

Reaction Data in PubChem

Tiejun Cheng¹, Evan Bolton¹ and Emma Schymanski²



¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

²Environmental Cheminformatics, Luxembourg Centre for Systems Biomedicine, University of Luxembourg

PubChem

An official website of the United States government [Here is how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information





PubChem About Docs Submit Contact

Explore Chemistry

Quickly find chemical information from authoritative sources

Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Use Entrez Compounds Substances BioAssays

 Draw Structure  Upload ID List  Browse Data  Periodic Table

115M Compounds 306M Substances 305M Bioactivities 36M Literature 43M Patents 928 Data Sources

[See More Statistics](#) [Explore Data Sources](#)

<https://pubchem.ncbi.nlm.nih.gov/>

PubChem is an open chemistry database with data on chemicals, bioactivities, proteins/genes, pathways, taxonomies, and cell lines.

Reaction data in PubChem

- 1.5 million from PubChem Pathways (Reactome, BioCyc, WikiPathways, PharmGKB, PathBank, Lipid Maps, etc)
- 14,500 from Rhea
- 1,400 from ChEMBL metabolism
- 5,500 from NORMAN Suspect List Exchange

Reaction data in PubChem

Detailed overview (> 1 K reactions)

Source	# Reactions	Source	# Reactions
PathBank	730 K	WikiPathways	5.3 K
Reactome	390 K	Pathway Interaction DB	3.5 K
PlantCyc	335 K	COVID-19 Disease Map	2.8 K
Plant Reactome	160 K	INOH	2.2 K
BioCyc	75 K	PANTHER	2.0 K
Rhea	14.5 K	ChEMBL	1.4 K*
NORMAN-SLE	5.5 K*	PharmGKB	1.0 K

*Transformations data

S60 [SWISSPEST19](#), S66 [EAWAGTPS](#), S68 [HSDBTPS](#), S73 [MetXBioDB](#),
S74 [REFTPS](#), S78 [SLUPESTTPS](#), S79 [UACCSCEC](#), S81 [THSTPS](#)

Access reaction data in PubChem

From a pathway page

PubChem resveratrol degradation (Pathway)

2 Interactions

1 item Download

Reaction	Control
<i>trans</i> -resveratrol + O ₂ → 4-hydroxybenzaldehyde + 3,4-dihydroxybenzaldehyde	activated by resveratrol cleavage oxygenase

▶ BioCyc

Cite Download

CONTENTS

- Title and Summary
- 1 Identity
- 2 Interactions**
- 3 Chemicals
- 4 Literature
- 5 Information Sources

Reaction data for the pathway resveratrol degradation ([PWY-6667](#))

https://pubchem.ncbi.nlm.nih.gov/pathway/BioCyc:META_PWY-6667#section=Interactions

Access reaction data in PubChem

From a target page – Enzyme

PubChem Receptor protein-tyrosine kinase (EC 2.7.10.1) (Enzyme)

6 Biochemical Reactions

1 item Download

Rhea Accession	Reaction	PubChem Enzyme	Evidence PMID
RHEA:10596	ATP + L-tyrosyl-[protein] = ADP + O-phospho-L-tyrosyl-[protein] + H ⁺	EC 2.7.10.1	15504335
		EC 2.7.12.1	11442633
		EC 2.7.12.2	9053841
		EC 2.7.10.2	12786602
			10954422
			...

[Rhea - annotated reactions database](#)

3,350 items Download

Search SORT BY Taxonomy - A to Z

Reaction	PubChem Pathway	Source	Taxonomy
ATP + PI(4,5)P2 → ADP + PI(3,4,5)P3	Signaling Pathways	Reactome	Homo sapiens (human)
RHOA:GTP + ROCK1,ROCK2 → RHOA:GTP:ROCK1,ROCK2	Signaling Pathways	Reactome	Homo sapiens (human)
ATP + VEGFA;p-6Y-VEGFR2;p-SHB;p-5Y-PTK2:SRC-1:HSP90AA1 → ADP + VEGFA;p-6Y-VEGFR2;pS-SHB;p-5Y,S732-PTK2:SRC-1:HSP90AA1	Signaling Pathways	Reactome	Homo sapiens (human)
ATP + VEGFA;p-6Y-VEGFR2;pS-SHB;p-5Y,S732-PTK2:SRC-1:HSP90AA1 → ADP + VEGFA;p-6Y-VEGFR2;p-SHB;p-6Y,S732-PTK2:SRC-1:HSP90AA1	Signaling Pathways	Reactome	Homo sapiens (human)
ATP + VEGFA;p-6Y-VEGFR2;p-SHB;p-7Y-PTK2:SRC-1:HSP90:PXN → ADP + VEGFA;p-6Y-VEGFR2;p-SHB;p-7Y-PTK2:SRC-1:HSP90;p-Y31,Y118-PXN	Signaling Pathways	Reactome	Homo sapiens (human)

