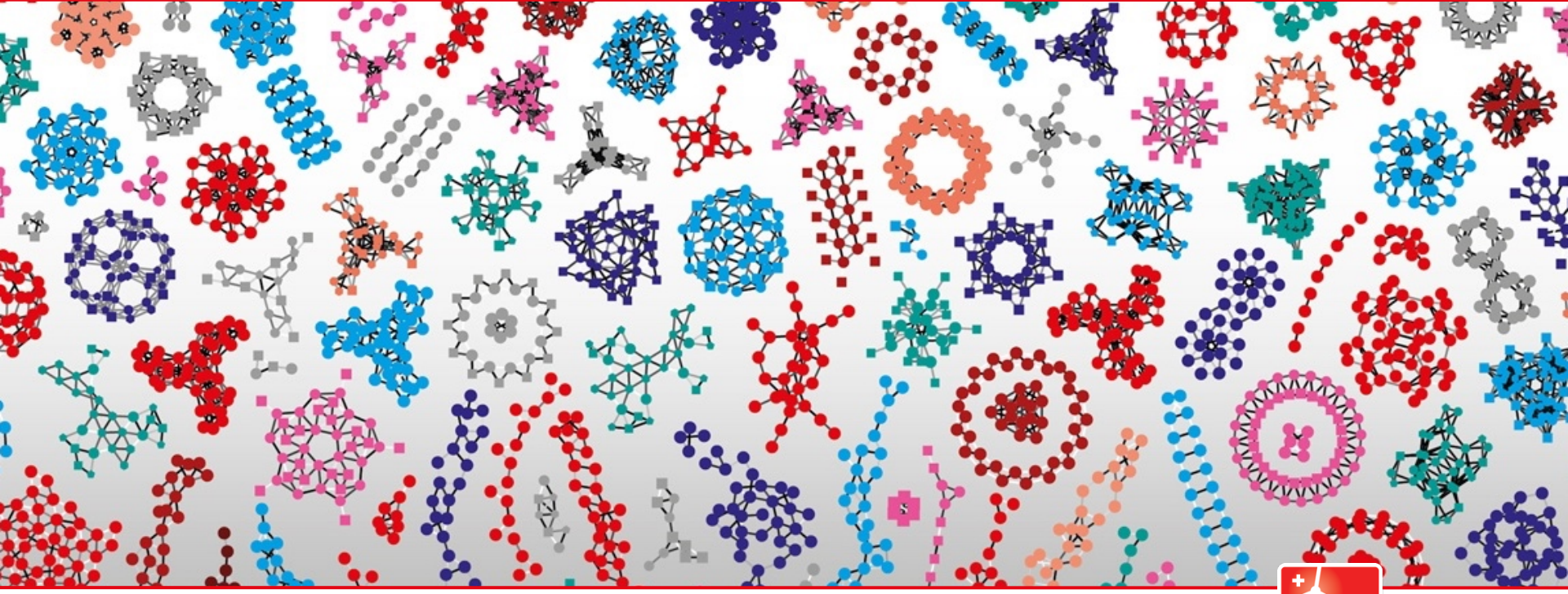


Swiss Institute of Bioinformatics

MetaNetX/MNXref

a resource for systems biology and metabolomics

SWAT4(HC)LS
Edinburgh
9th December 2019

Marco Pagni

elixir SWITZERLAND

www.sib.swiss

Genome-Scale Metabolic Network (GSMN)

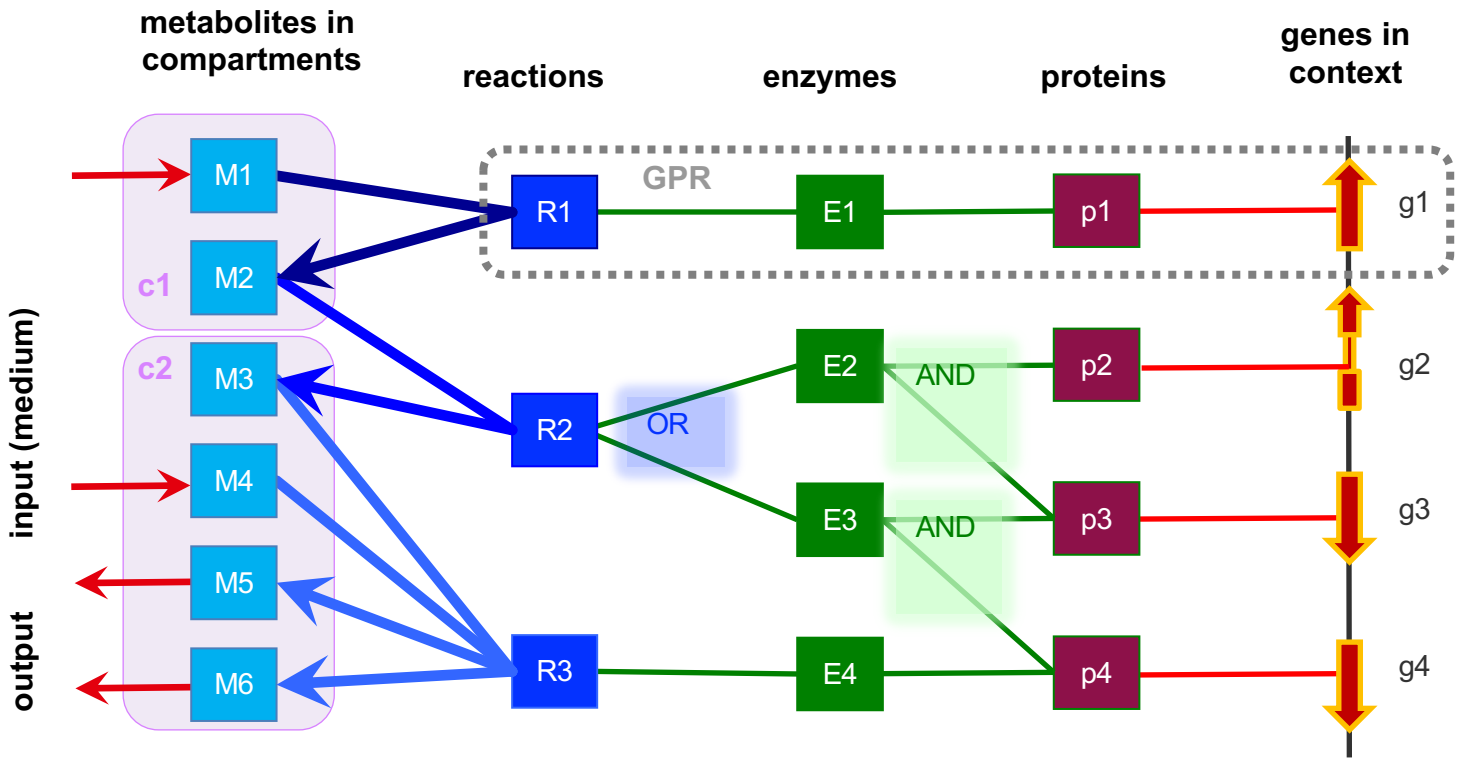
A genome-scale metabolic network (GSMN) is an *in silico* reconstruction of an organism metabolism.

GSMN has a dual nature:

- it is a **repository of knowledge** about an organism metabolism
- it is a **numerical model** that can be simulated, to generate predictions that can be tested experimentally

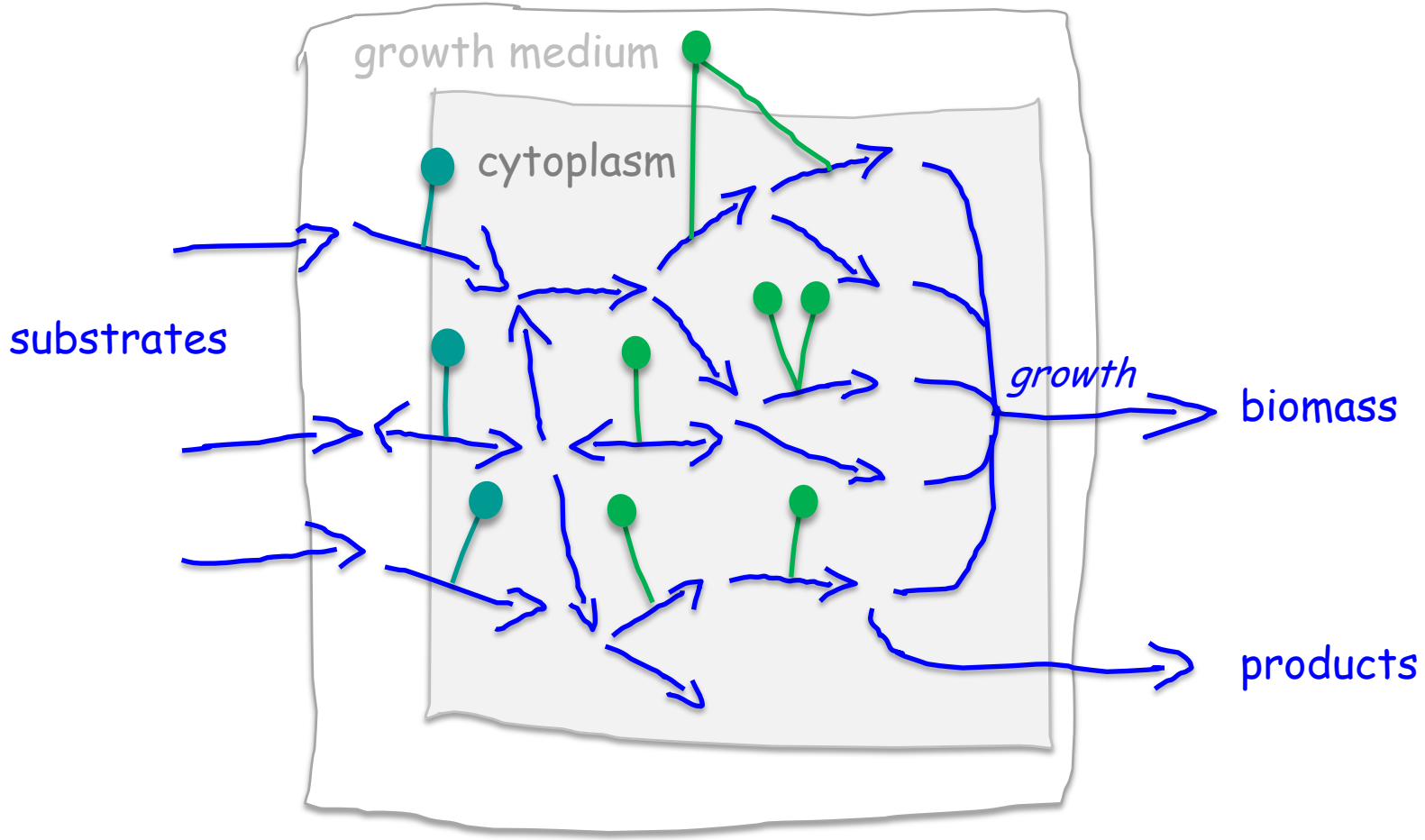
- Applications: metabollic engineering, synthetic biology

Genome Scale Metabolic Network (GSMN)

