


METHODOLOGY ARTICLE

Open Access



Model annotation and discovery with the Physiome Model Repository

Dewan M. Sarwar¹, Reza Kalbasi¹, John H. Gennari², Brian E. Carlson³, Maxwell L. Neal⁴, Bernard de Bono¹, Koray Atalag¹, Peter J. Hunter¹ and David P. Nickerson^{1*} 

Abstract

Background: Mathematics and Physics-based simulation models have the potential to help interpret and encapsulate biological phenomena in a computable and reproducible form. Similarly, comprehensive descriptions of such models help to ensure that such models are accessible, discoverable, and reusable. To this end, researchers have developed tools and standards to encode mathematical models of biological systems enabling reproducibility and reuse, tools and guidelines to facilitate semantic description of mathematical models, and repositories in which to archive, share, and discover models. Scientists can leverage these resources to investigate specific questions and hypotheses in a more efficient manner.

Results: We have comprehensively annotated a cohort of models with biological semantics. These annotated models are freely available in the Physiome Model Repository (PMR). To demonstrate the benefits of this approach, we have developed a web-based tool which enables users to discover models relevant to their work, with a particular focus on epithelial transport. Based on a semantic query, this tool will help users discover relevant models, suggesting similar or alternative models that the user may wish to explore or use.

Conclusion: The semantic annotation and the web tool we have developed is a new contribution enabling scientists to discover relevant models in the PMR as candidates for reuse in their own scientific endeavours. This approach demonstrates how semantic web technologies and methodologies can contribute to biomedical and clinical research. The source code and links to the web tool are available at <https://github.com/dewancse/model-discovery-tool>

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

Background

Over the years, computational models of human organ systems have been used to support and improve the diagnosis, treatment and prevention of diseases. These models help scientists investigate specific research questions and hypotheses which may be ethically inappropriate, expensive to study experimentally, or not practically achievable using human or animal subjects. Discovery and exploration of relevant computational models can help scientists more quickly and accurately test their clinical or experimental hypotheses which may involve a broad range of biophysical mechanisms and observations such as disease states, drug actions and vast number of clinical

observations. Ultimately this is expected to improve the quality, safety, effectiveness and reduce healthcare costs.

The International Union of Physiological Sciences (IUPS) Physiome Project [1] and the Virtual Physiological Human (VPH) initiative [2] aim at leveraging mathematics and physics-based computational models to represent functional systems of the human body in a computable form, and most importantly, apply our gained physiological insight into clinical practice. In order to achieve this, reproducibility and reuse of computational models are key for building such computational models [3, 4].

A key aspect of the Physiome Project and the VPH is to incorporate semantic annotation of computational models such that these models are accessible, comprehensible, reusable, and discoverable. Here we specifically use semantic annotation to refer to the association of biological knowledge with the relevant constituents of a

*Correspondence: d.nickerson@auckland.ac.nz

¹Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

Full list of author information is available at the end of the article



mathematical model in a computable and reproducible form. Such annotation does not, however, contribute to the mathematical content or interpretation of the model [5]. The biological knowledge added in this way consists of a wide range of contextual information such as species, genes, physiological processes, and anatomical locations. Integration of such information in the computational models requires a diverse set of reference ontologies and often combines concepts from multiple ontologies, i.e. composite annotation [6]. Manual annotation is a time consuming process. We therefore utilize SemGen [7], a semantic annotation tool, to annotate our cohort of exemplar models. To achieve this, SemGen combines concepts from a range of ontologies [8].

In order to discover relevant models, it is essential that models and their semantic annotations, along with other included metadata (e.g. authorship, original publication, etc), are findable and accessible by both people and computational services. In this manuscript, we focus on model discovery within the Physiome Model Repository (PMR) [9], but more broadly, our approach could be applied to a variety of model repositories or resources.

Knowledge discovery is essential within the pharmacology discipline. A recent initiative, Open PHACTS [10, 11], has used semantic querying to enable chemists to discover and explore chemical entities and/or structures. Similarly, other initiatives [12, 13] have developed discovery platforms to explore human diseases and associated genes, as well as other variants using the wealth of semantic resources now available. Pharmacogenomics [14] and biological [15] disciplines heavily depend on knowledge discovery with a view to accessing and integrating web of science resources.

In this paper we present a web-based tool for model discovery that uses the semantic annotations within PMR. Specifically, users can discover models encoded in the CellML format [16], which includes some useful biological information encapsulated within them: species, gene, anatomical location, etc. In the following section, we provide background on the tools and standards we leveraged to develop our web-based model discovery tool. Detailed analysis of the biological information as well as the web tool by which we retrieve this information are discussed in the “[Methods](#)” and “[Results](#)” sections.

Here we focus on models of epithelial transport as a way of scoping the implementation of our web application. The tools being assembled here into a workflow of model discovery and reuse are, however, of utmost importance when considering the increasingly vast amount of experimental and clinical data being made available. For instance, when considering epithelial Na⁺ transport, retrieving multiple computational models of the epithelial Na⁺ channel (ENaC) would be important when considering new experimental data of normal and

pathophysiological ENaC channel function since not all channel models may be able to represent the range of function observed. The ability to objectively select appropriate models both semantically and computationally and merge them across multiple scales of physiological complexity will allow researchers to build clinically-useful models that faithfully represent physiological function.

Tools and standards

Computational tools and standards [17, 18] have evolved over the years to store, exchange and comprehend computational models, including their comprehensive descriptions, i.e. semantic annotation, which make these resources more accessible and discoverable. Some of the software, tools and standards used in our model discovery approach are presented below.