Page 1 of 670

PubChem

<https://pubchem.ncbi.nlm.nih.gov/protein/EC:2.7.10.1#section=Biochemical-Reactions>

Access reaction data in PubChem

From a target page – Protein and Gene

PubChem Epidermal growth factor receptor (human) (Protein)

10 Biochemical Reactions

1 item Download

Rhea Accession	Reaction	PubChem Enzyme	Evidence PMID
RHEA:10596	ATP + L-tyrosyl-[protein] = ADP + O-phospho-L-tyrosyl-[protein] + H ⁺	EC 2.7.12.1 EC 2.7.10.1 EC 2.7.12.2 EC 2.7.10.2 ...	15504335 11442633 9053841 12786602 10954422 ...

[Rhea - annotated reactions database](#)

<https://pubchem.ncbi.nlm.nih.gov/protein/P00533#section=Biochemical-Reactions>

<https://pubchem.ncbi.nlm.nih.gov/gene/1956#section=Biochemical-Reactions>

PubChem EGFR - epidermal growth factor receptor (human) (Gene)

8 Biochemical Reactions

1 item Download

Rhea Accession	Reaction	PubChem Enzyme	Evidence PMID
RHEA:10596	ATP + L-tyrosyl-[protein] = ADP + O-phospho-L-tyrosyl-[protein] + H ⁺	EC 2.7.12.1 EC 2.7.10.1 EC 2.7.12.2 EC 2.7.10.2 ...	15504335 11442633 9053841 12786602 10954422 ...

[Rhea - annotated reactions database](#)

Access reaction data in PubChem

From a compound page – Resveratrol – Biochemical Reactions

9.8 Biochemical Reactions ? ↗

3 items ↓ Download

Rhea Accession	Reaction	PubChem Enzyme	Evidence PMID
RHEA:11936	4-coumaroyl-CoA + malonyl-CoA + H ⁺ = CO ₂ + CoA + <i>trans</i> -resveratrol	EC 2.3.1.95	6427224
RHEA:32103	<i>trans</i> -resveratrol + S-adenosyl-L-methionine = pterostilbene + S-adenosyl-L-homocysteine + H ⁺	EC 2.1.1.240	18799
RHEA:32111	<i>trans</i> -resveratrol + S-adenosyl-L-methionine = S-adenosyl-L-homocysteine + H ⁺ + 3-methoxy-4',5-dihydroxy- <i>trans</i> -stilbene	EC	18799

▶ Rhea - Annotated Reactions Database

<https://pubchem.ncbi.nlm.nih.gov/compound/445154#section=Biochemical-Reactions>

231 items ↓ Download

Search SORT BY Taxonomy - A to Z

Reaction	PubChem Pathway	Source	Taxonomy
<i>trans</i> -resveratrol + O ₂ → 4-hydroxybenzaldehyde + 3,4-dihydroxybenzaldehyde	resveratrol degradation	BioCyc	
H ⁺ + (E)-4-coumaroyl-CoA + malonyl-CoA → coenzyme A + CO ₂ + <i>trans</i> -resveratrol	resveratrol biosynthesis	BioCyc	
SAM + <i>trans</i> -resveratrol → H ⁺ + SAH + 3-methoxy-4',5-dihydroxy- <i>trans</i> -stilbene	pterostilbene biosynthesis	BioCyc	
H ⁺ + 4-coumaroyl-CoA + malonyl-CoA → coenzyme A + CO ₂ + <i>trans</i> -resveratrol	resveratrol biosynthesis	BioCyc	
Mal-CoA + 4-coumaroyl-CoA + coumaroyl-CoA → CoA-SH + carbon dioxide + resveratrol	Metabolism and regulation	Plant Reactome	Oryza sativa (Asian cultivated rice)