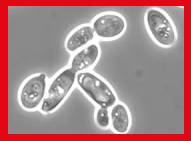


The cartoon network view of a GSMN

The cartoon network view of a GSMN

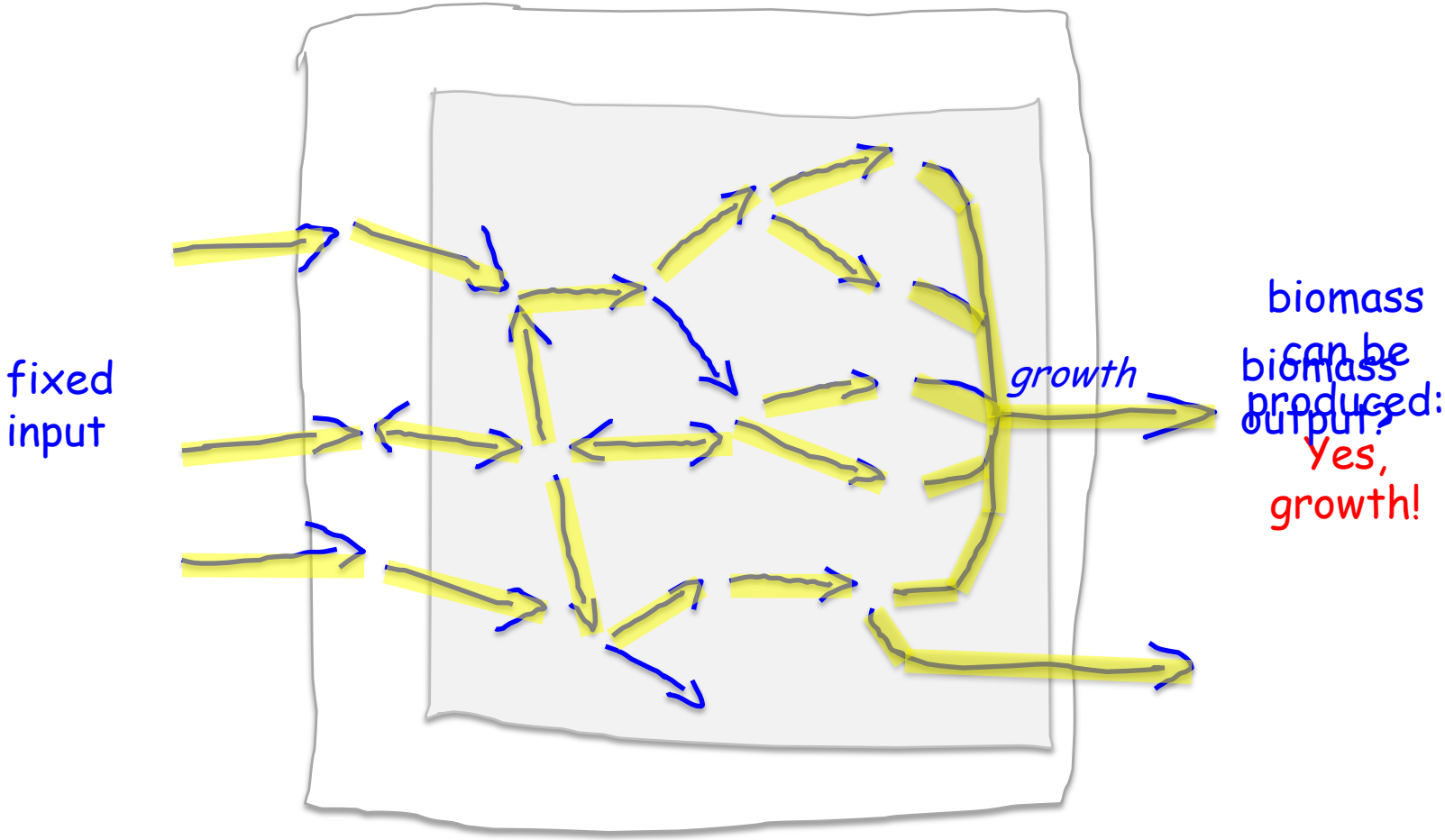


Cartoon network *vs* iMK735 *vs* iMM1415

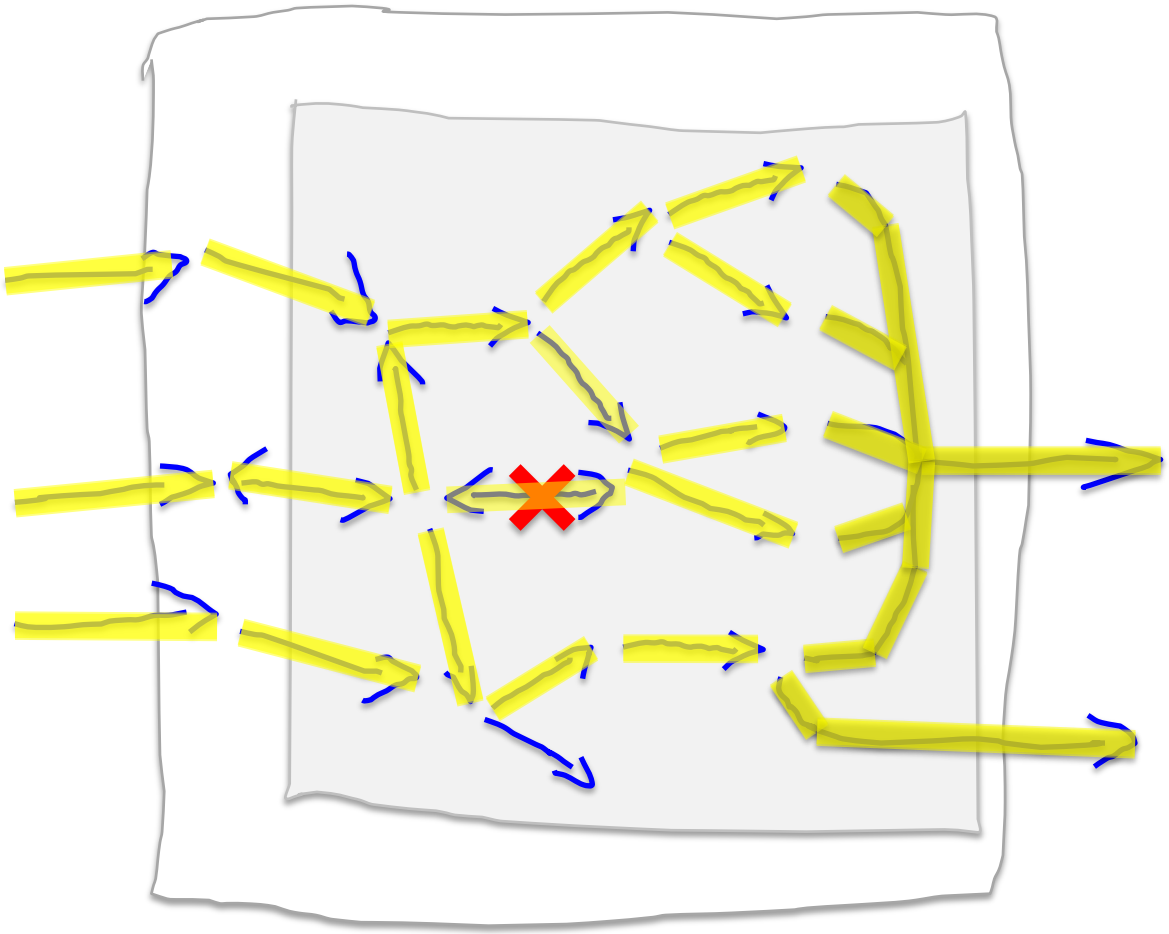
	cartoon network	iMK735  *Yarrowia lipolytica*	iMM1415*  Mouse
Compartments	2	9	9
Chemical compounds	12	678	1493
Species (chemical@compartment)	15	1224	3203
Enzymatic reactions (one compartment)	14	883	2040
Transport reactions (two compartments)	3	318	1188
Boundary reactions (source or sink)	4	121	450
Genes	8	709	1376
Substrates in growth reaction	5	43	63
number of reactions for the most connected metabolites	4 3 3	618 (H^+) 348 (H_2O) 207 (ATP)	1271 (H^+) 907 (H_2O) 319 (ATP)

*: with a modified growth equation

Linear-programming based simulations, *e.g.* FBA



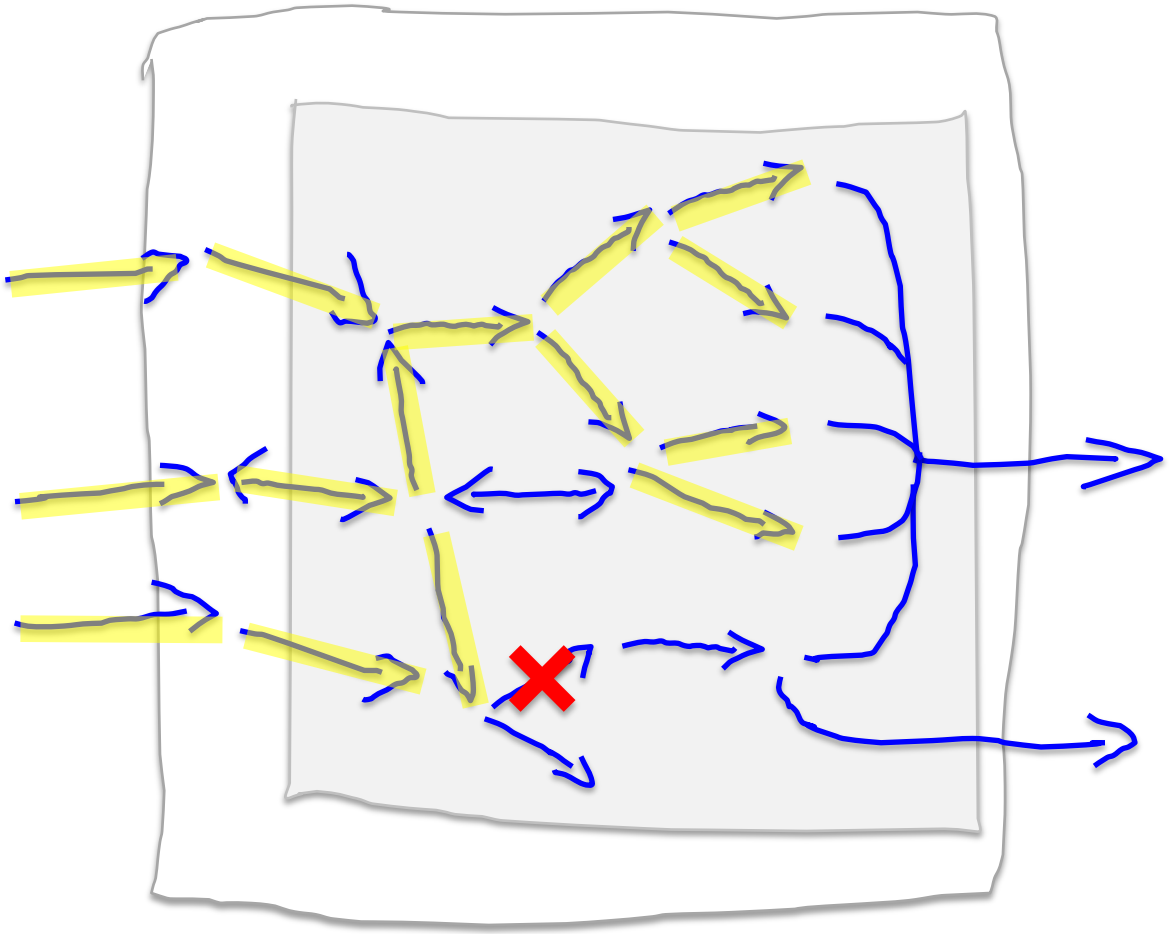
Reaction knockout simulation



biomass is produced?
Yes

=> this reaction is not essential

Growth recovery and blocked reactions

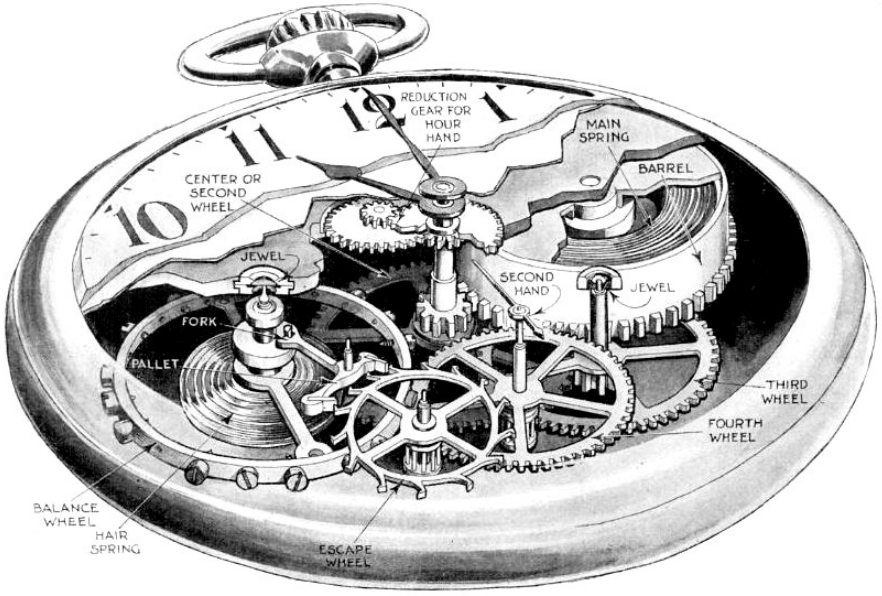


biomass is produced?
No

=> this reaction is essential

GSMNs are like mechanical Swiss watches

- They are unique and worth collecting ;-)
- But they are fragile: Typically, between 10 and 20 % of reactions are essential: The removal of any one of them kill the model!
- Great care must be taken while manipulating/modifying a model, for example during metabolite reconciliations, which is the main purpose of MetaNetX.



Mycobacterium tuberculosis GSMNs

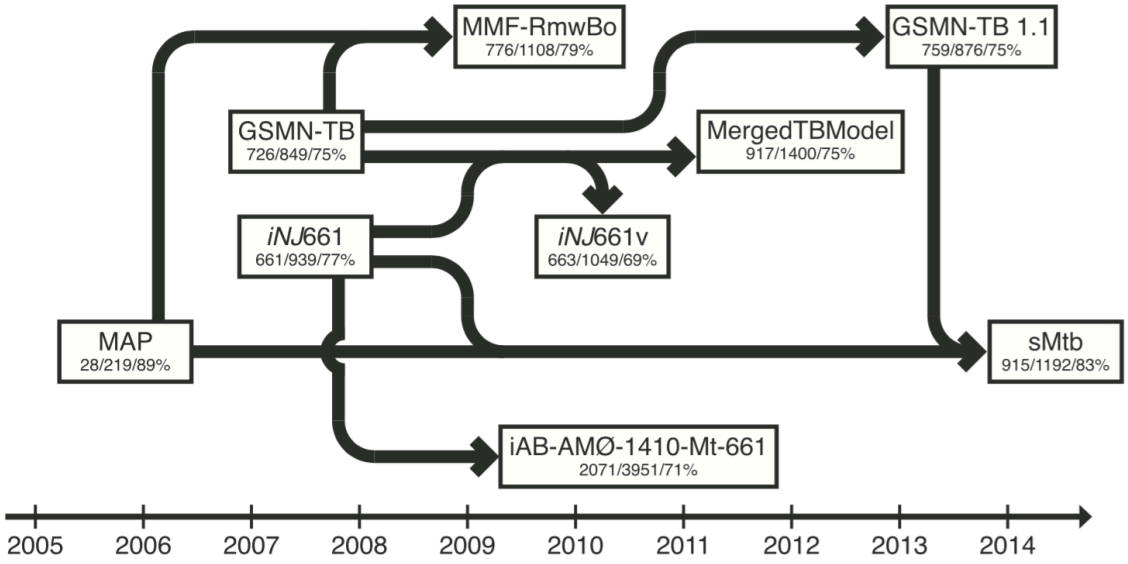
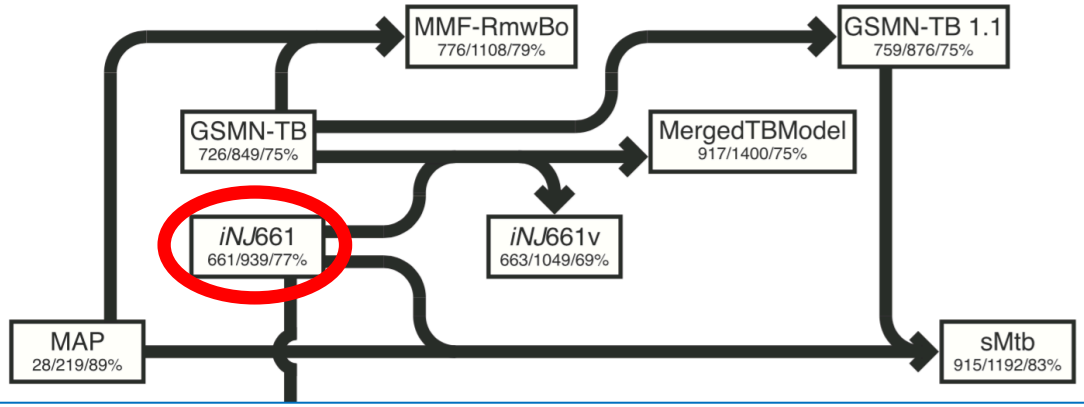


Fig. 2. Time line of CB metabolic models of Mtb. The numbers below every model name denote the number of genes, the number of reactions and the percentage of gene-associated reactions in the model. MAP: mycolic acid pathway [14], GSMN-TB: genome-scale metabolic network of *M. tuberculosis* [30], *iNJ*661: *in silico* Neema Jamshidi 661 genes [23], MMF-RmwBo: merged McFadden-Ramam with biomass objective [32], iNJ661v: *in vivo* compatible model based on iNJ661 [33], iAB-AMØ-1410Mt-661: *in silico* Aarash Bordbar alveolar macrophage 1410 genes *Mycobacterium tuberculosis* 661 genes [7], MergedTBModel: merged *Mycobacterium tuberculosis* model [35], GSMN-TB 1.1: a curated and extended version of GSMN-TB [24], sMtb: *in silico Mycobacterium tuberculosis*.

iNJ661



BMC Systems Biology

BioMed Central

Research article

Open Access

Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain *iNJ661* and proposing alternative drug targets

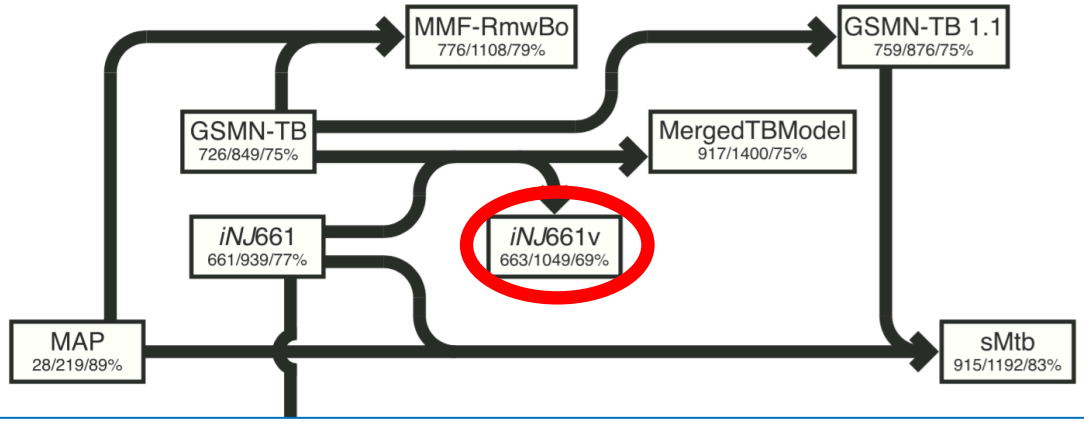
Neema Jamshidi and Bernhard Ø Palsson*

Address: Department of Bioengineering, University of California, San Diego, La Jolla, CA, 92093-0412, USA.

