# Semantic Annotation of Solute Carrier (SLC) Transporters

## Introduction

This document contains the annotation specification of the models of Solute Carrier (SLC) transporters and user interface provided for FAIRDOs project. The SLC model is encoded using CellML 2.0 specification.

## Standards and knowledge resources

The annotation approach complies with the OMEX metadata specification (version 1.2) (<https://doi.org/10.1515/jib-2021-0020>) and its suggested convention. The models and annotations are separated. Composite semantic annotation (CSA) is employed and resultant Resource Description Framework (RDF) graph is serialized and saved to Turtle (terse RDF triple language) format.

The namespace of model entities is the CellML model filename, while the namespace of the local entities defined within the RDF annotation is the RDF filename.

The ontology databases and terms are selected according to the SLC tables provided on BioParadigms (<http://slc.bioparadigms.org/>):

* Ontology of Physics for Biology (OPB) (<https://bioportal.bioontology.org/ontologies/OPB/?p=classes&conceptid=root>) is used to annotate physical property component of a CSA,
* UniProt (<https://www.uniprot.org/>) is used to annotate the SLC transporters (proteins),
* Chemical Entities of Biological Interest (ChEBI) (<https://www.ebi.ac.uk/chebi/>) is used to annotate the species that the SLC transports or involved in the transport as mediators,
* Gene Ontology (<https://www.ebi.ac.uk/QuickGO/>) is used to annotate the cellular compartment that the physical entity locates,
* SemSim quantifiers (No reference is found) are used to annotate physical process and its participants, while biology qualifiers (https://github.com/combine-org/combine-specifications/blob/main/specifications/qualifiers.md) are used link physical properties and the bearer of the property (a physical entity, process, energy differential or dependency).

The namespaces and prefixes for Uniform Resource Identifiers (URIs) of the above knowledge resources and standards follow the conventions specified in the OMEX metadata specification.

## Annotation specification

To retrieve the models for future reuse, e.g., compose the models for a particular application, we need to use annotation to capture the biological meaning of what a model simulates. Before we introduce the proposed annotation, we want to highlight what knowledge we want to capture to facilitate the future model discovery.

### Solute Carrier (SLC) transporters and relevant biological processes

Table 1 listed a family of glucose transporters and some relevant biological information, and the detailed knowledge of each transporter can be referred to the UniProt database, e.g., <https://www.uniprot.org/uniprotkb/P11166/entry#function>.

From Table 1 we can see:

1. Facilitative glucose transporter family has 14 proteins corresponding to 14 UniProt entries
2. Each transporter protein may be located in multiple cellular components, such as, plasma membrane (GO:0005886), cytoplasmic vesicle membrane (GO:0030659).
3. Each transporter protein may be involved multiple biological processes (chemical transport), e.g., glucose transmembrane transport, fructose transmembrane transport, dehydroascorbic acid transport.
4. The movement of species can be from extracellular space (GO: 0005615) to cytosol (GO: 0005829), from cytosol (GO: 0005829) to extracellular space (GO: 0005615), from cytosol (GO: 0005829) to cytoplasmic vesicle lumen (GO:0060205), etc.

If we want to retrieve a model simulating the transport of D-glucose (CHEBI:4167) from extracellular space (GO: 0005615) to cytosol (GO: 0005829) mediated by facilitated glucose transporter member 1 (UniProt: P11166) located in plasma membrane (GO:0005886), we need to annotate the model with all the relevant information.

### A bond graph model of facilitative glucose transporters

Now we look at the model of glucose transport. The bond graph template shown in Figure 1 can be used to model the biological process involved by majority of glucose transporters listed in Table 1.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 1 A bond graph template

The bond graph model has variables

1. *q* representing the molar amount of species involved in the transport process,
2. *u* representing the chemical potentials of the species,
3. *v* representing theflow rate of the transport, which is a property of the biological process

In terms of steady state of the above model, it only includes the variables *q* and *v*.