RDF and SPARQL

The Resource Description Framework (RDF) is a technology for encoding annotations [19]. RDF is machine readable and allows semantic interoperability between different systems. It represents annotations with a set of triples: subject, predicate, and object. The subject references the element being annotated, the predicate defines the assertion that is being made, and the object is the value being asserted. The COmputational Modeling in BIology NEtwork [20] community has recently recommended the use of RDF for storing annotations on models [21].

The SPARQL Protocol and RDF Query Language (SPARQL) is a query language used to retrieve information from collections of RDF triples [22]. SPARQL and RDF are both widely used semantic web technologies, which together enable “semantic querying”. Our model discovery tool relies extensively on the execution of such semantic queries.

CellML

CellML is an eXtensible Markup Language (XML) format that encodes mathematical models of biological phenomena [16]. CellML models are modular and hierarchical, with model authors able to define abstract interfaces to their models by encapsulating complex details beneath a well defined interface. Model components can be reused with an import mechanism allowing external models to be accessible within the importing model. CellML includes the mathematical markup language for representing equations; and makes use of the RDF for the description of metadata such as biological annotations. The CellML metadata specification [23] provides guidelines on how to annotate the biological semantics encapsulated in a CellML model using RDF.

Ontologies

For our purpose, we have utilized the Chemical Entities of Biological Interest (ChEBI) ontology [24] to represent solutes such as sodium or hydrogen, the Ontology of

Physics for Biology (OPB) [25] to represent biophysical properties such as chemical concentration or fluid flow, and the Foundational Model of Anatomy (FMA) [26] to represent anatomical entities such as the cell nucleus or membrane. By utilizing these knowledge resources, we can create composite annotations that capture, in a computable fashion, specific properties simulated by a model such as concentration of sodium in the collecting duct of the kidney. In addition, we have leveraged the European Bioinformatics Institute's Ontology Lookup Service (OLS) [27] to retrieve human readable labels from reference ontology URIs.

SemGen

SemGen is a toolset for annotating, merging, and decomposing models [7]. It is platform-independent and open source. It is capable of working with models encoded in a variety of modeling formats, including the Systems Biology Markup Language (SBML) [28] and CellML. With SemGen, users can add composite annotations to a model that capture the biological meaning of a model's contents [6], including the variables in CellML models. These composite annotations link reference terms from knowledge resources such as the OPB, ChEBI, and FMA to form precise, machine-readable descriptions of model variables.

The SemGen annotator interface of the sodium/hydrogen exchanger 3 (NHE3) [29] model is shown in Fig. 1 wherein the left panel lists 39 CellML variables and 4 CellML components, and the right panel shows a composite annotation for "sodium flux from proximal tubule to epithelial cell cytosol via sodium/hydrogen exchanger 3", as well as the declaration of that variable in the CellML model code.

The Physiome model repository

The Physiome Model Repository (PMR) was initially designed to store and manage CellML models [9] as part of the IUPS Physiome Project [1]. In the PMR, CellML models can be annotated, exchanged, and reused. PMR also supports a distributed version control system (DVCS) by which users can keep track of the changes made and thereby share and rollback to any committed point. By doing so, it enhances collaboration among model developers. Figure 2 shows an example of synchronizing processes between client and server/cloud applications. In this case, tools like OpenCOR [30] and SemGen maintain local copies of relevant files which are then pulled by the DVCS service (Git).

The PMR has two primary data containers: workspaces and exposures. A workspace is a DVCS repository for a user that contains related information in the PMR. As discussed above, using a DVCS provides the ability to keep track of a number of commit points where users can easily switch to any specific point in the history

of a workspace. A workspace can be shared with other users and thereby facilitates collaborative development. A workspace can be synchronized with other compatible DVCS repositories hosted anywhere on the Internet. Private workspaces can be made publicly accessible with the approval of a repository curator. A workspace can be assembled with another workspace, i.e. a hierarchical workspace, which resembles the modularity and reusability features of CellML. One of the key features in a PMR workspace is the indexing of RDF triples into an RDF store so that users can access these triples using SPARQL queries.

Exposures are equivalent to software releases – they provide a specialized view of a workspace at a specific point in the workspace's history. Users can view the workspace in many ways dependent on the types of content in the workspace. For example, CellML models will have the following views generated: documentation, model curation and metadata, mathematics, source XML, and open with OpenCOR. Exposures are private by default and can be accessible after approval from a repository curator. Exposures can be migrated from one revision of the associated workspace to another as well as a separate "clone" of the workspace that another user might have created when reusing a model.

The PMR also provides web services to address the growing complexity of cross-platform applications through the exchange of messages encapsulated in the HTTP protocol. PMR uses the Open Authorization (OAuth) standard to establish the authentication of a user or consumer and to manage access privileges for services provided by the PMR server. The PMR web services exchange information using the JavaScript Object Notation (JSON) and consumers have to follow specific guidelines [31] to fetch information from a PMR instance. One of the services provided by PMR is a read-only SPARQL endpoint, which can be utilised by application developers to submit SPARQL queries.