Email: Neema Jamshidi - neema@ucsd.edu; Bernhard Ø Palsson* - palsson@ucsd.edu

* Corresponding author

iNJ661v



Fang *et al. BMC Systems Biology* 2010, **4**:160
http://www.biomedcentral.com/1752-0509/4/160

BMC Systems Biology

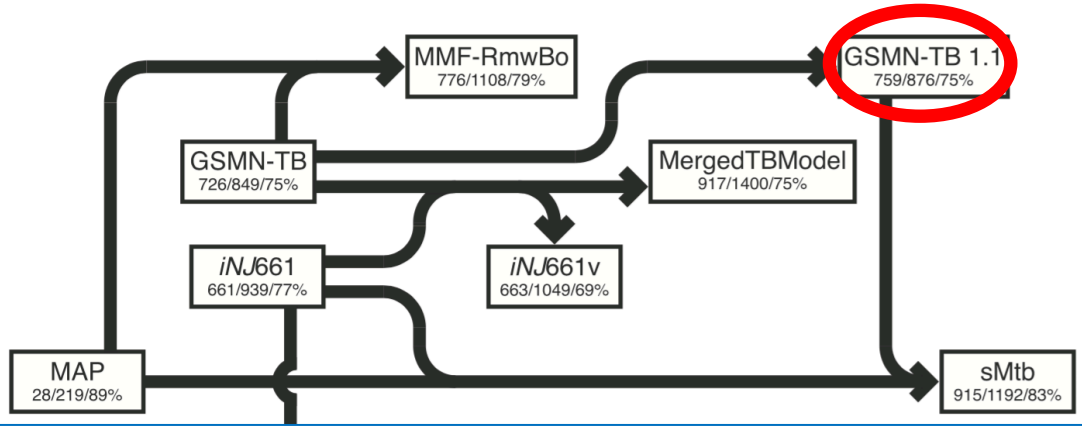
RESEARCH ARTICLE Open Access

Development and analysis of an *in vivo*-compatible metabolic network of *Mycobacterium tuberculosis*

Xin Fang, Anders Wallqvist, Jaques Reifman*

Abstract

GSMN-TB 1.1



OPEN ACCESS Freely available online

PLOS ONE

Systems-Based Approaches to Probing Metabolic Variation within the *Mycobacterium tuberculosis* Complex

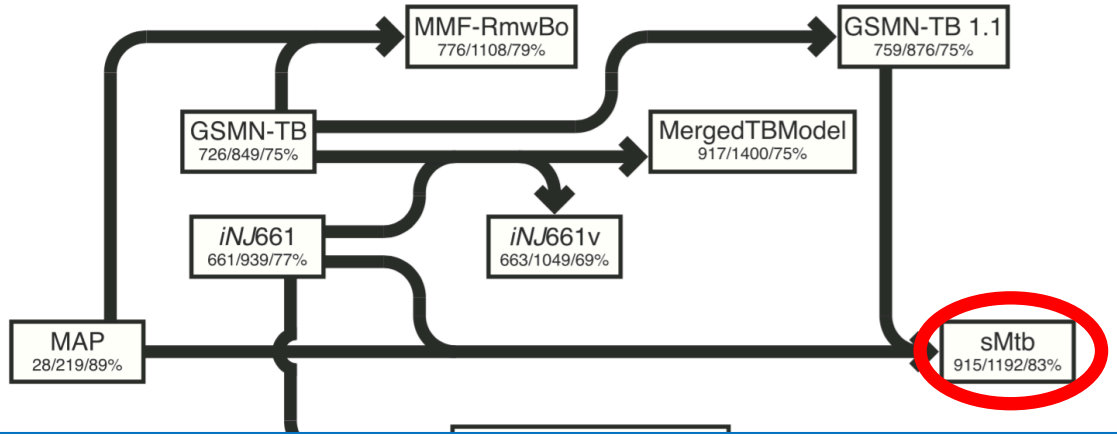
Emma K. Lofthouse[1,2], Paul R. Wheeler[1], Dany J. V. Beste[2], Bhagwati L. Khatri[1], Huihai Wu[2], Tom A. Mendum[2], Andrzej M. Kierzek[2], Johnjoe McFadden[2*]

1 Animal Health and Veterinary Laboratories Agency (Weybridge), Department for Bovine Tuberculosis, New Haw, Surrey, United Kingdom, **2** Department of Microbial and Cellular Sciences, Faculty of Health and Medical Sciences, University of Surrey, Stag Hill, Guildford, Surrey, United Kingdom

Abstract

The *Mycobacterium tuberculosis* complex includes bovine and human strains of the tuberculosis bacillus, including *Mycobacterium tuberculosis, Mycobacterium bovis* and the *Mycobacterium bovis* BCG vaccine strain. *M. bovis* has evolved from a *M. tuberculosis*-like ancestor and is the ancestor of the BCG vaccine. The pathogens demonstrate

iNJ661v



Seminars in Immunology 26 (2014) 610–622

Contents lists available at ScienceDirect

Seminars in Immunology

journal homepage: www.elsevier.com/locate/ysmim

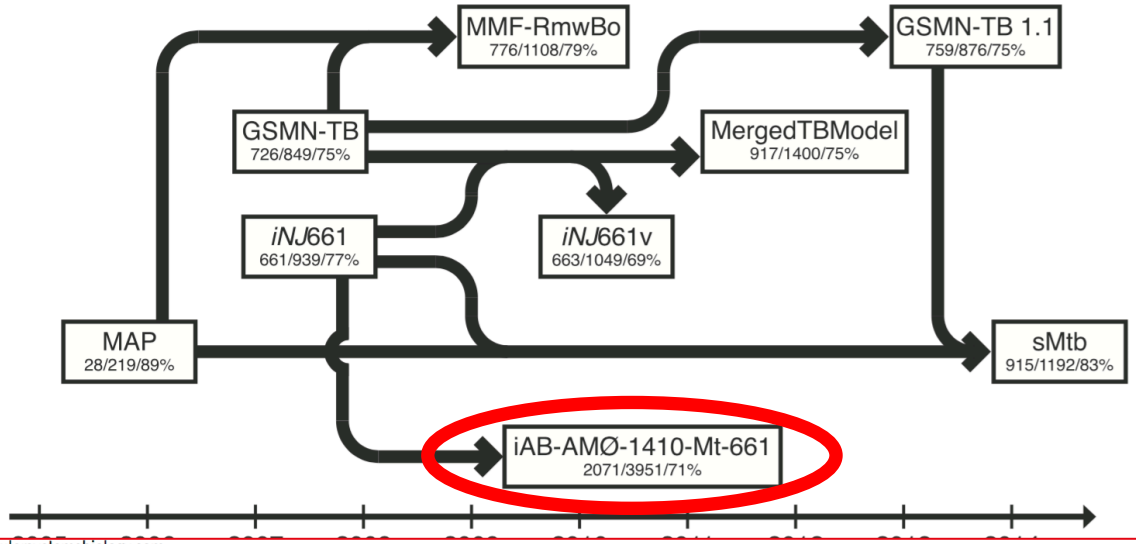
Review

Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets

Rienk A. Rienksma[a], Maria Suarez-Diez[a], Lucie Spina[b], Peter J. Schaap[a], Vitor A.P. Martins dos Santos[a,c,*]

[a] Laboratory of Systems and Synthetic Biology, Wageningen University and Research Centre, Dreijenplein 10, Wageningen 6703 HB, The Netherlands
[b] Centre National de la Rescherche Scientifique (CNRS), Institut de Pharmacologie et de Biologie Structurale (UMR 5089), Department of Tuberculosis and Infection Biology and Université de Toulouse (Université Paul Sabatier, Toulouse III), IPBS, 205 Route de Narbonne, BP 64182, F-31077 Toulouse, France
[c] Lifeglimmer GmbH, Markelstrasse 38, 12163 Berlin, Germany

iAB-AM0-1410-Mt-661



www.molecularsystemsbiology.com

Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions

Aarash Bordbar[1], Nathan E Lewis[1], Jan Schellenberger[2], Bernhard Ø Palsson[1] and Neema Jamshidi[1,*]

[1] Department of Bioengineering, University of California, San Diego, Powell-Focht Bioengineering Hall, La Jolla, CA, USA and [2] Department of Bioengineering, Bioinformatics Program, University of California, San Diego, Powell-Focht Bioengineering Hall, La Jolla, CA, USA
* Corresponding author. Department of Bioengineering, University of California, San Diego, 417 Powell Focht Bioengineering Hall, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA. Tel.: +1 858 822 1144; Fax: +1 858 822 3120; E-mail: neema@ucsd.edu

Received 22.2.10; accepted 30.7.10

Essays in Biochemistry (2019) **63** 267–284
https://doi.org/10.1042/EBC20180045

PORTLAND PRESS

Review Article

Genome-driven cell engineering review: *in vivo* and *in silico* metabolic and genome engineering

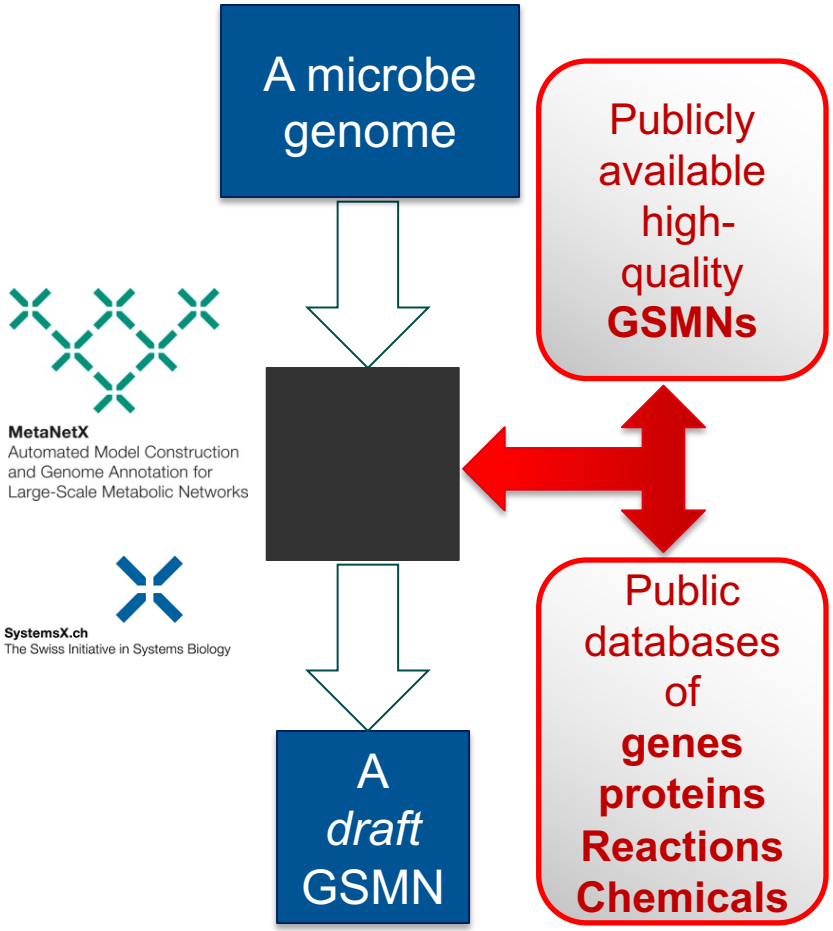
Sophie Landon[1,2,*], Joshua Rees-Garbutt[1,3,*], Lucia Marucci[1,2,4,†] and Claire Grierson[1,3,†]

[1]BrisSynBio, University of Bristol, Bristol BS8 1TQ, U.K.; [2]Department of Engineering Mathematics, University of Bristol, Bristol BS8 1UB, U.K.; [3]School of Biological Sciences, University of Bristol, Life Sciences Building, Bristol BS8 1TQ, U.K.; [4]School of Cellular and Molecular Medicine, University of Bristol, Bristol BS8 1UB, U.K.

Correspondence: Claire Grierson (claire.grierson@bristol.ac.uk) or Lucia Marucci (lucia.marucci@bristol.ac.uk)

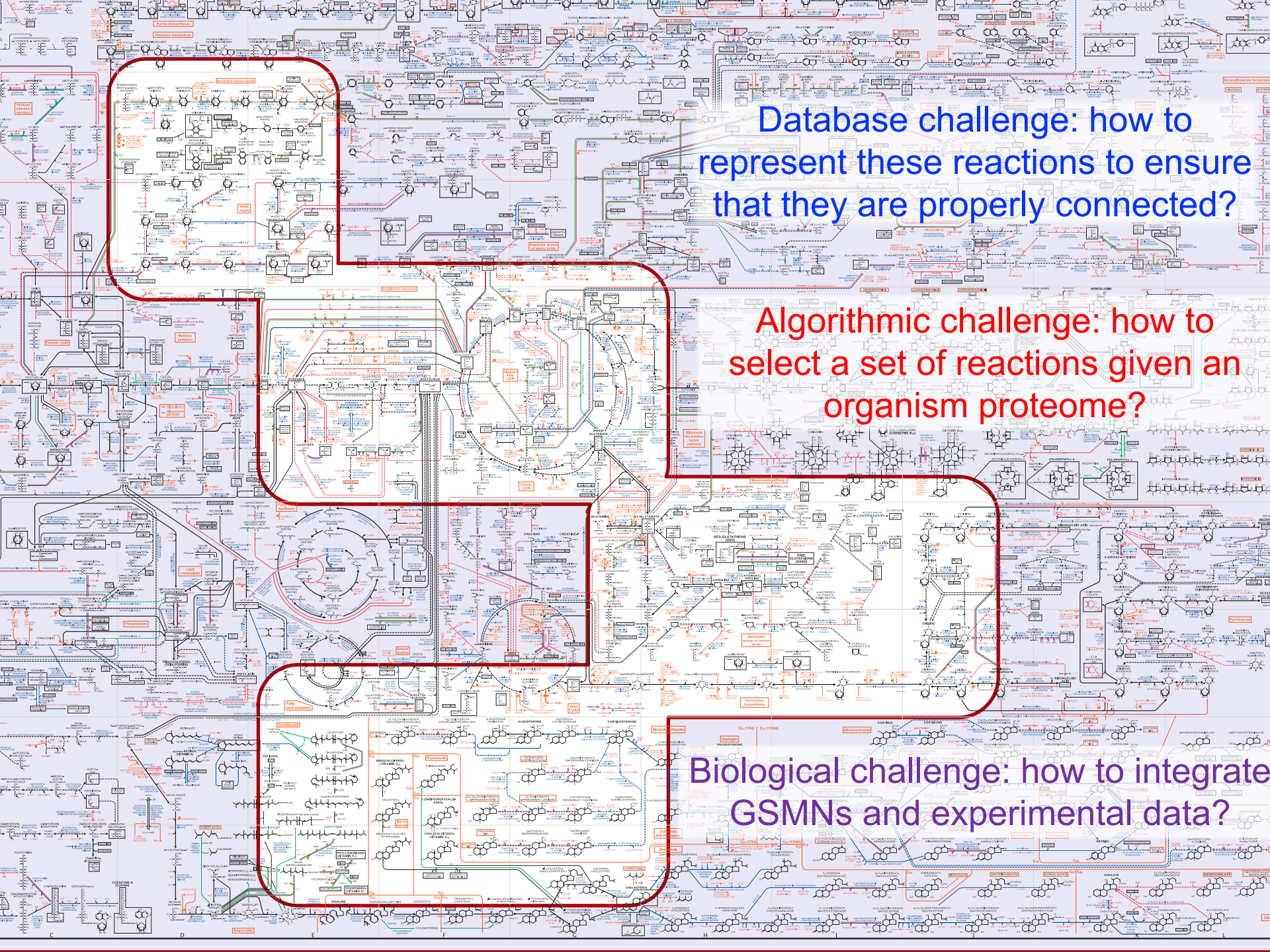
Producing 'designer cells' with specific functions is potentially feasible in the near future. Recent developments, including whole-cell models, genome design algorithms and gene editing tools, have advanced the possibility of combining biological research and mathematical modelling to further understand and better design cellular processes. In this review, we will explore computational and experimental approaches used for metabolic and genome design. We will highlight the relevance of modelling in this process, and challenges associated

The initially encountered difficulty (in 2010)

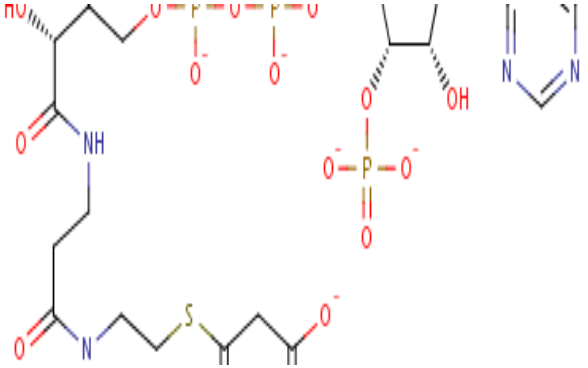


- About hundred high-quality GSMNs have been published by different groups, with specific nomenclature for metabolites and reactions and no information about molecular structure

- How can we compare the GSMNs published by different groups?
- How can we integrate them with public resources like KEGG, CHEBI, RHEA...?


Database challenge: how to represent these reactions to ensure that they are properly connected?
Algorithmic challenge: how to select a set of reactions given an organism proteome?
Biological challenge: how to integrate GSMNs and experimental data?

