Inference of Functions, Roles, and Applications of Chemicals Using Linked Open Data and Ontologies

Tatsuya Kushida[1], Kouji Kozaki[2], Takahiro Kawamura[3], Yuka Tateisi[1], Yasunori Yamamoto[4], and Toshihisa Takagi[1,5]

1 National Bioscience Database Center, Japan Science and Technology Agency, Japan
2 The Institute of Scientific and Industrial Research, Osaka Univ., Japan
3 Japan Science and Technology Agency, Japan
4 Database Center for Life Science, Research Organization of Information and Systems, Japan
5 Dept. Biological Sciences, Grad. School of Science, The Univ. of Tokyo, Japan

JIST2018, 26 - 28 November 2018
Awaji, Japan

Background, and Objectives

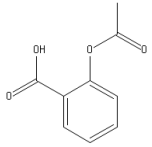
- **Background:** Information regarding chemicals' functions, roles, applications, and involvement in diseases is important to further promote drugs discovery, development of new medical treatments, appropriate leverage of chemicals, and new applications of materials. However, it is difficult for those who don't have highly specialized knowledge and skills of programming to efficiently acquire the information so far.

- **Aim:** A development of a method to easily and efficiently find and infer the chemical information using LOD, and ontologies for biologists on the internet.

Data sources

LOD/Ontology	Developer	Download site	License
A chemicals LOD: **NikkajiRDF**	National Bioscience Database Center (NBDC)	http://doi.org/10.18908/lsdba.nbdc01530-02-000.V008	CC BY-SA
Interlinking Ontology for Biological Concepts (**IOBC**)		https://bioportal.bioontology.org/ontologies/IOBC?p=classes	CC BY-NC
Chemical Entities of Biological Interest (**ChEBI**)	European Bioinformatics Institute (EBI)	https://www.ebi.ac.uk/chebi/downloadsForward.do	CC BY

