

CALCULUL BIOLOGIC – O PARADIGMĂ A INFORMATICII CONTEMPORANE

Dr. Constantin CIUBOTARU

Dr. hab. Svetlana COJOCARU

Dr. Galina MAGARIU

Dr. hab. Iurie ROGOJIN

Institutul de Matematică și Informatică, AȘM

BIOCOMPUTING

Biocomputing (also known as molecular or cell computing) is a new interdisciplinary area of science where computer science, chemistry, biology, and physics meet, that promises revolutionary changes in computations in the near future. It aims at developing new devices and computing systems to manipulate information operating at atomic or molecular scale on the basis of biological, chemical, electronic, photonic and/or mechanical principles. In the paper a short overview of biocomputing area and results obtained in the frame of STCU project 4032 "Power and efficiency of natural computing: neural-like P (membrane) systems" (project manager: Prof. Iurie Rogojin; coordinating institution: Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova) are presented. In particular new computational possibilities provided by computing with biomolecules in the following biocomputing models are investigated: Networks of Evolutionary Processors (NEPs) and Hybrid Networks of Evolutionary Processors (HNEPs), Communication P Systems, Transitional P Systems, P Systems with Active Membranes, Polymorphic P Systems, Insertion-Deletion Systems, Insertion-Deletion P Systems with Small Parameters, and Splicing P Systems. Promising results of application of biocomputing methods and results in computer algebra and computer linguistics are presented too.

Una dintre marile provocări ale informaticii moderne este ideea de a descrie lumea din jurul nostru în limbajul procesării informației. O modalitate de a face aceasta este studierea proceselor de calcul care au loc în natură.

Calculul natural reprezintă puntea de trecere dintre științele naturale și informaționale. Astăzi, majoritatea paradigmelor, principiilor și mecanismelor utilizate în calculele proiectate de o ființă umană pot fi urmărite în cadrul sistemelor naturale ca sursă de inspirație. În timp ce unele domenii ale calculului natural, cum ar fi calculul evolutiv și neuronal, au fost investigate o perioadă mai îndelungată, alte

domenii, precum calculul molecular și cuantic, sunt mult mai tinere și mult mai puțin investigate. Bazate pe fenomene naturale, calculul molecular folosește paradigme ale biologiei moleculare, iar calculul cuantic principiile fizicii cuantice și exploatează paralelismul cuantic.

În ultimii ani, utilizarea mecanismelor și tehnicilor de calcul natural au devenit principala forță de elaborare a tehnologiilor informaționale și de comunicare. Mai mult decât atât. Paradigmele și metodele calculului natural reprezintă o contribuție semnificativă în domeniul informaticii teoretice, ținând cont de faptul că perceperea esenței calculului constituie subiectul principal al acestei științe. Astfel, cercetările efectuate în domeniul calculului natural sunt pârghii importante ale progresului pentru ambele fațete ale informaticii (teorie și practică).

Calculul biologic (numit, de asemenea, calcul molecular sau celular) este un nou domeniu interdisciplinar de cercetare situat la intersecția informaticii, chimiei, biologiei și fizicii care promite schimbări revoluționare în domeniul calculatoarelor în viitorul apropiat. El urmărește dezvoltarea de noi echipamente și sisteme de calcul pentru procesarea informației la nivel atomic sau molecular în baza principiilor biologice, chimice, electronice, fotonice și/sau mecanice.

Cel mai vechi exemplu de model de calcul inspirat de natură este automatul celular, propus de Ulam și von Neumann în anii 1940. Aprofundarea cercetărilor în teoria automatelor a dat naștere domeniului de calcul neuronal, care îmbină cercetări din domeniile neuroștiințelor (neuroștiințe computaționale) și informaticii (rețele neuronale).

Evoluția darwiniană a inspirat paradigma de calcul evolutiv. În informatică, tehnicile evolutive sunt utilizate pe scară largă pentru rezolvarea problemelor de optimizare a parametrilor prin aplicarea strategiilor de evoluție sau a programării evolutive.

