

Epithelial Modelling Platform: A Tool for Model Discovery and Assembly with the Physiome Model Repository

Dewan M. Sarwar, Yuda Munarko and David P. Nickerson

Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

ABSTRACT

In this paper we present a web-based platform enabling scientists to construct a novel epithelial transport model to investigate their hypotheses, aided by building on existing models discovered in the Physiome Model Repository (PMR). We have comprehensively annotated a cohort of epithelial transport models deposited in the PMR as a seeding collection of building blocks that are freely available for reuse. On the platform, users are able to semantically display models for visualization, graphical editing, and model assembly. In addition, we leverage web services from the European Bioinformatics Institute (EBI) to help rank similar models based on the suggestions provided by the platform. In addition to potential use in biomedical and clinical research, novice modellers could use our platform as a learning tool. The source code and links to a live demonstration of the platform are available at <https://github.com/dewancse/epithelial-modelling-platform>.

Keywords: CellML, Physiome Model Repository, Semantic annotation, Model discovery, Model assembly, Epithelial transport

1 INTRODUCTION

The IUPS Physiome Project (Hunter and Borg, 2003) and the Virtual Physiological Human (VPH) (Hunter et al., 2010) have introduced tools and open source software for scientists to conduct biomedical research (Garny et al., 2010; Cooper et al., 2010). For example, biomedical engineers and clinicians often utilize computational models to investigate experimental or clinical hypotheses that are difficult or expensive to achieve experimentally using animal or human subjects. Modern tools and technologies help them to quickly and precisely test such hypotheses which may involve disparate biophysical symptoms and/or mechanisms such as clinical observations, diseases, and drug actions through physiology pathways (de Bono et al., 2016).

In this paper, we present a web-based platform, the Epithelial Modelling Platform (EMP), which will enable biomedical engineers and clinicians to discover epithelial transport models with a view to constructing a new epithelial model. This new model will help them to investigate specific research questions and hypotheses. To prepare a model discovery environment, mathematics- and physics-based epithelial transport models are integrated with biological information. Integration of such information in the context of biology is called model annotation. Model annotation captures semantic meanings of biological phenomena encapsulated in mathematics and thus hides underlying mathematical equations (Lister et al., 2009). In order to represent biological phenomena, model annotation makes use of a rich set of controlled vocabulary, i.e. ontologies (Courtot et al., 2011). This often requires multiple reference ontologies to represent such phenomena, which is called composite annotation (Gennari et al., 2011). Model annotation is a time-consuming process and thus requires automated process. For this, SemGen has been developed to annotate biosimulation models (Neal et al., 2018). We have deposited the annotated

information and the models in the Physiome Model Repository (PMR) (Yu et al., 2011) where information is discoverable, accessible, and persistent. PMR follows FAIR (Wilkinson and et al., 2016) principles; i.e. information in PMR is Findable, Accessible, Interoperable, and Reusable (FAIR).

Prior research in (Galdzicki et al., 2011; Neal et al., 2013) demonstrates that model discovery can be performed by utilizing semantic querying technologies. Similarly, with the provision of modern technologies, knowledge discovery has introduced a range of tools and open source software such as DisGeNET (Piñero et al., 2015), ApiNATOMY (de Bono et al., 2014), JWS Online (Wimalaratne et al., 2004), Saint (Lister et al., 2009), semanticSBML (Krause et al., 2010), eSolve (de Boer et al., 2017), Chalkboard (Cook et al., 2007), Physiomaps (Cook et al., 2013). A recent initiative in (Cooper et al., 2015a), the Cardiac Electrophysiology Web Lab, has introduced an online simulation environment to compare simulation experiments. Such virtual experiments are analogous to *in silico* environment of *in vivo* and *in vitro* experiments, or wet lab experiments (Cooper et al., 2015b).

In addition to model discovery, we have assembled the discovered models on the platform with a view to constructing a new epithelial model. Model construction integrates components from various source models. SemGen can compose models in CellML (Cuellar et al., 2003) and SBML (Hucka et al., 2003) format. CellML and SBML are XML-based languages to describe biological processes. Specifically, SemGen internally converts models into OWL format and then does model composition (Hoehndorf et al., 2011). In contrast, we have composed models by utilizing libCellML (libCellML, 2019). libCellML is a library to serialize, validate, and instantiate a CellML model. A great deal of previous research on model composition is investigated in (Neal et al., 2015; Gennari et al., 2008; Neal et al., 2009; Krause et al., 2010; Coskun et al., 2013).

Here we present a model discovery and assembly approach which is a new contribution to PMR. In addition to potential use in biomedical and clinical research, novice modellers could use our platform as a learning tool. This platform would demonstrate how semantic web technologies and methodologies can contribute to educational purpose. In the section Materials and Methods, we present our current workflow of the EMP, the standards and software tools we have used to encode the computational models and semantic annotations for our exemplar cohort of models, and the technologies that enable the implementation of our platform. This platform, which makes use of the encoded models and semantic annotations to discover, explore, and visualize models, is discussed in the Results section. The discussion section summarizes the model discovery and assembly of the EMP with some concluding remarks.

2 MATERIALS AND METHODS

Fig. 1 presents an overview of the knowledge management platform being developed and utilized in this work. On the client side, content creation and editing applications like OpenCOR (Garny and Hunter, 2015) are used to encode computational models and associated information in computable formats. Such applications work on local workspaces, which can be synchronized with cloud-based repositories such as the PMR. Synchronization with PMR ensures persistent, resolvable, and discoverable identifiers for all resources in the workspace. For composite annotation, we used the SemGen tool and deposited this annotation as metadata in PMR (Sarwar et al., 2018). Web-based applications such as our EMP are then able to leverage the knowledge stored in PMR to discover models, and semantically assemble these models on the platform for visualization, graphical editing, and model assembly and construction.

Our EMP has become a key demonstrator application of the knowledge management system. The aim was to discover models from the PMR and assemble these on the platform with the provision of the EBI