« First < Previous Page 1 of 47 Next > Last »

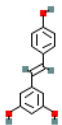
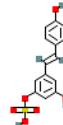
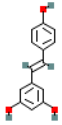
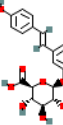
Access reaction data in PubChem

From a compound page – Resveratrol – Transformations

PubChem Resveratrol (Compound)

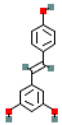
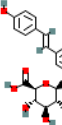
9.9 Transformations

2 items Download

Predecessor	Predecessor Name	Successor	Successor Name	Evidence PMID
	RESVERATROL		Trans-resveratrol-3-sulfate	23474649
	RESVERATROL		trans-resveratrol-3-O-glucuronide	23474649

[ChEMBL](#)

1 item Download

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidence DOI
	Resveratrol		Resveratrol 3-O-glucuronide	Aromatic-OH glucuronidation / Human Phase II	UGT1A9	10.1186/s13321-018-0324-5

[NORMAN Suspect List Exchange](#)

Cite Download

CONTENTS

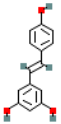
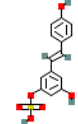
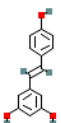
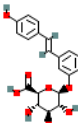
- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Food Additives and Ingredients
- 9 Pharmacology and Biochemistry
 - 9.1 Pharmacodynamics
 - 9.2 MeSH Pharmacological Classification
 - 9.3 Absorption, Distribution and Excretion
 - 9.4 Metabolism/Metabolites
 - 9.5 Biological Half-Life
 - 9.6 Mechanism of Action
 - 9.7 Human Metabolite Information
 - 9.7.1 Tissue Locations
 - 9.7.2 Cellular Locations
 - 9.8 Biochemical Reactions
 - 9.9 Transformations

Download & Programmatic Access Options

PubChem Resveratrol (Compound)

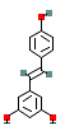
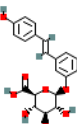
9.9 Transformations

2 items Download

Predecessor	Predecessor Name	Successor	Successor Name
	RESVERATROL		Trans-resveratrol
	RESVERATROL		trans-resveratrol

ChEMBL

1 item Download

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidence DOI
	Resveratrol		Resveratrol 3-O-glucuronide	Aromatic-OH glucuronidation / Human Phase II	UGT1A9	10.1186/s13321-018-0324-5

NORMAN Suspect List Exchange

<https://pubchem.ncbi.nlm.nih.gov/compound/445154#section=Transformations>

PubChem About Docs Submit Contact

DATA SOURCES > RHEA - ANNOTATED REACTIONS ... > ANNOTATIONS

Annotations from Rhea - Annotated Reactions Database

1 annotation topic Download

9,639 total annotation data items

Biochemical Reactions (Compound)

<https://pubchem.ncbi.nlm.nih.gov/source/23640#data=Annotations>

<https://pubchem.ncbi.nlm.nih.gov/docs/programmatic-access>

Download Annotations ×

JSON	Save	Display
XML	Save	Display
ASNT	Save	Display

*Hint: results from this PUG View service request may be paginated.
For more information, please refer to the [PUG View help document](#).*

Adding Missing Data to PubChem



NORMAN Suspect List Exchange

The NORMAN network enhances the exchange of information on emerging environmental measurement methods and monitoring tools so that the requirements of risk assessment (NORMAN-SLE) is a central access point to find suspect lists relevant for various environments.

