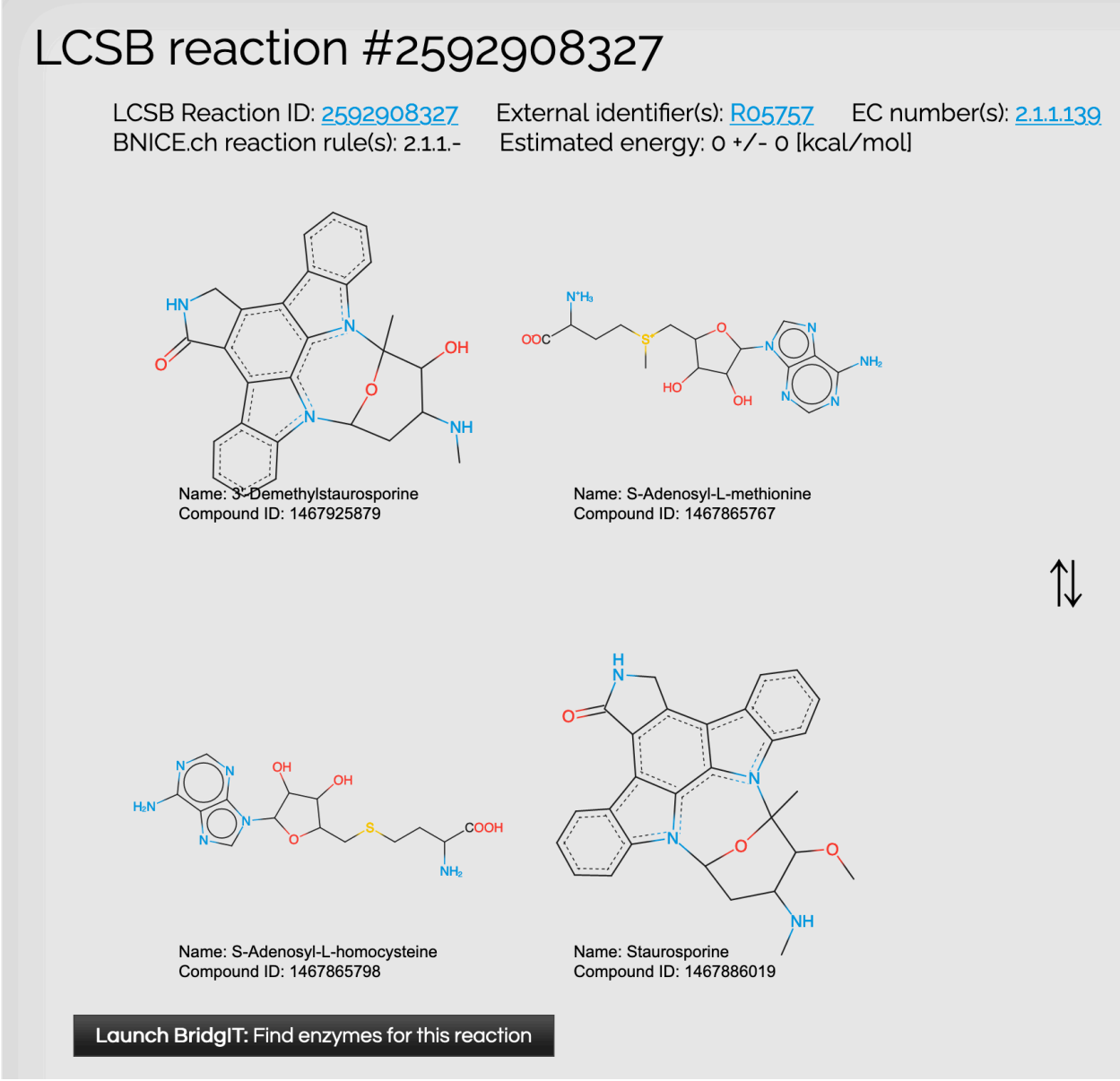
Database	Identifiers
KEGG	C02079
MetaNetX	MNXM4828, MNXM4828
MetaCyc	STAUROSPORINE
PubChem	44299148
ChEBI	15738
Chembl	CHEMBL162, CHEMBL364347, CHEMBL388978, CHEMBL1537489, CHEMBL1709437, CHEMBL1980995, CHEMBL2375177

d. Compound structure



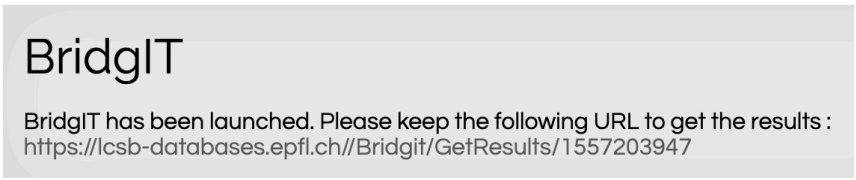
Find reactions

1. Search for one of the substrates in the compound search, save its LCSB ID.
2. Search for one of the products in the compound search, show all reactions.
3. Search (Ctrl+F) for the substrate ID within the reactions.
4. Click on the LCSB ID of the reaction for more information.



