

Increasing the understanding of Metabolomics data with ontologies

Denise Slenter, Chris Evelo, Egon Willighagen

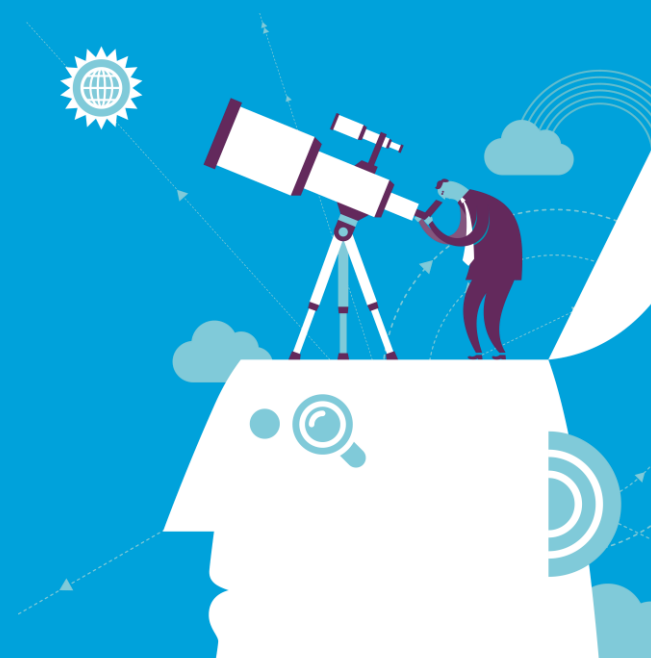
Twitter: @SMaLLCaT4Sci and @BiGCaT_UM

Blog: <http://smallcats4science.blogspot.nl>

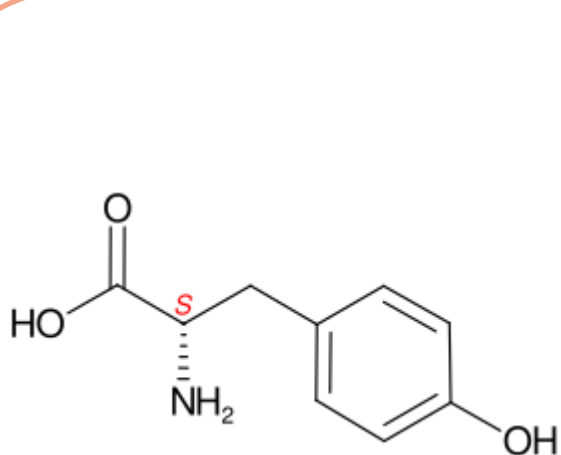
ORCID: 0000-0001-8449-1318

2018-11-12 Mainz #gcc2018mainz

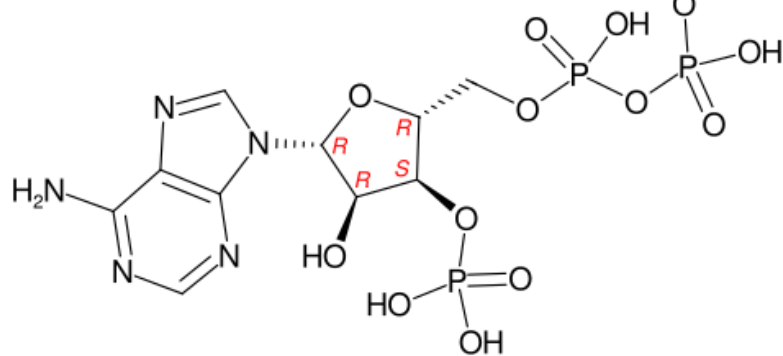
German Conference on Cheminformatics



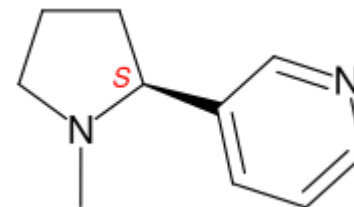
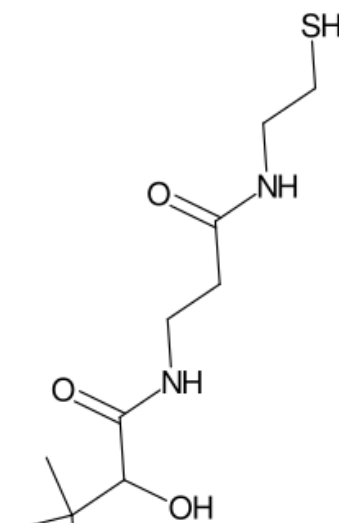
Linking metabolomics data to pathways...



