

Latvijas Lauksaimniecības universitāte
Informācijas Tehnoloģiju fakultāte
Datoru sistēmu katedra

Tatjana Rubina

**NEVIENDABĪGA SVARĪGUMA BIOĶĪMISKO TĪKLU
STRUKTŪRAS EVOLŪCIJAS DATORMODELĒŠANA**

Promocijas darba
KOPSAVILKUMS
doktora grāda ieguvei
Informācijas tehnoloģiju nozarē (Dr.sc.ing.)



IEGULDĪJUMS TAVĀ NĀKOTNĒ



Promocijas darba izstrāde un noformēšana līdzfinansēta no
Eiropas Savienības Sociālā fonda

Tatjana Rubina

Paraksts

Jelgava 2013

INFORMĀCIJA

Darba izpildes vieta: Latvijas Lauksaimniecības universitāte, Informācijas Tehnoloģiju fakultāte, Datoru sistēmu katedra, Liela iela 2, Jelgava, Latvija

Eksperimentālā darba izpildes vieta: Latvijas Lauksaimniecības universitāte, Informācijas Tehnoloģiju fakultāte, Datoru sistēmu katedra, Liela iela 2, Jelgava, Latvija

Promocijas darba zinātniskais vadītājs: Latvijas Lauksaimniecības universitātes asociētais profesors Dr.sc.ing. Egils Stalidzāns

Darbs akceptēts LLU Informācijas tehnoloģiju fakultātes Datoru sistēmu katedras paplašinātā akadēmiskā sēdē 2013. gada 15.maijā. Protokols Nr.6.

Promocijas darbs izstrādāts ar Eiropas Sociālā Fonda (ESF) projekta “Atbalsts LLU doktora studiju īstenošanai” vienošanās Nr.2009/0180/1DP/1.1.2.1.2./09/IPIA/VIAA/017 atbalstu.

Oficiālie recenzenti:

1. Rezeknes augstskolas profesors, Dr.sc.ing. Oļegs Užga-Rebrovs;
2. Rīgas Tehniskās universitātes profesore, Dr.habil. Gaļina Merkurjeva;
3. University College London (Apvienotā Karaliste) pēcdoktorantūras zinātniskā asistente, PhD Dace Rukliša.

Promocijas darba aizstāvēšana notiks LLU Informācijas tehnoloģiju nozares promocijas padomes atklātajā sēdē 2014. gada ____ . _____ Jelgavā, Lielā ielā 2, Informācijas Tehnoloģiju fakultātes ____ . auditorijā plkst. ____.

Ar promocijas darbu var iepazīties LLU Fundamentālajā bibliotēkā, Lielā ielā 2, Jelgavā un http://lufb.llu.lv/promoc_darbi.html

Atsauksmes sūtīt Promocijas padomes sekretārei – Lielā ielā 2, Jelgavā, LV-3001; tālrunis: 63 00 56 21; e-pasts: tatjana.tabunova@llu.lv. Atsauksmes vēlams sūtīt skenētā veidā ar parakstu.

Padomes sekretāre: LLU lektore, Mg.paed. Tatjana Tabunova.

ISBN 978-9984-861-68-5 (online)

SATURS

PROMOCIJAS DARBA APROBĀCIJA	4
IEVADS	6
Tēmas aktualitāte	6
Promocijas darba mērķis un uzdevumi	8
Pētījuma metodes	8
Zinātniskais jauninājums.....	9
Pētījuma tēzes	9
Praktiskā vērtība.....	10
Promocijas darba struktūra un apjoms	10
1. BIOĻĢISKO SISTĒMU EVOLŪCIJA	10
2. BIOĶĪMISKIE TĪKLI UN TO ANALĪZE	11
3. BIOĶĪMISKO TĪKLU EVOLUCIONĀRĀS ATTĪSTĪBAS MODELĒŠANA	12
Bioķīmisko tīklu struktūras evolūcijas modelēšanas procedūra.....	15
Bioķīmisko tīklu struktūras evolūcijas modelēšanas algoritms.....	18
4. PROGRAMMATŪRAS RĪKS BINESA	22
5. BIOĶĪMISKO TĪKLU MODEĻU PIELIETOJAMĪBAS NOVĒRTĒJUMS.....	23
6. BIOĶĪMISKO TĪKLU STRUKTŪRAS EVOLŪCIJAS DATORSIMULĀCIJAS.....	24
SECINĀJUMI	31
Galvenie darba rezultāti	31
Secinājumi un attīstības perspektīvas.....	33
LITERATŪRAS SARAKSTS.....	35

PROMOCIJAS DARBA APROBĀCIJA

Promocijas darbā veikto pētījumu rezultāti ir atspoguļoti šādās publikācijās:

- 1) **Rubina T.**, Stalidzans E. (2013) BINESA – a software tool for evolution modelling of biochemical networks' structure. In: Proceedings of the 14th International Symposium on „Computational Intelligence and Informatics” CINTI'2013, ISBN 978-1-4799-0194-4, November 19-21, 2013, Budapest, Hungary, p.345-350. (indeksēts SCOPUS un Web of Knowledge datubāzēs).
- 2) **Rubina T.**, Mednis M., Stalidzans E. (2013) Agreement assessment of biochemical pathway models by structural analysis of their intersection. In: Proceedings of the 14th International Symposium on „Computational Intelligence and Informatics” CINTI'2013, ISBN 978-1-4799-0194-4, November 19-21, 2013, Budapest, Hungary, p.411-418. (indeksēts SCOPUS un Web of Knowledge datubāzēs).
- 3) **Rubina T.** (2013) The procedure of evolution modelling of biochemical networks structure. Biosystems and Information Technology, Vol.2, Nr.2, ISSN 2255-8004, p.19-25.
- 4) **Rubina T.**, Stalidzans E. (2012) Evolution modeling algorithm of biochemical networks. In: Proceedings of the 10th Industrial Simulation Conference, ISC'2012. A publication of EUROSIS, ISBN 978-90-77381.71.7, June 4-6, 2012, Brno, Czech Republic, p.24-30. (indeksēts Thomson/Reuters Web of Knowledge datubāzē).
- 5) **Rubina T.**, Stalidzans E. (2012) Evolution of control loops of biological systems. In: Proceedings of the 5th international scientific conference „Applied Information and Communication Technologies” AICT'2012, ISBN 978-9984-48-065-7, April 26-27, 2012, Jelgava, Latvia, p.317-324.
- 6) **Rubina T.** (2012) Tools for analysis of biochemical network topology. Biosystems and Information Technology, Vol.1, Nr.1, ISSN 2255-8004, p.25-31.
- 7) **Rubina T.**, Stalidzans E. (2010) Software Tools for Structure Analysis of Biochemical Networks. In: Proceedings of the 4th international scientific conference „Applied Information and Communication Technologies” AICT'2010, ISBN 978-9984-48-022-0, April 22-23, 2010, Jelgava, Latvia, p.33-49. (indeksēts Thomson/Reuters Web of Knowledge datubāzē).
- 8) **Rubina T.**, Stalidzans E. (2010) Topological features and parameters of biochemical network structure. In: Proceedings of the 8th Industrial Simulation Conference ISC'2010. A publication of EUROSIS, ISBN 978-90-77381.5-57, June 7-9, 2010, Budapest, Hungary, p.228-236. (indeksēts Thomson/Reuters Web of Knowledge datubāzē).
- 9) Odzina I., **Rubina T.**, Rutkis R., Kalnenieks U., Stalidzans E. (2010) Structural model of biochemical network of *Zymomonas mobilis*

- adaptation for glycerol conversion into bioethanol. In: Proceedings of the 4th International Scientific Conference „Applied Information and Communication Technologies” AICT’2010, ISBN 978-9984-48-022-0, April 22-23, 2010, Jelgava, Latvia, p.50-54. (indeksēts Thomson/Reuters Web of Knowledge datubāzē).
- 10) **Rubina T.**, Brusbardis V. (2009) Applications of biochemical networks discovering control mechanisms in systems biology. In: Proceedings of the Annuals Students International Scientific Conference „Youth in science and Professional practice”, ISBN 978-9984-784-99-1, April 23, 2009, Jelgava, Latvia, p.1-7.
 - 11) **Zukova T.**, Stalidzans E. (2006) System biology – interaction science of biology and information technology. In: Proceedings of the International Scientific Conference „Information Technologies for Rural Development”, ISBN 9984-784-13-4, October 19-20, 2006, Jelgava, Latvia, p.120-130.
 - 12) Grunde-Zeiferts U., Mozga I., **Žukova T.**, Stalidzāns E. (2006) Therapy modelling combining methods of systems biology and automatic control theory. In: Proceedings of International Scientific Conference „Animals. Health. Food Hygiene.”, ISSN 1407-1754, November 10, 2006, Jelgava, Latvia, p.70-74.

Pētījumos iegūtie rezultāti ir prezentēti šādās konferencēs:

- 1) „Agreement assessment of biochemical pathway models by structural analysis of their intersection”. 14th IEEE International Symposium on *Computational Intelligence and Informatics CINTI’2013*, November 19-21, 2013, Budapest, Hungary
- 2) „BINESA – a software tool for evolution modelling of biochemical networks’ structure”. 14th IEEE International Symposium on *Computational Intelligence and Informatics CINTI’2013*, November 19-21, 2013, Budapest, Hungary
- 3) “Evolution modeling of biochemical network structure with BINESA”. 6th International Scientific Conference “*Applied Information and Communication Technologies AICT’2013*”, April 25-26, 2013, Jelgava, Latvia
- 4) “Evolution modeling algorithm of biochemical networks”. 10th *Industrial Simulation Conference, ISC’2012*, June 4-6, 2012, Brno, Czech Republic
- 5) “Evolution of control loops of biological systems”. 5th International Scientific Conference “*Applied Information and Communication Technologies AICT’2012*”, April 26-27, 2012, Jelgava, Latvia
- 6) „Topological features and parameters of biochemical network structure”. The European Multidisciplinary Society for Modelling and simulation Technology. 8th *Industrial Simulation Conference, ISC’2010*, June 7-9, 2010, Budapest, Hungary

- 7) „Software Tools for Structure Analysis of Biochemical Networks”. 4th International Scientific Conference “*Applied Information and Communication Technologies AICT'2010*”, April 22-23, 2010, Jelgava, Latvia
- 8) “Approach of biochemical-network structure analysis”. 4th International Scientific Conference „*Students on their Way to Science*”, May 14-15, 2009, Jelgava, Latvia
- 9) „Evolutionary modeling of cellular control networks”. International Scientific Conference „*Youth in Science and Professional Practice*”, April 23, 2009, Jelgava, Latvia

Pētījumos iegūtie rezultāti ir apspriesti šādos semināros:

- 1) Rubina T. Bioķīmisko tīklu veidi un to datormodelēšana, Biosistēmu grupas seminārs, Jelgava, Latvija, 2010.gada 13.janvārī
- 2) Rubina T. Struktūras parametru analīze, Biosistēmu grupas seminārs, Jelgava, Latvija, 2010.gada 3.februārī
- 3) Rubina T. Bioķīmisko tīklu struktūras vadības loki un to bioloģiskā nozīme. Datu formāti sistēmbioloģijas nozarē, Biosistēmu grupas seminārs, Jelgava, Latvija, 2010.gada 8.jūlijā
- 4) Rubina T. Novecošana kā ģenētiskā programma, Biosistēmu grupas seminārs, Jelgava, Latvija, 2010.gada 3.novembrī
- 5) Rubina T. Būla tīklu modelēšana, Biosistēmu grupas seminārs, Jelgava, Latvija, 2012.gada 4.janvārī
- 6) Rubina T. Matlab izmantošana zinātnisko pētījuma datu apstrādei un vizualizācijai, Biosistēmu grupas seminārs, Jelgava, Latvija, 2012.gada 21.novembrī
- 7) Rubina T. Bioķīmisko tīklu struktūras evolūcija ar modelēšanas rīku BINESA, Biosistēmu grupas seminārs, Jelgava, Latvija, 2013.gada 19.augustā

IEVADS

Tēmas aktualitāte

Informācijas tehnoloģiju strauja attīstība pagājušā gadsimta beigās ir ļāvusi tās izmantot daudzās nozarēs, tajā skaitā molekulārās bioloģijas, ģenētikas, populācijas ģenētikas nozarēs kā neaizstājamo instrumentu zinātnisko un praktisko pētījumu veikšanā. It īpaši tas skar dzīvo organismu izpēti, jo šis process ir sarežģīts dažādu iemeslu dēļ. Pirmkārt, tas prasa milzīgu skaitu eksperimentu izpildi laboratorijās (Chen u.c., 2009a) un laika resursus, lai iegūtu datus un izpratni par organisma iekšējo komponentu organizācijas, mijiedarbības, funkcionēšanas un evolūcijas atsevišķiem aspektiem noteiktajos apstākļos. Otrkārt, atsevišķos gadījumos dzīvo organismu izpēte pat nav iespējama pieņemto ētisko normu, dzīvo organismu sarežģītības, veselību vai

dzīvību apdraudošo apstākļu dēļ. Līdz ar to kā risinājumu šādos gadījumos var pielietot informācijas tehnoloģijas, modelēšanu un simulācijas zināšanu un izpratnes gūšanai un paplašināšanai par dzīvajiem organismiem, modeļu un algoritmu izstrādē pamatojoties uz esošajiem eksperimentālajiem datiem, zināšanām un faktiem.

Dzīvie organismi ir sarežģītas, nelineāras, pašorganizējošas un pašregulējošas bioloģiskās sistēmas. Jebkurā bioloģiskajā sistēmā (sūnā, audos, orgānos, organismā) darbojas dažāda veida procesi (regulācijas, fiziskie vai ķīmiskie), kas var tikt attēloti un raksturoti bioķīmisko ceļu veidā. Bioķīmiskie ceļi un to regulācija ir organizēti tīklu veidā, kuru pamatā ir savdabīga struktūra. Bioķīmisko tīklu struktūru veido dažādi elementi (proteīni, metabolīti, gēni, mazās molekulas) un to savienojumi, kas nosaka struktūras elementu mijiedarbību. Taču pašu bioķīmisko tīklu organizāciju un darbību nosaka, regulē un vada informācija un mehānismi, kas ir ieprogrammēti gēnos.

Organismu bioķīmiskajos tīklos ir dažāda svarīguma procesi. Ir daži bioķīmiskie procesi, piemēram, glikolīze, kas ir gandrīz visos zināmajos organismos un var tikt attiecināts uz vitāliem procesiem. Tam ir iespējams evolucionārs pamatojums: organismi ar izmaiņām svarīgajos procesos neizdzīvoja, atbrīvojot vietu organismiem, kuros izmaiņas būtiski neskāra to eksistences svarīgākos procesus. Tajā pašā laikā ir mazsvarīgi, rudimentāri procesi, kurus mutāciju radītie traucējumi var pat iznīcināt, tomēr organisma uzvedībā tas neatstās būtiskas sekas.

Stohastisku mutāciju rezultātā rodas nestohastiski veidoti tīkli. Tātad to veidošanās mehānismi ir netriviāli un dabiskā izlase veic sarežģītu darbu, ko ir nepieciešams pētīt dažāda svarīguma tīklos, jo tas ir būtisks tīklu veidošanās aspekts. Rodas jautājums - kā evolucionē dažāda svarīguma procesu bioķīmiskie tīkli, un kādas izmaiņas notiek to struktūrā atkarībā no dažādu mutāciju tipu ietekmes, mainoties mutāciju varbūtībai?

Ja dabiskajā izlasē netiek pieļautas būtiskas ģenētiskās informācijas kopēšanas kļūdas ļoti svarīgos procesos mutāciju ietekmē un organisma nākamā paaudze saglabā izdzīvošanai kritisko funkcionalitāti, tad mazāk svarīgos procesos izmaiņas ir pieļaujamas. Tādējādi esošie bioķīmiskie tīkli var tikt pārrauti vai var pat rasties jaunas struktūras. Līdz ar to evolūcijas procesā veidojošās bioķīmisko tīklu struktūras izmaiņas, pārrāvumi un struktūru atliekas var būtiski ietekmēt organisma noturību pret vides izmaiņām, kā arī pret ārstnieciskām terapijām radot alternatīvus dažādu svarīgu procesu izpildes variantus. Līdz ar to bioķīmisko tīklu struktūru evolūcijas modeļošana, pat vienkāršotā veidā, var radīt priekšstatu par organismu reakcijas daudzveidību, ietekmējot dažādus bioķīmiskos procesus.

Tā kā ģenētisko mutāciju ietekmes uz vienāda un jaukta svarīguma bioķīmisko tīklu struktūras izmaiņām izpēte laboratorijas apstākļos ir apgrūtināta un laikietilpīga, atsevišķos gadījumos pat neiespējama, šī jautājuma daudzpusīgai un detalizētai izpētei ir nepieciešama modeļošana un datorsimulācijas. Tā kā dabā notiekošie evolucionārie procesi ir ļoti sarežģīti un

visas to nianšes nav iespējams datorsimulācijās realizēt skaitļošanas jaudu dēļ, ir nepieciešams izstrādāt vienkāršotus formālistus, konkrētu jautājumu padziļinātai izpētei. Darba ietvaros plānots koncentrēties uz tīklu ar dažāda svarīguma reakcijām modelēšanu, kas prasa izstrādāt vienkāršotu un skaitļošanas jaudām piemērotu evolūcijas modelēšanas procedūru un algoritmu, kas ņemtu vērā bioķīmisko procesu svarīgumu, kā arī programmatūras rīku, ar kura palīdzību būtu iespējams izpildīt evolūcijas datoreksperimentus atbilstoši piedāvātajam algoritmam un procedūrai.

Promocijas darba mērķis un uzdevumi

Promocijas darba mērķis ir izstrādāt datormodelēšanas pieeju un ar tās palīdzību noteikt dažāda tipa mutāciju ietekmi uz neviendabīga svarīguma bioķīmisko tīklu struktūras evolūciju.

Promocijas darba sasniegšanai izstrādes procesā ir jāveic vairāki uzdevumi:

- 1) izpētīt bioķīmisko tīklu strukturālās analīzes metodes un to attēlošanas paņēmienus;
- 2) izpētīt bioķīmisko tīklu evolūcijas procesu un to ietekmējošos faktorus;
- 3) izpētīt bioķīmisko tīklu evolūcijas datormodelēšanas pieejas;
- 4) izpētīt esošus bioķīmisko tīklu datormodelēšanas un to struktūru analizējošus programmatūras rīkus;
- 5) izstrādāt genomam piesaistītas bioķīmisko tīklu struktūras evolūciju imitējošu vienkāršotu algoritmu un struktūras analīzes kritērijus, un ieviest tos programmatūras rīka prototipā;
- 6) izpildīt genomam piesaistītas bioķīmisko tīklu struktūras, kas apraksta vienāda un jaukta svarīguma procesus, evolūcijas datoreksperimentus;
- 7) ar datormodeļa palīdzību analizēt atsevišķu mutāciju tipu un vienlaicīgi vairāku mutāciju tipu ietekmi uz struktūras evolūciju;
- 8) noteikt bioķīmiskā tīkla modeļa pielietojamību raksturojošus topoloģiskos parametrus.

Pētījuma metodes

Programmatūras rīka prototipa algoritmu izstrāde tika realizēta *Visual Basic* programmēšanas valodā.

Bioķīmisko tīklu struktūras evolūcijas modelēšanā izmantoti dažādi mutāciju operatori, par kuriem aprakstīts molekulārās ģenētikas teorijā un kas atbilst daļai no dabā pastāvošajiem mutāciju veidiem.

Bioķīmiskā tīkla struktūras analīzei izmantota sistēmbioloģijas nozarē lietota topoloģiskā jeb strukturālā analīze, kura pamatā ir balstīta uz grafu teorijas jēdzieniem un metodēm. Darba izstrādē lietoti *SBML (Systems Biology Markup Language)*, *GML (Graph Modelling Language)* standarti modeļu

struktūras importam un eksportam, kā arī *GV* (programmatūras rīka *GraphViz* iebūvētais failu formāts) un *PNG* failu formāti tīkla struktūras attēlošanai.

Nelineārās regresijas analīze lietota labākas nelineāras likumsakarības noteikšanai starp bioķīmisko tīklu struktūras atbilstošo topoloģisko parametru un kādu no faktoriālām pazīmēm –tīkla elementu skaitu, elementa puspakāpi u.c.

Bioķīmisko tīklu struktūras evolūcijas algoritmā lietotas varbūtības teorijas metodes.

Aprakstošās statistikas metodes izmantotas eksperimentu datu analīzei vairāku evolūcijas eksperimentu rezultātu salīdzināšanā.

Regresijas analīze lietota mutāciju operatoru ietekmes analīzei un labākas likumsakarības noteikšanai starp bioloģiskās sistēmas izdzīvošanas ilgumu un atbilstošo mutācijas operatoru, kā arī starp atbilstošo struktūras topoloģisko parametru un paaudzes kārtas numuru.

Atbilstošo mutāciju operatoru ietekmes uz bioķīmiskā tīkla struktūru un izdzīvošanas ilgumu novērtēšanai un vizualizācijai, kā arī evolūcijas dinamikas attēlošanai izmantoti divdimensiju grafiki, tajā skaitā grafiki ar divām dažādā mēroga asīm.

Zinātniskais jauninājums

Promocijas darba zinātniskais jauninājums.

- Izstrādāta genomam piesaistītas neviendabīga svarīguma bioķīmisko tīklu struktūras evolūcijas uzdevuma formalizācijas procedūra.
- Izstrādāts neviendabīga svarīguma genomam piesaistītas bioķīmisko tīklu struktūras evolūcijas modelēšanas algoritms gēna garumu nemainošām mutācijām.
- Izstrādāts neviendabīga svarīguma bioķīmisko tīklu struktūras evolūcijas modelēšanas un analīzes programmatūras rīka prototips *BINESA* (**B**iochemical **N**etwork **S**tructure **A**nalys**E**r).
- Izstrādāta bioķīmisko tīklu modeļu pielietojamības novērtēšanas pieeja, izmantojot bioķīmisko tīklu struktūras topoloģiskos parametrus.

Pētījuma tēzes

- Bioķīmisko tīklu reakciju svarīguma ignorēšana būtiski ietekmē bioķīmisko tīklu struktūras evolūcijas datormodeļa uzvedību.
- Modeļu šķēluma pielietojamības ātru novērtēšanu var balstīt uz tīkla struktūras topoloģisko parametru analīzi.
- Evolucionārā spiediena ietekme atšķiras dažāda svarīguma viendabīgiem tīkliem.
- Struktūras topoloģisko parametru dinamika evolūcijas procesā ataino bioķīmisko tīklu strukturālā modeļa kvalitātes dinamiku.

Praktiskā vērtība

Darba praktisko vērtību nosaka izstrādātā bioķīmisko tīklu struktūras evolūcijas modelēšanas algoritma pielietojums dažādu organismu bioķīmisko tīklu modelēšanā ņemot vērā evolūcijas bojāto struktūru ietekmi uz tīkla darbības biotehnoloģiskām modifikācijām biotehnoloģiskajos vai terapeitiskajos nolūkos. Bioķīmisko tīklu evolūcijas datoreksperimentus var pielietot arī terapiju izstrādē personalizētajā medicīnā.

Izstrādāto brīvpieejas programmatūras rīka prototipu *BINESA* var izmantot neviendabīga svarīguma genomam piesaisītās bioķīmisko tīklu struktūras evolūcijas datoreksperimentu izpildē un struktūras evolūcijas dinamikas modelēšanā.

Neviidabīga svarīguma bioķīmisko tīklu modelēšanu var pielietot stabilu tīklu struktūru izveidei sintētiskajā bioloģijā.

Darba rezultāti ir izmantojami arī bioprocesu stehiometrisko modeļu izstrādes procesā, vērtējot pieejamo modeļu kvalitāti un vienprātības pakāpi.

Promocijas darba struktūra un apjoms

Promocijas darbs ir uzrakstīts latviešu valodā, satur anotāciju, ievadu, 6 nodaļas, secinājumus, literatūras sarakstu un 15 pielikumus, tajā skaitā 19 tabulas, 111 attēlus, 27 formulas, kopā 184 lappuses. Darbā izmantoti 224 literatūras avoti.

1. BIOLOĢISKO SISTĒMU EVOLŪCIJA

Šūna ir uzskatāma par bioloģisku sistēmu. Tās darbību nosaka sarežģīti procesi, lielākā daļa no kuriem ir saistīti ar molekulu mijiedarbību bioķīmiskajos tīklos. Bioloģiskās sistēmas darbojas pēc tās gēnos ieprogrammētām instrukcijām, kuras tiek pārmantotas no paaudzes paaudzē un pakļautas mainībai.

Neskatoties uz to, ka vizuāli dažādu sugu pārstāvji ir krasi atšķirīgi, bioloģijas un ģenētikas ziņā starp tiem pastāv daudz vairāk līdzību, piemēram, cilvēkam 98.5% genoma sekvences ir līdzīga šimpanzes genoma sekvencai (Volkers, 2004).

Organismu evolūciju nodrošina mainīgā vide mijiedarbībā ar dažāda veida gēnu mutācijām, kuras maina organisma morfoloģiskās, fizioloģiskās vai bioķīmiskās īpašības. Bioķīmiskie tīkli ir viena no organisma neatņemamām struktūrām, kas arī ir pakļauta evolucionārām pārmaiņām. Par noteicošajiem evolūcijas dzinējspēkiem tiek uzskatītas mutācijas, kuras ievieš izmaiņas ģenētiskajā materiālā, un dabiskā izlase, kas nodrošina spēcīgāko un pielāgotāko pēcnācēju atlasī tālākai evolūcijai. Mutācijas ir iedalāmas trijās galvenajās grupās pēc iedarbes avota: gēna mutācijas, hromosomu mutācijas un genoma mutācijas. To rašanās biežums un ietekme uz bioķīmiskajiem procesiem un tos attēlojošajiem tīkliem ir atkarīga no daudziem faktoriem.

Mutācijas parasti notiek samērā reti, tomēr to parādīšanās biežumu palielina fizisko, ķīmisko un bioloģisko mutagēno faktoru ietekme. Evolūcijas ietekmi uz bioķīmiskajiem procesiem, t.sk. bioķīmisko tīklu struktūrām, ir svarīgi pētīt, izstrādājot biotehnoloģiskas modifikācijas vai ārstnieciskās terapijas, lai novērtētu evolūcijai pakļauto tīklu iespējamo atbildes reakciju, kas atspoguļojas bioloģiskās sistēmas funkcionēšanā un uzvedībā.

2. BIOĶĪMISKIE TĪKLI UN TO ANALĪZE

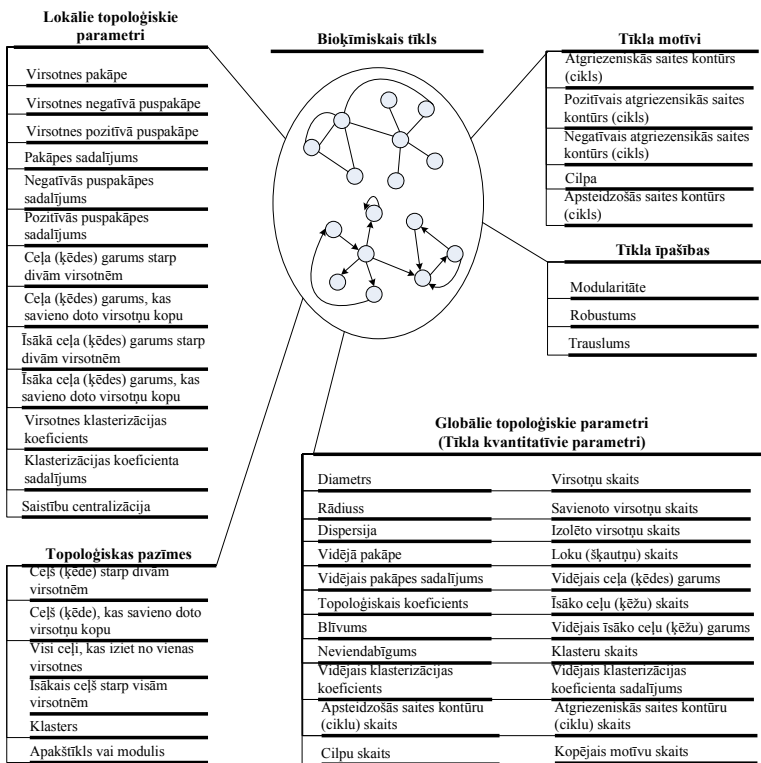
Dzīvajos organismos darbojas vielmaiņas, gēnu regulācijas un signāltīkli, kuri nosaka bioķīmiskās reakcijas, bioķīmiskos vai biofizikālos procesus. Bioķīmiskie tīkli var attēlot un atklāt mijiedarbību starp gēniem un to produktiem, proteīnu mijiedarbību, reaģentu, substrātu un to produktu mijiedarbību.

Bioķīmisko tīklu pamats ir struktūra. Lai izprastu bioloģiju sistēmas līmenī, sākumā ir nepieciešams izpētīt sistēmas struktūru, tajā skaitā arī bioķīmisko tīklu struktūru.

Bioķīmisko tīklu strukturālā jeb topoloģiskā analīze ir balstīta uz grafu teorijas jēdzieniem, metodēm un algoritmiem, un izmanto arī sistēmu teorijas jēdzienus, kā arī klasteru analīzes metodes un algoritmus. Strukturālā analīze sniedz ieskatu ne tikai par bioloģiskās sistēmas topoloģiskajām īpašībām, bet arī par bioloģiskās sistēmas dinamiskajām īpašībām, jo dinamiskas uzvedības fundamentālas īpašības bieži vada un nosaka tīkla struktūra (Klamt u.c., 2006).

Izpētot zinātnisko literatūru, publikācijas un analizējot sistēmbioloģijas nozarē izstrādātos programmatūras rīkus, kā arī to aprakstus, tika konstatētas vairākas tīkla topoloģiskās īpašības, parametri un pazīmes, kurus var iedalīt četrās grupās (Rubina, Stalidzans, 2010a): tīkla metrikas, tīkla motīvi, tīkla topoloģiskie parametri un topoloģiskās pazīmes (sk. 1.att.).

Tīkla topoloģiskie parametri atkarībā no pētāmā tīkla līmeņa (tīkla elementu līmenis, kad tiek pētīti atsevišķi individuālie tīkla elementi vai tīkla līmenis, kad tiek pētīts tīkls kā viens veselums) var tikt iedalīti lokālajos un globālajos parametros. Lokālie topoloģiskie parametri raksturo atsevišķus tīkla elementus vai komponentes, bet globālie parametri raksturo tīklu kopumā un to iegūšanai izmanto lokālos parametrus.



1.att. Biokīmisko tīklu topoloģiskie rādītāji

3. BIOĶĪMISKO TĪKLU EVOLUCIONĀRĀS ATTĪSTĪBAS MODELĒŠANA

Galvenais biokīmisko tīklu modelēšanas mērķis ir papildināt un attīstīt pētnieku izpratni par bioloģisko sistēmu lokālām un globālām īpašībām un uzvedību, kuru sistēma demonstrē kā atbildes reakciju uz dažāda veida kairinājumiem.

Izpratne par to, kā tīkli izveidojušies ir fundamentālā problēma reālās pasaules kompleksajos tīklos un var sniegt skaidrību par tīklu struktūru un to darbību (Chen u.c., 2009c).

Lai pētītu biokīmisko tīklu topoloģiskās īpašības pētnieki ir izmantojuši galvenokārt trīs tīklu modeļus, kurus raksturo noteikti virsošnes pakāpes un klasterizācijas koeficienta sadalījumi: nejaušais tīkls, bezmēroga un hierarhiskais tīkls. Turklāt, lai pētītu reālas pasaules tīklu topoloģiskās īpašības un īpatnības no evolucionārā viedokļa, pieņemot, ka tīkla topoloģija ir izveidota virkņu tīkla bloku un evolūcijas notikumu rezultātā, dažādas pētnieku grupas

piedāvāja arī dažādus tīkla attīstības modeļus, piemēram, duplikācijas mutāciju modeļi, duplikācijas mutāciju modeļi ar papildinājumu, duplikācijas novirzīšanās modeļi, delēcijas-duplikācijas-novirzīšanas modeļi, statistiskie gadījuma tīklu modeļi (*random static network models*), mazās pasaules tīkla modeļi, Būla tīklu modelis (*Boolean network model*). Minētie modeļi pēta tīklu attīstību ar mērķi izprast to izveides un organizācijas principus evolūcijas procesā.

Iepriekšminētajiem un praksē lietotajiem modeļiem bioķīmisko tīklu īpašību un to attīstības jeb evolūcijas izpētē ir vairāki trūkumi. Pirmkārt, šo modeļu pamatā izmantotajā evolūcijas pieeja pēta tīkla attīstību vispārīgi, neņemot vērā svarīgas bioķīmisko procesu īpašības. Viena no šādām īpašībām ir apskatāmo jeb pētāmo procesu svarīgums. Ne visi bioloģiskās sistēmas procesi ir vienādi svarīgi un dabiskajā atlasē pēcnācēji ar letālām mutācijām, kuras ir skārušas dzīvībai ļoti svarīgus procesus, tiek nošķirti, bet pēcnācējiem ar neitrālām vai labvēlīgām mutācijām tiek sniegtas lielākas iespējas piedalīties tālākā evolūcijā.