Organization	NORMAN Network (c/o UniLu)
Category	Research and Development
URL	https://www.norman-network.com/nds/SLE/
License Note	Data: CC-BY 4.0; Code (hosted by ECI, LCSB): Artistic-2.0
License URL	https://creativecommons.org/licenses/by/4.0/
Contact Name	Emma Schymanski
Address	6 avenue du Swing, Belvaux, Luxembourg, 4367
Data Source ID	23819
Data in PubChem	118,487 Live Substances 22,302 Annotations 1 Classification
Last Updated	2023/07/09

Browse NORMAN Suspect List Exchange Tree

- ▼ NORMAN Suspect List Exchange Classification ? ↗ 115,564
 - ▶ S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 3,933
 - ▶ S25 | OECDPFAS | List of PFAS from the OECD ? 3,678
 - ▶ S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances ? 254
 - ▶ S47 | ECHAPLASTICS | A list from the Plastic Additives Initiative Mapping Exercise by ECHA ? 241
 - ▶ S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 869
 - ▼ S60 | SWISSEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,358
 - Pesticide transformation products (TPs, metabolites, successors) ? 1,102
 - Pesticides (parents, predecessors) ? 267
 - ▶ S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
 - ▶ S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258
 - ▼ S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 740
 - Parents (Predecessors) ? 223
 - Transformation products (TPs, successors) ? 557
 - ▶ S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386
 - ▶ S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University ? 1,068
 - ▶ S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria ? 2,553

<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101>

Adding Missing Data to PubChem



7.2 Agrochemical Transformations

Terbutylazine has known environmental transformation products that include Terbutylazine-desethyl, Terbutylazine-2-hydroxy, and Terbutylazine-desethyl-2-hydroxy.

S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

► NORMAN Suspect List Exchange

<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101>

wise NORMAN Suspect List Exchange Tree

- ▼ NORMAN Suspect List Exchange Classification ? 115,564
 - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 3,933
 - S25 | OECDPFAS | List of PFAS from the OECD ? 3,678
 - S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances ? 254
 - S47 | ECHAPLASTICS | A list from the Plastic Additives Initiative Mapping Exercise by ECHA ? 241
 - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 869
 - ▼ S60 | SWISSEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,358
 - Pesticide transformation products (TPs, metabolites, successors) ? 1,102
 - Pesticides (parents, predecessors) ? 267
 - S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
 - S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258
 - ▼ S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 740
 - Parents (Predecessors) ? 223
 - Transformation products (TPs, successors) ? 557
 - S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386
 - S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University ? 1,068
 - S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria ? 2,553

PubChem Terbutylazine (Compound)

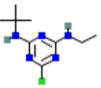
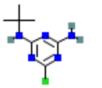
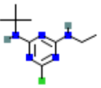
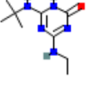
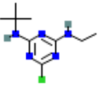
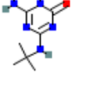
8.5 Transformations

30 items

Download

Search

SORT BY Predecessor Name - A to Z

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidence DO
	terbutylazine		desethyl-terbutylazine	Environmental		10.1021/acs.e
	terbutylazine		hydroxy-terbutylazine	Environmental		10.1021/acs.e
	terbutylazine		2-hydroxy-desethyl-terbutylazine	Environmental		10.1021/acs.e

FAIR Transformations in PubChem



- EAWAGTPS
- HSDB
- MetXBioDB**
- REFTPS
- SLUPESTTPS
- SWISSPEST19
- THSTPS
- UACCSCEC



6,641 chemicals

>double the
BioTransformer 3.0
library

zenodo Search Upload Communities

July 5, 2023 Dataset Open Access

Transformations in PubChem - Full Dataset

[ID](#) Schymanski, Emma; [ID](#) Bolton, Evan; [ID](#) Cheng, Tiejun; [ID](#) Thiessen, Paul; [ID](#) Zhang, Jian (Jeff); [ID](#) Helmus, Rick; [ID](#) Blanke, Gerd

This is an archive of the data contained in the "Transformations" section in PubChem for integration into patRoom and other workflows.

For further details see the ECI GitLab site: [README](#) and main "tps" folder.

Credits:

Concepts: E Schymanski, E Bolton, J Zhang, T Cheng;



Code (in R): E Schymanski, R Helmus, P Thiessen

Transformations: E Schymanski, J Zhang, T Cheng and many contributors to various lists!