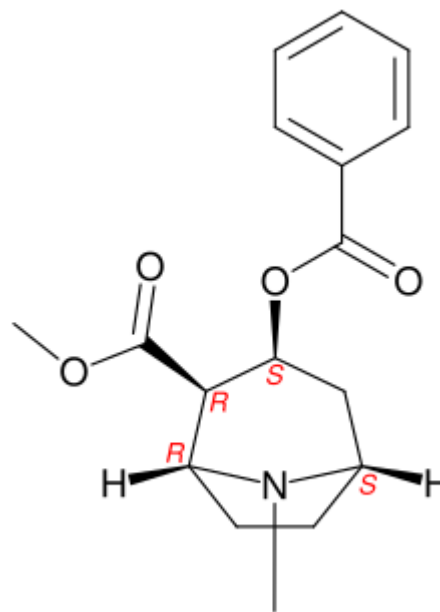
L-tyrosine



Co-A

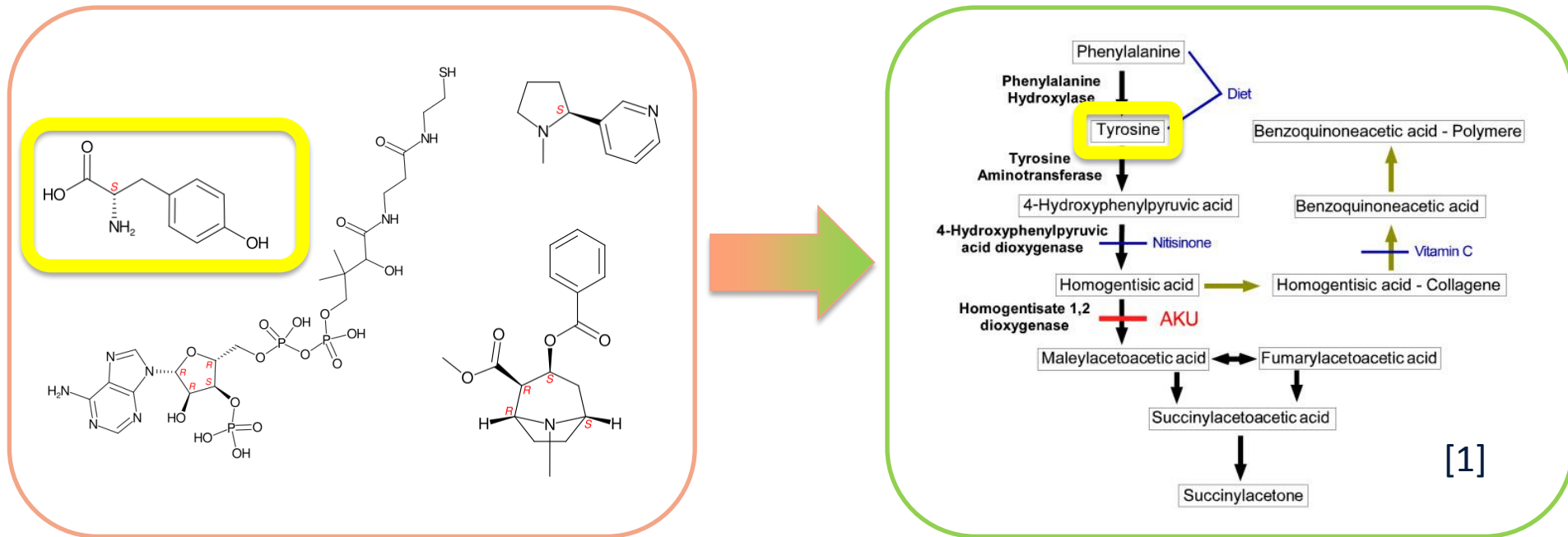


Nicotine



Cocaine

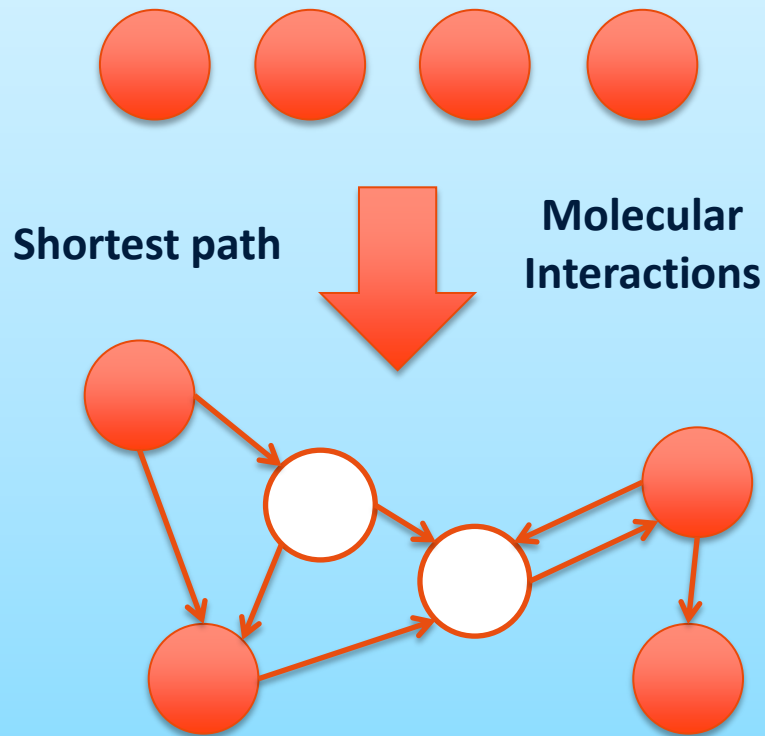