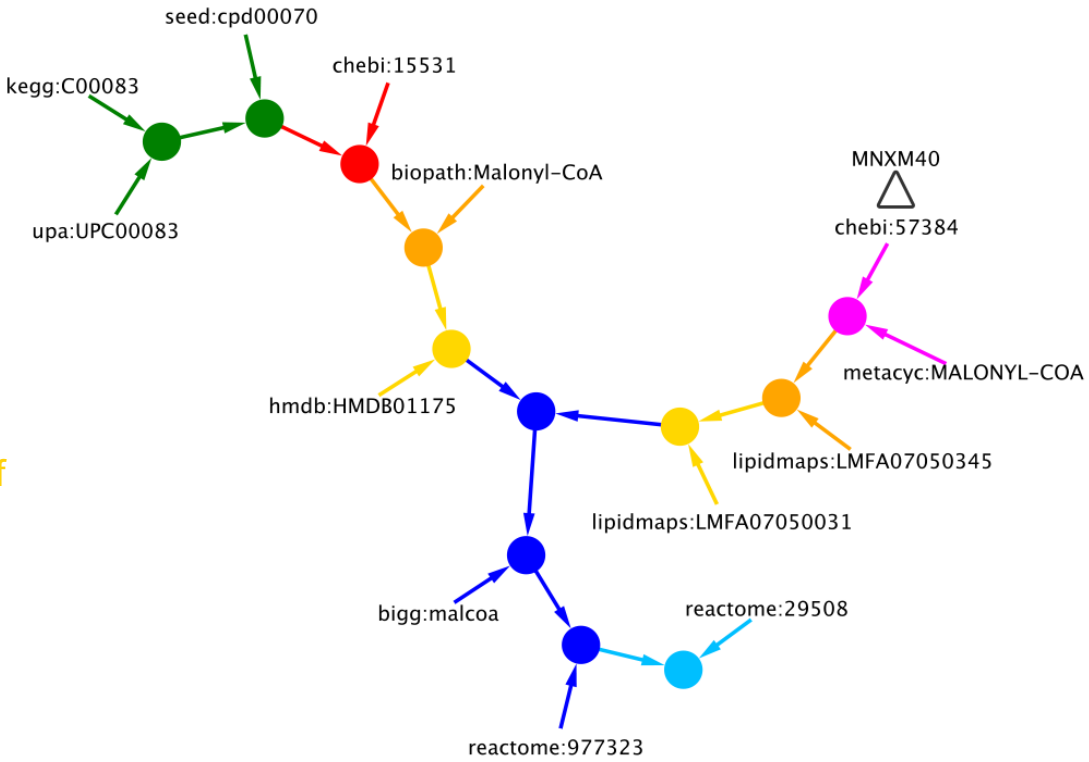
Automated reconciliation of malonyl-CoA



Evidences used in reconciliation:

- structure as supplied by the source databases
- recomputed structure
- recomputed structure protonated at pH 7.3
- recomputed structure protonated at pH 7.3 but ignoring the stereo layer of the InChI representation
- using the cross-references supplied by the source databases
- based on compound primary names
- based on compound synonym names

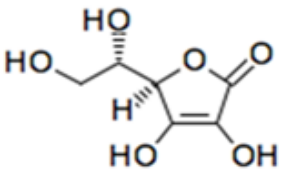
Triangle: in a post-processing step, chebi:57384 was chosen to best represent MNXM40



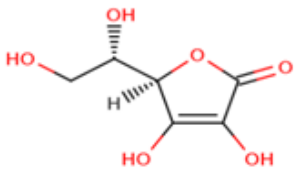
Different representation of a metabolite

a

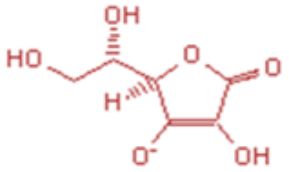
Different protonation states



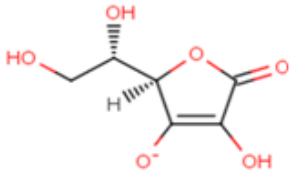
KEGG:C00072
ascorbate



CHEBI:29073
L-ascorbic acid



MetaCyc:ASCORBATE
L-ascorbate



CHEBI:38290
L-ascorbate

b

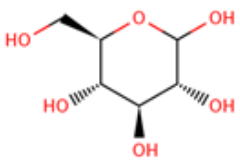
Di
tau

tone

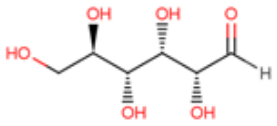
Automated solution:
using molecular structure (SMILES, InChI)
...and fix manually wrong molecular information

c

Circular / linear forms of carbohydrates



CHEBI:4167
D-glucopyranose

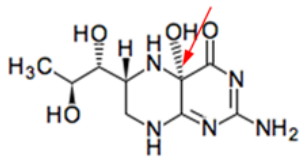


CHEBI:42758
aldehydo-D-glucose

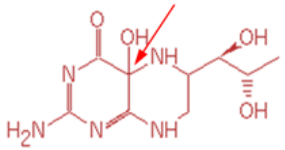
Mistakes, errors, imprecision...

d

Erroneous or unspecified stereochemistry



KEGG:C15522

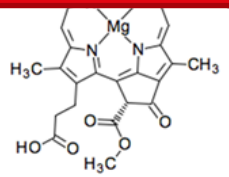


MetaCyc:CPD-5881

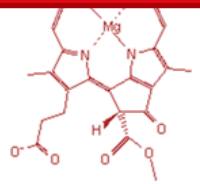
4α-hydroxy-tetrahydrobiopterin

Solution:
Trust some resources more than others
Consider the context given by the reactions…

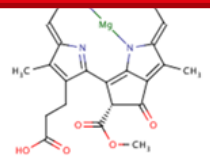
Representation of bonds in organo-metalic complexes



KEGG:C02880



METACYC: MONO-VINYL -PROTOCHLOROPHYLLIDE-A

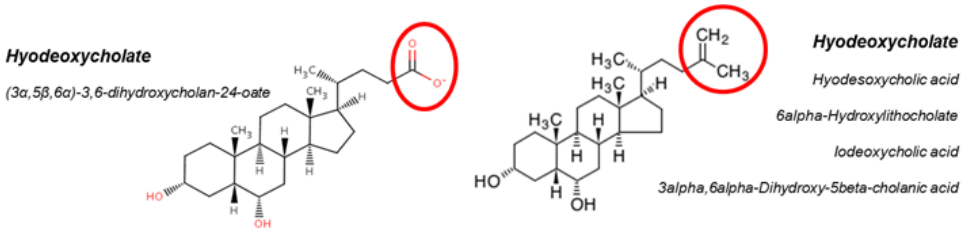


CHEBI:16673

protochlorophyllide

g

Incorrect naming and synonyms



Example of two different compounds with identical names found in two different resources

Reconciliation of reactions

rhea:26434 **rhea:26435** **rhea:26436** **rhea:26437**	chebi:16881 + 1 chebi:33384 = 1 chebi:15377 + 1 chebi:57912
kegg:R00674	C00065 + C00463 = C00078 + C00001
bigg:TRPS2	1 h2o@c + 1 trp__L@c = 1 indole@c + 1 ser__L@c
bigg:TRPS2h	1 h2o@h + 1 trp__L@h = 1 indole@h + 1 ser__L@h
metacyc:RXN0-2382	1 metacyc:SER + 1 metacyc:INDOLE = 1 metacyc:TRP + 1 metacyc:WATER
seed:rxn00474	1 cpd00054@c + 1 cpd00359@c = 1 cpd00001@c + 1 cpd00065@c

MNXR104946 **1 MNXM2@MNXD1 + 1 MNXM94@MNXD1 = 1 MNXM377@MNXD1 + 1 MNXM53@MNXD1**

H2O L-tryptophan indole L-serine

placeholder (generic) compartment

Metabolite reconciliation through reactions

Reactions of resource 1

R1: C1 + C2 ↔ C3

R2: C4 + C5 ↔ C1 + C6

R3: C4 + C7 ↔ C1 + C8

Structurally reconciled compounds:

C2 = c7 C4 = c1

C3 = c8 C7 = c5

Cross-references between reactions:

Xref1: R2 = r1

Xref1: R3 = r2

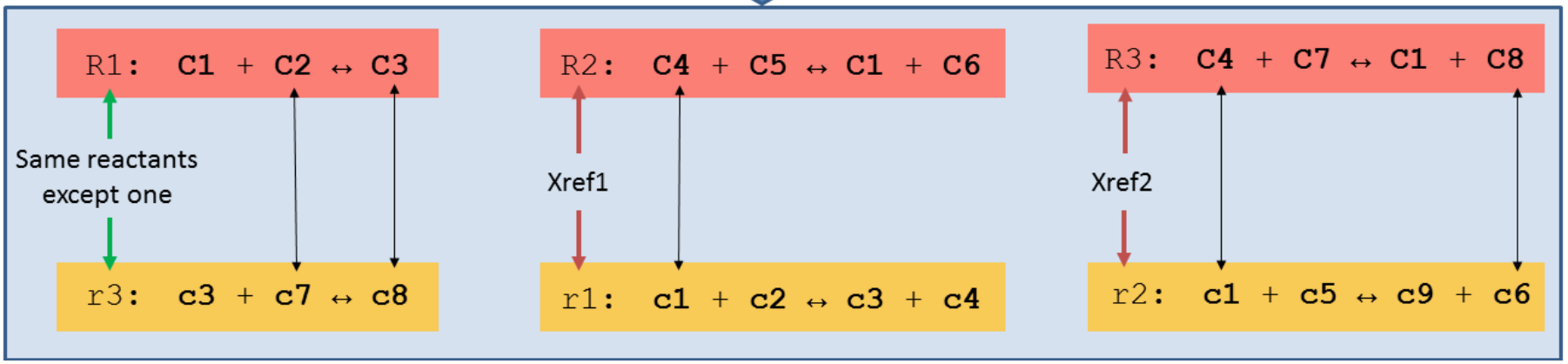
Reactions of resource 2

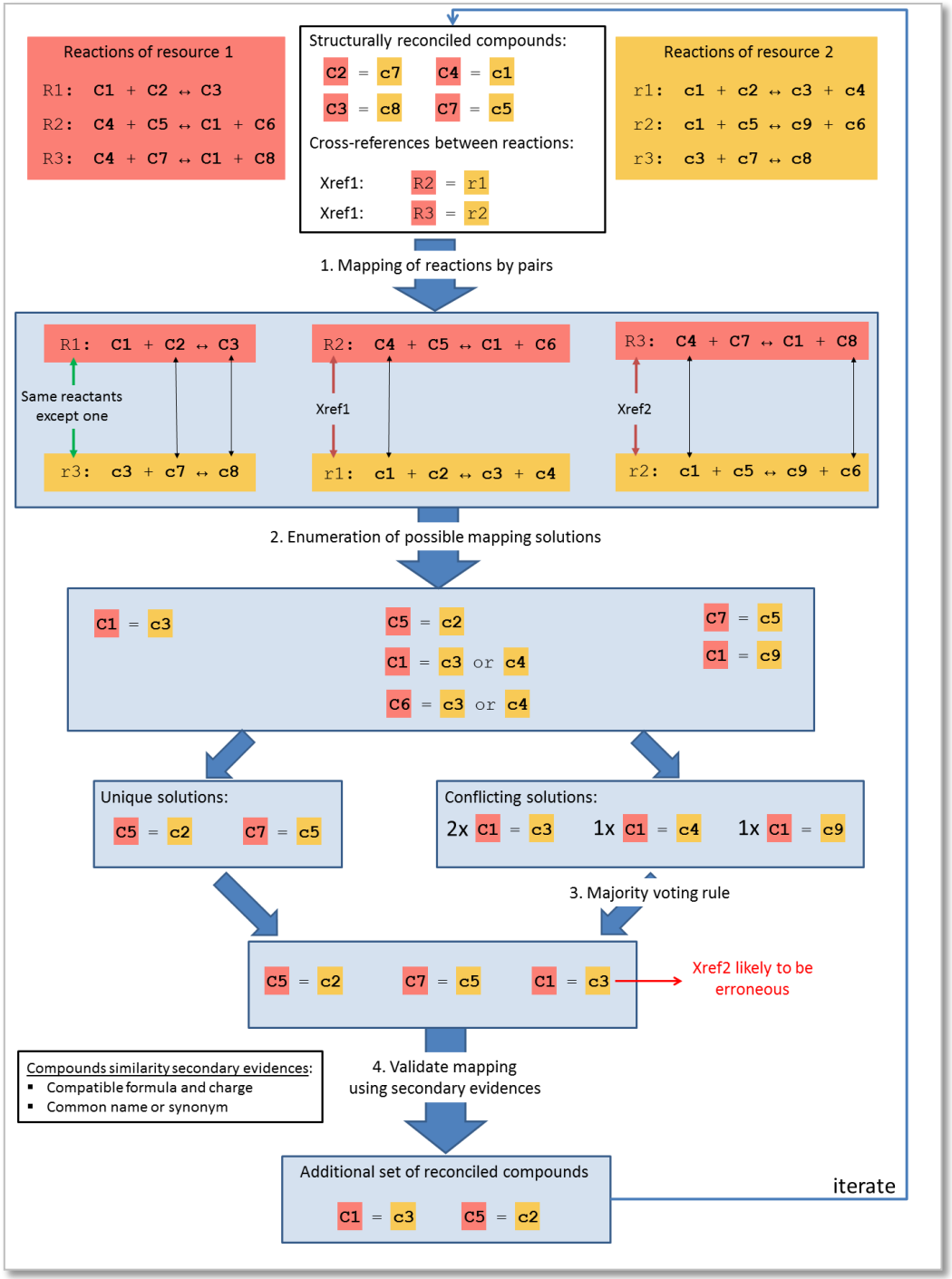
r1: c1 + c2 ↔ c3 + c4

r2: c1 + c5 ↔ c9 + c6

r3: c3 + c7 ↔ c8

1. Mapping of reactions by pairs





BRIEFINGS IN BIOINFORMATICS. VOL 15. NO 1. 123–135
Advance Access published on 19 November 2012
doi:10.1093/bib/bbs058

Reconciliation of metabolites and biochemical reactions for metabolic networks

Thomas Bernard, Alan Bridge, Anne Morgat, Sébastien Moretti, Ioannis Xenarios and Marco Pagni

Submitted: 15th May 2012; Received (in revised form): 7th August 2012

Abstract

Genome-scale metabolic network reconstructions are now routinely used in the study of metabolic pathways, their evolution and design. The development of such reconstructions involves the integration of information on reactions and metabolites from the scientific literature as well as public databases and existing genome-scale metabolic models. The reconciliation of discrepancies between data from these sources generally requires significant manual curation, which constitutes a major obstacle in efforts to develop and apply genome-scale metabolic network reconstructions. In this work, we discuss some of the major difficulties encountered in the mapping and reconciliation of metabolic resources and review three recent initiatives that aim to accelerate this process, namely BKM-react, MetRxn and MNXref (presented in this article). Each of these resources provides a pre-compiled reconciliation of