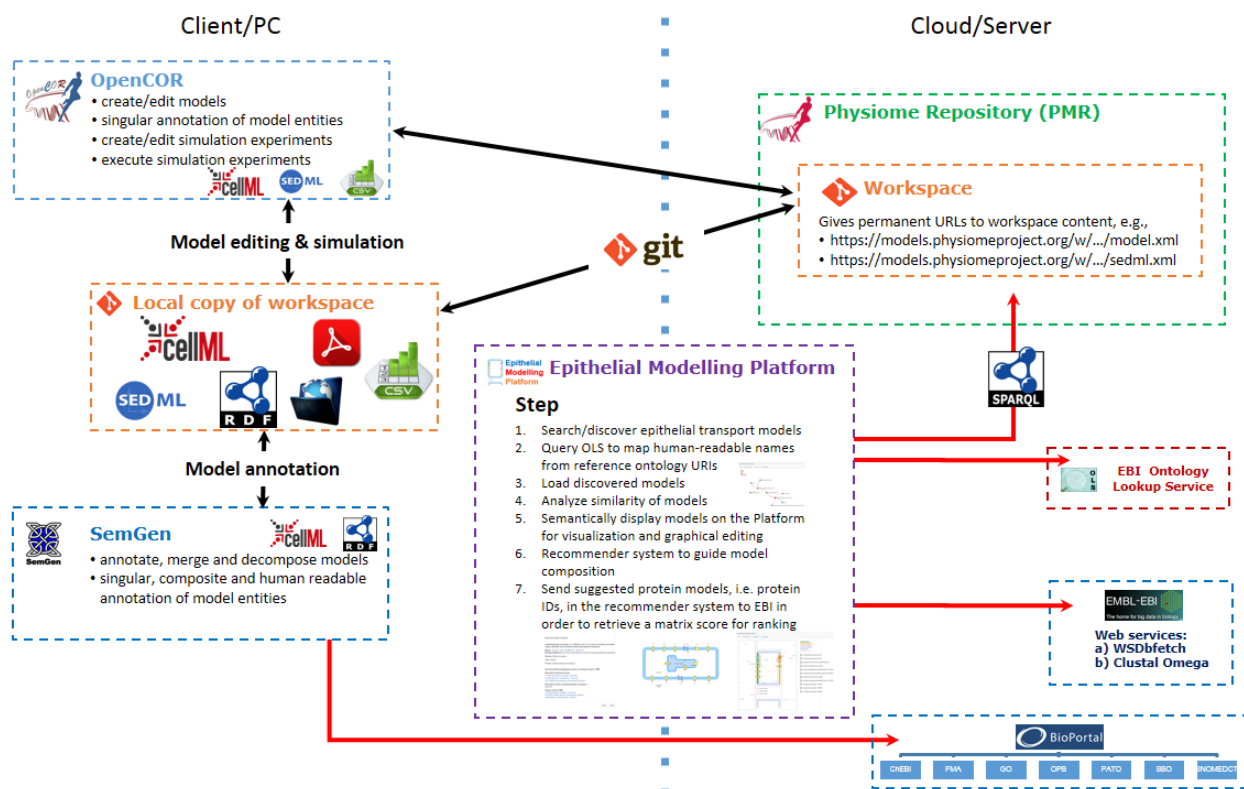


Figure 1. Illustrative example of current workflow for: 1) using OpenCOR to edit CellML models; 2) using SemGen to annotate the biological knowledge and deposit them in PMR; 3) using a git client to synchronize the local data to PMR; and 4) using our Epithelial Modelling Platform to discover CellML models and then semantically display these models on the platform for visualization, graphical editing, and model assembly and construction.

ontology lookup service (OLS) (OLS, 2019) and web services as shown in Fig. 1. The seven key steps in this process are described as follows. (1) Search/discover models of interest using the information extracted from the annotations deposited in PMR. (2) In this case, the EBI OLS is used to map human-readable names to reference ontology identifiers. (3) The user will then select models from the discovered model set and will make a shortlist of models for visualization and graphical editing on the platform. Similar to (2), the EBI OLS is used for name resolution. (4) In the shortlisted space, we have provided an option for the user to find similar components between a set of selected models, enabling users to identify overlap between models or points of integration. (5) We then semantically visualize the shortlisted models on a scalable vector graphics (SVG), which is an XML-based two-dimensional interactive display format, display panel for graphical editing. Specifically, the shortlisted models are placed on a specific membrane or compartment based on the annotation in PMR. (6) A recommender system is implemented to help the user enhance and extend the model they are building by providing links, for example, to models of similar transporters or transporters of the same solutes in different membrane locations. Such user-editing will automatically change the default settings of the models on the platform. Suggestions in this recommender system are provided from the available annotated apical or basolateral membrane models. (7) The EBI web services are used to rank the available models in the recommender system.

Computational tools and standards have been evolved over the years by The COmputational Modeling in BIology Network (COMBINE) (COMBINE, 2019) and the World Wide Web Consortium (W3C) in order

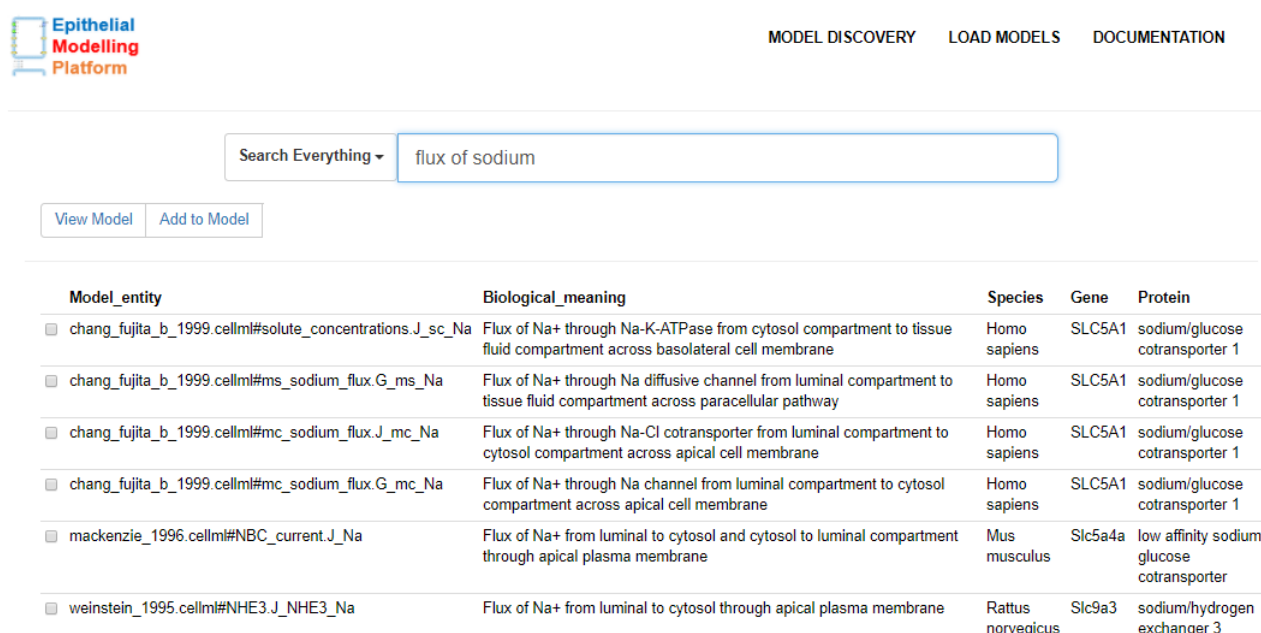
to collaborate and disseminate tools and standards. Some of the tools and standards used in our EMP are described in (Sarwar et al., 2018).

3 RESULTS

We have developed the EMP to discover and explore computational models of interest, as well as to semantically visualize the discovered models on the platform for graphical editing and model assembly with a view to constructing a new epithelial model. In particular, this involves annotation of mathematical models with biological information and generation of metadata using the SemGen tool so that our EMP could discover models from the metadata. Annotation of a cohort of epithelial transport models have been discussed in (Sarwar et al., 2018), therefore, following section presents a detailed workflow of the EMP.

Discover CellML Models

The first workflow of the EMP is to discover and explore models of interest extracted from the annotation in PMR. Fig. 2 shows a list of discovered models for a search query “flux of sodium” from PMR. From this list, the user can investigate various options such as CellML model entity – name of the model, component name and variable name; biological meaning deposited in PMR; protein name; and species and genes used during experiments of the associated models. In addition, the user can navigate to the “view” option to explore more information about a model. The user can make a shortlist of models by navigating through the “Add to Model” option and can return here in order to rediscover models.



The screenshot shows the Epithelial Modelling Platform interface. At the top, there is a navigation bar with the platform logo and three links: MODEL DISCOVERY, LOAD MODELS, and DOCUMENTATION. Below the navigation bar is a search bar with the text "Search Everything" and a dropdown arrow, followed by a text input field containing "flux of sodium". Below the search bar are two buttons: "View Model" and "Add to Model". Below these buttons is a table with five columns: Model_entity, Biological_meaning, Species, Gene, and Protein. The table contains six rows of search results.