Linking metabolomics data to pathways...



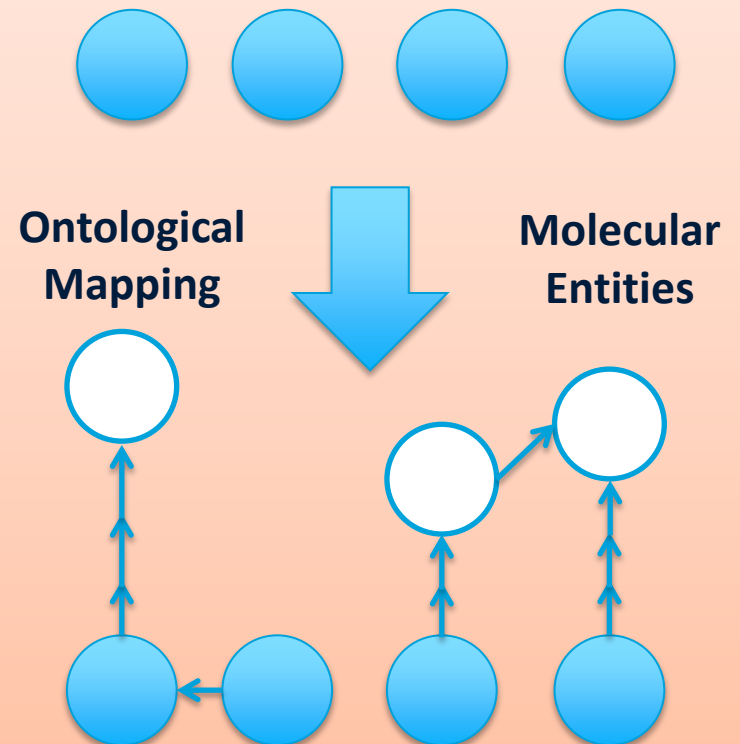
Sparseness of Data

Two approaches:

Network approach [1]