MNXref reconciliation of metabolites and biochemical reactions

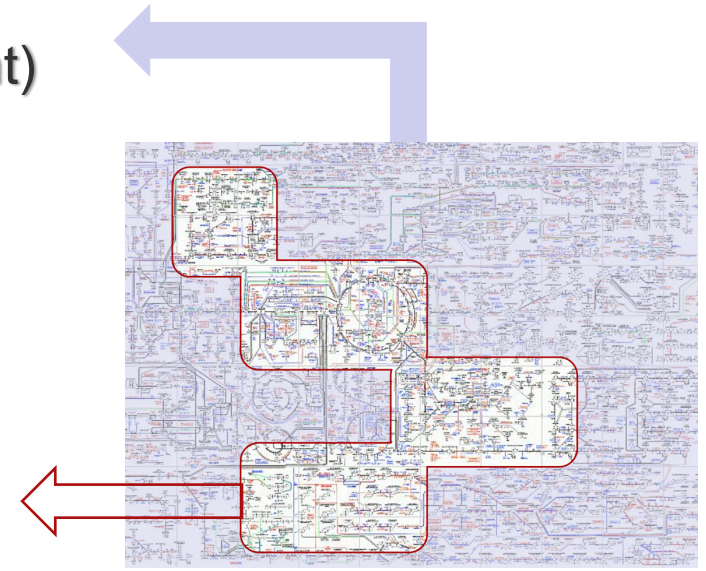
MNXref release	1.0 / 1.1	2.0	3.0
BioPath	2010/05/03	2010/05/03	
BiGG	2013/06/04	2015/09/02	2016/11/04
BKM/BRENDA	2013/04/12		
ChEBI	102	131	145
GO	2012/06/09	2015/09/02	2016/09/21
HMDB	3.0	3.6	3.6
KEGG	2013/04/12	75.1	79.0
LipidMaps	2013/04/12	2015/06/28	2016/09/21
MetaCyc	17.0	19.1	20.5
Reactome	2013/05/24	53	57
Rhea	39	64	75
SABIO-RK			2016/05/27
SwissLipids			2016/10/18
TheSEED	2013/04/12	2013/06/19	2016/12/16
UMBBD-EAWAG → enviPath		2014/06/30	2016/10/11
UniPathway	2012/04	2015/03	

MNXref reconciliation of metabolites and biochemical reactions

MNXref reconciliation:

691 · 10^3 metabolites

42 · 10^3 reactions (with generic compartment)



MetaNetX repository:

134 GSMNs and the like

55 · 10^3 reactions (in specific compartments)

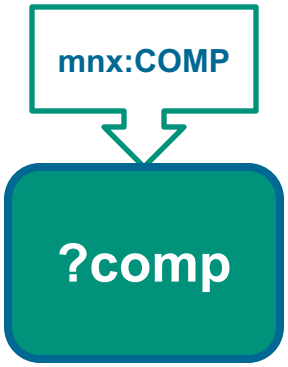
11 · 10^3 metabolites

63 · 10^3 genes or proteins

RDF: ~13.5*10^6 triples

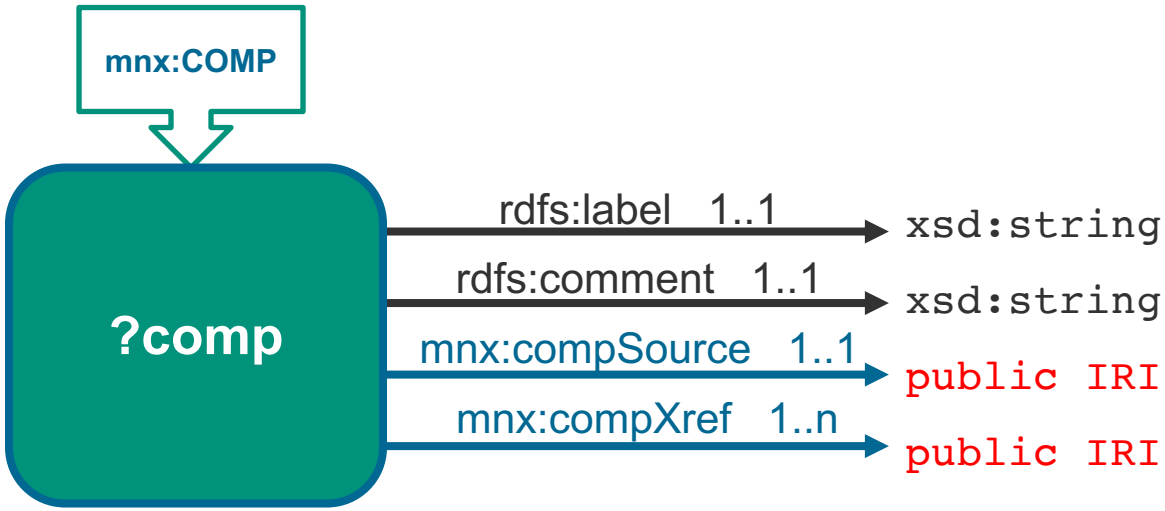
Notation

- IRI and blank nodes are systematically typed in MetaNetX RDF
- Graphical representation:



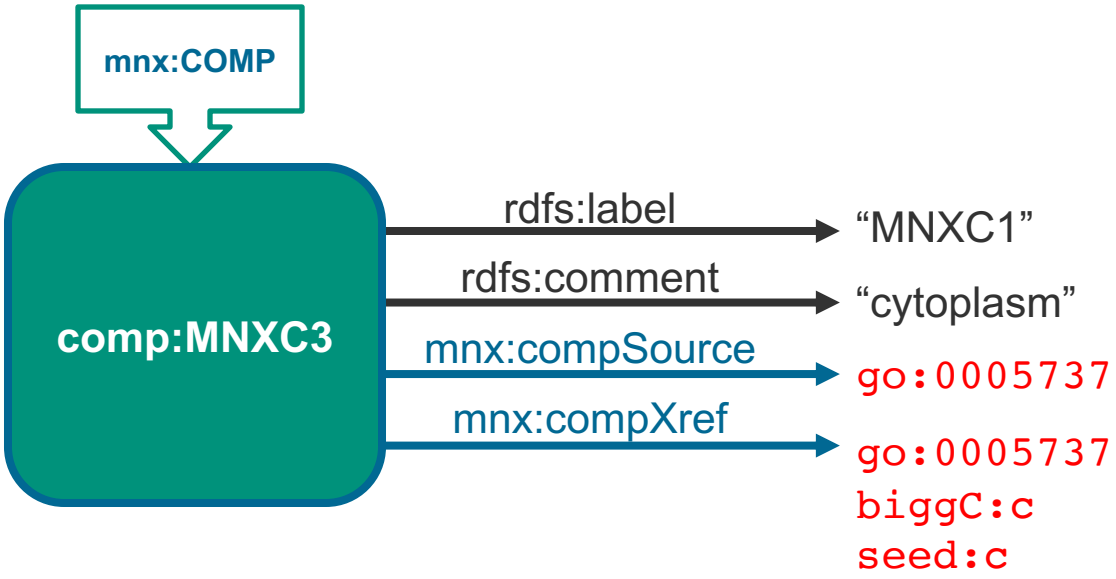
means `?comp rdf:type mnx:COMP`

mnx:COMP - sub-cellular compartment

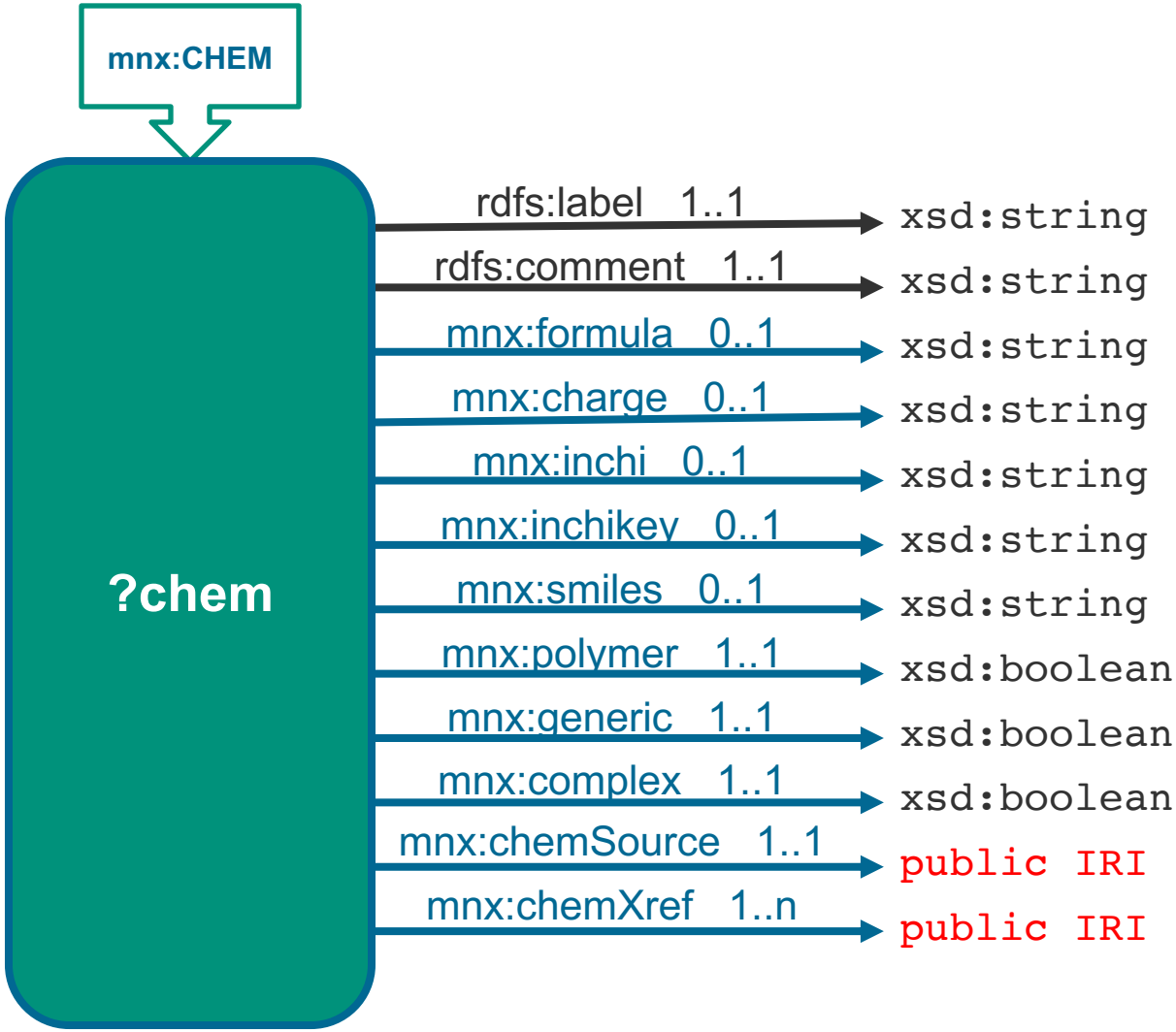


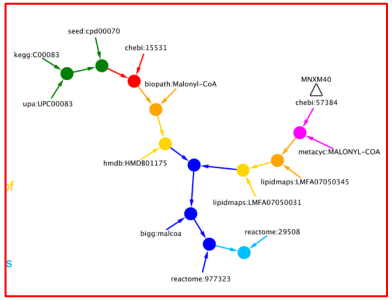
Example of a compartment instance: Cytoplasm

```
@PREFIX mnx:  <https://rdf.metanetx.org/schema/>
@PREFIX comp: <https://rdf.metanetx.org/comp/>
@PREFIX go:   <http://purl.obolibrary.org/obo/GO_>
@PREFIX biggC: <https://identifiers.org/bigg.compartment/>
comp:MNXC3 a mnx:COMP ;
        rdfs:label 'MNXC1' ;
        rdfs:comment 'cytoplasm' ;
        mnx:compSource go:0005737 ;
        mnx:compXref go:0005737 , biggC:c , seed:c .
```



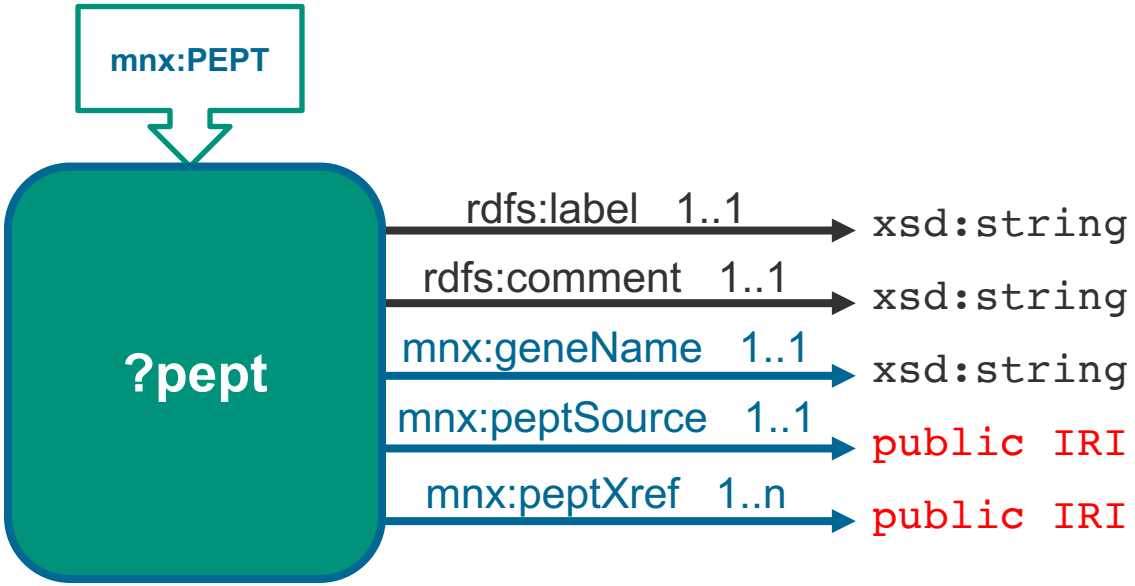
mnx:CHEM: - metabolite





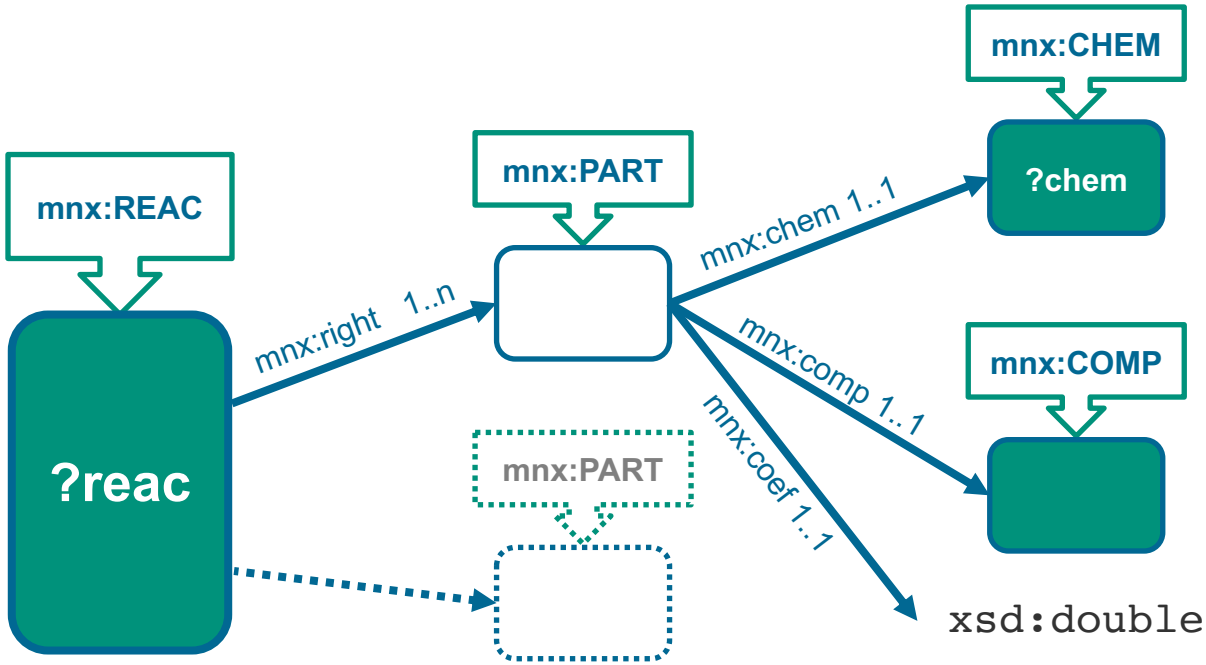
mnx:PEPT - gene or gene product (e.g. polypeptide)

- Most published GSMNs are using gene identifiers with organism-specific nomenclature
- The corresponding UniProt identifiers are recovered at MetaNetX, when possible