Model_entity	Biological_meaning	Species	Gene	Protein
<input type="checkbox"/> chang_fujita_b_1999.cellml#solute_concentrations.J_sc_Na	Flux of Na ⁺ through Na-K-ATPase from cytosol compartment to tissue fluid compartment across basolateral cell membrane	Homo sapiens	SLC5A1	sodium/glucose cotransporter 1
<input type="checkbox"/> chang_fujita_b_1999.cellml#ms_sodium_flux.G_ms_Na	Flux of Na ⁺ through Na diffusive channel from luminal compartment to tissue fluid compartment across paracellular pathway	Homo sapiens	SLC5A1	sodium/glucose cotransporter 1
<input type="checkbox"/> chang_fujita_b_1999.cellml#mc_sodium_flux.J_mc_Na	Flux of Na ⁺ through Na-Cl cotransporter from luminal compartment to cytosol compartment across apical cell membrane	Homo sapiens	SLC5A1	sodium/glucose cotransporter 1
<input type="checkbox"/> chang_fujita_b_1999.cellml#mc_sodium_flux.G_mc_Na	Flux of Na ⁺ through Na channel from luminal compartment to cytosol compartment across apical cell membrane	Homo sapiens	SLC5A1	sodium/glucose cotransporter 1
<input type="checkbox"/> mackenzie_1996.cellml#NBC_current.J_Na	Flux of Na ⁺ from luminal to cytosol and cytosol to luminal compartment through apical plasma membrane	Mus musculus	Slc5a4a	low affinity sodium-glucose cotransporter
<input type="checkbox"/> weinstein_1995.cellml#NHE3.J_NHE3_Na	Flux of Na ⁺ from luminal to cytosol through apical plasma membrane	Rattus norvegicus	Slc9a3	sodium/hydrogen exchanger 3

Figure 2. Model discovery interface to search for models in PMR that are relevant to the query “flux of sodium”. By querying the annotations stored in PMR, this interface retrieves components from the NHE3 (Weinstein, 1995), SGLT2 (Mackenzie et al., 1996), and an epithelial cell model (Chang and Fujita, 1999).

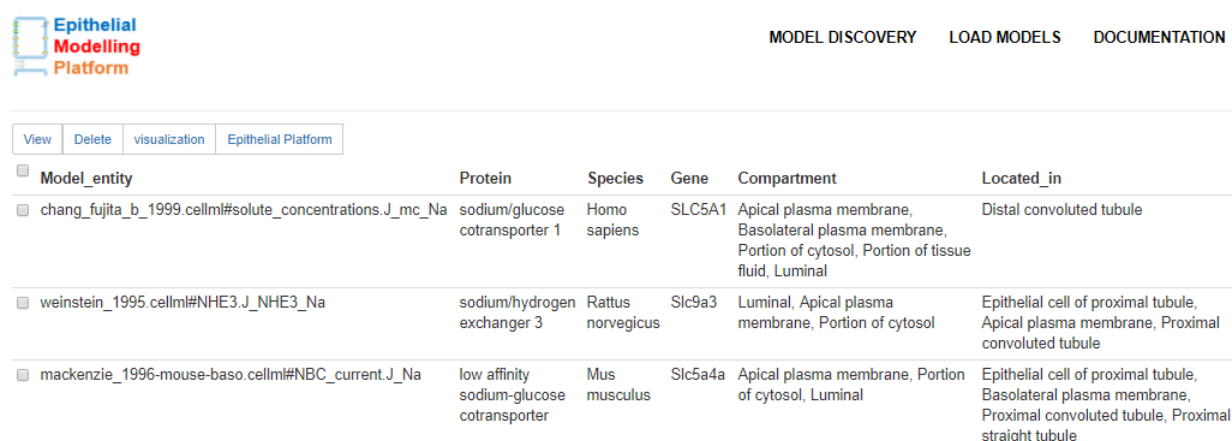
Users are able to enter human-readable search terms and phrases or choose to enter specific vocabulary terms to search for. When discovering resources in PMR relevant to a given search phrase, we first decompose the English prose into a series of semantic queries which are executed on the PMR knowledge base and ranked according to how well the discovered resource matches the initial search query. Continuing

our example, the phrase shown in Figure 2 will best match any resource associated directly with a sodium flux, but other sodium-related resources will also be returned with a lower ranking in the search results. To achieve that, we have maintained a static dictionary with key and value pairs to map searched text to reference ontology terms. For example, for a search query “flux of sodium”, the model discovery interface will map flux and sodium to OPB and ChEBI ontology terms, respectively, as defined in the dictionary.

In addition to model discovery, users are able to add protein models by querying to the bioportal web service. This begins with entering a protein name in the “Add Model” interface of the EMP and then the bioportal web service returns a protein ID from the protein ontology (PR). Afterwards, the interface sends this protein ID to the OLS to retrieve species and gene names. Next, users enter the rate of concentration of solutes (i.e. flux) moving between compartments across apical or basolateral membrane. After filling in the required information, metadata will be generated and a CellML model will be serialized. In the future, we will implement a notification system in order to notify this new CellML model to the CellML editorial board so that they can fill in the required ODE-based equations and metadata of this new CellML model.

Load Discovered Models

As mentioned above, the user can make a shortlist of models from the discovered models for visualization, graphical editing, and model assembly and construction. Fig. 3 presents an example where the Biological meaning in Fig. 2 is replaced with Compartment, and Located_in on the right of this example is added with a view to quickly provide more useful information to the user. In this space, we have provided some useful options: “view” option to get a detailed overview of a model; “delete” option to remove models if the user is not interested in considering these models on the platform; “visualization” option to find a similarity between a set of selected models; and “epithelial platform” option to navigate to the SVG-based interface and to semantically display the shortlisted models, which is discussed below.



The screenshot shows the Epithelial Modelling Platform interface. At the top, there is a logo and navigation links: MODEL DISCOVERY, LOAD MODELS, and DOCUMENTATION. Below the navigation links, there is a table of discovered models. The table has columns: Model_entity, Protein, Species, Gene, Compartment, and Located_in. The table contains three rows of model data.

Model_entity	Protein	Species	Gene	Compartment	Located_in
chang_fujita_b_1999.cellml#solute_concentrations.J_mc_Na	sodium/glucose cotransporter 1	Homo sapiens	SLC5A1	Apical plasma membrane, Basolateral plasma membrane, Portion of cytosol, Portion of tissue fluid, Luminal	Distal convoluted tubule
weinstein_1995.cellml#NHE3.J_NHE3_Na	sodium/hydrogen exchanger 3	Rattus norvegicus	Slc9a3	Luminal, Apical plasma membrane, Portion of cytosol	Epithelial cell of proximal tubule, Apical plasma membrane, Proximal convoluted tubule
mackenzie_1996-mouse-baso.cellml#NBC_current.J_Na	low affinity sodium-glucose cotransporter	Mus musculus	Slc5a4a	Apical plasma membrane, Portion of cytosol, Luminal	Epithelial cell of proximal tubule, Basolateral plasma membrane, Proximal convoluted tubule, Proximal straight tubule

Figure 3. List of models for visualization and graphical editing on the platform. These models are selected from the model discovery interface illustrated in Fig. 2.

Models of Similarity

Discovering similarity between models is vital to investigate coupling point(s), whereas visualization quickly conveys this investigation to the user. For this, we have developed models of similarity feature in the EMP. We aim to visually identify similar components, such as species, genes, and compartments between the selected models from the shortlist of models space. Fig. 4 shows a similar component between the

NHE3 (Weinstein, 1995) and SGLT2 (Mackenzie et al., 1996) models. In this example, both models have same compartment. Nodes and edges are represented with unique colors to make the model distinguishable. Overall, by using this feature, the user would be able to analyze models and could choose relevant models in order to construct an epithelial model.