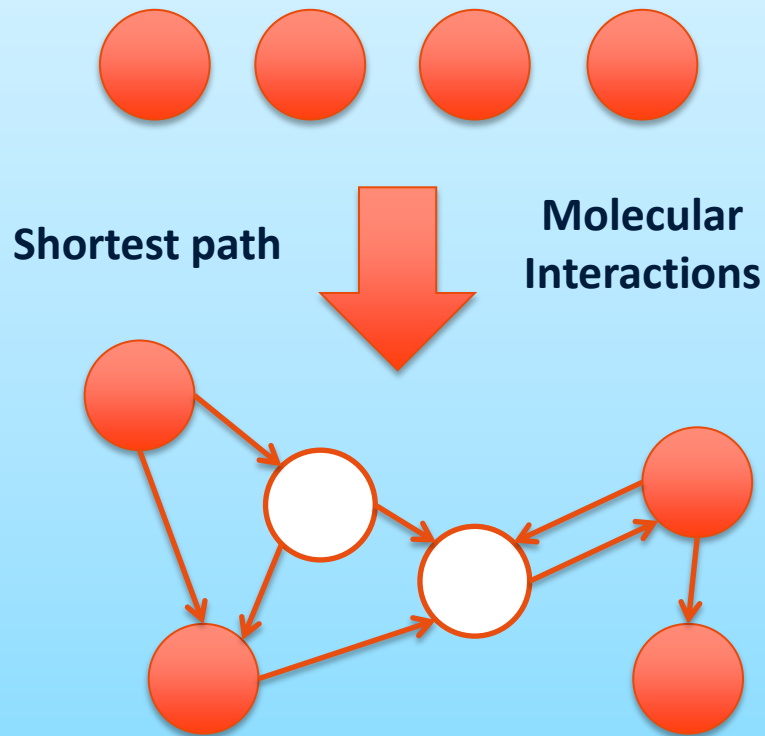


Ontological approach [2]

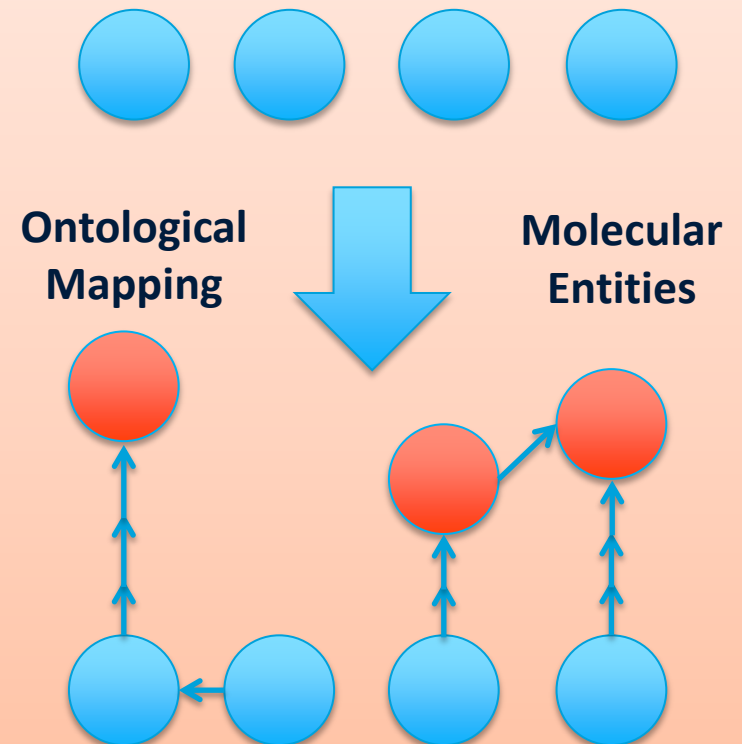


Two approaches:

Network approach [1]



Ontological approach [2]



Metabolomics Datasets [6-9]

BridgeDb

[11]



MetaboLights

[3]

Identifier

Name

HMDB00538

ATP

CHEBI:16761

CHEBI:16828

L-try

CHEBI:17895

L-ty

I. Mapping to databases [4]

Identifier

Name:

Xref

Database

HMDB00538

ATP

DB00171

Drugbank

CHEBI:16761

ADP

C00019353

KNAPSAcK

CHEBI:16828

L-tryptophan

6305

Pubchem CID

CHEBI:17895

L-tyrosine

Q188017

Wikidata

[3] Haug K; et al. Nucl. Acids Res. 2013

[6] MTBLS124, DOI: 10.1371/journal.pgen.1004801

[7] MTBLS265, DOI: 10.1073/pnas.1603023113

[8] MTBLS351, DOI: 10.1038/nature18646

[9] MTBLS404, DOI: 10.1021/acs.jproteome.5b00354

[11] <https://www.bridgedb.org/>



Maastricht University

Identifier	Name:	Xref	Database
HMDB00538	ATP	DB00171	Drugbank
CHEBI:16761	ADP	C00019353	KNAPSAcK
CHEBI:16828	L-tryptophan	6305	Pubchem CID
CHEBI:17895	L-tyrosine	Q188017	Wikidata



II. IDs in pathway(s) [5]



Pathway Analysis [6,10]

[2] Slenter D.; et al. Nucl. Acids Res. 2018

[5] WikiPathways pathways accession number:
<http://data.wikipathways.org/20180410/>

[6] MTBLS124, DOI: 10.1371/journal.pgen.1004801

[7] MTBLS265, DOI: 10.1073/pnas.1603023113

[8] MTBLS351, DOI: 10.1038/nature18646

[9] MTBLS404, DOI: 10.1021/acs.jproteome.5b00354

[10] WP4220_r97366, neurotransmitter disorders; Friesacher A., Slenter D., Willighagen E.

