

Zielona Góra, 18.02.2022.

Prof. dr hab. Mirosław R. Dudek  
Uniwersytet Zielonogórski  
Wydział Fizyki i Astronomii  
ul. Szafrana 4a, 65-516 Zielona Góra

**Recenzja rozprawy doktorskiej pt. „Badanie optymalności standardowego kodu genetycznego” napisanej przez Małgorzatę Wnętrzak.**

Przedstawiona do recenzji praca doktorska mgr Małgorzaty Wnętrzak dotyczy zagadnienia optymalności standardowego kodu genetycznego. W pracy, stosowany jest dla niego skrót anglojęzyczny SGC. Kod ten używany jest przez większość organizmów żywych i oznacza sposób przeniesienia informacji genetycznej 64 kodonów – trójek nukleotydów kodujących aminokwasy - na 20 aminokwasów. Z tych 64 kodonów 61 koduje aminokwasy a trzy to tzw. kodony STOP. W konsekwencji kod jest nadmiarowy w sensie niesionej informacji bo co najmniej jeden aminokwas musi być kodowany przez więcej niż jeden kodon. Dla tej nadmiarowości informacji powszechnie używane jest określenie, że kod genetyczny jest zdegenerowany. Właśnie ta nadmiarowość informacji oznacza, że własności fizykochemiczne aminokwasów muszą być skorelowane. Bardzo ładnie pokazane jest to w pracy doktorskiej na przykładzie Tab. 2 ze schematem standardowego kodu genetycznego wraz z zaznaczonymi wybranymi cechami fizykochemicznymi aminokwasów, gdzie z samego układu tabeli widać, że SGC minimalizuje efekty mutacji punktowych w sekwencji kodującej DNA. Badania związane z optymalizacją kodu genetycznego, które przedstawione zostały w pracy doktorskiej, pozwalają lepiej zrozumieć przyczyny nadmiarowości standardowego kodu genetycznego ze względu na właściwości fizykochemiczne kodowanych aminokwasów wskutek mutacji punktowych i błędów w procesie translacji. Jednocześnie pozwalają one zrozumieć istotę zmian kodu genetycznego w alternatywnych kodach genetycznych. Praca doktorska jest znakomicie napisana. Stanowi bardzo dobrą podstawę do monografii naukowej na temat optymalizacji kodu genetycznego.

Praca doktorska ma charakter pracy teoretycznej z odniesieniem się uzyskanych wyników do wyników doświadczalnych. Wykonana została w Zakładzie Bioinformatyki i Genomiki pod kierunkiem prof. dr hab. Pawła Mackiewicza oraz dr hab. Pawła Błażeja. Ukazało się sześć publikacji związanych bezpośrednio z pracą doktorską, które są bardzo dobrze cytowane. Dodatkowo ukazało się 10 publikacji spoza rozprawy doktorskiej. Doktorantka aktywna jest też od strony udziału w konferencjach naukowych (5 prezentacji wyników ustnie). Są też posterowe zgłoszenia konferencyjne. Dodatkowo zauważyłem przyznany w 2018 roku grant NCN Preludium 14.

Pierwsze trzy rozdziały pracy doktorskiej to przystępny wstęp, wyraźnie przedstawione cele pracy i etapy badań. Rozdział 4. to przedstawienie użytych metod badawczych, kryteriów optymalizacji kodu genetycznego i opis użytego oprogramowania. Rozdział 5 to przedstawienie wyników badawczych dotyczących optymalizacji kodu ze względu na minimalizację zmian w polarności aminokwasów, wyników optymalizacji ze względu na kilka wybranych cech aminokwasów, porównanie funkcji kosztów dla kodów zoptymalizowanych przy pomocy jednego i więcej kryteriów, wyniki dotyczące właściwości alternatywnych kodów genetycznych i porównania ich z SGC oraz innymi wygenerowanymi w symulacjach komputerowych kodami teoretycznymi. Rozdział 6. to dyskusja wyników przeprowadzonych badań, która podzielona została na trzy podrozdziały pokazujące rolę kodu genetycznego w ograniczaniu efektu mutacji. Są to rozdz. 6.1: Ograniczanie skutków mutacji punktowych poszczególnych pozycji w kodonie z punktu widzenia standardowego kodu genetycznego, rozdz. 6.2: Ograniczenie efektów mutacji ze względu na zmiany wielu własności podstawianych aminokwasów z punktu widzenia standardowego kodu genetycznego, rozdz. 6.3: Własności alternatywnych kodów genetycznych związane z ograniczeniem konsekwencji mutacji punktowych. Tytuły tych rozdziałów precyzyjnie wskazują które właściwości SGC zostały poddane analizie w rozprawie doktorskiej. Rozprawa kończy się Wnioskami, bibliografią, rozdziałem nt. osiągnięć naukowych doktorantki oraz dodatku z tabelami kodów genetycznych znalezionych przy pomocy optymalizacji z użyciem jednego kryterium i szerokiej dokumentacji programu symulacyjnego.