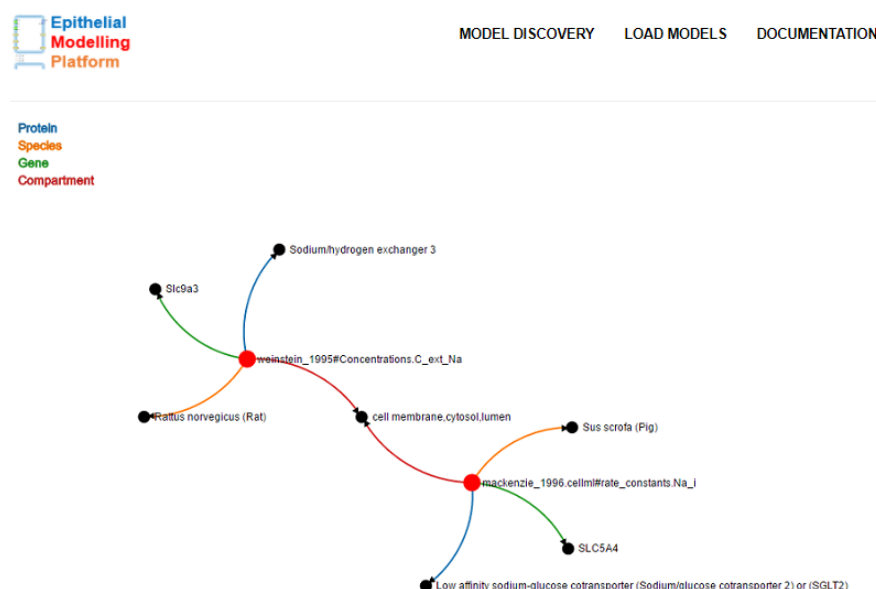


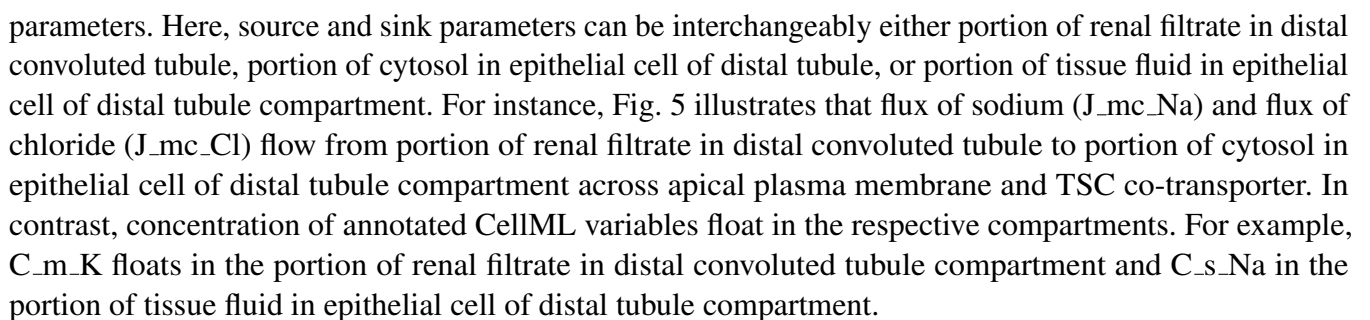
Figure 4. Illustrating the similarity between the NHE3 (Weinstein, 1995) and SGLT2 (Mackenzie et al., 1996) models. In this case, the transporter being modelled is located in the same compartment, labeled here with a unique colour. For convenience, we used “lumen” as a replacement of portion of renal filtrate in distal convoluted tubule and “cytosol” as a replacement of portion of cytosol in epithelial cell of distal tubule.

Modelling Platform

Fig. 5 presents an example of our SVG-based model visualization, graphical editing, and model assembly and construction platform. This platform has five compartments: portion of renal filtrate in distal convoluted tubule (luminal compartment), portion of cytosol in epithelial cell of distal tubule (cytosol compartment), paracellular, portion of tissue fluid in epithelial cell of distal tubule (interstitial fluid compartment), and blood capillary; and three membranes: apical, basolateral, and capillary. These compartments and membranes have been annotated with unique colors as shown in the top-right corner of this example. In order to distinguish between fluxes and channels, we have used various shapes, as described in Figure 5.

The paracellular compartment separates two epithelial cells where diffusive fluxes are semantically placed on this compartment. In a similar manner, fluxes and channels are placed on a membrane based on the biological semantics. Before placing these fluxes and channels on a membrane, the platform checks its mediator (either apical, basolateral, or capillary membrane ID, and a protein ID) and shape (circles for fluxes or polygons for channels). A co-transporter, for example, Na-Cl co-transporter (TSC) in Fig. 5, is constructed if both (Na^+ and Cl^-) fluxes have a mediator in common. On the right, checkboxes are generated for the visualized fluxes and channels in order to allow the user to move them across the membranes.

Direction of a flux is represented with a line, an arrow symbol, and a text to define the solute being transported. This direction is defined in the biological information deposited in PMR with source and sink



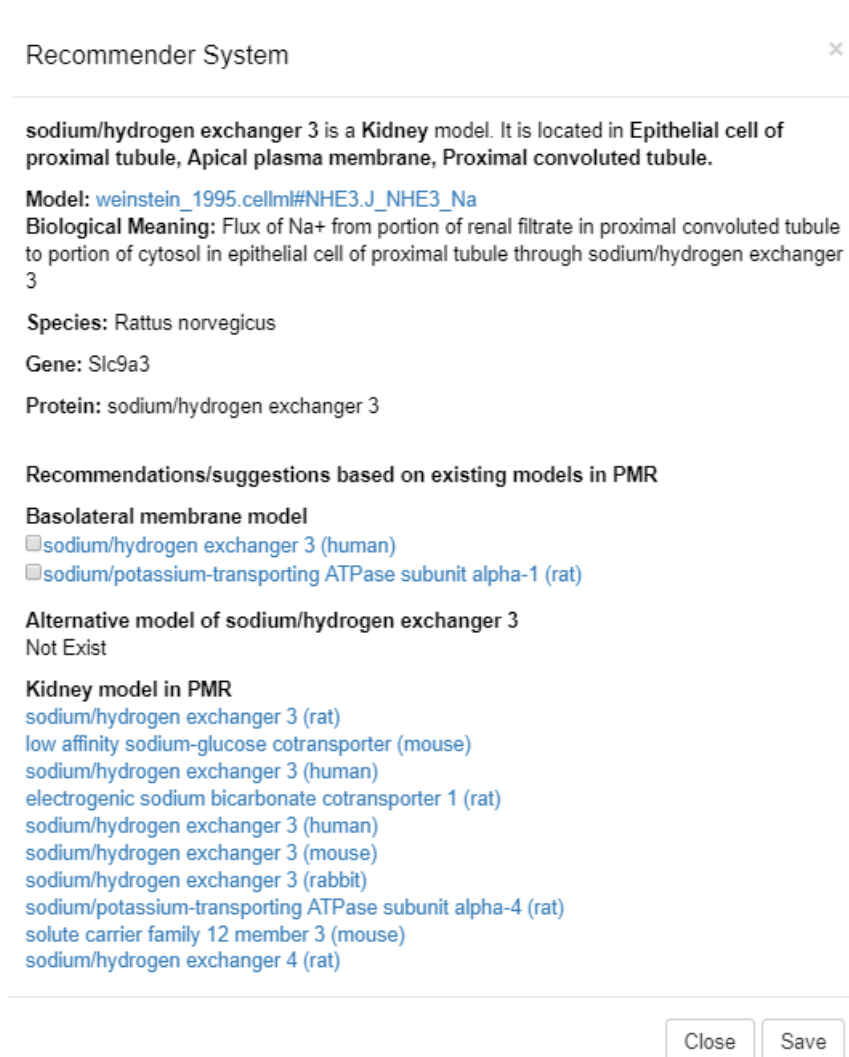


Figure 6. Recommender system as a pop-up window, with some information and suggestions, when a circle represented as “flux of Na^+ from portion of renal filtrate in distal convoluted tubule to portion of cytosol in epithelial cell of distal tubule through the sodium/hydrogen exchanger 3 (NHE3) (Weinstein, 1995) model” is moved from apical to basolateral membrane.

