

# SZAKMAI ZÁRÓJELENTÉS

a T042529. számú

„Biológiai indíttatású kiszámítás: formális nyelvi modellek „ című

OTKA kutatás keretében végzett munkáról

## 1. A kutatás háttere, célja, a munkatervben vállalt program

A biológiai indíttatású számítások elmélete egyike a jelenkori számítástudomány dinamikusan fejlődő, az érdeklődés középpontjában álló területeinek. A tárgykörben folytatott kutatások célja az, hogy az élő szervezetek és ezek közösségeinek felépítése és működési elveinek tanulmányozása alapján az ezen rendszerek sajátosságait hordozó, újszerű, nagy számítási erejű és nagy hatékonyságú számítási paradigmákat hozzanak létre, amelyek segítségével a számítástudomány és az informatika kurrens problémái hatékonyabban kezelhetők, illetve jobban megközelíthetők.

A formális nyelvek elméletében számos biológiai indíttatású modell létezik. Ilyen konstrukciók többek között a genom fejlődése által inspirált formális nyelvi műveleteken alapuló osztott eszközök (pl. az evolúciós processzorok hálózatai), az élő sejt felépítésének és működésének mintájára kialakított membrán rendszerek (P rendszerek) különböző változatai, valamint a molekuláris számítástudomány tárgykörébe is tartozó Watson-Crick D0L rendszerek és hálózataik.

Tervezett kutatásaink ezen tárgykörökhöz tartozó modellek számítási erejének és bonyolultságának (elsősorban méret- és számítási bonyolultság) leírására, illetve a modellek, valamint a tárgykörök matematikai eszköztárának továbbfejlesztésére, új elemekkel való bővítésére irányultak.

## 2. A tervezett program teljesítése

A kutatások során a felsorolt tárgykörök mindegyikében értünk el érdemi eredményeket, amelyek ismert modellek tulajdonságainak pontosabb leírásában, új, érdeklődésre számot tartó modellek létrehozásában, illetve a formális nyelvek bizonyos klasszikus fogalmainak újszerű interpretálásában nyilvánultak meg. Megállapításaink elsősorban a tanulmányozott modellek számítási erejére és (méret-, illetve számítási) bonyolultságára vonatkoztak, ezáltal lehetőséget adva a hagyományos és az attól eltérő formális nyelvi eszközök alaptulajdonságainak összevetésére. Eredményeink fogadtatása a nemzetközi kutatói közösségben pozitív volt, gondolataink, a kialakított módszerek további munkákra inspiráltak kutatókat.

Eredményeink nívós nemzetközi publikációs fórumokon (folyóiratokban, szerkesztett kötetekben, konferenciákon) jelentek meg. A publikációs fórumok között vezető folyóiratok és elismert folyóiratok vannak (Theoretical Computer Science, Acta Informatica, Fundamenta Informaticae, Natural Computing, Soft Computing), a szerkesztett kötetek a Springer kiadó Lecture Notes in Computer Science kötetei, a konferenciák pedig a membrán rendszerek évenként megrendezésre kerülő konferenciái, valamint a molekuláris számítástudomány DNA 10 vezető konferenciája. További folyóiratokhoz beküldött dolgozataink várnak elbírálásra, illetve elkészített dolgozataink folyóirathoz való beküldésre és elbírálásra. Az elfogadásról értesíteni szeretnénk a bizottságot. Ezen dolgozatokat elérhetővé tesszük a <http://www.sztaki.hu/tcs/otka-pub> webcímen.

A kutatás során a munka jelentős részét Csuha Varjú Erzsébet és Vaszil György végezte el, Csima Judit kisgyermekének gondozása miatt csak korlátozott mértékben tudott rendelkezésre állni, de az ő közreműködését csak az egyik tárgykörben terveztük. A munkavégzés során törekedtünk az intenzív nemzetközi együttműködésre, eredményeinknek a nemzetközi kutatói közösségben való megismertetésére és elismertetésére. Ennek megfelelően szorosan együttműködtünk a molekuláris számítástudomány és a membrán számítások olyan meghatározó egyéniségeivel mint Gh. Paun, G.

