



Wydział Farmaceutyczny
z Oddziałem Medycyny Laboratoryjnej
Gdański Uniwersytet Medyczny
Hallera 107, 80-416 Gdańsk

dr hab. Arkadiusz Piotrowski, prof. nadzw. GUMed
email: arkadiusz.piotrowski@gumed.edu.pl
Tel.: 662 02 69 59

GDAŃSKI UNIWERSYTET MEDYCZNY

**Recenzja rozprawy doktorskiej
mgr Edyty Czajkowskiej**

**pt. „Identyfikacja, klonowanie i ekspresja nowego genu *taqIIIRM* z termofilnej bakterii
Thermus aquaticus na podstawie analizy bioinformatycznej sekwencji genomu,
uzyskanej metodami sekwencjonowania nowej generacji”**

Termofilne bakterie *Thermus aquaticus*, a w szczególności izolowane z nich enzymy stanowią szeroko używane narzędzia naukowe w biologii molekularnej, a także w inżynierii genetycznej w przemyśle biotechnologicznym, przede wszystkim ze względu na ich odporność na wysokie temperatury.

Przedstawiona do oceny rozprawa doktorska jest bardzo solidna warsztatowo, ponieważ wpisuje się w rozwojowy obszar badawczy, w którym od wielu lat uznanymi ekspertami są promotor pracy dr hab. Agnieszka Żylicz-Stachula oraz kierownik zespołu badawczego prof. dr hab. Piotr Skowron.

Do recenzji dostarczono rozprawę doktorską liczącą 136 stron łącznie z wykazem piśmiennictwa. Układ pracy jest standardowy, przy czym rozbudowane cele pracy umieszczono na początku przed wstępem teoretycznym, a wyniki połączono z dyskusją. Taki układ jest spotykany w publikacjach naukowych i jak najbardziej dopuszczalny. Od strony redakcyjnej praca została przygotowana bardzo starannie, zawiera jedynie drobne błędy ortograficzne, a także anglicyzmy w kilku miejscach, które nie mają wpływu na ocenę merytoryczną rozprawy. Pod względem graficznym, zdjęcia żeli elektroforetycznych mogłyby zostać przedstawione w wyższej rozdzielczości. Zastosowana rozdzielczość jest wprawdzie wystarczająca do interpretacji, jednak zadanie byłoby łatwiejsze, jeśli zdjęcia byłyby wyższej jakości.

Modelem badawczym jest szczep YT-1 wspomnianego gatunku bakterii *Thermus aquaticus*, a doktorantka odważnie podjęła wyzwanie weryfikacji hipotezy zakładającej istnienie w genomie *T. aquaticus* YT-1 niezidentyfikowanego jeszcze genu, jak później wykazano w niniejszej rozprawie, ortologicznego do wcześniej zidentyfikowanego genu *taqIIIRM*, kodującego enzym o dwóch aktywnościach: metylotransferazy (MTazy) oraz endonukleazy restrykcyjnej (REazy). Hipoteza badawcza została postawiona na podstawie racjonalnej przesłanki, tj. opisanej przez promotorkę niniejszej rozprawy różnicy specyficzności substratowej natywnego oraz rekombinowanego enzymu RM.TaqII. Na podkreślenie zasługuje fakt, że była to hipoteza wysokiego ryzyka. Jak pokazują wyniki recenzowanej rozprawy, warto było podjąć wyzwanie,

ponieważ praca jest klasycznym przykładem udanej realizacji badania typu *high-risk / high-reward*.

Rozprawa zawiera siedem celów głównych. To duża liczba, ale w pełni uzasadniona przy tak szerokim spektrum prac badawczych, od identyfikacji hipotetycznego genu, poprzez sekwencjonowanie i analizę bioinformatyczną genomu, do klonowania, nadekspresji, oczyszczenia i scharakteryzowania sekwencji rozpoznawanej przez endonukleazę restrykcyjną RM.TaqIII.

