

Integrated Modelling, Simulation and Verification in P Systems and Synthetic Biology

Rezumatul Tezei de Doctorat

LAURENȚIU MARIAN MIERLĂ

Coordonatori:

Prof.univ.dr. FLORENTIN IPATE

Prof.univ.dr. TUDOR BĂLĂNESCU

Departamentul de Informatică

Universitatea din Pitești

Septembrie 2019

Cuprinsul Tezei

Capitolul 1: Introduction

Capitolul 2: Kernel P Systems

Capitolul 3: KP WORKBENCH - an Integrated Toolset for kernel P Systems

Capitolul 4: INFOBIOTICS WORKBENCH - an Integrated Toolset for Synthetic Biology

Capitolul 5: Case studies

Capitolul 6: Conclusions and Future Research Directions

1 Introducere

Calculul cu Membrane este o paradigma a calculului natural, care isi propune sa abstractizeze modele computationale din structura si functionarea celulelor vii, cat si din modul de cooperare al celulelor in cadrul tesuturilor sau al altor structuri celulare. Aceasta directie de cercetare a fost initiata de Gheorghe Păun in 1998, prin introducerea modelului computational denumit sisteme P. Diferite variante de sisteme P au fost alcatuite si o serie de aspecte teoretice au fost intens studiate: puterea computationala a diferitelor variante, aspecte de complexitate si ierarhii de clase de limbaje generate si/sau recunoscute de catre acestea. De-a lungul anilor, calculul cu membrane s-a dovedit util si in modelarea proceselor biologice, simularea ecosistemelor sau in diverse aplicatii in domeniul economiei, al graficii pe calculator etc.

Biologia Sintetica este o disciplina cu o crestere accelerata, care se ocupa cu proiectarea si implementarea de noi fenotipuri biologice. Desi aceste fenotipuri nu apar in mod normal in natura, pot fi construite prin combinatii rationale intre parti biologice naturale si sintetice. Cu toate ca biologia sintetica promite a fi o noua tehnologie revolutionara pentru ingineria sistemelor biologice, natura complexa a interactiunilor din cadrul retelelor genetice si de semnalizare face dificila proiectarea si analiza in laboratoarele clasice a acestor fenotipuri. De aceea, intr-o maniera similara cu cea a domeniului biologiei sistemice (disciplina inrudita, premergatoare), biologia sintetica depinde puternic de instrumente computationale sofisticate pentru specificarea, simularea, verificarea si compilarea schitelor biologice. Construirea modelelor acestor retele si analiza comportamentului lor prin metode computationale poate

furniza avantaje semnificative din punct de vedere timp, imbunatatirea intelegerii si a functionarii biologice.

Obiective

Teza de fata se concentreaza pe domeniul Calculului cu Membrane si al Biologiei Sintetice, avand ca obiectiv principal prezentarea de formalisme, proiectarea si implementarea de instrumente software care sa faciliteze o analiza integrata a unor studii de caz relevante din punct de vedere al modelarii, simularii si verificarii. In vederea atingerii acestui obiectiv, o serie de directii au fost concepute si abordate:

1. Prezentarea versiunii I a formalismului kernel P systems, capabil sa furnizeze un cadru coerent care cuprinde cele mai comune si mai puternice caracteristici ale diverselor tipuri de sisteme P;
2. Proiectarea, dezvoltarea si prezentarea unui set integrat de instrumente software, definite in contextul formalismului mai sus prezentat, oferind capabilitati de modelare, simulare si verificare;
3. Proiectarea, dezvoltarea si prezentarea unui set integrat de instrumente software care sa faciliteze modelarea, simularea, verificarea si biocompilarea diverselor sisteme din domeniul biologiei sintetice;
4. Demonstrarea capabilitatilor formalismelor si instrumentelor software de mai sus, pe baza unui set de studii de caz relevante.