Methods

To demonstrate the value of the semantics-based approaches we propose here, we have developed a platform which allows the user to discover and explore models of interest and then display the annotated information via a web-based interface. We do this in the context of epithelial transport, although as mentioned above these technologies are broadly applicable across the spectrum of biomedical modelling. To provide an initial cohort of models and model entities for use in developing our tool, we first annotated the biological semantics of a set of models in the CellML format using the SemGen tool. These semantic annotations were deposited in PMR to provide resources to discover and assemble on the web interface. Detailed analysis of this process is described below

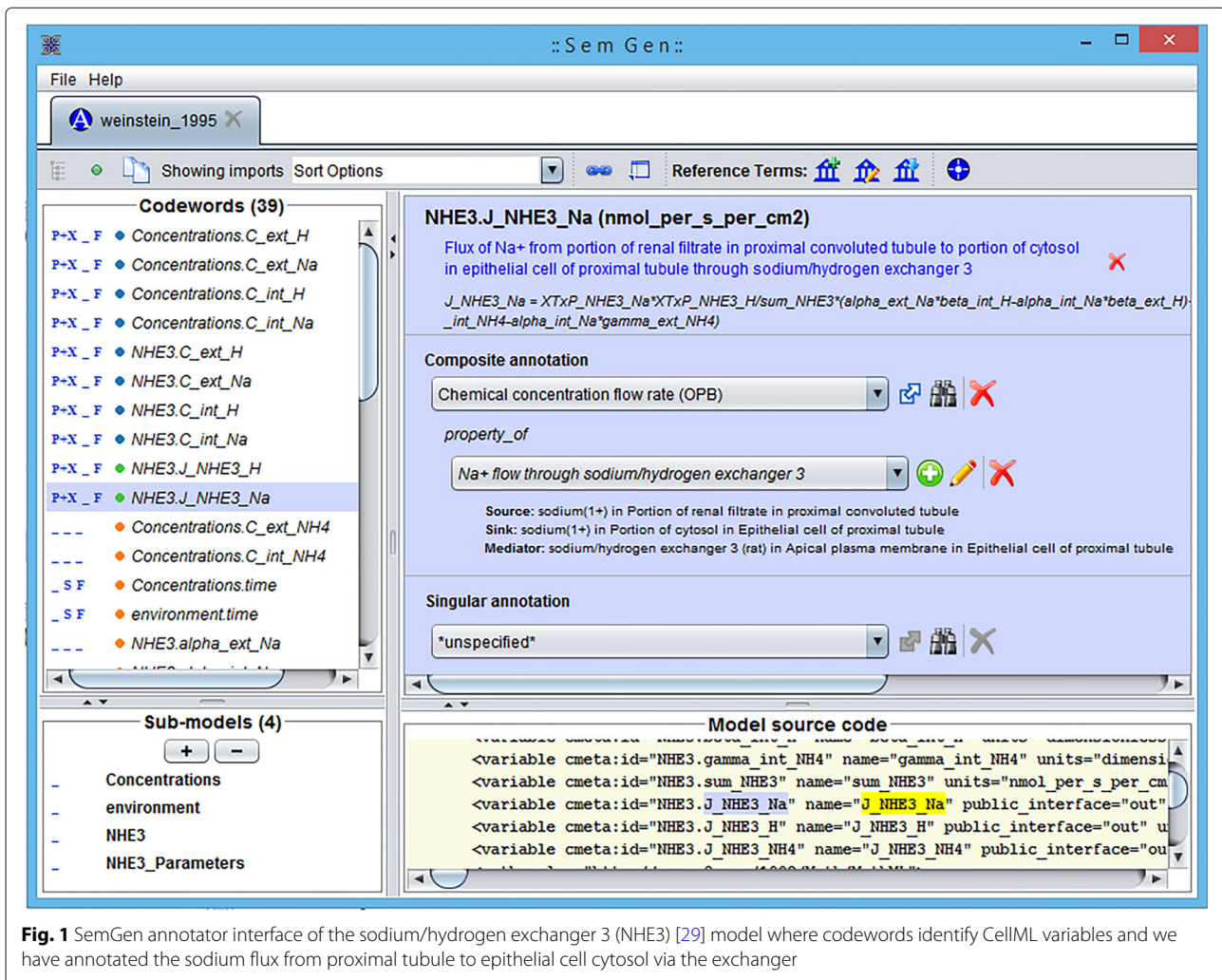


Fig. 1 SemGen annotator interface of the sodium/hydrogen exchanger 3 (NHE3) [29] model where codewords identify CellML variables and we have annotated the sodium flux from proximal tubule to epithelial cell cytosol via the exchanger

and the model discovery tool can be accessed via <https://github.com/dewancse/model-discovery-tool>.