Recommender System

On the SVG-based modelling platform of the EMP, we have provided a feature to make recommendations or suggestions for the user to perform graphical editing. In this case, users can drag a model from apical to basolateral membrane or vice versa, then a pop-up window will appear with a brief description of the dragged model extracted from the biological semantics in PMR, as well as some suggestions provided by the EBI web services. Here, we have applied our concept presented in (Sarwar et al., 2018) in regard to find similar models as suggestions.

Similar to Figure 4 in (Sarwar et al., 2018), a list of suggestions begins with a brief description of the dragged model: CellML model entity with an URI pointing to the PMR workspace, species and gene used during the experiment, and a protein name. Next, this recommender system provides some suggestions on existing annotated models as opposed to the dragged model’s membrane. Fig. 6 shows an example when “flux of sodium from portion of renal filtrate in distal convoluted tubule to portion of cytosol in

epithelial cell of distal tubule through the sodium/hydrogen exchanger 3 (NHE3) (Weinstein, 1995) model” is dragged from apical to basolateral membrane. Therefore, the recommender system extracts existing annotated models in the basolateral membrane from PMR. Users can choose one of the suggested models to replace with the dragged NHE3 model.

To populate the suggested models, we have leveraged two web services hosted at the EBI: WSDbfetch (WSDbfetch, 2019) and Clustal Omega (Clustal, 2019; Sievers et al., 2011). The WSDbfetch was used to get protein sequences for a list of protein identifiers. These protein sequences were formatted as JSON and then dispatched to the Clustal Omega web service to get a comparison score between the protein sequences in a representation of a similarity matrix. Fig. 7 illustrates a similarity matrix retrieved from the EBI Clustal Omega web service.

On the pop-up window, another recommendation is made as an alternative model of the dragged NHE3 model, i.e. same protein but different species, genes, or workspaces. In this case, “Not Exist” shows in Fig. 6. Lastly, a list of related organ models is shown for the users to investigate further.

```
Identity Matrix: #
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: sp|P06685|AT1A1_RAT    100.00   18.77   18.55
2: sp|P48764|SL9A3_HUMAN  18.77  100.00   89.49
3: sp|P26433|SL9A3_RAT   18.55   89.49  100.00
```

Figure 7. A similarity matrix retrieved from the EBI clustalo omega web service where P26433 is the dragged protein model and the remaining protein models are from Fig. 6, see list of proteins under “Basolateral membrane model” section header. Ranking of these proteins are based on IDs and matrix scores: P48764 (89.49) and P06685 (18.55), with respect to the dragged P26433 protein model (NHE3).

Model Assembly Service

The end result of our model discovery and assembly is a recipe that consists of models’ components and relevant information to help construct a new epithelial model. To achieve this, we have implemented a model assembly service which does model composition by utilizing the libCellML (libCellML, 2019) library. The source code of the model assembly service is available at <https://github.com/dewancse/epithelial-modelling-platform/tree/master/server>.

libCellML

libCellML is a library to serialize, validate, and instantiate a CellML model by utilizing convenience methods and objects in C++. It also provides Python bindings to access these convenience methods and objects in Python. The source code and documentation of libCellML are available at <https://github.com/cellml/libcellml> and <http://libcellml.readthedocs.io/>.

Model-recipe and Composition

The first step of the model assembly service was to iterate over the model entities in the model-recipe with a view to extracting models’ components and then importing these components and the associated variables in the new model. Fig. 8 shows an example of eleven models’ components which is reproduced from Fig. 5. For example, Na-Cl co-transporter (TSC) is located on the apical membrane where “Na⁺ and

Cl^- flow from portion of renal filtrate in distal convoluted tubule (“Luminal Compartment” indicated in the top-right corner in Fig. 8) to portion of cytosol in epithelial cell of distal tubule (“Cytosol Compartment” indicated in the top-right corner in Fig. 8) compartment”; and Na-K pump (P06685) is located on the basolateral membrane where “ Na^+ flows from portion of cytosol in epithelial cell of distal tubule to portion of tissue fluid in epithelial cell of distal tubule (“Interstitial Fluid” indicated in the top-right corner in Fig. 8) compartment” and “ K^+ flows from portion of tissue fluid in epithelial cell of distal tubule to portion of cytosol in epithelial cell of distal tubule compartment”. Relevant concentration of solutes float in the portion of renal filtrate in distal convoluted tubule, portion of cytosol in epithelial cell of distal tubule, and portion of tissue fluid in epithelial cell of distal tubule compartment.

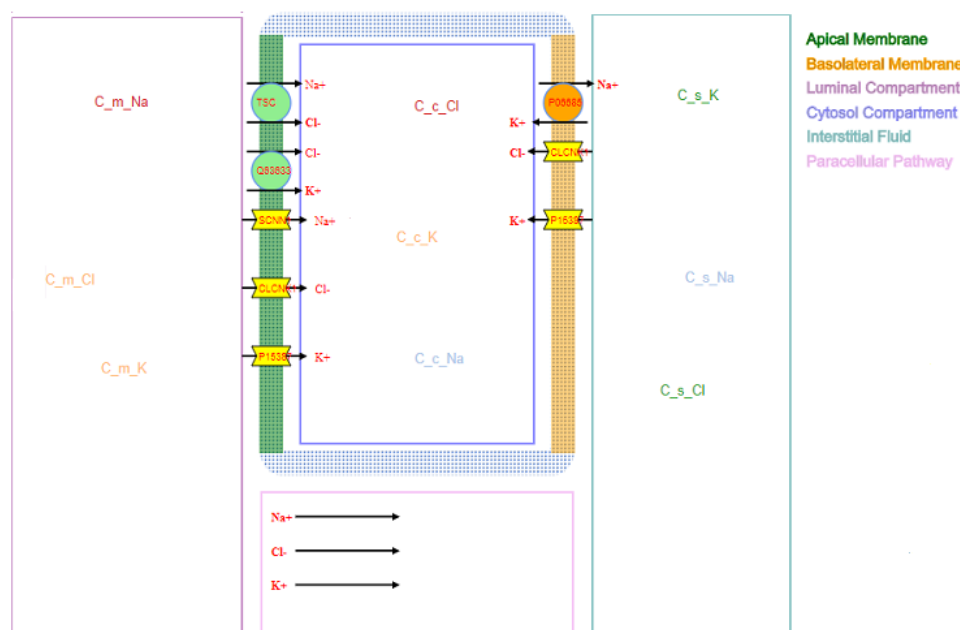


Figure 8. An example of the platform where Na-Cl and K-Cl co-transporters, Na channel, K channel, and Cl channel are located on the apical membrane; Na-K pump, K channel, and Cl channel are located on the basolateral membrane; and Na, K and Cl diffusive fluxes are located on the paracellular compartment. Respective concentration of solutes float in the portion of renal filtrate in distal convoluted tubule (“Luminal Compartment”), portion of cytosol in epithelial cell of distal tubule (“Cytosol Compartment”) and portion of tissue fluid in epithelial cell of distal tubule (“Interstitial Fluid”) compartment.

Fig. 9 shows a model-recipe, which is the background information of the visualized models in Fig. 8. This recipe has a detailed annotated information to construct a new epithelial model. In Fig. 9, med_fma denotes mediator FMA which has the URI of apical plasma membrane; med_pr denotes solute carrier family 12 member 3 or thiazide-sensitive sodium-chloride co-transporter; med_pr_text and med_pr_text_syn denote the textual representation and synonym (in short form) of the med_pr, respectively; model_entity combines model name (chang_fujita_b.1999.cellml), component name (total_transepithelial_sodium_flux) and variable name (J_{mc_Na}); model_entity2 is the second part of this co-transporter where model name is chang_fujita_b.1999.cellml, component name is solute_concentrations and variable name is J_{mc_Cl} ; protein_name, in this case, is a reference URI of the epithelial cell.