Ir vairāki bioķīmiskie procesi, kuri ir saglabājušies gandrīz visos dzīvajos organismos tādējādi norādot, ka organismi ar būtiskām novirzēm šos procesus realizējošo proteīnu gēnos evolūcijas gaitā iet bojā. Viens no šādiem procesiem ir glikolīze (Romano un Conway, 1996). Tipiskākais glikolīzes process ir Embdena-Meijerhofa-Parnas glikolīze. Vēl ir arī Entnera-Dudorova glikolīze, bet kopumā ar enerģijas nodrošināšanu saistītos procesus un īpaši glikolīzi var uzskatīt par vitāli svarīgiem procesiem, kuriem evolūcijas gaitā nav izveidojušās būtiskas alternatīvas visu dzīvības formu spektrā. Pastāv arī virkne citu procesu, kas nopietni ietekmē organisma dzīvotspēju. Piemēram, tādi ir proteīnu, nukleīnskābju, lipīdu un citu komponentu sintēzes bioķīmiskie ceļi (Copley, 2000). Tajā pašā laikā ir bioķīmiskie ceļi, kuru izmaiņas var kompensēt citi mehānismi un šādu bioķīmisko ceļu bojājumi būtisku iespaidu uz organisma dzīvotspēju neatstāj.

Autore piedāvā vienkāršoti sadalīt procesus trijās grupās pēc to svarīguma. Pirmā procesu grupa ir vitāli svarīgi un bez tiem bioloģiskā sistēma nevar nodrošināt savu darbību un būt dzīvotspējīga. Otrās grupas procesi nosaka eksistences kvalitāti, un tikai daudzu procesu traucējumu dēļ var ietekmēt dzīvotspēju. Trešās grupas procesi skar tikai nebūtiskas bioloģiskās sistēmas īpašības, piemēram, dažas vizuālās pazīmes vai alternatīvus bioķīmiskos procesus. Šī iemesla dēļ, šo procesu grupu īpašības ir jādefinē un jāizdala atsevišķi. Tātad izmaiņām genoma sekvencē, kuras nosaka un regulē pirmās un citu grupu procesus, nebūs vienlīdzīga ietekme uz sistēmu. Pat nebūtiskas mutācijas vitāli svarīga procesa regulējošajā gēnā var izrādīties daudz bīstamākas. Tās var ieviest būtiskas izmaiņas bioloģiskās sistēmas uzvedībā un izraisīt novirzes tās darbībā. Autore pieņem, ka vitāli svarīgi procesi ir līdzīgi lielā daļā bioloģisko sistēmu un pieļauj tikai nelielas atšķirības ģenētiskajā materiālā jeb gēnu sekvencēs, kuri šos procesus regulē un nosaka.

Otrs izmantojamo metožu trūkums ir piesaistīta genoma neesamība, izņemot mākslīgo genoma modeli. Izmaiņas tīkla līmenī tiek tikai teorētiski paskaidrotas ar genoma līmenī notiekošajām izmaiņām, nevis tiek ieviestas notikušo mutāciju ietekmē.

Trešais izmantojamo metožu trūkums ir tālums no reālā evolūcijas procesa un galvenokārt tikai divu vai trīs mutāciju tipu iekļaušana modeļos: gēnu duplikācija (visa genoma duplicēšana, lokāli norobežoto gēnu duplicēšana, retrotranspozīcija) un delēcija (Farid, Christensen, 2006, Yamada u.c., 2009, Yamada, Bork, 2009, Wagner, 2009), kā arī novirzīšanās (*divergence*), tajā skaitā punktveida mutāciju ietekmē (Aldana u.c., 2007, Gibson, Goldberg, 2011). Šīs mutācijas ir atzītas par būtiskākajiem tīkla evolūcijas dzinējspēkiem. Gēna duplikācija nozīmē tīkla virsotnes ievietošanu, kā arī visu dublētas virsotnes saišu pievienošanu. Zaudējot kādu gēnu, tiek zaudēta ne tikai virsotne, bet arī visas tās saites tīklā.

Otrā notikumu grupa – saišu diverģences notikumi - atspoguļojas saites pievienošanā vai izdzēšanā, teorētiski pamatojoties uz tādām ģenētiskajām izmaiņām, kuras nemaina gēna sekvenci pilnībā vai maina gēna regulāciju. Šādas ģenētiskas mutācijas var būt punktveida mutācijas, nukleotīdu insercijas vai delēcijas (Wagner, 2003), vai mutācijas, kuras skar gēnu regulāciju (Noort u.c., 2004, Yamada, Bork, 2009, Yamada u.c., 2009). Tātad ir vairāki citi mutāciju tipi, kuri var ietekmēt tīkla evolūciju. Tie nemaina gēna sekvenci kopumā, bet gan modificē gēnu daļēji vai tā regulāciju, kas var izpausties tīkla līmenī kā saites pievienošana vai izdzēšana. Kā uzsver Yamada ar kolēģiem (Yamada u.c., 2009), izpētot vielmaiņas un proteīnu mijiedarbības tīklus, tīkla virsotņu un saišu evolūcija ir saistīta ar šūnas ģenētisko materiālu, bet saites var mainīties visu laiku. Ja virsotnes netiek skartas, saišu aizstāšana var notikt bez gēnu duplikācijas un saišu izmaiņas mēdz notikt ar lielāku frekvenci nekā virsotņu izmaiņas. Arī Bergs ar kolēģiem ir uzsvēris (Berg u.c., 2004), ka saišu ieviešanas vai dzēšanas koeficients mutāciju iespaidā ir par pakāpi augstāks nekā tīkla attīstības koeficients ar duplikācijām. Lēnāka gēnu duplikācija (Berg u.c., 2004), kā arī delēcija (Wagner, 2009) iespaido tikai tīkla izmēru. Tieši saišu dinamika darbojas kā dominējošais evolūcijas dzinējspēks bezmēroga tīklu struktūras veidošanā, nevis virsotņu duplikācija (Farid, Christensen, 2006).

Darba autore piedāvā bioķīmisko tīklu struktūras evolūcijas modelēšanas algoritmu, kas novērš analizēto tīklu evolucionāras attīstības modeļos lietojamo algoritmu minētos trūkumus. Pirmkārt, evolūcijas algoritmā vērā tiek ņemta procesu svarīguma īpašība un tīkla procesiem (reakcijām jeb saitēm) ir iespējams definēt svarīguma pakāpi. Otrkārt, tīkla virsotņu un saišu evolūcija ir saistīta ar šūnas ģenētisko materiālu, līdz ar to autores piedāvātajā algoritmā tīkla struktūras evolūcija tiek balstīta uz piesaistīta genoma evolucionārām izmaiņām, kas ir mutāciju un dabiskās izlases rezultāts. Treškārt, tā kā saišu izmaiņas mēdz notikt biežāk nekā virsotņu duplikācija un saišu dinamika ir atzīta par primāro evolūcijas virzītājspēku bezmēroga tīklu

struktūras veidošanā, autores piedāvātajā bioķīmisko tīklu struktūras evolūcijas algoritmā tīkla virsotņu skaits netiek mainīts, galveno uzmanību pievēršot saišu dinamikai.

Bioķīmisko tīklu struktūras evolūcijas modelēšanas procedūra

Darba ietvaros izstrādātajā bioķīmisko tīklu struktūras evolūcijas procedūrā un algoritmā tiek lietota šāda terminoloģija:

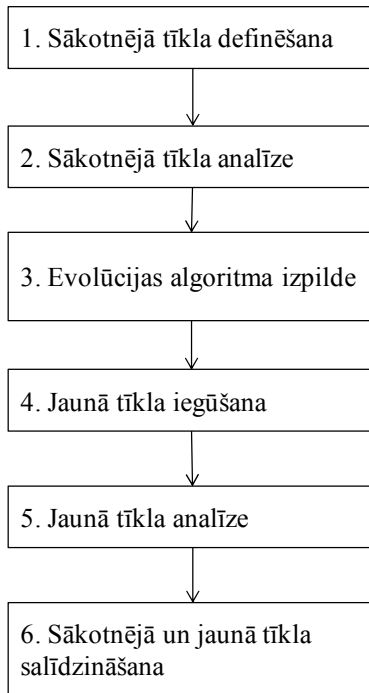
- bioķīmiskais tīkls – reakciju kopums, kas attēlo kādu bioķīmisko procesu(s);
- bioķīmiskā tīkla struktūra – tīkla elementu un to mijiedarbības attēlojums;
- gēns – ģenētiskās informācijas vienība, kas tiek attēlota četru teksta simbolu virknes veidā un piesaistīta katrai atsevišķai reakcijai;
- genoms – gēnu kopums, kas tiek piesaistīts konkrētajam bioķīmisko tīklu modelim;
- mutācija – gēna sekvenču izmaiņa, kas notiek ar noteiktu varbūtību;
- mutāciju process – process, kurā katrs genoma gēns tiek pakļauts mutācijām;
- paaudze – viena skaitļošanas iterācija, kas ietver mutāciju procesu un nākamās paaudzes pēcteča izvēli.

Evolucionārās bioķīmisko tīklu struktūras izmaiņas notiek genoma līmenī notiekošo mutāciju ietekmē. Tā kā gēni nosaka un regulē katra bioķīmiskā tīkla uzbūvi un darbību, tad, lai pētītu bioķīmisko tīklu evolūciju, kas atspoguļojas tīkla struktūrā, ir nepieciešams piesaistīt tīkla saitēm un procesiem gēnus nukleotīdu sekvenču veidā un definēt noteiktos genoma izmaiņu novērtēšanas kritērijus. Tādējādi bioķīmisko tīklu evolūcijas modelēšana ir realizējama:

- 1) sasaistot gēnus ar to nukleotīdu sekvenču un ar tīkla reakcijām, kuras gēni nosaka,
- 2) izpildot genoma evolūciju jeb gēnu sekvenču izmaiņu, kas var rasties dažāda tipa mutāciju ietekmē, izmantojot datormodeli;
- 3) ģenerējot struktūras izmaiņas, pamatojoties uz gēnu sekvenču izmaiņām, izmantojot datormodeli.

Bioķīmisko tīklu struktūras evolūcijas modelēšanas procedūra ietver sešus secīgus etapus (sk. 2.att.).

Pirmais procedūras etaps ir sākotnējās tīkla struktūras datu definēšana, kas iekļauj trīs galvenos apakšetapus: 1) tīkla virsotņu un saišu definēšana manuāli vai ielādējot no esošiem tīkla modeļiem, piemēram, *SBML* modeļa; 2) sākotnējā organisma genoma datu definēšana, ģenerējot automātiski gēnu testa sekvenču vai izmantojot genoma datus no publiski pieejamām datu bāzēm; 3) piesaistot genomu jeb atsevišķos gēnus tīkla reakcijām, t. i., saitēm.



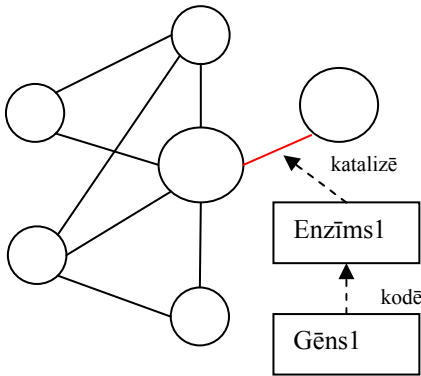
2.att. Bioķīmisko tīklu struktūras evolūcijas modelēšanas procedūra

Tīkla evolūcija balstās uz ģenētiskā materiāla izmaiņām. Izstrādātajā algoritmā evolūcija tiek realizēta ģenoma sekvences līmenī un tiek transformēta tīkla struktūras izmaiņās. Atbilstoši molekulāras bioloģijas centrālajai dogmai (sk. 1.2.apakšnodaļu) katrai tīkla saitei ir iespējams piesaistīt gēnu (Rodrigues, Wagner, 2011, Wagner, 2011), kurš kodē attiecīgās reakcijas norisi nodrošinošo enzīmu (sk. 3.att.). Katram gēnam atbilst noteikta nukleotīdu sekvenca, kura tiek definēta četrus simbolus A, C, T un G virknes veidā.

Struktūras evolūcijas modelēšanas algoritmā tiek pieņemts, ka visiem gēniem ir vienāds garums, un evolūcijas algoritma izpildes procesā gēna garums paliek nemainīgs, piemēram, 1000 nukleotīdu gari gēni.

Otrais procedūras etaps ietver sākotnējā tīkla struktūras analīzi. Sākotnējās struktūras analīze ir nepieciešama, lai noteiktu izvēlēta tīkla modeļa piemērotību pētījuma mērķim. Sākotnējās tīkla struktūras analīze sniedz vairākus struktūras topoloģiskos rādītājus un motīvus, kurus var izmantot struktūras novērtēšanā. Tīkla struktūras analīzē var tikt iegūti šādi topoloģiskie parametri: virsotnes pakāpe un puspakāpes (negatīvā, pozitīvā), pakāpes vai puspakāpju sadalījumi, elementu skaits, saistīto un izolēto virsotņu skaits, saišu skaits, vidējais klasterizācijas koeficients, kaimiņu skaita sadalījums,

klasterizācijas koeficienta sadalījums, tiek noteikti tādi tīkla motīvi kā atgriezeniskās saites (*feedback loop*) jeb vadības cikli, cilpas (*self-loop*).



ACTGCGCTGCGTGTGCACTGCAAATGTT

3.att. **Genoma piesaiste struktūrai**

Trešais un ceturtais evolūcijas procedūras etaps iekļauj evolūcijas algoritma izpildi, kuru veido divi posmi: genoma evolūcija un struktūras evolūcija (sk. 3.2.apakšnodaļu). Lai realizētu evolūcijas algoritmu, kas nodrošina nākamo paaudžu organisma genomu iegūšanu, sākumā ir nepieciešams definēt evolūcijas parametrus. Evolūcijas algoritms tiek izpildīts n reizes, kur n ir lietotāja definētais paaudžu skaits un kura izpildes gaitā tiek ģenerēti n paaudžu genomi. Piemēram, lai iegūtu pirmās paaudzes organisma genomu, tiek izmantots sākotnējais organisma genoms, bet, lai iegūtu i -tās paaudzes organisma genomu, tiek izmantots $i-1$ paaudzes organisma genoms.

Jaunās paaudzes genoms tiek iegūts, realizējot divus bioloģisko sistēmu evolūciju noteicošos procesus: mutāciju un dabiskās izlases procesus, kuru realizācija ir balstīta uz literatūrā sniegto informāciju par ģenētiskā materiāla mutācijām, un dabisko izlasi (sk. 1.2.apakšnodaļu).

Lai iegūtu jaunas t -ās paaudzes genomu, tiek ģenerētas iepriekšējās paaudzes $t-1$ organisma genoma 10 kopijas, kuras tālāk tiek pakļautas mutāciju procesam. No 10 genoma kopijām tiek atlasīti iespējamie pēcteču kandidāti, novērtējot genomu atbilstību sākotnējam etalona genomam. Mutāciju procesā un kandidātu atlases procesā tiek iegūti M genomu kandidāti ar alternatīvām gēnu formām, no kuriem tālākai evolūcijai tiek izvēlēts viens genoms ar noteiktu varbūtību Pe_i . Katra atsevišķa jaunizveidota genoma varbūtība Pe_i tikt izvēlētam tālākai evolūcijai ir atkarīga no genoma kandidāta atbilstības sākotnējam genomam, pieņemot, ka sākotnējais genoms ir etalona genoms ar augstākajām iespējamajām dzīvotspējas pazīmēm.

Izvēlētajām paaudzēm tiek ģenerētas tīkla struktūras izmaiņas, balstoties uz atbilstošās paaudzes genoma evolucionārajām izmaiņām.

Piektais procedūras etaps ietver evolūcijā iegūtas tīkla struktūras topoloģisko analīzi, kuras rezultāti tiek izmantoti nākamajā procedūras etapā.

Sestais procedūras etaps iekļauj sākotnējās un jauniegūtās struktūras salīdzināšanu. Tīklu strukturālajā analīzē iegūstamie topoloģiskie rādītāji kļūst par vairāku struktūru salīdzināšanas kritērijiem, t.i., evolūcijā radušos un sākotnējās struktūras salīdzināšanā radušos izmaiņu noteikšanai un novērtēšanai. Salīdzinot sākotnējās un jauniegūtās struktūras parametrus, to izmaiņas dinamiku un motīvus, tiek noteiktas līdzības un atšķirības starp tām, lai formulētu secinājumus par izvēlēto mutāciju tipu ietekmi uz tīkla struktūru.

Bioķīmisko tīklu struktūras evolūcijas modelēšanas algoritms

Evolūcijas algoritms (sk. 2.att. 3.etapu) ietver divus posmus: genoma evolūcijas un struktūras evolūcijas posmu. **Genoma evolūcija** tiek veikta n reizes, ģenerējot n paaudžu genomus. Katru reizi izpildot evolucionāro algoritmu, lai iegūtu t -ās paaudzes genomu ($t=1,2,..n$), tiek ģenerētas iepriekšējās paaudzes genoma 10 kopijas. Genoma kopiju skaits ir maināms, tomēr genoma kopiju mazāks skaits samazina jaunās paaudzes pēcnācēja genoma kandidātu skaitu, kas var ietekmēt evolūcijas procesa rezultātus. Genoma kopijas tiek pakļautas mutāciju procesam, kas var ietvert vairāku mutāciju operatoru izpildi.

- **Punktveida mutāciju operatora** iedarbē gēna sekvencē nejauši izvēlēts nukleotīds tiek mainīts pret citu, piemēram, A tiek mainīts pret C.
- **Nukleotīdu inversijas operatora** izpildē nejauši izvēlēts gēna iecirknis no k -tās līdz l -tai pozīcijai tiek sakārtots otrādi, t.i., nukleotīdi tiek izkārtoti apgrieztā secībā, piemēram, nukleotīdu secība ACTGTGATCGCGTAATGGC no 7 līdz 11 pozīcijai tiek pārveidota par ACTGTGCGCTAGTAATGGC sekvenci.
- **Misensmutāciju operators** tiek izmantots genoma kopiju atbilstības etalona genomam novērtēšanā. Ja tiek izvēlēts šis mutācijas tips, tad genoma kopijas tiek salīdzinātas kodonu veidā jeb pa trīs secīgiem nukleotīdiem jeb tripletiem, kas kodē atbilstošās aminoskābes (sk. 2.3.apakšnodāļu). Kā tika minēts iepriekš, vienu aminoskābi var kodēt vairāki dažādi nukleotīdu tripleti. Punktveida mutācijā, pat ja nukleotīds ir mutējies, kodējamā aminoskābe var palikt tā pati. Genoma sekvencu salīdzināšanas procesā (salīdzinot jaunu evolūcijā iegūto sekvenci ar etalona genoma šī gēna sekvenci) tiek pārbaudīts, vai kodējamā aminoskābe ir palikusi tā pati vai ir mainījusies.
- **Nonsensmutācijas operators** līdzīgi misensmutācijas operatoram, tiek lietots genoma kopiju atbilstības etalona genomam novērtēšanai. Šis operators tiek izmantots noteikšanai, vai atbilstošo kodonu kodējamā aminoskābe ir mainījusies (mainījusies par stop kodonu) vai nav salīdzinājumā ar etalona genoma kodētajām aminoskābēm.

- **Duplikācijas operatora** izpildē tiek kopēts nejausi izvēlēts gēns, un gēnu skaits genomā palielinās par vienu gēnu.
- **Delēcijas operatora** izpildē tiek dzēsts nejausi izvēlēts gēns, un gēnu skaits genomā samazinās par vienu gēnu.
- **Inversijas operatora** iedarbē visa gēna nukleotīdu virkne tiek pārkārtota apgriezta secībā.
- **Translokācijas operatora** izpilde nosaka dažādās nehomologajās hromosomās atrodošos gēnu pārraušanu nejausi izvēlētā gēna k -tajā pozīcijā un hromosomu atrauto galu apmaiņu starp nehomologajām hromosomām.

Evolūcijas modelēšanas procedūras pirmajā etapā ir paredzēts izvēlēties atbilstoši konkrēta pētījuma mērķim, kādus no mutāciju operatoriem iekļaut evolūcijas procesā un ar kādu varbūtību mutāciju operatoram darboties.

Pēc mutāciju procesa izpildes ***t*-tās paaudzes iegūtie 10 genomi tiek** salīdzināti ar sākotnējo genomu un **novērtēti**. Katrs jauniegūtais genoma gēns tiek salīdzināts ar sākotnējo tā sekvenci (sākotnējā genoma atbilstošā gēna sekvenci) un ar visu pārējo sākotnējo gēnu sekvencēm, aprēķinot atbilstības koeficientus: Rgk_i – atbilstības koeficients *i*-tā gēna sākotnējai sekvencei un Qgk_{ij} – atbilstības koeficienti pārējiem sākotnējo *j*-to gēnu sekvencēm, kur *j* ir gēna kārtas numurs. Ja ir izvēlēti misensmutāciju un/vai nonsensmutāciju operatori, tad sekvences tiek salīdzinātas pa tripletiem, t.i. pa trīs secīgiem nukleotīdiem. Ja neviens no minētajiem operatoriem nav izvēlēts, tad gēnu sekvences tiek salīdzinātas pa vienam nukleotīdam. Salīdzināšanā aprēķinātais Rgk_i koeficients raksturo alternatīvas *i*-tā gēna formas atbilstības daļu sākotnējam gēnam un var pieņemt vērtības robežās [0; 1]. Savukārt Qgk_{ij} koeficienti var pieņemt vērtības no [0; 1] un raksturo alternatīvas gēna formas atbilstības daļu visiem pārējiem etalona genoma gēniem, kur *i* ir salīdzināmā gēna kārtas numurs, bet *j* ir pārējo gēnu kārtas numurs..

Ja *i*-tā gēna atbilstības koeficients $Rgk_i < 0.2$ un tas nosaka mazsvarīgo reakciju, tad tiek noteikts maksimālais no atbilstības koeficientiem $\max Qgk_i = \max(Qgk_{ij})$. Ja $\max Qgk_i > 0.2$, tad tiek pieņemts, ka *i*-tais gēns ir mutējis par *j*-to gēnu un sācis pildīt tā funkcijas.

Ja *i*-tā gēna atbilstības koeficients $Rgk_i < 0.5$ un tas nosaka kvalitatīvo reakciju, tad tiek noteikts maksimālais no atbilstības koeficientiem $\max Qgk_i = \max(Qgk_{ij})$. Ja $\max Qgk_i > 0.5$, tad tiek pieņemts, ka *i*-tais gēns ir mutējis par *j*-to gēnu un sācis pildīt tā funkcijas.

Ja *i*-tā gēna atbilstības koeficients $Rgk_i < 0.7$ un tas nosaka vitālo reakciju, tad tiek noteikts maksimālais no atbilstības koeficientiem $\max Qgk_i = \max(Qgk_{ij})$. Ja $\max Qgk_i > 0.7$, tad tiek pieņemts, ka *i*-tais gēns ir mutējis par *j*-to gēnu un sācis pildīt tā funkcijas.

Pēc mutāciju procesa notiek **jaunās paaudzes genoma kandidātu atlases** process. Atbilstoši dabiskajai atlasei, izdzīvo spēcīgākais un pielāgotākais. Autores piedāvātajā algoritmā, par stiprāko tiek uzskatīts organisms ar tādu genomu, kuram ir lielāka atbilstība etalona genomam un kurš

saglabā visus vitālo reakciju regulējošos gēnus, un pēc iespējas vairāk kvalitatīvo reakciju regulējošo vai nosakošo gēnu, bet vismaz vienu obligāti. Tādējādi piešķirot spēcīgākajam lielākas iespējas tikt izvēlētam par jaunas paaudzes genomu.

Jaunās paaudzes genoma kandidātam ir jāatbilst kādam no tālākminētajiem nosacījumiem.

- Ja modelējamais bioķīmiskais tīkls ietver tikai vitāli svarīgas reakcijas, tad genoma visu gēnu atbilstības koeficientiem jābūt $Rgk > 0.7$.
- Ja modelējamais tīkls ietver tikai kvalitatīvas reakcijas, tad genoma vismaz viena gēna atbilstības koeficientam jābūt $Rgk > 0.5$.
- Ja modelējamais tīkls ietver mazsvarīgas reakcijas, tad genoma vismaz viena gēna atbilstības koeficientam jābūt $Rgk > 0.2$.
- Ja modelējamais tīkls ietver jauktas svarīguma pakāpes reakcijas saites, tad genoma visu vitālo reakciju nosakošo gēnu atbilstības koeficientiem jābūt $Rgk > 0.7$, vismaz viena kvalitatīvo reakciju nosakošā gēna atbilstības koeficientam jābūt $Rgk > 0.5$.

Tātad no 10 jaunās paaudzes kandidātu pretendentiem tiek izslēgti tie pretendenti, uz kuriem attiecas kāds no tālākminētajiem nosacījumiem:

- 1) genoms, kurā vismaz viens vitāli svarīgo reakciju nosakošs gēns atšķiras no etalona genoma vairāk nekā par 30% ar atbilstības koeficientu $Rgk < 0.7$;
- 2) genoms, kurā visi kvalitatīvās reakcijas nosakošie gēni atšķiras no etalona genoma vairāk nekā par 50% ar atbilstības koeficientu $Rgk < 0.5$.

Lai nodrošinātu spēcīgākā kandidāta izvēli, tiek izpildīta virkne pārveidojumu, nosakot katra kandidāta varbūtības koeficientu tikt izvēlētam par jaunas paaudzes pēcnācēju.

Katram kandidātam tiek aprēķināts genoma summārais atbilstības koeficients $sumRgk_j$, kas ir vienāds ar visu j -tā genoma gēnu atbilstības koeficientu Rgk_i summu.

Genoma kandidāta summārais atbilstības koeficients $sumRgk_j$ tiek normēts, izdalot tā summāro atbilstības koeficientu ar tā kopējo gēnu skaitu m un tiek saukts par normēto jeb genoma atbilstības koeficientu $gRgk_j$.

Tiek aprēķināta visu jaunās paaudzes genoma kandidātu normēto atbilstības koeficientu summa $TotalRgk$.

Katra genoma kandidāta varbūtības koeficients Pe_j tiek aprēķināts, izdalot tā normēto atbilstības koeficientu $gRgk_j$ ar visu kandidātu normēto atbilstības koeficientu summu $TotalRgk$.

Ja modelējamais tīkls ietver kvalitatīvas reakcijas, tad genoma kandidāta varbūtības koeficients tiek aprēķināts, izdalot tā normēto atbilstības koeficientu $gRgk_j$, kas ir reizināts ar aktīvo kvalitatīvo reakciju skaita daļu, ar visu kandidātu normēto atbilstības koeficientu summu $TotalRgk (I)$:

$$Pe_j = \frac{gRgk_j \cdot \frac{q}{Tq}}{TotalRgk} , \quad (1)$$

kur q – aktīvo kvalitatīvas reakcijas nosakošo gēnu skaits (bez atkārtojumiem),

Tq – kopējais kvalitatīvas reakcijas nosakošo gēnu skaits tīklā (bez atkārtojumiem).

Mutāciju procesa un kandidātu atlasē procesa izpildē tiek iegūti M genomu kandidāti ar alternatīvām gēnu formām, no kuriem tālākai evolūcijai tiek izvēlēts viens genoms ar varbūtību Pe_j .

Pamatojoties uz izvēlēta genoma gēnu atbilstības koeficienta vērtību un tā regulējamās reakcijas svarīguma pakāpi, tiek aprēķināts saites intensitātes rādītājs $Ints_j$. Intensitātes rādītājs nosaka, cik intensīvi saite darbojas. Lai ievērotu reakciju svarīgumu, tiek pieņemts, ka vitālas, kvalitatīvas un mazsvarīgas reakciju intensitāte mainās pēc dažādiem likumiem.

1.svarīguma pakāpes reakciju $Ed=1$ (vitālas reakcijas jeb to saites) reakcijas intensitāte mainās pēc pakāpes likuma $f(x)=x^{-2}$;

2.svarīguma pakāpes reakciju $Ed=2$ (kvalitatīvas reakcijas jeb to saites) reakcijas intensitāte mainās pēc lineārā likuma $f(x)=x$;

3.svarīguma pakāpes reakciju $Ed=3$ (mazsvarīgas reakcijas jeb to saites) reakcijas intensitāte mainās pēc polinomiāla likuma $f(x)=1.3x^3-3.42x^2+3.12x$.

Otrais evolūcijas algoritma posms ietver struktūras evolūciju, kurā tiek iegūta jaunās paaudzes tīkla struktūra. Balstoties uz katra gēna atbilstības koeficientu Rgk_i un reakcijas svarīguma pakāpi, tiek ģenerētas tīkla struktūras izmaiņas.

- Ja vitālas reakcijas regulējošā gēna atbilstības koeficients $Rgk_i < 0.7$, tad reakcija jeb tās veidotās saites tiek dzēstas.
- Ja vitālas reakcijas regulējošā gēna atbilstības koeficients ir robežās $0.7 \leq Rgk_i < 0.9$, tad reakcijas intensitāte tiek pavājināta atbilstoši iepriekšminētajam likumam.
- Ja kvalitatīvas reakcijas regulējošā gēna atbilstības koeficients $Rgk_i < 0.5$, tad reakcija jeb tās veidotās saites tiek dzēstas.
- Ja kvalitatīvas reakcijas regulējošā gēna atbilstības koeficients ir robežās $0.5 \leq Rgk_i < 0.8$, tad reakcijas intensitāte tiek pavājināta atbilstoši iepriekšminētajam likumam.
- Ja mazsvarīgas reakcijas regulējošā gēna atbilstības koeficients $Rgk_i < 0.2$, tad reakcija jeb tās veidotās saites tiek dzēstas.
- Ja mazsvarīgas reakcijas regulējošā gēna atbilstības koeficients ir robežās $0.2 \leq Rgk_i < 0.5$, tad reakcijas intensitāte tiek pavājināta atbilstoši iepriekšminētajam likumam.

4. PROGRAMMATŪRAS RĪKS BINESA

Promocijas darba ietvaros tika izstrādāts arī programmatūras rīka prototips *BINESA*, kurā ir realizēts autores piedāvātais bioķīmisko tīklu struktūras evolūcijas algoritms. Rīks *BINESA* ļauj izpildīt visus darba ietvaros piedāvātas bioķīmiskā tīkla struktūras evolūcijas procedūras etapus. Turklāt ar rīka *BINESA* palīdzību ir iespējams izpildīt evolūcijas datoreksperimentus un novērtēt iegūtos rezultātus.

BINESA modelēšanas rīka prototipu var lietot šādiem mērķiem:

- 1) struktūras evolucionāro izmaiņu modelēšanai, ņemot vērā procesu svarīguma pakāpi;
- 2) tīklam piesaistītā genoma evolūcijas imitācijai (sk.4.1.nodaļu);
- 3) jaunu tīkla modeļu izveidei;
- 4) *SBML* un *GML* formātu tīkla modeļu struktūras lejuplādēšanai;
- 5) bioķīmiskā tīkla struktūras topoloģiskajai analīzei (lokālo un globālo topoloģisko parametru aprēķināšanai, motīvu noteikšanai);
- 6) divu tīkla modeļu struktūru analīzei un salīdzinājumam pēc to topoloģiskajiem rādītājiem;
- 7) vairāku struktūru topoloģisko parametru aprēķināšanai un eksportam *CSV* faila formātā tālākas apstrādes, analīzes un vizualizācijas nolūkiem;
- 8) sākotnējās un evolūcijā iegūtās tīkla struktūras analīzei, pamatojoties uz evolūcijas procesa datiem un evolūcijas rezultātiem, kā arī struktūras topoloģiskajiem parametriem;
- 9) bioķīmisko tīklu struktūras vizualizācijai ar dažāda svarīguma un intensitātes saišu izcelšanu (sk. 8.pielikumu).

Rīka prototips *BINESA* ir programmēts *Visual Basic* programmēšanas valodā un realizēts *MS Access*, izmantojot *DAO* datu pieejas tehnoloģiju. *BINESA* var izmantot kā patstāvīgu programmatūru, kuras darbībai ir nepieciešams šāds programmnodrošinājums: *Windows XP* operētājsistēma un *Microsoft Access 2007* vai tā jaunāka versija, un dators, kura tehniskie rādītāji atbilst programmnodrošinājuma stabilas darbības rekomendējamiem rādītājiem.

BINESA pētāmais bioķīmiskā tīkla modelis ir saistīts ar atbilstošu genomu, un tajā ir iespējama trīs līmeņu procesu/reakciju svarīguma iestatīšana. *BINESA* ir realizētas tikai tādi mutāciju operatori, kuri nemaina gēnu garumu, kas samazina gēnu salīdzināšanas skaitļošanas izmaksas. Piedāvātais genomam piesaistītais bioķīmiskā tīkla struktūras evolūcijas algoritms sniedz iespēju novērtēt vienas evolūcijas mehānismu daļas ietekmi uz bioķīmiskā tīkla struktūru.

Izstrādātais programmatūras rīka prototips *BINESA* ļauj ne tikai modelēt bioķīmiskā tīkla struktūras izmaiņas atbilstoši piedāvātajam evolucionārajam algoritmam un izstrādātajai evolūcijas modelēšanas procedūrai, bet arī iegūt un novērtēt modelējamās tīkla struktūras evolūcijas dinamiku.