Wstęp teoretyczny jest napisany przystępnie i ciekawie, stanowiąc konsekwentny potok myślowy. Przy szerokim zakresie tematycznym, ta część rozprawy zajmuje tylko 27 stron, ale zawiera wszystkie potrzebne informacje i świadczy o zdolności doktorantki do przedstawiania zdobytej wiedzy w sposób syntetyczny, a jednocześnie zrozumiały. Doktorantka umiejętnie rozwija wątki tematyczne od kwestii podstawowych po zaawansowane. Autorka rozprawy szczegółowo wyjaśnia skąd wynika odporność m.in. na wysoką temperaturę enzymów izolowanych z organizmów ekstremofilnych i mechanizmy przystosowawcze termofili do życia w wysokich temperaturach, stopniowo zawężając temat do charakterystyki bakterii *Thermus aquaticus*. Analogicznie przedstawiona jest tematyka genomiki i metagenomiki: od kwestii podstawowych po zaawansowane, z następującym zawężeniem do kwestii związanych z *T. aquaticus*, z nawiązaniami do filogenetyki molekularnej rodzaju *Thermus* i płynnym przejściem, poprzez aspekty ewolucji, do horyzontalnego transferu genów i opisu w tym kontekście systemów restrykcyjno-modyfikacyjnych, tj. procesów biologicznych, które stanowią fundament koncepcyjny recenzowanej rozprawy. Wprawdzie w części wstępu, doktorantka wprowadza pojęcia bardzo podstawowe, wręcz podręcznikowe (np. definicja plazmidu), ale z drugiej strony wskazuje to, że zaawansowana wiedza doktorantki osadzona jest na solidnych podstawach.

Mam tylko kilka drobnych uwag do tej części pracy. Opis lokalizowania niescharakteryzowanych genów poprzez analizę sekwencji jest nieco archaiczny. Oczywiście, narzędzia, czy zasoby takie jak ORFfinder, BLAST, czy NCBI GenBank są nadal z powodzeniem używane, jednak warto byłoby wspomnieć nowsze narzędzia i bazy danych ukierunkowane na metody wysokoprzepustowe i automatyzację zadań. Podobną uwagę mam do metod badania transkryptomu. Owszem, metody takie jak SAGE lub konstrukcja bibliotek cDNA i sekwencjonowanie każdego klonu są nadal stosowane, ale stopniowo wypierane przez sekwencjonowanie całego transkryptomu z użyciem wysokoprzepustowego sekwencjonowania równoległego.

Legenda do Ryc. 2.4 zawiera nieprecyzyjne stwierdzenie dotyczące sposobu detekcji sygnału z mikromacierzy: „uzyskuje się określone barwy, które zależne są od intensywności sygnału”. Zasadniczo różne „barwy” w sposób sztuczny są nakładane na monochromatyczne obrazy mikromacierzy przez oprogramowanie do analizy danych, aby umożliwić wizualną, wstępną ocenę jakościową doświadczenia. Różne kolory mogą rzeczywiście podkreślać różnice w intensywności sygnału, ale równie dobrze mogą służyć do zobrazowania sygnału z różnych

kanatów, jeśli w doświadczeniu prowadzi się pomiar względny, z użyciem więcej niż jednego znacznika fluorescencyjnego, o różnych długościach fali wzbudzenia. Intensywność sygnału emisji („świecenia”) to niezależny czynnik umożliwiający przeprowadzanie właściwej analizy ilościowej.

Przedstawione na stronie 17 przykłady zastosowania proteomiki są jedynie wąskim wycinkiem możliwości jakie daje proteomika, tym niemniej są powiązane z dziedziną, której dotyczy rozprawa.

W części dotyczącej genomiki doktorantka mogła bardziej rozwinąć opis rearanzacji genomowych, jako że poza mutacjami punktowymi, są one główną siłą napędową ewolucji (*vide* paralogi) i obejmują swym zakresem znacznie większą część genomu niż sumarycznie mutacje punktowe.

Cześć metodyczna rozprawy jest zasadniczo poprawna, starannie przygotowana, również od strony redakcyjnej – zawiera liczne, przejrzyste i bogate informacyjnie ryciny oraz zestawienia w tabelach. Protokoły laboratoryjne są zwięzłe, ale w zupełności wystarczające. Mam jednak kilka komentarzy. Opis sekwencjonowania w rozdziale 4.3 (a także innych metod w rozdz. 4.4 i 4.5) jest wprowadzeniem teoretycznym w technologię i jako taki powinien znaleźć się w części teoretycznej. Ogólnie, omówienie skoncentrowane jest na aspektach technicznych, obejmujących szczegóły rozwiązań sprzętowych/inżynierskich. To świadczy o zrozumieniu technologii przez doktorantkę. Niedosyt pozostawia jednak opis różnych sposobów fragmentacji genomu, ich zalet i wad, w szczególności, że ten etap procedury ma istotny wpływ na jakość uzyskiwanych wyników.