As shown in Fig. 8, “ Na^+ and Cl^- flow from portion of renal filtrate in distal convoluted tubule to portion of cytosol in epithelial cell of distal tubule compartment across apical plasma membrane and Na-Cl co-transporter (TSC)”, thus, source_fma is a reference URI of portion of renal filtrate in distal convoluted

```

▼ 0:
  med_fma: "http://purl.obolibrary.org/obo/FMA_84666"
  med_pr: "http://purl.obolibrary.org/obo/PR_P55018"
  med_pr_text: "solute carrier family 12 member 3 (rat)"
  med_pr_text_syn: "TSC"
  model_entity: "chang_fujita_b_1999.cellml#total_transepithelial_sodium_flux.J_mc_Na"
  model_entity2: "chang_fujita_b_1999.cellml#solute_concentrations.J_mc_Cl"
  model_entity3: ""
  protein_name: "http://purl.obolibrary.org/obo/CL_0000066"
  sink_fma: "http://purl.obolibrary.org/obo/FMA_66836"
  sink_fma2: "http://purl.obolibrary.org/obo/FMA_66836"
  sink_fma3: ""
  solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29101"
  solute_chebi2: "http://purl.obolibrary.org/obo/CHEBI_17996"
  solute_chebi3: ""
  solute_text: "Na+"
  solute_text2: "Cl-"
  solute_text3: ""
  source_fma: "http://purl.obolibrary.org/obo/FMA_74550"
  source_fma2: "http://purl.obolibrary.org/obo/FMA_74550"
  source_fma3: ""
  variable_text: "J_mc_Na"
  variable_text2: "J_mc_Cl"
  variable_text3: ""
  ▶ __proto__: Object
  ▶ 1: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29101", solute_text: "Na+"
  ▶ 2: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29101", solute_text: "Na+"
  ▶ 3: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_17996", solute_text: "Cl-"
  ▶ 4: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29103", solute_text: "K+"
  ▶ 5: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_17996", solute_text: "Cl-"
  ▶ 6: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29101", solute_text: "Na+"
  ▶ 7: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_17996", solute_text: "Cl-"
  ▶ 8: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29103", solute_text: "K+"
  ▶ 9: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29103", solute_text: "K+"
  ▶ 10: {solute_chebi: "http://purl.obolibrary.org/obo/CHEBI_29103", solute_text: "K+"

```

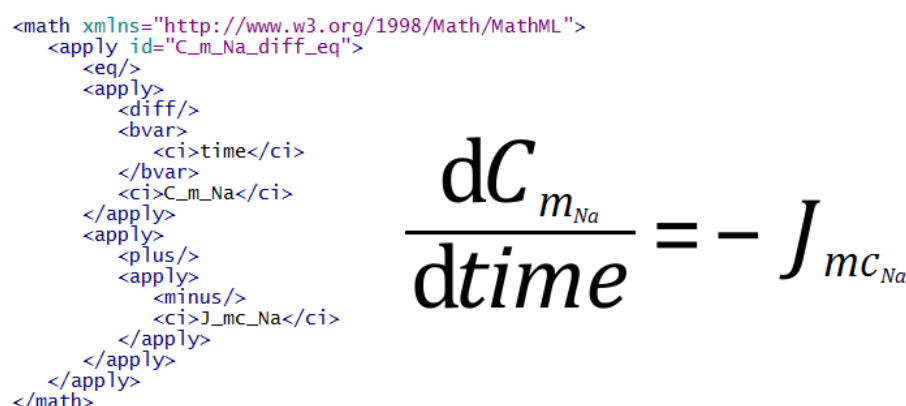
Figure 9. A detailed annotated information encapsulated inside a Na-Cl co-transporter (TSC) object visualized on the apical membrane in Fig. 8, which is a recipe to perform model composition by utilizing the libCellML (libCellML, 2019). The remaining co-transporters and channels visualized in Fig. 8 are kept hidden for convenience.

tubule compartment, sink_fma is a reference URI of portion of cytosol in epithelial cell of distal tubule compartment, solute_chebi is a reference URI of sodium, solute_text is a short form of sodium, variable_text is a CellML variable name of sodium, source_fma2 is a reference URI of portion of renal filtrate in distal convoluted tubule compartment, sink_fma2 is a reference URI of portion of cytosol in epithelial cell of distal tubule compartment, solute_chebi2 is a reference URI of chloride, solute_text2 is a short form of chloride, and variable_text2 is a CellML variable name of chloride. Similarly, Na-K pump in Fig. 8 has a detailed annotated information which is kept hidden in Fig. 9 for convenience.

In addition to importing components and the associated variables from the source models in the model-recipe, we encapsulated the imported components inside an epithelial component. Next, we created an environment component that contains a time variable, and also created three components - lumen (portion of renal filtrate in distal convoluted tubule), cytosol (portion of cytosol in epithelial cell of distal tubule), and interstitial fluid (portion of tissue fluid in epithelial cell of distal tubule) to store relevant variables from the model-recipe. The epithelial component contains additional variables that are required in the lumen, cytosol, and interstitial fluid component.

Following step involves iterating over the model entities in the model-recipe and then extracting variable and component names. As mentioned above, we encapsulated the extracted components inside an epithelial

component. On the other hand, variable names were used to represent fluxes, channels, and diffusive fluxes. We constructed ordinary differential equations (ODE) to represent fluxes. However, for channels and diffusive fluxes, we extracted the full mathematical equations from the source models in the model-recipe. To construct these equations, we instantiated a time variable in the lumen, cytosol, and interstitial fluid component.



```

<math xmlns="http://www.w3.org/1998/Math/MathML">
  <apply id="C_m_Na_diff_eq">
    <eq/>
    <apply>
      <diff/>
      <bvar>
        <ci>time</ci>
      </bvar>
      <ci>C_m_Na</ci>
    </apply>
    <plus/>
    <apply>
      <minus/>
      <ci>J_mc_Na</ci>
    </apply>
  </apply>
</math>

```

$$\frac{dC_{m_{Na}}}{dtime} = - J_{mc_{Na}}$$

Figure 10. A MathML (MathML, 2019) representation of “flux of sodium” (J_{mc_Na}) which is an ordinary differential equation for “concentration of sodium” (C_{m_Na}) with respect to time.

To construct an ODE equation, we need to have concentration of solutes in the respective compartments - portion of renal filtrate in distal convoluted tubule, portion of cytosol in epithelial cell of distal tubule, or portion of tissue fluid in epithelial cell of distal tubule. For example, to construct an ODE equation for “flux of sodium from portion of renal filtrate in distal convoluted tubule to portion of cytosol in epithelial cell of distal tubule compartment across apical plasma membrane”, we need to have “concentration of sodium in the portion of renal filtrate in distal convoluted tubule” and “concentration of sodium portion of cytosol in epithelial cell of distal tubule” compartment, then we can make an ODE equation which will measure “flux of sodium from portion of renal filtrate in distal convoluted tubule to portion of cytosol in epithelial cell of distal tubule” compartment, i.e. “sodium flows from portion of renal filtrate in distal convoluted tubule to portion of cytosol in epithelial cell of distal tubule” compartment. To get the respective concentration of solutes, we made a SPARQL call to PMR for a given solute (sodium, chloride, etc) and compartment (portion of renal filtrate in distal convoluted tubule, portion of cytosol in epithelial cell of distal tubule, or portion of tissue fluid in epithelial cell of distal tubule). The result of this call was a list of concentration of the specified solute in the specified compartment. Then we iterated over each concentration variable in the result and checked that whether both the flux and concentration variable reside in the same component. In the ODE equation, we marked with plus or minus sign notation to denote in which direction the concentration of the solute flows. After getting the respective concentration variable of the flux and the notation, we constructed an ODE equation by utilizing a MathML (MathML, 2019). Fig. 10 illustrates an ODE equation - concentration of sodium in the portion of renal filtrate in distal convoluted tubule compartment (C_{m_Na}) with respect to time is equal to negative flow of concentration of sodium (J_{mc_Na}). This flow is negative because concentration of sodium is moving from portion of renal filtrate in distal convoluted tubule to portion of cytosol in epithelial cell of distal tubule compartment.