5. BIOĶĪMISKO TĪKLU MODEĻU PIELIETOJAMĪBAS NOVĒRTĒJUMS

Sekvencēšanas tehniku ātruma pieauguma dēļ ir kļuvis iespējams rekonstruēt bioķīmisko reakciju tīklus daudzos organismos. Tas būtu liels atvieglojums un priekšrocība modelēšanā, ja pētniekam būtu iespēja novērtēt pieejamo modeļu kvalitāti un pielietojamību, izvēlēties labāko vai piemērotākās daļas no publicēto tīkla modeļu jauna, pētāmajai problēmai atbilstoša modeļa izveidei. Atšķirības vai pretrunas rekonstrukcijās, it īpaši genoma izmēra rekonstrukcijās, var atspoguļot vai norādīt uz dažādu autoru lietotu modeļa izveides stilu, atšķirīgiem viedokļiem par modelētajiem procesiem vai arī viedokļu saskaņotību par interesējošo tēmu.

Lai novērtētu dažādu bioķīmisko modeļu saskaņotību, līdzības un atšķirības, darba ietvaros tiek piedāvāta pieeja, kuras pamatā tiek izmantota strukturālā analīze. Divu modeļu saskaņotības novērtēšanai tiek piedāvāts veidot šķēluma modeli, kas ietvertu paraugmodeļu pārklāšanās daļu, veikt šķēluma un paraugmodeļu topoloģisko analīzi un salīdzināt to topoloģiskos rādītājus.

Darbā tiek sniegts divu šķēluma modeļu novērtējums, salīdzinot šķēlumu ar paraugmodeļiem, ar mērķi noteikt līdzības un atšķirības, lietojot strukturālo analīzi, un noteikt topoloģiskos rādītājus, kas var tikt izmantoti modeļu kvalitātes un vienotības pakāpes, kā arī pielietojamības novērtēšanā. Veiktā pētījuma ietvaros (Rubina u.c., 2013) tika salīdzināti divi genoma izmēra modeļu pāri no *BioCyc* publiskās datubāzes: 1) baktērijas *Escherichia coli* modeļi “*ecol199310cyc*” un “*ecol316407cyc*”, 2) rauga *Saccharomyces cerevisiae* modeļi *iND750* un *iLL672*. Modeļu pāru šķēluma iegūšanai tika izmantots rīks *Moderator* (Mednis, Brusbardis, & Galvanauskas, 2012). Savukārt modeļu strukturālās analīzes izpildei tika izmantots promocijas darba ietvaros izstrādātais programmatūras rīka prototips *BINESA*. Katra organisma divu paraugmodeļu pāru šķēlumi strukturālā analīzē demonstrē gan ļoti līdzīgus, gan atšķirīgus parametrus.

Veiktais pētījums sniedz pavisam atšķirīgus *E.coli* un *S.cerevisiae* modeļu pāru šķēlumu topoloģiskos rādītājus. Modeļi, kas ir vienas autoru grupas veidoti, kā, piemēram, baktērijas *E.coli* modeļi, var tikt izmantoti kā paraugi ar augstu sakritības pakāpi starp modeļiem un var tikt interpretēti kā divu modeļu saskanošā jeb vienādā daļa. Līdzīgi topoloģiskie parametri nozīmē to, ka šķēluma modelis var funkcionēt kā patstāvīgs tīkla modelis pat bez būtiskas tālākas pilnveides. Tajā pašā laikā rauga *S.cerevisiae* modeļu šķēlums demonstrē ļoti atšķirīgas strukturālās īpašības un visdrīzāk šķēluma modelis nespētu funkcionēt pat pēc būtiskas pilnveides.

Daži topoloģiskie parametri, tādi kā metabolītu skaits, reakciju un saišu skaits, nosaka šķēluma modeļa izmēru salīdzinājumā ar paraugmodeļiem. To var interpretēt divējādi: kā augstu sakritību mazā modeļu pārklāšanās daļā vai zemu vienotības pakāpi lielākajā modeļu daļā. Šī iemesla dēļ minētie

topoloģiskie parametri vien nevar tikt izmantoti šķēluma modeļu kvalitātes novērtēšanā.

Metabolītu skaits, reakciju skaits, metabolītu puspakāpju aproksimācija, kā arī metabolītu klasterizācijas koeficients un tā sadalījums, vidējais klasterizācijas koeficients ir vāji šķēluma modeļa kvalitātes un atbilstības pakāpes noteicēji.

Veiktā pētījuma ietvaros par informatīviem strukturāliem rādītājiem modeļu atbilstības pakāpes noteikšanā ir atzīti kaimiņu skaita regresijas modelis, kaimiņu skaita sadalījums un metabolītu sasaistības (zema vai augsta saistības līmeņa metabolīti) procentuāls sadalījums. Par modeļu vāju līdzību liecina izolētas tīkla daļas. Par šķēluma modeļa zemas kvalitātes indikatoriem ir uzskatāmi vidējās pakāpes un puspakāpju, vidējā kaimiņu skaita zemās vērtības salīdzinājumā ar paraugmodeļiem.

Dažādu modeļu šķēlumu un apvienojumu strukturālā analīze ir svarīga gan salīdzinot dažādu modeļu evolūciju, gan veidojot jaunus uz pieejamo modeļu bāzes. Šķēluma analīze, salīdzinot oriģinālo modeli un modeli pēc evolūcijas, var parādīt, kuras no sākotnējām reakcijām ir izturējušas evolūciju un kuras tika aizstātas ar citām.

Šķēlumu analīze ļauj arī novērtēt dabiskās evolūcijas procesu noskaidrojot, kuras reakciju grupas un procesi palikuši kopēji dažādiem organismiem neskatoties uz ilgstošo evolūcijas procesu. Automātiskās salīdzināšanas lietojums un divu modeļu šķēluma ģenerēšana var sniegt ieskatu par modeļu iekšējo līdzību un noteikt dažādu modeļu un autoru viedokļu saskaņotības līmeni konkrētu organismu vielmaiņas modelēšanā. Šī pieeja var tikt izmantota, lai ātri noteiktu dažādu organismu modeļu līdzību.

6. BIOĶĪMISKO TĪKLU STRUKTŪRAS EVOLŪCIJAS DATORSIMULĀCIJAS

Bioķīmisko tīklu struktūras evolūcijas datoreksperimentu (turpmāk, eksperimentu) mērķis ir novērtēt atbilstošu mutāciju tipu ietekmi uz genomu un tam piesaistītu bioķīmiskā tīkla struktūru un noteikt bioķīmiskā tīkla struktūras evolūcijas dinamikas un struktūras izmaiņu īpatnības, kas rodas dažāda tipa mutāciju ietekmē vienāda un jaukta svarīguma pakāpes reakciju gadījumā. Datoreksperimenti tika izpildīti, izmantojot darba ietvaros izstrādāto modelēšanas rīka prototipu *BINESA*, kurā ir iestrādāts darbā piedāvātais genomam piesaistītas bioķīmisko tīklu struktūras evolūcijas algoritms.

Eksperimentu izpildei tika izvēlēti divi bioķīmiskā tīkla modeļi: maza izmēra testa modelis un baktērijas *Zymomonas mobilis* neliela izmēra bioķīmiskā tīkla modelis (turpmāk, *ZMO*). Testa modeļa struktūra tika piesaistīta mākslīgajam genomam ar 100 nukleotīdu gariem gēniem. Mazā izmēra bioķīmiska tīkla testa modelis sastāv no 9 virsotnēm un 14 reakcijām, kuras veido 19 saites. *ZMO* bioķīmiskā tīkla modelis (Pentjuss et al., 2013) iekļauj 81 metabolītu un 96 reakcijas, kuras veido 287 saites. Šī modeļa

struktūra tika piesaistīta mākslīgajam genomam ar 1300 nukleotīdu gariem gēniem, kas atbilst baktērijas *Z.mobilis* modelī iesaistīto gēnu vidējam garumam.

Lai novērtētu dažāda tipa mutāciju ietekmi vienāda un jaukta svarīguma pakāpes reakciju tīklos, evolūcijas datoreksperimenti tika realizēti, izmantojot šādus modeļus:

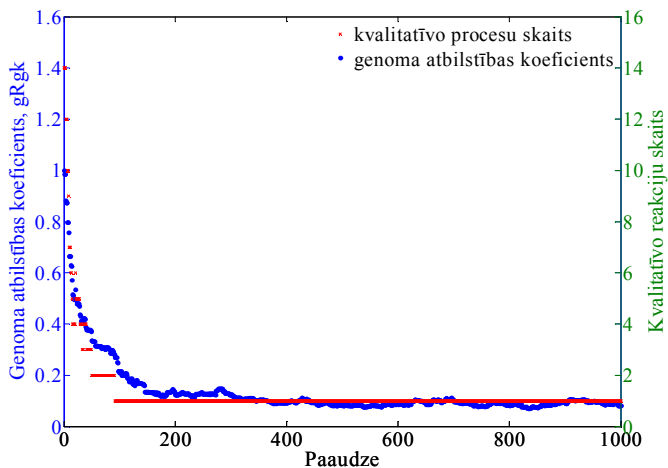
- 1) testa modelis, kurā visas reakcijas ir vitāli svarīgas (turpmāk, vitālo reakciju testa modelis jeb vitālais testa modelis);
- 2) testa modelis, kurā visas reakcijas ir kvalitatīvas (turpmāk, kvalitatīvo reakciju testa modelis jeb kvalitatīvais testa modelis);
- 3) testa modelis, kurš attēlo vitāli svarīgas, kvalitatīvas un mazsvarīgas reakcijas (turpmāk, jauktās svarīguma pakāpes reakciju testa modelis jeb jauktais testa modelis);
- 4) *ZMO* modelis, kurā visas reakcijas ir vitāli svarīgas (turpmāk, vitālo reakciju *ZMO* modelis jeb vitālais *ZMO* modelis);
- 5) *ZMO* modelis, kurā visas reakcijas ir kvalitatīvas (turpmāk, kvalitatīvo reakciju *ZMO* modelis jeb kvalitatīvais *ZMO* modelis);
- 6) *ZMO* modelis, kurš attēlo vitāli svarīgas, kvalitatīvas un mazsvarīgas reakcijas (turpmāk, jauktās svarīguma pakāpes reakciju *ZMO* modelis jeb jauktais *ZMO* modelis).

Nevienmērīga svarīguma tīklu gadījumā novērotas būtiskas atšķirības starp kvalitatīviem un jaukta svarīguma tīkliem, kā arī starp vitāli svarīgiem un kvalitatīviem tīkliem. To evolūcija atšķiras ar ievērojamām izdzīvošanas ilguma atšķirībām. Tas skaidrojams ar augstāku pieļaujamo atbilstības koeficienta novirzi un kvalitatīvo tīklu spēju pastāvēt pat ar vienu reakciju (sk. 4.att). Samazinoties kvalitatīvo reakciju skaitam, samazinās arī varbūtība, ka mutācija skars atlikušas reakcijas nosakošo gēnu visās 10 alternatīvajās genoma kopijās, kas ir nākamās paaudzes pēcnācēja kandidātu pretendenti. Līdz ar to kvalitatīva tīkla evolūcija var pārsniegt pat 100000 paaudžu ar vienu reakciju. Savukārt, vitālo reakciju klātbūtnē tīklā ir kritiska un to zaudēšana nav pieļaujama, kas atspoguļojas gan vitāla, gan jaukta tīkla struktūras evolūcijā un ievērojami ietekmē tās izdzīvošanas ilgumu.

Jaukta svarīguma tīklos novērots straujš kvalitatīvo un mazsvarīgo reakciju samazinājums jau pašas evolūcijas sākumā translokācijas, delēcijas, nukleotīdu inversijas (sk. 5.att.), un inversijas mutāciju ietekmē. Taču tīkls turpina evolucionēt ar vienu kvalitatīvo un visām vitālām reakcijām. Savukārt, nukleotīdu inversijas, inversijas (sk. 6.att.) un translokācijas ietekmē kvalitatīvie un mazsvarīgie procesi mēdz atjaunoties, notiekot apgrieztai mutācijai.

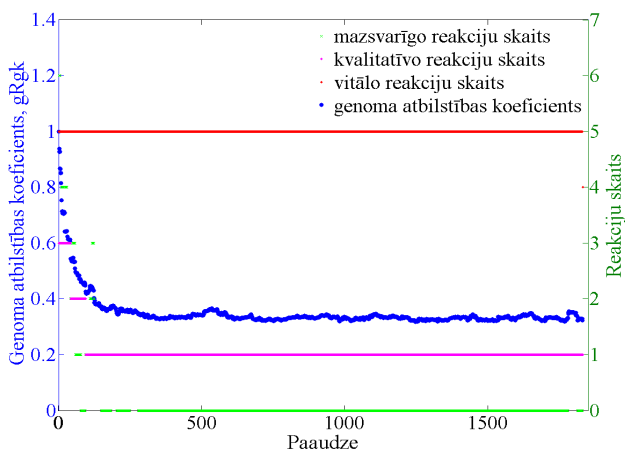
Analizējot atsevišķi punktveida mutācijas ietekmi uz tīkla struktūru, arī ir novērotas apgrieztās mutācijas, bet ar mazāku frekvenci. Neraugoties uz to, ka evolūcijas process turpinās ilgāk salīdzinājumā ar pārējiem mutāciju tipiem, punktveida mutācijas evolūcijas laikā uzkrājas un ar katru nākamo paaudzi

samazina genoma atbilstības koeficientu un padara katra nākamā pēcnācēja struktūru arvien traušlāku.

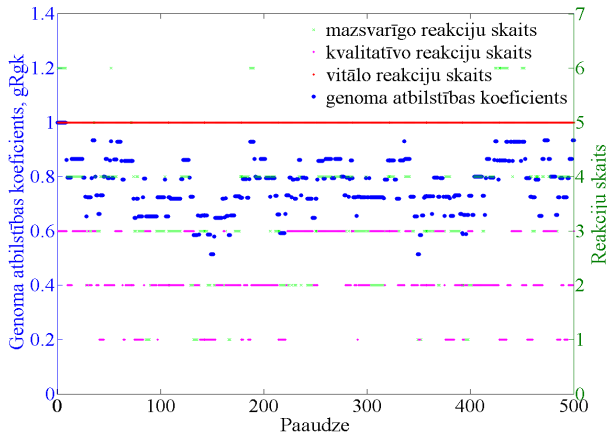


4.att. Kvalitatīva testa modeļa genoma atbilstības koeficienta un kvalitatīvo reakciju skaita izmaiņas nukleotīdu inversijas ietekmē ar varbūtību 10%

Attēlā ir parādīts genoma atbilstības koeficients, kas ir vienāds ar visu tā gēnu atbilstības koeficientu vidējo vērtību.



5.att. Jaukta testa modeļa genoma atbilstības koeficienta un dažāda svarīguma reakciju skaita izmaiņas nukleotīdu inversijas ietekmē ar varbūtību 10%

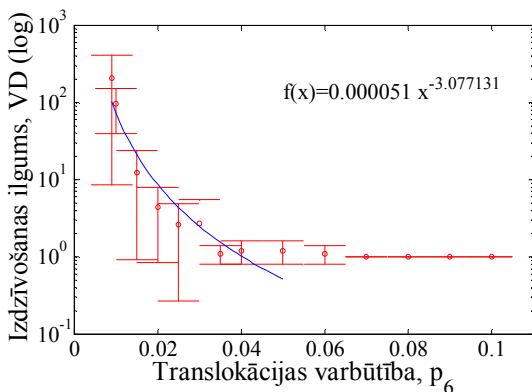


6.att. Jaukta testa tīkla genoma atbilstības koeficienta un dažāda svarīguma reakciju skaita izmaiņas inversijas ietekmē pie varbūtības 5%

Datorsimulācijās evolūcijas parametriem tika piešķirtas šādas vērtības: punktveida mutācijas varbūtība 10^{-7} , nukleotīdu inversijas varbūtība 10^{-7} , inversijas varbūtība 5%, bet pārējo mutāciju varbūtības sastāda 0%. Evolūcijas process tika ģenerēts līdz 100000 paaudzēm, kuras izpildes rezultātā tika saglabātas visas vitālas reakcijas un daļa kvalitatīvo un mazsvarīgo reakciju. Jāatzīmē, ka evolūcijas laikā kvalitatīvo un mazsvarīgo reakciju skaits un genoma atbilstības koeficients nepārtraukti mainās, kas ir rezultāts tiešajām, kad inversijas mutācija pilnībā maina gēna sekvenci, un apgriezām inversijas mutācijām, kad gēns atgriežas pie tā sākotnējās formas.

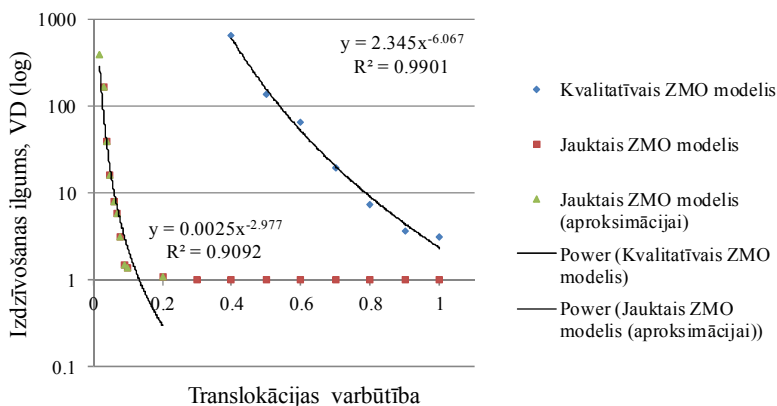
Nevienādīga svarīguma tīklos (gan vitālo, gan kvalitatīvo, gan jaukta svarīguma tīklu) izdzīvošanas ilgums mainās pēc pakāpes likuma ar negatīvu pakāpes rādītāju un ļoti augstu determinācijas koeficientu visu izpētīto mutāciju ietekmē. Konkrētā svarīguma tīkla veidam mainās tikai izdzīvošanas ilguma vērtības, kas ir atkarīga no konkrēta mutācijas tipa. Jo graujošāka ir mutācijas ietekmē, jo ātrāk evolūcijas process tiek pārtraukts un ir īsāks izdzīvošanas ilgums. Savukārt, jaukta svarīguma tīkls izrāda lielāku noturību pret mutācijām un izdzīvo ilgāk, salīdzinājumā ar vitāli svarīgo tīklu (sk. 7.att. un 8.att.) atsevišķā mutāciju tipa ietekmē. Savukārt, salīdzinot kvalitatīvo un jaukta svarīguma tīklus, jaukta svarīguma tīkla izdzīvošanas ilgums ir daudz mazāks nekā kvalitatīvajam tīklam.

Jaukta svarīguma tīkla struktūras evolūciju vairāku mutāciju ietekmes apstākļos visvairāk iespaido translokācijas, delēcijas, inversijas un nukleotīdu inversijas. Translokācijai un delēcijai ir būtiskākā ietekme uz bioķīmiskā tīkla struktūras izdzīvošanas ilgumu un struktūras topoloģiskajiem parametriem. Savukārt, palielinot duplikācijas varbūtību, tīkla izdzīvošanas ilgums palielinās un reakciju skaits tīklā pieaug eksponenciāli, analizējot šī mutāciju tipa ietekmi gan atsevišķi (sk. 9.att.), gan iedarbojoties vienlaicīgi vairākām mutācijām uz tīklu.



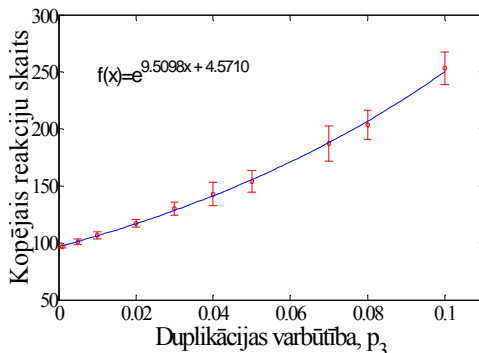
7.att. Vitāli svarīgo reakciju ZMO tīkla izdzīvošanas ilgums

7.attēla ir parādīts katras eksperimentu kopas vidējais izdzīvošanas ilgums un tā standartnovirze, katrā kopā izpildot pa 10 eksperimentiem. Kopā tika izpildītas 11 eksperimentu kopas ar dažādām translokācijas varbūtībām.



8.att. Izdzīvošanas ilguma salīdzinājums translokācijas ietekmē ZMO modelim

8.attēla ir parādīts katras eksperimentu kopas vidējais izdzīvošanas ilgums un tā standartnovirze pie dažādām translokācijas varbūtībām kvalitatīvajam un jauktam ZMO tīkla modelim. Katrā eksperimentu kopā tika izpildīts pa 10 eksperimentiem.



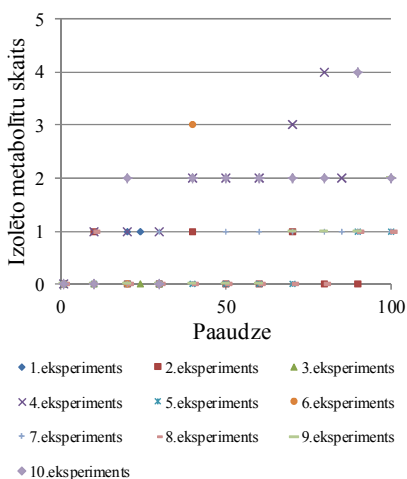
9.att. Kopējais reakciju skaits jaukta svarīguma reakciju ZMO tīkla modeli

9.attēla ir parādīts katras eksperimentu kopas vidējais izdzīvošanas ilgums un tā standartnovirze, katrā kopā izpildot pa 10 eksperimentiem. Kopā tika izpildītas 7 eksperimentu kopas ar dažādām translokācijas varbūtībām.

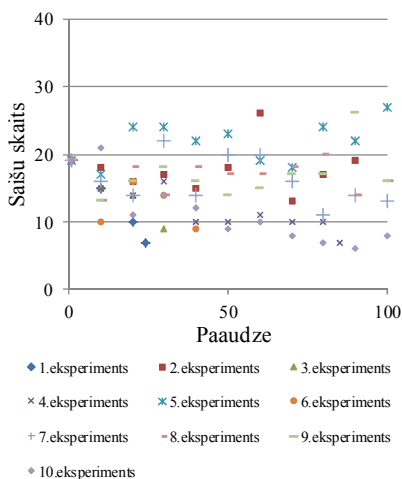
Vairāku mutāciju vienlaicīgas iedarbes uz jaukta tīkla struktūras evolūciju analīze. Piemēram, pie translokācijas 10% varbūtības, kad pārējo mutāciju varbūtība ir 1%, struktūras vidējais izdzīvošanas ilgums ir 18 paaudzes ar standartnovirzi 11 paaudzes (minimāli izdzīvo 4 paaudzes, maksimāli evolūcija turpinās līdz 47 paaudzei), evolūcijā zaudējot vismaz vienu vitālo reakciju (100%) un visas kvalitatīvas reakcijas (10%), un saglabājot reakciju dublikātus. Evolūcijas pārtraukšanas brīdī vidējais genomu atbilstības koeficients sastāda 0.604 ar standartnovirzi 0.149. Tajā pašā laikā, delēcijas ar 10% varbūtību ietekmē ir konstatēts, ka struktūra izdzīvo vidēji 38 paaudzes ar standartnovirzi 20 paaudzes (minimāli izdzīvo 6 paaudzes, maksimāli struktūras evolūcija turpinās līdz 77 paaudzei), zaudējot kādu vitālo reakciju (80%) vai visas kvalitatīvas reakcijas (20%), ar vidējo genomu atbilstības koeficientu 0.92 un tā standartnovirzi 0.077. Bet inversijas varbūtību palielinot līdz 10%, ir konstatēts, ka struktūra izdzīvo vairāk ka 100 paaudzes 60% gadījumos, rezultātā saglabājot vienu kvalitatīvo un visas vitālas reakcijas. 4 eksperimentos evolūcija tika pārtraukta nesasniedzot 100 paaudzes (minimālais izdzīvošanas ilgums 24 paaudzes, maksimālais – 85 paaudzes), zaudējot visas kvalitatīvas reakcijas, ar vidējo genomu atbilstības koeficientu 0.532 un izkliedi 0.069. Savukārt, nukleotīdu inversija ar 10% varbūtību ietekmē struktūra izdzīvo ilgāk par 100 paaudzēm 80% gadījumos, zaudējot vitālo reakciju vai visas kvalitatīvas reakcijas, un saglabājot reakciju dublikātus. Evolūcijai beidzoties ātrāk, 31 un 71 paaudzē, vidējais genomu atbilstības koeficients 10 eksperimentu kopai sastāda 0.349 ar standartnovirzi 0.06.

Nukleotīdu inversijas un inversijas ietekmē tīkla struktūras izdzīvošanas ilgumam ir ļoti augsta izkliede un inversijai mēdz būt apgrieztais efekts, kā rezultāta mutējošais gēns atgriežas pie savas sakontējās formas un reakcija,

kuru tās nosaka, atjaunojas. Līdz ar to inversijas ietekmē struktūras topoloģiskie rādītāji (sk. 10.att., 11.att., 12.att.), kā arī genoma atbilstības koeficients mēdz mainīt savas vērtības lēcienveida, t.i. tā vērtības mēdz palielināties un samazināties evolūcijas gaitā. Ja nukleotīdu inversijas un inversijas, kā arī translokācijas mutācija mēdz būt apgrieztas, tad delēcijas sekas ir neatgriezeniskās. Palielinot delēcijas varbūtību no 1% uz 10% struktūras izdzīvošanas ilgums strauji samazinās, bet genoma atbilstības koeficients evolūcijas pārtraukšanas brīdī saglabā augstās vērtības.



10.att. Izolēto metabolītu skaita izmaiņas inversijas ietekmē



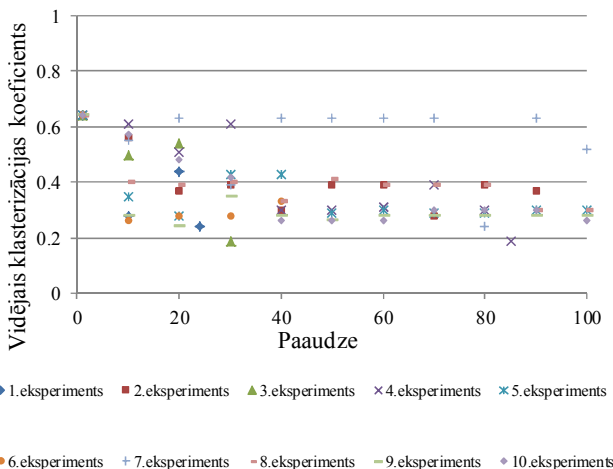
11.att. Saišu skaita izmaiņas inversijas ietekmē

Vienlaicīgi iedarbojoties vairākiem mutāciju tiptiem uz jaukta svarīguma reakciju tīkla struktūru un palielinot vienas mutācijas varbūtību, lielāko ietekmi uz izdzīvošanas ilgumu atstāj translokācija, delēcija un inversija. Savukārt genoma atbilstības koeficientu visvairāk ietekmē nukleotīdu inversijas, inversijas un punktveida mutācija.

Izolēto virsotņu skaitu visvairāk ietekmē delēcija, translokācija un nukleotīdu inversija. Savukārt punktveida mutācijas, nukleotīdu inversijas un inversijas ietekmē, izolēto virsotņu skaits evolūcijas gaitā mēdz palielināties, bet saglabā tendenci samazināties.

Vidējo tīkla pakāpi visvairāk ietekmē delēcija, punktveida mutācija un nukleotīdu inversija. Savukārt punktveida mutācijas, nukleotīdu inversijas un inversijas ietekmē, vidējās pakāpes vērtība evolūcijas gaitā mēdz svārstīties vairāk, bet saglabā tendenci samazināties.

Vidējo kaimiņu skaitu visvairāk ietekmē delēcija, translokācija un nukleotīdu inversija. Savukārt punktveida mutācijas, nukleotīdu inversijas un inversijas ietekmē, vidējais kaimiņu skaits evolūcijas gaitā mēdz palielināties un samazināties evolūcijas gaitā divreiz it īpaši inversijas ietekmē, bet saglabā tendenci samazināties.



12.att. **Vidējā klasterizācijas koeficienta izmaiņas inversijas ietekmē**
 10., 11. un 12. Attēlos ir parādīts 10 eksperimentu kopas rezultāti, t.i., izolēto virsotņu skaita, saišu skaita un vidējā klasterizācijas koeficienta izmaiņas evolūcijas gaitā, kad inversijas varbūtība ir 10%, bet pārējo mutāciju varbūtības ir 1%.

Saišu skaitu visvairāk ietekmē delēcija, translokācija un nukleotīdu inversija. Savukārt punktveida mutācijas, nukleotīdu inversijas un inversijas ietekmē, izolēto saišu skaits evolūcijas gaitā mēdz palielināties un samazināties divreiz it īpaši punktveida mutācijas ietekmē, bet tas saglabā tendenci samazināties.

Savukārt vidējo klasterizācijas koeficientu visvairāk ietekmē delēcija, translokācija un inversija. Taču punktveida mutācijas, nukleotīdu inversijas un inversijas ietekmē, vidējais klasterizācijas evolūcijas gaitā mēdz palielināties un samazināties divreiz vai pat vairākkārt it īpaši inversijas ietekmē, bet saglabā tendenci samazināties.

SECINĀJUMI

Galvenie darba rezultāti

Ir izstrādāta datormodelēšanas pieeja un ar tās palīdzību ir noteikta dažāda tipa mutāciju ietekme uz neviendabīga svarīguma bioķīmisko tīklu struktūras evolūciju.

1) Ir izpētītas bioķīmisko tīklu strukturālās analīzes metodes un tīklu attēlošanas paņēmieni.

Bioķīmisko tīklu attēlošanai, reprezentācijai un modelēšanai praksē tiek izmantoti bioķīmisko ceļu kartes, strukturālie, stohastiskie un dinamiskie datormodeļi. Bioķīmisko tīklu attēlojumi balstās uz kādu no sistēmbioloģijas

nozārē pieņemtajiem standartiem, piemēram, *SBML*, *SBGN* vai formālismiem, piemēram, *Būla* vai *Petri* tīklu formālismu.

Bioķīmisko tīklu struktūras pētniecībā tiek izmantotas galvenokārt divas metodes: 1) strukturālā jeb topoloģiskā analīze, kas ir pamatota uz grafu teorijas jēdzieniem un metodēm, un 2) klasteru analīze.

2) Ir izpētīta bioķīmisko tīklu evolūcijas procesa norise un tās dinamiku ietekmējošie faktori.

Bioķīmisko tīklu evolūcija pamatā balstīta uz bioloģiskās sistēmas molekulāras evolūcijas, kas noris dzīva organisma genoma līmenī. Izmaiņas gēnos ietekmē mijiedarbību starp gēniem, gēnu produktiem, un to ietekmētajiem procesiem.

3) Ir izpētītas bioķīmisko tīklu evolūcijas modelēšanas pieejas.

Dominē divas bioķīmisko tīklu evolūcijas modelēšanas pieeju klases: tīkla attīstības duplikāciju un delēciju pieejas.

Taču praksē lietojamajām bioķīmisko tīklu un to evolūcijas izpētes pieejām ir vairāki trūkumi. Pirmais trūkums ir būtisku procesu īpašību ignorēšana, imitējot tīkla struktūras evolūciju. Viena no šīm īpašībām ir procesu svarīgums. Otrais trūkums ir tīkla struktūras izmaiņu pamatojums ar mutāciju parādīšanos genoma līmenī bez genoma piesaistīšanas. Trešais trūkums ir attālināts reālā evolūcijas procesa attēlojums un galvenokārt tikai divu mutāciju tipu iekļaušana modeļos – gēnu duplikācija un/vai delēcija, kas izpaužas virsotnes ievietošanā tīklā, un punktveida mutācija, kas teorētiskā līmenī izskaidro saišu aizstāšanas notikumus.

4) Ir izpētīti bioķīmisko tīklu modelēšanas programmatūras rīki.

Izpētē ir konstatēts, ka pēdējos gados sistēmbioloģijas nozarē tika izstrādāti daudzi gan šauras, gan plašas specializācijas programmatūras rīki. Daļa no bioķīmisko tīklu modelēšanas rīkiem tiek lietoti atsevišķu tīkla tipu modelēšanā, piemēram, vielmaiņas tīklu vai gēnu regulācijas tīklu modelēšanā, izmantojot kādu formālismu vai standartu. Netika atrasti programmatūras rīki, kas modelētu bioķīmisko tīklu struktūras evolūciju, kura būtu piesaistīta genomam, un evolūcijas imitācijā ņemtu vērā bioķīmisko procesu svarīgumu.

5) Ir izstrādāts genomam piesaistītas bioķīmisko tīklu struktūras evolūciju imitējošs algoritms.

Bioķīmisko tīklu evolūcijas algoritms ļauj modelēt bioķīmisko tīklu struktūras izmaiņas, kas ir ģenētiskā materiāla izmaiņu sekas, kuras notiek genoma līmenī dažāda tipa mutāciju ietekmē. Piedāvātais genomam piesaistītas bioķīmisko tīklu struktūras evolūcijas algoritms ņem vērā punktveida mutācijas, misensmutācijas, nonsensmutācijas, nukleofīdu inversiju, duplikācijas, delēcijas, inversijas un translokācijas. Izstrādātais bioķīmisko tīklu struktūras evolūcijas algoritms piesaista bioķīmiskajam tīklam gēnus un ņem vērā reakciju atšķirīgo svarīgumu.

6) Ir izstrādāts programmatūras rīka prototips *BINESA*, kas realizē izstrādāto genomam piesaistītas bioķīmisko tīklu struktūras evolūciju imitējošo algoritmu.

