

Computer Support for Physiological Cell Modelling using an Ontology on Cell Physiology

SHIMAYOSHI Takao, KOMURASAKI Kazuhiro, AMANO Akira,
IWASHITA Takeshi, KANAZAWA Masanori and MATSUDA Tetsuya

Abstract—The development of electrophysiological whole cell models to support the understanding of biological mechanisms is increasing rapidly. Due to the complexity of biological systems, comprehensive cell models, which are composed of many imported sub-models of functional elements, can get quite complicated as well, making computer modification difficult. Here, we propose a computer support to enhance structural changes of cell models, employing the markup languages CellML and our original PMSML (Physiological Model Structure Markup Language), in addition to a new ontology for cell physiological modelling. In particular, a method to make references from CellML files to the ontology and a method to assist manipulation of model structures using markup languages together with the ontology are reported. Using these methods three software utilities, including a graphical model editor, are implemented. Experimental results proved that these methods are effective for the modification of electrophysiological models.

I. INTRODUCTION

To elucidate the mechanisms of life, it is necessary to understand the functions of each single cell which is an elementary unit of biological activities. For the representation of cellular functions, several electrophysiological models of whole cells [1]–[4] have recently been developed. Whole cell models are constructed of multiple component models of functional elements, e.g. functional protein complexes such as membrane channels and pumps, and abstract cellular functions such as background currents and ATP production. Generally, these whole cell models are expanded by the following process. First, an existing component model of a new functional element is inserted, and then the whole cell model is adjusted to accommodate the new element. Next, the behaviour of the whole cell model is examined by simulation. Depending on these simulation results, the component model is improved or exchanged.

Comprehensive whole cell models can get quite complicated because they are an assembly of many functional elements. Since many researchers in the physiological field find it hard to edit such complex model structures in conventional computer languages, a tool, which allows an

easy modification of comprehensive models, is important to accelerate physiological modelling.

Here, we propose novel methods using markup languages and an ontology, which support the modification of complicated electrophysiological models by importing multiple component models. First, the specifications of the markup languages and an ontology suitable to deal with multiple models in an integrated way are discussed. Then, methods to identify components of existing models to an entry of the ontology and to assist manipulations of model structures are explained and the implementations described. Finally, experimental results of the methods for electrophysiological models are reported.

II. BACKGROUND

Several kinds of software are available to edit electrophysiological models [5], [6]. However, the interfaces of some existing editors, such as character-based ones are not intuitive for physiological researchers. Other editors require enormous manual operations for modifications of model structures with importing multiple component models, e.g. for the insertion or replacement of a functional element, lacking variables need to be added and new connections need to be made to existing variables which influence the element. These demanding tasks induce errors such as wrong connections of an element to its variables. Moreover, complicated models, which have many elements and a complex relationship among them, have a low visibility, making editing even more difficult.

On the other hand, an XML-based description language for cell physiological models, CellML [7] was proposed by Cueller, et al. A CellML file includes a definition and bibliographic metadata of a model. CellML is an useful representation technique for an editor to process and store models. In fact, several software packages [5], [6], [8], [9] in the cell modelling field support CellML. Moreover, a model repository described by CellML has been constructed. However, since the original purpose of CellML was the description of published models, it is not well suitable for addition and replacement of model elements.

Finally, integration of multiple models is further complicated through the fact that different terms are often assigned to the physiologically same element, and especially many CellML files use different identification names for the physiologically same element. It is essential to identify all names given to a model component.

This work was supported by Leading Project for Biosimulation, Ministry of Education, Culture, Sports, Science and Technology, Japan.

Shimayoshi T. is with ASTEM Research Institute of Kyoto, Kyoto, 600–8813, Japan (corresponding author to provide phone: +81-75-326-1367; fax: +81-75-315-3198; e-mail: simayosi@astem.or.jp).

Komurasaki K. was with Graduate School of Informatics, Kyoto University, Kyoto, 600–8501, Japan.

Amano A. and Matsuda T. are with Graduate School of Informatics, Kyoto University, Kyoto, 600–8501, Japan.

