



Międzyuczelniany Wydział  
Biotechnologii

Uniwersytetu Gdańskiego  
i Gdańskiego Uniwersytetu Medycznego

---

## ROZPRAWA DOKTORSKA

---

*mgr Milena Stolarska*

# Zastosowanie metod filogenetycznych do badania pochodzenia, funkcji oraz oddziaływań pomiędzy białkami systemów opiekuńczych Hsp70/JDP

Phylogenetic approach to study the origin, function and  
interactions between proteins of JDP/Hsp70 systems

Praca przedstawiona  
Radzie Dyscypliny Biotechnologia Uniwersytetu Gdańskiego  
celem uzyskania stopnia doktora  
w dziedzinie nauk ścisłych i przyrodniczych  
w dyscyplinie naukowej biotechnologia

Promotor: prof. dr hab. Jarosław Marszałek  
Promotor pomocniczy: dr inż. Bartłomiej Tomiczek  
Zakład Biochemii Ewolucyjnej

GDAŃSK 2023

## OŚWIADCZENIE

Ja, niżej podpisana oświadczam, iż przedłożona praca dyplomowa została wykonana przeze mnie samodzielnie, nie narusza praw autorskich, interesów prawnych i materialnych innych osób.

.....  
data

.....  
własnoręczny podpis

Chciałabym serdecznie podziękować:

**Profesorowi Jarosławowi Marszałkowi**  
*za nieocenioną opiekę mentorską, cenne uwagi oraz stworzenie możliwości rozwoju.*

**dr inż. Bartłomiejowi Tomiczkowi**  
*za całą przekazaną wiedzę, pasję i wsparcie.*

**Zespołowi Zakładu Biochemii Ewolucyjnej**  
*za pomoc w realizacji badań i przyjemną atmosferę pracy,*  
*a w szczególności **Anecie Grabińskiej-Rogała***  
*za pomoc, na którą zawsze mogłam liczyć i wyjątkową przyjaźń, która przetrwała całą*  
*doktorancką podróż.*

**Moim przyjaciołom: Jeroenowi, Kaji, Michalinie i Izie**  
*na których zawsze mogę polegać, za wyrozumiałość, wiarę w moje możliwości i*  
*zarażanie mnie radością nawet w ciężkich chwilach.*

**Pracę dedykuję moim Rodzicom Ewie i Bogusławowi**  
*którzy byli dla mnie nieocenionym wsparciem i wiem, że byliby ze mnie bardzo dumni.*

Badania wykonano w ramach  
programu TEAM Fundacji na Rzecz Nauki Polskiej  
„Cykl wiązania substratu przez białka opiekuńcze Hsp70/J: molekularne  
mechanizmy i funkcjonalne konsekwencje”

nr POIR.04.04.00-00-4114 /17-00



oraz

grantu Narodowego Centrum Nauki - Opus 21

„Pochodzenie aktywności dezagregacji amyloidów w rodzinie białek  
pomocniczych J z podwójna baryłka beta.”

nr 2021/41/B/NZ8/02835



Część badań przedstawionych w poniższej pracy została opublikowana w:

Tomiczek, B., Delewski, W., Nierzwicki, L., **Stolarska, M.**, Grochowina, I., Schilke, B., Dutkiewicz, R., Uzarska, M. A., Ciesielski, S. J., Czub, J., Craig, E. A., & Marszalek, J. (2020). Two-step mechanism of J-domain action in driving Hsp70 function. *PLoS computational biology*, 16(6), e1007913.

Kleczewska, M., Grabinska, A., Jelen, M., **Stolarska, M.**, Schilke, B., Marszalek, J., Craig, E. A., & Dutkiewicz, R. (2020). Biochemical Convergence of Mitochondrial Hsp70 System Specialized in Iron-Sulfur Cluster Biogenesis. *International journal of molecular sciences*, 21(9), 3326.



# Spis treści

Streszczenie .....	7
Summary .....	8
1. Wstęp .....	9
2. Cel pracy.....	16
3. Ewolucja eukariotycznych JDP spokrewnionych z bakteryjnym DnaJ. ....	17
3.1 Wstęp .....	17
3.1.1 Klasyfikacja JDP .....	19
3.1.2 Oddziaływanie JDP należących do klas A i B z Hsp70 .....	22
3.2 Metody .....	26
3.2.1 Przygotowanie zbioru danych do analiz filogenetycznych JDP należących do klas A i B .....	26
3.2.2 Rekonstrukcja drzewa filogenetycznego JDP .....	28
3.2.3 Rekonstrukcja sekwencji przodków .....	30
3.3 Wyniki .....	30
3.3.1 Hipotetyczny scenariusz ewolucji JDP zgodny z podziałem na klasy A i B. ....	30
3.3.2 Analiza filogenetyczna JDP klasy A i B .....	31
3.3.3 Pochodzenie ewolucyjne eukariotycznych JDP klasy A i B .....	32
3.3.5 Analiza filogenetyczna ML JDP bez białek klasy B' .....	46
3.3.6 Analiza Bayesowska.....	47
3.3.7 Ewolucja reszt aminokwasowych zaangażowanych w oddziaływanie JDP klasy B z sekwencją EEVD Hsp70. ....	50
3.4 Podsumowanie i dyskusja .....	54
4. Ewolucja systemów JDP/Hsp70 wyspecjalizowanych w biogenezie centrów FeS	59
4.1 Wstęp .....	59
4.1.1 Centra żelazo-siarkowe .....	59
4.2 Metody .....	61
4.2.1 Rekonstrukcja drzewa filogenetycznego Hsp70 zaangażowanych w biogenezę centrów FeS metodą ML .....	61
4.2.2 Dystrybucja filogenetyczna bakteryjnych białek HscA oraz HscB .....	61
4.3 Wyniki .....	62
4.3.1 Analiza filogenetyczna Hsp70 zaangażowanych w proces biogenezy centrów FeS .....	62
4.3.2 Dystrybucja filogenetyczna HscA oraz HscB .....	65
4.4 Dyskusja .....	67
5. Ewolucja miejsca oddziaływania domeny-J JDP z partnerskim Hsp70 .....	70
5.1 Wstęp .....	70
5.2 Metody .....	71
5.2.1 Rekonstrukcja filogenezy Hsp70.....	71