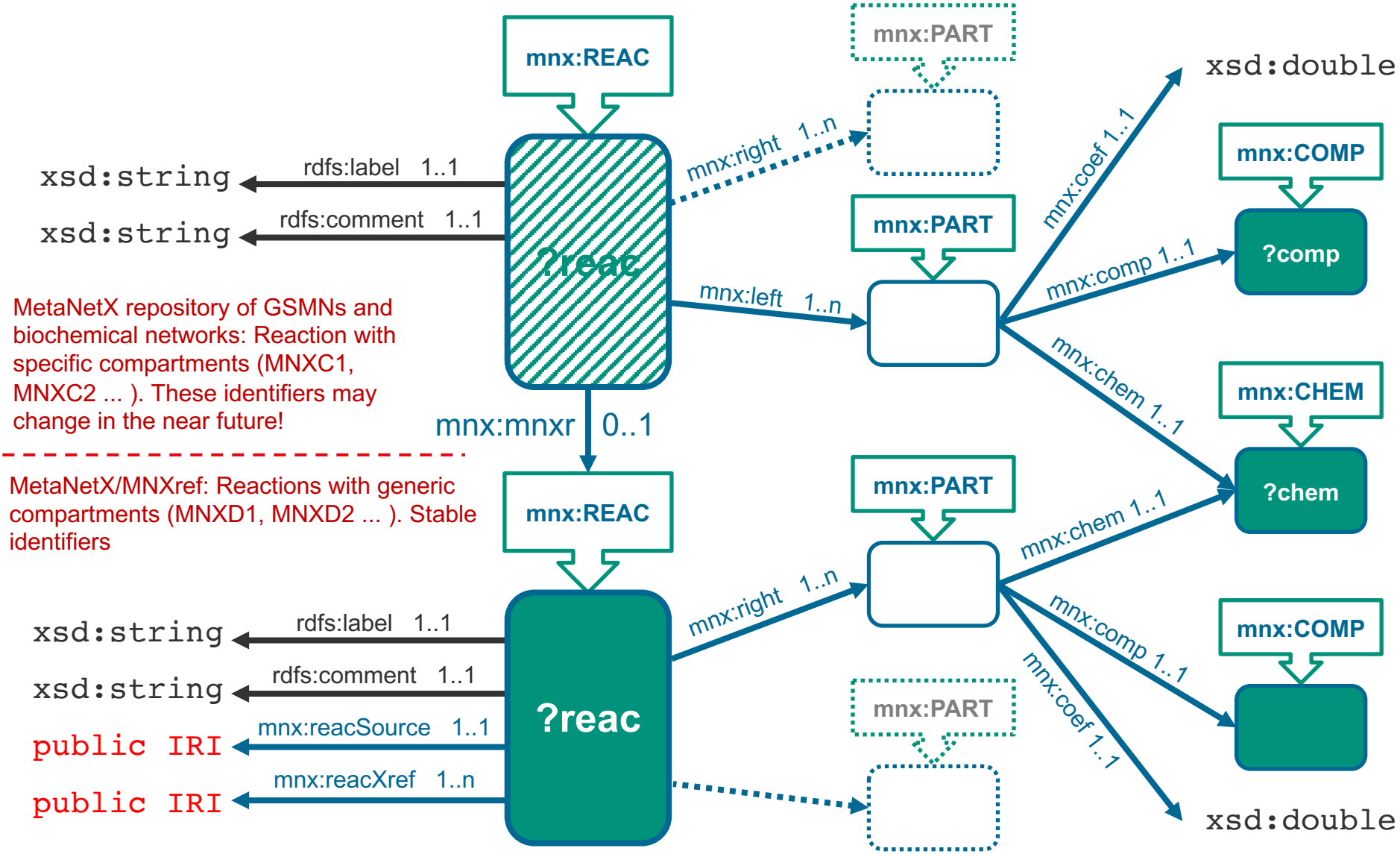


mnx:REAC - reactions

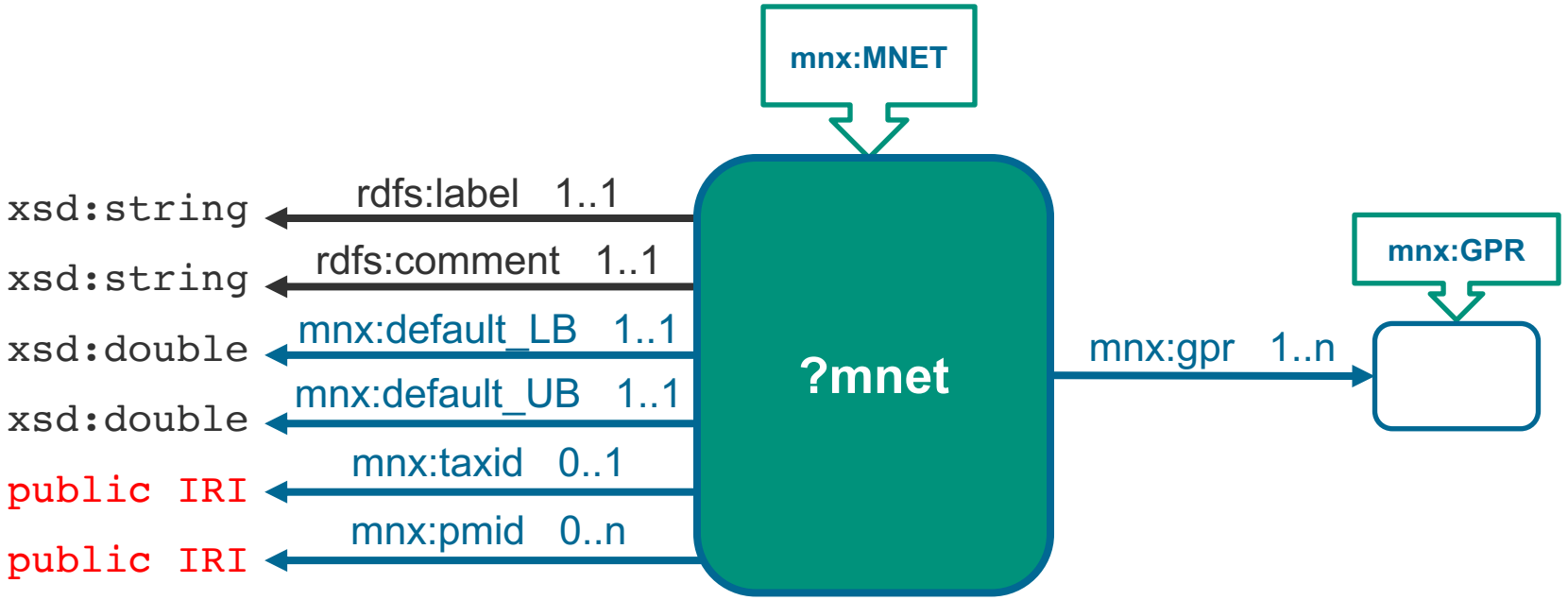
Example here



mnx:REAC - reactions, two flavours



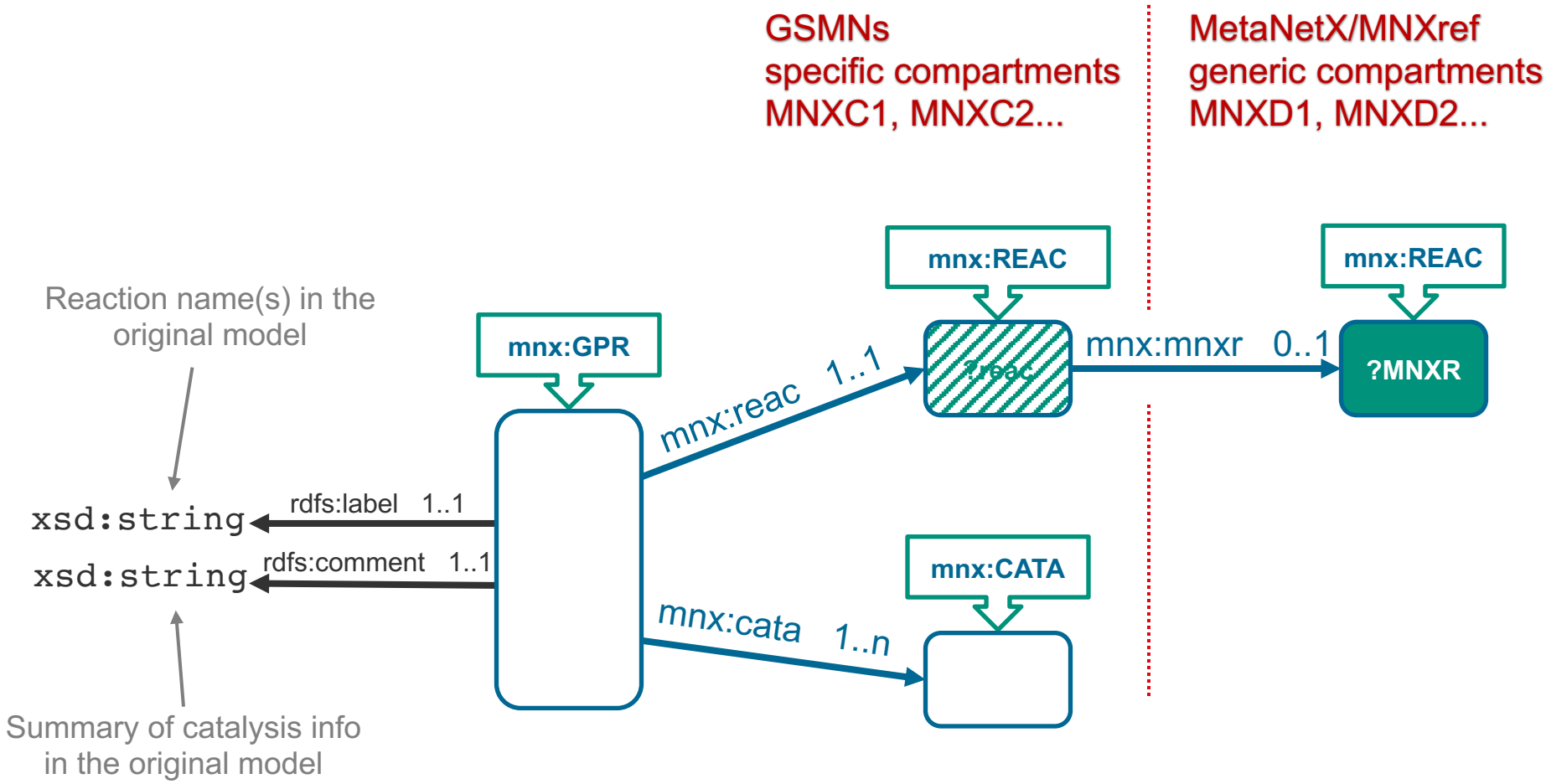
mnx:MNET - GSMN and other metabolic networks



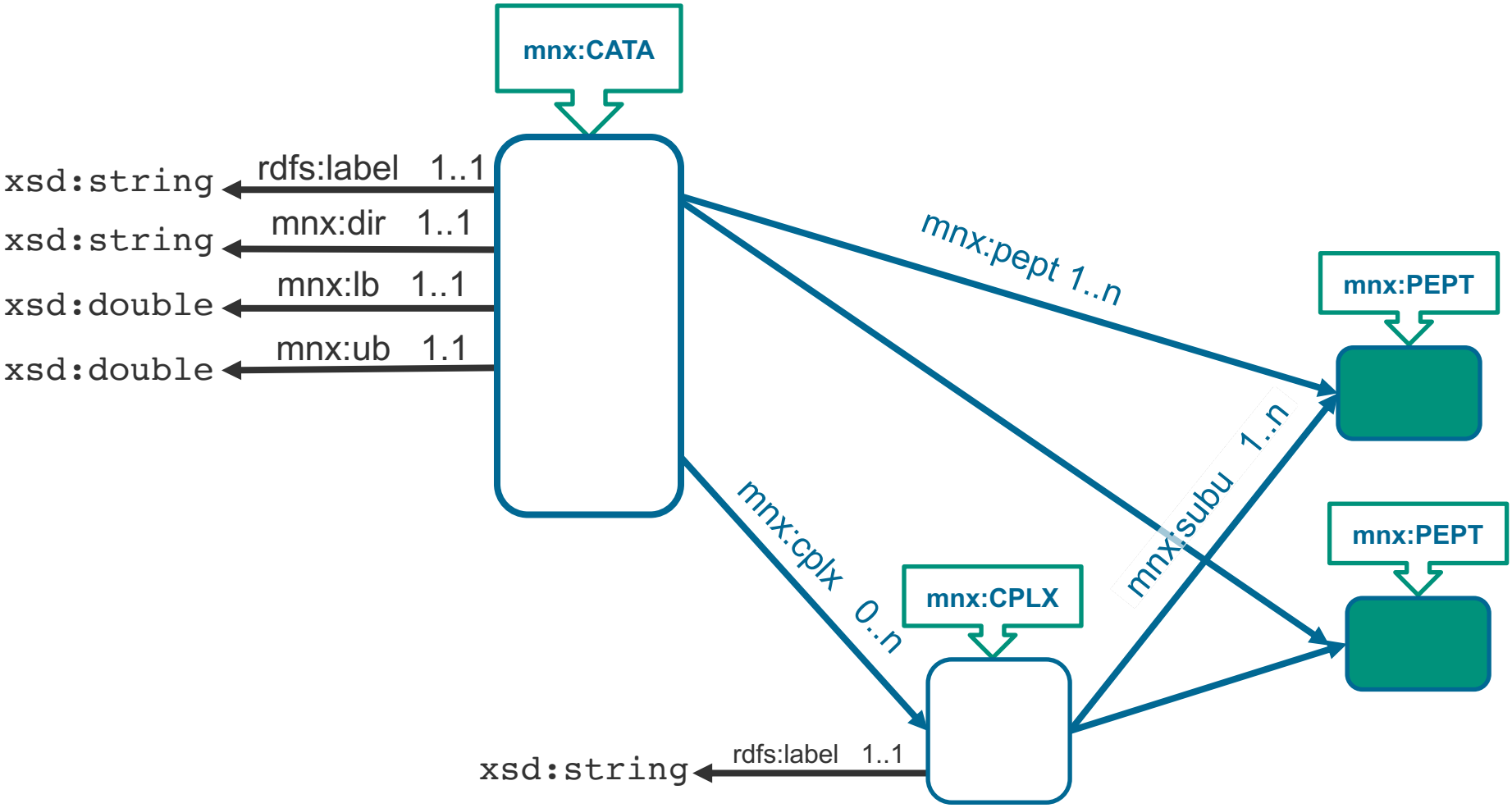
GPR stands for Gene-Protein-Reaction

Large GSMNs contain thousands of GPR

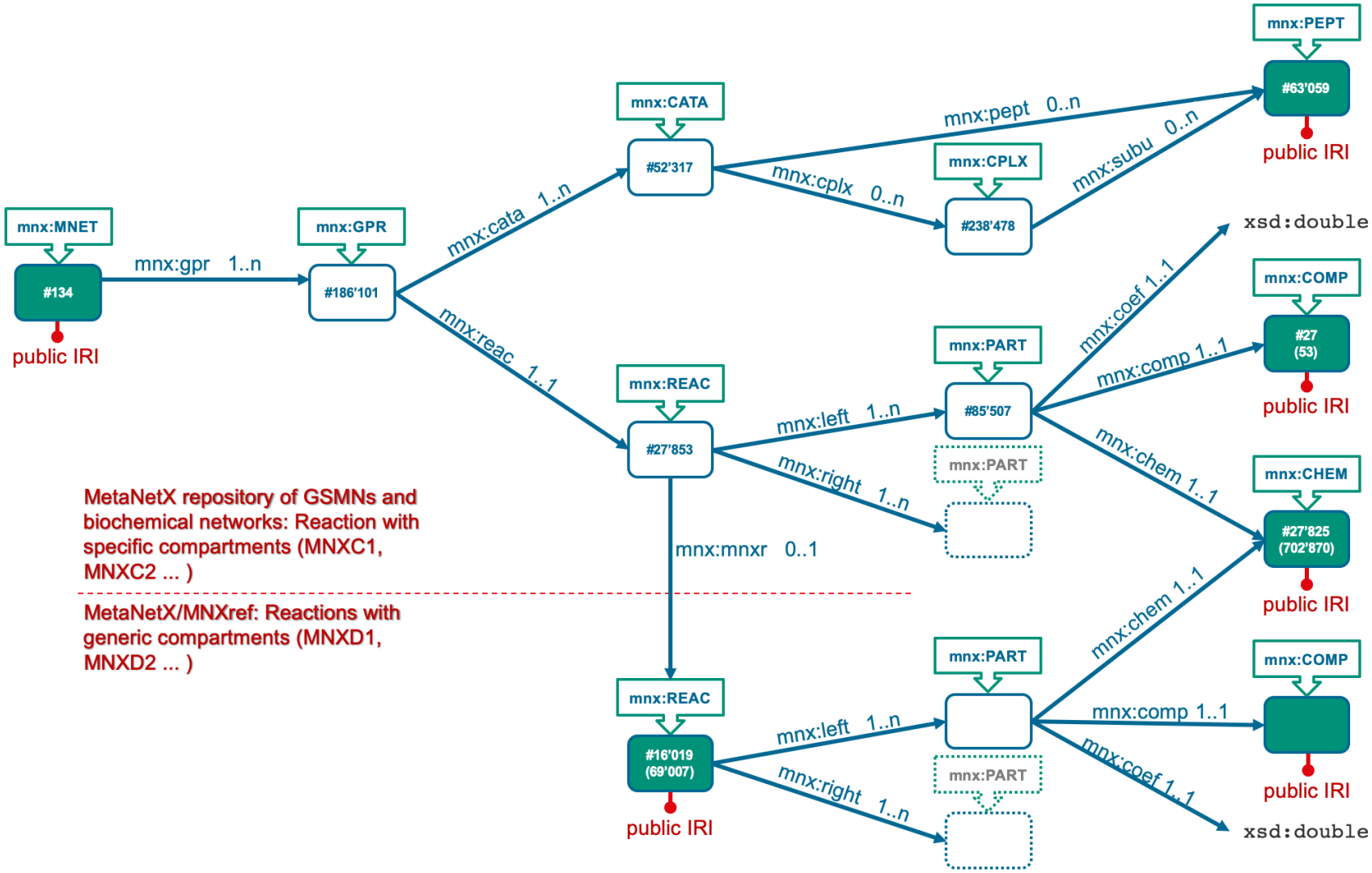
mnx:GPR – GPR are the building block of GSMN