PubChem infrastructure: PubChem team

Reaction InChI (RInChI) calculations (v1.0): Gerd Blanke (previous versions of these files)

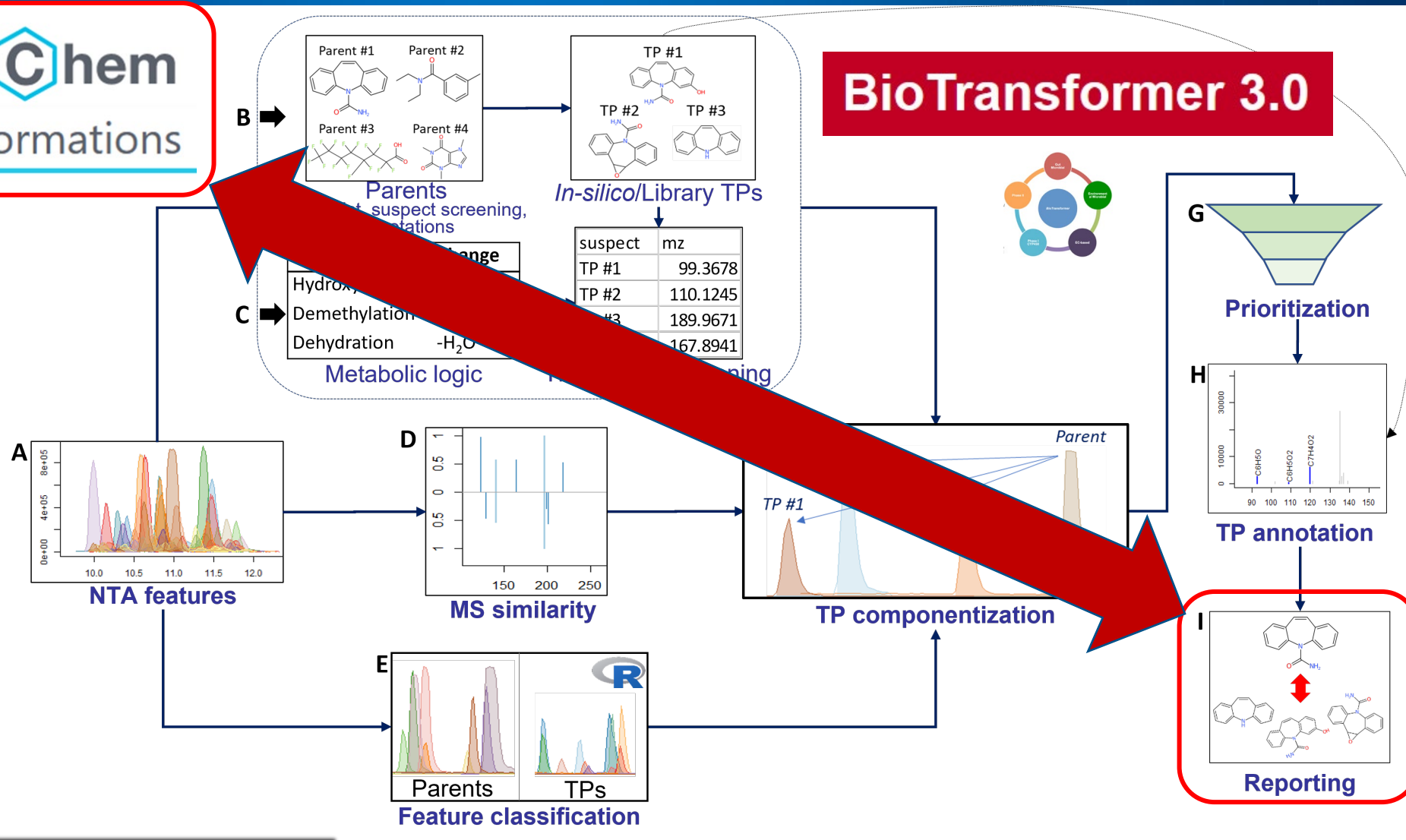
Acknowledgements: ECI team who contributed to related efforts, especially: J. Krier, A. Lai, M. Narayanan, T. Kondic, P. Chirsir, E. Palm. All contributors to the NORMAN-SLE transformations!