Title: Neurotransmitter Disorders
Organism: Homo sapiens

Pathway Analysis [6,10]

Identifier

Name

HMDB00538

CHEBI:16761

CHEBI:16828

L-tryptophan

CHEBI:17895

L-tyrosine

Tyrosine

L-tryptophan

Tyrosine 3-mono
oxygenase

Tyrosine Hydroxylase
Deficiency

Vanillic acid

3-O-methyldopa

L-Dopa

5-Hydroxytryptophan

Q188017

Wikidata

[6] MTBLS124, DOI: 10.1371/journal.pgen.1004801

[7] MTBLS265, DOI: 10.1073/pnas.1603023113

[8] MTBLS351, DOI: 10.1038/nature18646

[9] MTBLS404, DOI: 10.1021/acs.jproteome.5b00354

[10] WP4220_r97366, neurotransmitter disorders; Friesacher A., Slenter D., Willighagen E.



Identifier	Name:	Xref	Database
HMDB00538	ATP	DB00171	Drugbank
CHEBI:16761	ADP	C00019353	KNAPSAcK
CHEBI:16828	L-tryptophan	6305	Pubchem CID
CHEBI:17895	L-tyrosine	Q188017	Wikidata

III. IDs not in pathway [5]

Ontology
Mapping [6,10]



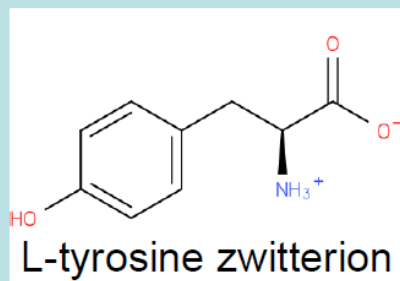
- [1] Hastings J. et al. Nucl. Acids Res. 2016
- [5] WikiPathways pathways accession number:
<http://data.wikipathways.org/20180410/>
- [6] MTBLS124, DOI: 10.1371/journal.pgen.1004801
- [7] MTBLS265, DOI: 10.1073/pnas.1603023113
- [8] MTBLS351, DOI: 10.1038/nature18646
- [9] MTBLS404, DOI: 10.1021/acs.jproteome.5b00354
- [10] WP4220_r97366, neurotransmitter disorders; Friesacher A., Slenter D., Willighagen E.

Ontological approach: AIM

Connect annotated metabolites != in pathway,
to related metabolites = in a pathway

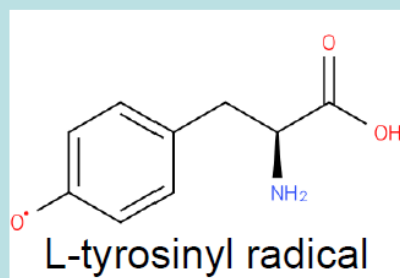
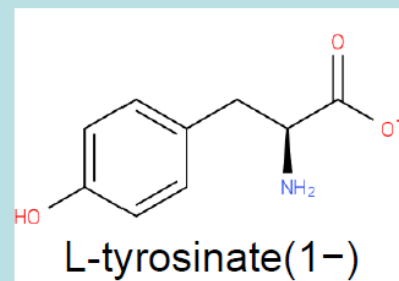
6 relationship types defined in ChEBI[1]:

- Child compounds → “Subclass”
- Parent compounds/groups → “SuperClass”
- Charged forms (acid/base) → “Charge state”
- Proton(s) swopped internally → “Tautomers”
- Diff. stereochemistry → “Enantiomers”
- Biological roles → “Roles”



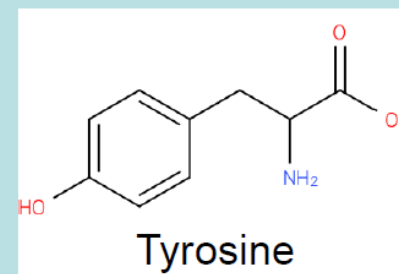
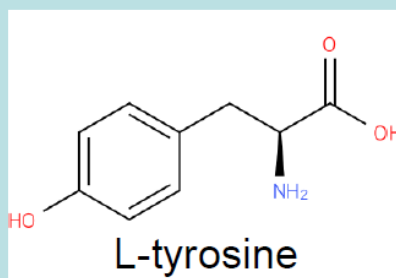
Tautomer