Rīks *BINESA* sniedz arī iespēju analizēt bioķīmiskā tīkla struktūru un aprēķināt vai noteikt dažādus topoloģiskos rādītājus, kā arī iegūt dažāda veida evolūcijas procesa datus un rezultātus: izdzīvošanas ilgumu jeb paaudžu skaitu, genoma atbilstības koeficientu, reakciju vidējo intensitāti, vitālo reakciju skaitu, kvalitatīvo reakciju skaitu, mazsvarīgo reakciju skaitu, katras evolūcijā saglabātas un analizētas struktūras topoloģiskos rādītājus, t.i., kopējo virsotņu, izolēto virsotņu skaitu, savienoto virsotņu skaitu, kopējo reakciju skaitu, saišu skaitu, tīkla vidējo pakāpi, tīkla vidējās puspakāpes, vidējo kaimiņu skaitu, vidējo klasterizācijas koeficientu, ciklu skaitu, cilpu skaitu, pakāpes un puspakāpju sadalījumus, kaimiņu skaita sadalījumu, klasterizācijas koeficienta sadalījumu.

7) Ir veiktas genomam piesaistītas neviendabīga svarīguma bioķīmisko tīklu struktūras evolūcijas datoreksperimenti dažāda tipa mutāciju ietekmē un iegūto rezultātu analīze.

Darba ietvaros tika veiktas bioķīmiskā tīkla struktūras evolūcijas datorsimulācijas vitāli svarīgo reakciju, kvalitatīvo un jaukta svarīguma pakāpes reakciju bioķīmiskajiem tīkliem. Lai noteiktu katra atsevišķa mutācijas tipa ietekmi uz tīkla evolūcijas dinamiku un struktūras topoloģisko rādītāju izmaiņas īpatnībām tika veiktas vairākas datoreksperimentu kopas ar dažādām punktteida mutācijas, nukleotīdu inversijas, duplikācijas, delēcijas, inversijas un translokācijas mutāciju varbūtībām.

8) Ir noteikti bioķīmiskā tīkla modeļa pielietojamību raksturojošie topoloģiskie parametri.

Veiktā pētījuma ietvaros par informatīviem strukturāliem rādītājiem ir atzīti kaimiņu skaita aproksimācija un metabolītu sasaistības (zemu vai augsti saistīti metabolīti) procentuāls sadalījums. Par zemu modeļu atbilstības, kas izpaužas fragmentos, un šķeluma modeļa zemas kvalitātes indikatoriem ir uzskatāmi vidējās pakāpes un puspakāpju, vidējā kaimiņu skaita zemās vērtības salīdzinājumā ar paraugmodeļiem.

Dažādu modeļu šķelumu un apvienojumu strukturālā analīze ir svarīga gan salīdzinot dažādu modeļu evolūciju, gan veidojot jaunus modeļus par pamatu izmantojot vairākus jau pieejamus modeļus. Šķeluma analīze, salīdzinot oriģinālo modeli un modeli pēc evolūcijas, var analizēt ne tikai pašu modeļu, bet arī to šķelumu struktūru, lai novērtētu, kuras no sākotnējām reakcijām ir izturējušas evolūciju un kuras ir aizstātas ar citām.

Secinājumi un attīstības perspektīvas

Datormodelēšanas un datorsimulāciju izpildes rezultātā radušies vairāki secinājumi par bioķīmisko tīklu evolūcijas īpatnībām.

1. Salīdzinot jaukta svarīguma un kvalitatīvu reakciju bioķīmiskos tīklus, ir novērojama jaukta svarīguma tīklu ātrāka bojāeja. Sasniegto paaudžu skaits atšķiras 10-10000 kārtīgi gan neliela izmēra (14 reakcijas un 9 metabolīti) mākslīgi veidotam testa modelim, gan arī lielāka izmēra (96 reakcijas un 81

metabolīts) *Zymomonas mobilis* centrālā metabolisma modelim. Līdz ar to atšķirīga reakciju svarīguma iekļaušana modeļos būtiski maina evolūcijas datorsimulāciju rezultātus.

2. Datormodeļa simulāciju rezultātā ir noteikts, ka tīklu struktūras evolūcijas ilgumu līdz bojāejai visvairāk ietekmē translokācijas, delēcijas un inversijas mutāciju operatori.

3. Tīkla struktūras evolūcijā vairāku mutāciju operatoru iedarbē tipiska tendence ir izolēto metabolītu skaita palielināšanās, vidējās metabolītu pakāpes un saišu skaita, kā arī vidējā kaimiņu skaita un klasterizācijas koeficienta samazināšanās. Parametru izmaiņas lielākoties notiek lineāri, saglabājot vienādas mutāciju varbūtības vai palielinot punktveida mutācijas, delēcijas vai translokācijas varbūtību. Turklāt, palielinot nukleotīdu inversijas vai inversijas varbūtību, topoloģisko parametru vērtības samazinās, bet to izmaiņas notiek lēcienvēidā. Krasi atšķiras eksperimenti ar paaugstinātām delēciju un translokāciju varbūtībām, kas noved pie pāātrinātas bojāejas.

4. Genomam piesaistīta bioķīmiskā tīkla dzīvotspēju atkarībā no mutāciju varbūtības atsevišķo mutācijas tipu ietekmē gan vienāda, gan jaukta svarīguma tīklos vislabāk raksturo pakāpes likums ar negatīvo pakāpes rādītāju.

5. Eksperimentu ilgums ir atkarīgs no modeļa izmēra un piesaistīto gēnu garuma, kā arī saglabājamo struktūru skaita evolūcijā. Turklāt, eksperimenta ilgums nedaudz pieaug līdz ar modelējamo paaudžu skaitu, rēķinot laiku, kas tiek patērēts vienas paaudzes simulācijai. Piemēram, translokācijas ietekmes uz kvalitatīvo *ZMO* tīklu ar varbūtību 50% novērtēšanai vienas paaudzes simulācija aizņem vidēji 10.75 sekundes (vienas simulācijas ar 10000 paaudzēm izpildei ir nepieciešamas 29 stundas 51 minūte 40 sekundes), kamēr kvalitatīvā testa tīkla vienas paaudzes simulācija ar tiem pašiem parametriem aizņem 0.31 sekundi (vienas simulācijas ar 100000 paaudzēm izpildei ir nepieciešamas 8 stundas 36 minūte 40 sekundes), izmantojot datoru ar *Intel Core2* 2GHz procesoru, 1Gb operatīvo atmiņu un *Microsoft Access* 2010.

Kā **attīstības perspektīvas** var iezīmēt vairākus darbības virzienus.

- 1) Ieviest genomam piesaistītas bioķīmiskā tīkla evolūcijas algoritmā mutācijas operatorus, kuri maina gēnu garumu un nosacījumu kopu, kas nosaka jaunu, līdz šim vēl nebijušu, saišu veidošanos tīklā.
- 2) Pētīt bioķīmisko tīklu evolūciju ne tikai struktūras līmenī, bet arī bioķīmiskā tīkla funkcionēšanas līmenī un novērtēt bioķīmiskā tīkla dinamikas izmaiņas evolūcijā, izmantojot evolūcijā iegūto tīkla struktūru.
- 3) Analizēt bioķīmisko tīklu struktūras motīvu izmaiņas, atgriezeniskās un apsteidzošās saites ciklu jeb vadības ciklu evolūciju dažādu mutācijas tipu ietekmē, kā arī alternatīvo vadības ciklu rašanās likumsakarības.

LITERATŪRAS SARAKSTS

- Albert R., Barabasi A.-L. (2002) Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, Vol. 74, p.47-97.
- Albert I., Thakar J., Li S., Zhang R., Albert R. (2008) Boolean network simulations for life scientists. *Source code for Biology and Medicine*, Vol.3, p.1-16.
- Albert R., Jeong H., Barabasi A.-L. (2000) Error and attack tolerance of complex networks. *Nature*, Vol. 406, p.378-382.
- Aldana M., Balleza E., Kauffman S., Resendiz O. (2007) Robustness and evolvability in genetic regulatory networks. *Journal of Theoretical Biology*, Vol. 245, p.433-448.
- Arbabi-Ghahroudi M., Tanha J., MacKenzie R. (2005) Prokaryotic expression of antibodies. *Cancer Metastasis Rev*, Vol. 24, Issue 4, p.501-519.
- Assenov Y., Ramirez F., Schelhorn S.-H., Lengauer T., Albrecht M. (2008) Computing topological parameters of biological networks. *Bioinformatics Applications Note*, Vol. 24, Nr. 2, p.282-284.
- Baitaluk M., Sedova M., Ray N., Gupta A. (2006) BiologicalNetworks: visualization and analysis tool for systems biology. *Nucleic Acids Research*, Vol. 34, D466- 471.
- Barabasi A.-L., Oltvai Z.N. (2004) Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, Vol. 5, p.101-113.
- Barabasi A.-L., Albert R. (1999) Emergence of scaling in random networks. *Science*, Vol. 286, p.509-512.
- Barrat A., Weigt M. (2000) On the properties of small-world network models. *The European Physical Journal B*, Vol. 13, p.547-560.
- Becker S.A., Feist A.M., Mo M.L., Hannum G., Palsson B.Ø., Herrgard M.J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. *Nature Protocols*, Vol.2, No.3, p.727-738.
- Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Wheeler D.L. (2008) GenBank. *Nucleic Acids Research*, Januar, p.25-30.
- Berg J., Lassing M., Wagner A. (2004) Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evolutionary Biology*, Vol. 4, No. 51, p.1-12.
- Bersini H., Lenaerts T., Santos F.C. (2005) Growing biological networks: Beyond the gene-duplication model. *Journal of Theoretical Biology*, Vol. 241, p.488-505.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U. (2006) Complex networks: Structure and dynamics. *Physics Reports*, Vol. 424, p.175-308.
- Brazhnik P., Fuente A., Mendes P. (2002) Gene networks: How to put the function in genomics. *Trends in Biotechnology*, Vol. 20, No. 11, p.467-472.

- Breitling R., Gilbert D., Heiner M., Orton R. (2008) A structured approach for the engineering of biochemical network models, illustrates for signalling pathways. *Briefings in Bioinformatics*, Vol. 9, No. 5, p.404-421.
- Browna C.T., Rustb A.G., Clarke P.J.C., Panb Z., Schilstrab M.J., De Buyschera T., Griffingb G., Wolda B.J., Cameron R.A., Davidson E.H., Bolourib H. (2002) New Computational Approaches for Analysis of cis-Regulatory Networks. *Developmental Biology*, Vol. 246, Issue 1, p.86-102.
- Callaway D.S., Hopcroft J.E., Kleinberg J.M., Newman E.J., Strogatz S.H. (2001) Are randomly grown graphs really random? *Physical Review E*, Vol. 64 No.4, p.1-7.
- Cara A., Garg A., Micheli G., Xenarios I., Mendoza L., (2007) Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinformatics*, Vol. 8, p.462.
- Chaouiya C. (2007) Petri net modelling of biological networks. *Briefings in Bioinformatics*, Vol. 8, No. 4, p.210-219.
- Chaves M., Albert R., Sontag E.D. (2005) Robustness and fragility of Boolean models for genetic regulatory networks. *Journal of Theoretical Biology*, Vol. 235, p.431-449.
- Christogianni A., Douka E., Koukkou A.I., Hatziloukas E., Drainas C. (2005) Transcriptional Analysis of a Gene Cluster Involved in Glucose Tolerance in *Zymomonas mobilis*: Evidence for an Osmoregulated Promoter. *Journal of Bacteriology*, Vol. 187, No. 15, p. 5179-5188.
- CellDesigner Tutorial. (2008) [Tiešsaite] [skafīts 2009.g. 24.okt.] Pieejams: <http://www.celldesigner.org/tutorial/CellDesignerTutorialICSB2008.pdf>
- Chen L., Wang R.-S., Zhang X.-S. (2009a) Introduction. Transcription regulation: networks and models. **In:** Chen L., Wang R.-S., Zhang X.-S. *Biomolecular networks: Methods and Applications in System Biology*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.1-45.
- Chen L., Wang R.-S., Zhang X.-S. (2009b) Signaling Networks: Modeling and Inference. **In:** Chen L., Wang R.-S., Zhang X.-S. *Biomolecular networks: Methods and Applications in System Biology*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.313-321.
- Chen L., Wang R.-S., Zhang X.-S. (2009c) Topological structure of Biomolecular Networks. **In:** Chen L., Wang R.-S., Zhang X.-S. *Biomolecular networks: Methods and Applications in System Biology*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.169-204.
- Chen L., Wang R., Zhou T., Aihara K. (2005) Noise-induced cooperative behavior in a multi-cell system. *Bioinformatics*, Vol. 21, p.2722-2729.
- Clark D.P., Pazdernik N.J. (2012) *Molecular Biology: Understanding the Genetic Revolution*: 2-nd eds. USA, Waltham, Elsevier, p.274-307.
- Cline M.S., Smoot M., Cerami E., Kuchinsky A., Landys N., Workman C., Christmas R., u.c. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols*, Vol.2, p.2366 – 2382.

- Cohen R., Havlin S. (2003) Scale-free networks are ultra small. *Physical Review Letters*, Vol. 90, No. 5, p.1-4.
- Croft D., O'Kelly G., Wu G., Haw R., Gillespie M., Matthews L., Caudy M. u.c. (2011), Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research*, Vol. 39, D691-D697.
- Conklin Lab. *GenMapp Concept*. [Tiešsaite] [skatīts 2009.g. 25.okt.] Pieejams: <http://www.genmapp.org/concept.html>
- Cutter A.D. (2010) *Molecular evolution inferences from the C.elegans genome*. WormBook, May 5, p.1-14. [Tiešsaite]. [skatīts 2012.g. 4.dec.]. Pieejams: http://wormbook.org/chapters/www_molecularevol/molecularevol.html
- Dahlquist K.D. (2004) Using GenMAPP and MAPPFinder to view microarray data on biological pathways and identify global trends in the data. *Current Protocols in Bioinformatics*. May, Chapter 7, Unit 7, p.1-5.
- Dahlquist K.D., Salomonis N., Vranizan K., Lawlor S.C., Conklin B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genetics*, Vol. 31, p.19-20.
- Dambītis J. (2002) *Modernā grafu teorija: mācību līdzeklis*. Rīga: Datorzinību centrs, 17.-30.lpp.
- DDBJ — DNA Data Bank of Japan*. [Tiešsaite]. [skatīts 2012.g. 3.jūn.]. Pieejams: <http://www.ddbj.nig.ac.jp/>
- Demin O.V., Plysnina T.Y., Lebedeva G.V., Zobova E.A., Metelkin E.A., Kolupaev A.G., Goryanin I.I., Tobin F. (2005) Kinetic modelling of the E.Coli. In: Alberghina L., Westerhoff H. (Eds.) *Systems Biology. Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.31-67.
- Doniger S.W., Salomonis N., Dahlquist K.D., Vranizan K., Lawlor S.C., Conklin B.R. (2003) MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology*, Vol.4, Issue 1, Art.R7, p.1-12.
- Drake J.W., Charlesworth B., Charlesworth D., Crow J.F. (1998) Rates of Spontaneous Mutation. *Genetics*, Vol. 148, Nr. 4, p.1667-1686.
- Duarte N.C., Becker S.A., Jamshidi N., Thiele I., Mo M.L., Vo T.D., Srivas R., Palsson B.O. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences on the USA*, Vol.104, No.6, p.1777-1782.
- Duarte N.C., Herrgard M.J., Palsson B.O. (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, Vol. 14, p.1298-1309.
- ENA — The European Nucleotide Archive*. [Tiešsaite]. [skatīts 2012.g. 3.jūn.]. Pieejams: <http://www.ebi.ac.uk/ena/>
- EMBL — European Molecular Biology Laboratory*. [Tiešsaite]. [skatīts 2012.g. 3.jūn.]. Pieejams: <http://www.embl.de/>

- Farid N., Christensen K. (2006) Evolving networks through deletion and duplication. *New Journal of Physics*, Vol. 8, No. 9, p.1-17.
- Feist A.M., Henry C.S., Reed J.L., Krummenacker M., Joyce A.R., Karp P.D., Broadbelt L.J., Hatzimanikatis V., Palsson B. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, Vol. 3, No. 121, p.1-18.
- Fell D.A. (2005) Metabolic control Analysis. In: Alberghina L., Westerhoff H. (Eds.) *Systems Biology. Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.69-80.
- Fell D.A., Wagner A. (2000) The small world of metabolism. *Nature Biotechnology*, Vol. 18, p.1121-1122.
- Felsenstein J. (2013) Theoretical Evolutionary Genetics. Genome 562. USA: Washington, University of Washington, Department of Genome Sciences and Department of Biology, p.45-152.
- Fields S. (2001) Proteomics in genomeland. *Science*, Vol. 291, No. 5507, p.1221-1224.
- Finkel T., Gutking J.S. (2003) Signal transduction and Human Disease. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.23-29.
- Förster J., Famili I., Fu P., Palsson B. Ø., Nielsen J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome research*, Vol. 13, Issue 2, p.244–253.
- Freeman W.H. u.c.. (2000) *Mechanism of DNA replication*. NCBI, Bookshelf ID NBK21862. [Tiešsaite]. Pieejams: <http://www.ncbi.nlm.nih.gov/books/NBK21862>
- Funahashi A., Matsuoka Y., Jouraku A., Morohashi M., Kikuchi N., Kitano H. (2008) CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proceedings of the IEEE*, Vol. 96, Issue 8, p.1254-1265.
- Funahashi, A., Tanimura, N., Morohashi, M., and Kitano, H. (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, Vol. 1, p.159-162.
- Fuhrer T., Fischer E., Sauer U. (2005) Experimental identification and quantification of glucose metabolism in seven bacterial species. *Journal of Bacteriology*, March, p.1581-1590.
- GenBank. [Tiešsaite]. [Skatīts 2012.g. 1.jūl.]. Pieejams: <http://www.ncbi.nlm.nih.gov/genbank/>
- Gibson T.A., Goldberg D.S. (2011) Improving evolutionary models of protein interaction networks. *Bioinformatics*, Vol. 27, No. 3, p.376-382.
- GML – *Graph modelling language* (2009) University of Passau. [Tiešsaite] [skatīts 2009.g. 20.dec.]. Pieejams: <http://www.infosun.fim.uni-passau.de/Graphlet/GML/>
- Gonzalez A., Naldi A., Sánchez L., Thieffry D., Chaouiya C., (2006) GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems*, Vol.84, p.91-100.

- Grassi L., Tramontano A. (2011) Horizontal and vertical growth of *S. Cerevisiae* metabolic network. *BMC Evolutionary Biology*, Vol. 11, No.301, p.1-9.
- Grunde-Zeiferts U., Mozga I., Žukova T., Stalidzāns E. (2006) Therapy modelling combining methods of systems biology and automatic control theory. **No: International Scientific Conference „Animals. Health. Food Hygiene.”: proceedings**, November 11, 2006, Latvia, Jelgava, p.70-74.
- Grundspenķis J. *Sistēmu teorija un vadība: ESF projekta ietvaros izveidots metodiskais materiāls*. [Tiešsaite] [skatīts 2010.g. 05.aug.]. Pieejams: http://estudijas.itf.llu.lv/esf_materiali/default.aspx
- Grundspenķis J. *Sistēmu teorijas metodes: lekciju konspēkts*. [Tiešsaite]. [skatīts 2010.g. 04.aug.]. Pieejams: stpk.cs.rtu.lv/read_write/file/materiali/stm/stm.ppt
- Guelzim N., Bottani S., Bourguin P., Kepes F. (2002) Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics*, Vol. 31, p.60-63.
- Han J.-D. J. (2008) Understanding biological functions through molecular networks. *Cell Research*, Vol. 18, p.224-237.
- Han J.D., Bertin N., Hao T., Goldberg D.S., Berriz G.F., Zhang L.V., Dupuy D., Walhout A. J., Cusick M.E., Roth F.P., Vidal M. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, Vol. 430, p.88-93.
- Hallinan J.S., Jackway P.T. (2005) Network Motifs, Feedback Loops and The Dynamic of Genetic Regulatory Networks. **In: IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology: proceedings, 2005**, IEEE Press, p.90 – 96.
- Hallinan, J. Bradley, D. & Wiles, J. (2006). Effects of constitutive gene activation on the dynamics of genetic regulatory networks. **In: IEEE World Congress on Computational Intelligence: proceedings**, BC Canada, Vancouver, July 16-21, 2006, p.1-7.
- Hartl D.L., Clark A.G. (1997) Introduction. **In: Hartl D.L., Clark A.G. Principles of Population Genetics: 3rd ed.** USA: Sunderland, Sinauer Associates, p.1-11.
- Hartl D.L., Jones E.W. (2001a) Introduction to Molecular Genetics and Genomics **In: Hartl D.L., Jones E.W. Genetics: Analysis of Genes and Genomes: 5-th ed.** Canada: Jones and Bartlett Publishers, p.1-35.
- Hartl D.L., Jones E.W. (2001b) DNA Structure and DNA Manipulation. **In: Hartl D.L., Jones E.W. Genetics: Analysis of Genes and Genomes: 5-th ed.** Canada: Jones and Bartlett Publishers, p.36-49.
- Hase T., Niimura Y. (2012) *Protein-Protein Interaction Networks: Structures, Evolution, and Application to Drug Design*. Protein-Protein Interactions – Computational and Experimental Tools, p.405-426. [Tiešsaite]. [Skatīts 2013.g. 25.jūn.] Pieejams: http://bioinfo.tmd.ac.jp/~niimura/Hase&Niimura_2012.pdf

- Hase T., Niimura Y., Tanaka H. (2010) Difference in gene duplicability may explain the difference in overall structure of rotein-protein interaction networks among eukaryotes. *BMC Evolutionary Biology*, Vol. 10, No. 358, p.1-15.
- Hase T., Tanaka H., Suzuki Y., Nakagawa S., Kitano H. (2009) Structure of Protein Interaction Networks and Their Implications on Drug Design. *PLoS Computational Biology*, Vol. 5, Issue 10, p.1-9.
- Helikar T., Kochi N., Konvalina J., Rogers A. (2011) Boolean Modeling of Biochemical Networks. *The Open Bioinformatics Journal*, Vol.5, p.16-25.
- Helikar T., Rogers J. (2009) ChemChains: a platform for simulation and analysis of biochemical networks aimed to laboratory scientists. *BMC Systems Biology*, Vol. 3, p.58.
- Herbert M. (2004) *Reference and Tutorial Manual*. An Introduction to Biochemical Modeling using Jdesigner. Assisted by Abhishek Agrawal and Brian Gates. (Revision 1, Oct 10th, 2004) [Tiešsaite]. [skatīts 2009.g. 2.okt.]. Pieejams: <http://sbw.kgi.edu/software/jdesigner.htm>
- Herrgard M., Lee B.S., Portnoy V., Palsson B.O. (2006) Integrated analyses of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Research*, Vol. 16, No.5, p.627-635.
- Himsolt M. *GML: A portable Graph File Format*. [Tiešsaite]. [skatīts 2010.g. 11.nov.]. Pieejams: <http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/brandenburg/projekte/gml/gml-technical-report.pdf>
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. (2006) COPASI — a Complex Pathway Simulator. *Bioinformatics*, Vol. 22, p.3067-3074.
- Hu Z., Mellor J., Wu J., Yamada T., Holloway D., DeLisi C. (2005) VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Research*, Vol. 33, p.352-357.
- Hu Z., David M. Ng., Yamada T., Chen C., Kawashima S., Mellor J., Linghu B., Kanehisa M., Stuart J.M., DeLisi C. (2007) VisANT 3.0: new modules for pathway visualization, editing, prediction and construction. *Nucleic Acids Research*, Vol.35, p.625-632.
- Hu Z., Hung J.-H., Wang Y., Chang Y.-C., Huang C.-L., Huyck M., DeLisi C. (2009) VisANT 3.5 multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Research*, Vol. 37, p.115.-121.
- Hucka M., Hoops S., Keating S.M., Le Novère N., Sahle S., Wilkinson D.J.. (2008) Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions. SBML Level 2 Version 4 specification, Release 1. *Nature Precedings*, p.1-166.
- Hucka M., Finney A., Sauro H. M., Bolouri H., Doyle J. C., Kitano H. (2003) The systems biology markup language (SBML): a medium for representation

- and exchange of biochemical network models. *Bioinformatics*, Vol. 19, No.4, p.524–531.
- National Human Genom Research Institute (2007) *A Guide to Your Genome*. [Tiešsaite]. [skatīts 2013.g. 12.jūl.]. Pieejams: <http://www.genome.gov/Pages/Education/AllAbouttheHumanGenomeProject/GuidetoYourGenome07.pdf>
- Jana S., Deb J.K. (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Application in Microbiology and Biotechnology*, Vol. 67, Issue 3, p.289-298.
- Jarboe L.R., Zhang X., Wang X., Moore J.C., Shanmugam K.T., Ingram L.O. (2010) Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. *Journal of Biomedicine and Biotechnology*, p.1-18.
- Jong H., Geiselman J., Hernandez C., Page M. (2003) Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, Vol. 19, Issue 3, p.336-344.
- Jeong H., Tombor B., Albert R., Oltvai Z.N., Barabasi A.-L. (2000) The large-scale organization of metabolic networks. *Nature*, Vol. 407, p.651-654.
- Jeong H., Mason S.P., Barabasi A.-L., Oltvai Z.N., (2001) Lethality and centrality in protein networks. *Nature*, Vol. 411, p.41-42.
- Kanehisa M., Goto S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, Vol. 28, p.27-30.
- Kanehisa M., Goto S., Hattori M., Aoki-Kinoshita K.F., Itoh M., Kawashima S., Katayama T., Araki M., Hirakawa M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*. Vol. 34, p.354-357.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. (2011) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, Vol. 40, D109–114.
- Karp, P. D., Ouzounis, C. A, Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., Tsoka, S., u.c. (2005). Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research*, Vol. 33, Issue 19, p.6083–6089.
- Kauffman S. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, Vol. 22, p.437-467.
- Keating S. M., Bornstein B. J., Finney A., Hucka M. (2006) SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics*, Vol. 22, No. 10, p.1275–1277.
- KEGG Gene Database*. [Tiešsaite]. [skatīts 2013.g. 21.mai.]. Pieejams: <http://www.genome.jp/kegg/genes.html>
- Keseler I. M., Collado-Vides J., Santos-Zavaleta A., Peralta-Gil M., Gama-Castro S., Muñiz-Rascado L., Bonavides-Martinez C., u.c. (2011). EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Research*, Vol. 39, D583–590.

- Kholodenko B.N., Bruggeman F.J., Sauro H.M. (2005) Mechanistic and modular approaches to modeling and inference of cellular regulatory networks. **In:** Alberghina L., Westerhoff H.V. (eds.) *Systems Biology: Definitions and Perspectives*. London: Springer, p.143-162.
- Kim J.R., Yoon Y., Cho K.H. (2008) Coupled Feedback Loops Form Dynamic Motifs of Cellular Networks. *Biophysical Journal*, Vol. 94, No.2, p.359–365.
- Kim W.K., Marcotte E.M. (2008) Age-dependent evolution of the yeast protein interaction network suggests a limited role of gene duplication and divergence. *PLoS Computational Biology*, Vol. 4, Issue 11, p.1-10.
- Kimura M. (1984) *The Neutral Theory of Molecular Evolution*. UK: Cambridge, Cambridge University Press. 367 p.
- Kitano H. (2007) The theory of biological robustness and its implication in cancer. **In:** *Ernst Schering Foundation Symposium: proceedings*, Vol.61, p.69-88.
- Kitano H. (2004) Biological robustness. *Nature Reviews*, Vol. 5, p.826-837.
- Kitano H. (2005) Scientific and technical challenges for systems biology. **In:** Alberghina L., Westerhoff H. (Eds.) *Systems Biology. Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.373-385.
- Klamt S., Saez-Rodriguez J., Lundquist J., Simeoni L., Gilles E. (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, Vol. 7. 56 p.
- Klamt S., Saez-Rodriguez J., Gilles E. D. (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, Vol. 1. [Tiešsaite]. [skafīts 2009.g. 2.okt.]. Pieejams: <http://dx.doi.org/10.1186/1752-0509-1.2>
- Klipp E., Herwig R., Kowald A., Wierling C., Lehrach H. (2005) *Systems biology in practice. Concept, Implementation and Application*. Weinheim: WILEY-VCH Verlag GmbH & Co, KgaA. 486 p.
- Koschutski D., Junker B.H., Schwender J., Schreiber F. (2010) Structural analysis of metabolic networks based on flux centrality. *Journal of Theoretical biology*, Vol. 265, p.261-269.
- Kostromins A., Stalidzans E. (2012) Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism. *Biosystems*, Vol. 109, Issue 2, p.233-239.
- Krapivsky P.L., Redner S., Leyvraz F. (2000) Connectivity of Growing Random Networks. *Physical Review Letters*, Vol. 85, Issue 21, p.4629-4632.
- Krishna S., Semsey S., Sneppen K. (2007) Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proceedings of the National Academy of Sciences*, Vol. 104, Nr.52, p.20815-20819.
- Kuepfer L., Sauer U., Blank L. M. (2005). Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Research*, Vol. 15, Issue 10, p.1421–1430.

- Kwon Y.K., Cho K.H. (2007) Analysis of feedback loops and robustness in network evolution based on Boolean models. *BMC Bioinformatics*, Vol.8, 9 p.
- Le Novère N., Hucka M., Mi H., Moodie S., Schreiber F., Sorokin A., Demir E., Wegner K., Aladjem M.I., Wimalaratne S.M., Bergman F.T., u.c. (2009) The Systems Biology Graphical Notation. *Nature Biotechnology*, Vol. 27, No. 8, p.735-741.
- Lee K.Y., Park J.M., Kim T.Y., Yun H., Lee S.Y. (2010) The genome-scale metabolic network analysis of *Zymomonas mobilis* ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. *Microbial Cell Factories*, Vol. 9, p.1-12.
- Lee H., Popodi E., Tang H., Foster P.L. (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceeding of the National Academy of Sciences in USA*, Vol. 109, No. 41, p.E2774–E2783.
- Lewin R. (1996) *Patterns in evolution: the new molecular view*. USA: New York, Scientific American library. 246 p.
- Li D., Li J., Quyang S., Wang J., Wu S., Wan P., Zhu Y. u.c. (2006) Protein interaction networks of *Saccharomyces cerevisia*, *Caenorhabditis elegans* and *Drosophila melanogaster*: Large-scale organization and robustness. *Proteomics*, Vol. 6, p.456-461.
- Lodish H., Berk A., Matsudaira P., Kaiser C.A., Krieger M., Scott M.P., Zipursky L., Darnell J. (2004) *Molecular Cell Biology*: 5-th. USA: New York, Eds. W. H. Freeman, p.101-125.
- Longabaugh W., Davidson E., Bolouri H. (2009) Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. *Biochim. Biophys. Acta*, Vol. 1789, Issue 4, p.363-374.
- Longabaugh W., Davidson E., Bolouri H. (2005) Computational representation of developmental genetic regulatory networks. *Developmental Biology*, Vol. 283, p.1-16.
- Lucock M. (2007) *Molecular Nutrition and Genomics*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.1-9.
- Lynch M. (2007) The evolution of genetic networks by nonadaptive processes. *Nature Reviews Genetics*, Vol. 8, p.803-813.
- Madera S. (1998a) *Bioloģija 1.daļa*. Rīga: Zvaigzne ABC. 298 lpp.
- Madera S. (1998b) *Bioloģija 2.daļa*. Rīga: Zvaigzne ABC. 372 lpp.
- Massanori A. (2004) The metabolic world of *Escherichia coli* is not small. *Proceeding of National Academy of Sciences in USA*, Vol. 101, No. 6, p.1543-1547.
- Mazein A. *User Manual EPE3.0.0-alpha6*. [Tiešsaite]. [skafīts 2009.g. 24.okt.]. Pieejams: <http://garr.dl.sourceforge.net/project/epe/EPE/Documentation/Manual EPE3 .0.0 —alpha6.pdf>
- Mednis M., Aurich M.K. (2012) Application of string similarity ratio and edit distance in automatic metabolite reconciliation comparing

- reconstructions and models. *Biosystems and Information Technology*, Vol. 1, Issue 1, p.14-18.
- Mednis M., Brusbardis V., Galvanauskas V. (2012) Comparison of genome-scale reconstructions using ModeRator. **In:** *13th IEEE International Symposium on Computational Intelligence and Informatics*: proceedings, Hungary, Budapest, p.79–84.
- Mendes P. (1993) GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput. Appl. Biosci.*, Oct, 1993, Vol. 9, Issue 5, p.563-571.
- Mensonides F., Schuurmans J., Mattos M., Hellingwerf K., Brul S. (2002) The Metabolic Response of *Saccharomyces Cerevisiae* to Continuous Heat Stress, *Molecular Biology Reports*, Vol.1, p.103-106.
- Milo R., Shen-Orr S., Itzkovitz S., Kashtan N., Chklovskii D., Alon U. (2002) Network motifs: Simple Buildings Blocks of Complex Networks. *Science*, Vol. 298, No. 5594, 824-827.
- Moran N.A., McLaughlin H.J., Sorek R. (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*, Vol. 323, No. 5912, p.379-382.
- Myers C. J. (2010) *Engineering Genetics Circuits*. USA: New York, Chapman & HALL/CRC Mathematical and computational Biology Series. Taylor and Francis Group, LLC, p.1-278.
- Müssel C., Hopfensitz M., Kestler H. (2010) BoolNet—an R package for generation, reconstruction and analysis of Boolean networks, *Bioinformatics*, Vol.26, p.1378-1380.
- Nachman M.W., Crowell S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics*, Vol. 156, Issue 1, p.297-304.
- Natal A.W., Riel V. (2006) Dynamic modeling and analysis of biochemical networks: mechanism-based models and model-based experiments. Briefings in *Bioinformatics*, Vol. 7, No. 4, p.364-374.
- NCBI – The National Center for Biotechnological Information (2010) *What is a genome?* [Tiešsaite]. [Skatīts 2010.g. 4.feb.]. Pieejams: http://www.ncbi.nlm.nih.gov/About/primer/genetics_genome.html
- NetBuilders Concept* (2006) [Tiešsaite]. [skatīts 2009.g. 2.okt.]. Pieejams: <http://strc.herts.ac.uk/bio/maria/Apostrophe/Pdf/NetBuilder-prime.pdf>
- Newman M.E.J. (2003) The structure and function of complex networks. *SIAM Review*, Vol. 45, p.167-256.
- Nishio Y., Usuda Y., Matsui K., Kurata H. (2008) Computer-aided rational design of the phosphotransferase system for enhanced glucose uptake in *Escherichia coli*. *Molecular Systems Biology*, Vol. 4, p.1-12.
- Noort V., Snel B., Huynen M.A. (2004) The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Reports*, Vol. 5, p.280-284.
- Novak B., Chen K.C., Tyson J.J. (2005) Systems biology of the yeast cell cycle engine. **In:** Alberghina L., Westerhoff H. (Eds.) *Systems Biology*.

- Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.305-323.
- Osis J.J. (1969) *Automātiskā vadība un regulēšana*. Rīga: Zvaigzne, 1-16.lpp.
- Ossowski S., Schneeberger K., Lucas-Lledo J., Warthmann N., Clark R.M., Shaw R.G., Weigel D., Lynch M. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, Vol. 327, p.92–94.
- Palsson, B. Ø. (2006) *Systems Biology: Properties of reconstructed networks*. UK: Cambridge, Cambridge University Press. 334 p.
- Pastor-Satorras R., Smith E., Solé R.V. (2003) Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, Vol. 222, p.199-210.
- Paszek E. (2007) *Boolean networks*. Connexions module: m12394. [Tiešsaite]. [skatīts 2010.g. 21.dec.]. Pieejams: http://cnx.org/content/m12394/1.5/content_info
- Patil A., Nakamura H. (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Letters*, Vol. 580, p.2041–2045.
- Paxson R., Zannella K. (2007) System biology: Studying the world's most complex dynamic systems. *Journal The Math Works News & Notes*, June, p.4-7.
- Philips N. Salomon M., Custer A., Ostrow D., Baer C.F. (2009) Spontaneous Mutational and Standing Genetic (Co)variation at Dinucleotide Microsatellites in *Caenorhanditis briggsae* and *Caenorhabditis elegans*. *Molecular Biology and Evolution*, Vol. 26, No.3, p.659-669.
- Pentjuss A., Odzina I., Kostromins A., Fell D.A., Stalidzans E., Kalnenieks U. (2013) Biotechnological potential of respiring *Zymomonas mobilis*: A stoichiometric analysis of its central metabolism. *Journal of Biotechnology*, Vol.165, p.1-10.
- Qin H., Lu H.H.S., Wu W.B., Li W.-H. (2003) Evolution of the yeast protein interaction networks. *Proceedings of the National Academy of Sciences in USA*, Vol. 100, No. 22, p.12820-12824.
- Ravasz E., Somera A.L., Mongru D.A., Oltvai Z.N., Barabasi A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, Vol. 297, p.1551-1555.
- Raipulis J. (2002) *Ģenētikas pamati: tālmācības līdzeklis*. Rīga: Izdevniecība RaKa, 9.-157.lpp.
- Reil, T. (1999) Dynamics of gene expression in an artificial genome: Implications for biological and artificial ontogeny. **In:** *5th European Conference on Artificial Life*: Floreano, D., Mondada, F. & Nicoud, J. D. (eds.) proceedings, Berlin: Springer Verlag, p.457–466.
- Rodrigues J.F.M., Wagner A. (2011) Genotype networks, innovation, and robustness in sulfur metabolism. *BMC Systems Biology*, Vol. 5, Issue 39, p.1-13.

- Roy S., Lane T., Werner-Washburne M. (2007) A Simulation Framework for Modeling Combinatorial Control in Transcription Regulatory Networks. *UNM Computer Science Technical Report*, TR-CS-2007-06, p.1-10.
- Rubina T. (2013) The procedure of evolution modelling of biochemical networks structure. *Biosystems and Information Technology*, Vol.2, No.2, p.19-25.
- Rubina T., Stalidzans E. (2010a) Topological features and parameters of biochemical network structure. **In:** *8-th Industrial Simulation Conference*: publication of EUROSIS, proceedings, June 7-9, 2010, Hungary, Budapest, p.228-236.
- Rubina, T., & Stalidzans, E. (2010b) Software Tools for Structure Analysis of Biochemical Networks. **In:** *4-th International Scientific Conference „Applied Information and Communication Technologies”*: proceedings, April 22-23, 2010, Latvia, Jelgava, p.33-49.
- Rubina T., Stalidzans E. (2012a) Evolution modeling algorithm of biochemical networks. **In:** *10-th Industrial Simulation Conference*: a publication of EUROSIS, proceedings, June 4-6, 2012, Czech Republic, Brno, p.24-30.
- Rubina T., Stalidzans E. (2012b) Evolution of control loops of biological systems. **In:** *5-th International Scientific Conference „Applied information and Communication Technologies”*: proceedings, April 26-27, Latvia, Jelgava, p.317-324.
- Rubina, T. (2012) Tools for analysis of biochemical network topology. *Biosystems and Information Technology*, Vol.1, No.1, p.25-31.
- Robins G, Pattison P., Koskinen J. (2008) *Network degree distribution. Technical report to Australian Defence Science and Technology Organisation*. University of Melbourne, p.1-9.
- Rubina T., Brusbardis V. (2009) Applications of biochemical networks discovering control mechanisms in systems biology. **In:** *Annuals Students International Scientific Conference “Youth in Science and Profession Practice”*: proceedings, April 23, 2009, Latvia, Jelgava, p.1-7.
- Sausiņa L. (2010) *Bioloģija vidusskolai 4. Daļa*. Rīga: Zvaigzne ABC, 208 lpp.
- Schellenberger J., Park J. O., Conrad T. M., Palsson B. Ø. (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, Vol. 11, Issue 1, p.213.
- Schellenberger J., Que R., Fleming R. M. T., Thiele I., Orth J. D., Feist A. M., Zielinski D. C., u.c. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, Vol. 6, Issue 9, p.1290-1307.
- Schilstra M.J., Bolouri H. (2002) The logic of gene regulation. **In:** *3-rd International Conference on Systems Biology*: poster abstract, 2002.
- Shannon P., Markiel A., Ozier O., Baliga NS., Wang JT., Ramage D., Amin N., Schwikowski B., Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, Vol. 13, No. 11, p.2498-2504.

- Sharan R., Ideker T. (2006) Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, Vol. 24, No. 4, p.427-433.
- Selga T. (2008) *Šūnu bioloģija*. Rīga: LU Akadēmais kais apgāds. 343 lpp.
- Seo J.S., Chong H., Park H.S., Yoon K.O., Kim J.J., Hong J.H., Kim H., Kim J.H. Kil J.I., Park C.J., Oh H.M., Lee J.S., Jin S.J., Um H.W. u.c.. (2005) The genome sequence of the ethanogenic bacterium *Zymomonas mobilis* ZM4. *Nature Biotechnology*, Vol. 23, No.1, p.63-68.
- Snoep J., Westerhoff H. (2005) From isolation to integration, a systems biology approach for building the Silicon Cell. In: Westerhoff H.V. Alberghina L., (eds.) *Systems Biology: Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.13-30.
- Solé V.R., Pastor-Satorras R., Smith E., Kepler T.B. (2008) A model of large-scale proteome evolution. *WSPC/Guidelines, Advances in Complex Systems*, p.1-12.
- Sole R.V., Pastor-Satorras R., Smith E., Kepler T. (2002) A model of large-scale proteome evolution. *Advances in Complex Systems*, Vol. 5, p.43-54.
- Sontag E., Vilz-Cuba A., Laubenbacher R., Jarrar A.S. (2008) The Effect of Negative Feedback Loops on the Dynamics of Boolean Networks. *Biophysical Journal*, Vol. 95, p.528-526.
- Sorensen H.P., Mortensen K.K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*, Vol. 115, Issue 2, p.113-128.
- Strazewski P., Tamm C. (1990) Replication Experiments with Nucleotide Base Analogues. *Angewandte Chemie*, Vol. 29, Issue 1, p.36-57.
- Strazdiņš I. (2001) Diskrētā matemātika:mācību grāmata programmēšanas, tehnisko un ekonomikas specilitāšu studentiem. Rīga: Zvaigzne ABC, 101.-107. lpp.
- Suderman M., Hallett M. (2006) Tools for visually exploring biological networks. *Bioinformatics*, Vol. 23, Issue 20, p.2651-2659.
- Swings J., De Ley J. (1977) The biology of *Zymomonas*. *Bacteriological Reviews*, Vol. 41, p.1-46.
- Thiele I., Palsson B. Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, Vol. 5, Issue 1, p.93-121.
- Thiele I, Swainston N., Fleming R., Hoppe A., Sahoo S. U.c. (2013) A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, Vol. 31, p.419-425.
- Vazquez A. (2003) Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, Vol. 67, Art.No. 056104, p.1- 15.
- Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O.u.c. (2001) The Sequence of the Human Genome. *Science*, Vol. 291, No.5507, p.1304-1351.
- Volkers R. (2004) *Gēni un DNS*. / Walker R. (2004) *Genes and DNA*. Stīva

- Džonsaw priekšv. [tulk. Frīdenberga A.] Rīga: Eve, 63 lpp.
- Wagner A. (2011) The molecular origins of evolutionary innovations. *Trends in Genetics*, Vol.27, p.397-410.
- Wagner A. (2009) Molecular evolution, Networks in. **In:** Meyers R.A. (ed.) *Encyclopedia of Complexity and System Science*. London: Springer, p.1-21.
- Wagner A. (2005) *Robustness and Evolvability in Living Systems*. USA: New Jersey, Princeton University Press, p.15-38.
- Wagner A., Fell D. (2001) The small world inside large metabolic networks. *Proceedings of the Royal Society B*, Vol. 268, No. 1478, p.1803-1810.
- Wagner A. (2003) How the global structure of protein interaction networks evolves. *Proceedings of the Royal Society B*, Vol. 270, p.457-466.
- Walsh B. (2003) Population-genetics models of the fates of duplicate genes. *Genetica*, Vol. 118, p.279-294.
- Watson J., Wiles J., Hanan J. (2003) Towards more Relevant Evolutionary Models: Integrating an Artificial Genome with a Developmental Phenotype. **In:** *1st Australian Conference on Artificial Life*: Abbas H., Wiles J. (eds.) proceedings, 2003, Australia, Canberra, p.288-298.
- Watson J.D., Crick F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, Vol. 171, Art.No. 4356, p.737–738.
- Watts D.J., Strogatz S.H. (1998) Collective dynamics of ‘small-world’ networks. *Nature*, Vol. 393, p.440-442.
- Weir B.S. (1996) Genetic data analysis: 2nd ed. Sinauer Associates, p.8-15.
- Wegner K., Knabe J., Robinson M., Egri-Nagy A., Schilstra M., Nehaniv C. (2007) The NetBuilder' project: development of a tool for constructing, simulating, evolving, and analysing complex regulatory networks. *BMC Systems Biology*, P72, 2 p.
- Westerhoff H., Hofmeyr J-H. (2005) What is system biology? From genes to functions and back. **In:** Alberghina L., Westerhoff H.V. (eds.) *Systems Biology: Definitions and Perspectives*. London: Springer, p.119-141.
- Whitaker J. W., Letunic I., McConkey G. A., Westhead D. R. (2009). metaTIGER: a metabolic evolution resource. *Nucleic Acids Research*, Vol. 37, D531–D538.
- Wilson R.J. (1972) *Introduction to Graph Theory*. New York: Oliver and Boyd, Edinburgh Academic Press, p.9-21.
- Wolfe K.H., Shields D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, Vol. 387, p.708-713.
- Wuensche A. (2009) Discrete dynamics lab: tools for investigating cellular automata and discrete dynamical networks. Artificial life models in software. In: Adamatzky A, Kosinski M, Eds.: London: Springer, p.215-258.
- Yamada T., Bork P. (2009) Evolution of biomolecular networks – lessons from metabolic and protein interactions. *Nature Reviews, Molecular Cell*

Biology, Vol. 10, p.791-803.

- Yamada T., Sugiyama T., Tamaki N., Kawakita A., Kato M. (2009) Adaptive radiation of gobies in the interstitial habitats of gravel beaches accompanied by body elongation and excessive vertebral segmentation. *BMC Evolutionary Biology*, Vol. 9, Issue 145, p.1-14.
- Yang S., Pan C., Tschaplinski T.J., Hurst G.B., Engle N.L., Zhou W., Dam P., Xu Y., Jr M.R., Dice L., Johnson C.M., Davison B.H., Brown A.D. (2013) Systems Biology Analysis of *Zymomonas mobilis* ZM4 Ethanol Stress Responses. *PLoS One*, Vol. 8, Issue 7, p.1-14.
- Yook S.-H., Oltvai Z.N., Barabasi A.-L. (2004) Functional and topological characterization of protein interaction networks. *Proteomics*, Vol. 4., p.928-942.
- Zhang A. (2009) *Protein Interaction Networks: Computational analysis*. Cambridge: Cambridge University Press, p.1-20.
- Zhang J., Shakhnovich E.-I. (2008) Sensitivity-dependent model of protein-protein interaction networks. *Physical Biology*, Vol. 5, p.1-6.
- Zhao J., Ding G.-H., Tai L., Yu H., Yu Z.-H., Luo J.-H., Cao Z.-W., Li Y.-X. (2007) Modular co-evolution of metabolic networks. *BMC Bioinformatics*, Vol. 8, No. 311, p.1-12.
- Zhao J., Yu H., Luo J., Cao Z.W., Li Y. (2006) Complex networks theory for analysing metabolic networks. *Chinese Science Bulletin*, Vol. 51, No.13, p.1529-1537.
- Zheng J., Zhang D., Przytycki P., Zielinski R., Capala J., Przytycka T. (2010) SimBoolNet—a Cytoscape plugin for dynamic simulation of signaling networks. *Bioinformatics*, Vol. 26, p.141-142.
- Zinovyev A., Viara E., Calzone L., Barillot E. (2008) BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics Applications Note*, Vol. 24, No. 6, p.876-877.
- Zinovyev A., Calzone L. *Binom Manual Version 1.0*. InstitutCurie, Service de Bioinformatique. [Tiešsaite]. [skatīts 2010.g. 7.janv.]. Pieejams: http://bioinfo-out.curie.fr/projects/binom/docs/Binom_Manual_v1.0.pdf
- Бакай А.В., Кочиш И.И., Скрипниченко Г.Г. (2006с) **В кн.:** Бакай А.В., Кочиш И.И., Скрипниченко Г.Г. *Генетика: учебник для студентов высших учебных заведений*. Москва: КолосС, 448 с.
- Давидич М.И., Постников Е.Б. (2007) Булевская модель цикла деления клетки дрожжей *Schizosaccharomyces pombe*: динамика в случае нормальных и возмущенных начальных условий. В м.к.: *2-ая Международная Конференция “Математическая Биология и биоинформатика”*: материалы конференции, том 2, 2007, с.377-386.
- USTUA (2002) Термодинамика живых систем. Жизнь как информационный процесс: Лекция из курса Концепции современного естествознания. (Concepts of modern natural sciences. Lecture: Thermodynamics of living systems. Life as information process.). The Ufa state technical university of aviation (USTUA).

[Tiešsaite]. [Skatīts 2011.g. 5.dec.].
<http://www.ugatu.ac.ru/ddo/KSE/01/0118/ks011800.htm>

Pieejams:

Latvia University of Agriculture
Faculty of Information Technologies
Department of Computer Systems

Tatjana Rubina

**STRUCTURE EVOLUTION COMPUTER MODELLING
OF BIOCHEMICAL NETWORKS OF NON-COHERENT
IMPORTANCE**

SUMMARY

of the Thesis for acquiring Doctoral Degree in the Field
of Information Technologies (Dr.sc.ing.)



IEGULDĪJUMS TAVĀ NĀKOTNĒ



Development and Design of the Thesis financed by
the European Social Fund

Tatjana Rubina

Signature

Jelgava 2013

PARTICULARS

Research was executed at: Latvia University of Agriculture, Faculty of Information Technologies, Department of Computer Systems, Lielā 2, Jelgava, Latvia.

Experimental research was executed at: Latvia University of Agriculture, Faculty of Information Technologies, Department of Computer Systems, Lielā 2, Jelgava, Latvia.

Scientific Adviser of the Doctoral Thesis: Dr.sc.ing. Egils Stalidzans, Associate Professor, Latvia University of Agriculture.

The thesis was approved at the expanded academic session of the Department of Computer Systems, Faculty of Information Technologies of the Latvia University of Agriculture on 15 May, 2013. Minutes No.6.

The doctoral thesis was developed with the assistance of the European Social Fund (ESF) project “Support for the Implementation of Doctoral Studies at Latvia University of Agriculture” agreement No. 2009/0180/1DP/1.1.2.1.2./09/IPIA/VIAA/ 017.

Official Reviewers:

1. Professor of the Rezekne Higher Education Institution, Dr.sc.ing. Oļegs Užga-Rebrovs;
2. Professor of the Riga Technical University, Dr.habil. Gaļina Merkurjeva;
3. Postdoctoral scientific assistant of the University College London, PhD Dace Rukliša.

The defence of the doctoral thesis will take place at the open session of the Promotion Council in the field of Information Technologies of LUA at __ a.m. on _____, 2014, Room __, Faculty of Information Technologies. Lielā 2, Jelgava, Latvia.

The thesis can be accessed at the LUA Fundamental Library, Lielā 2, Jelgava, and online at http://lufb.llu.lv/promoc_darbi.html.

You are welcome to send your comments, signed and scanned to secretary of Promotion Council – Lielā 2, Jelgava, LV-3001; phone (+371) 63005621; e-mail: tatjana.tabunova@llu.lv.

Council Secretary: lecturer, Mg.Paed. Tatjana Tabunova.

TABLE OF CONTENTS

APPROBATION OF PHD THESIS.....	4
INTRODUCTION.....	6
Theme topicality	6
The aim and the tasks of the PhD thesis	8
Research methods	9
Scientific novelty	9
Research theses.....	10
Practical novelty.....	10
PhD thesis structure and volume.....	10
1. EVOLUTION OF BIOLOGICAL SYSTEMS.....	11
2. BIOCHEMICAL NETWORKS AND THEIR ANALYSIS.....	11
3. MODELLING OF EVOLUTIONARY GROWTH OF BIOCHEMICAL NETWORKS	13
Modelling procedure of structure evolution of biochemical networks.....	15
Modelling algorithm of structure evolution of biochemical networks	19
4. SOFTWARE TOOL BINESA.....	23
5. APPLICABILITY ASSESSMENT OF BIOCHEMICAL NETWORK MODELS	24
6. COMPUTER SIMULATIONS OF STRUCTURE EVOLUTION OF BIOCHEMICAL NETWORKS	26
CONCLUSIONS	34
The main PhD thesis results.....	34
Conclusions and development prospects	36
BIBLIOGRAPHY	38

APPROBATION OF PHD THESIS

The research results are presented in the following publications:

- 1) **Rubina T.**, Stalidzans E. (2013) BINESA – a software tool for evolution modelling of biochemical networks' structure. In: Proceedings of the 14th IEEE International Symposium on Computational Intelligence and Informatics, CINTI'2013, ISBN 978-1-4799-0194-4, November 19-21, 2013, Budapest, Hungary, p.345-350. (Indexed in the Web of Knowledge and SCOPUS database).
- 2) **Rubina T.**, Mednis M., Stalidzans E. (2013) Agreement assessment of biochemical pathway models by structural analysis of their intersection. In: Proceedings of the 14th IEEE International Symposium on Computational Intelligence and Informatics, CINTI'2013, ISBN 978-1-4799-0194-4, November 19-21, 2013, Budapest, Hungary, p.411-418. (Indexed in the Web of Knowledge and SCOPUS database).
- 3) **Rubina T.** (2013) The procedure of evolution modelling of biochemical networks structure. Biosystems and Information Technology, Vol.2, No.2, ISSN 2255-8004, p.19-25.
- 4) **Rubina T.**, Stalidzans E. (2012) Evolution modeling algorithm of biochemical networks. In: Proceedings of the 10th Industrial Simulation Conference, ISC'2012. A publication of EUROSIS, ISBN 978-90-77381.71.7, June 4-6, 2012, Brno, Czech Republic, p.24-30. (Indexed in the Thomson/Reuters Web of Knowledge).
- 5) **Rubina T.**, Stalidzans E. (2012) Evolution of control loops of biological systems. In: Proceedings of the 5th International Scientific Conference "Applied Information and Communication Technologies" AICT'2012, ISBN 978-9984-48-065-7, April 26-27, 2012, Jelgava, Latvia, p.317-324.
- 6) **Rubina T.** (2012) Tools for analysis of biochemical network topology. Biosystems and Information Technology, Vol.1, No.1, ISSN 2255-8004, p.25-31.
- 7) **Rubina T.**, Stalidzans E. (2010) Software Tools for Structure Analysis of Biochemical Networks. In: Proceedings of the 4th International Scientific Conference "Applied Information and Communication Technologies" AICT'2010, ISBN 978-9984-48-022-0, April 22-23, 2010, Jelgava, Latvia, p.33-49. (Indexed in the Thomson/Reuters Web of Knowledge).
- 8) **Rubina T.**, Stalidzans E. (2010) Topological features and parameters of biochemical network structure. In: Proceedings of the 8th Industrial Simulation Conference ISC'2010. A publication of EUROSIS, ISBN 978-90-77381.5-57, June 7-9, 2010, Budapest, Hungary, p.228-236. (Indexed in the Thomson/Reuters Web of Knowledge).
- 9) Odzina I., **Rubina T.**, Rutkis R., Kalnenieks U., Stalidzans E. (2010) Structural Model of biochemical network of *Zymomonas mobilis*

- adaptation for glycerol conversion into bioethanol. In: Proceedings of the 4th International Scientific Conference "Applied Information and Communication Technologies" AICT'2010, ISBN 978-9984-48-022-0, April 22-23, 2010, Jelgava, Latvia, p.50-54. (Indexed in the Thomson/Reuters Web of Knowledge).
- 10) **Rubina T.**, Brusbardis V. (2009) Applications of biochemical networks discovering control mechanisms in systems biology. In: Proceedings of the Annuals Students International Scientific Conference "Youth in science and Professional practice", ISBN 978-9984-784-99-1, April 23, 2009, Jelgava, Latvia, p.1-7.
 - 11) **Zukova T.**, Stalidzans E. (2006) Systems biology – interaction science of biology and information technology. In: Proceedings of the International Scientific Conference "Information Technologies for Rural Development", ISBN 9984-784-13-4, October 19-20, 2006, Jelgava, Latvia, p.120-130.
 - 12) Grunde-Zeiferts U., Mozga I., **Žukova T.**, Stalidzāns E. (2006) Therapy modelling combining methods of systems biology and automatic control theory. In: Proceedings of International Scientific Conference "Animals. Health. Food Hygiene.", ISSN 1407-1754, November 10, 2006, Jelgava, Latvia, p.70-74.

The research results were presented at the following conferences:

- 1) „Agreement assessment of biochemical pathway models by structural analysis of their intersection”. 14th IEEE International Symposium on Computational Intelligence and Informatics CINTI'2013, November 19-21, 2013, Budapest, Hungary
- 2) „BINESA – a software tool for evolution modelling of biochemical networks’ structure”. 14th IEEE International Symposium on Computational Intelligence and Informatics CINTI'2013, November 19-21, 2013, Budapest, Hungary
- 3) “Evolution modeling of biochemical network structure with BINESA”. 6th International Scientific Conference “Applied Information and Communication Technologies AICT'2013”, April 25-26, 2013, Jelgava, Latvia
- 4) “Evolution modeling algorithm of biochemical networks”. 10th Industrial Simulation Conference, ISC'2012, June 4-6, 2012, Brno, Czech Republic
- 5) “Evolution of control loops of biological systems”. 5th International Scientific Conference “Applied Information and Communication Technologies AICT'2012”, April 26-27, 2012, Jelgava, Latvia
- 6) “Topological features and parameters of biochemical network structure”, The European Multidisciplinary Society for Modelling and simulation Technology. 8th Industrial Simulation Conference, ISC'2010, June 7-9, 2010, Budapest, Hungary

- 7) “Software Tools for Structure Analysis of Biochemical Networks”. 4th International Scientific Conference “Applied Information and Communication Technologies AICT’2010”, April 22-23, 2010, Jelgava, Latvia
- 8) “Approach of biochemical-network structure analysis”. 4th International Scientific Conference “Students on their Way to Science”, LUA TF, May 14-15, 2009, Jelgava, Latvia
- 9) “Evolutionary modeling of cellular control networks”. International Scientific Conference “Youth in Science and Professional Practice”, April 23, 2009, Jelgava, Latvia

The research results were presented at the following seminars:

- 1) **Rubina T.** The Types and Computer Modeling of Biochemical Networks, Seminar of LUA Biosystems group, January 13, 2010, Jelgava, Latvia
- 2) **Rubina T.** Analysis of Structure Parameters, Seminar of LUA Biosystems group, January 3, 2010, Jelgava, Latvia
- 3) **Rubina T.** Control Loops of Biochemical Networks and their Biological Meaning. Data formats in Systems Biology, Seminar of LUA Biosystems group, July 8, 2010, Jelgava, Latvia
- 4) **Rubina T.** Senescence as Genetic Program, Seminar of LUA Biosystems group, November 3, 2010, Jelgava, Latvia
- 5) **Rubina T.** Modeling of Boolean Networks, Seminar of LUA Biosystems group, January 4, 2012, Jelgava, Latvia
- 6) **Rubina T.** Research data processing and visualization using Matlab, Seminar of LUA Biosystems group, November 21, 2012, Jelgava, Latvia
- 7) **Rubina T.** Structure’s Evolution Modeling of Biochemical Networks using Software Tool BINESA, Seminar of LUA Biosystems group, August 19, 2013, Jelgava, Latvia

INTRODUCTION

Theme topicality

At the end of the last century the explosion of Information Technologies (IT) enables to apply them as an irreplaceable instrument for scientific and practical research implementation in different areas including molecular biology, genetics, and population genetics. It is especially related to the research of living organisms while this process is difficult for several reasons. Firstly, this process requires execution of a huge number of experiments in laboratories (Chen et al., 2009a) and time resources to obtain the data and understanding about separate aspects of organisation of organism components, its functioning and evolution in certain conditions. Secondly, in several cases examination of living organisms is impossible due to the accepted ethical rules,

complexity of living organisms, the health and life-threatening circumstances. Therefore, as a solution in such cases IT, modelling and simulations can be applied to gain knowledge, build and expand the understanding of living organisms, develop models and algorithms based on existing experimental data, knowledge and facts.

Living organisms are complex, nonlinear, self-organised and self-regulated biological systems. Inside any biological system (in the cell, tissue, organ and organism) under different conditions processes of different types (regulation, physical and chemical processes) operate that can be presented and described in a form of biochemical pathways. Biochemical pathways and their regulation are organised in a form of networks. A peculiar structure lies at the basis of a network which is formed from different elements (proteins, metabolites, genes, small molecules) and their connections determining interaction of structure elements. But the organisation and operation of the network define, regulate and control the information and mechanisms that are coded in genes.

There are processes of different importance in biochemical networks. Some of biochemical processes such as glycolysis are present in almost every known organism and can be referred to as the vital processes. It can be explained and has an evolutionary basis: organisms with changes in important processes did not survive freeing up a place for other organisms, in which processes that are most important for existence were not significantly affected by the changes. At the same time these insignificant, rudimentary processes that mutations can even destroy will have no significant effect on the behaviour of the organism. Stochastic mutations lead to the formation of nonstochastically formed networks. So the mechanisms of their formation are non-trivial and the natural selection executes a complex task that needs to be explored in networks of different importance since it is an essential aspect of network formation.

The following question arises – how the biochemical networks of different importance evolve and which changes in the structure depending on the influence of different types of mutations occur when the mutation probability ratio changes?

If significant copying errors in vital processes by the influence of mutations are not permitted during natural selection and the next generation of organism maintains the critical functionality needed for the survival, then the changes in less important processes are allowed. Consequently, the existing biochemical networks can be broken or even new structures may emerge. So, some structure changes of biochemical network, gaps and remains of the structure emerging during evolution process can influence the resistance of the organism both to the environmental changes and medical therapies provoking emergence of the alternative enforcement variants of various important processes. Therefore, the structure's evolution modelling of biochemical networks even in a simplified manner can provide an insight in the diversity of organism response by influencing different biochemical processes.

While the exploration of genetic mutations influence on the structure changes of biochemical networks of equal and mixed importance in laboratory conditions is difficult and time-consuming, in some cases even impossible, the computer modelling and computer simulations are necessary for multi-faceted and detailed examination of this issue. It is necessary to develop simplified formalisms in order to investigate thoroughly questions under consideration while evolutionary processes occurring in nature are very complex and all their nuances cannot be realized by the computer simulations due to computational capacity. The author of the PhD thesis intends to focus on the network modelling with reactions of different importance that requires development of simplified and suitable for the computational capacity an evolution modelling procedure and an algorithm that takes into account the importance of biochemical processes, as well as software tool which will be able to execute the computer experiments of structure evolution according to the offered algorithm and procedure.

The aim and the tasks of the PhD thesis

The aim of the thesis is to develop a modelling approach and determine the influence of different types of mutations on the structure evolution of biochemical networks of non-coherent importance by means of the developed modelling approach.

In order to achieve the aim of the PhD thesis several tasks were defined:

- 1) examine the methods of the structural analysis of biochemical networks and the methods of their representation;
- 2) study the evolution process of biochemical networks and the factors influencing it;
- 3) analyse the computer modelling approaches of structure evolution of biochemical networks;
- 4) explore the existing software tools designated for the computer modelling of biochemical networks and for the analysis of their structure;
- 5) develop the simplified algorithm and the structure analysis criteria and implement them in software tool prototype for the purpose of imitating to a genome attached structure evolution of biochemical networks;
- 6) carry out the experiments aimed to imitate the evolution of to a genome attached structure of biochemical networks that describes the processes of equal or different importance;
- 7) analyse the influence on structure evolution of each separate type of mutation and the influence of simultaneously operating several mutation types using computer model;
- 8) determine the topological parameters which characterise the applicability of the model of biochemical networks.

Research methods

The programming language *Visual Basic* was used in algorithm development for software tool prototype BINESA.

Different mutation operators were used in structure evolution modelling of biochemical networks which are described in molecular genetics theory and correspond to the part of mutation types that exist in nature.

The topological or structural analysis was applied to perform the structural analysis of biochemical networks that is grounded on the concepts and methods of graph theory and is used in the field of systems biology. During development of the PhD thesis SBML (*Systems Biology Markup Language*), GML (*Graph Modelling Language*) standards for the import and export of model structure of biochemical network, as well as file formats GV (built-in file format of software tool *GraphViz*) and PNG for visualisation of network structure were used.

The non-linear regression analysis was applied to determine the best non-linear functional dependence between the corresponding topological parameter of biochemical network structure and some of the factorial characteristics, for example, the number of network elements, the element incoming or outgoing degree.

The methods of probability theory were used in the offered modelling algorithm of structure evolution, for example, for the selection of the offspring of next generation.

Statistical methods were used for analysis of the results of network structure evolution.

The regression analysis was applied to determine the best functional dependence between the viability duration of biochemical network and the corresponding mutation operator, as well as the two-dimensional graphs (errorbars) including two-dimensional plots with two Y-axis with different scaling were used to display evolution dynamics.

Scientific novelty

The scientific novelty of the PhD thesis.

- The formalisation procedure for the structure evolution task of biochemical networks of non-coherent importance was developed.
- The algorithm for evolution dynamics modelling of to a genome attached structure of non-coherent biochemical networks was developed in case of mutations which do not change the length of genes.
- The prototype of software tool *BINESA* (**B**iochemical **N**etwork **S**tructure **A**nalys**E**r) was developed for structure evolution modelling and topological analysis of the structure of biochemical networks of non-coherent importance.

- The approach for assessing applicability of biochemical network models was developed using topological parameters of biochemical network structure.

Research theses

- Disregard of the reaction importance of biochemical networks has a significant impact on the computer model behaviour of structure evolution of biochemical networks.
- The fast assessment of the applicability of models intersection can base on the analysis of topological parameters of biochemical network structure.
- The influence of evolutionary pressure differs in coherent networks of different importance.
- The dynamic of structure topological parameters illustrates the dynamic of structural model quality of biochemical networks during the evolution process.

Practical novelty

Practical novelty is provided by the application of the proposed algorithm of structure evolution modelling of biochemical networks for modelling biochemical networks of organisms of different types taking into account the influence of the damaged structures on the biotechnological modifications of network operation with a biotechnological or therapeutic purpose. Computer simulations of biochemical network evolution can also be used for developing therapies in personalised medicine.