A chemicals LOD: NikkajiRDF

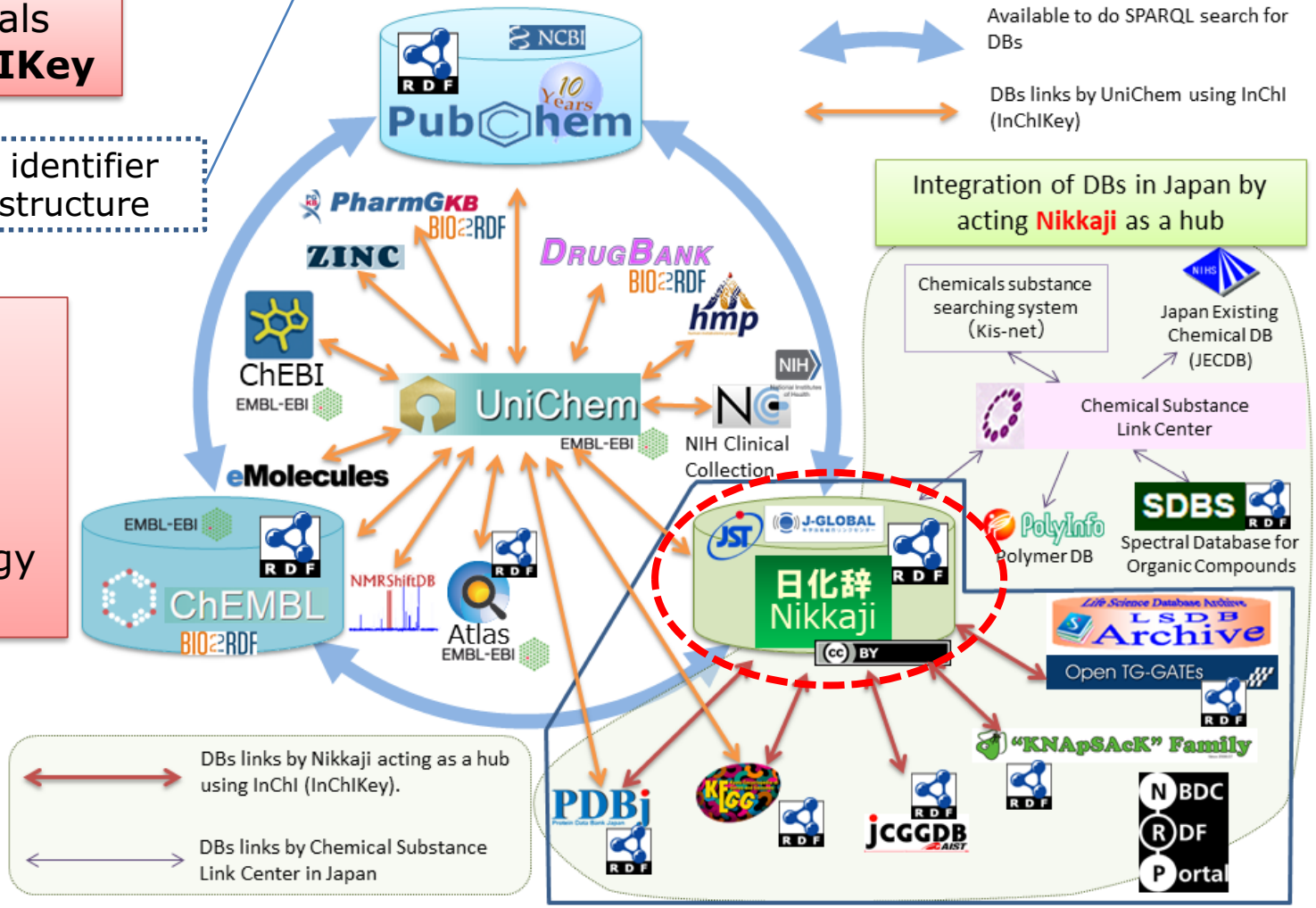


Name: Aspirin
Molecular formula: C9H8O4
InChI: InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)
InChI key: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

3.5 million chemicals with **InChI/InChIKey**

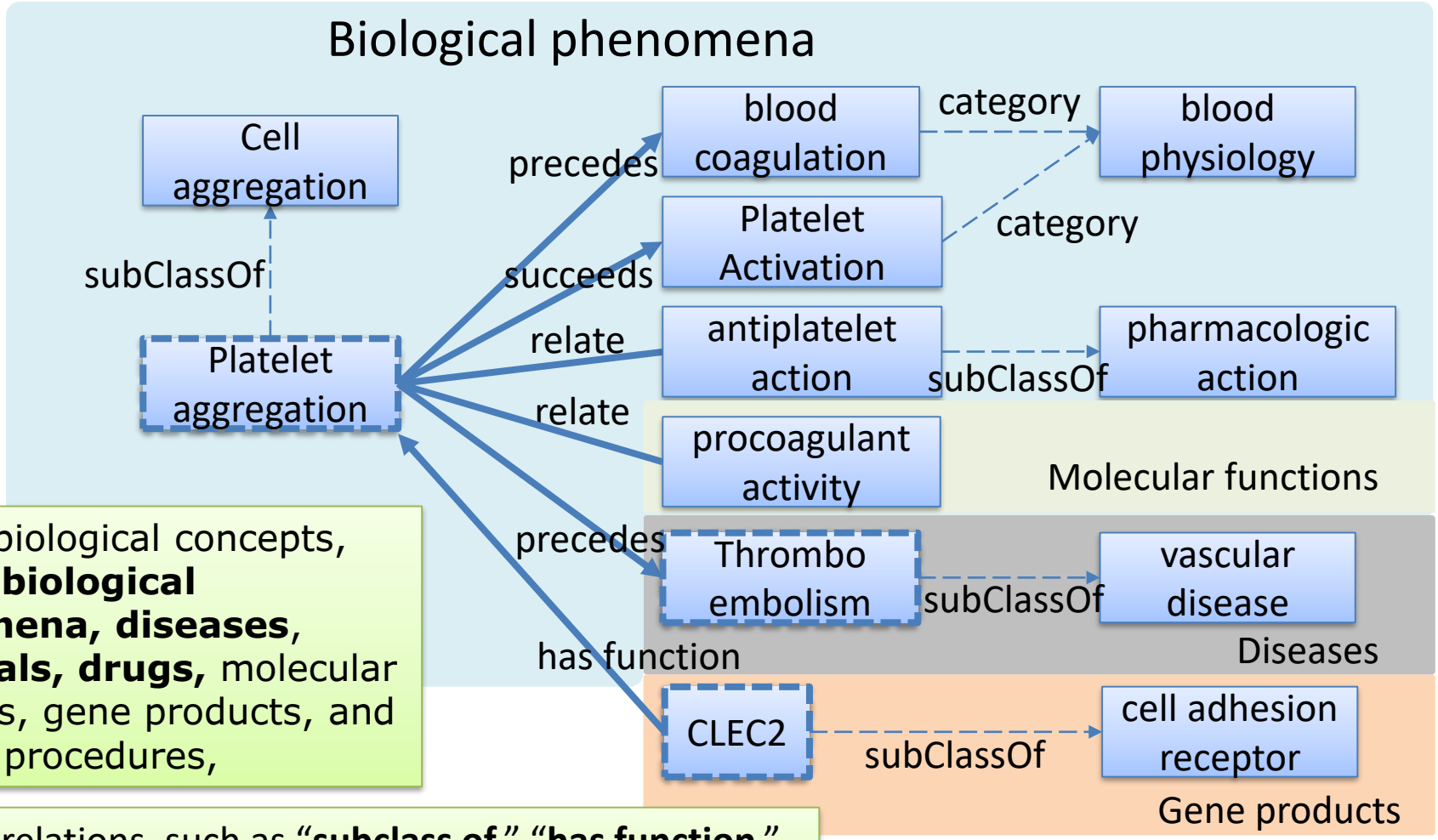
Unique chemical identifier based on steric structure