Z kolei właściwy opis metodyczny analizy danych, tworzenia kontigów i uzyskania pełnej sekwencji DNA uznaję za szczytkowy, a jest to przecież istotny aspekt wykonanej pracy. Co prawda w rozdziale 4.4 doktorantka przedstawiła metody i narzędzia on-line do analizy bioinformatycznej genomu, ale ponownie jest to wprowadzenie teoretyczne. Brakuje szczegółowego opisu procedury analizy bioinformatycznej wykonanej w ramach recenzowanej rozprawy. Jak można wnioskować w dalszej części z opisu wyników, sekwencjonowanie zostało zlecone zewnętrznemu wykonawcy. To wyjaśnia niedostateczny opis metodyki tej części doświadczeń, nie usprawiedliwia jednak jego braku. Nie podano chociażby gdzie i przez kogo zostało wykonane sekwencjonowanie, a taką informację podaje się standardowo w publikacjach naukowych, w przypadku prac zleconych.

W opisie metody nested-PCR (rozdz. 4.5.1) znajduje się nieścisłość. Tak jak stwierdza doktorantka, ta metoda rzeczywiście może służyć zwiększeniu czułości standardowej reakcji PCR, ale przede wszystkim swoistości, a o tym nie wspomniano w tekście.

Ponadto, mam uwagę dotyczącą terminologii, która odnosi się z resztą również do pozostałych części pracy. Doktorantka konsekwentnie używa określenia „sekwencjonowanie nowej generacji” (ang. Next Generation Sequencing). Jest to wprawdzie nazwa używana powszechnie, tym niemniej potoczna. Bardziej właściwe jest np. określenie „wysokoprzepustowe sekwencjonowanie równoległe” (ang. massively parallel sequencing) tym bardziej, że metoda jest znana od ponad dekady, stąd trudno nadal uznawać ją za nową.

Wyniki pracy i dyskusja są bardzo szczegółowo przedstawione i wyczerpująco omówione, z jednym wyjątkiem. Ten aspekt pracy, który dotyczy wysokoprzepustowego sekwencjonowania równoległego genomu na platformie Illumina (we wszystkich jej częściach, tj. metodach, wynikach i dyskusji), oceniam słabiej niż pozostałe elementy rozprawy. Rozumiem, że wykonanie sekwencjonowania i powiązanych analiz bioinformatycznych zostało zlecone wyspecjalizowanemu wykonawcy (zapewne, bo nie podano stosownej informacji) i stanowiło cel o charakterze instrumentalnym, którego realizacja umożliwiła osiągnięcie innych, bardziej wyeksponowanych w rozprawie celów. Jako że cele te zostały osiągnięte, pozostaje założyć, że sekwencjonowanie i powiązana analiza danych zostały wykonane mimo wszystko poprawnie. Tym niemniej, moim zadaniem jest krytyczna ocena rozprawy w oparciu o dostępne informacje, a nie domniemanie.

Zarówno w części metodycznej jak i w wynikach, sposób analizy danych, tworzenia kontigów i uzyskania pełnej sekwencji DNA został zaopatrzony jedynie w szątkowy opis. Podobną uwagę mam do opisu technicznych parametrów sekwencjonowania. Doktorantka nie podała np. długości odczytów, czy sekwencjonowanie zostało wykonane jako *single-end* lub *paired-end*, jaka była głębokość i wydajność sekwencjonowania, oraz istotność statystyczna predykcji przypuszczalnej funkcji zidentyfikowanych genów. To są parametry ważne i standardowo podawane w publikacjach wykorzystujących wysokoprzepustowe sekwencjonowanie równoległe.

Ponadto, nie jest jasne czy wszystkie uzyskane sekwencje zostały zdeponowane w publicznie dostępnych bazach danych, a jest to rutynowa czynność przed publikacją tego typu wyników. Doktorantka podaje, że „uzyskano 108 kontigów o różnej długości”. To bardzo ogólna informacja. Nie wiadomo jakiej wielkości kontigi uzyskano. Wskazane byłoby umieszczenie w rozprawie stosownej tabeli.