Na początku odniosę się do metod badawczych użytych przy generowaniu teoretycznych kodów genetycznych. Ponieważ różnych teoretycznych kodów genetycznych z jednym warunkiem, żeby każdy z 20 podstawowych aminokwasów i sygnał terminacji translacji były kodowane przez co najmniej jeden kodon jest olbrzymia bo ponad  $10^{84}$  to niemożliwym jest zbadanie właściwości każdego z nich. Dlatego doktorantka ograniczyła się do czterech klas kodów genetycznych i omawiane są właściwości klas tych kodów. Są to: (1) klasa DEG, która zachowuje zarówno strukturę bloków kodonów i degenerację SGC, (2) klasa BLO, jak klasa DEG ale bez zachowania degeneracji SGC, (3) klasa NUM zachowująca liczbę kodonów dla danego aminokwasu, (4) klasa GEN, gdzie jedynym ograniczeniem jest że każdy z 20 aminokwasów i sygnał terminacji translacji są reprezentowane co najmniej przez jeden kodon. Odpowiednie opublikowane prace wspólnie z zespołem promotorów dotyczące klas i metod optymalizacji są też wskazane na liście publikacji. W badaniach korzystano z wersji algorytmu ewolucyjnego o nazwie Strength Pareto Evolutionary Algorithm (SPEA2), który umożliwia optymalizację z wieloma kryteriami. Jest to uznany algorytm a opisane w pracy doktorskiej poszczególne kroki użycia algorytmu i przedstawiona dokumentacja kodu programu przekonują, że algorytm został poprawnie zaimplementowany. W dodatku D2 jest m.in. dokumentacja programów w języku C++ do znajdowania kodów genetycznych klasy BLO optymalnych ze względu na dowolną zadaną liczbę kryteriów. Pierwszym współautorem kodów jest pani mgr Małgorzata Wnętrzak – autorka rozprawy doktorskiej. Drugim autorem jest dr hab. Paweł Błażej.

Przy optymalizacji kodu wykorzystane zostało 8 indeksów aminokwasowych: hydrofobowość, właściwości elektryczne, udział aminokwasów w białkach, skłonności do tworzenia struktur beta, własności wewnętrzne jak potencjał hydracyjny, własności optyczne, elastyczne i inne, własności reszt (masa molekularna, mutabilność, powierzchnia dostępności do rozpuszczalnika i inne) oraz indeks „różne własności fizykochemiczne”. Indeksy były wzięte z bazy Aaindex 1 (publikacja Kawashima, Pokarowski et al. 2008). Ich wybór wskazuje, że użyte w algorytmach ewolucyjnych parametry do badań właściwości fizykochemicznych podstawianych aminokwasów są bardzo miarodajne i pozwalają na ogólniejsze wnioski badawcze.

**Pytanie 1:** Jaki był używany generator liczb losowych?

**Pytanie 2:** Do określenia podobieństw przy przypisaniu aminokwasów do kodonów SGC i kodów zoptymalizowanych użyta była miara odległości między kodami genetycznymi oparta na odległości Hamminga (str. 45 w rozprawie doktorskiej). Prosiłbym o doszczegółowienie jak były zliczane wagi różnic pomiędzy kodami genetycznymi korzystając z tej miary. Pytanie związane jest tylko z tym, czy również liczba sekwencji o zadanej odległości Hamminga wchodzi do algorytmu optymalizującego? Liczba sekwencji o długości  $k$  w odległości Hamminga  $d \leq k$  nie będzie rosła monotonicznie z wartością  $d$ . Na przykład dla alfabetu czteroliterowego A,T,G,C dla  $k=6$  (dwa kodony) oraz  $d=1$  mamy  $N_1=18$  sekwencji, dla  $d=2$  mamy  $N_2=135$  sekwencji, ..., dla  $d=5$  mamy  $N_5=1458$  sekwencji ale dla  $d=6$  już tylko  $N_6=729$  sekwencji.

**Pytanie 3:** Dotyczy algorytmu SPEA2 (schemat na Rys. 4) i redukowania oraz uzupełniania tymczasowego zbioru Pareto w związku z zadaniem rozmiarem archiwum (pkt. 4a i 4b algorytmu). W jaki sposób są redukowane korelacje na zbiorach populacji? Czy populacje są przetasowywane?

Układ rozprawy doktorskiej jest wyważony ale biorąc pod uwagę dyskusję wyników badawczych dotyczącą ewolucji kodu genetycznego można byłoby trochę poszerzyć rozdz. 1.5 pt. „Badanie ewolucji standardowego kodu genetycznego” o więcej szczegółów dotyczących prezentowanych hipotez ale też o informacje na temat obecnego stanu prac badawczych na świecie nad poszerzeniem kodu genetycznego – w jakim kierunku te badania są prowadzone. Tutaj chciałbym zadać ogólniejsze pytanie dotyczące ewolucji SGC.

