



UNIwersytet
Warszawski

Wydział Biologii
Instytut Genetyki i Biotechnologii
prof. dr hab. Paweł Golik



Warszawa, 10.11.2023

Recenzja rozprawy doktorskiej mgr Mileny Stolarskiej pt. „Zastosowanie metod filogenetycznych do badania pochodzenia, funkcji oraz oddziaływań pomiędzy białkami systemów opiekuńczych Hsp70/JDP”

Białka opiekuńcze i towarzyszące im białka pomocnicze stanowią jeden podstawowych mechanizmów zapewniających prawidłowe funkcjonowanie proteomu. Pod wieloma względami szczególnie interesujące są systemy oparte na białkach klasy Hsp70 i towarzyszących im białkach z domeną J (JDP). Systemy te obecne są we wszystkich znanych gałęziach drzewa życia, wykazując przy tym duże zróżnicowanie będące rezultatem licznych duplikacji, które doprowadziły do powstania szeregu różnorodnych form dostosowanych do pełnienia różnych funkcji w różnych przedziałach komórki. Recenzowana praca ogólnie podejmuje się rekonstrukcji różnych aspektów ewolucji białek systemu Hsp70/JDP wykorzystując dostępne bazy sekwencji aminokwasowych i różnorodne metody analizy filogenetycznej. Ze względu na dużą liczbę różnych białek z tej grupy i ich znaczne zróżnicowanie, jest to zagadnienie niełatwe i dalekie od pełnego zrozumienia. Praca spełnia zatem najważniejsze kryterium stawiane nie tylko przed rozprawami doktorskimi, ale przed każdym przedsięwzięciem naukowym – podejmuje się rozwiązania naprawdę istotnego poznawczo problemu naukowego.

Praca ma formę tradycyjnej rozprawy i została podzielona na trzy części, z których każda dotyczy odrębnego zagadnienia w obrębie szerokiego problemu, jakim jest ewolucja systemu Hsp70/JDP. Zostały one poprzedzone wspólnym, ogólnym wstępem, natomiast każda z trzech części posiada odrębną strukturę pracy badawczej, z wydzielonym bardziej szczegółowym wstępem, opisem metod, wynikami oraz dyskusją. Z jednej strony ułatwia to śledzenie (i tak niekiedy złożonego) rozumowania w każdej z analiz, z drugiej prowadzi jednak do występowania pewnych powtórzeń, zwłaszcza w opisie metod. Poza zamieszczonym na początku krótkim streszczeniem brakuje też wspólnej dyskusji, w której można by zawrzeć ogólne wnioski ze wszystkich trzech szczegółowych analiz.

Jak w każdym złożonym tekście, tak i tu nie udało się uniknąć pewnej liczby błędów edytorskich i niezręczności językowych, z obowiązku wymienię przykładowo „DANJA3” zamiast DNAJA3 (s. 19), "sekwencji" zamiast "sekwencje" i "zestawieni" zamiast "zestawienie (s. 26), „wheited” zamiast

weighted (dwukrotnie na s. 29), „eukarityczne” zamiast eukariotyczne (s. 35), „Wyniki badań biochemicznych [69], że trzy reszty lizyny są niezbędne [...]” - brak orzeczenia (s. 53), „JPD” zamiast JDP (s. 54), „protistita” zamiast Protista (s. 61). Odniosłem też wrażenie, że prezentacja graficzna wyników, zwłaszcza w postaci drzew filogenetycznych w drugiej i trzeciej części jest bardziej staranna i przemyślana, niż w części pierwszej, prawdopodobnie ze względu na to, że wyniki tych części zostały przygotowane do publikacji i opublikowane. Mimo tych drobnych uwag, nie mam wątpliwości, że struktura i forma rozprawy spełniają wszystkie wymagania stawiane pracom naukowym.

Pierwsza, najobszerniejsza część pracy zawiera bardzo interesujące i (na ile mogłem stwierdzić) jeszcze nieopublikowane analizy dotyczące pokrewieństw prokariotycznych i eukariotycznych białek JDP. Białka te dzielone są ze względu na strukturę domenową na dwie główne klasy: klasę A, której prototypem jest bakteryjne DnaJ, oraz klasę B, której prototypem jest bakteryjne CbpA. Na tej podstawie można było spodziewać się, że białka eukariotyczne zaliczane do klasy A powinny wywodzić się od DnaJ, natomiast te o budowie typowej dla klasy B – od CbpA. Tymczasem bardzo staranne i wnikliwe analizy filogenetyczne przeprowadzone przez Doktorantkę całkowicie odrzucają tę hipotezę i wspierają alternatywny, nieoczekiwany scenariusz: ortologi CbpA występują jedynie u bakterii, natomiast wszystkie eukariotyczne JDP, niezależnie od ich budowy domenowej są bliżej spokrewnione z DnaJ. W tym miejscu pozostaje mi tylko Doktorantce pogratulować – uzyskanie zupełnie nieoczekiwanego rezultatu, czyli *bona fide* odkrycie, nie zdarza się w pracy naukowej aż tak często. Oczywiście takie nieoczekiwane odkrycie wymaga starannej weryfikacji. Autorka stanęła tu na wysokości zadania, przeprowadzając analizy szeregiem różnych metod filogenetycznych, za każdym razem wspierających odkryte przez nią zależności.

