

Data:  
L. Dz.

Toruń, 23 stycznia, 2018

Prof. dr hab. Wiesław Nowak  
Instytut Fizyki  
Uniwersytet M.Kopernika w Toruniu  
87-100 Toruń, ul. Grudziądzka 5  
(wiesiek@fizyka.umk.pl)

RECENZJA rozprawy doktorskiej  
Mgr Natalii Szóstak pt.

## “Bioinformatyczne metody modelowania i weryfikacji hipotezy Świata RNA”

### 1. Problem badawczy i jego znaczenie

W nauce współczesnej nie ma aż tak wielu fundamentalnych pytań dotyczących natury wszechświata i przyrody jak to było np. na początku XX wieku. Jednak kilka takich problemów można wyróżnić, np. Czy jest jeszcze jakaś inna forma inteligentnego życia w kosmosie? Jak pracuje mózg? Jaki jest mechanizm zachowania informacji w mózgu ludzkim? Co to jest ciemna materia? Czy istnieje kod RNA? Jak powstało życie? ... Autorka rozprawy podjęła próbę „zaatakowania” tego ostatniego problemu. Jest on w pewnością ważny w nauce, poważne laboratoria i grupy teoretyków zmagają się z nim od lat. Z pewnością jest to jeden z najbardziej ambitnych problemów, jeśli nie najambitniejszy, z jakimi się dotychczas zetknąłem w długoletniej praktyce recenzenta i uczestnika wielu przewodów doktorskich. Istotą problemu rozwiązywanego w doktoracie jest zbudowanie modelu komputerowego procesu początków życia, rozumianego jako samoorganizacja materii pozwalająca molekułom kopiować informację do następných pokoleń. Zagadnienie ogólne nie jest nowe, proponowano w przeszłości algorytmy symulujące proces życia, jednak model zaproponowany w rozprawie, z racji swojej hierarchiczności i zakresu stosowanych formalizmów jest nowy i oryginalny.

Dyskutowane zagadnienie z pewnością należy do głównego nurtu biologii teoretycznej, interesuje chemików i fizyków, z racji formalizmu matematycznego i narzędzi informatycznych ma charakter informatyczny. Nie widzę (na tym etapie) w wynikach osiągniętych przez doktorantkę aspektów praktycznych, chyba, że do osiągnięć praktycznych zaliczymy stworzenie schematu modelowania różnych wariantów wczesnej ewolucji prekursorów układów biologicznych. Narzędzie to może być przydatne dla innych uczonych do testowania hipotez badawczych.

### 2. Wkład autorki

Wkład mgr Natalii Szóstak w powstanie recenzowanej rozprawy nie budzi wątpliwości. Autorka brała udział w tworzeniu modelu, ułożeniu hierarchii modeli, wyborze formalizmu informatycznego stosowanego do modelowania na każdym etapie, programowała szereg istotnych części modelu, nadzorowała programowanie innych fragmentów modelu. Wykonała przeszukanie aktualnej literatury przedmiotu badań, napisała obszerny wstęp stanowiący przegląd dotychczasowych osiągnięć i wiedzy w dziedzinie odkrycia mechanizmów początków życia i modelowania komputerowego tego procesu. Miała wiodący udział w przygotowaniu kilku publikacji z tematyki doktoratu. Prezentowała wielokrotnie wyniki cząstkowe na konferencjach krajowych i międzynarodowych, głównie informatycznych i bioinformatycznych.

Głównym osiągnięciem doktoratu było sformułowanie cyklu modeli o rosnącej złożoności, zdolnych prezentować w formie matematycznej i numerycznej pewne aspekty samoorganizacji materii nieożywionej. Wykonano poważną pracę koncepcyjną, analityczną, implementacyjną i obliczeniową. Uzyskano ciekawe i czasem sprzeczne z dotychczasowymi danymi literaturowymi wnioski na temat wielu czynników mających potencjalne znaczenie dla procesu powstawania życia. Wyniki sugerują, że na wczesnym etapie musiały pojawić się replikazy o dwóch typach funkcji – katalityczne i informacyjne. Stworzenie modelu (a raczej modeli) i jego analiza jest istotny wkładem Autorki do pewnego nurtu bioinformatyki.