Mauri, O. H. Ibarra, A. Salomaa, I. Petre, M. J. Pérez-Jiménez, V. Mitrana, valamint az European Molecular Computing Consortium kutatóival. Munkánkat versenykörnyezetben végeztük, amely a tanulmányozott tárgykörök dinamikus fejlődéséből adódott. A kutatási időszak alatt vizsgálataink aktuális súlypontjának kialakításánál figyelembe vettük a nemzetközi kutatási trendek, a nemzetközi érdeklődés és a saját eredményeink iránt megnyilvánuló kifejezett érdeklődés irányát, ezért munkánk jelentős részét a membrán rendszerek tanulmányozására fordítottuk.

### **3. Az elért eredmények**

#### **3.1. Molekuláris számítástudomány**

##### *3.1.1. Evolúciós processzorok hálózatai*

Az evolúciós processzorok hálózata egy virtuális gráf csúcspontjaiban elhelyezkedő, szavak halmazain, illetve multihalmazain működő átíró rendszerekből (evolúciós processzorokból) áll, amely processzorok kizárólag szimbólumok beszúrására, vagy törlésére, vagy kicserélésére alkalmasak. A szavak DNS sorozatokat, a műveletek elemi genetikai műveleteket reprezentálnak. A működés során az egyes processzorok szinkronizált módon átírják a hozzájuk tartozó szavakat, majd az így nyert szimbólumsorozatokat szövegfeltételekkel adott szűrőkön keresztül egymásnak közvetítik. A processzor elemi, ha a műveletet csak egy bizonyos szimbólumra képes alkalmazni. A hálózatot hibridnek nevezzük, ha az egyes processzorok különbözhetnek működési módjukban (a művelet végrehajtási helye) és szűrőjükben. A fogalmat sejtbiológiai motivációk alapján 2001-ban vezették be J. Castellanos és szerzőtársai, inspirálva a témavezető egyik korábbi, A. Salomaa-val közös nyelvprocesszor-hálózatokról szóló dolgozata által is. Vizsgálataink hibrid evolúciós processzorhálózatokkal kapcsolatos nyitott kérdések eldöntésére irányultak. C. Martín-Vide és V. Mitrana társszerzőinkkel megmutattuk, hogy az elemi evolúciós processzorok hibrid hálózatai a Turing gépekkel egyenlő számítási erejű eszközök, továbbá bebizonyítottuk, hogy a nem elemi processzorok esetén minden rögzített ábécé felett értelmezett rekurzíven felsorolható nyelv meghatározható hasonló topológiájú hálózat segítségével (közleményjegyzék, [4]). Az elért eredmények azt bizonyítják, hogy rendkívül egyszerű számítási eszközök hálózatai is maximális számítási erőt képviselhetnek, sőt, ez az erő bizonyos esetekben hasonló architektúrájú hálózatokkal is elérhető.

##### *3.1.2. Watson-Crick D0L rendszerek hálózatai*

A Watson-Crick D0L rendszerek hálózatai (az NWD0L rendszerek) olyan párhuzamos átírást alkalmazó rendszerek (determinisztikus Lindenmayer rendszerek; D0L rendszerek) hálózatai, ahol a csúcspontokban elhelyezkedő szavak ún. DNS típusú ábécé elemeiből képzettek. A hálózat csúcsaiban elhelyezkedő Lindenmayer rendszerek a hozzájuk tartozó szavakon szinkronizált módon átírást végeznek, majd az így nyert szimbólumsorozatokat egy, a DNS jellemző tulajdonságát, a Watson-Crick komplementaritás elvét alapul vevő kommunikációs protokollt használva egymásnak közvetítik. A fogalmat a témavezető és A. Salomaa vezette be 2000-ben. A témavezető egy megelőző dolgozatában bebizonyította, hogy az ún. kiterjesztett standard Watson-Crick D0L rendszerek hálózatai a Turing gépekkel egyenlő számítási erejű eszközök. A projekt keretében végzett kutatásaink során megmutattuk, hogy ez a számítási erő akkor is fennáll, ha a hálózat csúcspontjaiban levő Watson-Crick D0L rendszerek nem feltétlenül teljes információt közvetítenek egymásnak, azaz, a rendelkezésükre álló szavaknak valamely elő-, illetve utótagját is elküldhetik a másik csúcsba (közleményjegyzék, [2]). Később az eredmény érvényességét további nem teljes információközlés és kommunikációs protokoll esetére is megmutattuk („E. Csehaj-Varjú, J. Csima: On Information Communication in Networks of Standard Watson-Crick D0L Systems”, kézirat). Az eredmények azt jelentik, hogy Watson-Crick D0L rendszerek hálózatai képesek a nem teljes információ felismerésére és kiszűrésére is. A. Salomaa társszerzőnkkel egy korábbi dolgozatunkban megmutattuk, hogyan lehet az NWD0L rendszereket NP teljes problémák megoldására használni. A projekt keretében megvizsgáltuk ezen problémák megoldásának lehetőségét a nem teljes információkövetítés