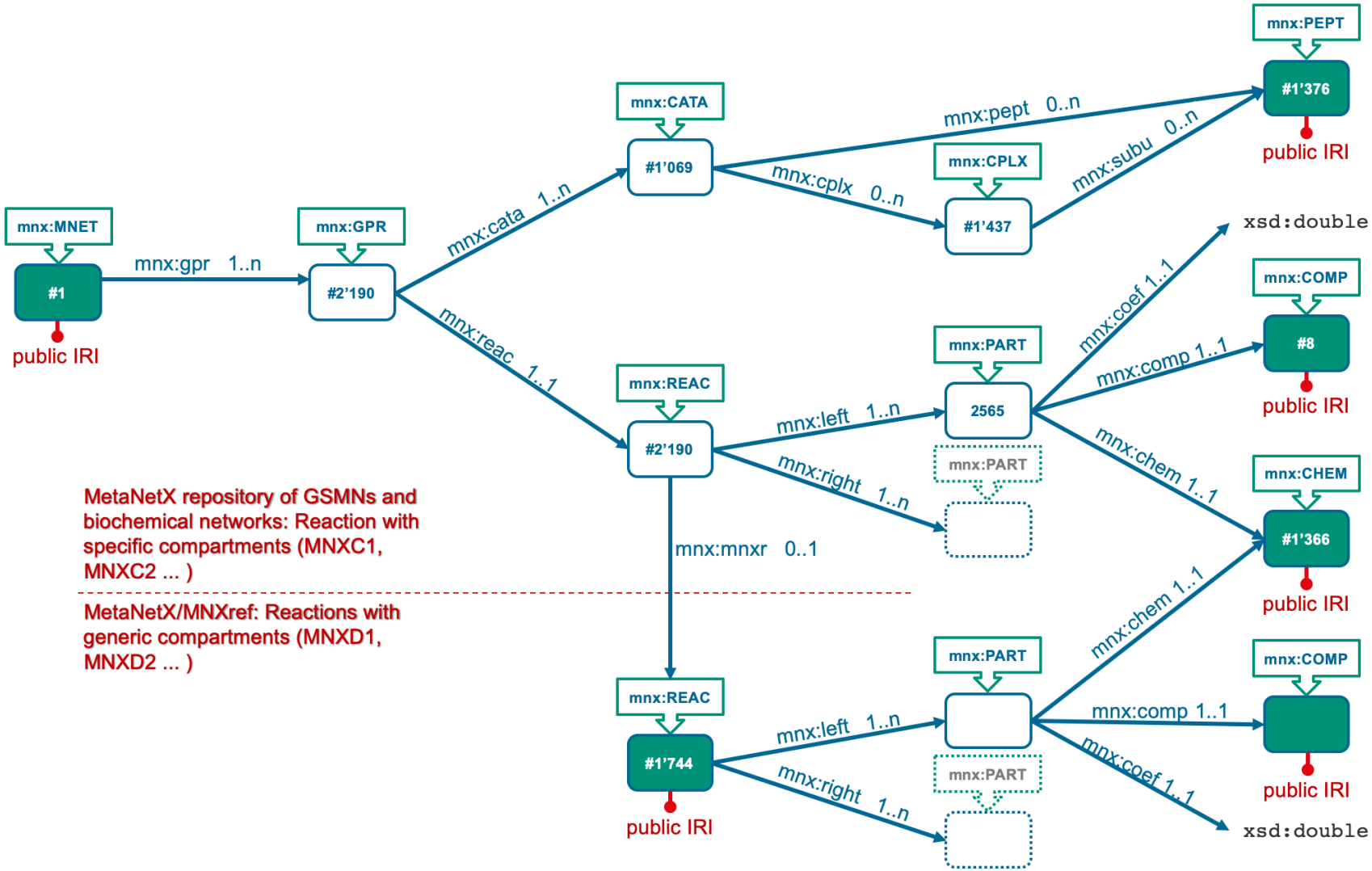
mnx:CATA – catalyst and complex description



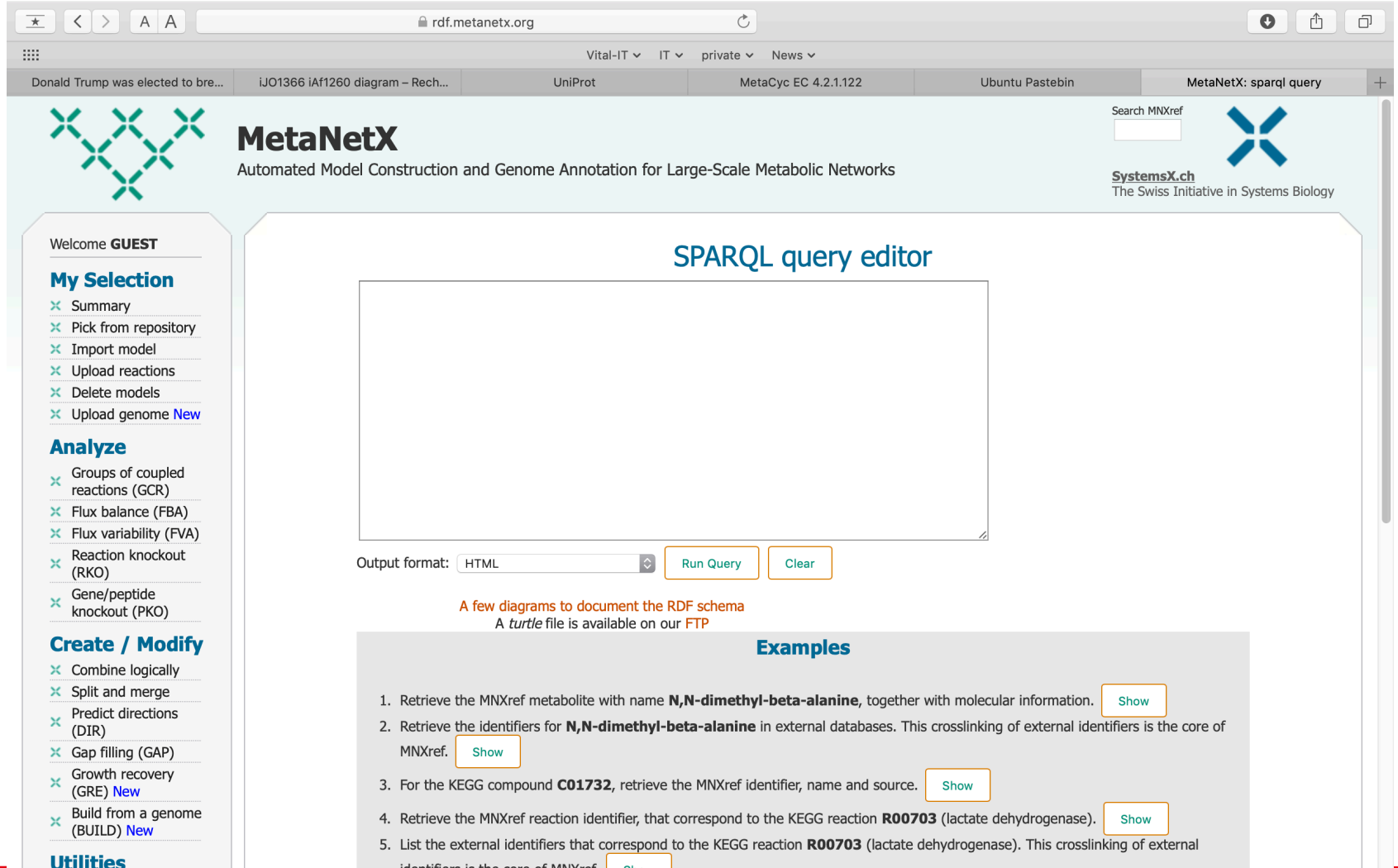
All subunits are required to produce a functional protein complex


mnx:MNET
#134
public IRI
mnx:gpr 1..n
mnx:GPR
#186'101
mnx:cata 1..n
mnx:reac 1..1
mnx:CATA
#52'317
mnx:cplx 0..n
mnx:pept 0..n
mnx:CPLX
#238'478
mnx:subu 0..n
mnx:PEPT
#63'059
public IRI
xsd:double
mnx:REAC
#27'853
mnx:left 1..n
mnx:right 1..n
mnx:PART
#85'507
mnx:PART
mnx:coef 1..1
mnx:comp 1..1
mnx:chem 1..1
mnx:COMP
#27
(53)
public IRI
mnx:CHEM
#27'825
(702'870)
public IRI
mnx:mnxr 0..1
MetaNetX repository of GSMNs and biochemical networks: Reaction with specific compartments (MNXC1, MNXC2 ...)
MetaNetX/MNXref: Reactions with generic compartments (MNXD1, MNXD2 ...)
mnx:REAC
#16'019
(69'007)
public IRI
mnx:left 1..n
mnx:right 1..n
mnx:PART
mnx:PART
mnx:chem 1..1
mnx:comp 1..1
mnx:coef 1..1
mnx:COMP
public IRI
xsd:double

iMM1415


mnx:PEPT
#1'376
public IRI
mnx:pept 0..n
mnx:subu 0..n
mnx:CATA
#1'069
mnx:cata 1..n
mnx:cplx 0..n
mnx:CPLX
#1'437
xsd:double
mnx:MNET
#1
public IRI
mnx:gpr 1..n
mnx:GPR
#2'190
mnx:reac 1..1
mnx:REAC
#2'190
mnx:PART
2565
mnx:coef 1..1
mnx:COMP
#8
public IRI
mnx:comp 1..1
mnx:left 1..n
mnx:right 1..n
mnx:PART
mnx:chem 1..1
mnx:CHEM
#1'366
public IRI
MetaNetX repository of GSMNs and biochemical networks: Reaction with specific compartments (MNXC1, MNXC2 ...)
mnx:mnxr 0..1
MetaNetX/MNXref: Reactions with generic compartments (MNXD1, MNXD2 ...)
mnx:REAC
#1'744
public IRI
mnx:left 1..n
mnx:right 1..n
mnx:PART
mnx:PART
mnx:chem 1..1
mnx:COMP
mnx:comp 1..1
public IRI
mnx:coef 1..1
xsd:double

rdf.metanetx.org



SBML

SBML is the *de facto* standard format to exchange GSMN

- XML based
- Internal non-standard formats are used by most research groups
- RDF is just another format, here

- Our RDF schema was designed to be compatible with SBML representation of a GSMN, *i.e.* no essential information should be lost.

Our complementary editorial policies...

	MetaNetx / MNXref	ChEBI / RHEA / SwissLipids
Evidences from scientific publications	Indirect	Primary goal
Metabolite compartmentalization; Balance vs Transported protons	Enforced (even arbitrarily)	Not systematic (often unknown)
Hypothetical reactions Artificial reactions	Many	None
Choice of metabolite representations	To ensure connectivity and preserve model predictions	To reflect scientific publication
Favored common identifiers	ChEBI + RHEA + UniProt	

- RHEA at the core of the MetaNetX reconciliation as the most favoured resources for reactions
- Importing the latest RHEA data from the SPARQL endpoint turned to be very convenient

Federated query with RHEA/UniProt

Given that

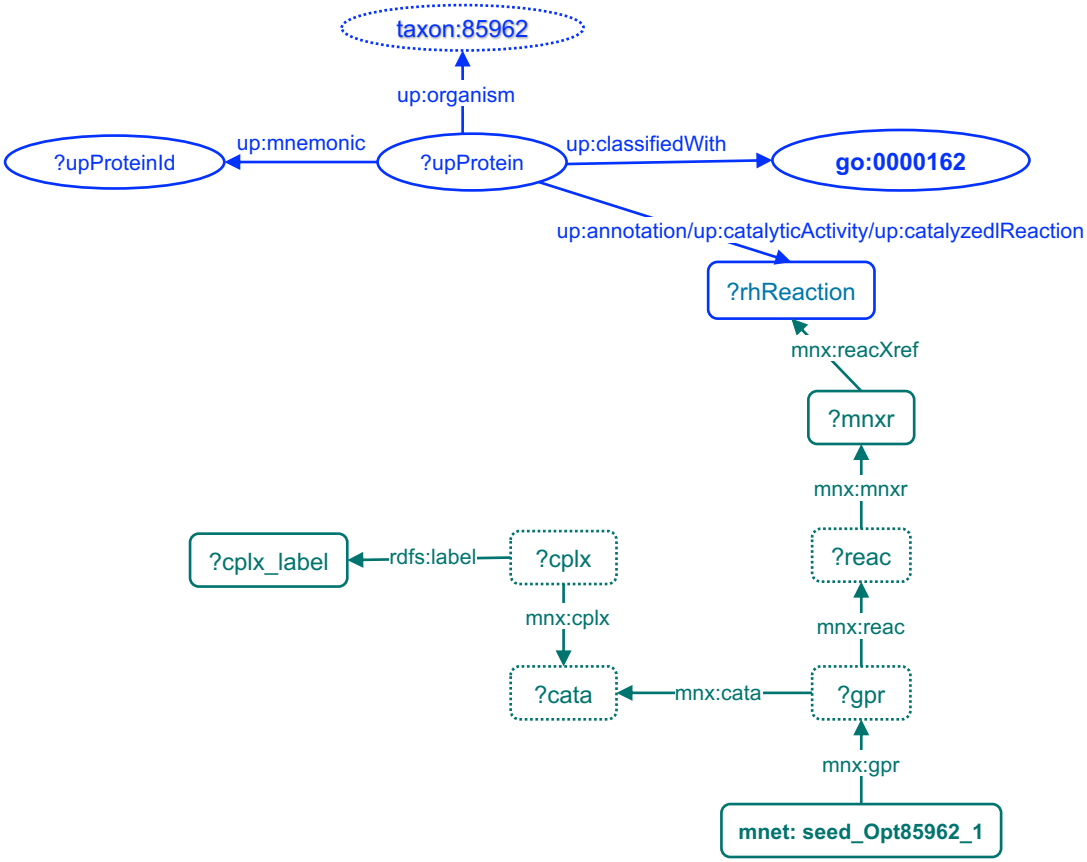
- https://sparql.uniprot.org/sparql provides links between biochemical pathways and RHEA entries
- https://rdf.metanetx.org contains the descriptions of some enzymatic complexes that are missing in UniProt

What is the complexes organisation for a given pathway?