### 3. Poprawność

- a. Główna hipoteza badawcza „*możliwe jest informatyczne modelowanie początków życia*” jest moim zdaniem dość trywialna. Oczywiście doktorat dowodzi słuszności tej tezy, ale i bez badań prowadzonych przez Autorkę większość specjalistów zgadza się z tym twierdzeniem, wynika to zresztą z przeglądu literatury zawartego w rozprawie. Trzy przedstawione w podrozdziale ze str. 11 bardziej szczegółowe hipotezy badawcze (nie przytaczam z braku miejsca) już wg mnie są sensowne i oryginalne.
- b. Szkoda, że nie podano argumentów przeciwko Hipotezie Świata RNA, są tylko argumenty „za”. Można było wykorzystać np. informacje zawarte w pracy : „The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)”, Harold S Bernhardt, *Biology Direct* 2012;7:23 (<https://doi.org/10.1186/1745-6150-7-23>).
- c. Na bardzo pozytywną ocenę zasługuje zaproponowanie hierarchicznego podejścia do problemu badania Hipotezy Świata RNA (Rozdział 5), wykorzystanie stosunkowo dużej prostoty formalizmu równań różniczkowych do sprecyzowania warunków brzegowych w bardziej złożonych obliczeniowo modelach automatów komórkowych, czy później w modelach wieloagentowych. W opisie tej koncepcji (5.3, str. 59) nie spodobała mi się jej nazwa: „koncepcja rozwiązania problemu powstania życia na Ziemi na podstawie modelu RP”. To, że powstał użyteczny model pozwalający testować wpływ rozmaitych parametrów na ewolucje układu R+P na powierzchni 2D nie oznacza jeszcze, że „rozwiązano problem powstania życia”.
- d. Praca napisana jest na ogół poprawnym językiem, ale zdarzają się niepoprawne sformułowania i wyrażenia żargonowe. Np. (d.1) Rys. 1 „numeryczna symulacja układu rr” – powinno być „numerycznej symulacji ROZWIĄZAŃ układu rr”; (d.2) rozumiem „mutowanie cząsteczek”, nie rozumiem „mutowania parametrów”; (d.3) str. 49 jest „epiderma” a ma być „naskórek”; (d.4) str 48 i inne – zwykle w lit. polskiej mówi się o dynamice brownowskiej, a nie „dynamice Browna”. (d.5) str. 42 nie spotkałem jeszcze „refleksyjnie autokatalitycznego zbioru” (?kalka z ang.); str. 43 chemicy mówią zwykle „reagenty” a nie „reaktanty”.
- e. Rozumiem celowość i znaczenie wprowadzenia parametru  $l_p$  do modelu. Nie znalazłem jednak uzasadnienia dlaczego nie wprowadzono analogicznego parametru  $l_r$ ? Czy replikazy nie ulegają „sfoldowaniu”?
- f. Wg mnie cytat przytoczony przy wprowadzeniu Hipotezy świata RNA nie jest poprawny (Cech 2012). „Wszyscy” podają, iż hipotezę tę wprowadził Gilbert w 1986 roku („*The RNA World*”, *Nature* **319**:618).

#### 4. Wiedza kandydatki

Szereg rozdziałów i podrozdziałów dowodzą dobrej wiedzy kandydatki w zakresie bioinformatyki (lub szerzej - informatyki). W Rozdziale 2, w dość elementarny co prawda sposób, podane są podstawy równań różniczkowych, opis teorii gier i elementy ewolucyjnej teorii gier, wprowadzenie do idei wykorzystania automatów komórkowych i systemów wieloagentowych do modelowania. Nie są to tylko proste opisy „encyklopedyczne”, rozdział zawiera autorskie uwagi nt. związku omawianych formalizmów z modelowaniem układów biologicznych. Bardziej dokładnie zastosowania omówione są w Rozdziale 4 („Modelowanie systemów biologiczno-chemicznych”) w tym, modele deterministyczne, stochastyczne, dynamiki molekularnej, modele oparte o automaty komórkowe i systemy wieloagentowe. W każdym podrozdziale podane są zwykle konkretne przykłady wykorzystania modeli informatycznych do opisu zjawisk chemicznych czy biologicznych.

Rozdziały 5-9 zawierają własne wyniki Autorki, zatem w oczywisty sposób ilustrują dobrą wiedzę i orientację w temacie. O wiedzy doktorantki świadczy też Rozdział 10, w którym zawarto krytyczne omówienie własnych wyników obliczeń i modelowania na tle innych rezultatów wydobytych z opublikowanych prac. Szczególne znaczenie ma dyskusja stabilności modelowanego systemu, roli pasożytów oraz ewolucji replikaz.

Spis literatury jest obszerny, zawiera wiele pozycji trudnych matematycznie, dowodzi b. dobrej orientacji w dziedzinie modelowania, rozprawa może stanowić cenny punkt wyjścia do dalszych badań/projektów w tej ważnej tematyce.