The template is generic, which can be used to model a variety of biological processes with proper parameters, such as

1. the transport of D-glucose (CHEBI:4167) from extracellular space (GO: 0005615) to cytosol (GO: 0005829) mediated by facilitated glucose transporter member 1 (UniProt: P11166) located in plasma membrane (GO:0005886), or
2. the transport of D-fructose (CHEBI:37721) from cytosol (GO: 0005829) to cytoplasmic vesicle lumen (GO:0060205) mediated by facilitated glucose transporter member 8 (UniProt: Q9NY64) located in cytoplasmic vesicle membrane (GO:0030659).

To retrieve the models for future reuse, e.g., compose the models for a particular application, we need to use annotation to capture the biological meaning of what a model simulates. The more information the annotation encodes, the more accurate the model discovery result will be. However, we want a trade-of between the annotation effort and the model retrieval performance. We now introduce the composite annotation construction for the steady state CellML model of SLC transporters.

### Composite annotation construction

Computational models usually capture the quantitative properties of physical entities instead of physical entities themselves. Hence composite annotation is needed to link the annotations.

The annotated variables in the CellML model include:

1. the flow rate of the transport, annotated with opb: OPB\_00592 (Chemical amount flow rate),
2. the molar amount of SLC transporter protein, annotated with opb: OPB\_00425 (Chemical molar amount),
3. the molar amount of reactants, annotated with opb: OPB\_00425 (Chemical molar amount),
4. the molar amount of products, annotated with opb: OPB\_00425 (Chemical molar amount).

The local entities that needed to link the annotations are:

1. Process, local entity representing the biological process, whereas flow rate is property of the process,
2. SourceParticipant, corresponding to reactants, the entity name is in form of *source\_i*, where *i* is the id of the source if multiple source participants are present,
3. SinkParticipant, corresponding to products, the entity name is in form of *sink\_i*, where *i* is the id of the sink if multiple sink participants are present,
4. MediatorParticipant, corresponding to the transporter that mediates the transport process.

The SourceParticipant and SinkParticipant are annotated with:

1. chemical entities, such as D-glucose (CHEBI:4167), K+ (CHEBI:29103), Na+ (CHEBI:29101),
2. subcellular location, such as extracellular space (GO: 0005615), cytosol (GO: 0005829), cytoplasmic vesicle lumen (GO:0060205),
3. stoichiometric coefficient (using the semsim:hasMultiplier qualifier and Literal (1.0, datatype=XSD.double) ).

The MediatorParticipant is annotated with:

1. Protein entities, such as Excitatory amino acid transporter 3 (UniProt: P43005), Electrogenic sodium bicarbonate cotransporter 1 (UniProt: Q9Y6R1),
2. subcellular location, such as plasma membrane (GO:0005886), cytoplasmic vesicle membrane (GO:0030659)

### RDF example using Turtle syntax

The CellML model and rdf file can be found here: https://models.physiomeproject.org/workspace/aec/file/d5be1c99bcfb01db284a91ae7ca261336c921ced/SLC.

### RDF graph visualization

The visualized RDF graph is shown in Figure 2.