lehetősége mellett („E. Csuhaj-Varjú, J. Csima: On the Power of Networks of Watson-Crick DOL Systems with Incomplete Information Communication”, kézirat).

### 3.1.3. Nanostruktúrákat reprezentáló szimbólumsorozatok önfelépítő mechanizmusai

A molekuláris számítástudomány területén végzett kutatásaink további eredménye az a I. Petre társszerzőnkkel készített dolgozatunk, amely matematikai megközelítést adja nanostruktúrák (pl. DNS sorozatok) bizonyos önfelépítő eljárásának (self-assembly) (közleményjegyzék, [17]). Az önfelépítő eljárások során komplex struktúrák épülnek fel egyszerű elemekből. Az ilyen eljárások természetének megismerése jelenleg a molekuláris számítástudomány egyik kurrens problémaköre. Modellünkben a struktúrákat szavakkal, a szavak felépítési eljárását a szavak utó-, illetve előtagjainak átfedésével reprezentáltuk. (Azaz, az  $uv$  és a  $vw$  szóból az  $uvw$  szó építhető fel.) Megvizsgáltuk különböző típusú (reguláris, környezetfüggő, rekurzív, stb.) kiindulási szavak halmazai esetén a felépíthető szavakból álló nyelvek osztályait, azok zártági tulajdonságait, továbbá meghatároztuk egy adott nyelvhez minimális kiinduló szóhalmaz léte eldönthetőségének kérdését reguláris nyelvek esetében. A kutatások lehetséges folytatásaként további méret-, illetve ún. levezetési (elérhetőségi) bonyolultsági problémák tisztázását javasoltuk. Munkánk iránt máris mutatkozott érdeklődés.

## 3.2. Membrán rendszerek (P rendszerek)

Érdemi eredményeket értünk el a membrán rendszerek (P rendszerek) tulajdonságainak vizsgálatában is. A membrán rendszerek az élő sejt felépítésének és működésének mintájára kialakított konstrukciók, amelyek hierarchikusan egymásba ágyazott (fa struktúrát követő) ún. membránokkal határolt régiókból állnak, amely régiók objektumokat tartalmaznak. Az objektumok molekuláknak, kémiai alkotórészeknek felelnek meg. A rendszer működése során az objektumok változni (fejlődni) tudnak, valamint bizonyos szabályok szerint közlekedni képesek a szomszédos régiók között. A P rendszerek régiók tartalmát, azaz, a bennük levő objektumok összességét szimbólumok vagy komplex struktúrák (pl. szavak) multihalmazaival azonosítják, a fejlődési, illetve a közlekedési szabályokat pedig formális nyelvi átírási szabályokkal adják meg. A membrán rendszer fogalmát Gh. Paun vezette be 1998-ban. Azóta az elmélet a biológiai indíttatású számítástudomány egyik, az érdeklődés homlokterében álló ágává fejlődött (pl. az Institute for Scientific Information 2003. októberében ún. „Fast Emerging Research Front in Computer Science” besorolásban részesítette).