Charge state



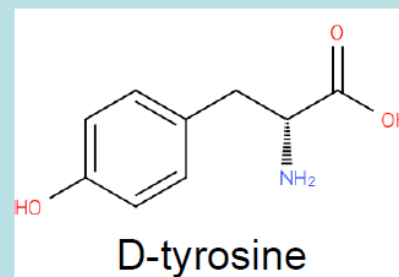
SubClass

SuperClass



Roles

Enantiomer

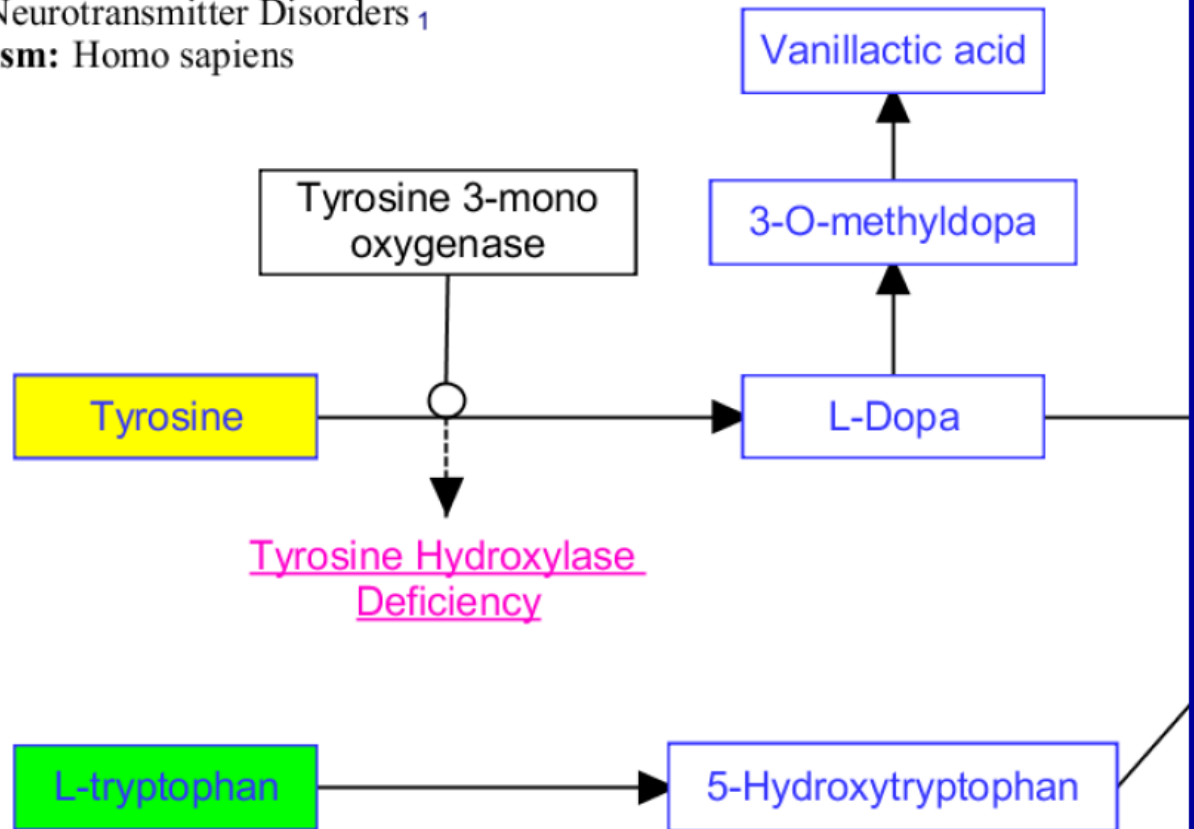


- EC 1.3.1.43 (arogenate dehydrogenase) inhibitor
- Micronutrient

Ontology Mapping [6,10]