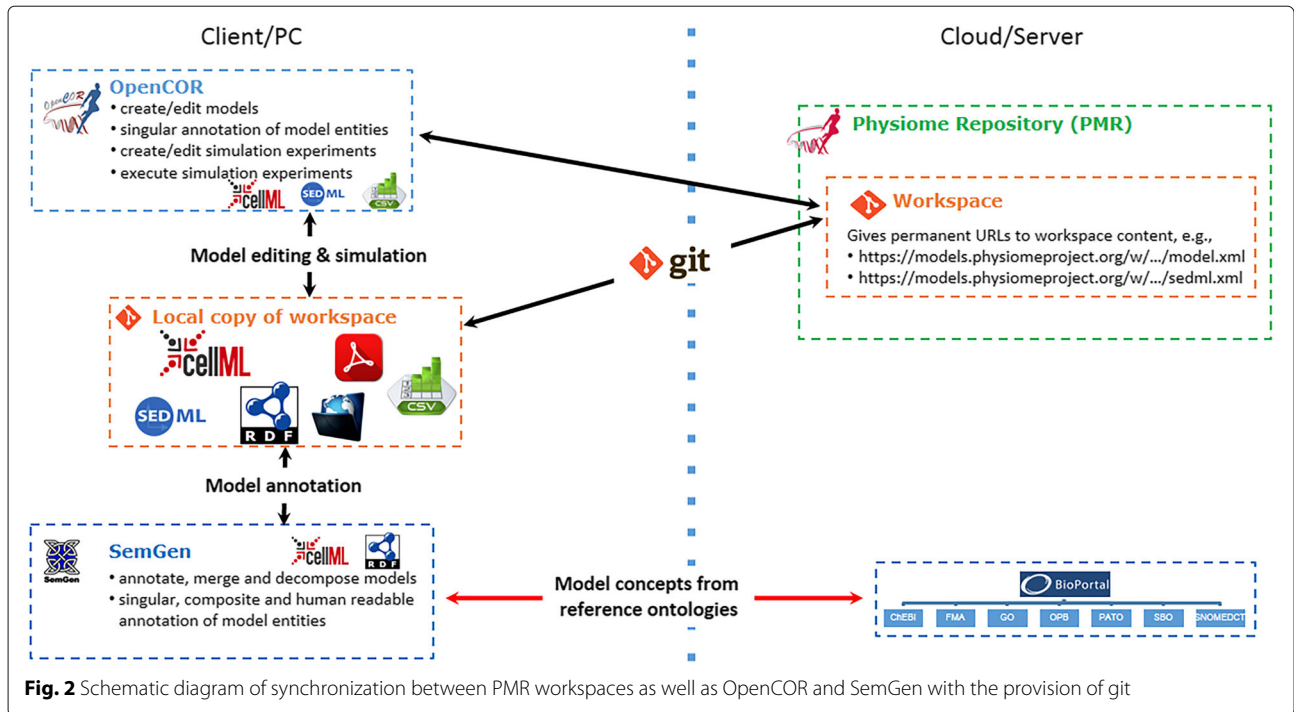
Biological annotation

We annotated a cohort of 33 models on the PMR with biological semantics. This included 12 renal, 11 cardiac, 4 lung, 2 musculoskeletal and 4 miscellaneous epithelial transport models. In particular, we extracted each of the mathematical variables from the models and associated them with biologically meaningful information so that computational modellers and clinicians can discover models relevant to their investigations. These models are all encoded in the CellML format and can be accessed via <https://models.physiomeproject.org/workspace/527>.

This initial cohort of well-annotated models serves as a proof of concept for testing our implementation. According to the recent community agreement on the harmonization of annotation across the computational modelling in biology community [21], we expect the repository of available annotated models to rapidly grow

as the community begins to populate this repository. With community adoption of the harmonized annotation guidelines, we will be able to utilize our platform to query across all available repositories and modelling formats.

Resources (variables, components, etc.) in these models can be uniquely identified with URIs composed of the CellML file name (including relative or absolute paths in the PMR) and the document-unique identifiers defined by the CellML format (so-called metadata identifiers). For example, when storing annotations in the same PMR workspace as the NHE3 [29] model, the URI `weinstein_1995.cellml#NHE3.J_NHE3_Na` unambiguously links to a variable defining the sodium flux through the NHE3 transporter. This example has been annotated in human readable text as “sodium flux from proximal tubule to epithelial cell cytosol via sodium/hydrogen exchanger 3”. Figure 1 shows the SemGen annotator interface for this example annotation formulated as a composite annotation [6]. This demonstration combines concepts from

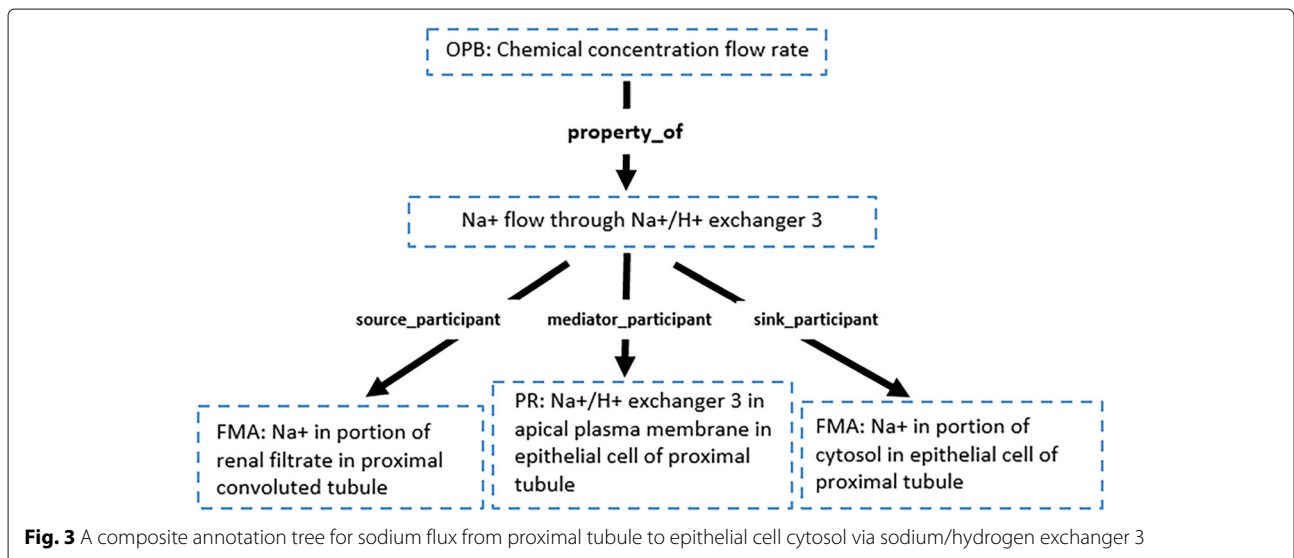



multiple reference ontologies – OPB, FMA and the Protein Ontology (PR). The detailed composite annotation has been presented in Fig. 3.

Results

In this section we present our web-based interface, Model Discovery Tool, that enables users to discover and explore the annotated biological information in order to investigate specific questions and hypotheses.