Lektura tej części pracy nasunęła mi nieco uwag i pytań:

- Analiza obejmuje 725 proteomów bakterii, archeonów, grzybów, zwierząt oraz roślin. Dla lepszego odtworzenia filogenezy białek eukariotycznych cenne byłoby włączenie do analizy sekwencji pochodzących z różnych bazalnych linii jednokomórkowych Eukaryota, takich jak bliskie Opisthokonta grupy Breviata, Apusomonada i Amoebozoa, a także odległe od Opisthokonta i roślin kłady SAR czy CRuMS. Cenne byłyby też sekwencje archeonów z grupy Lokiarchaeota, najbliższej gospodarzowi endosymbiozy, która dała początek eukariontom. Sekwencje takie są dostępne, jednak zwykle w postaci sekwencji genomów (DNA), w których trzeba odnaleźć geny kodujące interesujące białka (adnotacja białek nie nadąża za sekwencjonowaniem kolejnych genomów).
- W jaki sposób identyfikowano sekwencje białek JDP o lokalizacji mitochondrialnej? Czy na podstawie występowania w sekwencji sygnału kierującego do mitochondriów, na

podstawie homologii ze znanymi JDP mitochondrialnymi, czy też na podstawie adnotacji w bazie danych (która u organizmów niemodelowych jest zapewne oparta na ortologii do białek mitochondrialnych z organizmów modelowych)?

- Autorka podaje, że pierwsze zestawienie liczyło 1588 sekwencji (s. 27), brak natomiast informacji o tym, ile sekwencji trafiło do kolejnych zestawień (A/B-B' i C-B'/ER).
- Drzewa filogenetyczne przedstawione na ryc. 3.5-3.7 są mało czytelne, ponieważ nie da się na nich stwierdzić, jakie sekwencje odpowiadają poszczególnym gałęziom. Oczywiście przy takiej liczbie sekwencji niemożliwe byłoby dodanie etykiet do każdej gałęzi drzewa (rozwiązaniem byłoby udostępnienie go w formie elektronicznej jako suplementu), ale może dałoby się połączyć sekwencje z bliskich sobie organizmów w jednostki wyższego rzędu i je oznaczyć tak, jak w pracy zrobiono to w dalszych częściach, np. na Ryc. 4.1, 4.2 i 5.4.
- Autorka stwierdza, że „Kład mitochondrialnych JDP znajduje się na politomii i nie można na tej podstawie stwierdzić, z którą grupą bakteryjnych DnaJ mają one najbliższego wspólnego przodka”. Jest jednak w drzewie kilka grup sekwencji bliższych kładowi mitochondrialnych JDP niż pozostałe, ale wspomniany powyżej brak jakichkolwiek etykiet uniemożliwia stwierdzenie, jakie to są białka. Czy są wśród nich np. ortologi DnaJ z Alphaproteobacteria? A jeżeli nie, to gdzie na drzewie znajdują się te, tak istotne dla pochodzenia mitochondriów sekwencje? Czy Alphaproteobacteria w ogóle znalazły się w zestawieniu – znowu elektroniczny suplement pozwolił by na obejście ograniczeń formy drukowanej i zamieszczenie pełnej listy sekwencji, które zostały analizowane na drzewie. Pochodzenie białek wchodzących w skład proteomu mitochondriów było analizowane w literaturze (Pittis i Gabaldon, 2016) – jestem ciekaw, czy w tej analizie można znaleźć jakieś wskazówki odnośnie pochodzenia mitochondrialnych JDP.