Odnośnie tabeli 5.1 „Podsumowanie wyników poszukiwania sekwencji kodujących białka, zlokalizowanych w kontigu jcf180000008186”, w nagłówku bardziej właściwe byłoby użycie określenia „przypuszczalna funkcja kodowanego białka” zamiast „nazwa białka”, ponieważ nie wykonano rzeczywistej, biochemicznej analizy funkcjonalnej. Ponadto, z opisu nie wynika czy sprawdzono potencjalne geny pod kątem występowania w sekwencji kodonów *nonsense*, co wskazywałoby na pseudogeny. Ta tabela jest również zbyt uboga w informacje. Żeby ocenić wartość podanych predykcji przydałoby się chociażby umieszczenie w tabeli większej ilości danych z analiz BLAST (analogiczne uwagę mam do tabel 5.2 i 5.3).

Tym niemniej, zasadniczy cel, tj. identyfikacja przypuszczalnych genów systemu RM został osiągnięty, a negatywny wydźwięk powyższych komentarzy niweluje ponowne wykonanie sekwencjonowania z użyciem technologii SMRT (Pacific Biosciences), we współpracy z renomowanym partnerem, tj. New England Biolabs. W konsekwencji otrzymano sekwencję sześciu megaplazmidów oraz sekwencję DNA chromosomowego. Dane uzyskane z platformy PacBio zostały zdeponowane w NCBI w bazie sekwencji nukleotydowych, tym niemniej podane przez doktorantkę numery dostępu są odnośnikami do sekwencji tylko dwóch plazmidów. Czy podjęto próbę złożenia szkicu sekwencji całego, kompletnego genomu? Taka informacja

znajduje się na s. 83 rozprawy, jednak nie podano odnośników do stosownych sekwencji nukleotydowych.

Kolejne wyniki, wraz z dyskusją, nie budzą już większych zastrzeżeń i są bardzo dobrej jakości – demonstrują ogrom wykonanej przez doktorantkę pracy, skrupulatność i głęboki wgląd w tematykę badawczą.

Wysoko oceniam pełne wykorzystanie potencjału technicznego metody SMRT, w znaczeniu analizy modyfikacji DNA (tu: metylacji DNA). Ta analiza dostarczyła pośrednich dowodów na istnienie aż sześciu funkcjonalnych metylotransferaz lub systemów RM w badanym szczepie bakteryjnym, z których to systemów wcześniej poznano tylko trzy. W tym kontekście mam następujące pytanie: czy podjęto próbę szerszego scharakteryzowania pozostałych trzech systemów RM?

Doceniam również ostrożne, krytyczne podejście doktorantki do własnych wyników. W szczególności, na podstawie wyników sekwencjonowania na platformie PacBio, doktorantka wykryła kontaminację próby innym gatunkiem bakterii i z powodzeniem podjęła wieloetapowe działania zaradcze. Analogicznie, podczas klonowania genu *taqIIIRM*, ze względu na bardzo wysokie podobieństwo do sekwencji *taqIIRM* doktorantka wdrożyła kontrolę jakości przygotowanej wstawki, aby wykluczyć ryzyko kontaminacji niewłaściwym fragmentem DNA. To wszystko w mojej ocenie świadczy o dojrzałości naukowej doktorantki.

Podobnie, na zaradność doktorantki wskazuje zastosowanie prostego, aczkolwiek skutecznego rozwiązania problemu podobieństwa sekwencji genów *taqIIRM* i *taqIIIRM* uniemożliwiającego swoistą amplifikację indywidualnej sekwencji każdego z tych genów w standardowej reakcji PCR. W takiej sytuacji, można by uciekać się do złożonych, zaawansowanych i przez to kosztownych, czasochłonnych metod, jednak wystarczające okazało się zastosowanie *nested-PCR* po odpowiedniej optymalizacji warunków reakcji.

Pytania i uwagi odnośnie rozdziałów elektroforetycznych:

- 1) Fot. 5.8, panel, ścieżka 7: brak interpretacji czwartego, dodatkowego prążka (o najwyższej MW), którego nie przewiduje teoretyczny wzór trawienia – proszę o wyjaśnienie.
- 2) Fot. 5.9, ścieżki 1-4 i 8: brak komentarza do produktu w ścieżce nr 2 (klon negatywny), tym bardziej, że w innych klonach negatywnych ten produkt (być może niespecyficzny?) nie występuje – proszę o wyjaśnienie.
- 3) Fot. 5.12, ścieżka K - białko RM.TaqII: czy widoczny podwójny prążek – może wynikać z niedoczyszczenia wzorca, błędu rozdziału lub naniesienia na żel, czy z innej przyczyny?
- 4) Fot. 5.19: Uważam, że to doświadczenie powinno zostać powtórzone, ze względu na słaby, w zasadzie nieudany rozdział elektroforetyczny w szczególności większych fragmentów (np. nietrawiony produkt PCR powinien być widoczny jako pojedynczy, ostry prążek). Czy stężenie i wielkość żelu, oraz warunki i parametry rozdziału zostały właściwie dobrane?