The developed prototype of software tool *BINESA* can be used for carrying out computer experiments of to a genome attached structure evolution of biochemical networks of non-coherent importance and for modelling evolution dynamics.

The modelling of biochemical networks of non-coherent importance can be applied to establish stable network structures in synthetic biology.

The PhD results can also be used for the development of stoichiometric models by assessing quality and agreement degree of the available models of biochemical networks.

PhD thesis structure and volume

The PhD thesis is written in Latvian containing abstract, introduction, 6 chapters, conclusions, bibliography, and 15 annexes, including 19 tables, 111 figures, 27 formulae, 184 pages in total. 224 literature sources were used.

1. EVOLUTION OF BIOLOGICAL SYSTEMS

The cell is considered as biological system. Its function determines complex processes while the most part of them are related to the interaction of molecules in biochemical networks. Biological systems operate according to the instructions encoded in genes which are passed through the generations and subject to the mutability.

Despite the fact that the agents of different species differs drastically they have much more similarities in case of biology and genetics, for example, 98.5% of the sequence of human genome is similar to the sequence of chimpanzee genome (Vokers, 2004).

The changing environment and genes mutations of different types changing the morphological, physiological and biochemical properties provide the evolution of organisms. Biochemical networks are one of integral structures of organism that is subject to the evolutionary changes. The main driving forces of evolution are the mutations that introduce the changes of genetic material and natural selection performs the selection of the more adaptable and strongest offspring of the next generation. The mutations are divided in three groups by the influence source: gene mutations, chromosomal mutations and genome mutations. Their appearance frequency and influence on the biochemical processes and networks describing these processes depends on several factors. Mutations generally occur relatively infrequently, but the influence of physical, chemical and biological mutagenic factors increases their occurrence frequency. The influence of evolution on the biochemical processes including the biochemical network structure that emerges in the function and behaviour of biological system is important to study for development of biotechnological modifications and medical therapies in order to assess possible response of the networks subjected to the evolution.

2. BIOCHEMICAL NETWORKS AND THEIR ANALYSIS

Insight living organisms operate metabolic, gene regulation and signal transduction networks that describe biochemical reactions, biochemical and biophysical processes. Biochemical networks can illustrate and describe interaction of genes and their products, interaction of proteins, as well as interaction of reactants, substrates and their products.

The basis of biochemical networks is the structure. To understand the biology at the system level, it is necessary to study the system structure before including structure of biochemical networks.

Structural or topological analysis of biochemical networks is grounded on the concepts, methods, and algorithm of graph theory and uses also the concepts of system theory, as well as methods and algorithms of clustering analysis. Structural analysis provides insights not only in the topological properties of biochemical networks of biological system, but also in the

dynamic properties of biological system while fundamental properties of dynamic behaviour are often controlled and established by the network structure (Klamt et al., 2006).

During examination of scientific literature, publications and analysis of software tools developed in the field of systems biology, as well as their description and user manuals, several topological properties, measurements and features of networks were identified that can be divided in four groups (Rubina, Stalidzans, 2010a): network metrics, network motifs, topological parameters and topological features of the network (see Figure 1).

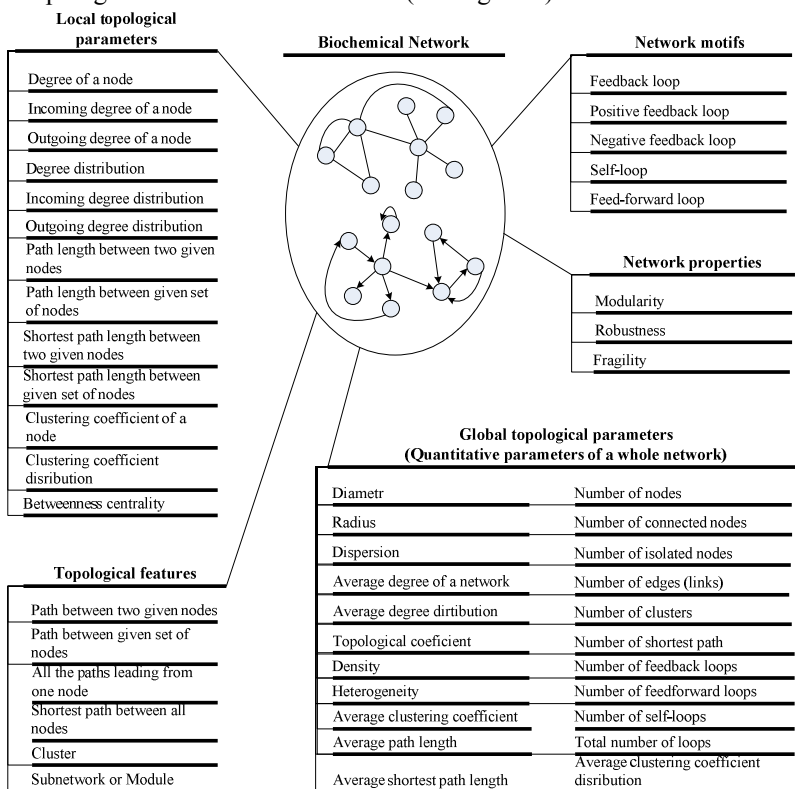


Figure 1. Topological measurements of biochemical networks

Topological parameters depending on the examined level of the network (level of network elements when the network elements are examined separately or network level when the network is examined as a whole) can be divided in local and global parameters. The local topological parameters characterise separate network elements or components, but global parameters describe the network as a whole and are calculated using local parameters.

3. MODELLING OF EVOLUTIONARY GROWTH OF BIOCHEMICAL NETWORKS

The key aim of biochemical network modelling is to supplement and expand researchers' understanding of the local and global properties and behaviour of biological systems which demonstrate the system in response to different stimuli. Understanding how networks evolve is a fundamental issue in real-life complex networks and can provide clarity and insights into the structure and function of the networks (Chen et al., 2009c).

To study topological properties of biochemical networks researchers have used mainly three types of network models which are characterised by the certain distribution of network degree and clustering coefficient: random network, scale-free network, and hierarchical network. Furthermore, to establish the topological properties and peculiarities of real-life networks from an evolutionary perspective by assuming that the current topology of a network is formed through a series of network assembly events and network evolution events (Chen et al., 2009c), different researchers groups have also developed several network growth models, for example, duplication-mutation (with complement) models, duplication-divergence models, random growing network models, random static network models, preferential attachment or scale-free model, small world network models, Boolean network model, duplication-deletion-divergence models, and others. These growth models investigate network growth for the purpose of defining the principles of network establishment and organisation in the course of evolution.

These models and other models used in practice for studying topological properties and growth or evolution of networks have several shortcomings. Firstly, evolutionary approach that is used in the mentioned models investigates network growth commonly and does not consider essential properties of biochemical processes. One of such property is process importance (Rubina, Stalidzans, 2012, Rubina, 2013). Not all processes of biological system are equally important, and the offspring with lethal mutations affecting essential processes necessary for viability is separated during natural selection. But the offspring with neutral or beneficial mutations gets a better chance to participate in further evolution.

There are several biochemical processes that remain intact in almost all organisms. These indicate the fact that organisms with substantial deviation in the genes sequences of proteins that catalyse essential processes die off. One of such processes is glycolysis (Romano, Conway, 1996). Typical glycolysis process is Embden-Meyerhof-Parnas glycolysis. Another process is Entner-Dudorova glycolysis. In total, processes that are related with energy provision and in particular glycolysis can be regarded as vital processes which do not obtain (gain) significant alternatives in spectrum of all living forms. There are other processes that seriously influence viability of organism, for example, biochemical pathways that describe synthesis of proteins, nucleic acids, lipids

and other components (Copley, 2000). At the same time there are such biochemical pathways whose changes can be recompensed by other mechanisms and damage of these pathways do not significantly affect the viability of organism.

The author of the PhD thesis offers a simplified division of the processes in three groups by process importance. The processes of the first group are vital and without them a biological system cannot ensure self-operation and retain viability (living). The processes of the second group identify (provide) the quality of existence and can affect viability of a biological system only due to many defects of such processes. The processes of the third group affect inessential properties of a biological system, for example, some visual features or alternative biochemical processes. For this reason, the properties of these groups of processes should be marked out and defined. So, the changes of genes sequences that establish or regulate the processes of the first and other groups will have different effect on the biological system. For this reason, even insignificant changes that occur in gene sequence regulating vital processes can be more dangerous. They can introduce significant changes in system's behaviour and provoke deviations in its function.

The author of the PhD thesis presumes that vital processes are similar among the greatest part of biological systems and permits poor differences in genetic material or genes sequences that regulate and establish these processes.

The second shortcoming of growth models is lack of underlying genome except *Artificial genome* model. The network level changes are only theoretically explained by the existence of different mutation types that act on the underlying genome instead of being introduced as a result of occurring mutations.

The third shortcoming of the used growth models is remoteness from real process of evolution and introduction of only two or three mutation types in models: gene duplication (whole genome duplication, local genes duplication, retrotransposition) and/or deletion (Farid, Christensen, 2006, Yamada et al, 2009, Yamada, Bork, 2009, Wagner, 2009), and divergence that arises due to the influence of point mutations (Aldana et al., 2007, Gibson, Goldberg, 2011). These mutations are acknowledged as main driving forces of network evolution. Gene duplication at the network level arises as node addition with all the links of duplicated node. When some gene is lost, some node is deleted with all these links in the network.

The second group of events appears as link addition or deletion that is theoretically grounded on genetic changes that do not change gene completely or change gene regulation (Noort et al., 2004, Yamada, Bork, 2009, Yamada et al., 2009). Such genetic mutations can be point mutations, nucleotide insertions or deletions (Wagner, 2003). So, there are many other mutation types that can influence network evolution. They do not change gene completely, but modify gene partly or its regulation that can arise as link addition or deletion in the network level. As Yamada with colleagues (Yamada et al., 2009) studying

metabolic and protein interaction networks note, the network nodes and links evolution is related with cell genetic material, but links can change all the time. If nodes are not affected, the links rewiring can occur without gene duplication and links change with higher frequency than nodes change. Also Berg with colleagues note (Berg et al., 2004) that coefficient of links addition or deletion by the mutation influence is n time higher than coefficient of network growth with duplications. The slowest gene duplication (Berg et al., 2004), as well as deletion (Wagner, 2009) influences the network size only. Directly links dynamics act as dominant driving force of evolution in the structure formation of scale-free networks than nodes duplication (Farid, Christensen, 2006).

The author of the PhD thesis offers the algorithm of structure evolution modelling of biochemical networks that eliminate shortcoming of evolution algorithms used in the considered models of network growth. Firstly, evolution algorithm takes into account the property of processes importance and allows defining the importance level of the processes (reactions or links). Secondly, the evolution of network nodes and links is related with cell genetic material, hence the evolution of network structure is grounded on the evolutionary changes of connected genome that occur as a result of mutations and natural selection in the offered algorithm. Thirdly, since links change with higher frequency than nodes duplication and links dynamics is acknowledged as a key driving force of evolution in the structure formation of scale-free networks, then the number of nodes remains unchanged in the offered algorithm paying central attention to the links dynamics.

Modelling procedure of structure evolution of biochemical networks

Within the PhD thesis for the developed algorithm and procedure of structure evolution of biochemical networks the following terminology is used:

- biochemical network – set of reactions that describes some biochemical process(es);
- structure of biochemical network – representation of network elements and their interaction;
- gene – the unit of genetic information that is displayed as text string including four-character and is attached to each separate reaction;
- genome – set of genes that is attached to the specific model of biochemical network;
- mutation – alteration of gene sequence that occurs with a certain probability;
- mutation process – process in which each gene of genome is subjected to the mutations;
- generation – one computational iteration that includes mutation process and selection of an offspring of the next generation.

Evolutionary changes of biochemical networks structure emerge under the influence of occurring mutations at the level of a genome. As genes define and regulate the structure (architecture) and function of the network, that way in order to explore the evolution of biochemical networks on their structure, it is necessary to connect genes to the processes and network links in the form of nucleotide sequence, as well as to define the assessment criteria of genome changes. So, the evolution modelling of biochemical networks structure can be realized:

- 1) connecting genes to the nucleotide sequence and network reactions which define corresponding reactions,
- 2) executing simulations of genome evolution or alteration of genes sequences which can occur by pressure of various mutation types,
- 3) generating structure changes based on the changes of genes sequences.

The procedure of structure evolution modelling of biochemical networks includes six consecutive stages (see Figure 2).

The first stage of the procedure is definition of initial structure data that includes three main sub-stages: 1) definition of network nodes and links, manually or by loading an existing model, for example, SBML model; 2) initial genome definition of an organism generating automatically the test sequences of genes or using genome from existing database; 3) genome or separate genes connection to network attaching the genes to the network reactions (i.e. links).

The network evolution is based on the changes of genetic material. The evolution of the offered algorithm of structure evolution modelling is realised at the level of genome sequence and is transformed into network structure changes. According to the central dogma of molecular biology (see Sub-section 1.2) it is possible to connect gene to each network link (Rodrigues, Wagner, 2011, Wagner, 2011) that encodes enzyme which ensures course of the corresponding chemical reaction (see Figure 3). Each gene has nucleotide sequence that is defined in the form of text string of four characters A, C, T, and G.

Within the evolution algorithm it is assumed that all genes have equal length and in the course of evolution algorithm the genes length remain constant, for example, 1,000 nucleotides long genes.

The second stage of the procedure includes topological analysis of initial structure of a network. The analysis of initial network structure is necessary for the determination of applicability of network model to the research purpose. The analysis of initial network structure provides several topological measurements and motifs of the structure which are used for structure assessment. The analysis of network structure can provide such topological parameters as a node degree, an incoming and outgoing degree of a node, a degree distribution or an incoming/outgoing degree distribution, a number of nodes, a number of connected and isolated nodes, a number of links, an average clustering coefficient, a distribution of neighbour's number, a

distribution of clustering coefficient, and can determine such network motifs as feedback loops or control loops and self-loops.

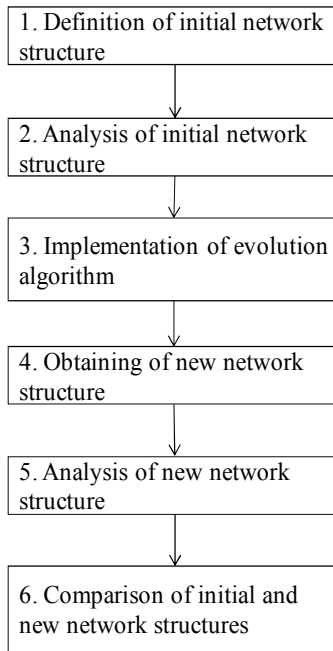


Figure 2. **The procedure of structure evolution modelling of biochemical networks**

The third and fourth stages of the procedure include implementation of evolution algorithm that consists of two stages: evolution of genome and evolution of structure (see Sub-section 3.2). To implement evolution algorithm that ensures the obtaining of next generation genomes of an organism, at the beginning the evolution parameters should be defined.

Evolution algorithm is generated n times where n is user-defined number of generations and in its course n generations genomes are generated. For example, to obtain the genome of the first generation of an organism, the initial genome of an organism is used, but to obtain the genome of the i -th generation of an organism, the genome of $(i-1)$ -th generation of an organism is used.

The genome of new generation is obtained via realisation of two processes which determine the evolution of biological systems: mutation process and natural selection. Implementation of these processes is grounded on information provided in literature about mutations of genetic material and natural selection (see Sub-section 1.2).

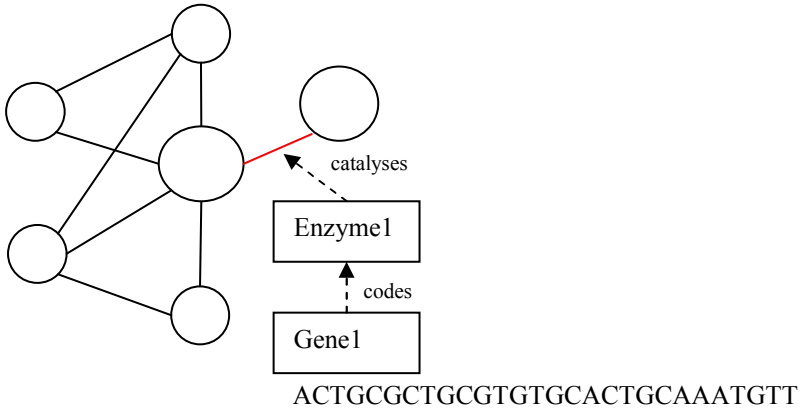


Figure 3. **Attachment of genome to the network structure or reaction**

To obtain the genome of new t -th generation, 10 genome copies of $(t-1)$ -th generation are generated which are further subject to mutation process. From 10 genome copies the possible candidates of the offspring are selected evaluating the concordance of genomes to the benchmark-genome. As a result of performance of mutation process and candidate selection the M genome candidates are obtained from which one genome is chosen with probability Pe_i . The probability Pe_i of each separate genome candidate to be chosen for further evolution depends on the correspondence (compliance) of genome candidate to the initial genome assuming that the initial genome is benchmark-genome with the highest possible characteristics of vitality.

The structure changes can be generated for previously chosen generations based on evolutionary changes of genome of the corresponding generation.

The fifth stage of the procedure includes topological analysis of the new structure of a network received as a result of evolution.

The sixth stage of the procedure includes comparison of the initial structure and the newly obtained structure. During structural analysis the computed topological parameters become the comparing criteria of several structures which are used for comparison of the initial structure and structure that was erased in evolution for the purpose of detecting and assessing the undergone changes. In order to draw conclusions about the influence of chosen types of mutations on network structure, the similarities and differences are detected between the examined structures, comparing parameters, dynamics of their changes and motifs of the initial structure and the newly obtained structure.

Modelling algorithm of structure evolution of biochemical networks

Evolution algorithm (see Stage 3 of Figure 2) consists of two main stages: genome evolution and structure evolution (Rubina, 2013). Genome evolution executes n times and genomes of n generations are created during it. Each time step 10 genome copies of the previous ($t-1$)-th generation genome are generated executing evolution algorithm in order to obtain the genome of t -th generation ($t=1,2,..n$). The number of genome copies can be changed, but smaller number of genome copies decreases the number of potential candidates of the next generation offspring and it can have an impact on the evolution results. Genome copies are subjected to a mutation process which can include implementation of several mutation operators.

- **Point mutation** is an alteration of a single nucleotide in the gene sequence that is chosen uniformly at random, for example, nucleotide A is replaced by C.
- Under influence of **nucleotide inversion operator** the gene section from k -th to l -th position chosen uniformly at random is rearranged in reverse order, i.e. nucleotides of gene sequence are rearranged in reverse order, for example, nucleotide sequence ACTGTGATCGCGTAATGGC from position 7 to position 11 is transformed to a sequence ACTGTGCGCTAGTAATGGC.
- **Missense mutation operator** is used in genome copies assessment in order to determine the concordance to the benchmark-genome. When this mutation operator is chosen then the genome sequences are compared by codons or by three consecutive nucleotides which encode the corresponding amino acids (see Sub-section 2.3). One amino acid can be encoded by several nucleotide triplets. Even if one nucleotide is mutated under influence of point mutation, the encoded amino acid can remain the same. During comparison of genome sequences (comparison of the newly obtained sequence and sequence of benchmark-genome) it is checked whether an amino acid is the same or is changed comparing to the benchmark-genome.
- **Nonsense mutation operator** is used in genome copies assessment in order to determine the concordance to the benchmark-genome similarly to the missense mutation operator. This operator is used to determine whether the amino acid encoded by the corresponding codon was changed or not (changed to the stop codon) comparing to the benchmark-genome.
- **Duplication operator** copies gene that is chosen uniformly at random, and the number of genes increases by one unit in genome.
- **Deletion operator** removes gene that is chosen uniformly at random, and the number of genes decreases by one unit in genome.

- Under influence of **inversion operator** the nucleotide sequence of gene is rearranged in reverse order.
- **Translocation operator** breaks genes of two nonhomologous (two different) chromosomes on *k-th* position and exchanges the distracted ends of chromosomes.

On the first stage of evolution modelling procedure we choose which mutation operators should be included in evolution process according to the research purpose and define mutation probability to each operator.

After the mutation process execution **10 obtained genomes of *t-th* generation are compared** to the benchmark-genome and **estimated**. Each gene of the newly obtained genome is compared to its initial sequence in the benchmark-genome and to the all other genes sequences of benchmark-genome calculating coefficients: Rgk_i – concordance coefficient of *i-th* gene to its initial sequence and Qgk_{ij} – concordance coefficient of *i-th* gene to all other *j-th* genes sequences of the benchmark-genome, where *j* is the serial number of gene. When the missense mutation and/or nonsense mutation operators are chosen then sequences are compared by triplets, i.e. by three consecutive nucleotides. If no operator from these is chosen then gene sequences are compared by one nucleotide. The calculated coefficient Rgk_i characterises a part of correspondence of gene alternative form and initial gene, and it can take values in the range [0; 1]. In turn, coefficients Qgk_{ij} characterise a part of correspondence of gene alternative form and all other initial genes in benchmark-genome, and it can also take values in the range [0; 1], where *i* is the serial number of the compared gene in the newly obtained genome and *j* is the serial number of genes in benchmark-genome.

If concordance coefficient of *i-th* gene $Rgk_i < 0.2$ and it regulates insignificant reaction (essentiality level 3) then the maximum coefficient is detected $\max Qgk_i = \max(Qgk_{ij})$ from coefficients Qgk_{ij} . If $\max Qgk_i > 0.2$ then it is accepted that the *i-th* gene is mutated on the *j-th* gene and has begun exercising *j-th* gene functions.

If concordance coefficient of *i-th* gene $Rgk_i < 0.5$ and it regulates qualitative reaction (essentiality level 2) then the maximum coefficient is detected $\max Qgk_i = \max(Qgk_{ij})$ from coefficients Qgk_{ij} . If $\max Qgk_i > 0.5$ then it is accepted that the *i-th* gene is mutated on the *j-th* gene and has begun exercising *j-th* gene functions.

If concordance coefficient of *i-th* gene $Rgk_i < 0.7$ and it regulates important reaction (essentiality level 1) then the maximum coefficient is detected $\max Qgk_i = \max(Qgk_{ij})$ from coefficients Qgk_{ij} . If $\max Qgk_i > 0.7$ then it is accepted that the *i-th* gene is mutated on the *j-th* gene and has begun exercising functions similar to the *j-th* gene functions.

After execution of the mutation and estimation processes the **process of candidate selection of new generation genome** follows. According to the process of natural selection the strongest and better adapted individual survives. In the proposed algorithm such individual is considered as the strongest one

which has a genome with higher similarity to the benchmark-genome and has kept all genes that regulate vital reactions and as many genes as possible that regulate qualitative reactions, but at least one qualitative reaction is obligatory. In this way the strongest individual gets better chance of being chosen as the new generation offspring.

The candidate of new generation offspring should correspond to the one of the following conditions.

- If the modelled biochemical network includes only vital links then the concordance coefficients of all genes should be $Rgk \geq 0.7$.
- If the modelled biochemical network includes only qualitative links then the concordance coefficients of all genes should be $Rgk \geq 0.5$.
- If the modelled biochemical network includes only insignificant links then the concordance coefficients of all genes should be $Rgk \geq 0.2$.
- If the modelled biochemical network includes links of different importance then the concordance coefficients of all vital genes should be $Rgk \geq 0.7$, but at least one gene regulating quality links should be $Rgk \geq 0.5$.

Consequently, from 10 renderers of the next generation candidates those renderers are separated which are subject to one of the following conditions:

- 1) genome that includes at least one gene that regulates vital reaction and differs from benchmark-genome by more than 30% with concordance coefficient $Rgk < 0.7$;
- 2) genome in which all genes that regulate qualitative reactions differ from benchmark-genome by more than 50% with concordance coefficient $Rgk < 0.5$.

To select the strongest candidate, a series of transformations is executed defining probability ratio of each genome candidate being chosen for the offspring of the next generation.

The cumulative concordance coefficient $sumRgk_j$ called genome concordance coefficient is calculated for each genome candidate that is equal to the sum of concordance coefficients Rgk_i of all genes (in total, m genes) of the corresponding j genome.

The concordance coefficient of genome candidate called normalised concordance coefficient of a genome $gRgk_j$ is normalised dividing cumulative concordance coefficient $sumRgk_j$ by total number of genes m of the corresponding genome.

The sum $TotalRgk$ of normalised concordance coefficients of all genome candidates is calculated.

The probability ratio of each genome candidate is calculated dividing normalised concordance coefficient $gRgk_j$ of a genome j by the sum $TotalRgk$ of normalised concordance coefficients of all genome candidates.

If the modelled network includes qualitative links then the probability ratio of genome candidate is calculated dividing its normalised concordance coefficient $gRgk_j$ by the sum $TotalRgk$ of normalised concordance coefficients

of all genome candidates and multiplying outcome by the part of number of active qualitative reactions (1):

$$Pe_j = \frac{gRgk_j \cdot \frac{q}{Tq}}{TotalRgk} , \quad (1)$$

where q – number of genes that regulate active qualitative reactions in the network (without repetitions),
 Tq – total number of genes that regulate qualitative reactions in the network (without repetitions).

The M genome candidates with alternative gene forms are obtained executing the processes of mutations and candidates selection from which one genome is chosen with probability Pe_j .

The intensity indicator $Ints_i$ of each reaction i is calculated based on the essentiality level of reaction and on the concordance coefficient (below denoted as variable x) of the gene of the chosen genome that regulates the corresponding reaction. Intensity indicator characterises how intensely reaction occurs or links work. In order to take into account the importance of processes, it is assumed that:

- intensity of reactions of the essentiality level 1, $Ed=1$ (vital processes or reactions) changes by the power law $f(x)=x^2$;
- intensity of reactions of the essentiality level 2, $Ed=2$ (qualitative processes or reactions) changes by the linear law $f(x)=x$;
- intensity of reactions of the essentiality level 3, $Ed=3$ (vital processes or reactions) changes by the polynomial law $f(x)=1.3x^3-3.42x^2+3.12x$.

The second stage of the algorithm includes structure evolution which provides the structure of the new generation. The structure changes are generated based on each separate gene concordance coefficient Rgk_i of the offspring genome and the essentiality level of reaction which is regulated by the corresponding gene.

- Vital reaction and all these links are removed from network if the concordance coefficient of its regulating gene is less than $Rgk_i < 0.7$.
- The intensity of vital reaction is reduced according to the above-described law if concordance coefficient of its regulating gene is in range $0.7 \leq Rgk_i < 0.9$.
- Qualitative reaction and all these links are removed from network if the concordance coefficient of its regulating gene is less than $Rgk_i < 0.5$.
- The intensity of quality reaction is reduced according to the above-described law if the concordance coefficient of its regulating gene is in range $0.5 \leq Rgk_i < 0.8$.
- Insignificant reaction and all these links are removed from network if the concordance coefficient of its regulating gene is less than $Rgk_i < 0.2$.

- The intensity of insignificant reaction is reduced according to the above described law if the concordance coefficient of its regulating gene is in range $0.2 \leq Rgk_i < 0.5$.

4. SOFTWARE TOOL BINESA

Within the scope of the PhD thesis the prototype of software tool BINESA was developed which includes implementation of the proposed algorithm of structure evolution. Software tool *BINESA* allows to realise evolution procedure developed during work on the PhD thesis. Furthermore, it is possible to carry out computer experiments of structure evolution using BINESA and to estimate results corresponding to the offered procedure of structure evolution of biochemical networks.

The prototype of software tool *BINESA* can be used for the following purposes:

- 1) to model structure evolutionary changes that arise as a result of genome evolution taking into account the importance level of processes;
- 2) to imitate genome evolution that is connected to the network structure;
- 3) to create structure of new network models;
- 4) to import structure of network models in *SBML* and *GML* formats;
- 5) to perform topological analysis of biochemical network structure (to calculate local and global topological parameters, to determine network motifs);
- 6) to compare two network structures by the topological parameters;
- 7) to calculate topological parameters of many structures and to export these analysis results in *CSV* file format aimed for further processing, analysis and visualisation;
- 8) to analyse and compare initial structure and structure that was derived after evolution or during evolution based on evolution process data and evolution results, as well as on topological parameters of the structures;
- 9) to visualise structure of biochemical network marking out links of different importance and intensity (see Annex 8).

The prototype of software tool *BINESA* is written in *Visual Basic* programming language and developed on *Microsoft Access* using *DAO* data access technology. *BINESA* can be used as standalone software tool that requires the following software: *Windows XP* operating system and *Microsoft Access 2007* or higher version and a computer with technical parameters corresponding to the recommended parameters for stable software operation.

The examined model of biochemical network in *BINESA* is connected to the certain genome, and *BINESA* allows defining and attributing of three importance levels to the processes/reactions. *BINESA* includes only such mutations which do not change genome length in order to reduce the computational costs of genes comparison. The offered algorithm of to a genome attached structure evolution of biochemical network allows estimating

the influence of one part of evolution mechanisms on the structure of biochemical network.

The developed prototype of software tool *BINESA* allows not only modelling of structure changes of biochemical network according to the evolution algorithm and to the offered evolution modelling procedure, but also ensures receiving and estimation possibility of modelled evolution dynamics.

5. APPLICABILITY ASSESSMENT OF BIOCHEMICAL NETWORK MODELS

The fast development of the sequencing techniques enables relatively fast reconstruction of biochemical reaction networks in many organisms. In case of model development, it would be useful to assess the quality of the available models looking for the best one or to find suitable parts of a published model to build a new one. The differences or contradictions in reconstructions, especially genome scale reconstructions, give an insight in the scope of models and the level of agreement among different authors about the topic of interest.

Within the PhD thesis in order to assess the coherence, similarities and differences of the models an approach is offered in which structural analysis is used. To assess the coherence of two models, it is proposed to build an intersection model which includes the intersecting part of initial models, then to perform topological analysis of the intersection and initial models and, finally, to compare their topological parameters.

Within the PhD thesis the assessment of two intersection models is demonstrated comparing the current intersection and its initial models for the purpose of establishing similarities and differences using topological analysis and determining topological measurements that can be used to assess the quality of the model and the level of agreement and determining the applicability of the model.

Within the study (Rubina et al., 2013) two pairs of scale-free models were compared from *BioCyc* public database: 1) bacterium *Escherichia coli* models “*ecol199310cyc*” and “*ecol316407cyc*”, 2) yeast *Saccharomyces cerevisiae* models *iND750* and *iLL672*. The comparison tool of the stoichiometric models *ModeRator* (Mednis et al., 2012) was used to compare models and generate their intersection model. But the prototype of software tool *BINESA* (www.biosystems.lv/binesa), which was developed while working on the PhD thesis, was used for the topological analysis of the structure. The structural analysis of model pairs of the same organism and its intersection model demonstrates cases of highly similar and different models.

The study reveals very different topological parameters of intersections of *E.coli* and *S.cerevisiae* model pairs. The models built by the same group of authors like in case of *E.coli* can be taken as example of high agreement between models and can be interpreted as consensus part of two models. Similar topological parameters give indication that the intersection model may

function as a standalone network model even without further improvements. At the same time the intersection of *S.cerevisiae* models demonstrates very different structural properties and most probably the intersection model would not be able to function even after significant improvement.

Some topological parameters such as number of metabolites, reactions and links indicate a size of the intersection model compared to the initial one. That can be interpreted as a high agreement in a small part or a low agreement in a larger part of the models. Therefore, the intersection model parameters, like a number of the metabolites and a number of the reactions, cannot only be used to assess the level of agreement.

A number of metabolites, a number of reactions, an approximation of incoming/outgoing degrees, a distribution of neighbours and a distribution of clustering coefficients of the intersection model are weak indicators of the agreement level of two initial models.

In this study the topological parameters such as an incoming and outgoing degree distribution, a percentage of the low or high interconnectivity metabolites (low or highly interconnected), the distribution of neighbour's number and an approximation of neighbour's number are recognised as informative structural parameters. A low agreement of the model pair resulting in a fragmented, poor quality model can be indicated by low values of an average degree, an average incoming degree, an average outgoing degree and the average number of the neighbours. A low agreement of the model pair can be detected also by the distribution of the incoming and outgoing degrees of the metabolites: a high percentage of the low interconnectivity metabolites (one or two links) and a low percentage of the hubs (more than ten links).

The structural analysis of the intersection and association of different models is important both in comparing evolution of models and building the new models on the basis of several available models. In intersection analysis when comparing the original model with the model received during or after the evolution, not only models but also their intersection can be investigated for the purpose of estimating which reactions have survived evolution and which were replaced.

The intersection analysis allows also estimating the process of natural evolution and determining which groups of reactions and processes remain common for different organisms despite the long evolution process. The automated generation of an intersection of two models combined with its structural analysis can give indication about the agreement level between models created by different authors and between metabolic models of a particular organism.

This approach can be used to rapidly determine the similarities of models of different organisms.

6. COMPUTER SIMULATIONS OF STRUCTURE EVOLUTION OF BIOCHEMICAL NETWORKS

The aim of computer experiments of structure evolution of biochemical networks (further – experiments) is to assess the influence of the corresponding mutation types on the genome and on the attached to a genome structure and to establish the peculiarities of the evolution dynamics of biochemical network structure and the peculiarities of structure changes that arise as a result of the influence of several types of mutations in case of reactions of equal and different importance. Experiments were carried out using the prototype of modelling tool *BINESA* which includes implementation of algorithm of to a genome attached structure evolution that was developed during work on the PhD thesis.