Jakość opisu w rozdziałach czy sekcjach jest nierówna. Na przykład, podobał mi się dobrze napisany podrozdział 3.9 „Kinetyka chemiczna”, ale w podrozdziale 4.3 „Metody dokładne” metody dynamiki molekularnej (MD) w jej różnych wariantach zastały opisane wg mnie zbyt ogólnikowo i niestarannie, np. zabrakło tam podstawowego równania (II zasada dynamiki Newtona) oraz wzmianki o metodach numerycznych całkowania układów równań stosowanych w metodach MD (np. Algorytm Verleta czy Leapfrog).

Ważnym aspektem badań było włączenie czynnika przestrzennego i dyfuzji do rozważań, co prawda postulowano to już we wcześniejszych publikacjach, jednak dopiero na poziomie analizy modelu wieloagentowego przedstawionego w doktoracie można docenić znaczenie i rolę tych czynników w ewolucji prostych replikujących się cząsteczek chemicznych. Autorka prezentuje umiejętność krytycznego wyciągania wniosków z dość „zaszumionych” i trudnych w interpretacji danych wynikających z symulacji ewolucji modeli.

#### 5. Inne uwagi

- 5.1 Czy słusznie możemy utożsamiać początki ewolucji, prezentowane niewątpliwie w tej rozprawie, z pytaniem o początki życia? Czy to jest to samo? A może po pojawieniu się ewoluujących samoreplikujących się układów potrzeba czegoś więcej by mówić o życiu, takim jakie znamy z autopsji? (zob. problem dyskutowany w: *Conceptualizing the origin of life in terms of evolution*; N. Takeuchi, P. Hogeweg, K. Kaneko Published 13 November 2017. DOI: 10.1098/rsta.2016.0346 )
- 5.2 W pracy Autorka w zasadzie nie podejmuje próby oceny czy stosowane w modelach parametry są fizyczne. Część osób może być z tego powodu sceptycznie nastawiona do modelowania. Zauważam, że Autorka zdaje sobie sprawę z tego ograniczenia, wspomina, że można by było, mając dane doświadczalne, zawężyć zakresy niektórych parametrów. Ja akceptuję na potrzeby tej pracy i tego etapu badań podejście zaprezentowane w rozprawie.

Wstępny etap eksploracji przestrzeni parametrów jest niezbędny do dalszego doskonalenia modeli samoorganizacji materii i jest równie wartościowy naukowo. Prace tego typu są możliwe właśnie dzięki metodom komputerowym – wydajnym implementacjom i szybkim komputerom.

### 5.3 Uwagi redakcyjne:

- Str. 31: Rys. 3.2 Podany jest termin „pętla terminalna” na strukturę drugorzędową RNA. Jest to nieładna kalka językowa (z ang.). Już na rysunku 3.4 autorka stosuje lepszy wg mnie termin „pętla końcowa” – na ten sam obiekt.
- Str. 93: Rysunek 8.5 jest nieczytelny, dane należało przedstawić np. na 3 wykresach.
- Załączony dysk CD warto byłoby podpisać.

Uwagi redakcyjne i stylistyczne nie umniejszają znacząco mojej wysokiej oceny rozprawy. Błędów redakcyjnych jest stosunkowo mało, układ treści b. czytelny, forma graficzna dobra, spis literatury przygotowany jest niezwykle starannie.

## 6. Podsumowanie

Biorąc po uwagę moje opinie przedstawione w punktach 1-5 i wymagania zdefiniowane przez art. 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami) moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

- A. Uważam, że rozprawa zawiera oryginalne rozwiązanie problemu naukowego (*zdecydowanie tak*).
- B. Po przeczytaniu rozprawy uważam, że kandydatka do stopnia doktora posiada ogólną wiedzę teoretyczną w dyscyplinie informatyka, a w szczególności w bioinformatyce (*zdecydowanie tak*).
- C. Sądzę, że rozprawa dowodzi, iż mgr Natalia Szóstak posiada umiejętność samodzielnego prowadzenia pracy naukowej (*zdecydowanie tak*)

Rozprawa spełnia zatem wymagania formalne [Ustawa z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki ze zmianami późniejszymi (Dz. U. Nr 65, poz. 595, ze zm. w Dz. U. z 2005 r. Nr 164, poz. 1365 oraz w Dz. U. z 2011 r. Nr 84, poz. 455)] i zwyczajowe stawiane rozprawom doktorskim.

### **Wnoszę o dopuszczenie Pani magister Natalii Szóstak do dalszych etapów przewodu doktorskiego.**

Ponadto, biorąc pod uwagę znaczenie poznawcze podjętego problemu; pomysłowe, hierarchiczne podejście do modelowania; różnorodność zastosowanych metod informatycznych; opublikowane wyniki; interdyscyplinarny charakter badań wnoszę o wyróżnienie rozprawy zgodnie z panującymi na Wydziale zasadami.

W. Nowak