Następnie Autorka skoncentrowała się na JDP funkcjonujących w cytoplazmie i ER, przedstawiając prawdopodobny scenariusz duplikacji, które dały początek repertuarowi roślinnych, drożdżowych i zwierzęcych białek JDP. Ta analiza także przyniosła bardzo cenne i ciekawe rezultaty i w związku z tym nasunęła kilka uwag i pytań:

- Ryc. 3.8 (zwłaszcza panel B) byłaby czytelniejsza, gdyby zaznaczono nie tylko podział na klasę A i B, ale wyodrębniono też graficznie podklasę B'. To samo dotyczy Ryc. 3.12.
- Jak już wskazałem powyżej, więcej informacji i dokładniejsze odtworzenie scenariuszy ewolucyjnych można by uzyskać włączając sekwencje lepiej próbujące różnorodność linii ewolucyjnych Eukaryota (bazalne jednokomórkowe eukarionty, np. kłady SAR i CRuMS). Podobnie, uwzględniając różnorodność Metazoa można by stwierdzić, kiedy

doszło do duplikacji, które dały w rezultacie cztery ludzkie białka klasy B. Czy np. miały one związek z całogenomowymi duplikacjami, które zaszły w ewolucji kręgowców (hipoteza 2R) – tu pomocny byłby np. znany genom lancetnika.

- Na s. 41 w tekście mowa jest o tym, że „sekwencje ludzkie DNAJB13 [...] oraz roślinne ERDJ3 tworzą siostrzane grupy monofiletyczne”, tymczasem na rycinie 3.8 siostrzana dla ERDJ3 jest DNAJB11.

W dalszych analizach Doktorantka wykazała, że usunięcie sekwencji podklasy B' zwiększyło statystyczne wsparcie drzewa JDP nie zmieniając w istotny sposób jego topologii. Podobnie, różnice w topologii drzew Bayesowskich (na mniejszej, ale nie podano jakiej, liczbie białek) względem drzewa ML LG+I+G4 nie są znaczące i nie dotyczą kluczowych dla tej analizy węzłów drzewa wskazujących na niezależne pochodzenie JDP klasy B w cytoplazmie i ER od białek klasy A.

Następnie podjęto próbę sprawdzenia na jakim etapie ewolucji białka klasy B uzyskały dodatkowe miejsce wiązania Hsp70 poprzez rekonstrukcję sekwencji ancestralnych. Okazuje się, że dwie z trzech takich reszt najprawdopodobniej występowały już u ostatniego wspólnego przodka cytoplazmatycznych JDP klas A i B, zaś trzecia występuje tylko u drożdży (możliwe, że u innych organizmów w wiązaniu bierze udział sąsiednia reszta).

W podsumowującej tę część pracy dyskusji Doktorantka wyczerpująco i kompetentnie przedstawia możliwe ewolucyjne scenariusze, które mogą wyjaśnić uzyskane rezultaty pokrewieństw. Znowu (co potwierdza, jak interesujące i inspirujące są to odkrycia), mam do tej dyskusji kilka pytań i uwag:

- Na s. 55 Autorka pisze o „genach plastydowych”, mając zapewne na myśli geny jądrowe kodujące białka plastydowe (sformułowanie „geny plastydowe” sugeruje raczej geny kodowane w genomie plastydowym).
- Wynik sugerujący pochodzenie plastydowych JDP od Chloroflexi a nie Cyanobacteria jest nieoczekiwany i zasługuje na pełniejsze przedyskutowanie. Znowu, brak jakichkolwiek oznaczeń na drzewach nie pozwala stwierdzić, jak mają się do nich homologi DnaJ pochodzące z Cyanobacteria (gdzie są na drzewie).
- Relacje białek DNAJB13, ERDJ3 i Scj1 nie są zgodne z drzewem gatunków. Autorka przedstawia interesujący scenariusz wczesnej duplikacji (na jakim etapie – ostatniego wspólnego przodka Eukaryota (LECA)?) i późniejszej utraty. Ponownie, wiele wyjaśniłoby włączenie do analiz innych niż rośliny i Opisthokonta gałęzi eukariotycznych.
- Podobnie, analiza występującej tylko u Metazoa podklasy B' zyskałaby po włączeniu większej liczby sekwencji z różnych przedstawicieli tej grupy, a także z grup siostrzanych,

takich jak Choanoflagellata czy Filasterea (mogło by to też rozwiązać problem artefaktów przyciągania długich gałęzi).

- Dyskusja o sekwencjach pochodzących od archeonów bardzo zyskałaby na uwzględnieniu wspomnianych już Lokiarchaeota, których genomy zawierają wiele genów nieobecnych u innych Archaea a typowych dla eukariontów.
- Ogólna uwaga i hipoteza: skoro struktura domenowa białek JDP nie odzwierciedla ich filogenezy, czy uzyskanych wyników nie można wyjaśnić tasowaniem domen? Tu można by zobaczyć, czy drzewa uzyskane dla poszczególnych domen odpowiadają drzewom całych białek, a także przyrzeć się temu, jak domeny w białkach eukariotycznych korelują z eksonami w kodujących je genach.