Open TP Workflows in patRoou 2.0



PubChem
Transformations



BioTransformer 3.0



Transforming PubChemLite with

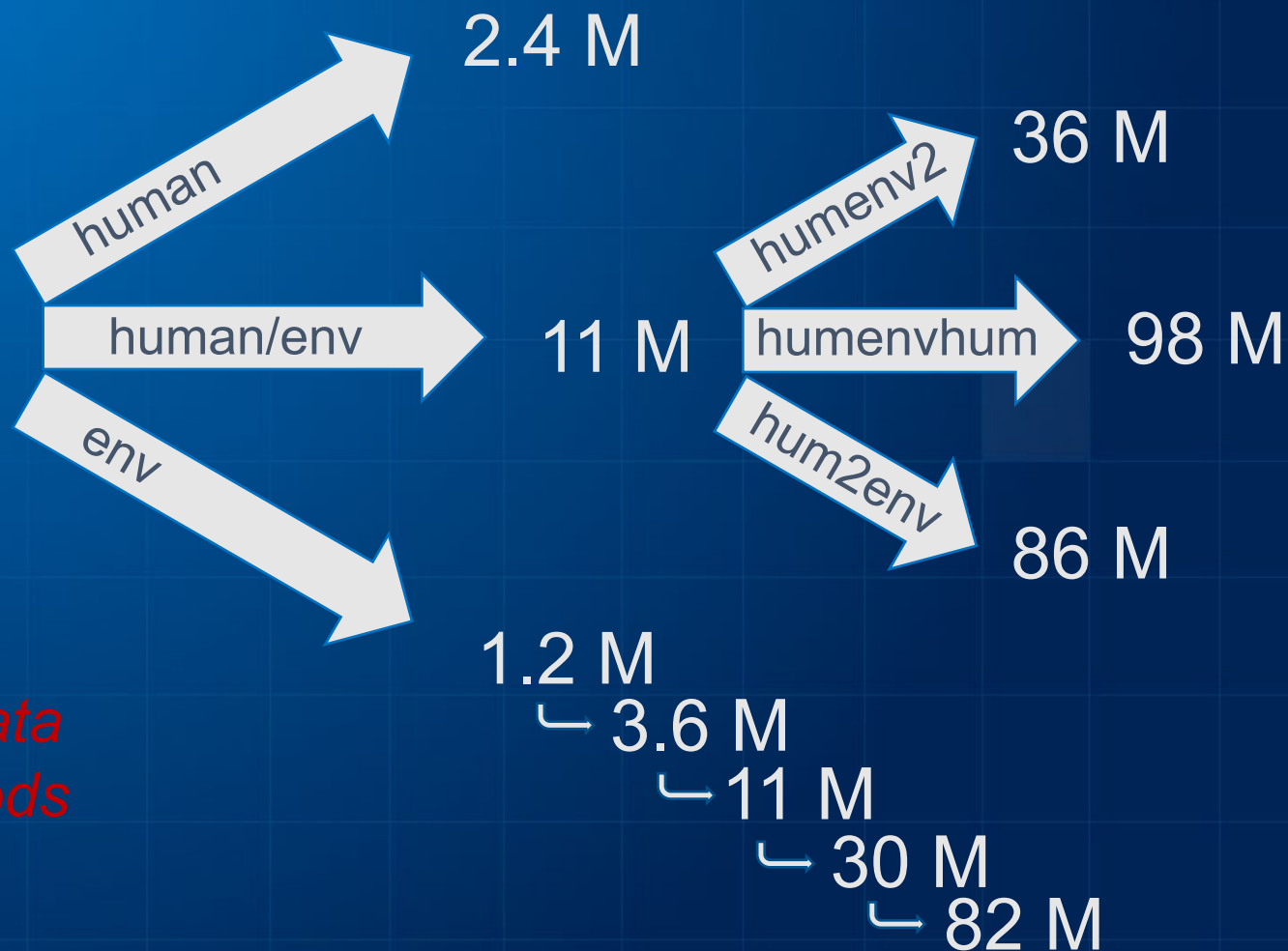
BioTransformer 3.0



379,199 entries

version 1.0.0; DOI: [10.5281/zenodo.5995886](https://doi.org/10.5281/zenodo.5995886)

Combinatorial explosion – more data needed to predict reaction likelihoods



Perspectives

- >10 Million reactions from Google Patents
- Dedicated reaction summary pages
- More data from literature mining