Un alt grup prevalent de modele de calcul inspirate de natură este cunoscut sub numele de algoritmi genetici. Algoritmii genetici sunt aplicați cu succes la soluționarea problemelor combinatorice și de optimizare.

Printre paradigmele, inspirate de inteligența colectivă prezentă în natură, trebuie menționați algoritmi furnicii (ant algorithms) și inteligența roiului (swarm intelligence). Algoritmii furnicii sunt modele inspirate de comportamentul coloniei de furnici, iar algoritmul inteligența roiului – de comportamentul unui roi de organisme biologice mobile (cum ar fi bacteriile, furnicile, termitelile, albinele, păianjenii, peștii, păsările), în care fiecare individ comunică cu

ceilalți fie direct, fie indirect, prin acțiuni asupra mediului său local. Cea mai răspândită aplicație a inteligenței roiului este optimizarea roi a particulei (particle swarm optimization).

Viața artificială (ALife) constă în încercarea de a crea sisteme artificiale care, în mod normal, manifestă proprietăți specifice doar organismelor vii. Un exemplu de model al vieții artificiale a fost propus în 1968: sistemele Lindenmayer (L-sisteme). Acestea sunt sisteme cu rescriere paralelă care, pornind de la un cuvânt inițial, continuă aplicând regulile de rescriere în paralel la toate literele cuvântului generând, astfel, cuvinte noi.

Sistemele imunitare artificiale sunt, de asemenea, modele de calcul care utilizează inteligența colectivă, dar pot efectua calcule mai complexe într-o manieră paralelă și distribuită. Ele folosesc procedee de învățare, memorizare, de recuperare asociațivă și alte mecanisme pentru a soluționa probleme ce țin de recunoaștere și clasificare.

Paradigma calculului biomolecular se bazează pe ideea că datele pot fi codificate prin biomolecule, iar metodele biologiei moleculare pot fi utilizate pentru a le transforma efectuând operații aritmetice sau logice.

Modelul de calcul ADN (DNA computing) presupune că biomoleculele sunt ADN-uri. Nașterea acestui domeniu este marcată prin experimentul realizat în anul 1994 de către Leonard Adleman, care a reușit să rezolve un caz particular al problemei căii hamiltoniene într-un graf orientat exclusiv prin manipularea eprubetelor cu ADN-uri. Ulterior s-au propus și alte metode pentru construirea dispozitivelor de calcul bazate pe ADN.

Domeniul de cercetare calculul membranal, având la origine încercarea de a formula un model de calcul inspirat de structura și funcționarea unei celule vii, este introdus de Gh.Păun în 1998. Astfel, primele modele propuse aveau la bază o structură asemănătoare amplasamentului celular al membranelor (ierarhică), care delimitează niște compartimente în care mulțimile de chimicale (numite obiecte) evoluează în conformitate cu regulile de evoluție specificate. Aceste reguli fie modelează reacțiile chimice și reprezintă niște reguli de rescriere (pentru mulțimuri), fie au fost inspirate de alte procese biologice, cum ar fi penetrarea obiectelor prin membrane. Modelele au fost mai apoi modificate prin încorporarea unor caracteristici suplimentare inspirate de idei înrădăcinate în biologie, matematică sau informatică.

În cadrul proiectului STCU 4032 „Puterea și eficiența calculului natural: P sisteme (membrana-

re) neuronale“ („Power and efficiency of natural computing: neural-like P (membrane) systems”) (managerul proiectului – prof. Iurie Rogojin, instituția coordonatoare – Institutul de Matematică și Informatică al Academiei de Științe a Moldovei), s-au cercetat noile posibilități de calcul oferite de mai multe modele importante din domeniul calculului biologic. S-a ținut cont atât de aspectele teoretice, cât și de cele practice ale acestor modele.