Example: Tryptophan biosynthetic process (GO:0000162) in *Helicobacter pylori* (taxon:85962) as documented in a GSMN from TheSEED (mnet:seed_Opt85962_1)

Problem: there are more than one way to do it!

UNIPROT ⇔ MetaNetX federated query

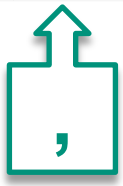


H. pylori enzyme complexes for tryptophan biosynthesis pathway from GSMN seed_Opt85962_1

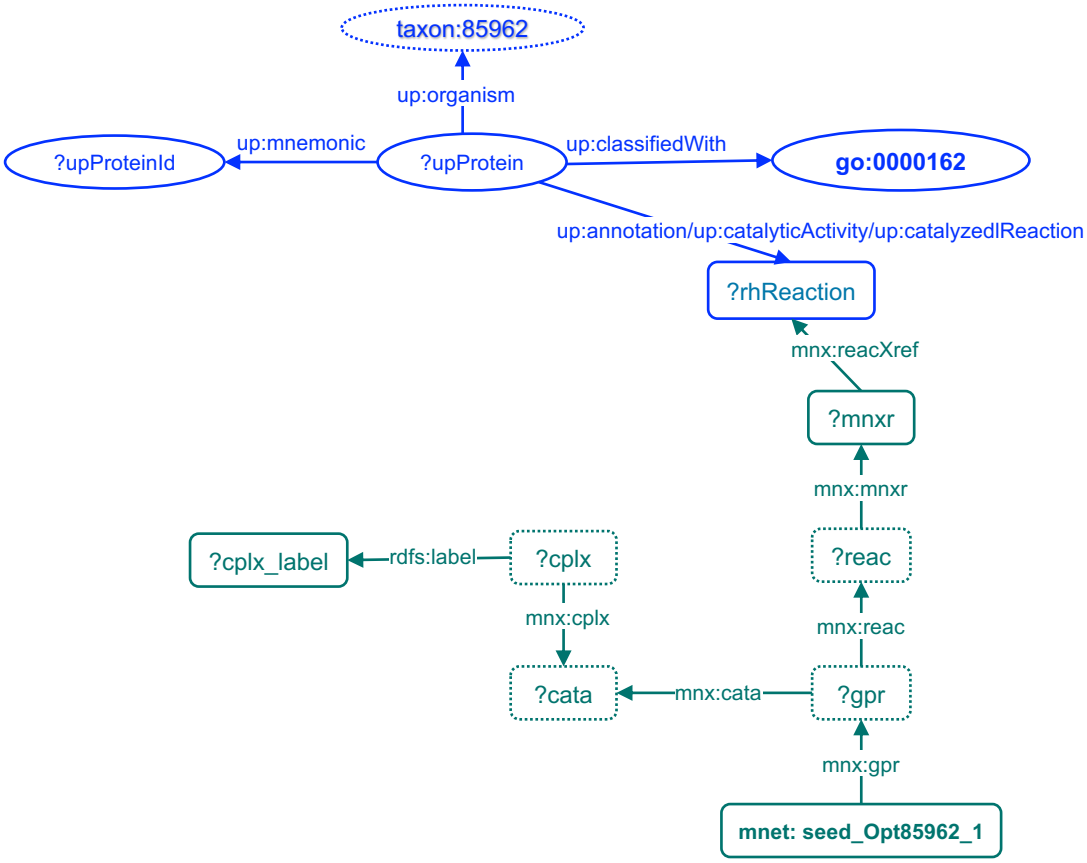
```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX mnx:<https://rdf.metanetx.org/schema/>
PREFIX mnet:<https://rdf.metanetx.org/mnet/>
PREFIX GO:<http://purl.obolibrary.org/obo/GO_>

SELECT
    ( GROUP_CONCAT( DISTINCT ?rhReaction; separator=', ' ) AS ?rh_list )
    ( GROUP_CONCAT( DISTINCT ?upProteinId; separator=', ' ) AS ?id_list )
    ?mnxr
    ( GROUP_CONCAT( DISTINCT ?cplx_label; separator=', ' ) AS ?cplx_list )
          ?mnet
WHERE{
    ?upProtein up:reviewed true .
    ?upProtein up:mnemonic ?upProteinId .
    ?upProtein up:organism taxon:85962 .
    ?upProtein up:classifiedWith GO:0000162.
    ?upProtein up:annotation/up:catalyticActivity/up:catalyzedReaction ?rhReaction .
    SERVICE <https://rdf.metanetx.org/sparql> {
        ?mnxr mnx:reacXref ?rhReaction .
        ?reac mnx:mnxr     ?mnxr       .
        ?gpr  mnx:reac     ?reac       ;
              mnx:cata     ?cata       .
        ?cata mnx:cplx     ?cplx       .
        ?cplx rdfs:label   ?cplx_label .
        ?mnet mnx:gpr      ?gpr        .
        VALUES ?mnet {mnet:seed_Opt85962_1}
    }
}
GROUP BY ?mnxr ?mnet
```

MOST mappings here are not one-to-one

rh_list	id_list	mnxr	cplx_list	mnet
"http://rdf.rhea-db.org/21540"^^xsd:string	"TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR103156	"bact:TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/11768"^^xsd:string	"TRPD_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR95842	"bact:TRPD_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/10532"^^xsd:string	"TRPA_HELPY, TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR104343	"bact:TRPA_HELPY+bact:TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/23476"^^xsd:string	"TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR100814	"bact:TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/21732"^^xsd:string	"TRPE_HELPY, TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR95843	"bact:TRPE_HELPY+bact:TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1

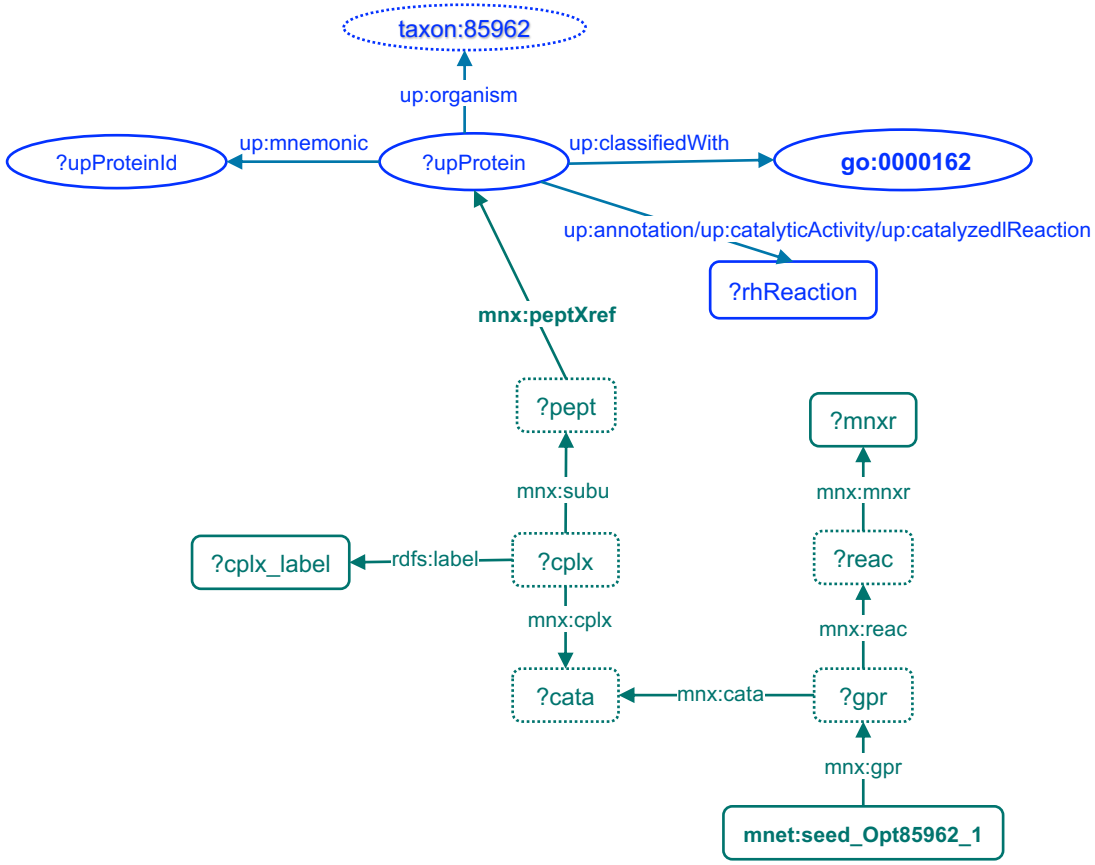


UNIPROT ⇔ MetaNetX federated query



ylori enzyme
nplexes for
tophan

UNIPROT ⇔ MetaNetX federated query



```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX mnx:<https://rdf.metanetx.org/schema/>
PREFIX mnet:<https://rdf.metanetx.org/mnet/>
PREFIX GO:<http://purl.obolibrary.org/obo/GO_>

SELECT
    ( GROUP_CONCAT( DISTINCT ?rhReaction; separator=', ' ) AS ?rh_list )
    ( GROUP_CONCAT( DISTINCT ?upProteinId; separator=', ' ) AS ?id_list )
    ?mnxr
    ( GROUP_CONCAT( DISTINCT ?cplx_label; separator=', ' ) AS ?cplx_list )
          ?mnet
WHERE{
    ?upProtein up:reviewed true .
    ?upProtein up:mnemonic ?upProteinId .
    ?upProtein up:organism taxon:85962 .
    ?upProtein up:classifiedWith GO:0000162.
    ?upProtein up:annotation/up:catalyticActivity/up:catalyzedReaction ?rhReaction .
    SERVICE <https://rdf.metanetx.org/sparql> {
        ?reac mnx:mnxr     ?mnxr       .
        ?gpr  mnx:reac     ?reac       ;
              mnx:cata     ?cata       .
        ?cata mnx:cplx     ?cplx       .
        ?cplx rdfs:label   ?cplx_label ;
              mnx:subu     ?pept       .
        ?pept mnx:peptXref ?upProtein  .
        ?mnet mnx:gpr      ?gpr        .
        VALUES ?mnet {mnet:seed_Opt85962_1}
    }
}
GROUP BY ?mnxr ?mnet
```

rh_list	id_list	mnxr	cplx_list	mnet
"http://rdf.rhea-db.org/21732"^^xsd:string	"TRPE_HELPY, TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR95843	"bact:TRPE_HELPY+bact:TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/21540, http://rdf.rhea-db.org/23476"^^xsd:string	"TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR103156	"bact:TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/10532"^^xsd:string	"TRPA_HELPY, TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR104946	"bact:TRPA_HELPY+bact:TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/21732"^^xsd:string	"TRPE_HELPY, TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR95844	"bact:TRPE_HELPY+bact:TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/21732"^^xsd:string	"TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR95440	"bact:TRPG_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/11768"^^xsd:string	"TRPD_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR95842	"bact:TRPD_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/10532"^^xsd:string	"TRPA_HELPY, TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR104343	"bact:TRPA_HELPY+bact:TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/21540, http://rdf.rhea-db.org/23476"^^xsd:string	"TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR100814	"bact:TRPC_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"http://rdf.rhea-db.org/10532"^^xsd:string	"TRPA_HELPY, TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/reac/MNXR99889	"bact:TRPA_HELPY+bact:TRPB_HELPY"^^xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1

MetaNetX ⇔ OMA

Idea: Produce a GSMN (or at least a set of GPRs) for a target organism, starting from the MetaNetX repository and utilizing orthology/paralogy relationships

Difficulty: with federated query: One way or another, a large number of rows (> 50‘000) has to be exchanged between the two endpoints, causing serious performance issues.

Solutions:

1. Address the faulty/poor performances of endpoints
2. Download the relevant data at once, and work on the client side
3. LOOP with SPARQL queries involoving few rows at a time

Summary

- MetaNetX is a domain-specific resource, the compiled dataset is taylored for particular applications
- A major release (v4.0) is in preparation with lots of bug fixes and improvements
- The RDF-SPARQL exercise helped us to clarify many concepts and provide us with a REST service with clean specifications (among others)
- Integration with other SIB resources, using RDF/SPARQL is ongoing. What will be the maintenance cost?

Swiss Institute of Bioinformatics

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

BioHackathon 2018 in Matsue

SystemsX.ch
The Swiss Initiative in Systems Biology

elixir SWITZERLAND

Acknowledgments

Sebastien MORETTI

Thomas BERNARD
Olivier MARTIN
Fréderic BURDET
Nicolas GUEX
Ioannis XENARIOS
Mark IBBERSON

Alan BRIDGE
Anne MORGAT
Jerven BOLLEMAN

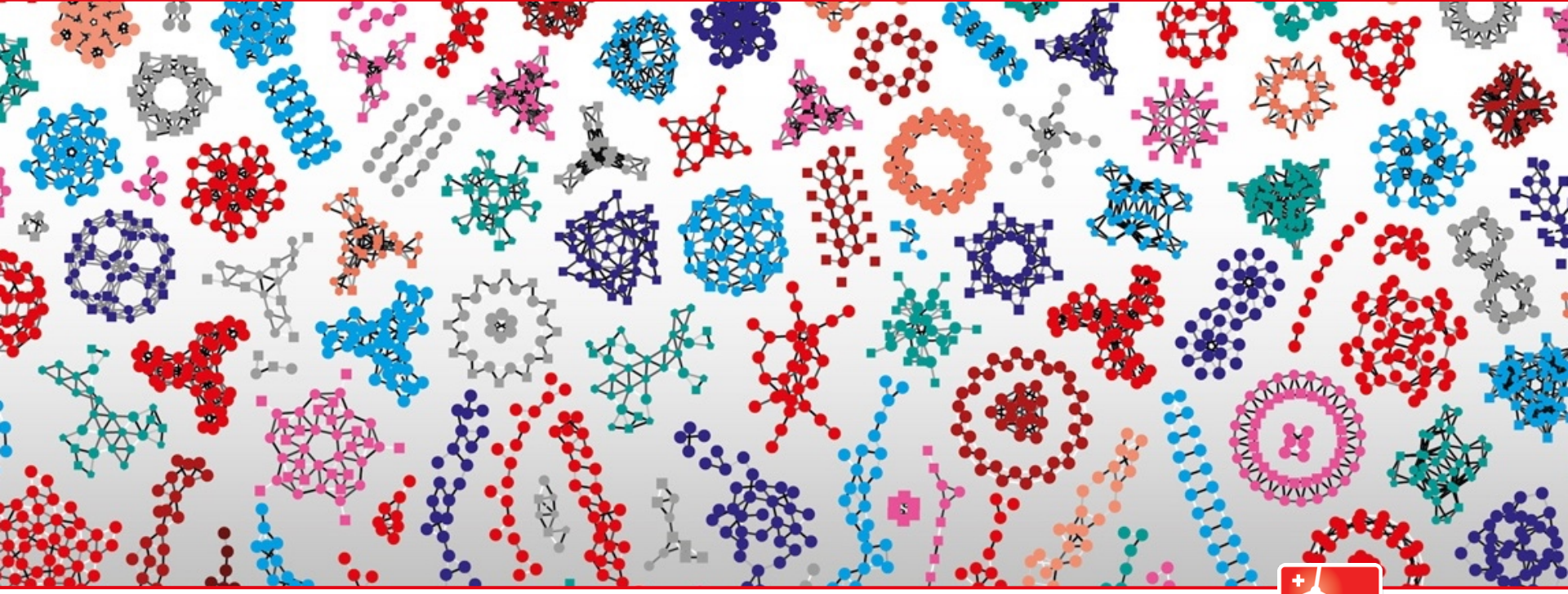
D-BSSE
Department of Biosystems Science and Engineering

Mathias GANTER
Joerg STELLING

UNIL | Université de Lausanne
Béatrice DESVERGNE
Tiziana CAPUTO

UNIL | Université de Lausanne
Sara VAZ
Alix COSTE
Dominique SANGLARD

UNIVERSITÉ DE GENÈVE
Thierry SOLDATI
Louise LEFRANCOIS



SIB
Swiss Institute of Bioinformatics

Thank you