Iwashita T. and Kanazawa M. are with Academic Center for Computing and Media Studies, Kyoto University, Kyoto, 600–8501, Japan.

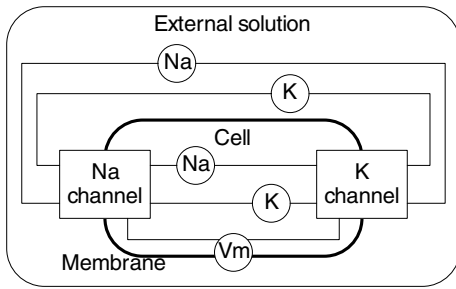


Fig. 1. An Example of a Model Structure

```
<?xml version="1.0"?>
<model>
  <compartment name="xsol" desc="external solution">
    <reactant name="Na" desc="sodium concentration"/>
    <reactant name="K" desc="potassium concentration"/>
  </compartment>
  <compartment name="cell" desc="ventricular cell">
    <reactant name="Na" desc="sodium concentration"/>
    <reactant name="K" desc="potassium concentration"/>
  </compartment>
  <separator name="membrane" desc="cell membrane">
    <reactant name="Vm" desc="membrane potential"/>
  </separator>
  <instance type="km2003.xml#sodium_current"
    name="sodiumChannel">
    <reference name="Vm" link="[@name='Vm']"/>
    <reference name="Na_i" link="../[@name='Na']"/>
    <reference name="Na_o" link="../..[@name='Na']"/>
    <reference name="K_i" link="../[@name='K']"/>
    <reference name="K_o" link="../..[@name='K']"/>
  </instance>
  <instance type="km2003.xml#pottasium_current"
    name="potassiumChannel">
    <reference name="Vm" link="[@name='Vm']"/>
    <reference name="Na_i" link="../[@name='Na']"/>
    <reference name="Na_o" link="../..[@name='Na']"/>
    <reference name="K_i" link="../[@name='K']"/>
    <reference name="K_o" link="../..[@name='K']"/>
  </instance>
</separator>
</compartment>
</model>
```

Fig. 2. An Example of a PMSML file

III. METHODS

A. Description Format

To improve CellML for the usage of model modifications, we are developing the description language PMSML (Physiological Model Structure Markup Language) as a supplement to CellML, which is extracted from CPSML [10] for a clear objective and improved. In a PMSML file, the relations of model components, which include functional elements and variables such as ion concentrations and the membrane potential, and the anatomical hierarchy of a model, which is important for the recognition of the model structure, is clearly defined with a separation from the details of the model.

PMSML describes an anatomical hierarchy as an XML tree composed of 'compartment' and 'separator' tags. A functional element is marked in the tree by an 'instance' tag with a link to a CellML file that defines the element. Global variables, which connect multiple functional elements, are enclosed by 'reactant' tags. Referred variables of a functional element are specified by 'reference' tags. Fig. 1 shows an example of a simple model. Fig. 2 depicts an example of the description in PMSML.

B. Ontology

A sensible solution of ambiguous terminologies is ontology. An ontology semantically defines a terminology and

```
<component name="fast_sodium_current" cpmo:id="520">
  <variable name="i_Na" public_interface="out" units="microA_per_microF" cpmo:id="220"/>
  <variable name="E_Na" public_interface="out" units="millivolt" cpmo:id="203"/>
  <variable name="g_Na" initial_value="16.0" units="milliS_per_microF"/>
  <variable name="time" public_interface="in" private_interface="out" units="millisecond" cpmo:id="010"/>
  <variable name="v" public_interface="in" private_interface="out" units="millivolt" cpmo:id="202"/>
  <variable name="R" public_interface="in" units="joule_per_kilomole_kelvin" cpmo:id="006"/>
  <variable name="F" public_interface="in" units="coulomb_per_mole" cpmo:id="008"/>
  <variable name="T" public_interface="in" units="kelvin" cpmo:id="007"/>
  <variable name="Nai" public_interface="in" units="millimolar" cpmo:id="211"/>
  <variable name="Nao" public_interface="in" units="millimolar" cpmo:id="101"/>
  <variable name="m" private_interface="in" units="dimensionless"/>
  <variable name="h" private_interface="in" units="dimensionless"/>
  <variable name="j" private_interface="in" units="dimensionless"/>
</component>
```