Using Chemical Information Ontology (**CHEMINF**) and Semanticscience Integrated Ontology (**SIO**)



Integration of NikkajiRDF with major databases of chemicals using InChI/InChIKey.

Interlinking Ontology for Biological Concepts (IOBC), [formerly known as "Refined JST thesaurus"]



80,000 biological concepts, such as **biological phenomena, diseases, chemicals, drugs,** molecular functions, gene products, and medical procedures,

Using 35 relations, such as "**subclass of**," "**has function**," "**has role**," "**precedes**," and "**is participant in**"

Chemical Entities of Biological Interest (ChEBI)

90,000* chemicals, with **InChI/InChIKey**

1,000 kinds of **roles** and **applications** (e.g. protein kinase inhibitor, environmental contaminant, and psychotropic drug).

*: The number is few than that of PubChem, ChEMBL, and NikkajiRDF.

Related works

The number of chemicals stored in five LODs, and ontologies, and the properties used for describing chemical functions/roles/applications.

:Positive point

Ontology

	DBpedia	Wikidata	NikkajiRDF	IOBC	ChEBI
No. of chemicals	18,000	150,000	3.5M	17,180 (in 80,000 biological concepts)	90,000
InChI/ InChIKey	InChI/ InChIKey	InChI/ InChIKey	InChI/ InChIKey	No	InChI/ InChIKey
Properties used for describing chemical functions, roles, applications	“dcterms:subject” “rdfs:seeAlso”	“wdt :P31 (instance of)”	“jst:use-for”	“sio:SIO_000225 (has function)” “sio:SIO_000228 (has role)” “sio:SIO_000062 (is participant in)”	“obo:RO_0000087 (has role)”

annotation properties

specific properties

dcterms: <http://purl.org/dc/terms/>
rdfs: <http://www.w3.org/2000/01/rdf-schema#>
wdt: <http://www.wikidata.org/prop/direct/>

jst: <http://vocab.jst.go.jp/terms/sti#>
sio: <http://semanticscience.org/resource/>
obo: <http://purl.obolibrary.org/obo/>