Two models of biochemical networks were chosen for the execution of experiments: test model of small size and real model of small size of bacterium *Zymomonas mobilis* (further – *ZMO*). The structure of test model was attached to the artificial genome with 100 nucleotides long genes. Test model of biochemical network consists of 9 metabolites and 14 reactions that form 19 links. *ZMO* model of biochemical network (Pentjuss et al, 2013) consists of 81 metabolites and 96 reactions that form 287 links. The structure of *ZMO* model was attached to the artificial genome with 1,300 nucleotides long genes where the length of genes corresponds to the average length of genes which take part in examined model of bacterium *Z.mobilis*.

In order to estimate the influence of mutations in case of reactions of equal and different importance evolution experiments were executed using the following models:

- 1) test model of small size which has only vital reactions (further – test model of vital reactions or vital test model);
- 2) test model of small size which has only qualitative reactions (further – test model of qualitative reactions or qualitative test model);
- 3) test model of small size which includes vital and qualitative reactions, as well as insignificant reactions (further – test model of reactions of different importance or mixed test model);
- 4) *ZMO* model which has only vital reactions (further – *ZMO* model of vital reactions or vital *ZMO* model);
- 5) *ZMO* model which has only qualitative reactions (further – *ZMO* model of qualitative reactions or qualitative *ZMO* model);
- 6) *ZMO* model which includes vital and qualitative reactions, as well as insignificant reactions (further – *ZMO* model of reactions of different importance or mixed *ZMO* model).

There are significant differences observed (noticed) in case of network of non-coherent importance between qualitative networks and mixed networks, as well as between vital networks and qualitative networks. Their evolution

varies by marked differences of viability duration. It is explained by the higher allowed deviation of concordance coefficient and by the ability of qualitative networks to last even with one reaction (see Figure 4). When the number of qualitative reactions decreases, the probability that mutation will affect genes which provide (establish) the remaining reactions in all 10 alternative genome copies decreases, i.e. in all renderers of candidates of next generation offspring. Hence the evolution of qualitative network can overachieve even 100,000 generations with one reaction. In its turn, the presence of vital reactions in the network is marginal and the loss of vital reactions is inadmissible. It arises both in structure evolution of vital network and in structure evolution of mixed network, and noticeably affects the viability duration of the structure.

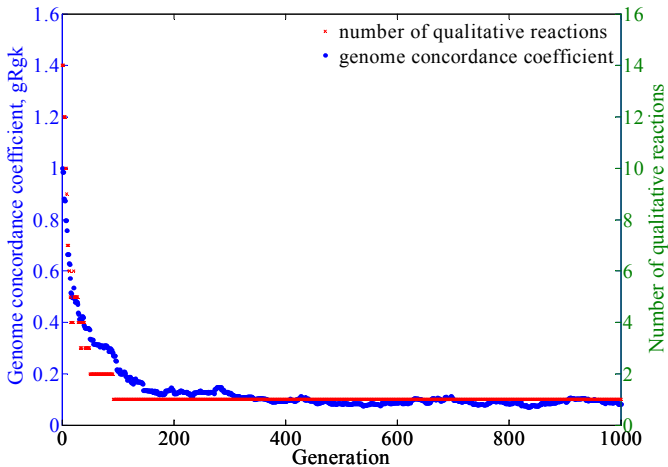


Figure 4. The changes of the Genome concordance coefficient and changes of the number of qualitative reactions by influence of nucleotide inversion with probability 10%

In the figure genome concordance coefficient that is equal to average concordance coefficient of all its genes is shown.

The rapid decrease in the number of qualitative and insignificant reactions is observed on the mixed network at the evolution beginning by the influence of translocation, deletion, nucleotide inversion (see Figure 5) and inversion. However, the network continues evolution with one qualitative reaction and all vital reactions. In their turn, the qualitative and insignificant reactions can be renewed by the influence of nucleotide inversion, inversion (see Figure 6) and translocation when the reversion (back) mutation occurs.

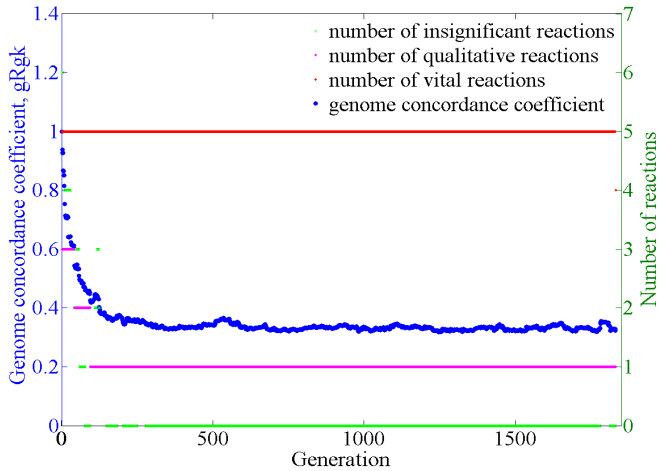


Figure 5. The changes of the Genome concordance coefficient and changes of the number of reactions of different importance by influence of nucleotide inversion with probability 10%

Back mutation, but with smaller frequency, is noticed while analysing separately the influence of point mutation on the network structure. Despite the fact that the evolution process continues longer than the evolution process by the influence of other mutation types, point mutations are accumulated and with each successive generation they decrease the genome concordance coefficient and make the structure of each successive offspring more fragile.

The viability duration changes by the power law with negative exponent and high determination coefficient in the networks of non-coherent importance (in vital, qualitative and mixed networks) by the influence of all considered mutation types. In case of a separate network type only values of viability duration that depends on the certain mutation type varies. The more disruptive mutation influence is, the faster the evolution process ends and the shorter the viability duration is. In their turn, the mixed network shows better resilience to the mutations and survives longer comparing to the vital network (see Figures 7 and 8) considering the influence of each type of mutation separately. But the viability duration of mixed network is much shorter comparing to the viability duration of qualitative network.

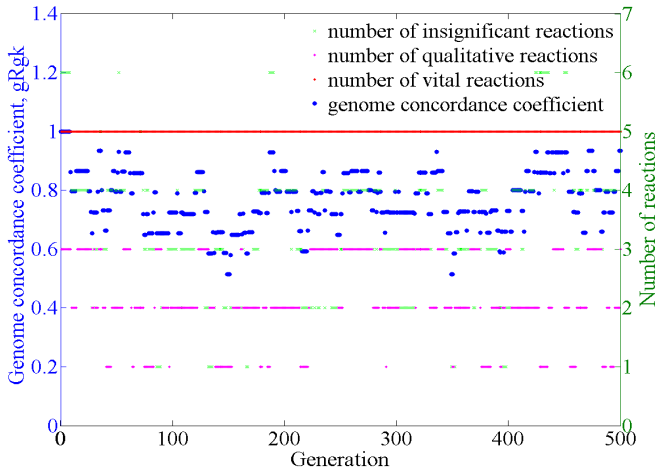


Figure 6. The changes of the Genome concordance coefficient and changes of the number of reactions of different importance by influence of inversion with probability 5%

The values of evolution parameters were given as following: probability of point mutation 10^{-7} , probability of nucleotide inversion 10^{-7} , probability of inversion 5%, probabilities of other mutation operators were 0%. Evolution process was executed up to 100,000 generations maintaining all vital reactions and a part of qualitative and insignificant reactions in the network. It should be noted that the number of qualitative and insignificant reactions, as well as genome concordance coefficient were changing all the time (continuously decreasing and then increasing) during evolution. It is a result of direct mutations when the whole gene sequence by the inversion influence changes and result in reversion or back mutations when gene reverts to its initial form.

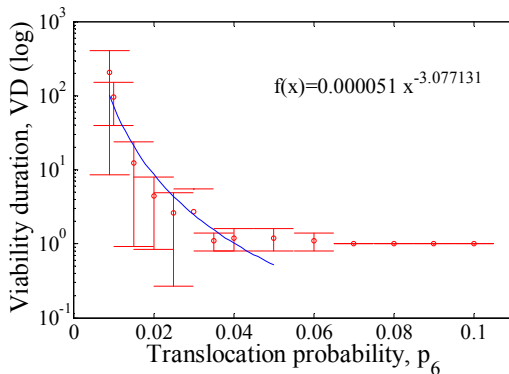


Figure 7. The viability duration of vital ZMO network model

In Figure 7 the average value of viability duration and its standard deviation for each set of 10 experiments are depicted. In total, 11 sets of experiments were executed with different values of translocation probability.

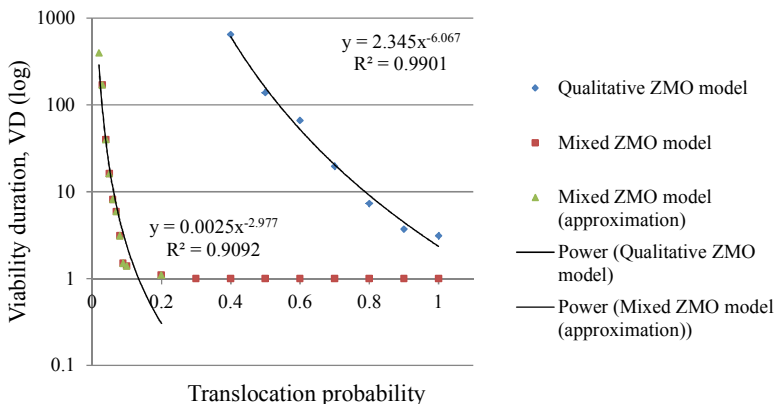


Figure 8. **The comparison of viability duration of ZMO models under influence of translocation**

Figure 8 shows the average value of viability duration for each set of 10 experiments of ZMO model. In total, 7 sets of experiments in case of qualitative ZMO model and 18 sets of experiments in case of mixed ZMO model with different values of translocation probability were executed.

The structure evolution of mixed network by the influence of several mutations is mostly affected by the translocations, deletions, inversions, and nucleotide inversions. Translocation and deletion have the most significant effect on the viability duration of the biochemical network structure and on the topological parameters of structure. Whereas, the viability duration of the network increases and the number of reactions increases exponentially by increasing probability of duplication mutation and analysing the influence of duplication both separately (see Figure 9) and together with other mutations, i.e. estimating the effect of several simultaneous mutations on the structure evolution of the network.

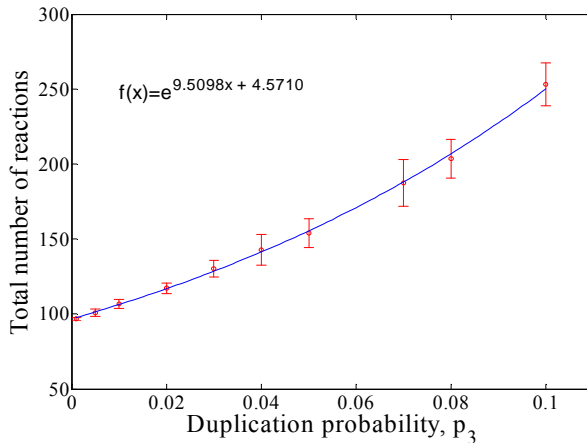


Figure 9. **The total number of reactions in ZMO network model with reactions of different importance**

Figure 9 shows the average viability duration and its standard deviation of each set of experiments. In total, 7 sets of 10 experiments with different probabilities of translocation were executed.

Analysis of the influence of several mutation types. For example, the average viability duration of the network structure is 18 generations with standard deviation 11 generations by the translocation probability 10% and probabilities of other mutations 1% (min viability duration is 4 generations, max viability duration is 47 generations). During evolution at least one vital reaction (100% of experiments) and all qualitative reactions (10% of experiments) are lost, and duplicates of reactions are maintained. The average concordance coefficient of a genome is 0.604 and standard deviation 0.149 when the evolution is interrupted. At the same time, the average viability duration of the network structure is 38 generations with standard deviation 20 generations by the influence of deletion with probability 10% and probabilities of other mutations 1% (min viability duration is 6 generations, max viability duration is 77 generations). During evolution at least one vital reaction (80% of experiments) and all qualitative reactions (20% of experiments) were lost. The average concordance coefficient of genomes is 0.92 and standard deviation 0.077 when the evolution is interrupted. But, when all mutation probabilities are 1% and inversion probability is increased to 10%, network structure survives more than 100 generations in 60% of cases maintaining one qualitative reaction and all vital reactions. In 4 of 10 experiments evolution was interrupted (min viability duration 24 generations, max viability duration 85 generations) when all qualitative reactions were lost with average concordance coefficient of genomes 0.526 and standard deviation 0.069. In its turn, network structure survives more than 100 generations in 80% of cases by the nucleotide

inversion influence with probability 10% (probabilities of all other mutation types are 1%) losing one vital reaction and all qualitative reactions and maintaining duplicates of reactions. In cases when evolution was interrupted (on the 31st generation and on the 71st generation) the average concordance coefficient of genomes was 0.349 with standard deviation 0.06.

By the influence of nucleotide inversion and inversion mutation the viability duration has a very high deviation, and inversion can have reverse effect when a gene returns to its initial form and reaction that is established by the corresponding gene is renewed. Therefore, topological parameters of the structure (see Figures 10, 11 and 12), as well as genome concordance coefficient change their values in a saltatory manner. Whereas nucleotide inversion and inversion, as well as translocation can have a reverse effect, then deletion is irreversible. If the deletion probability is changed from 1% to 10%, the viability duration of the network structure decreases rapidly, but the genome concordance coefficient keeps a high value when the evolution is interrupted.

Translocation, deletion and inversion has the most significant effect on the viability duration when the network structure of different importance is affected by several mutation types and the probability of the above-mentioned mutations is increased. In their turn, the genome concordance coefficient is affected mostly by the nucleotide inversion, inversion, and point mutation.

The number of isolated metabolites is affected mostly by the deletion, translocation, and nucleotide inversion. In their turn, its value can increase by the influence of point mutation, nucleotide inversion, and inversion, but the tendency of a decrease remains.

The average degree of a network is affected mostly by the deletion, point mutation, and nucleotide inversion. In their turn, its value can range many times during evolution by the influence of point mutation, nucleotide inversion, and inversion.

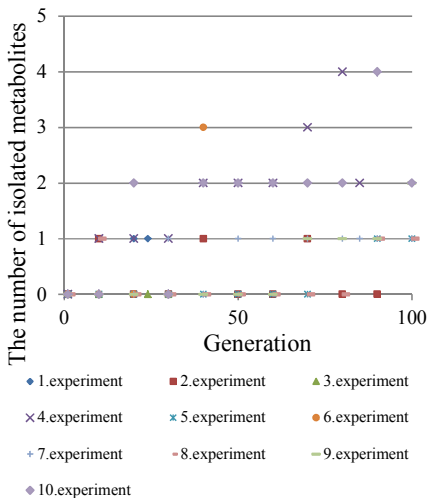


Figure 10. The changes of the number of isolated metabolites under influence of the inversion

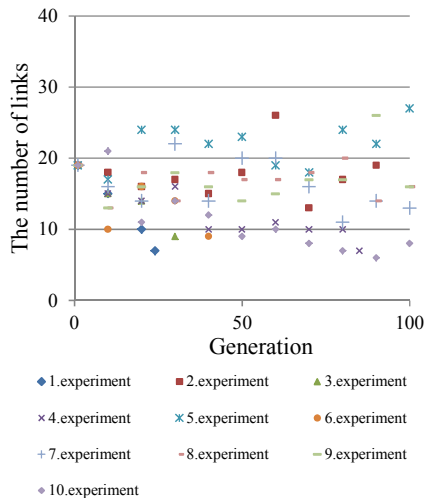


Figure 11. The changes of the number of links under influence of the inversion

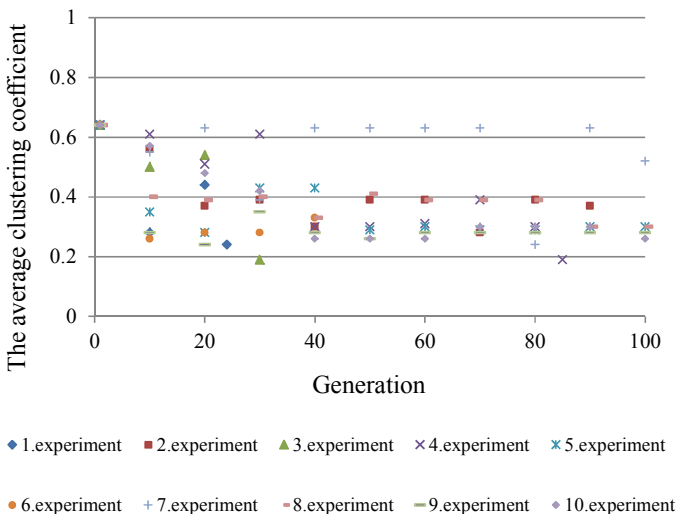


Figure 12. The changes of the number of average clustering coefficient under influence of the inversion

Figures 10, 11 and 12 display the results of 10 experiments, i.e. value changes of the corresponding topological parameter during evolution when the probability of inversion is 10%, but probabilities of all other mutations are 1%. In Figure 10 the number of isolated metabolites is shown, while Figure 11 depicts the number of links, while Figure 12 shows an average clustering coefficient.

The average number of neighbours is affected mostly by the deletion, translocation, and nucleotide inversion. In their turn, its value can increase and decrease two times during evolution by the influence of point mutation, nucleotide inversion, and inversion, especially by the influence of inversion, but the tendency of a decrease remains.

The number of links is affected mostly by the deletion, translocation, and nucleotide inversion. In their turn, its value can increase and decrease two times during evolution by the influence of point mutation, nucleotide inversion, and inversion, especially by the influence of point mutation, but the tendency of a decrease remains.

The average clustering coefficient is affected mostly by the deletion, translocation, and inversion. In their turn, its value can increase and decrease two and more times during evolution by the influence of point mutation, nucleotide inversion, and inversion, especially by the influence of inversion, but the tendency of a decrease remains.

CONCLUSIONS

The main PhD thesis results

The computer modelling approach was developed and the influence of different mutation types on the structure evolution of biochemical networks of non-coherent importance was determined by means of the developed modelling approach.

1) The methods of the structural analysis of biochemical networks and the methods of their representation were examined.

For visualisation, representation and modelling of biochemical networks the maps of biochemical networks, structural, stoichiometric and dynamic computer models are used in practice. The representation of a biochemical network based on certain standard, for example, *SBML*, *SBGN* standard, or based on formalism such as *Boolean* network, *Petri Nets* that are accepted in the field of systems biology.

In research of biochemical network structure two methods are generally applied: 1) structural or topological analysis that is based on the concepts and methods of graph theory and 2) clustering analysis.

2) The evolution process of biochemical networks and the factors influencing it were studied.

The evolution of biochemical networks is based on the molecular evolution of biological system which occurs in living organism at the level of underlying genome. The alterations in genes influence the interaction of genes, their products, and the interaction of processes which are subject to the influence of these genes.

3) The modelling approaches of structure evolution of biochemical networks were analysed.

Two classes of modelling approaches of biochemical network evolution dominate: duplication approach and deletion approach of network growth.

But the applied approaches for the studying of biochemical networks and their evolution have several shortcomings. The first shortcoming is disregard of essential properties of processes imitating evolution of biochemical network. One of such properties is process importance. The second shortcoming is justification of network structure changes by the appearance of mutations at the genome level without genome attachment to the network structure. The third shortcoming is remoteness from real process of evolution and introduction of only two or three mutation types in models – gene duplication and/or deletion that arises as node addition in network, and point mutation that theoretically explains occurrence of links rewiring.

4) The existing software tools designated for the modelling of biochemical networks and for the analysis of their structure were explored.

While researching, it was ascertained that during the last few years many software tools have been developed suited for the needs of both narrow and wide specialisation. Part of the modelling tools of biochemical networks are applied in modelling of separate types of networks, for example, in modelling of metabolic or gene regulation networks using some formalism or standard. No software tools were found which model evolution of to a genome attached structure of biochemical networks and take into consideration the process importance in evolution imitation.

5) The algorithm which imitates to a genome attached structure evolution of biochemical networks was developed.

The algorithm of structure evolution of biochemical networks enables modelling of structure changes of biochemical networks which are the consequences of genetic material changes that occur at the genome level by the influence of different types of mutations. The proposed algorithm of to a genome attached structure evolution includes implementation of point mutation, missense mutation, nonsense mutation, nucleotide inversion, duplication, deletion, inversion, and translocation. The algorithm of structure evolution of biochemical networks developed within the scope of the PhD thesis attaches genes to the biochemical network and takes into consideration different importance of processes.

6) A prototype of software tool *BINESA* was developed that implements the algorithm of to a genome attached structure evolution of biochemical networks developed during work on the PhD thesis.

The prototype of software tool *BINESA* allows also to analyse structure of biochemical network and to calculate different topological parameters, as well as to obtain data of the evolution process of different types and evolution results: a viability duration or a number of generation, a number of vital reactions, a number of qualitative reactions, a number of insignificant reactions, topological parameters of each analysed structure, i.e. a total number of nodes, a number of isolated nodes, a number of connected nodes, a total number of

reactions, a number of links, an average degree of network, an average incoming degree, an average outgoing degree, an average number of neighbours, an average clustering coefficient, a number of oriented cycles (loops), a number of self-loops, a distribution of degree, a distribution of incoming degree, a distribution of outgoing degree, a distribution of neighbours' number, a distribution of clustering coefficient.

7) The evolution experiments of to a genome attached structure of biochemical networks by the influence of several types of mutations were carried out, and the received evolution results were analysed.

Within the PhD thesis computer simulations of network structure evolution were carried out in case of biochemical networks which include only vital reactions, only qualitative reactions, and reactions of different importance. To establish the influence of each mutation type separately on the network evolution dynamics and on the changes in the peculiarities of topological measurements of a structure, several sets of experiments were carried out by different mutation probabilities of point mutation, nucleotide inversion, duplication, deletion, inversion, and translocation.

8) The topological parameters which characterise the applicability of the model of biochemical networks were determined.

In this study the topological parameters such as an incoming and outgoing degree distribution, a percentage of the low or high interconnectivity metabolites (low or highly interconnected), and an approximation of neighbour's number are recognised as informative structural parameters. A low agreement of the model pair resulting in a fragmented, poor quality model can be indicated by low values of an average degree, an average incoming degree, an average outgoing degree, and an average number of the neighbours compared to the initial models.

The structural analysis of the intersection and association of different models is important both in comparing evolution of models and building the new models on the basis of several available models. In intersection analysis when comparing the original model with the model received during or after the evolution, not only models but also their intersection can be investigated for the purpose of estimating which reactions have survived evolution and which were replaced by others.

Conclusions and development prospects

As a result of execution of computer modelling and simulations, several conclusions on evolution peculiarities of biochemical networks were drawn.

1. While comparing mixed biochemical networks and qualitative reaction networks a faster interruption of evolution process of mixed network was noticed. The number of the reached generations differs by 10-10,000 times both for network model of small size (14 reactions and 9 metabolites) and for model of real network of *Zymomonas mobilis* central metabolism of the bigger size

(96 reactions and 81 metabolites). Hence inclusion of attribute of different importance of reactions significantly influences the results of evolution simulations.

2. The duration of network structure evolution up to destruction is mostly affected by mutations of translocation, deletion, and inversion.

3. A typical observed tendency is an increase in a number of isolated metabolites, and reduction of average degree of metabolites, a number of links, as well as reduction of average neighbours' number and average clustering coefficient. The changes of parameters mainly occur linearly maintaining equal probabilities of mutations or increasing probability of point mutation, deletion or translocation. In addition to this increasing probability of nucleotide inversion or inversion, the values of topological parameters decrease, but their changes occur in a saltatory manner. Experiments with higher probability of deletion and translocation differ substantially bringing the network structure to accelerated destruction.

4. The viability of to a genome attached structure of biochemical network is better described by power law with negative exponent depending on the mutation probability by the influence of separate types of mutations both in case of networks of equal importance and networks of different importance.

5. The duration of experiments depends on the size of a model and the length of attached genes, as well as on the number of structures that should be saved in the course of evolution. In addition, the experiment duration slightly increases during evolution along with the serial number of generation calculating the average time that is consumed during one generation simulation. For example, to measure the influence of translocation with probability 50% on the qualitative *ZMO* network simulation of one generation, on average 10.75 seconds are required (execution of simulation with 10,000 requires 29 hours 51 minutes 40 seconds), while simulation of one generation in case of qualitative test network with the same simulation parameters requires 0.31 seconds (execution of simulation with 100,000 generations requires 8 hours 36 minutes 40 seconds) using the following hardware: computer processor *Intel Core2* 2GHz, 1Gb RAM and *Microsoft Access 2010* software package.

Several directions can be highlighted as **future development prospects**.

- 1) To introduce mutation operators which change the length of gene and a set of conditions that establish the formation of novel links in the algorithm of to a genome attached structure evolution of biochemical networks.
- 2) To study the evolution of network not only at the level of structure, but also at the level of biochemical network operation (function), and to estimate the changes of dynamics of biochemical network during evolution using network structure received or obtained in evolution.
- 3) To analyse the changes of network motifs, i.e. feedback and feed-forward loops, or the evolution of control loops by the influence of

different types of mutations, as well as the emergence regularities of alternative control loops.

BIBLIOGRAPHY

- Albert R., Barabasi A.-L. (2002) Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, Vol. 74, p.47-97.
- Albert I., Thakar J., Li S., Zhang R., Albert R. (2008) Boolean network simulations for life scientists. *Source code for Biology and Medicine*, Vol.3, p.1-16.
- Albert R., Jeong H., Barabasi A.-L. (2000) Error and attack tolerance of complex networks. *Nature*, Vol. 406, p.378-382.
- Aldana M., Balleza E., Kauffman S., Resendiz O. (2007) Robustness and evolvability in genetic regulatory networks. *Journal of Theoretical Biology*, Vol. 245, p.433-448.
- Arbabi-Ghahroudi M., Tanha J., MacKenzie R. (2005) Prokaryotic expression of antibodies. *Cancer Metastasis Rev*, Vol. 24, Issue 4, p.501-519.
- Assenov Y., Ramirez F., Schelhorn S.-H., Lengauer T., Albrecht M. (2008) Computing topological parameters of biological networks. *Bioinformatics Applications Note*, Vol. 24, No. 2, p.282-284.
- Baitaluk M., Sedova M., Ray N., Gupta A. (2006) Biological Networks: visualization and analysis tool for systems biology. *Nucleic Acids Research*, Vol. 34, D466- 471.
- Barabasi A.-L., Oltvai Z.N. (2004) Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, Vol. 5, p.101–113.
- Barabasi A.-L., Albert R. (1999) Emergence of scaling in random networks. *Science*, Vol. 286, p.509-512.
- Barrat A., Weigt M. (2000) On the properties of small-world network models. *The European Physical Journal B*, Vol. 13, p.547-560.
- Becker S.A., Feist A.M., Mo M.L., Hannum G., Palsson B.Ø., Herrgard M.J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. *Nature Protocols*, Vol.2, No.3, p.727-738.
- Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Wheeler D.L. (2008) GenBank. *Nucleic Acids Research*, January, p.25-30.
- Berg J., Lassing M., Wagner A. (2004) Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evolutionary Biology*, Vol. 4, No. 51, p.1-12.
- Bersini H., Lenaerts T., Santos F.C. (2005) Growing biological networks: Beyond the gene-duplication model. *Journal of Theoretical Biology*, Vol. 241, p.488-505.
- Boccaletti, S., Latora, V., Moreno, Y.,Chavez, M., Hwang, D.-U. (2006)

- Complex networks: Structure and dynamics. *Physics Reports*, Vol. 424, p.175-308.
- Brazhnik P., Fuente A., Mendes P. (2002) Gene networks: How to put the function in genomics. *Trends in Biotechnology*, Vol. 20, No. 11, p.467-472.
- Breitling R., Gilbert D., Heiner M., Orton R. (2008) A structured approach for the engineering of biochemical network models, illustrations for signalling pathways. *Briefings in Bioinformatics*, Vol. 9, No. 5, p.404-421.
- Browna C.T., Rustb A.G., Clarke P.J.C., Panb Z., Schilstrab M.J., De Buysschera T., Griffinb G., Wolda B.J., Camerona R.A., Davidsona E.H., Bolourib H. (2002) New Computational Approaches for Analysis of cis-Regulatory Networks. *Developmental Biology*, Vol. 246, Issue 1, p.86-102.
- Callaway D.S., Hopcroft J.E., Kleinberg J.M., Newman E.J., Strogatz S.H. (2001) Are randomly grown graphs really random? *Physical Review E*, Vol. 64 No.4, p.1-7.
- Cara A., Garg A., Micheli G., Xenarios I., Mendoza L., (2007) Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinformatics*, Vol. 8, p.462.
- Chauuiya C. (2007) Petri net modelling of biological networks. *Briefings in Bioinformatics*, Vol. 8, No. 4, p.210-219.
- Chaves M., Albert R., Sontag E.D. (2005) Robustness and fragility of Boolean models for genetic regulatory networks. *Journal of Theoretical Biology*, Vol. 235, p.431-449.
- Christogianni A., Douka E., KoukkouA.I., Hatziloukas E., Drainas C. (2005) Transcriptional Analysis of a Gene Cluster Involved in Glucose Tolerance in *Zymomonasmobilis*: Evidence for an Osmoregulated Promoter. *Journal of Bacteriology*, Vol. 187, No. 15, p.5179-5188.
- CellDesigner Tutorial. (2008) [Online] [viewed on October 24, 2009] Available at: <http://www.celldesigner.org/tutorial/CellDesignerTutorialICSB2008.pdf>
- Chen L., Wang R.-S., Zhang X.-S. (2009a) Introduction. Transcription regulation: networks and models. **In:** Chen L., Wang R.-S., Zhang X.-S. *Biomolecular networks: Methods and Applications in System Biology*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.1-45.
- Chen L., Wang R.-S., Zhang X.-S. (2009b) Signaling Networks: Modeling and Inference. **In:** Chen L., Wang R.-S., Zhang X.-S. *Biomolecular networks: Methods and Applications in System Biology*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.313-321.
- Chen L., Wang R.-S., Zhang X.-S. (2009c) Topological structure of Biomolecular Networks. **In:** Chen L., Wang R.-S., Zhang X.-S. *Biomolecular networks: Methods and Applications in System*

- Biology*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.169-204.
- Chen L., Wang R., Zhou T., Aihara K. (2005) Noise-induced cooperative behavior in a multi-cell system. *Bioinformatics*, Vol. 21, p.2722-2729.
- Clark D.P., Pazdernik N.J. (2012) *Molecular Biology: Understanding the Genetic Revolution*: 2nd eds. USA, Waltham, Elsevier, p.274-307.
- Cline M.S., Smoot M., Cerami E., Kuchinsky A., Landys N., Workman C., Christmas R. et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols*, Vol.2, p.2366 – 2382.
- Cohen R., Havlin S. (2003) Scale-free networks are ultra small. *Physical Review Letters*, Vol. 90, No. 5, p.1-4.
- Croft D., O’Kelly G., Wu G., Haw R., Gillespie M., Matthews L., Caudy M. et al. (2011), Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research*, Vol. 39, D691-D697.
- Conklin Lab. *GenMapp Concept*. [Online] [viewed on October 25, 2009] Available at: <http://www.genmapp.org/concept.html>
- Cutter A.D. (2010) *Molecular evolution inferences from the C.elegans genome*. WormBook, May 5, p.1-14. [Online]. [viewed on December 4, 2012]. Available at: http://wormbook.org/chapters/www_molecularevol/molecularevol.html
- Dahlquist K.D. (2004) Using GenMAPP and MAPPFinder to view microarray data on biological pathways and identify global trends in the data. *Current Protocols in Bioinformatics*. May, Chapter 7, Unit 7, p.1-5.
- Dahlquist K.D., Salomonis N., Vranizan K., Lawlor S.C., Conklin B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genetics*, Vol. 31, p.19-20.
- Dambītis J. (2002) *Modernā grafu teorija: mācību līdzeklis*. Rīga: Datorzinību centrs, 17.-30.lpp.
- DDBJ — DNA Data Bank of Japan. [Online]. [viewed on June 3, 2012]. Available at: <http://www.ddbj.nig.ac.jp/>
- Demin O.V., Plysnina T.Y., Lebedeva G.V., Zobova E.A., Metelkin E.A., Kolupaev A.G., Goryanin I.I., Tobin F. (2005) Kinetic modelling of the E.Coli. In: Alberghina L., Westerhoff H. (Eds.) *Systems Biology. Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.31-67.
- Doniger S.W., Salomonis N., Dahlquist K.D., Vranizan K., Lawlor S.C., Conklin B.R. (2003) MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology*, Vol.4, Issue 1, Art.R7, p.1-12.
- Drake J.W., Charlesworth B., Charlesworth D., Crow J.F. (1998) Rates of