Title: Neurotransmitter Disorders₁

Organism: Homo sapiens



[6] MTBLS124, DOI: 10.1371/journal.pgen.1004801

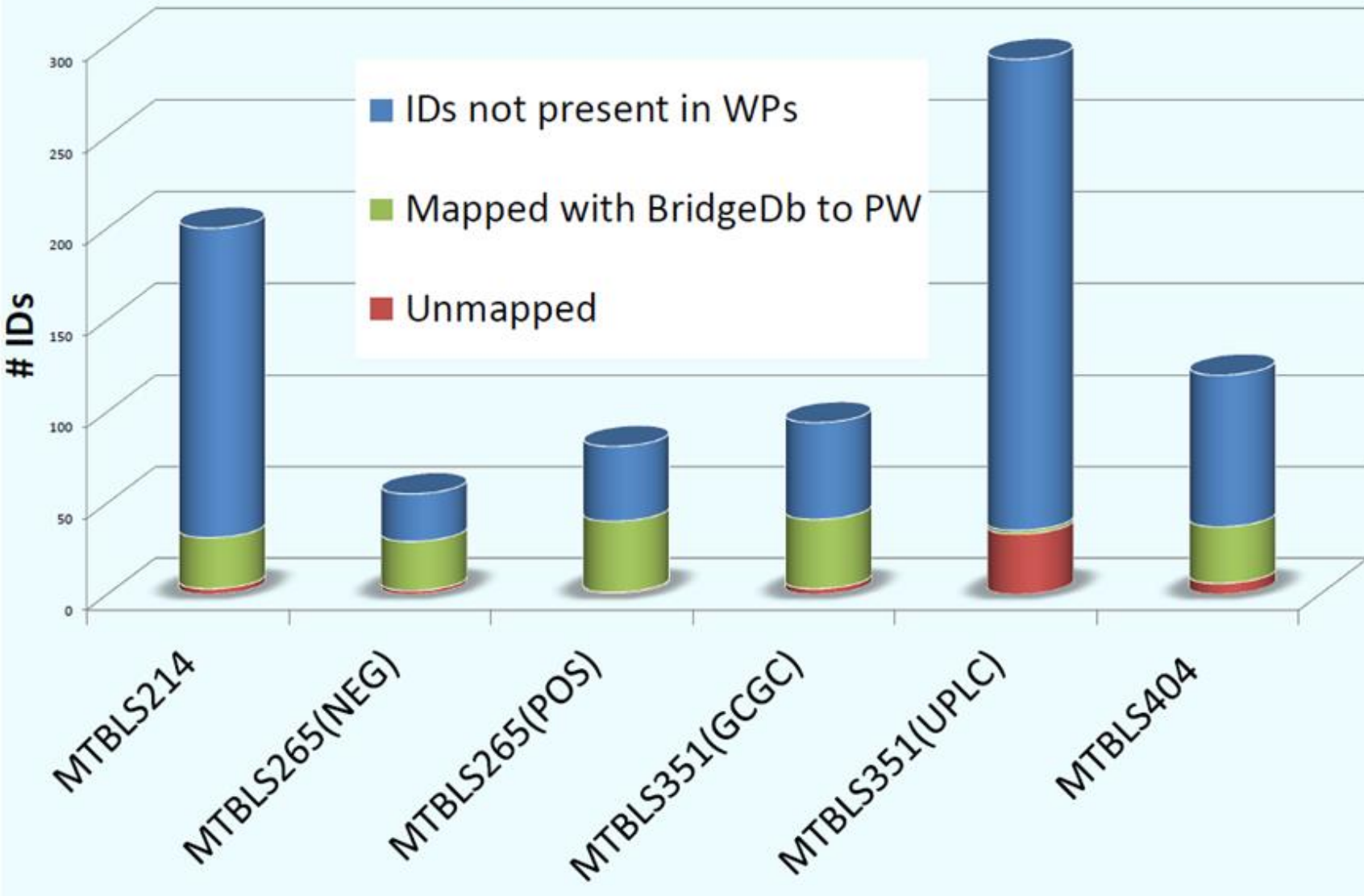
[7] MTBLS265, DOI: 10.1073/pnas.1603023113