Structura Tezei

Teza de fata este organizata dupa cum urmeaza:

- Introduction - Capitolul 1 sintetizeaza obiectivele tezei si prezinta contributiile principale;
- Kernel P Systems - Chapter 2 prezinta versiunea I a sistemelor kernel P, definind formal diversele concepte, precum compartimente, reguli, garzi, strategii de executie etc. O analiza comparativa cu alte variante de sisteme P este furnizata, urmata de un studiu de caz teoretic. La finalul acestui capitol, este prezentat limbajul de specificare al sistemelor kernel P si cel al sistemelor kernel P simplificate;

- KP WORKBENCH - an Integrated Toolset for kernel P Systems - Capitolul 3 prezinta instrumentul software construit in jurul versiunii I a formalismului kernel P systems, oferind capabilitati precum limbajul de modelare kP Lingua, limbajul kP Queries pentru specificarea de proprietati, simulare si verificare formala automata. Elemente de nivel inalt ale arhitecturii software aferente instrumentului software KP WORKBENCH sunt prezentate la finalul acestui capitol;
- INFOBIOTICS WORKBENCH - an Integrated Toolset for Synthetic Biology - Capitolul 4 prezinta un instrument software care faciliteaza modelarea, simularea, verificarea si biocompilarea diferitelor ansamble din biologia sintetica, oferind limbajul InfoBiotics Language, implementari ale diferitelor variante ale algoritmului de simulare Gillespie, capabilitati de analiza cantitativa si calitativa bazata de model checking, biocompilare si mecanisme de conversie a modelelor catre diferite standarde din domeniu;
- Case studies - Capitolul 5 demonstreaza utilitatea formalismelor si a instrumentelor software introduse, pe baza unor studii de caz reprezentative din domeniul informaticii si al biologiei sintetice;
- Conclusions and Future Research Directions - Finally, Capitolul 6 traseaza concluziile si prezinta viitoarele directii de cercetare, pe baza rezultatelor elaborate in teza de fata.

Contributiile Tezei

In ultimii ani a avut loc o dezvoltare semnificativa in folosirea paradigmei sistemelor P in modelarea simularea si verificarea formala a diferitelor mecanisme, printre care se numara probleme si algoritmi bine cunoscuti. In multe dintre cazuri, formalismele rezultate au necesitat caracteristici si constrangeri aditionale, in comparatie cu definitia variantei originale de sisteme P. Desi extrem de folositoare in modelare in anumite situatii, adaugarea ad-hoc de caracteristici aditionale conduce catre un efect advers in ceea ce priveste capabilitatea sistemelor P in a furniza un cadru coerent pentru analiza si verificare. In adresarea acestor probleme, contributiile principale ale tezei sunt urmatoarele:

- Proiectarea si implementarea limbajului kP Lingua, care acopera intregul set de caracteristici ale sistemelor kernel P. Definit in jurul principiilor de claritate,

expresivitate si minimalism, kP Lingua reflecta paradigma tip-instanta a sistemelor kernel P, ilustrand doua categorii de elemente in modelare: 1. Definitia de tipuri de compartimente si 2. Instantierea si legaturile dintre acestea;

- Proiectarea si implementarea platformei software kP WORKBENCH, care integreaza un set de instrumente ce permit modelarea, simularea si verificarea sistemelor kernel P, incluzand limbajul kPLingua, de mai sus. kP WORKBENCH este scris in limbajul C#, folosind framework-ul .NET Core, rezultand intr-o implementare cross-platform atat a interfetei in linie de comanda, cat si a interfetei grafice. Desi, anterior introducerii kP WORKBENCH, au existat eforturi concetrate in jurul integrarii unui subset al sistemelor kernel P (simple kernel P systems) in P-Lingua si MeCoSim, aceste implementari omit o serie larga de caracteristici puse la dispozitie de platforma software prezentate in teza de fata;
- Extinderea platformei kP WORKBENCH cu un nou modul de simulare implementat in C# si .NET Core. Simulatorul permite modelarea detaliilor din output printr-un set de parametri. Desi in afara ariei de prezentare din cadrul acestei teze, kP WORKBENCH a fost extins, de asemenea, pentru a include un simulator bazat pe FLAME;
- Proiectarea si implementarea unui modul de verificare in cadrul kP WORKBENCH, integrand instrumentele de model checking Spin si NuSMV. In vederea facilitarii acestei integrari, au fost implementate mecanisme de mapare si translatate automata dinspre elementele limbajului de modelare kP Lingua, catre cele ale limbajelor Promela si SMV, corespunzatoare instrumentelor de model checking Spin, respectiv NuSMV;
- Extinderea mecanismului de verificare al kP WORKBENCH prin introducerea kP Queries – un limbaj independent de specificare a proprietarilor unui sistem modelat in kP Lingua, bazat pe formalismele LTL si CTL. Modulul kP Queries faciliteaza, de asemenea, o translatate automata a acestor proprietati catre limbajele de model checking ale diferitelor instrumente folosite (Spin, NuSMV).