We have developed a web-based tool to discover and explore models of interest using information extracted from the PMR. Figure 4 shows a list of discovered models for a search query "flux of sodium" from the PMR. From this list, the user can investigate various options such as CellML model entities – model name, component name and variable name, biological meaning deposited in PMR, protein name, and species and genes used during experiments of the associated models.




DOCUMENTATION

Discover CellML Models

Model Entity (click a model entity to expand)	Biological Meaning	Species	Gene	Protein
chang_fujita_b_1999.cellml#sc_sodium_flux.J_sc_Na	Flux of Na+ from portion of cytosol in epithelial cell of distal tubule to portion of tissue fluid in epithelial cell of distal tubule through Na(+)/K(+) ATPase alpha-1 subunit	Rattus norvegicus	Atp1a1	sodium/potassium-transporting ATPase subunit alpha-1
mackenzie_1996.cellml#NBC_current.J_Na	Flux of Na+ from portion of renal filtrate in proximal convoluted tubule to portion of cytosol in epithelial cell of proximal tubule and portion of cytosol in epithelial cell of proximal tubule to portion of renal filtrate in proximal convoluted tubule through apical plasma membrane	Mus musculus	Slc5a4a	low affinity sodium-glucose cotransporter
weinstein_1995.cellml#NHE3.J_NHE3_Na	Flux of Na+ from portion of renal filtrate in proximal convoluted tubule to portion of cytosol in epithelial cell of proximal tubule through sodium/hydrogen exchanger 3	Rattus norvegicus	Slc9a3	sodium/hydrogen exchanger 3

sodium/hydrogen exchanger 3 is a **Kidney** model. It is located in **epithelial cell of proximal tubule, apical plasma membrane, proximal convoluted tubule**.

Model: [weinstein_1995.cellml#NHE3.J_NHE3_Na](#)

Biological Meaning: Flux of Na+ from portion of renal filtrate in proximal convoluted tubule to portion of cytosol in epithelial cell of proximal tubule through sodium/hydrogen exchanger 3

Species: Rattus norvegicus

Gene: Slc9a3

Protein: sodium/hydrogen exchanger 3

Compartment: Portion of cytosol, Apical plasma membrane, Luminal

Location: Epithelial cell of proximal tubule, Apical plasma membrane, Proximal convoluted tubule

Recommendations/suggestions based on existing models in PMR

Basolateral membrane model

- [sodium/hydrogen exchanger 3 \(human\)](#)
- [low affinity sodium-glucose cotransporter \(mouse\)](#)
- [sodium/potassium-transporting ATPase subunit alpha-1 \(rat\)](#)

Alternative model of sodium/hydrogen exchanger 3

Not Exist

Kidney model in PMR

- [sodium/hydrogen exchanger 3 \(human\)](#)
- [low affinity sodium-glucose cotransporter \(mouse\)](#)
- [epithelial cell](#)

Protein name: epithelial cell
 Protein uri: http://purl.obolibrary.org/obo/CL_0000066
 Mediator name: sodium/potassium-transporting ATPase subunit alpha-1 (rat)
 Mediator uri: http://purl.obolibrary.org/obo/PR_P06685
 Similarity value: 13.2
 Model entity: chang_fujita_b_1999.cellml#sc_sodium_flux.J_sc_Na
 Model entity2:

Fig. 4 A use case application to search for models in PMR which are relevant to the query 'flux of sodium'. For convenience, we have displayed here top three components and the user can scroll down to see other components. By querying the knowledge in PMR this tool retrieves components from the NHE3 model [29], SGLT2 model [32], and the epithelial cell model [33]. Recommendations of models similar to the NHE3 model [29] are displayed as a result of the user selecting the NHE3 sodium flux in the returned results. See text for further details

Users are able to enter human-readable search terms and phrases or choose to enter specific vocabulary terms to use in searches. When discovering resources in PMR relevant to a given search phrase, we first decompose the text into a series of semantic queries which are executed on the PMR knowledgebase and ranked according to how well the discovered resource matches the initial search query. Continuing our example, the phrase shown in Fig. 4 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 tool will map flux and sodium to OPB and ChEBI ontology terms, respectively, as defined in the dictionary.

In addition, the user can click on a CellML model entity to explore more information as a recommendation or suggestion. As the name implies, this recommendation

system suggests for similar or alternative models they users wish to explore or utilize in their model. Machine learning based recommender systems are widely used to build knowledge management systems enabling the user to interact with the system [34–40]. However, our recommender system works in a different manner: it suggests relevant models based on semantic annotations within the PMR and the model the user is currently working with.

Specifically, when the user clicks on a model-entity, the recommender system displays relevant models in the PMR. Initially, it provides a brief overview of the selected model entity: its anatomical location, biological meaning, species, gene, and protein name, as shown in Fig. 4. In this system, we have added a tooltip feature to quickly view some useful information such as protein name and its URI, mediator name and its URI if it is a co-transporter, CellML model entity, and a link to navigate to the associated CellML model.

This information is followed by a list of relevant models in the PMR with respect to the selected model. For example, if the selected model is located in the basolateral membrane, then existing basolateral model(s) are listed as similar models. This populated list is ranked in an attempt to ensure more relevant models are closer to the top of the list.

As discussed above, we focus here on epithelial transport models which are often models of specific membrane-bound transport proteins. To determine which models in the repository are more relevant, we use the WSDbfetch [41] and Clustal Omega [42, 43] services freely available from the European Bioinformatics Institute. For our purpose, we used WSDbfetch to retrieve a protein sequence for a given protein identifier, and the Clustal Omega service to obtain alignment metrics among multiple protein sequences calculated as a similarity matrix.