Fig. 3. A Fragment of a CellML file with Ontology IDs

includes a taxonomy and relations of the terms. For model structure modification, ontology is powerful for not only identifying model components, but also allocating model components to an anatomical hierarchy as well as integrating information about major dependencies of functional elements on global variables. Although many ontologies in biology are constructed, these ontologies are not suitable for our purpose. For example, Gene Ontology [11], a major ontology includes a database with numerous protein molecules, which is inadequate for physiological models, which often include abstract functional elements and ion currents with not yet specified molecular substances.

Here, a new ontology for cell physiological modelling is introduced. This ontology describes the information about identifications, major dependencies and hierarchical locations of model components common in physiological models. Each entry contains the information for one physiological model component and includes the identification number, the standard name, the standard unit, and terminological keywords of the model component, as well as ID numbers of a model component that anatomically contains the model component, and ID numbers of variables that the model component generally depends on.

C. Identification of CellML Elements

In CellML files physiological model components marked by 'component' or 'variable' tags do not have standardized names as mentioned earlier. Therefore it is necessary to associate each element of a CellML file with an entry in the ontology. This is achieved by assigning an ontology ID number to each CellML element. An ID number is appended as an 'cpmo:id' attribute to a 'component' tag or a 'variable' tag defined with a 'public_interface' attribute. However, an ID is not allocated to model-specific components and variables. Fig. 3 shows an example of a CellML file containing ontology ID numbers.

The identification of CellML elements is supported by estimations using lexical and structural analysis. The estimations of 'component' elements are achieved by the following

algorithm. First, all ontology entries are scored for each component if the name of the component includes the keywords of the entry, and/or if the name of ‘public.interfaces’ of the component includes the keywords of entries in the list of major dependencies of the entry. Then, the reliability of the first candidate of each component is calculated as the ratio of the score of the first candidate to that of the second. If the reliability value is higher than a threshold value, the component is provisionally fixed to the first candidate. This fixed entry is then removed from candidates of other components. After recalculation of the reliabilities for all non-fixed components, the same process is repeated while further components are fixed. In contrast to ‘component’ elements, ‘variable’ elements are estimated as follows. All ontology entries are scored for each variable, which might belong to multiple components through a ‘connection’ tag in the CellML file. An entry acquires points if the name of the variable includes the keywords of the entry, if the unit of the variable corresponds to the unit of the entry, or if the variable belongs to components with the entry in the list of dependencies. Then, variables are fixed by the same way as defined for the components.

D. Assistance of Model Structure Manipulation

To ease model structure modifications, a method to intelligently assist manipulations using informations of the CellML file and the ontology are developed. The method supports the following tasks: restricted allocation of an inserted model component to the anatomical hierarchy, intelligent addition of variables required for an inserted functional element, automatic connection of an inserted functional element to variables that the functional element depends on, and extraction of functional elements that have a major dependency on a specified variable. When a new functional element is inserted into an existing model structure, its anatomical location is determined by the anatomical information registered in the ontology. Variables that need to be connected are recognized through the ontology ID numbers of the corresponding variables defined in the CellML file. If some variables do not exist in the model, these variables are automatically added. For example, if a membrane transporter with ATP and ion concentrations as public interfaces is inserted into a model, the missing ATP variable is added first, and then all public interfaces are connected automatically. Furthermore, functional elements that depend on a physiological variable are discriminated by searching entries with the ID number of the variable in the list of major dependencies.