### 3.2.1. P automaták

Jelentős eredményeket értünk el a P automaták, azaz, az elfogadó P rendszerek tulajdonságainak meghatározásában is. A P automata kifejezett érdeklődést keltett fogalmát Csuhaj Varjú Erzsébet és Vaszil György vezette be 2002-ben. Ezek a P rendszerek a klasszikus automatákhoz hasonlóan viselkednek, működésük során a külvilágból lépésről-lépésre objektumok véges multihalmazait fogadják inputként, amely kommunikáció befolyásolja működésüket. Korábban bebizonyítottuk, hogyan kaphatók meg a rekurzíven felsorolható nyelvek szavai P automaták által elfogadott véges inputsorozatok képeként. Vizsgálatainkat folytatva, O. H. Ibarra társszerzőnkkel együtt megmutattuk, hogy a P automaták által elfogadott inputsorozatok halmazai az automata működési módjától függően (szekvenciális vagy ún. maximálisan párhuzamos) bizonyos, a Turing gépek által elfogadott, a logaritmikusnál kisebb tárcorlátos nyelvek osztályát, illetve a környezetfüggő nyelvek osztályát alkotják (közleményjegyzék, [6],[10]). Ezek az eredmények a természetes rendszerek bonyolultságáról adnak információkat, hiszen egy ilyen rendszer esetében a számításához mindig csak a rendelkezésünkre álló, ténylegesen létező objektumokat használhatjuk. Ugyancsak fontos megjegyezni, hogy a P automaták segítségével a környezetfüggő nyelvek osztálya természetes módon reprezentálható, amely nem mondható el minden klasszikus megközelítésről. Eredményeinket a DNA 10, 2004 vezető konferencián ismertettük, ahol kifejezett érdeklődést mutatkozott irántuk. A P automaták további tanulmányozásához újabb szempontokat is javasoltunk (közleményjegyzék, [1],[5]), illetve felvetettük a P automaták mint a végtelen ábécé felett értelmezett nyelvek reprezentálására alkalmas eszköz voltát (közleményjegyzék, [8]).

A P automaták, a L. Cardelli által 2004-ben megfogalmazott, folyamatalkébrái (process algebra) alapokon nyugvó ún. membrán kalkulusok (Brane Calculi) elmélete, valamint a klasszikus automataelmélet fogalmai által inspirálva, kialakítottunk egy olyan automata fogalmat membrán rendszerekre, amely nemcsak a régiókban levő objektumok tartalmát, hanem a membránokhoz csatolt multihalmazokból álló címkéket is használ működése során. A fogalom biológiai háttéréül a membrán felületéhez tapadó, onnan leváló fehérjéknek a sejt életében játszott szerepe szolgált. Megmutattuk, hogy ez a P automata változat is a Turing gépekkel egyenlő számítási erejű. A dolgozat jelentőségét az adja, hogy a fogalom kapcsolatot teremt a membrán rendszerek két kalkulusa, valamint a klasszikus automataelmélet között, továbbá kapcsolódási pontot jelent az ún. rendszerbiológia (systems biology) területéhez. Az eredményt tartalmazó dolgozatunkat folyóirathoz közlésre benyújtottuk, jelenleg bírálattal áll (közleményjegyzék, [9]).

### 3.2.2. *Antiport P rendszerek*

A membrán rendszerek egyik fontos részsztyálját alkotják a biológiai szempontból is jól motivált, ún. antiport szabályokkal rendelkező P rendszerek, melyek működése kizárólagosan csak objektum-multihalmazoknak a szomszédos régiók közötti cseréjén alapul. A fogalmat A. Paun és Gh. Paun vezette be 2002-ben. M. Margenstern és S. Verlan társszerzőinkkel megmutattuk, hogy ezek a P rendszerek nemcsak objektumaik korlátozott száma, hanem szabályaik számának konstans korlátja mellett is a Turing gépekkel egyenlő számítási erőt képviselnek. A bizonyítás eszközeül egy olyan erős értelemben vett univerzális regiszter gép szimulálását használtuk, amely szabályainak száma egy bizonyos konstans alatt marad. Az antiport P rendszer szabályainak száma mellett új, a konstrukcióra jellemző méretbonyolultsági mértékeket is bevezettünk, amelyekre és további mértékekre is igazoltuk a korlátos mérték melletti univerzális kiszámítási erő meglétét (közleményjegyzék, [13]). Kialakított szimulációs technikánkat azóta a P rendszerek más modelljeire is alkalmazták.