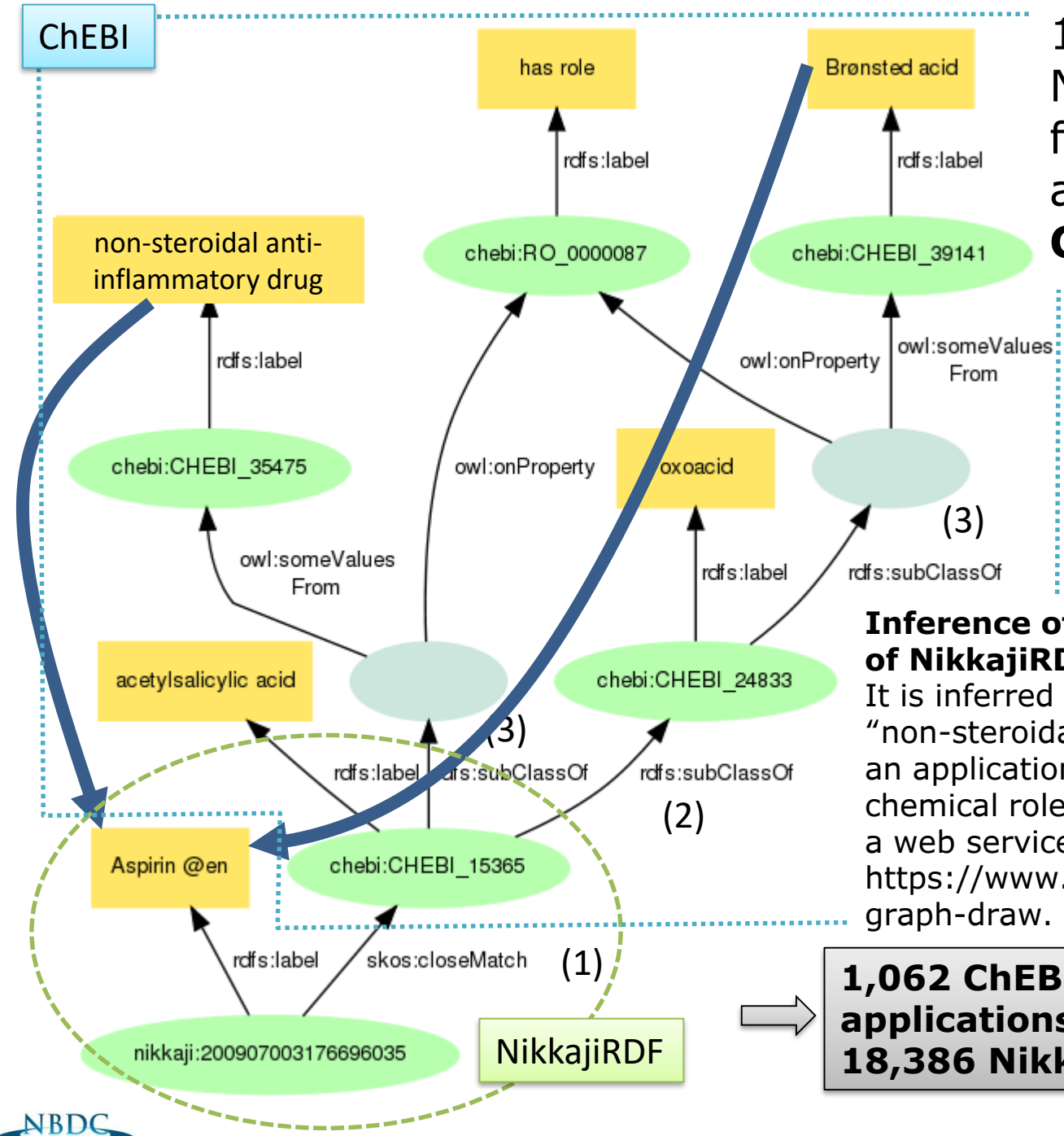
Inference of chemicals' functions, roles, applications, and involvement in diseases using NikkajiRDF, IOBC, and ChEBI