In contrast, to construct mathematical equations for channels and diffusive fluxes, we extracted the full math equation from the source model in the model-recipe. For example, Fig. 11 shows an example equation of a sodium channel from the Chang and Fujita model (Chang and Fujita, 1999). We extracted

$$G_{mc_{Na}} = P_{mc_{Na}} \cdot \frac{F \cdot (\psi_m - \psi_c)}{RT} \cdot \frac{C_{m_{Na}} - C_{c_{Na}} \cdot e^{-\frac{F}{RT} \cdot (\psi_m - \psi_c)}}{1 - e^{-\frac{F}{RT} \cdot (\psi_m - \psi_c)}}$$

Figure 11. Equation of a “sodium channel” ($G_{mc_{Na}}$) visualized as polygons (SCNN1) on the apical membrane in Fig. 8.

this channel’s equation and traversed through the variables of this equation in order to instantiate them in the epithelial component of the new model.

```

<math xmlns="http://www.w3.org/1998/Math/MathML">
  <apply id="J_Na_lumen_calculation">
    <eq/>
    <ci>J_Na_lumen</ci>
    <apply>
      <plus/>
      <apply>
        <minus/>
        <ci>J_mc_Na</ci>
      </apply>
    </apply>
  </apply>
  <apply id="J_Cl_lumen_calculation">
    <eq/>
    <ci>J_Cl_lumen</ci>
    <apply>
      <plus/>
      <apply>
        <minus/>
        <ci>J_mc_Cl</ci>
      </apply>
    </apply>
  </apply>
</math>

```

$$J_{Na_{lumen}} = - J_{mc_{Na}}$$

$$J_{Cl_{lumen}} = - J_{mc_{Cl}}$$

Figure 12. A MathML (MathML, 2019) representation of “total flux of sodium” ($J_{Na_{lumen}}$) and “total flux of chloride” ($J_{Cl_{lumen}}$) in the portion of renal filtrate in distal convoluted tubule (Luminal Compartment) compartment in Fig. 8.

In terms of total fluxes, channels and diffusive fluxes equations, we evaluated out-going and in-coming fluxes of components and then constructed equations. For example, Fig. 12 illustrates “total flux of sodium” and “total flux of chloride” in the portion of renal filtrate in distal convoluted tubule (luminal compartment) compartment of Fig. 8. In the final step of the model assembly service, we mapped variables between the epithelial component and its encapsulated components. We also mapped variables between epithelial component’s encapsulated siblings.

4 DISCUSSION

We have developed a novel web-based platform, the Epithelial Modelling Platform (EMP), for scientists to discover and explore mathematical models relevant to their work. On this platform, they can semantically visualize models for graphical editing and model assembly with a view to constructing a new model. In particular, this involves three steps: annotation of mathematical models with biological information; generation of metadata using the SemGen tool; and the EMP to discover models from the metadata in order to visualize models for graphical editing and model assembly and construction.

To perform biological annotation, we have comprehensively annotated a cohort of epithelial transport models: twelve renal and four lung models, which are available at <https://models.physiomeproject.org/workspace/527>. In particular, we pulled out mathematical terms from

each model and annotated with biological information. In addition to this, we curated the models with protein identifiers from the UniProt database, species and gene names used during the experiments of the models, compartment names, and anatomical locations. Next, we fed these biological information onto the SemGen tool to generate metadata. This metadata is then deposited in PMR which are discoverable, accessible and persistent.

To discover and explore the epithelial transport models, we have developed the EMP that utilizes cascaded SPARQL calls to iteratively retrieve information from PMR. First, we make a query with a protein identifier to the OLS in order to get a human-readable name. Subsequent call retrieves protein's species and gene name. In addition to model discovery, users are able to add protein models and generate metadata in PMR. The EMP allows users to semantically display a collection of discovered models for graphical editing and model assembly. This “graphical editing” has been integrated with a recommender system where users can rank similar models based on the suggestions provided by the EBI web services.

The end result of the graphical editing and model assembly is the generation of a model-recipe, which is a collection of components from disparate discovered models. Fig. 9 illustrates an example model-recipe that has the required information to perform model composition. This recipe is then dispatched to our model assembly service in order to compose a new epithelial model. To perform model composition, we have utilized convenience methods and objects of the libCellML library.

The integration and discovery of the semantic annotation visualized on the SVG-based epithelial platform is a novel application of the services provided by PMR that would inspire users to reuse the data and models. Integration of data and models with the provision of semantic web technologies will integrate computational model of various organ systems, which is the ultimate effort of the Physiome Project and the VPH. We believe this approach with the aid of modern tools and technologies would provide an excellent epithelial modelling platform. Novice modellers could use this platform as a learning tool.

The “Add Model” feature enables scientists to create a new CellML model and metadata. In the future, we would like to extend this feature and implement a notification system between CellML editorial board and the scientists so that the CellML editorial board can incorporate mathematical equations and validate the new CellML model.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING

DMS was supported by the Medical Technologies Centre of Research Excellence's Doctoral Scholarship. DPN was supported by an Aotearoa Foundation Fellowship. JHG, BEC, and MLN were supported by the National Institutes of Health grant R01LM011969.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Centre for eResearch at the University of Auckland for their help in facilitating this research. The availability of our live web application demonstration is made possible by use of the Nectar Research Cloud, a collaborative Australian research platform supported by the

National Collaborative Research Infrastructure Strategy (NCRIS). Thanks to Sean Matheny at the Center for eResearch for his cooperation to setup our tool and service in Center for eResearch at the University of Auckland and Nectar.

AVAILABILITY OF DATA AND MATERIALS

The source code and links to the live demonstration of the Epithelial Modelling Platform are available at <https://github.com/dewancse/epithelial-modelling-platform>.