Principalele obiective ale proiectului preconizau aplicarea metodelor și rezultatelor oferite de calculul natural (membranal) în algebra computațională și lingvistica matematică, în scopul soluționării eficiente a unor probleme importante din aceste domenii, elaborarea unui instrumentar software pentru simularea P sistemelor și contribuirea la dezvoltarea teoriei P sistemelor.

Au fost cercetate noile posibilități oferite de calculul cu biomolecule. În special, Rețelele de Procesoare Evoluționiste (RPE) și Rețelele Hibrice de Procesoare Evoluționiste (RHPE), diverse variante de P sisteme (comunicaționale, tranziționale, cu membrane active, polimorfe, splicing), sisteme cu inserție-deleție, P sisteme cu inserție-deleție cu un număr mic de parametri. Multe dintre aceste modele pot fi considerate și modele de sisteme neuronale (în special, P sistemele tissue-like, RPE și alte exemple concrete de dispozitive universale de calcul elaborate în cadrul proiectului). S-a mai examinat un model intramolecular pentru procesul de asamblare a genelor în ciliate, care poate fi considerat o paradigmă bio-inspirată potrivită pentru reprezentarea P sistemelor tissue-like.

S-a cercetat frontiera de intractabilitate (indecidabilitate, în cazul nostru) a problemelor propuse spre rezolvare pentru majoritatea acestor modele în scopul de a înțelege mai bine posibilitățile lor de calcul (universalitatea, eficiența, complexitatea și altele). Astfel, s-a demonstrat că orice limbaj recursiv enumerabil poate fi generat de o RHPE completă cu 7 noduri, obținând un răspuns afirmativ la întrebarea formulată de E.Csuhaj-Varjú, C.Martín-Vide și V. Mitrana (2005) cu privire la posibilitatea de a genera un limbaj recursiv enumerabil arbitrar peste un vocabular V cu o RHPE completă de dimensiuni mai mici decât $27+3 \cdot \text{card}(V)$. S-a arătat, de asemenea, că familia RHPE cu 2 noduri nu poate genera limbajele recursiv enumerabile.

S-a introdus o nouă variantă de sisteme de calcul cu eprubete, sisteme ADN cu separarea lungimii, bazat pe operații splicing și comunicare între eprubete prin cuvinte filtrate conform lungimii, inspirate de tehnica de laborator electroforeză în gel.

S-a arătat că sistemele ADN cu separarea lungimii, chiar și cu parametri foarte limitați, sunt capabile să simuleze mașini Turing. Astfel, în baza lor s-ar putea construi (teoretic) un biocalculator universal.

Calculul membranar reprezintă un model formal de calcul distribuit și celular. S-a definit o variantă nouă de P sisteme cu reguli de rescriere, numite P sisteme polimorfe, care admit ca regulile oricărei regiuni să fie definite prin conținutul regiunilor interioare, dar nu prin specificare explicită la descrierea sistemului. Această idee este inspirată de conceptul lui von Neumann „program is data“, dar și de direcția nouă de cercetare propusă de Gh. Păun privind nucleul celulei. Pentru P sisteme splicing s-a demonstrat o afirmație remarcabilă: 6 reguli splicing sunt suficiente pentru universalitatea sistemului (am obținut astfel P sistemul universal cu cel mai mic număr cunoscut de reguli).

Metodele de calcul și rezultatele calculului natural (membranar) s-au aplicat în algebra computațională. Acest domeniu ridică probleme extrem de complicate, care impun căutarea unor metode eficiente de rezolvare. Pe parcurs, au fost analizate sisteme de calcul utilizate atât în algebra computațională comutativă, cât și în cea necomutativă. Rezultatul obținut a fost prezentat sub forma unei baze de date. Sistemele analizate ținneau cont de eficiență, de terminarea calculelor, dar și de unele detalii tehnice privind implementarea acestora. De asemenea, s-au cercetat metodele existente de calcul paralel utilizate în sistemele algebrei computaționale. S-a studiat o serie de probleme din teoria matricelor (calculul permanentului), teoria bazelor Gröbner (determinarea dacă algebra are dimensiune finită, verificarea dacă un set dat de polinoame formează baza Gröbner pentru algebra dată). Pentru a simula algoritmi elaborați a fost folosit sistemul P-lingua. S-a constatat că acest sistem nu poate fi aplicat pentru a simula manipulări cu monoame. S-a propus ca acest sistem să fie extins cu un mecanism de parametrizare a regulilor.