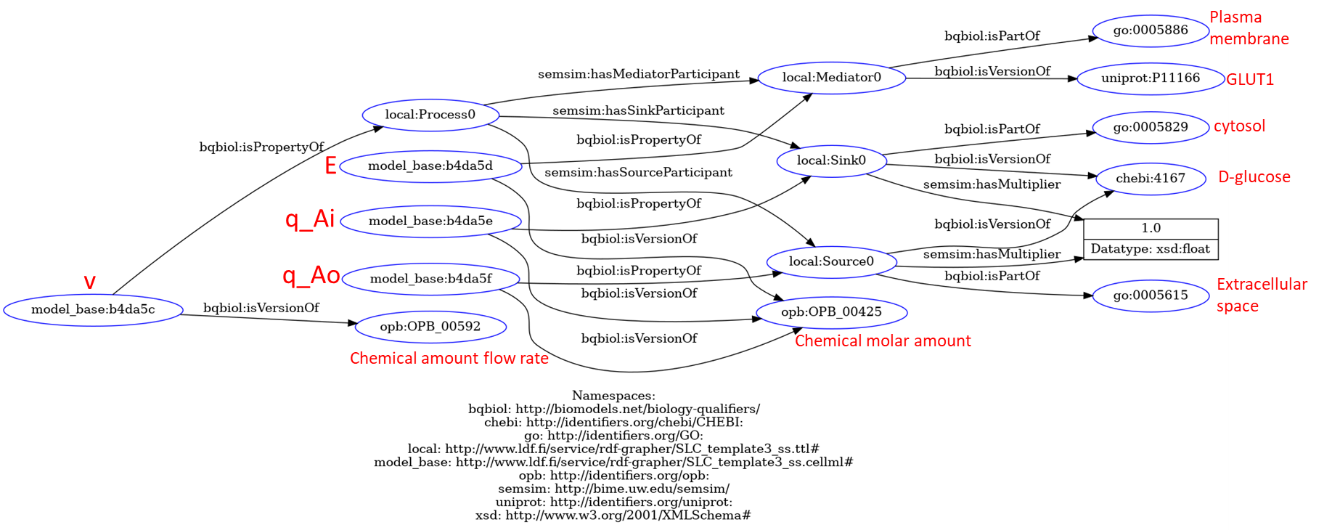


Figure 2 RDF graph

## User inputs (specific to the students enrolled in the FAIRDOs project)

To construct the RDF in Figure 2, we need the inputs:

1. which variable is the flow rate of the transport
2. which variables are the molar amount of reactants, products and transporter proteins
3. the species of the reactants and products
4. which subcellular locations where the species are
5. the name of the transporter and its subcellular location

The inputs 1 and 2 could be given by the prior knowledge of the template, while the inputs 3, 4 and 5 are from the students because they are specific to the model and experimental data that the student want to fit.