Diferite metode computationale si instrumente (precum GEC, Eugene and Proto) au fost dezvoltate in ultimii ani pentru a facilita proiectarea si analiza in biologia sintetica. Cu toate acestea, tehnicile de analiza sunt limitative, majoritatea acestor instrumente bazandu-se predominant pe simulare. De asemenea, diferitele abordari

in modelare nu iau in considerare natura compartimentala a sistemelor biologice, ale caror procese sunt puternic distribuite nu doar in timp, ci si in spatiu. Pentru a adresa aceste probleme, teza de fata include urmatoarele contributii:

- • Proiectarea si implementarea platformei software INFOBIOTICS WORKBENCH (IBW), ca fiind un Mediu Integrat de Dezvoltare (IDE) pentru domeniul biologiei sintetice. IBW este bazat pe platforma Eclipse si pune la dispozitia utilizatorilor instrumente de modelare, simulare, verificare, biocompilare si conversie a modelelor biologice;
- • Introducea, proiectarea si implementarea limbajului InfoBiotics Language (IBL), prin a carui sintaxa poate fi modelata structura si interactiunea diferitelor entitati biologice, cat si specificate proprietati pentru a fi verificate folosind diverse instrumente de model checking;
- • Proiectarea si implementarea unui modul de model checking in cadrul IBW, capabil sa furnizeze atat o analiza cantitativa, cat si una calitativa a sistemelor biologice, prin integrarea cu diferite instrumente de model checking, precum Prism, MC2, NuSMV. In vederea facilitarii acestor integrari, au fost implementate mecanisme de mapare si translatare automata dinspre elementele limbajului IBL, catre limbajele Prism si SMV.

Pentru a demonstra capabilitatile si eficacitatea formalismelor si instrumentelor introduse, o serie de rezultate sunt prezentate in baza unor studii de caz:

- Problema 3-Col, prezentand expresivitatea in modelare a kernel P Systems, implementarea in Promela si rezultatele experimentelor efectuate in Spin;
- Lantul respirator al E. Coli, punand accentul pe modelarea in kernel P systems si verificarea in Spin;
- Problema Subset Sum, demonstrand expresivitatea limbajelor kP Lingua si kP Queries in cadrul analizelor efectuate folosind platforma kP WORKBENCH;
- Generating Square Numbers, demonstrand capabilitatile de modelare ale kP Lingua si cele de verificare ale kP WORKBENCH;
- Broadcasting with Acknowledgement, prezentand modelul kP Lingua si rezultatele verificarii diferitelor proprietati LTL si CTL folosind kP WORKBENCH;

- Quorum Sensing si Pulse-Generator – studii de caz bine-cunoscute in biologia sintetica, pe baza carora sunt prezentate rezultatele unei analize atat cantitative, cat si calitative, folosind instrumentele de model checking Prism si MC2;
- Toggle Switch si The Repressilator – modele biologice sintetice folosite in demonstrarea capabilitatilor de modelare, simulare si verificare ale INFOBIOTICS WORK-BENCH.