Find enzymes for novel reactions

1. Click on the button “Launch BridgIT: Find enzymes for this reaction”.
2. You will be redirected to a waiting page, while BridgIT finds enzymes.



3. Click on the link to get the results (be patient, it may take a while). As soon as the results are ready, your browser asks you to download a results file.
4. Open the file in a spreadsheet editor (Microsoft Excel, Numbers, …):

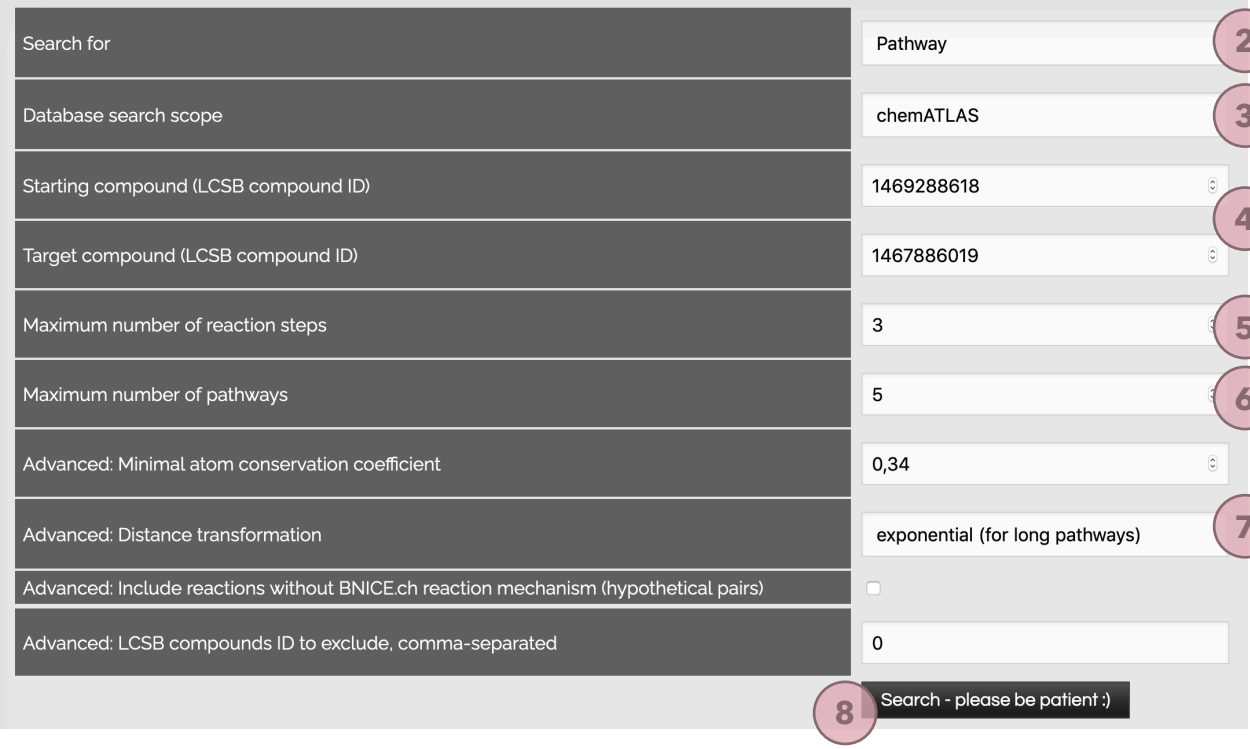
	A	B	C	D	E	F
1	**rank**	**score**	**predicted_similar_KEGG**	**predicted_similar_EC**	**reaction_rule**	**hyperlink**
2	1	0.4539	R09441		1.14.13.-	LINK
3	2	0.4333	R05655		1.14.13.-	LINK
4	3	0.4136	R08274		1.14.13.-	LINK
5	4	0.4136	R08271		1.14.13.-	LINK
6	5	0.4101	R06916		1.14.13.-	LINK
7	6	0.4076	R04708	1.14.13.57	1.14.13.-	LINK
8	7	0.4062	R06926		1.14.13.-	LINK
9	8	0.403	R04702	1.14.13.56	1.14.13.-	LINK
10	9	0.4023	R05422	1.14.12.12	1.14.13.-	LINK
11	10	0.3862	R11134	1.14.13.217	1.14.13.-	LINK

5. You get a list of similar known reactions (*predicted_similar_KEGG*), ranked by their BridgIT score (*score*), and associated with the EC number of the reaction (*predicted_similar_EC*) and the BNICE.ch reaction rule. The LINK field redirects you to the corresponding database entry in KEGG.