1. Inference of **NikkajiRDF** chemicals' functions, roles, and applications using **ChEBI**

2. Inference of **NikkajiRDF** chemicals' involvement in diseases using **IOBC**

3. Inference of **IOBC** chemicals' involvement in diseases using knowledge graphs (KGs) constructed from **IOBC**

1. Inference of NikkajiRDF chemicals' functions, roles, and applications using **ChEBI**



Inference of the roles and applications of NikkajiRDF's chemicals using ChEBI. It is inferred that "Aspirin" had not only "non-steroidal anti-inflammatory drug" as an application but also "Brønsted acid" as a chemical role. This diagram is visualized by a web service: https://www.kanzaki.com/works/2009/pub/graph-draw.

1,062 ChEBI's roles and applications were assigned to **18,386 NikkajiRDF's chemicals**

A SPARQL query for inferring NikkajiRDF chemicals' functions, role, and applications (**f/r/a**) using **ChEBI**.

```
prefix obo: <http://purl.obolibrary.org/obo/>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix nikkaji: <http://nikkaji.biosciencedbc.jp/>
SELECT distinct ?chebi ?role
WHERE
{
# NikkajiRDF.ttl
graph nikkaji:link2OtherDBs_basedOnUniChem20180515 {
        nikkaji:20090700317669603 5 skos:closeMatch ?chebi .
      }
# ChEBI.owl
graph nikkaji:ChEBI2018001.owl.gz {
        ?upperchebi rdfs:subClassOf
            [owl:someValuesFrom ?role ;
             owl:onProperty obo:RO_0000087] .
        ?chebi rdfs:subClassOf* ?upperchebi .
      }
}
```

Finding ChEBI's chemicals having the same **InChIKey** as Aspirin

Aspirin

Inferring Aspirins' **f/r/a** using **ChEBI's ontological structure.**

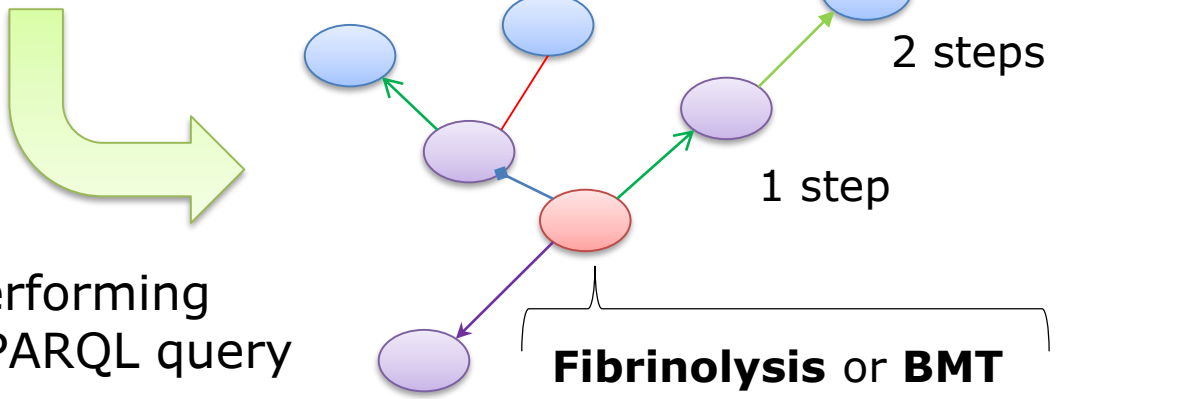
has role

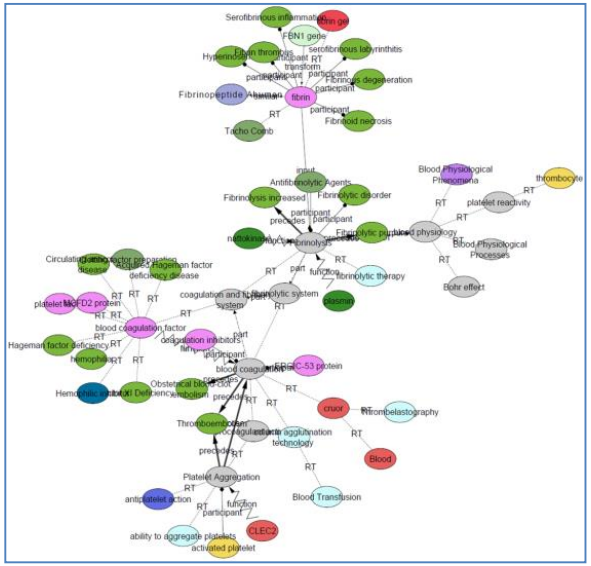
3. Inference of IOBC chemicals' involvement in diseases using knowledge graphs (**KGs**)