[8] MTBLS351, DOI: 10.1038/nature18646

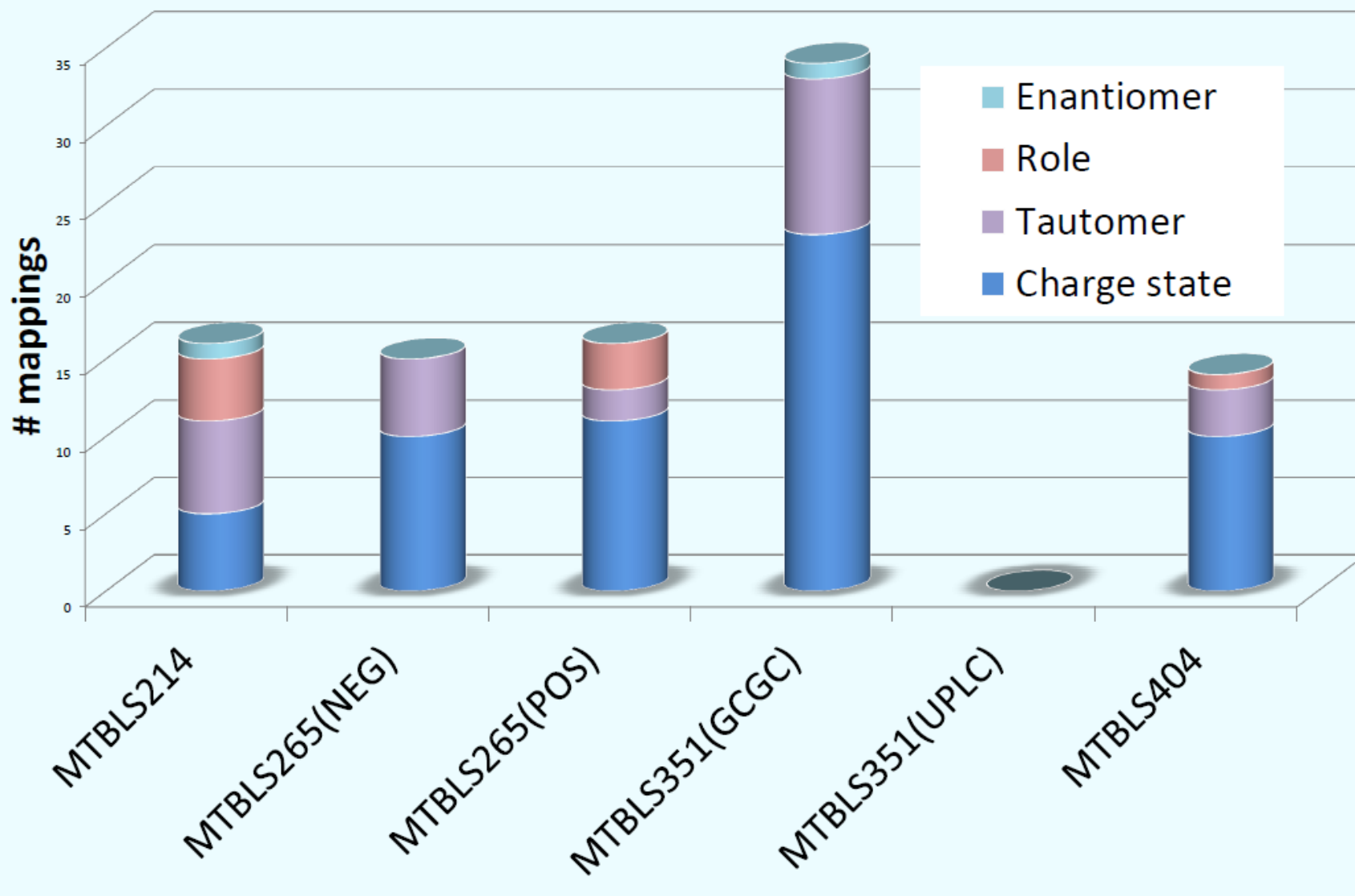
[9] MTBLS404, DOI: 10.1021/acs.jproteome.5b00354

[10] WP4220_r97366, neurotransmitter disorders; Friesacher A., Slenter D., Willighagen E.

Mappings through BridgeDb and WikiPathways



Additional Mappings through Ontologies



Results compound classes: Biocrates [1]



Group:	# IDs	# Directly Mapped	# Ontologically mapped
Hexoses	1	1 (100%)	-
Amino acids & Biogenic amines	50	50 (100%)	-
Steroids	17	15 (88%)	16 (94%)
Lysophosphatidyl-cholines	19	2 (10%)	15 (79%)
Acylcarnitines	50	4 (8%)	48 (96%)
Sphingomyelins	15	0 (0%)	15 (100%)

Future Directions:



- Import ChEBI database into Neo4j
- Add mappings to pathway information
- Query metabolites in dataset (with set length for child-parent relationships)
- Visualise mappings automatically in Cytoscape
- For unmapped metabolites, find important substructures (with CDK), add ontological relationship to Wikidata
- Expand ontological database with LipidMaps and Wikidata info