To get this matrix, we first fetched relevant protein models' identifier located in the basolateral membrane. Next, we sent these protein identifiers to the WSDbfetch service to retrieve protein sequences. Finally, we sent these protein sequences to the Clustal Omega service to get the similarity matrix. An example of such a similarity matrix is shown in Fig. 5. Along with the biological semantics, the similarity matrix provides a quantitative measure to help rank related models. In this manner, we hope to present the user with the most relevant models first. While these metrics are specific to membrane-bound transport proteins, we envision similar domain-relevant metrics can be developed as our methods are applied within of other domains.

An alternative option shown in Fig. 4 is also provided for the user to explore the same selected model with different

species and from various workspaces. The user could also explore relevant organ models and could navigate to the associated workspaces for further investigation.

Discussion

We have developed a novel semantic annotation and model discovery methodology with the aid of modern tools and standards and implemented a web-based tool build on this methodology. We have made available in PMR an initial cohort of models in the CellML format annotated with biological semantics for use in demonstrating the utility of our methodology and implemented platform.

In our methodology, we make cascading SPARQL calls to explore CellML models in the PMR. These calls first explore model name and biological meaning of a search text string. This text then maps to a static dictionary as key and value pairs with the help of reference ontologies such as the OPB and ChEBI. Subsequent calls fetch protein identifiers associated with a model, and these identifiers are used to retrieve the species and associated gene names for the protein through calls to the OLS. We have also developed a recommender system which leverages semantic annotations in the PMR to identify biologically-similar models.

A limitation of our discovery implementation is that we use a static dictionary with key and value pairs to map searched text to reference ontology terms, i.e. exact matching. However, this could be enhanced with Natural Language Processing (NLP) techniques described previously [44, 45] and by making SPARQL queries to the services provided by the European Bioinformatics Institute. For example, for the text “flux of sodium”, NLP will parse the text into pieces with tokenization, lemmatization, and part of speech techniques and then SPARQL queries could map these pieces onto reference ontology terms. For convenience, the application could also recommend similar words when the user misspells.

Conclusion

Using semantic annotations to discover models takes into account more contextual information when querying the PMR to provide more relevant results to the user. Using the standard text-based search functionality of PMR, a query for “flux of sodium” returns seven items (5 CellML models and 2 exposure documentation pages). In each case the user is required to navigate to each of those items and determine if they are relevant to their needs. Using our discovery tool, the same query returns over 40 specific entities in our exemplar cohort of CellML models that have been annotated to have some association with “flux of sodium”. These 40 entities, from 12 different models, are ranked so that model variables representing a specific flow of sodium through a cell membrane are nearest the top

```

Identity Matrix: #
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: sp|Q9ET37|S5A4A_MOUSE 100.00 22.86 17.86 21.86
2: sp|P48764|SL9A3_HUMAN 22.86 100.00 19.15 89.49
3: sp|P06685|AT1A1_RAT 17.86 19.15 100.00 18.20
4: sp|P26433|SL9A3_RAT 21.86 89.49 18.20 100.00

```

Fig. 5 Similarity matrix from the EBI Clustal omega service where protein with UniProt ID P26433 is in the selected model and the rest of the proteins are the basolateral membrane models shown in Fig. 4, (under the "Basolateral membrane model" heading) whose IDs and matrix scores are as follows: P48764 (89.49), Q9ET37 (21.86) and P06685 (18.20)

of the list, and less specific fluxes are nearer the bottom. As shown in Fig. 4, the semantic annotations make it possible to display to the user key information to help them determine relevancy to their work as well as providing recommendations on similar models they may want to investigate.

In addition to the discovery approach, we plan to extend the web application into an Epithelial Modelling Platform that facilitates the reuse and recomposition of epithelial models and their components on the PMR. Our ultimate goal is to create a platform whereby researchers can easily assemble customized models tailored for their specific research needs. The model-discovery tool described here, along with its underlying methodology, represent crucial first steps in the development of such a platform: they offer a solution for identifying and prioritizing modeling components based on a user's research interests.

Abbreviations

ChEBI: Chemical entities of biological interest; COMBINE: Computational Modeling in Biology Network; DVCS: Distributed version control system; ENaC: Epithelial sodium channel; FMA: Foundational Model of Anatomy; IUPS: International Union of Physiological Sciences; JSON: JavaScript object notation; NHE3: Sodium/hydrogen exchanger 3; NLP: Natural language processing; OAuth: Open authorization; OLS: Ontology lookup service; OPB: Ontology of physics for biology; PMR: Physiome model repository; PR: Protein ontology; RDF: Resource description framework; SBML: Systems biology markup language; SPARQL: SPARQL protocol and RDF query language; URI: Uniform resource identifier; VPH: Virtual Physiological Human; XML: eXtensible markup language

Acknowledgements

We would like to thank the Semantics of Biological Processes lab at the University of Washington for valuable discussions and feedback during the April 2018 meeting in Seattle. 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.

Authors' contributions

DMS, DPN and KA prepared the manuscript; DMS and DPN implemented the model discovery tool; DMS, DPN, KA, JHG, BEC, BdB, and MLN prepared the biological annotation; DPN, RK, PJH, JHG, BEC, and MLN edited the manuscript. All authors read and approved the final manuscript.

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. The funding bodies did not play any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

Project name: Model Discovery Tool Project home page: <https://github.com/dewancse/model-discovery-tool> Operating system(s): Platform independent Programming language: JavaScript, HTML License: Apache 2.0.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand. ²Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington, USA. ³Molecular & Integrative Physiology, University of Michigan, Ann Arbor, Michigan, USA. ⁴Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, Washington, USA.