S-au aplicat metode și rezultate ale calculului natural (membranar) în lingvistica matematică. Rezultatele principale obținute în acest domeniu sunt algoritmi originali elaborați pentru modelarea procesului de divizare a cuvintelor în silabe, metodele de reprezentare, creare și gestionare a dicționarelor electronice, algoritmi de paralelizare a procesului de flexionare, formalismele noi de generare a limbajelor cu sisteme membranare tranzitionale.

În continuare a fost elaborat un instrumentar software pentru construirea simulatoarelor P sistemelor. Astfel, a fost propusă versiunea beta a instrumentarului care permite crearea simulatoarelor pentru P sisteme orientate la soluționarea unor pro-

bleme specifice. Folosind acest instrumentar, s-au construit simulatoare pentru câteva P sisteme în scopul de a demonstra și obține soluții pentru problemele din algebra computațională și lingvistică cu aplicarea paralelismului masiv inspirat de calculul biomolecular. Crearea simulatoarelor servește drept test pentru aplicabilitatea instrumentarului elaborat. Simulatoarele s-au aplicat pentru imitarea dezvoltării P sistemelor, propuse pentru modelarea soluțiilor unor probleme din algebra computațională și lingvistică. Astfel, simulatoarele au permis validarea acestor soluții.

În cadrul proiectului s-au obținut mai multe rezultate importante (teoretice și practice) publicate în reviste de prestigiu internațional, și anume, în *Theoretical Computer Science*, Elsevier (5 articole); *Natural Computing*, Springer (4 articole); *Lecture Notes in Computer Science*, Springer (9 articole); *Fundamenta Informaticae*, IOS Press (2 articole); *Biosystems*, Elsevier (1 articol); *Discrete Applied Mathematics*, Elsevier (1 articol); *International Journal of Computers, Communications and Control*, Agora University Editing House (1 articol); *International Journal of Foundations of Computer Science*, World Scientific (1 articol); *Romanian Journal of Information Science and Technology*, Romanian Academy (1 articol); *Romanian Biological Science* (1 articol); *Computer science Journal of Moldova*, (4 articole). Membrii proiectului sunt coautori la două capitole din monografia *The Oxford Handbook of Membrane Computing*, ed. by G.Paun, G.Rozenberg, A.Salomaa, Oxford University Press, 2010. Membrul proiectului Vladimir Rogojin a susținut teza PhD „Gene Assembly in Stichotrichous Ciliates. Elementary Operations, Parallelism and Computation” la Abo Academy University, Turku, Finland, TUCS Dissertation, No 117, May 2009. O serie de rezultate din teza sa au fost incluse în rapoartele proiectului. De asemenea, membrii proiectului au prezentat comunicări și rapoarte invitate la numeroase conferințe internaționale dedicate calculului natural.

Un rezumat al cercetărilor în domeniul calculului natural efectuate în cadrul Institutului de Matematică și Informatică a fost publicat în: Alhazov A., Boian E., Burtseva L., Ciubotaru C., Cojocaru S., Colesnicov A., Demidova V., Ivanov S., Macari V., Magariu G., Malahova L., Rogojin V., Rogozhin Yu., Tofan T., Verlan S., Verlan T. *Investigations on Natural Computing in the Institute of Mathematics and Computer Science*. In: *Computer Science Journal of Moldova*. 2010, V.18, N.2(53), 101-138. ISSN 1561-4042.