Find pathways

Input

1. Using the compound search, find the LCSB IDs of your precursor and your target compound (see *Find compounds*).
2. In the search interface, select “Pathways” from the drop-down menu.
3. Select the search space. By default, all the reactions in ATLASx are considered (chemATLAS).
4. Paste the IDs of your precursor and target compounds into the fields for “Starting compound” and “Target compound”, respectively.
5. Set the maximum pathway length (number of reaction steps).
6. Set the maximum number of pathways you want to find.
7. *Advanced:* Use defaults, or check ATLASx paper for more details.
8. Launch query



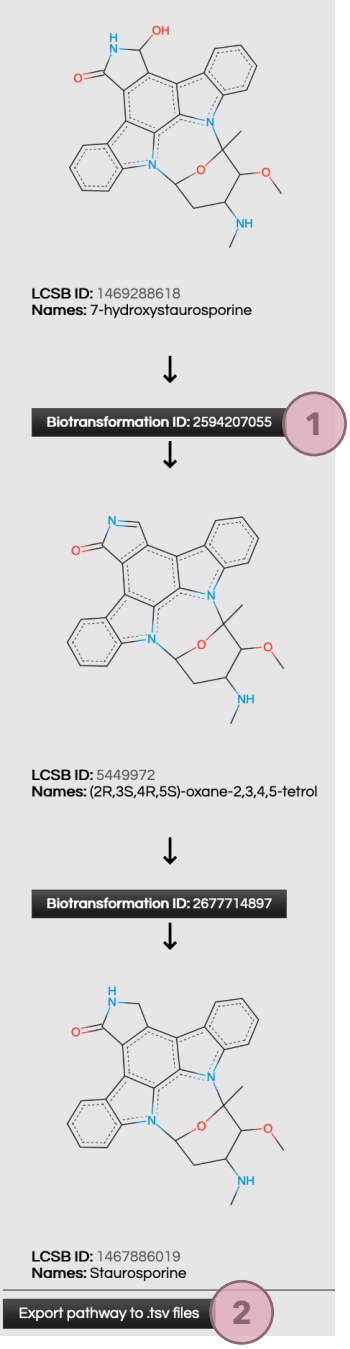
Output

List of pathways found to match the query criteria:

2 results found

Pathway ID	Length	Intermediates (refV)	Intermediates (names)	Reactions	Conserved Atom Ratio (CAR)	Average CAR	Pathway weight score
1	1	1469288618 => 1467886019	7-hydroxystaurosporine => Staurosporine	2677718711	0.95	0.95	1.05
2	2	1469288618 => 5449972 => 1467886019	7-hydroxystaurosporine => (2R,3S,4R,5S)-oxane-2,3,4,5-tetrol => Staurosporine	2594207055, 2677714897	0.95, 1.0	0.98	2.05

The pathway weight score is the total weight of the pathway, i.e. the sum of edge weights assigned to each reaction step.



Click on pathway ID for visualization:

1. Click on “Biotransformation ID" to retrieve all possible reactions for this biotransformation.
2. *Optional:* you can export the pathway information as a folder with:

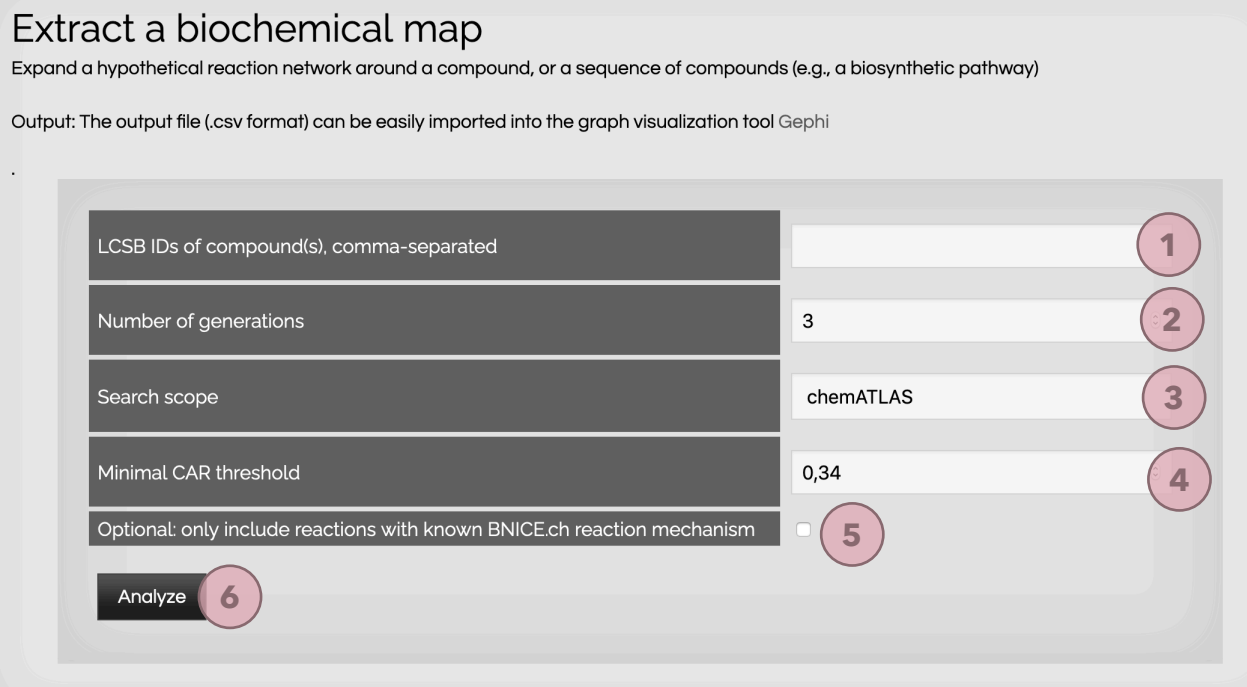
- /molfiles/ : folder containing one SDF molfile for each compound
- compounds.tsv: information on compounds
- reactions.tsv: information on reactions
- pathways.tsv: information on pathway

Analysis platform

Extract a biochemical map

For the network expansion around a compound or pathway, go to the ANALYSIS tab.

Input



1. Enter the LCSBs identifier(s) of input compound(s) from where you want to expand a reaction network.
2. Define how many reaction steps to explore around the input compound.
3. Define the reaction network you want to search
4. All edges with a CAR below the indicated threshold will be removed from the result network.
5. Restrict the search to reactions that have a BNICE.ch reaction mechanism assigned.
6. Run analysis

Output

Once the network extraction is completed, you can download the results as a .csv file:

	A	B	C	D	E
1	**Source**	**Target**	**Generation**	**Biotransformation_ID**	**CAR**
2	1468160340	1109598609	1	2594589111	0.79
3	1468160340	116298274	1	2602987880	0.79
4	1468160340	122001844	1	2603044295	0.55
5	1468160340	125054860	1	2603041273	0.55
6	1468160340	1467915328	1	2602993076	0.76
7	1468160340	1467993870	1	2603004311	0.79
8	1468160340	1468117144	1	2603021274	1
9	1468160340	1468352445	1	2739319792	0.83
10	1468160340	1468357651	1	2602991628	0.79

The network is formatted as an edge table, meaning that each line represents an edge in the network.

- *Source:* LCSB ID of source compound
- *Target:* LCSB ID of target compound
- *Generation:* Number of reaction step away from input compound(s). Here, the input compound is 1468160340, and the compound 1109598609 is found 1 reaction step away from the input compound.
- *Biotransformation_ID:* LCSB ID of the biotransformation (or reactant pair)
- *CAR:* Conserved Atom Ratio of the reactant pair (weight of the edge)

To visually explore the network, you can import the output file into the free network visualization tool Gephi (https://gephi.org/), or any other graph visualization software.

Additional notes

SMILES viewer

ATLASx uses SmilesDrawer, a JavaScript library developed by Probst and Reymond[1], to visualize compounds directly from SMILES format. The SMILES tab allows you to visualize and play with compounds using SMILES strings.

How to cite ATLASx

Please cite the following reference for using ATLASx:

TO BE ADDED

---

[1] **Probst D**, **Reymond J-L**. 2018. SmilesDrawer: Parsing and Drawing SMILES-Encoded Molecular Structures Using Client-Side JavaScript. *J. Chem. Inf. Model.* **58**: 1–7.