Received: 24 March 2019 Accepted: 9 July 2019

Published online: 06 September 2019

References

- Hunter PJ, Borg TK. Integration from proteins to organs: the Physiome Project. *Nat Rev Mol Cell Biol.* 2003;4:237–43. <https://doi.org/10.1038/nrm1054>.
- Hunter P, Coveney PV, de Bono B, Diaz V, Fenner J, Frangi AF, Harris P, Hose R, Kohl P, Lawford P, McCormack K, Mendes M, Omholt S, Quarteroni A, Skår J, Tegner J, Thomas SR, Tollis I, Tsamardinos I, van Beek JHGM, Viceconti M. A vision and strategy for the virtual physiological human in 2010 and beyond. *Philos Trans A Math Phys Eng Sci.* 2010;368:2595–614. <https://doi.org/10.1098/rsta.2010.0048>.
- Nielsen PF, Nickerson DP. VPHi Webinar: Reproducibility and reuse in the Virtual Physiological Human. 2017. <http://www.vph-institute.org/webinar/reproducibility-and-reuse-in-the-virtual-physiological-human.html>. Accessed 24 Oct <https://doi.org/10.17608/k6.auckland.5552746.v1>.
- Neal ML, Cooling MT, Smith LP, Thompson CT, Sauro HM, Carlson BE, Cook DL, Gennari JH. A reappraisal of how to build modular, reusable models of biological systems. *PLOS Comp Biol.* 2014;10:1003849. <https://doi.org/10.1371/journal.pcbi.1003849>.