Lista de Publicatii

- [1] M. Gheorghe, F. Ipate, L. Mierlă, and S. Konur, “kPWorkbench: A Software Framework for Kernel P Systems,” in *Thirteenth Brainstorming Week on Membrane Computing, BWMC 2015, Sevilla, Spain, February 2-6, 2015*, pp. 179–194, 2015
- [2] M. Gheorghe, S. Konur, F. Ipate, L. Mierla, M. E. Bakir, and M. Stannett, “An Integrated Model Checking Toolset for Kernel P Systems,” in *Membrane Computing*, vol. LNCS 9504 of *Lecture Notes in Computer Science*, pp. 153–170, Springer International Publishing, 2015
- [3] C. Dragomir, F. Ipate, S. Konur, R. Lefticaru, and L. Mierla, “Model Checking Kernel P Systems,” in *Membrane Computing*, vol. LNCS 8340 of *Lecture Notes in Computer Science*, pp. 151–172, Springer Berlin Heidelberg, 2014
- [4] F. Ipate, C. Dragomir, R. Lefticaru, L. Mierlă, and M. Pérez-Jiménez, “Using a Kernel P System to Solve the 3-Col Problem,” in *Proceedings of the 13th International Conference on Membrane Computing*, pp. 243–258, Computer and Automation Research Institute, Hungarian Academy of Sciences, 2012
- [5] A. Țurcanu, L. Mierlă, F. Ipate, A. Ștefanescu, H. Bai, M. Holcombe, and S. Coakley, “Modelling and Analysis of E. coli Respiratory Chain,” in *Applications of Membrane Computing in Systems and Synthetic Biology*, vol. ECC 7 of *Emergence, Complexity and Computation*, pp. 247–266, Springer International Publishing, 2013
- [6] S. Konur, L. Mierlă, F. Ipate, and M. Gheorghe, “kPWorkbench: A Software Suite for Membrane Systems,” *Submitted to SoftwareX*, 2019
- [7] S. Konur, M. Gheorghe, C. Dragomir, L. Mierla, F. Ipate, and N. Krasnogor, “Qualitative and Quantitative Analysis of Systems and Synthetic Biology Constructs using P Systems,” *ACS Synthetic Biology*, vol. 4, no. 1, pp. 83–92, 2015
- [8] S. Konur, H. Fellermann, L. Mierlă, D. Sanassy, C. Ladroue, S. Kalvala, M. Gheorghe, and N. Krasnogor, “An Integrated In Silico Simulation and Biomatter Compilation Approach to Cellular Computation,” in *Advances in Unconventional Computing*, vol. ECC 23 of *Emergence, Complexity and Computation*, pp. 655–676, Springer International Publishing, 2016
- [9] S. Konur, C. Ladroue, H. Fellermann, D. Sanassy, L. Mierlă, F. Ipate, S. Kalvala, M. Gheorghe, and N. Krasnogor, “Modeling and analysis of genetic boolean gates using the Infobiotics Workbench,” in *Proceedings of the 2014 Workshop on Verification of Engineered Molecular Devices and Programs*, 2014
- [10] H. Fellermann, S. Konur, C. Ladroue, L. Mierlă, B. Dun, D. Sanassy, S. Kalvala, M. Gheorghe, and N. Krasnogor, “Integrated Specification, Simulation, Verification

- and Compilation for Rapid Prototyping in Synthetic Biology,” *To be Submitted to ACS Synthetic Biology*, 2020
- [11] M. Gheorghe, F. Ipate, C. Dragomir, L. Mierlă, L. Valencia-Cabrera, M. García-Quismondo, and M. J. Pérez-Jiménez, “Kernel P Systems - Version I,” *Eleventh Brainstorming Week on Membrane Computing (11BWMC)*, pp. 97–124, 2013
- [12] M. E. Bakir, F. Ipate, S. Konur, L. Mierla, and I. Niculescu, “Extended Simulation and Verification Platform for Kernel P Systems,” in *Membrane Computing*, vol. LNCS 8961 of *Lecture Notes in Computer Science*, pp. 158–178, Springer International Publishing, 2014
- [13] M. Gheorghe, F. Ipate, R. Lefticaru, M. J. Pérez-Jiménez, A. Țurcanu, L. Valencia Cabrera, M. García-Quismondo, and L. Mierlă, “3-Col Problem modelling using Simple Kernel P Systems,” *International Journal of Computer Mathematics*, vol. 90, no. 4, pp. 816–830, 2013
- [14] F. Ipate, R. Lefticaru, L. Mierlă, L. Valencia-Cabrera, H. Han, G. Zhang, C. Dragomir, M. J. Pérez-Jiménez, and M. Gheorghe, “Kernel P systems: Applications and implementations,” in *Proceedings of The Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2013*, vol. 212 of *Advances in Intelligent Systems and Computing*, pp. 1081–1089, Springer Berlin Heidelberg, 2013
- [15] R. Lefticaru, L. F. Macías-Ramos, I. M. Niculescu, and L. Mierlă, “Agent-based Simulation of Kernel P Systems with Division Rules using FLAME,” in *Membrane Computing*, vol. LNCS 10105 of *Lecture Notes in Computer Science*, pp. 286–306, Springer International Publishing, 2017
- [16] D. Sanassy, H. Fellermann, N. Krasnogor, S. Konur, L. Mierlă, M. Gheorghe, C. Ladroue, and S. Kalvala, “Modelling and Stochastic Simulation of Synthetic Biological Boolean Gates,” in *Proceedings of the 2014 IEEE 16th International Conference on High Performance Computing and Communication, HPCC '14, (Paris, France)*, pp. 404–408, 2014
- [17] I. Dincă, F. Ipate, L. Mierlă, and A. Ștefănescu, “Learn and Test for Event-B – a Rodin Plugin,” in *Abstract State Machines, Alloy, B, VDM, and Z*, pp. 361–364, Springer Berlin Heidelberg, 2012