### 3.2.3. *P kolóniák*

Megvizsgáltuk rendkívül egyszerű P rendszerek együtteseinek, az ún. P kolóniáknak számítási erejét és méretbonyolultságát. A P kolónia fogalmát J. Kelemen, A. Kelemenová és Gh. Paun vezette be 2004-ben. Ezek a P rendszerek ún. sejtekből épülnek fel, amelyek egy membránból álló, a működés során mindig konstans számú objektumot tartalmazó konstrukciók. A sejtek egy szimbólumok végtelen tárházával reprezentált környezetben helyezkednek el. A sejt programokat hajt végre, a program a sejt egy objektumának a környezet egy objektumával való kicserélését, illetve saját többi objektumának átírását jelenti. J. Kelemen, A. Kelemenová és Gh. Paun társszerzőinkkel megmutattuk, hogy bár ezen rendkívül egyszerű membrán rendszerek között közvetlen kapcsolat nincs, környezetükkel együtt mégis a Turing gépekkel egyenlő számítási erőt képviselnek, még bizonyos méretparamétereik korlátozása esetén is (közleményjegyzék, [12]). M. Margenstern társszerzőnkkel együtt igazoltuk továbbá azt is, a P kolóniák a sejtek számának és azok programjának konstans korlátjai együttes fennállása esetén is univerzális erejűek (közleményjegyzék, [11]). Bár a modell motivációja különbözött, a P kolóniák multihalmazokon értelmezett evolúciós műveletekre (pont mutációk) épített osztott rendszereknek is tekinthetők, így az eredmények az evolúciós modellek tárgykörében is interpretálhatóak.

### 3.2.4. *P rendszerek komplex objektumokkal*

A P rendszerek fontos válfajait jelentik azok a rendszerek, amelyek esetében az objektumok nem szimbólumok, hanem komplex struktúrák, például szavak.

Ilyen, biológiai szempontból jól motivált P rendszer az ún. sarjadzó (gemmating) membrán rendszer. Ezek rendszerek szavakból álló objektumokon vannak értelmezve, amely objektumok közül egyesek a működés során nemcsak a szomszédos régiók egyikébe, hanem egy bizonyos - akár távoli-, kijelölt régióba is távozhatnak. Ezek a szavak kivételesen nagy molekulákat reprezentálnak, amelyek túlságosan nagy méretük miatt nem képesek áthatolni a membrán falán, ezért - mintegy mobil

membránt alkotva - egy meghatározott régióba távoznak speciális, fejlődéssel kombinált közlekedési szabályok segítségével. Az objektumok fejlődése mutáció, replikáció, illetve hasadás segítségével megy végbe. A mobil membránok tartalma a célrégió tartalmával egyesül. A fogalmat D. Besozzi és szerzőtársai vezették be 2001-ben, és megmutatták, hogy ezek a rendszerek a Turing gépekkel egyenlő számítási erejűek. D. Besozzi, G. Mauri és C. Zandron társszerzőinkkel csökkentettünk az univerzális számítási erő elérését biztosító részosztály elemeinek méretére adott korlátot (közleményjegyzék, [3]). Későbbi dolgozatunkban (Csuha Varjú Erzsébet és Vaszil György) pontos értéket adtunk meg ezen korlátra (közleményjegyzék, [7]).

Természetes megfontolásokból és formai hasonlóságokból kiindulva, interpretáltunk a P rendszerek elméletében a formális nyelvek egyik részterületén, a grammatikarendszerek elméletében használatos fogalmakat. Ezen munkáinkat Gh. Paun társszerzőnkkel készítettük. Így – például - bevezettük szavakon értelmezett membrán rendszerekre a kommunikáció egy új fajtáját, az ún. t-kommunikációt. A t-kommunikáció értelmében egy objektum akkor hagyja el tartózkodási régióját, ha ott már nem képes továbbfejlődni, azaz, nincs olyan fejlődési szabály, amely rá alkalmazható, illetve, ha erre közvetlen utasítást kap. Ez azt jelenti, hogy az egyedek elhagyják környezetüket, ha az további fejlődésükre alkalmatlan. Vizsgálataink során különböző kommunikációs altípusokat véve figyelembe meghatároztuk ezen rendszereknek számítási erejét, és megállapítottuk, hogy már viszonylag kevés - hat vagy kilenc, típustól függően - membránnal rendelkező P rendszerek is elérhetik a Turing gépek számítási erejét (közleményjegyzék, [15]).