- Spontaneous Mutation. *Genetics*, Vol. 148, No. 4, p.1667-1686.
- Duarte N.C., Becker S.A., Jamshidi N., Thiele I., Mo M.L., Vo T.D., Srivas R., Palsson B.O. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences on the USA*, Vol.104, No. 6, p.1777–1782.
- Duarte N.C., Herrgard M.J., Palsson B.O. (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, Vol. 14, p.1298–1309.
- ENA – *The European Nucleotide Archive*. [Online]. [viewed on June 3, 2012]. Available at: <http://www.ebi.ac.uk/ena/>
- EMBL — *European Molecular Biology Laboratory*. [Online]. [viewed on June 3, 2012]. Available at: <http://www.embl.de/>
- Farid N., Christensen K. (2006) Evolving networks through deletion and duplication. *New Journal of Physics*, Vol. 8, No. 9, p.1-17.
- Feist A.M., Henry C.S., Reed J.L., Krummenacker M., Joyce A.R., Karp P.D., Broadbelt L.J., Hatzimanikatis V., Palsson B. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, Vol. 3, No. 121, p.1-18.
- Fell D.A. (2005) Metabolic control Analysis. In: Alberghina L., Westerhoff H. (Eds.) *Systems Biology. Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.69-80.
- Fell D.A., Wagner A. (2000) The small world of metabolism. *Nature Biotechnology*, Vol. 18, p.1121-1122.
- Felsenstein J. (2013) Theoretical Evolutionary Genetics. Genome 562. USA: Washington, University of Washington, Department of Genome Sciences and Department of Biology, p.45-152.
- Fields S. (2001) Proteomics in genomeland. *Science*, Vol. 291, No. 5507, p.1221-1224.
- Finkel T., Gutking J.S. (2003) Signal Transduction and Human Disease. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.23-29.
- Förster J., Famili I., Fu P., Palsson B. Ø., Nielsen J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome research*, Vol. 13, Issue 2, p.244–253.
- Freeman W.H. et al. (2000) *Mechanism of DNA replication*. NCBI, Bookshelf ID NBK21862. [Online]. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK21862>
- Funahashi A., Matsuoka Y., Jouraku A., Morohashi M., Kikuchi N., Kitano H.(2008) CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proceedings of the IEEE*, Vol. 96, Issue 8, p.1254-1265.
- Funahashi, A., Tanimura, N., Morohashi, M., and Kitano, H. (2003)

- CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, Vol. 1, p.159-162.
- Fuhrer T., Fischer E., Sauer U. (2005) Experimental identification and quantification of glucose metabolism in seven bacterial species. *Journal of Bacteriology*, March, p.1581-1590.
- GenBank. [Online]. [viewed on July 1, 2012]. Available at: <http://www.ncbi.nlm.nih.gov/genbank/>
- Gibson T.A., Goldberg D.S. (2011) Improving evolutionary models of protein interaction networks. *Bioinformatics*, Vol. 27, No. 3, p.376-382.
- GML – Graph modelling language (2009) University of Passau. [Online] [viewed on December 20, 2009]. Available at: <http://www.infosun.fim.uni-passau.de/Graphlet/GML/>
- Gonzalez A., Naldi A., Sánchez L., Thieffry D., Chaouiya C., (2006) GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems*, Vol.84, p.91-100.
- Grassi L., Tramontano A. (2011) Horizontal and vertical growth of *S. Cerevisiae* metabolic network. *BMC Evolutionary Biology*, Vol. 11, No. 301, p.1-9.
- Grunde-Zeiferts U., Mozga I., Žukova T., Stalidzāns E. (2006) Therapy modelling combining methods of systems biology and automatic control theory. **No:** *International Scientific Conference “Animals. Health. Food Hygiene.”: proceedings*, November 11, 2006, Latvia, Jelgava, p.70-74.
- Grundspenķis J. *Sistēmu teorija un vadība*: ESF projekta ietvaros izveidots metodiskais materiāls. [Online] [viewed on August 5, 2010]. Available at: http://estudijas.itf.llu.lv/esf_materiali/default.aspx
- Grundspenķis J. *Sistēmu teorijas metodes:lekciju konspekts*. [Online]. [viewed on August 4, 2010]. Available at: stpk.cs.rtu.lv/read_write/file/materiali/stm/stm.ppt
- Guelzim N., Bottani S., Bourguin P., Kepes F. (2002) Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics*, Vol. 31, p.60-63.
- Han J.-D. J. (2008) Understanding biological functions through molecular networks. *Cell Research*, Vol. 18, p.224-237.
- Han J.D., Bertin N., Hao T., Goldberg D.S., Berriz G.F., Zhang L.V., Dupuy D., Walhout A. J., Cusick M.E., Roth F.P., Vidal M. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, Vol. 430, p.88-93.
- Hallinan J.S., Jackway P.T. (2005) Network Motifs, Feedback Loops and The Dynamic of Genetic Regulatory Networks. **In:** *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*: proceedings, 2005, IEEE Press, p.90-96.
- Hallinan, J. Bradley, D. & Wiles, J. (2006). Effects of constitutive gene

- activation on the dynamics of genetic regulatory networks. **In:** *IEEE World Congress on Computational Intelligence*: proceedings, BC Canada, Vancouver, July 16-21, 2006, p.1-7.
- Hartl D.L., Clark A.G. (1997) Introduction. **In:** Hartl D.L., Clark A.G. *Principles of Population Genetics*: 3rd ed. USA: Sunderland, Sinauer Associates, p.1-11.
- Hartl D.L., Jones E.W. (2001a) Introduction to Molecular Genetics and Genomics **In:** Hartl D.L., Jones E.W. *Genetics: Analysis of Genes and Genomes*: 5th ed. Canada: Jones and Bartlett Publishers, p.1-35.
- Hartl D.L., Jones E.W. (2001b) DNA Structure and DNA Manipulation. **In:** Hartl D.L., Jones E.W. *Genetics: Analysis of Genes and Genomes*: 5th ed. Canada: Jones and Bartlett Publishers, p.36-49.
- Hase T., Niimura Y. (2012) *Protein-Protein Interaction Networks: Structures, Evolution, and Application to Drug Design*. Protein-Protein Interactions – Computational and Experimental Tools, p.405-426. [Online]. [viewed on June 25, 2013] Available at: http://bioinfo.tmd.ac.jp/~niimura/Hase&Niimura_2012.pdf
- Hase T., Niimura Y., Tanaka H. (2010) Difference in gene duplicability may explain the difference in overall structure of rotein-protein interaction networks among eukaryotes. *BMC Evolutionary Biology*, Vol. 10, No. 358, p.1-15.
- Hase T., Tanaka H., Suzuki Y., Nakagawa S., Kitano H. (2009) Structure of Protein Interaction Networks and Their Implications on Drug Design. *PLoS Computational Biology*, Vol. 5, Issue 10, p.1-9.
- Helikar T., Kochi N., Konvalina J., Rogers A. (2011) Boolean Modeling of Biochemical Networks. *The Open Bioinformatics Journal*, Vol.5, p.16-25.
- Helikar T., Rogers J. (2009) ChemChains: a platform for simulation and analysis of biochemical networks aimed to laboratory scientists. *BMC Systems Biology*, Vol. 3, p.58.
- Herbert M. (2004) *Reference and Tutorial Manual*. An Introduction to Biochemical Modeling using Jdesigner. Assisted by Abhishek Agrawal and Brian Gates. (Revision 1, Oct 10th, 2004) [Online]. [viewed on October 2, 2009]. Available at: <http://sbw.kgi.edu/software/jdesigner.htm>
- Herrgard M., Lee B.S., Portnoy V., Palsson B.O. (2006) Integrated analyses of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Research*, Vol. 16, No.5, p.627-635.
- Himsolt M. *GML: A portable Graph File Format*. [Online]. [viewed on November 11, 2010]. Available at: <http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/ brandenburg/projekte/gml/gml-technical-report.pdf>
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M.,

- Xu, L., Mendes, P., and Kummer, U. (2006) COPASI — a Complex Pathway Simulator. *Bioinformatics*, Vol. 22, p.3067-3074.
- Hu Z., Mellor J., Wu J., Yamada T., Holloway D., DeLisi C. (2005) VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Research*, Vol. 33, p.352-357.
- Hu Z., David M. Ng, Yamada T., Chen C., Kawashima S., Mellor J., Linghu B., Kanehisa M., Stuart J.M., DeLisi C. (2007) VisANT 3.0: new modules for pathway visualization, editing, prediction and construction. *Nucleic Acids Research*, Vol.35, p.625-632.
- Hu Z., Hung J.-H., Wang Y., Chang Y.-C., Huang C.-L., Huyck M., DeLisi C. (2009) VisANT 3.5 multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Research*, Vol. 37, p.115.-121.
- Hucka M., Hoops S., Keating S.M., Le Novère N., Sahle S., Wilkinson D.J.. (2008) Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions. SBML Level 2 Version 4 specification, Release 1. *Nature Proceedings*, p.1-166.
- Hucka M., Finney A., Sauro H. M., Bolouri H., Doyle J. C., Kitano H. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, Vol. 19, No. 4, p.524–531.
- National Human Genom Research Institute (2007) *A Guide to Your Genome*. [Online]. [viewed on July 12, 2013]. Available at: <http://www.genome.gov/Pages/Education/AllAbouttheHumanGenomeProject/GuidetoYourGenome07.pdf>
- Jana S., Deb J.K. (2005) Strategies for efficient production of heterologous proteins in Escherichia coli. *Application in Microbiology and Biotechnology*, Vol. 67, Issue 3, p.289-298.
- Jarboe L.R., Zhang X., Wang X., Moore J.C., Shanmugam K.T., Ingram L.O. (2010) Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. *Journal of Biomedicine and Biotechnology*, p.1-18.
- Jong H., Geiselmann J., Hernandez C., Page M. (2003) Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, Vol. 19, Issue 3, p.336-344.
- Jeong H., Tombor B., Albert R., Oltvai Z.N., Barabasi A.-L. (2000) The large-scale organization of metabolic networks. *Nature*, Vol. 407, p.651-654.
- Jeong H., Mason S.P., Barabasi A.-L., Oltvai Z.N., (2001) Lethality and centrality in protein networks. *Nature*, Vol. 411, p.41-42.
- Kanehisa M., Goto S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, Vol. 28, p.27-30.
- Kanehisa M., Goto S., Hattori M., Aoki-Kinoshita K.F., Itoh M.,

- Kawashima S., Katayama T., Araki M., Hirakawa M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, Vol. 34, p.354-357.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. (2011) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, Vol. 40, D109–114.
- Karp, P. D., Ouzounis, C. A, Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., Tsoka, S. et al. (2005). Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research*, Vol. 33, Issue 19, p.6083–6089.
- Kauffman S. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, Vol. 22, p.437-467.
- Keating S. M., Bornstein B. J., Finney A., Hucka M. (2006) SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics*, Vol. 22, No. 10, p.1275–1277.
- KEGG Gene Database. [Online]. [viewed on May 21, 2013]. Available at: <http://www.genome.jp/kegg/genes.html>
- Keseler I. M., Collado-Vides J., Santos-Zavaleta A., Peralta-Gil M., Gama-Castro S., Muñiz-Rascado L., Bonavides-Martinez C. et al. (2011). EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Research*, Vol. 39, D583–590.
- Kholodenko B.N., Bruggeman F.J., Sauro H.M. (2005) Mechanistic and modular approaches to modeling and inference of cellular regulatory networks. **In:** Alberghina L., Westerhoff H.V. (eds.) *Systems Biology: Definitions and Perspectives*. London: Springer, p.143-162.
- Kim J.R., Yoon Y., Cho K.H. (2008) Coupled Feedback Loops Form Dynamic Motifs of Cellular Networks. *Biophysical Journal*, Vol. 94, No.2, p.359–365.
- Kim W.K., Marcotte E.M. (2008) Age-dependent evolution of the yeast protein interaction network suggests a limited role of gene duplication and divergence. *PloS Computational Biology*, Vol. 4, Issue 11, p.1-10.
- Kimura M. (1984) *The Neutral Theory of Molecular Evolution*. UK: Cambridge, Cambridge University Press. 367 p.
- Kitano H. (2007) The theory of biological robustness and its implication in cancer. **In:** *Ernst Schering Foundation Symposium: proceedings*, Vol. 61, p.69-88.
- Kitano H. (2004) Biological robustness. *Nature Reviews*, Vol. 5, p.826-837.
- Kitano H. (2005) Scientific and technical challenges for systems biology. **In:** Alberghina L., Westerhoff H. (Eds.) *Systems Biology. Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag,

p.373-385.

- Klamt S., Saez-Rodriguez J., Lundquist J., Simeoni L., Gilles E. (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, Vol. 7. 56 p.
- Klamt S., Saez-Rodriguez J., Gilles E. D. (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, Vol. 1. [Online]. [viewed on October 2, 2009]. Available at: <http://dx.doi.org/10.1186/1752-0509-1.2>
- Klipp E., Herwig R., Kowald A., Wierling C., Lehrach H. (2005) *Systems biology in practice. Concept, Implementation and Application*. Weinheim: WILEY-VCH Verlag GmbH & Co, KgaA. 486 p.
- Koschutski D., Junker B.H., Schwender J., Schreiber F. (2010) Structural analysis of metabolic networks based on flux centrality. *Journal of Theoretical biology*, Vol. 265, p.261-269.
- Kostromins A., Stalidzans E. (2012) Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism. *Biosystems*, Vol. 109, Issue 2, p.233-239.
- Krapivsky P.L., Redner S., Leyvraz F. (2000) Connectivity of Growing Random Networks. *Physical Review Letters*, Vol. 85, Issue 21, p.4629-4632.
- Krishna S., Semsey S., Sneppen K. (2007) Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proceedings of the National Academy of Sciences*, Vol. 104, Nr.52, p.20815-20819.
- Kuepfer L., Sauer U., Blank L. M. (2005). Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Research*, Vol. 15, Issue 10, p.1421-1430.
- Kwon Y.K., Cho K.H. (2007) Analysis of feedback loops and robustness in network evolution based on Boolean models. *BMC Bioinformatics*, Vol.8, 9 p.
- Le Novère N., Hucka M., Mi H., Moodie S., Schreiber F., Sorokin A., Demir E., Wegner K., Aladjem M.I., Wimalaratne S.M., Bergman F.T. et al. (2009) The Systems Biology Graphical Notation. *Nature Biotechnology*, Vol. 27, No. 8, p.735-741.
- Lee K.Y., Park J.M., Kim T.Y., Yun H., Lee S.Y. (2010) The genome-scale metabolic network analysis of *Zymomonas mobilis* ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. *Microbial Cell Factories*, Vol. 9, p.1-12.
- Lee H., Popodi E., Tang H., Foster P.L. (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceeding of the National Academy of Sciences in USA*, Vol. 109, No. 41, p.E2774-E2783.
- Lewin R. (1996) *Patterns in evolution: the new molecular view*. USA:

- New York, Scientific American library. 246 p.
- Li D., Li J., Quyang S., Wang J., Wu S., Wan P., Zhu Y. et al. (2006) Protein interaction networks of *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster*: Large-scale organization and robustness. *Proteomics*, Vol. 6, p.456-461.
- Lodish H., Berk A., Matsudaira P., Kaiser C.A., Krieger M., Scott M.P., Zipursky L., Darnell J. (2004) *Molecular Cell Biology*: 5th. USA: New York, Eds. W. H. Freeman, p.101-125.
- Longabaugh W., Davidson E., Bolouri H. (2009) Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. *Biochim. Biophys. Acta*, Vol. 1789, Issue 4, p.363-374.
- Longabaugh W., Davidson E., Bolouri H. (2005) Computational representation of developmental genetic regulatory networks. *Developmental Biology*, Vol. 283, p.1-16.
- Lucock M. (2007) *Molecular Nutrition and Genomics*. USA: New Jersey, John Wiley & Sons, Inc., Hoboken, p.1-9.
- Lynch M. (2007) The evolution of genetic networks by nonadaptive processes. *Nature Reviews Genetics*, Vol. 8, p.803-813.
- Madera S. (1998a) *Bioloģija 1.daļa*. Rīga: Zvaigzne ABC. 298 lpp.
- Madera S. (1998b) *Bioloģija 2.daļa*. Rīga: Zvaigzne ABC. 372 lpp.
- Massanori A. (2004) The metabolic world of *Escherichia coli* is not small. *Proceeding of National Academy of Sciences in USA*, Vol. 101, No. 6, p.1543-1547.
- Mazein A. *User Manual EPE3.0.0-alpha6*. [Online]. [viewed on October 24, 2009]. Available at: [http://garr.dl.sourceforge.net/project/epe/EPE/Documentation/Manual EPE3 .0.0 —alpha6.pdf](http://garr.dl.sourceforge.net/project/epe/EPE/Documentation/Manual%20EPE3%20.0.0%20—alpha6.pdf)
- Mednis M., Aurich M.K. (2012) Application of string similarity ratio and edit distance in automatic metabolite reconciliation comparing reconstructions and models. *Biosystems and Information Technology*, Vol. 1, Issue 1, p.14-18.
- Mednis M., Brusbardis V., Galvanauskas V. (2012) Comparison of genome-scale reconstructions using ModeRator. **In:** *13th IEEE International Symposium on Computational Intelligence and Informatics*: proceedings, Hungary, Budapest, p.79–84.
- Mendes P. (1993) GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput. Appl. Biosci.*, Oct, 1993, Vol. 9, Issue 5, p.563-571.
- Mensonides F., Schuurmans J., Mattos M., Hellingwerf K., Brul S. (2002) The Metabolic Response of *Saccharomyces Cerevisiae* to Continuous Heat Stress, *Molecular Biology Reports*, Vol.1, p.103-106.
- Milo R., Shen-Orr S., Itzkovitz S., Kashtan N., Chklovskii D., Alon U. (2002) Network motifs: Simple Building Blocks of Complex Networks. *Science*, Vol. 298, No. 5594, p.824-827.

- Moran N.A., McLaughlin H.J., Sorek R. (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*, Vol. 323, No. 5912, p.379-382.
- Myers C. J. (2010) *Engineering Genetics Circuits*. USA: New York, Chapman & HALL/CRC Mathematical and computational Biology Series. Taylor and Francis Group, LLC, p.1-278.
- Müssel C., Hopfensitz M., Kestler H. (2010) BoolNet—an R package for generation, reconstruction and analysis of Boolean networks, *Bioinformatics*, Vol. 26, p.1378-1380.
- Nachman M.W., Crowell S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics*, Vol. 156, Issue 1, p.297-304.
- Natal A.W., Riel V. (2006) Dynamic modeling and analysis of biochemical networks: mechanism-based models and model-based experiments. Briefings in *Bioinformatics*, Vol. 7, No. 4, p.364-374.
- NCBI – The National Center for Biotechnological Information (2010) *What is a genome?* [Online]. [viewed on February 4, 2010]. Available at: http://www.ncbi.nlm.nih.gov/About/primer/genetics_genome.html
- NetBuilders Concept* (2006) [Online]. [viewed on October 2, 2009]. Available at: <http://strc.herts.ac.uk/bio/maria/Apostrophe/Pdf/NetBuilder-prime.pdf>
- Newman M.E.J. (2003) The structure and function of complex networks. *SIAM Review*, Vol. 45, p.167-256.
- Nishio Y., Usuda Y., Matsui K., Kurata H. (2008) Computer-aided rational design of the phosphotransferase system for enhanced glucose uptake in *Escherichia coli*. *Molecular Systems Biology*, Vol. 4, p.1-12.
- Noort V., Snel B., Huynen M.A. (2004) The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Reports*, Vol. 5, p.280-284.
- Novak B., Chen K.C., Tyson J.J. (2005) Systems biology of the yeast cell cycle engine. In: Alberghina L., Westerhoff H. (Eds.) *Systems Biology. Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.305-323.
- Osis J.J. (1969) *Automātiskā vadība un regulēšana*. Rīga: Zvaigzne, 1.-16.lpp.
- Ossowski S., Schneeberger K., Lucas-Lledo J., Warthmann N., Clark R.M., Shaw R.G., Weigel D., Lynch M. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, Vol. 327, p.92–94.
- Palsson, B. Ø. (2006) *Systems Biology: Properties of reconstructed networks*. UK: Cambridge, Cambridge University Press. 334 p.
- Pastor-Satorras R., Smith E., Solé R.V. (2003) Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*,

- Vol. 222, p.199-210.
- Paszek E. (2007) *Boolean networks*. Connexions module: m12394. [Online]. [viewed on December 21, 2010]. Available at: http://cnx.org/content/m12394/1.5/_content_info
- Patil A., Nakamura H. (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Letters*, Vol. 580, p.2041–2045.
- Paxson R., Zannella K. (2007) System biology: Studying the world's most complex dynamic systems. *Journal The Math Works News & Notes*, June, p.4-7.
- Philips N. Salomon M., Custer A., Ostrow D., Baer C.F. (2009) Spontaneous Mutational and Standing Genetic (Co)variation at Dinucleotide Microsatellites in *Caenorhanditis briggsae* and *Caenorhabditis elegans*. *Molecular Biology and Evolution*, Vol. 26, No.3, p.659-669.
- Pentjuss A., Odzina I., Kostromins A., Fell D.A., Stalidzans E., Kalnenieks U. (2013) Biotechnological potential of respiring *Zymomonas mobilis*: A stoichiometric analysis of its central metabolism. *Journal of Biotechnology*, Vol.165, p.1-10.
- Qin H., Lu H.H.S., Wu W.B., Li W.-H. (2003) Evolution of the yeast protein interaction networks. *Proceedings of the National Academy of Sciences in USA*, Vol. 100, No. 22, p.12820-12824.
- Ravasz E., Somera A.L., Mongru D.A., Oltvai Z.N., Barabasi A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, Vol. 297, p.1551-1555.
- Raipulis J. (2002) *Ģenētikas pamati: tālmācības līdzeklis*. Rīga: Izdevniecība RaKa, 9.-157.lpp.
- Reil, T. (1999) Dynamics of gene expression in an artificial genome: Implications for biological and artificial ontogeny. **In:** *5th European Conference on Artificial Life*: Floreano, D., Mondada, F. & Nicoud, J. D. (eds.) proceedings, Berlin: Springer Verlag, p.457–466.
- Rodrigues J.F.M., Wagner A. (2011) Genotype networks, innovation, and robustness in sulfur metabolism. *BMC Systems Biology*, Vol. 5, Issue 39, p.1-13.
- Roy S., Lane T., Werner-Washburne M. (2007) A Simulation Framework for Modeling Combinatorial Control in Transcription Regulatory Networks. *UNM Computer Science Technical Report*, TR-CS-2007-06, p.1-10.
- Rubina T. (2013) The procedure of evolution modelling of biochemical networks structure. *Biosystems and Information Technology*, Vol.2, No.2, p.19-25.
- Rubina T., Stalidzans E. (2010a) Topological features and parameters of biochemical network structure. **In:** *8th Industrial Simulation Conference*: publication of EUROSIS, proceedings, June 7-9, 2010,

- Hungary, Budapest, p.228-236.
- Rubina, T., & Stalidzans, E. (2010b) Software Tools for Structure Analysis of Biochemical Networks. **In:** *4th International Scientific Conference "Applied Information and Communication Technologies"*: proceedings, April 22-23, 2010, Latvia, Jelgava, p.33-49.
- Rubina T., Stalidzans E. (2012a) Evolution modeling algorithm of biochemical networks. **In:** *10th Industrial Simulation Conference: a publication of EUROSIS*, proceedings, June 4-6, 2012, Czech Republic, Brno, p.24-30.
- Rubina T., Stalidzans E. (2012b) Evolution of control loops of biological systems. **In:** *5th International Scientific Conference "Applied information and Communication Technologies"*: proceedings, April 26-27, Latvia, Jelgava, p.317-324.
- Rubina, T. (2012) Tools for analysis of biochemical network topology. *Biosystems and Information Technology*, Vol.1, No.1, p.25-31.
- Robins G, Pattison P., Koskinen J. (2008) *Network degree distribution. Technical report to Australian Defence Science and Technology Organisation*. University of Melbourne, p.1-9.
- Rubina T., Brusbardis V. (2009) Applications of biochemical networks discovering control mechanisms in systems biology. In: *Annuals Students International Scientific Conference "Youth in Science and Profession Practice"*: proceedings, April 23, 2009, Latvia, Jelgava, p.1-7.
- Sausiņa L. (2010) *Bioloģija vidusskolai 4. Daļa*. Rīga: Zvaigzne ABC, 208 lpp.
- Schellenberger J., Park J. O., Conrad T. M., Palsson B. Ø. (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, Vol. 11, Issue 1, p.213.
- Schellenberger J., Que R., Fleming R. M. T., Thiele I., Orth J. D., Feist A. M., Zielinski D. C. et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, Vol. 6, Issue 9, p.1290–1307.
- Schilstra M.J., Bolouri H. (2002) The logic of gene regulation. **In:** *3rd International Conference on Systems Biology*: poster abstract, 2002.
- Shannon P., Markiel A., Ozier O., Baliga NS., Wang JT., Ramage D., Amin N., Schwikowski B., Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, Vol. 13, No. 11, p.2498-2504.
- Sharan R., Ideker T. (2006) Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, Vol. 24, No. 4, p.427-433.

- Selga T. (2008) *Šūnu bioloģija*. Rīga: LU Akadēmiskais apgāds. 343 lpp.
- Seo J.S., Chong H., Park H.S., Yoon K.O., Kim J.J., Hong J.H., Kim H., Kim J.H. Kil J.I., Park C.J., Oh H.M., Lee J.S., Jin S.J., Um H.W. et al. (2005) The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. *Nature Biotechnology*, Vol. 23, No.1, p.63-68.
- Snoep J., Westerhoff H. (2005) From isolation to integration, a systems biology approach for building the Silicon Cell. In: Westerhoff H.V. Alberghina L., (eds.) *Systems Biology: Definitions and Perspectives*. Berlin: Weidelberg, Springer Verlag, p.13-30.
- Solē V.R., Pastor-Satorras R., Smith E., Kepler T.B. (2008) A model of large-scale proteome evolution. *WSPC/Guidelines, Advances in Complex Systems*, p.1-12.
- Sole R.V., Pastor-Satorras R., Smith E., Kepler T. (2002) A model of large-scale proteome evolution. *Advances in Complex Systems*, Vol. 5, p.43-54.
- Sontag E., Vilz-Cuba A., Laubenbacher R., Jarrah A.S. (2008) The Effect of Negative Feedback Loops on the Dynamics of Boolean Networks. *Biophysical Journal*, Vol. 95, p.528-526.
- Sorensen H.P., Mortensen K.K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*, Vol. 115, Issue 2, p.113-128.
- Strazewski P., Tamm C. (1990) Replication Experiments with Nucleotide Base Analogues. *Angewandte Chemie*, Vol. 29, Issue 1, p.36-57.
- Straziņš I. (2001) Diskrētā matemātika:mācību grāmata programmēšanas, tehnisko un ekonomikas specilitāšu studentiem. Rīga: Zvaigzne ABC, 101.-107. lpp.
- Suderman M., Hallett M. (2006) Tools for visually exploring biological networks. *Bioinformatics*, Vol. 23, Issue 20, p.2651-2659.
- Swings J., De Ley J. (1977) The biology of *Zymomonas*. *Bacteriological Reviews*, Vol. 41, p.1-46.
- Thiele I., Palsson B. Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, Vol. 5, Issue 1, p.93-121.
- Thiele I, Swainston N., Fleming R., Hoppe A., Sahoo S. et al. (2013) A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, Vol. 31, p.419-425.
- Vazquez A. (2003) Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, Vol. 67, Art.No. 056104, p.1-15.
- Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O. et al. (2001) The Sequence of the Human Genome. *Science*, Vol. 291, No.5507, p.1304-1351.
- Volkers R. (2004) *Gēni un DNS*. / Walker R. (2004) *Genes and DNA*.

- Stīva Džonsaw priekšv. [tulk. Frīdenberga A.] Rīga: Eve, 63 lpp.
- Wagner A. (2011) The molecular origins of evolutionary innovations. *Trends in Genetics*, Vol.27, p.397-410.
- Wagner A. (2009) Molecular evolution, Networks in. **In:** Meyers R.A. (ed.) *Encyclopedia of Complexity and System Science*. London: Springer, p.1-21.
- Wagner A. (2005) *Robustness and Evolvability in Living Systems*. USA: New Jersey, Princeton University Press, p.15-38.
- Wagner A., Fell D. (2001) The small world inside large metabolic networks. *Proceedings of the Royal Society B*, Vol. 268, No. 1478, p.1803-1810.
- Wagner A. (2003) How the global structure of protein interaction networks evolves. *Proceedings of the Royal Society B*, Vol. 270, p.457-466.
- Walsh B. (2003) Population-genetics models of the fates of duplicate genes. *Genetica*, Vol. 118, p.279-294.
- Watson J., Wiles J., Hanan J. (2003) Towards more Relevant Evolutionary Models: Integrating an Artificial Genome with a Developmental Phenotype. **In:** *1st Australian Conference on Artificial Life*: Abbass H., Wiles J. (eds.) proceedings, 2003, Australia, Canberra, p.288-298.
- Watson J.D., Crick F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, Vol. 171, Art. No. 4356, p.737-738.
- Watts D.J., Strogatz S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, Vol. 393, p.440-442.
- Weir B.S. (1996) Genetic data analysis: 2nd ed. Sinauer Associates, p.8-15.
- Wegner K., Knabe J., Robinson M., Egri-Nagy A., Schilstra M., Nehaniv C. (2007) The NetBuilder' project: development of a tool for constructing, simulating, evolving, and analysing complex regulatory networks. *BMC Systems Biology*, P72, 2 p.
- Westerhoff H., Hofmeyr J-H. (2005) What is system biology? From genes to functions and back. **In:** Alberghina L., Westerhoff H.V. (eds.) *Systems Biology: Definitions and Perspectives*. London: Springer, p.119-141.
- Whitaker J. W., Letunic I., McConkey G. A., Westhead D. R. (2009). metaTIGER: a metabolic evolution resource. *Nucleic Acids Research*, Vol. 37, D531-D538.
- Wilson R.J. (1972) *Introduction to Graph Theory*. New York: Oliver and Boyd, Edinburgh Academic Press, p.9-21.
- Wolfe K.H., Shields D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, Vol. 387, p.708-713.
- Wuensche A. (2009) Discrete dynamics lab: tools for investigating cellular automata and discrete dynamical networks. Artificial life models in

- software. **In:** Adamatzky A, Kosinski M, Eds.: London: Springer, p.215-258.
- Yamada T., Bork P. (2009) Evolution of biomolecular networks – lessons from metabolic and protein interactions. *Nature Reviews, Molecular Cell Biology*, Vol. 10, p.791-803.
- Yamada T., Sugiyama T., Tamaki N., Kawakita A., Kato M. (2009) Adaptive radiation of gobies in the interstitial habitats of gravel beaches accompanied by body elongation and excessive vertebral segmentation. *BMC Evolutionary Biology*, Vol. 9, Issue 145, p.1.14.
- Yang S., Pan C., Tschaplinski T.J., Hurst G.B., Engle N.L., Zhou W., Dam P., Xu Y., Jr M.R., Dice L., Johnson C.M., Davison B.H., Brown A.D. (2013) Systems Biology Analysis of *Zymomonas mobilis* ZM4 Ethanol Stress Responses. *PloS One*, Vol. 8, Issue 7, p.1-14.
- Yook S.-H., Oltvai Z.N., Barabasi A.-L. (2004) Functional and topological characterization of protein interaction networks. *Proteomics*, Vol. 4, p.928-942.
- Zhang A. (2009) *Protein Interaction Networks: Computational analysis*. Cambridge: Cambridge University Press, p.1-20.
- Zhang J., Shakhnovich E.-I. (2008) Sensitivity-dependent model of protein-protein interaction networks. *Physical Biology*, Vol. 5, p.1-6.
- Zhao J., Ding G.-H., Tai L., Yu H., Yu Z.-H., Luo J.-H., Cao Z.-W., Li Y.-X. (2007) Modular co-evolution of metabolic networks. *BMC Bioinformatics*, Vol. 8, No. 311, p.1-12.
- Zhao J., Yu H., Luo J., Cao Z.W., Li Y. (2006) Complex networks theory for analysing metabolic networks. *Chinese Science Bulletin*, Vol. 51, No.13, p.1529-1537.
- Zheng J., Zhang D., Przytycki P., Zielinski R., Capala J., Przytycka T. (2010) SimBoolNet—a Cytoscape plugin for dynamic simulation of signaling networks. *Bioinformatics*, Vol. 26, p.141-142.
- Zinovyev A., Viara E., Calzone L., Barillot E. (2008) BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics Applications Note*, Vol. 24, No. 6, p.876-877.
- Zinovyev A., Calzone L. *Binom Manual Version 1.0*. Institut Curie, Service de Bioinformatique. [Online]. [viewed on January 7, 2010]. Available at: http://bioinfo-out.curie.fr/projects/binom/docs/Binom_Manual_v1.0.pdf
- Бакай А.В., Кочиш И.И., Скрипниченко Г.Г. (2006с) **В кн.:** Бакай А.В., Кочиш И.И., Скрипниченко Г.Г. *Генетика: учебник для студентов высших учебных заведений*. Москва: КолосС, 448 с.
- Давидич М.И., Постников Е.Б. (2007) Булевская модель цикла деления клетки дрожжей *Schizosaccharomyces pombe*:

динамика в случае нормальных и возмущенных начальных условий. В м.к.: *2-ая Международная Конференция “Математическая Биология и биоинформатика”*: материалы конференции, том 2, 2007, с.377-386.

USTUA (2002) Термодинамика живых систем. Жизнь как информационный процесс: Лекция из курса Концепции современного естествознания. (Concepts of modern natural sciences. Lecture: Thermodynamics of living systems. Life as information process.). The Ufa state technical university of aviation (USTUA). [Online]. [viewed on December 5, 2011]. Available at: <http://www.ugatu.ac.ru/ddo/KSE/01/0118/ks011800.htm>