Method for developing KGs from IOBC

IOBC.owl

Ho - zo an Environment for Building/Using Ontologies 法造 www.hozo.jp



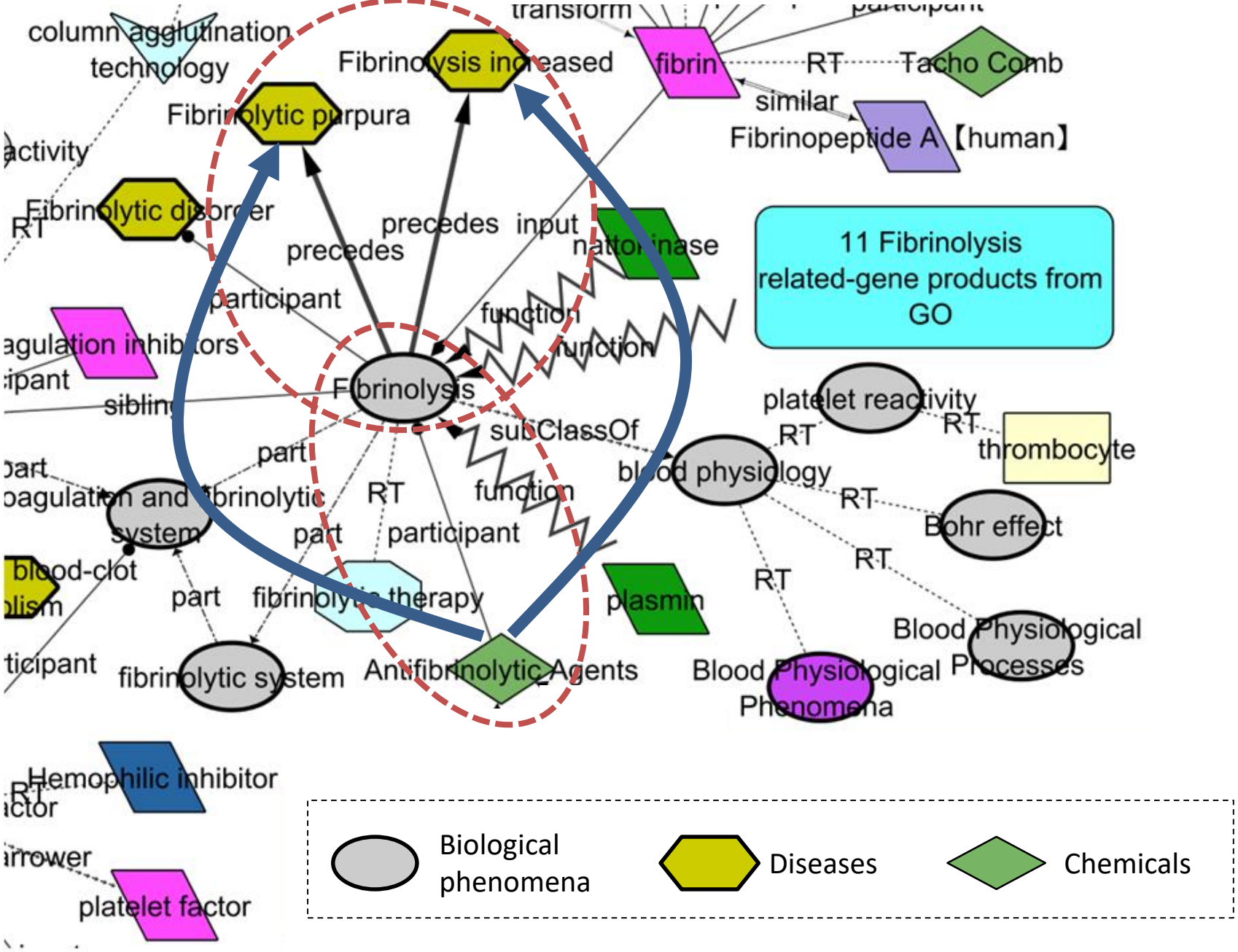


Knowledge graph (KG)