E. Implementations

Three kinds of software were implemented. The first is an interactive tool to append ontology ID numbers to a CellML file. This utility first provides a list of the estimated entry of components. Then, an estimated list of variables is presented. The user can interactively correct them in both steps. The second software is a graphical editor for model structures. This editor visualizes a model structure, which the user can manipulate. User operations are automatically assisted with

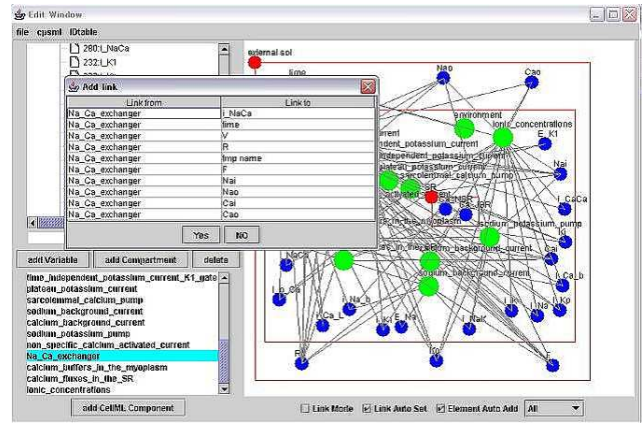


Fig. 4. Screen Shot of the Graphical Model Structure Editor

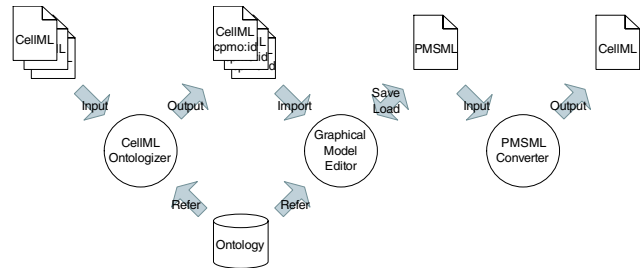


Fig. 5. Data Flow of Implemented Tools

confirmations. Fig. 4 shows an example of the screen shot. The third tool converts PMSML files to CellML files. This converter creates a new CellML file from a PMSML file by merging all referred CellML files. The data flow of these tools is shown in Fig. 5.

IV. EXPERIMENTS

A. Estimation of CellML Elements

Experimental results for the estimation of ontology ID numbers for elements of CellML files are shown in Table I. Public CellML files of the LRd¹, Kyoto², Jafri³, and ten

TABLE I
ESTIMATIONS FOR CELLML ELEMENTS

(a) Estimations for Components			
	Targets	Correct estimations	Valid estimations
LRd	15	13 (87%)	15 (100%)
Kyoto	24	18 (75%)	21 (88%)
Jafri	14	12 (86%)	14 (100%)
ten Tusscher	13	12 (92%)	13 (100%)

(b) Estimations for Variables			
	Targets	Correct estimations	Valid estimations
LRd	35	27 (77%)	33 (94%)
Kyoto	50	38 (76%)	48 (96%)
Jafri	27	24 (89%)	27 (100%)
ten Tusscher	24	21 (88%)	22 (92%)

TABLE II
OPERATION COUNTS FOR MODEL MANIPULATIONS

Task	Assisted	Non-assisted
A	5	13
B	11	17
C	10	16

Tusscher⁴ models are estimated using an ontology that is constructed only from physiological elements of the LRd and Kyoto models. The estimations of variables are performed under the condition that all components are assigned the correct ID numbers. Here, estimation is defined as correct if the component or variable are fixed to the correct entry, and as valid if the first candidate is the correct entry.

The ratios of valid estimations are over 80% for all files. In addition, the difference between models used for the construction of the ontology, i.e. LRd and Kyoto model, and those which are not used, i.e. Jafri and ten Tusscher model, is negligible small. This confirms that the proposed method is effective in appending ID numbers to CellML files. Nevertheless, three components of the Kyoto model are estimated as wrong entries, because the names of these components include abbreviations of physiological terms, for instance the “non-selective cation current” is named as ‘background_NSC_current’. For such a case it is extremely difficult to determine the correct entry. Moreover, two cases of wrong variable estimations are detected. In one case, the variable belongs to a component that is not registered in the ontology. This problem can be solved by expanding the ontology. In the other case, the name of the variable cannot be distinguished from others only by keyword matching. Although an improvement of the estimation method is necessary to solve this problem, current recognition rates are considered high enough for practical use.

B. Efficiency of the Manipulation Assistance

The numbers of assisted and non-assisted user operations for three different model manipulations are shown in Table II. All tasks are applied to the LRd model structure. In task A and B, respectively the time independent potassium current and the sodium channel of the LRd model are exchanged to those of the Kyoto model. In task C, the transient outward current of the Kyoto model is inserted into the LRd model.