5. Lister AL, Pocock M, Taschuk M, Wipat A. Saint: a lightweight integration environment for model annotation. *Oxf Bioinforma*. 2009;25:3026–7.
6. Gennari JH, Neal ML, Galdzicki M, Cook DL. Multiple ontologies in action: Composite annotations for biosimulation models. *J Biomed Informa*. 2011;44:146–54. <https://doi.org/10.1016/j.jbi.2010.06.007>.
7. Neal ML, Thompson CT, Kim KG, James RC, Cook DL, Carlson BE, Gennari JH. Semgen: a tool for semantics-based annotation and composition of biosimulation models. *Bioinformatics*. 2018;32:29. <https://doi.org/10.1093/bioinformatics/bty829>.
8. Neal ML. Modular, semantics-based composition of biosimulation models. Ph.D. Dissertation: University of Washington; 2010. <https://search.proquest.com/docview/753931061>. Accessed 21 Jan 2019.
9. Yu T, Lloyd CM, Nickerson DP, Cooling MT, Miller AK, Garry A, Terkildsen JR, Lawson J, Britten RD, Hunter PJ, Nielsen PMF. The Physiome Model Repository 2. *Bioinformatics*. 2011;27:743–4. <https://doi.org/10.1093/bioinformatics/btq723>.
10. JWilliams A, Groth LHP, Pettifer S, Chichester C, Willighagen EL, Evelo CT, Blomberg N, Ecker G, Goble C, Mons B. Open PHACTS: semantic interoperability for drug discovery. *Drug Discov Today*. 2012;17:1188–98. <https://doi.org/10.1016/j.drudis.2012.05.016>.
11. Gray AJG, Groth P, Loizou A, Askjaer S, Brenninkmeijer C, Burger K, Chichester C, Evelo CT, Goble C, Harland L, Pettifer S, Thompson M, Waagmeester A, Williams AJ. Applying linked data approaches to pharmacology: Architectural decisions and implementation. *Semant Web*. 2014;5(2):101–13. <https://doi.org/10.3233/SW-2012-0088>.
12. Piñero J, Àlex Bravo, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2017;45:833–9. <https://doi.org/10.1093/nar/gkw943>.
13. Piñero J, Queralt-Rosinach N, Àlex Bravo, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxf)*. 2015;2015:028. <https://doi.org/10.1093/database/bav028>.
14. Dumontier M, Villanueva-Rosales N. Towards pharmacogenomics knowledge discovery with the semantic web. *Brief Bioinform*. 2009;10(2): 153–63. <https://doi.org/10.1093/bib/bbn056>.
15. Goble C, Wolstencroft K, Goderis A, Hull D, Zhao J, Alper P, Lord P, Wroe C, Belhajjame K, Turi D, Stevens R, Oinn T, Roure DD. *Knowledge Discovery for Biology with Taverna*. Boston: Springer; 2007, pp. 355–95. https://doi.org/10.1007/978-0-387-48438-9_17.
16. Cuellar AA, C.M. L., P.F. N., D.P. B., Nickerson DP, P.J. H. An overview of CellML 1.1, a biological model description language. *SIMULATION*. 2003;79:740–7. <https://doi.org/10.1177/0037549703040939>.
17. Nickerson D, Atalag K, de Bono B, Geiger J, Goble C, Hollmann S, Lonien J, Müller W, Regierer B, Stanford NJ, Golebiewski M, Hunter P. The Human Physiome: how standards, software and innovative service infrastructures are providing the building blocks to make it achievable. *Interface Focus*. 2016;6. <https://doi.org/10.1098/rsfs.2015.0103>.
18. Nickerson DP, Atalag K, de Bono B, Hunter PJ. *The Physiome Project, openEHR Archetypes, and the Digital Patient*. Hoboken: Wiley; 2016, pp. 101–25. <https://doi.org/10.1002/9781118952788.ch9>.
19. Klyne G, Carroll JJ, McBride B. RDF 1.1 Concepts and Abstract Syntax. W3C Recomm. 2019. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>. Accessed 21 Jan.
20. COMBINE. The COmputational Modeling in Biology NEtwork (combine). 2019. <http://co.mbine.org/>. Accessed 21 Jan.
21. Neal ML, König M, Nickerson D, Mısırlı G, Kalbasi R, Dräger A, et al. Harmonizing semantic annotations for computational models in biology. *Brief Bioinforma*. 2018;087. <https://doi.org/10.1093/bib/bby087>.
22. Prud'hommeaux E, Seaborne A. SPARQL Query Language for RDF. W3C Recomm. 2019. <https://www.w3.org/TR/rdf-sparql-query/>. Accessed 21 Jan.
23. Cooling MT, Hunter P. The CellML Metadata Framework 2.0 Specification. *J Integr Bioinforma JIB*. 2015;12(2). <https://doi.org/10.2390/biecoll-jib-2015-260>.
24. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res*. 2015;44(D1):1214–9. <https://doi.org/10.1093/nar/gkv1031>.
25. Cook DL, Bookstein FL, Gennari JH. Physical Properties of Biological Entities: An Introduction to the Ontology of Physics for Biology. *PLoS ONE*. 2011;6:28708. <https://doi.org/10.1371/journal.pone.0028708>.
26. Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003;36(6):478–500. <https://doi.org/10.1016/j.jbi.2003.11.007>.
27. OLS. EBI Ontology Lookup Service. 2019. <https://www.ebi.ac.uk/ols/index>. Accessed 21 Jan.
28. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, the rest of the SBML Forum: Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-H, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Nov'ere NL, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinforma Oxf J*. 2003;19:524–31.
29. Weinstein AM. A kinetically defined Na⁺/H⁺ antiporter within a mathematical model of the rat proximal tubule. *J Gen Physiol*. 1995;105(1): 617–41. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2216949/>.
30. Garry A, Hunter PJ. OpenCOR: a modular and interoperable approach to computational biology. *Front Physiol*. 2015;6(26). <https://doi.org/10.3389/fphys.2015.00026>.
31. Yu T. Webservice for the Physiome Model Repository. 2019. <http://aucklandphysiomerepository.readthedocs.io/en/latest/webservice.html>. Accessed 21 Jan.
32. Mackenzie M, Loo DDF, Panayotova-Heiermann M, Wright EM. Biophysical characteristics of the pig kidney na/glucose cotransporter sglT2 reveal a common mechanism for sglT1 and sglT2. *J Biol Chem*. 1996;271:32678–83. <https://doi.org/10.1074/jbc.271.51.32678>.
33. Chang H, Fujita T. *Am J Physiol*. 1999;276(6):931–51. <https://www.ncbi.nlm.nih.gov/pubmed/10362782>.
34. Bobadilla J, Ortega F, Hernando A, Gutiérrez A. Recommender systems survey. *Knowl Based Syst*. 2013;46:109–32. <https://doi.org/10.1016/j.knsys.2013.03.012>.
35. Almazro D, Shahatah G, Albdulkarim L, Khrees M, Martinez R, Nzoukou W. A survey paper on recommender systems. *Cornell Univ Libr*. 2010. <https://doi.org/10.1006/1006.5278> [cs.IR].
36. Kywe SM, Lim E.-P., Zhu F. A Survey of Recommender Systems in Twitter. *Berlin, Heidelberg: Springer*; 2012, pp. 420–33. https://doi.org/10.1007/978-3-642-35386-4_31.
37. Carrer-Neto W, Hernández-Alcaraz ML, Valencia-García R, García-Sánchez F. Social knowledge-based recommender system. Application to the movies domain. *Expert Syst Appl*. 2012;39:10990–1000. <https://doi.org/10.1016/j.eswa.2012.03.025>.
38. Ricci F, Rokach L, Shapira B. *Introduction to Recommender Systems Handbook*. Boston, MA: Springer; 1990, pp. 1–35. https://doi.org/10.1007/978-0-387-85820-3_1.
39. Wiesner M, Pfeifer D. *Health Recommender Systems: Concepts, Requirements, Technical Basics and Challenges*. *Int J Environ Res Public Health*. 2014;11(3):2580–607. <https://doi.org/10.3390/ijerph110302580>.
40. Lémdani R, Polaillon G, Bennacer N, Bourda Y. A semantic similarity measure for recommender systems. In: *I-Semantics '11 Proceedings of the 7th International Conference on Semantic Systems*; 2011. p. 183–86. <https://doi.org/10.1145/2063518.2063545>.
41. WSDbfetch. EBI webservice to get entries from various biological databases. 2019. <https://www.ebi.ac.uk/seqdb/confluence/pages/viewpage.action?pageId=48923608>. Accessed 21 Jan.
42. Clustal. EBI webservice to get multiple sequence alignment. 2019. https://www.ebi.ac.uk/seqdb/confluence/display/WEBSERVICES/clustalo_rest. Accessed 21 Jan.
43. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol*. 2011;7(1). <https://doi.org/10.1038/msb.2011.75>.
44. Sfakianaki P, Koumakis L, fakianakis S, Tsiknakis M. Natural Language Processing for Biomedical Tools Discovery: A Feasibility Study and Preliminary Results. *Cham: Springer*; 2014, pp. 134–145. https://doi.org/10.1007/978-3-319-06695-0_12.

45. Sfakianaki P, Koumakis L, Sfakianakis S, Iatraki G, Zacharioudakis G, Graf N, Marias K, Tsiknakis M. Semantic biomedical resource discovery: a Natural Language Processing framework. *BMC Med Inf Decis Mak.* 2015;15(1):77. <https://doi.org/10.1186/s12911-015-0200-4>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