Nie chciałbym, aby tak duża liczba uwag i pytań do pierwszej części przedstawionej pracy została potraktowana jako jej krytyka. Przeciwnie, wynika ona z tego, że uzyskane wyniki są nieoczekiwane, i przez to inspirujące i zachęcające do dalszych dociekań. Przedstawione wyniki mogą zainspirować jeszcze wiele analiz, pozwoliłem sobie jedynie na kilka sugestii co do ich możliwych kierunków.

Kolejne dwie części pracy weszły w skład dwóch wieloautorskich publikacji. W drugiej części zawarta jest analiza pochodzenia ewolucyjnego białek Hsp70, w szczególności tych, które są zaangażowane w syntezę grup FeS, u eukariontów zlokalizowaną w mitochondriach. Wyniki analizy potwierdziły hipotezę, że białko mtHsp70, będące wielofunkcyjnym mitochondrialnym Hsp70 pochodzi od bakteryjnego DnaK, a bakteryjne Hsp70 wyspecjalizowane w syntezie FeS (HscA) nie występuje u eukariontów. Wyspecjalizowane w biogenezie Hsp70 występujące u drożdży powstało w wyniku duplikacji genu kodującego ogólne Hsp70 z linii DnaK. Ten wynik nie jest aż tak nieoczekiwany, jak przedstawione w pierwszej części, ale wciąż bardzo interesujący. Do tej części mam następujące pytania i uwagi:

- Na drzewie z Ryc. 4.1 dwa białka DnaK i HscA występują u przedstawicieli Bacteria, u eukariontów brak homologów HscA. Jak wygląda sytuacja u przedstawicieli Archaea? Ta informacja pozwoli na nakreślenie hipotezy odnośnie tego, czy wyspecjalizowane w syntezie FeS białko typu HscA istniało u LUCA i zostało utracone w ewolucji eukariontów (a może już ich przodków z Lokiarchaeota), czy może powstało wyłącznie w linii Bacteria?
- W sekcji 4.2.2 (s. 61) Autorka pisze, że wykorzystwała m. in. proteomy „38 gatunków pierwotniaków (protista)”. Jednokomórkowe eukarionty zwyczajowo nazywane protistami reprezentują bardzo różne linie ewolucyjne, często bardziej odległe od siebie niż zwierzęta i rośliny, nie można więc, zwłaszcza w analizach filogenetycznych, traktować ich jako jednej grupy. Do jakich gałęzi drzewa eukariontów należały te

organizmy i gdzie są na drzewie zamieszczonym na Ryc. 4.1? Z Tabeli 2 wynika, że u badanych protistów nie ma homologów HscA, ale co z homologami DnaK, dlaczego nie ma ich na drzewie?

- Duplikację, która doprowadziła do powstania wyspecjalizowanego paraloga Ssq1 można by precyzyjniej umieścić wykorzystując to, że znane jest już ponad 330 genomów reprezentujących różne linie drożdży pączkujących (Shen et al., 2018).

W ostatniej, trzeciej części przedstawiona jest analiza koewolucji reszt aminokwasowych odpowiadających za oddziaływania pomiędzy domeną J białka JDP a białkiem Hsp70. Doktorantka najpierw przetestowała hipotezę zerową (brak koewolucji) i wykazała, że dane filogenetyczne wskazują na koewolucję w obrębie 16 wysoce zmiennych par reszt aminokwasowych. Na wyróżnienie w tej części zasługuje interesująca hipoteza wyjaśniająca zaobserwowaną zmienność, przy zachowanych oddziaływaniach międzybiałkowych.

W podsumowaniu stwierdzam, że przedstawiona mi do recenzji praca doktorska Pani mgr Mileny Stolarskiej pt. „Zastosowanie metod filogenetycznych do badania pochodzenia, funkcji oraz oddziaływań pomiędzy białkami systemów opiekuńczych Hsp70/JDP” stanowi oryginalne rozwiązanie istotnego problemu naukowego i potwierdza ogólną wiedzę teoretyczną Kandydatki oraz umiejętność samodzielnego prowadzenia pracy naukowej. Do rozprawy doktorskiej dołączone zostały streszczenia w języku polskim i angielskim. Na tej podstawie stwierdzam, że wszystkie wymagania prawne zapisane w Ustawie z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668 z późn. zm.) stawiane rozprawom doktorskim i Kandydatom ubiegającym się o nadania stopnia doktora zostały spełnione i może stanowić podstawę do nadania stopnia doktora. Zwracam się zatem do Rady Dyscypliny Biotechnologia Uniwersytetu Gdańskiego o dopuszczenie mgr Mileny Stolarskiej do dalszych etapów postępowania o nadanie stopnia doktora w dziedzinie nauk ścisłych i przyrodniczych w dyscyplinie biotechnologia.

prof. dr hab. Paweł Golik