## Appendix

Table 1 Facilitative GLUT transporter family and relevant ontology terms

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gene | Protein, location | UniProt entry | Catalytic activity, Participant ID, locations | CellML models |
| SLC2A1 | Solute carrier family 2, facilitated glucose transporter member 1, [plasma membrane (GO:0005886)](https://www.ebi.ac.uk/QuickGO/term/GO:0005886) | P11166 | D-glucose(out) = D-glucose(in), [CHEBI:4167](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=4167)  extracellular space (GO: 0005615) 🡪cytosol (GO: 0005829) | SLC2A1.cellml |
| SLC2A2 | Solute carrier family 2, facilitated glucose transporter member 2, [plasma membrane (GO:0005886)](https://www.ebi.ac.uk/QuickGO/term/GO:0005886) | P11168 | D-glucose(out) = D-glucose(in), [CHEBI:4167](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=4167)  D-fructose(out) = D-fructose(in), [CHEBI:37721](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=37721)  L-dehydroascorbate(out) = L-dehydroascorbate(in), [CHEBI:58539](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=58539)  D-galactose(in) = D-galactose(out), [CHEBI:4139](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=4139)  cytosol (GO: 0005829) 🡪 extracellular space (GO: 0005615) | SLC2A2\_1.cellml  SLC2A2\_2.cellml  SLC2A2\_3.cellml  SLC2A2\_4.cellml |
| SLC2A3 | Solute carrier family 2, facilitated glucose transporter member 3, [plasma membrane (GO:0005886)](https://www.ebi.ac.uk/QuickGO/term/GO:0005886) | P11169 | D-glucose(out) = D-glucose(in), [CHEBI:4167](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=4167)  extracellular space (GO: 0005615) 🡪cytosol (GO: 0005829)  D-galactose(in) = D-galactose(out), [CHEBI:4139](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=4139)  cytosol (GO: 0005829) 🡪 extracellular space (GO: 0005615) |  |
| SLC2A4 | Solute carrier family 2, facilitated glucose transporter member 4,   * [plasma membrane (GO:0005886)](https://www.ebi.ac.uk/QuickGO/term/GO:0005886) * [perinuclear region of cytoplasm (GO:0048471)](https://www.ebi.ac.uk/QuickGO/term/GO:0048471) | P14672 | D-glucose(out) = D-glucose(in), [CHEBI:4167](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=4167)  extracellular space (GO: 0005615) 🡪cytosol (GO: 0005829) |  |
| SLC2A5 | Solute carrier family 2, facilitated glucose transporter member 5,   * [apical plasma membrane (GO:0016324)](https://www.ebi.ac.uk/QuickGO/term/GO:0016324) * [plasma membrane (GO:0005886)](https://www.ebi.ac.uk/QuickGO/term/GO:0005886) * [sarcolemma (GO:0042383)](https://www.ebi.ac.uk/QuickGO/term/GO:0042383) | P22732 | D-fructose(out) = D-fructose(in) |  |
| SLC2A6 | Solute carrier family 2, facilitated glucose transporter member 6 | Q9UGQ3 | Probable sugar transporter that acts as a regulator of glycolysis in macrophages (Probable).  Was initially thought to act as a glucose transporter (PubMed:[10970791](https://www.uniprot.org/citations/10970791)). However, later studies demonstrated that it does not transport glucose (PubMed:[30431159](https://www.uniprot.org/citations/30431159)).2 Publications  Was named GLUT9 by a report, but this gene name has already been used for SLC2A9. |  |
| SLC2A7 | Solute carrier family 2, facilitated glucose transporter member 7 | Q6PXP3 | D-glucose(out) = D-glucose(in)  Probable sugar transporter  According to some reports, mediates transmembrane transport of glucose and fructose (PubMed:[15033637](https://www.uniprot.org/citations/15033637), PubMed:[16186102](https://www.uniprot.org/citations/16186102), PubMed:[29548810](https://www.uniprot.org/citations/29548810)). However, another group could not confirm transporter activity for glucose or fructose (PubMed:[28083649](https://www.uniprot.org/citations/28083649)). |  |
| SLC2A8 | Solute carrier family 2, facilitated glucose transporter member 8  [cytoplasmic vesicle membrane (GO:0030659)](https://www.ebi.ac.uk/QuickGO/term/GO:0030659) | Q9NY64 | D-glucose(out) = D-glucose(in)  D-fructose(out) = D-fructose(in)  L-dehydroascorbate(out) = L-dehydroascorbate(in) cytoplasmic vesicle lumen GO:0060205  cytosol (GO: 0005829) |  |
| SLC2A9 | Solute carrier family 2, facilitated glucose transporter member 9   * [plasma membrane (GO:0005886)](https://www.ebi.ac.uk/QuickGO/term/GO:0005886) * [apical plasma membrane (GO:0016324)](https://www.ebi.ac.uk/QuickGO/term/GO:0016324) * [basolateral plasma membrane (GO:0016323)](https://www.ebi.ac.uk/QuickGO/term/GO:0016323) |  | urate(out) = urate(in)  CHEBI:17775  High-capacity urate transporter that was first described as a fructose and glucose transporter. Also described in the literature as high-affinity and low-capacity glucose and fructose transporter (PubMed:[18327257](https://www.uniprot.org/citations/18327257), PubMed:[17710649](https://www.uniprot.org/citations/17710649), PubMed:[18842065](https://www.uniprot.org/citations/18842065)). However, another group could not confirm transporter activity for glucose or fructose (PubMed:[28083649](https://www.uniprot.org/citations/28083649)). |  |
| SLC2A10 | Solute carrier family 2, facilitated glucose transporter member 10,   * perinuclear endoplasmic reticulum membrane, GO:1990578 | O95528 | D-glucose(out) = D-glucose(in) perinuclear endoplasmic reticulum lumen, GO:0099020 |  |
| SLC2A11 | Solute carrier family 2, facilitated glucose transporter member 11 | Q9BYW1 | Has been described as GLUT10 in literature, but this gene name has already been used for SLC2A10. |  |
| SLC2A12 | Solute carrier family 2, facilitated glucose transporter member 12 | Q8TD20 | D-glucose(out) = D-glucose(in) |  |
| SLC2A13 | Proton myo-inositol cotransporter | Q96QE2 | H+(out) + myo-inositol(out) = H+(in) + myo-inositol(in) |  |
| SLC2A14 | Solute carrier family 2, facilitated glucose transporter member 14 | Q8TDB8 | D-glucose(out) = D-glucose(in)  L-dehydroascorbate(out) = L-dehydroascorbate(in)  GLUT14 is a recent (less than 5 M year old) duplication of GLUT3. |  |