The operation counts with assistance are reduced by more than one-third compared to those without assistance for all tasks. Required variables are inserted and connected automatically expect for model-specific variables, which are not registered in the ontology. These variables have to be inserted and connected manually. Therefore, the insertion of a functional element that contains many model-specific

variables, such as the sodium channel and the transient outward current of the Kyoto model used in task B and C, requires comparatively large number of operations.

V. CONCLUSION

This paper reports methods to assist modifications of complex cell model structures using markup languages and an ontology. Experimental results for the estimation of CellML elements show that the proposed ontology and estimation method is effective for physiological cell models. In addition, the assistant methods for model structure manipulations are shown to be practical. Edited models can be converted to CellML files and utilized by other tools. Future work includes the expansion of the proposed ontology. An efficient development of complex cell models using these methods and implementations will promote the further progress of cell physiological research.

REFERENCES

- [1] C.-H. Luo and Y. Rudy, “A dynamic model of the cardiac ventricular action potential. I. simulations of ionic currents and concentration changes,” *Circulation Research*, vol. 74, no. 6, pp. 1071–1096, June 1994.
- [2] M. S. Jafri, J. J. Rice, and R. L. Winslow, “Cardiac Ca^{2+} dynamics: The roles of ryanodine receptor adaptation and sarcoplasmic reticulum load,” *Biophysical Journal*, vol. 74, no. 3, pp. 1149–1168, Mar. 1998.
- [3] S. Matsuoka, N. Sarai, K. S., K. Ono, and A. Noma, “Role of individual ionic current systems in ventricular cells hypothesized by a model study,” *Jpn J Physiol*, vol. 53, no. 2, pp. 105–123, Apr. 2003.
- [4] K. H. ten Tusscher, D. Noble, P. J. Noble, and A. V. Panfilov, “A model for human ventricular tissue,” *American Journal of Physiology. Heart and Circulatory Physiology*, vol. 286, no. 4, pp. 1573–1589, Apr. 2004.
- [5] L. M. Loew and J. C. Schaff, “The virtual cell: a software environment for computational cell biology,” *Trends in biotechnology*, vol. 19, no. 10, pp. 401–406, Oct. 2001.
- [6] A. Garny, P. Kohl, and D. Noble, “Cellular open resource (COR): a public cellml based environment for modelling biological function,” *International Journal of Bifurcation and Chaos*, vol. 13, no. 12, pp. 3579–3590, Dec. 2003.
- [7] A. A. Cuellar, C. M. Lloyd, P. F. Nielsen, D. P. Bullivant, D. P. Nickerson, and P. J. Hunter, “An overview of CellML 1.1, a biological model description language,” *SIMULATION*, vol. 79, no. 12, pp. 740–747, Dec. 2003.
- [8] G. R. Christie, P. D. Bullivant, A. S. Blackett, and P. J. Hunter, “Modelling and visualising the heart,” *Computing and Visualization in Science*, vol. 4, no. 4, pp. 227–235, July 2002.
- [9] S. Missan and T. F. McDonald, “CESE: Cell electrophysiology simulation environment,” *Applied Bioinformatics*, vol. 4, no. 2, pp. 155–156, 2005.
- [10] T. Shimayoshi, A. Amano, and T. Matsuda, “A description language for cell physiological simulation,” in *IFMBE Proceedings APCMBE 2005 “6th Asian-Pacific Conference on Medical and Biological Engineering”*, vol. 8, Apr. 2005, pp. PA–3–09.
- [11] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, “Gene Ontology: tool for the unification of biology,” *Nature Genetics*, vol. 25, no. 1, pp. 25–29, May 2000.

¹http://www.cellml.org/examples/models/LR_II_model_1994.xml

²http://www.cellml.org/examples/models/matsuoka.ventricular.cell_model_2003.xml

³http://www.cellml.org/examples/models/jafri_rice_winslow_model_1998.xml

⁴http://www.cellml.org/examples/models/ten_tusscher_model_2004.xml