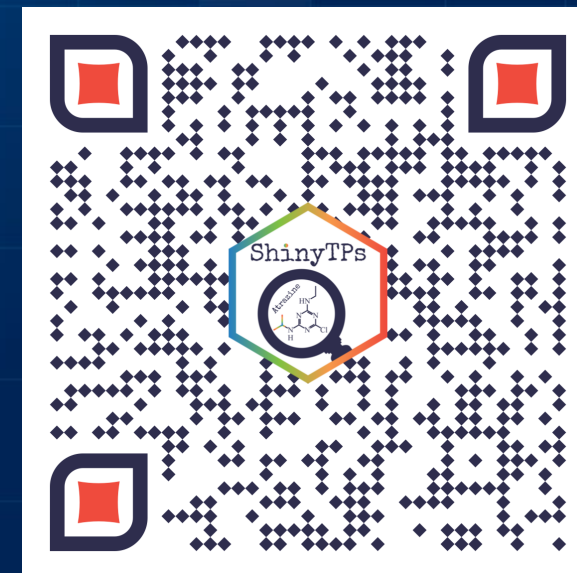
PubChem Acyclovir (Compound)

8.6 Metabolism/Metabolites

Acyclovir is metabolized partially to 9-carboxymethoxymethylguanine and minimally to 8-hydroxy-9-(2-hydroxyethoxymethyl)guanine. In vitro, acyclovir also is metabolized to acyclovir monophosphate, diphosphate, and triphosphate in cells infected with herpes viruses, principally by intracellular phosphorylation of the drug by virus coded thymidine kinase and several cellular enzymes.

American Society of Health-System Pharmacists 2013; Drug Information 2013. Bethesda, MD. 2013, p. 769

► Hazardous Substances Data Bank (HSDB)



Further reading (selection)

- Kim *et al.* (2023) PubChem 2023 update. *Nucleic Acids Res.* 51(D1):D1373-D1380. DOI: [10.1093/nar/gkac956](https://doi.org/10.1093/nar/gkac956)
- Kim *et al.* (2022) PubChem Protein, Gene, Pathway, and Taxonomy Data Collections. *J Mol Biol.* 434(11):167514. DOI: [10.1016/j.jmb.2022.167514](https://doi.org/10.1016/j.jmb.2022.167514)
- Zaslavsky *et al.* (2021) Discovering and Summarizing Relationships Between Chemicals, Genes, Proteins, and Diseases in PubChem. *Front. Res. Metr. Anal.* 6:689059. DOI: [10.3389/frma.2021.689059](https://doi.org/10.3389/frma.2021.689059)
- Schymanski *et al.* (2021) Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. *J. Cheminform.* 13:19. DOI: [10.1186/s13321-021-00489-0](https://doi.org/10.1186/s13321-021-00489-0)
- Krier *et al.* (2022) Discovering pesticides and their TPs. *Environ. Int.* 158:106885. DOI: [10.1016/j.envint.2021.106885](https://doi.org/10.1016/j.envint.2021.106885)
- Mohammed Taha *et al.* (2022) The NORMAN Suspect List Exchange (NORMAN-SLE). *Environ. Sci. Eur.* 34, 104. DOI: [10.1186/s12302-022-00680-6](https://doi.org/10.1186/s12302-022-00680-6)
- Talavera Andújar *et al.* (2023) Exposome of MCI and AD on CSF. *ChemRxiv*. DOI: [10.26434/chemrxiv-2023-6j2gm](https://doi.org/10.26434/chemrxiv-2023-6j2gm)

Acknowledgements

- PubChem team
- LCSB-ECI team
- PubChem contributors
- NORMAN-SLE contributors

- You all for listening!



Today's slides: DOI:
[10.5281/zenodo.8143832](https://doi.org/10.5281/zenodo.8143832)

Email:
emma.schymanski@uni.lu
bolton@ncbi.nlm.nih.gov

Twitter:
[@ESchymanski](https://twitter.com/ESchymanski)
[@EvanBolton](https://twitter.com/EvanBolton) [@pubchem](https://twitter.com/pubchem)



ZerO²PM



FNR
(A18/BM/12341006)
EU Horizon 2020
(Grant No. 101036756)
for [ZeroPM](#)