- **Fibrinolysis network**
- **Bone metabolism turnover (BMT) network**
- etc.

Visualization using Cytoscape

Cytoscape



A part of Fibrinolysis network. This graph is visualized using Cytoscape (http://www.cytoscape.org/).

A SPARQL query for inferring IOBC's chemicals' involvement in diseases using **KGs**.

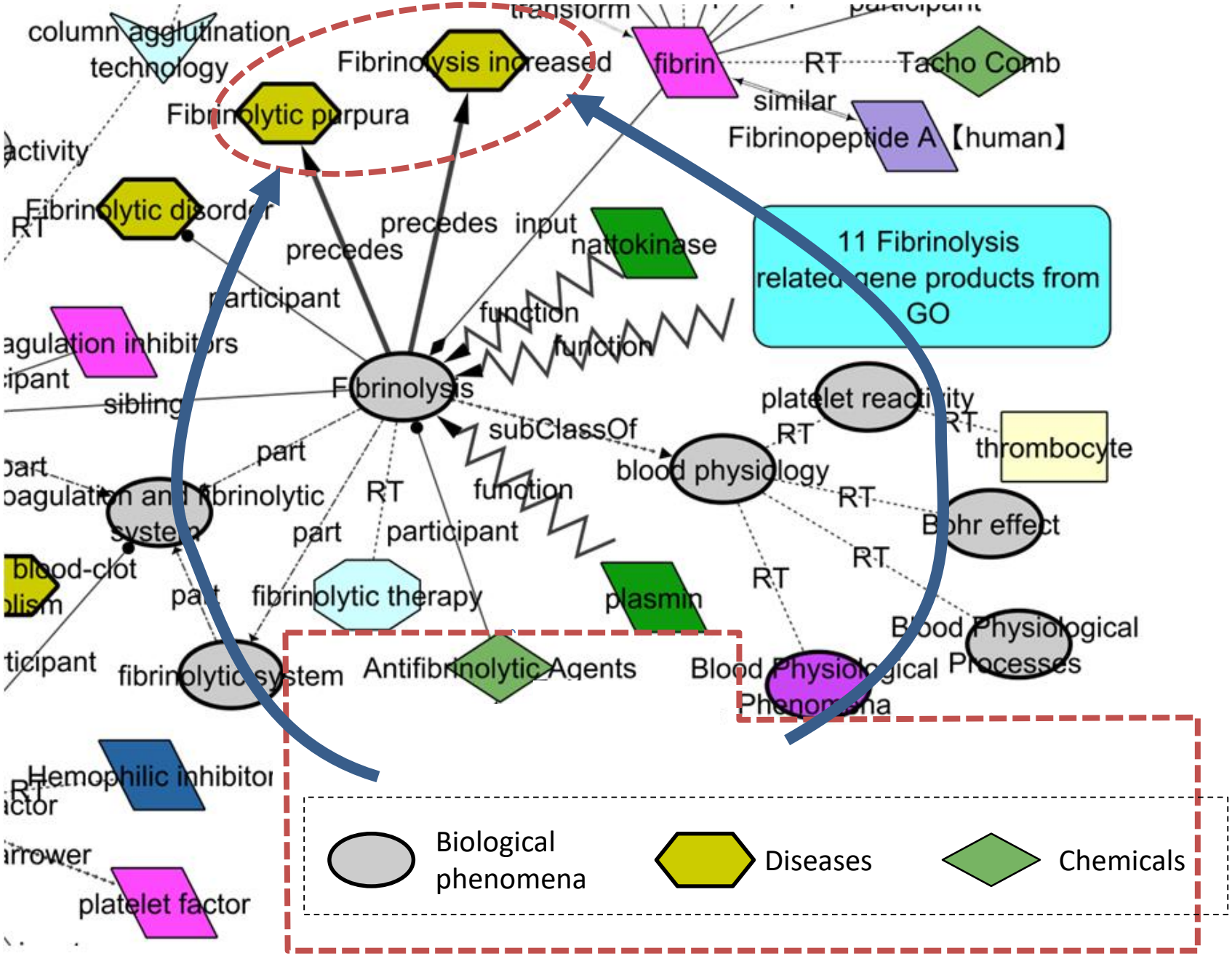
```
prefix dcterms: <http://purl.org/dc/terms/>
prefix sio: <http://semanticscience.org/resource/>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix subject: <http://purl.jp/bio/4/subject/>
prefix xkos: <http://rdf-vocabulary.ddialliance.org/xkos#>
SELECT distinct ?label_chem ?label_event ?label_disease
{
   {?event xkos:precedes ?disease .}
   UNION {?event sio:SIO_000062 ?disease .}
#finding pre-biological events of diseases
   {?chem sio:SIO_000062 ?event .} # SIO_000062 => "is participant in"
   UNION {?chem sio:SIO_000225 ?event .}
#finding chemicals participating in biological events
   ?disease dcterms:subject subject:LS51 . #diseases
   {?chem dcterms:subject subject:CA06 .}
#organic compounds
   UNION {?chem dcterms:subject subject:CA05 .}
#inorganic compounds
   UNION {?chem dcterms:subject subject:LS44 .} #drugs
   ?chem skos:prefLabel ?label_chem.
   ?event skos:prefLabel ?label_event.
   ?disease skos:prefLabel ?label_disease.
}
```