- 5) Fot. 5.20: Nieoptymalny rozdział; przydałby się wzór trawienia *in silico*, tak jak w przypadku innych rozdziałów zaprezentowanych w pracy. Ponadto, nie zostało wyjaśnione niecałkowite trawienie substratu. Proszę o komentarz.

Ponadto, mam jeszcze poniższe drobne uwagi krytyczne, które jednak nie wynikają z istotnych niedostatków pracy, a raczej z braku precyzji w przekazie myśli naukowych i niewielkich potknięć redakcyjnych.

Na stronie 89 rozprawy znajduje się nieprecyzyjnie sformułowane zdanie: „Dzięki zjawisku HGT i rearanzacji genomu bakterie z rodzaju *Thermus* posiadają tzw. plastyczny genom, który umożliwia im bytowanie w warunkach wysokotemperaturowych”. Uważam, że bardziej właściwym sformułowaniem byłoby np.: „HGT i rearanzacje genomu są mechanizmami ewolucyjnymi zapewniającymi plastyczność genomu i w konsekwencji wykształcenie przystosowania do warunków wysokotemperaturowych”.

Przy wielostronicowych rycinach wskazane byłoby powtórzenie legendy na poszczególnych stronach. Niektóre legendy są również niekompletne np. do rys. 5.4, jednak zawartość ryciny jest mimo to zrozumiała.

Opis powiększeń mikroskopii elektronowej w głównym tekście nie jest w pełni zgodny z legendami do Fot. 5.2 - 5.5. Ponadto, proszę o odpowiedź na pytanie czy poszczególne panele fotografii z mikroskopów elektronowych (np. Fot. 5.2) przedstawiają ten sam obszar pod różnym powiększeniem czy różne obszary preparatu? Niezależnie od odpowiedzi, to powinno być zaznaczone bezpośrednio na fotografiach lub w legendzie.

Proszę również o odpowiedź na pytanie, czy artefakty widoczne na Fot. 5.5 (TEM) mogą być efektem niewłaściwego utrwalenia preparatu i w konsekwencji obkurczenia cytozolu?

Ostatnią częścią rozprawy jest „podsumowanie”, które ze względu na w dużej mierze techniczny i opisowy charakter wykonanych badań, nie zawiera typowych wniosków, tylko listę wykonanych zadań. Podsumowanie dobrze współgra z listą siedmiu ambitnych celów postawionych w pracy, które zostały z powodzeniem osiągnięte. Jedyne zastrzeżenie mam do sposobu w jaki został sformułowany i zrealizowany cel dotyczący analizy bioinformatycznej genomu. Uważam, że doktorantka powinna była ograniczyć tę część pracy do wybranych megaplazmidów zawierających geny badanych systemów restrykcyjno-modyfikacyjnych. Tym bardziej, że powiązany cel miał charakter instrumentalny dla innych, nadrzędnych celów pracy. W ten sposób doktorantka uniknęłaby niedostatków opisanych w niniejszej recenzji. Tym niemniej, jest to jedyna główna uwaga o negatywnym wydźwięku na temat pracy bardzo dobrej pod innymi względami.

Ponadto, wyniki badań zostały opublikowane w *Nucleic Acids Research*. Doktoranta jest wprawdzie trzecim autorem tego raportu badawczego, ale jest to bardzo dobra praca, opublikowana w prestiżowym czasopiśmie, stąd zapewne efekt oddziaływania tej publikacji będzie duży i trwały.

W świetle wyżej przedstawionej, pozytywnej oceny recenzowanej rozprawy doktorskiej, stwierdzam, że zostały spełnione wymogi formalne i merytoryczne stawiane rozprawom doktorskim, zawarte w ustawie o stopniach naukowych i tytule naukowym. Wnoszę do Rady Wydziału Chemii Uniwersytetu Gdańskiego o dopuszczenie mgr Edyty Czajkowskiej do dalszych etapów przewodu doktorskiego.



dr hab. Arkadiusz Piotrowski, prof. nadzw. GUMed

Gdańsk, 24 marca 2019 roku