**Pytanie 4:** Między innymi na str. 10. doktorantka napisała: „Mówimy, że kod ten jest uniwersalny, ponieważ wykorzystywany jest przez przeważającą większość współcześnie żyjących organizmów.” Jak rozumieć to stwierdzenie? Czy to oznacza, że kod genetyczny nie rozwija się, jest „zamrożony”? Jak więc interpretować pojawianie się kodów alternatywnych?

Jeden z pierwszych wyników przedstawionych w rozprawie w rozdz. 5, dotyczy optymalizacji jednokryterialnej związanej z odpornością kodu na zmiany w polarności podstawianych

aminokwasów, które wynikają z mutacji w poszczególnych pozycjach w kodonie i sumy kosztów zmian we wszystkich pozycjach kodonu. Optymalizację przeprowadzono dla czterech klas kodów: DEG, BLO, NUM i GEN. Otrzymane wyniki wskazują, że odporność na mutację w poszczególnych pozycjach kodonu jest zróżnicowana. Pokazano, że SGC jest optymalnym kodem ze względu na skutki mutacji w trzeciej pozycji kodonu ale już jego odporność na skutki mutacji w pierwszej i drugiej pozycji kodonu jest dużo mniejsza. Z drugiej strony przeprowadzone symulacje komputerowe sugerują, że względnie większa odporność na skutki mutacji w pierwszej pozycji kodonu w porównaniu z drugą pozycją może sugerować, że w trakcie kształtowania się kodu genetycznego pierwsza pozycja w kodonie mogła być odczytywana podobnie niejednoznacznie jak trzecia. Jest to bardzo ciekawy wynik. Przeprowadzone przez doktorantkę również inne symulacje jednokryterialne ze względu na potencjał hydracyjny, refrakcję, aktywność optyczną czy elastyczność również wykazują tendencję SGC do ograniczania zmian w cechach aminokwasów podobnie jak w przypadku polarności podstawianych aminokwasów. Inaczej jest z właściwościami elektrycznymi aminokwasów gdzie zmiany podstawianych aminokwasów są maksymalizowane. Z drugiej strony symulacje wielokryterialne wciąż wskazują na tendencję SGC do ograniczania zmian we własnościach aminokwasów. Wyniki symulacyjne pokazują też, że teoretyczne kody genetyczne znalezione przez zastosowane algorytmy ewolucyjne bardziej ograniczają te zmiany niż SGC. Zgadza się tutaj z interpretacją doktorantki i grupy badawczej w której ona pracuje, że jedną z przyczyn jest to że SGC nie wyewoluował tylko ze względu na optymalizację jednej wybranej cechy lub kilku z nich. W końcowej części rozdz. 6.2 doktorantka sugeruje że przyszłe badania optymalizacji kodu genetycznego powinny być wykonane z optymalizacją wielokryterialną z uwzględnieniem włączania aminokwasów do kodu genetycznego w trakcie jego ewolucji.

Badania optymalności SGC ze względu na ograniczanie konsekwencji podstawień punktowych poszerzone zostały również o analizę właściwości alternatywnych kodów genetycznych. Celem badań w tej części doktoratu było zweryfikowanie hipotezy, że SGC znajduje się w lokalnym optimum rozwiązań problemu minimalizacji efektów mutacji i błędnych odczytów kodonów podczas translacji. Otrzymane wyniki sugerują, że alternatywne kody genetyczne ograniczają skutki mutacji punktowych i błędów translacji dużo bardziej efektywnie niż w przypadku SGC i odpowiednio też w symulacjach komputerowych znaleziono dużo mniejszy procent „lepszych” generowanych kodów teoretycznych. Doktorantka napisała (str. 102): „Fakt, że alternatywne kody genetyczne ograniczają skutki zmian w polarności aminokwasów w większym stopniu niż standardowy kod genetyczny i większość rozważanych teoretycznych kodów genetycznych o podobnej strukturze, nie świadczy automatycznie, że powstały one w wyniku selekcji na minimalizację efektów mutacji i błędów w translacji.”

**Pytanie 5:** Jakie znaczenie dla ewolucji kodu może mieć długość genomów?

Główne wnioski będące odpowiedzią na postawione cele badawcze w rozprawie są podane w rozdz. 7.

Podsumowując, rozprawa doktorska spełnia wszystkie wymagania stawiane rozprawom doktorskim. Duża część wyników opublikowana jest w prestiżowych czasopismach i publikacje te są już bardzo dobrze cytowane. Bardzo proszę o dopuszczenie rozprawy doktorskiej do dalszych etapów przewodu doktorskiego.

Biorąc pod uwagę otrzymane wyniki badawcze, bardzo dobrze przemyślaną strukturę pracy i bardzo przystępny sposób przekazu otrzymanych wyników badawczych wnioskuję o nagrodę dla tej rozprawy jako rozprawy wyróżniającej.

*Mirosław Dudka*