Results of the inferring chemicals' involvement with diseases using two KGs: "Fibrinolysis network" and "BMT network."

KGs	Chemicals or drugs	Pre-biological events of diseases	Diseases
Fibrinolysis network	Antifibrinolytic Agents	Fibrinolysis	Fibrinolysis increased Fibrinolytic disorder Fibrinolytic purpura
BMT network	Bone Density Conservation Agents	osteogenesis	Abnormal bone metabolism Bony cataract Enostosis Hyper-ALP-emia Lipoma ossificans Osteoplastic sarcoma Osteoplastica Tuberculous dactylitis
	nano hydroxyapatite		
	osteogenesis promoter		
	Teriparatide		
	bone resorption inhibitor	Bone Resorption	Alveolar Bone Loss Osteitis Fibrosa Cystica Osteolysis Osteolysis, Essential Osteolytic lesion
	denosumab		

7 chemicals/drugs

16 diseases



Inferring the **ChEBI's**, and **MeSH's** chemicals' involvement in diseases using the KGs.

Conclusions

- DBpedia and Wikidata are currently neither reasonable nor suitable for efficiently collecting information on chemicals.
- This study developed a method of inferring chemical information on the internet, particularly **chemicals' involvement in diseases**, as well as the functions, roles, and applications.
- Data sources should become more findable, accessible, interoperable, and reusable based on the **FAIR principle**.
- The preparation of **InChI/InChIKey** is effective for integrating chemicals among different data sources.
- Enabling the **federated search on SPARQL endpoints** is also important.
- In the future, the leverage of information on the **interactions between chemicals and gene products** and **metabolic and signal transduction pathways** will enable more extensive, and precise collection and prediction of information regarding chemicals' associations with biological phenomena, along with the corresponding side effects.

Acknowledgments

- This study was supported by an operating grant from the Japan Science and Technology Agency and JSPS KAKENHI Grant Number JP17H01789.
- A part of this study was progressed and discussed in Japan BioHackathon 2016 (BH16.12), which served as a research and development meeting. We are grateful to all participants who gave us their valuable advice and constructive comments.

Thank you for your attention.

kushida@biosciencedbc.jp