REFERENCES

- Chang, H. and Fujita, T. (1999). A numerical model of the renal distal tubule. *American Journal of Physiology* 276, F931–F951. doi:10.1152/ajprenal.1999.276.6.F931
- Clustal (2019). EBI webservice to get multiple sequence alignment. [https://www.ebi.ac.uk/seqdb/confluence/display/WEBSERVICES/clustalo\\$_\\$rest](https://www.ebi.ac.uk/seqdb/confluence/display/WEBSERVICES/clustalo$_$rest). Accessed 21 January.
- COMBINE (2019). The COmputational Modeling in Biology NEtwork (combine). <http://co.mbine.org/>. Accessed 21 January.
- Cook, D. L., Neal, M. L., Hoehndorf, R., Gkoutos, G. V., and Gennari, J. H. (2013). Representing physiological processes and their participants with PhysioMaps. *Biomedical Semantics* 4. doi:10.1186/2041-1480-4-S1-S2
- Cook, D. L., Wiley, J. C., and Gennari, J. H. (2007). Chalkboard: ontology-based pathway modelling and qualitative inference of disease mechanisms. *Pacific Symposium on Biocomputing*, 16–27
- Cooper, J., Cervenansky, F., De Fabritiis, G., Fenner, J., Friboulet, D., Giorgino, T., et al. (2010). The Virtual Physiological Human toolkit. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 368, 3925–3936. doi:10.1098/rsta.2010.0144
- Cooper, J., Scharm, M., and Mirams, G. R. (2015a). The Cardiac Electrophysiology Web Lab. *PeerJ PrePrints* doi:10.7287/peerj.preprints.1338v1
- Cooper, J., Vik, J. O., and Waltemath, D. (2015b). A call for virtual experiments: Accelerating the scientific process. *Progress in Biophysics and Molecular Biology* 117, 99–106. doi:10.1016/j.pbiomolbio.2014.10.001
- Coskun, S. A., Cicek, A. E., Lai, N., Dash, R. K., Ozsoyoglu, Z. M., and Ozsoyoglu, G. (2013). An online model composition tool for system biology models. *BMC Systems Biology* 7, 88. doi:10.1186/1752-0509-7-88
- Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., et al. (2011). Controlled vocabularies and semantics in systems biology. *Molecular Systems Biology* 7. doi:10.1038/msb.2011.77
- Cuellar, A., C.M., L., P.F., N., D.P., B., Nickerson, D., and P.J., H. (2003). An overview of CellML 1.1, a biological model description language. *SIMULATION* 79, 740–747. doi:10.1177/0037549703040939
- de Boer, T. P., van der Werf, S., Hennekam, B., Nickerson, D. P., Garny, A., Gerbrands, M., et al. (2017). esolv, a cellml-based simulation front-end for online teaching. *Advances in Physiology Education* 41, 425–427. doi:10.1152/advan.00127.2016
- de Bono, B., Grenon, P., Helvensteijn, M., Kok, J., and Kokash, N. (2014). ApiNATOMY: Towards multiscale views of human anatomy. *The 13th International Symposium on Intelligent Data Analysis (IDA)* 8819. doi:10.1007/978-3-319-12571-8_7
- de Bono, B., Helvensteijn, M., Kokash, N., Martorelli, I., Sarwar, D., Islam, S., et al. (2016). Requirements for the formal representation of pathophysiology mechanisms by clinicians. *Interface Focus Journal* 6, 20150099. doi:10.1098/rsfs.2015.0099

- Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M., and Gennari, J. H. (2011). Standard biological parts knowledgebase. *PLOS ONE* 6, e17005. doi:10.1371/journal.pone.0017005
- Garny, A., Cooper, J., and Hunter, P. J. (2010). Toward a VPH/Physiome toolkit. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 2, 134–147. doi:10.1002/wsbm.63
- Garny, A. and Hunter, P. J. (2015). Opencor: a modular and interoperable approach to computational biology. *Frontiers in Physiology* 6. doi:10.3389/fphys.2015.00026
- Gennari, J. H., Neal, M. L., Carlson, B. E., and Cook, D. L. (2008). Integration of Multi-scale Biosimulation Models via Light-weight Semantics. *Pacific Symposium on Biocomputing* , 414–425
- Gennari, J. H., Neal, M. L., Galdzicki, M., and Cook, D. L. (2011). Multiple ontologies in action: Composite annotations for biosimulation models. *Journal of Biomedical Informatics* 44, 146–154. doi:10.1016/j.jbi.2010.06.007
- Hoehndorf, R., Dumontier, M., Gennari, J. H., Wimalaratne, S., de Bono, B., Cook, D. L., et al. (2011). Integrating systems biology models and biomedical ontologies. *BMC Systems Biology* 5, 124. doi:10.1186/1752-0509-5-124
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., et al. (2003). The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics Oxford Journals* 19, 524–531
- Hunter, P., Coveney, P. V., de Bono, B., Diaz, V., Fenner, J., Frangi, A. F., et al. (2010). A vision and strategy for the virtual physiological human in 2010 and beyond. *Philos Transact A Math Phys Eng Sci* 368, 2595–614. doi:10.1098/rsta.2010.0048
- Hunter, P. J. and Borg, T. K. (2003). Integration from proteins to organs: the Physiome Project. *Nature Reviews Molecular Cell Biology* 4, 237–243. doi:10.1038/nrm1054
- Krause, F., Uhlendorf, J., Lubitz, T., Schulz, M., Klipp, E., and Liebermeister, W. (2010). Annotation and merging of sbml models with semanticsbml. *Bioinformatics* 26, 421–422. doi:10.1093/bioinformatics/btp642
- libCellML (2019). libCellML. <https://libcellml.readthedocs.io/en/latest/index.html>. Accessed 21 January.
- Lister, A. L., Pocock, M., Taschuk, M., and Wipat, A. (2009). Saint: a lightweight integration environment for model annotation. *Oxford Bioinformatics* 25, 3026–3027. doi:10.1093/bioinformatics/btp523
- Mackenzie, M., Loo, D., Panayotova-Heiermann, M., and Wright, E. M. (1996). Biophysical characteristics of the pig kidney na/glucose cotransporter sglt2 reveal a common mechanism for sglt1 and sglt2. *Journal of Biological Chemistry* 271, 32678–32683. doi:10.1074/jbc.271.51.32678
- MathML (2019). Mathematical Markup Language (mathml). <https://www.w3.org/TR/MathML3/>. Accessed 21 January.
- Neal, M. L., Carlson, B. E., Thompson, C. T., James, R. C., Kim, K. G., Tran, K., et al. (2015). Semantics-based composition of integrated cardiomyocyte models motivated by real-world use cases. *PLoS ONE* 10, e0145621. doi:10.1371/journal.pone.0145621
- Neal, M. L., Cook, D. L., and Gennari, J. H. (2013). An owl knowledge base for classifying and querying collections of physiological models: A prototype human physiome. *Proceedings of the 2013 International Conference on Biomedical Ontology*.
- Neal, M. L., Gennari, J. H., Arts, T., and Cook, D. L. (2009). Advances in Semantic Representation for Multiscale Biosimulation: A Case Study in Merging Models. *Pacific Symposium on Biocomputing* , 304–315

- Neal, M. L., Thompson, C. T., Kim, K. G., James, R. C., Cook, D. L., Carlson, B. E., et al. (2018). Semgen: a tool for semantics-based annotation and composition of biosimulation models. *Bioinformatics* , bty829doi:10.1093/bioinformatics/bty829
- OLS (2019). EBI Ontology Lookup Service. <https://www.ebi.ac.uk/ols/index>. Accessed 21 January.
- Piñero, J., Queralt-Rosinach, N., Àlex Bravo, Deu-Pons, J., Bauer-Mehren, A., Baron, M., et al. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* 2015, bav028. doi:10.1093/database/bav028
- Sarwar, D. M., Kalbasi, R., Gennari, J. H., Carlson, B. E., Neal, M. L., de Bono, B., et al. (2018). Model annotation and discovery with the physiome model repository. *bioRxiv* doi:10.1101/498501
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology* 7. doi:10.1038/msb.2011.75
- Weinstein, A. (1995). A kinetically defined Na⁺/H⁺ antiporter within a mathematical model of the rat proximal tubule. *The Journal of General Physiology* 105, 617–641
- Wilkinson, M. D. and et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3. doi:10.1038/sdata.2016.18
- Wimalaratne, S. M., Halstead, M. D. B., Lloyd, C. M., Cooling, M. T., Crampin, E. J., and Nielsen, P. F. (2004). Web-based kinetic modelling using JWS Online. *Applications Note in Bioinformatics Oxford Journal* 20, 2143–2144. doi:10.1093/bioinformatics/bth200
- WSDbfetch (2019). EBI webservice to get entries from various biological databases. <https://www.ebi.ac.uk/seqdb/confluence/pages/viewpage.action?pageId=48923608>. Accessed 21 January.
- Yu, T., Lloyd, C. M., Nickerson1, D. P., Cooling, M. T., Miller, A. K., Garny, A., et al. (2011). The Physiome Model Repository 2. *Bioinformatics Oxford Journals* 27, 743–744. doi:10.1093/bioinformatics/btq723