Ugyancsak sikeresen alkalmaztunk egy másik kommunikációs elvet is (szintén Gh. Paun társszerzőnkkel). Ezek a vizsgálatok ún. szövetszerű, szavak felett értelmezett P rendszerekre vonatkoznak (tissue-like P systems with string objects), ahol a membrán architektúra egy tetszőleges gráfnak felel meg. Modellünkben a szavak régiók közti kommunikációját speciális kommunikációs szimbólumoknak a szavakban való felbukkanása vezérli. A szimbólumok régiókat jelölnek. Hatásukra a megjelölt régió kommunikációs szimbólumtól mentes szavainak másolatai behelyettesítődnek a kommunikációs szimbólum helyére. Azok a szavak, amelyek kommunikációs igénye nem kielégíthető, elvesznek, azaz, kitörölődnek a rendszerből. Ezen kommunikáció két típusát vizsgáltuk meg, az ún. fertőzést (infekció; i-kommunikáció), amikor a bemásolt szó megmarad az eredeti régióban, valamint a parazitizmust (p-kommunikáció), amikor a bemásolt szó elvész, csak másodpéldányai „élnek tovább” az új szavakban az új régiókban. Megmutattuk, hogy mindkét kommunikációs mód esetében ezen P rendszerek a Turing gépekkel egyenlő erejű kiszámítási eszközök, i-kommunikáció esetében már hét membránból álló rendszerek esetében is. Bizonyítottuk, hogy a p-kommunikációval működő rendszerek szimulálják az i-kommunikációval működő rendszereket, továbbá megmutattuk, hogyan lehet p-kommunikációs P rendszerekkel a kielégíthetőségi problémát (SAT) lineáris időben megoldani. Dolgozatunkat közlésre benyújtottuk (közleményjegyzék, [16]).

### *3.2.5. P rendszerek konfigurációjának szerkeszthetősége*

Vizsgálatokat kezdtünk az ún. membrán kalkulus létrehozása lehetőségének irányában. A P rendszer konfigurációján a membránok által meghatározott architektúrát és az egyes membránokkal körülvett régiókban található objektumok multihalmazait értjük. A rendszer működése konfigurációjának változásában nyilvánul meg. A számítás sikeres, ha a működési folyamat megáll. Értelemszerűen felmerülő kérdés azonban az, hogy mit tudunk mondani a konfigurációk szerkeszthetőségéről, azaz, ha eltekintünk a számítási folyamat eredményétől és veszünk két tetszőleges konfigurációt, akkor az egyik konfigurációból a másik megkapható-e egy előre megadott szabályrendszerrel, és ha igen, meghatározható-e a két konfiguráció ún. szerkesztési távolsága. Megvizsgáltuk különböző szabályrendszerek esetén a szerkesztési távolság kiszámíthatóságát, az ún. generátor elemek létét bizonyos konfigurációhalmazok esetén. Ezt a munkánkat A. Di Nola, Gh. Paun és M.J. Pérez-Jiménez társszerzőinkkel együtt készítettük. Dolgozatunkat közlésre benyújtottuk (közleményjegyzék, [14]).

#### **4. A kutatás perspektívái, az eredmények felhasználásának, hasznosításának lehetőségei**

Kutatási témáink mindegyike továbbfejleszthető, perspektivikus vizsgálatok, hiszen a jelenkori számítástudomány egyik fő vonulata a hagyományos elvektől eltérő, hatékony kiszámítási eszközök létrehozása. Elért eredményeink hasznosítása elsősorban a számítástudományon belül képzelhető el. Általánosságban azt mondhatjuk, hogy kutatásaink ahhoz járulnak hozzá, hogy pontosabban megismerhessük a számítás és a számítási eszközök természetét, ezen fogalmak határait, különös tekintettel a formális nyelvi modellekre.