5.2.2	Analiza koewolucji reszt aminokwasowych zaangażowanych w oddziaływanie domena-J/Hsp70.....	72
5.2.3	Konserwacja ewolucyjna aminokwasów zaangażowanych w oddziaływanie domena-J/Hsp70 .....	73
<b>5.3</b>	<b>Wyniki .....</b>	<b>74</b>
5.3.1	Ewolucja miejsca oddziaływania domena-J /Hsp70 .....	74
5.3.2	Koewolucja oddziaływania domena-J/Hsp70 .....	77
<b>5.4</b>	<b>Dyskusja .....</b>	<b>84</b>
<b>6.</b>	<b>Bibliografia.....</b>	<b>86</b>

## Streszczenie

Systemy JDP/Hsp70 składają się z białka opiekuńczego Hsp70 oraz białek posiadających domenę-J (JDP) są kluczowymi elementami proteostazy komórkowej pełniącymi ważne funkcje zarówno w warunkach fizjologicznych, jak i podczas stresu lub patologii. Funkcje systemów JDP/Hsp70 obejmują m.in. fałdowanie nowo syntetyzowanych białek, transport polipeptydów przez błony, a także zapobieganie agregacji oraz dezagregacja agregatów białkowych, chroniąc w ten sposób komórki przed szkodliwymi efektami stresu i starzenia. Funkcje JDP/Hsp70 warunkowane są przez cykliczne wiązanie i uwalnianie substratów białkowych, kontrolowane przez wiązanie i hydrolizę ATP. JDP dostarczają substrat do partnerskiego Hsp70 lub kierują Hsp70 do określonej lokalizacji warunkując specyficzność substratową systemu. W niniejszej pracy przedstawiam wyniki zastosowania metod filogenetycznych takich jak rekonstrukcja drzew filogenetycznych, analiza koewolucji oraz rekonstrukcja sekwencji przodków do badania pochodzenia, funkcji i oddziaływań pomiędzy białkami wchodzącymi w skład systemów JDP/Hsp70.

JDP zostały podzielone na trzy klasy (A, B, C) ze względu na podobieństwo strukturalne do bakteryjnego JDP DnaJ z *Escherichia coli*; Białka klasy B mają unikatową zdolność zapobiegania agregacji amyloidów.

W pierwszej części pracy analizuję pokrewieństwa filogenetyczne JDP klasy A i B z 725 prokariotycznych i eukariotycznych proteomów. Uzyskane wyniki wspierają scenariusz ewolucyjny, w którym wszystkie eukariotyczne JDP zarówno klasy A jak i klasy B są spokrewnione z bakteryjnym DnaJ (klasa A). JDP klasy B ewoluowały niezależnie w wyniku duplikacji genów kodujących DnaJ (klasa A) u bakterii, w cytoplazmie komórki eukariotycznej i w retikulum endoplazmatycznym zwierząt. Wyniki te pokazują, że klasyfikacja JDP nie odzwierciedla ich zależności ewolucyjnych i powinna być zrewidowana.

W drugiej części analizuję pochodzenie ewolucyjne mitochondrialnego mtHsp70 zaangażowanego wraz z białkiem JDP (Hsc20) w proces biogenezy centrów żelazo siarkowych (FeS) - grup prostetycznych niezbędnych do funkcjonowania wielu białek. Moje wyniki pokazują, że mtHsp70 pochodzi od bakteryjnego wielofunkcyjnego Hsp70 (DnaK), które u bakterii nie funkcjonuje w procesie biogenezy FeS. Pokazują także, że bakteryjne Hsp70 wyspecjalizowane w biogenezie FeS (HscA) nie występuje w proteomach eukariotycznych sugerując, że HscA zostało utracone w procesie ewolucji mitochondriów i zastąpione w procesie biogenezy FeS przez potomka DnaK.

W trzeciej części pokazuję, że specyficzność oddziaływania pomiędzy domeną J białka JDP a partnerskim Hsp70 uwarunkowana jest przez koewolucję oddziałujących reszt aminokwasowych. Oddziaływanie domena-J/Hsp70 jest kluczowe dla stymulacji hydrolizy ATP i prawidłowego funkcjonowania systemów JDP/Hsp70. Dotychczas nie było jednak wiadomo czy reszty zaangażowane w to oddziaływanie są zmienne i czy koewoluja - to znaczy czy zmiana sekwencji w obrębie domeny J kompensowana jest przez zmianę sekwencji w obrębie Hsp70 i odwrotnie. Stosując dwa dobrze poznane systemy JDP/Hsp70 (DnaJ/DnaK) oraz Hsc20/mtHsp70) wykazałam, że reszty zaangażowane w te oddziaływania zmieniają się w procesie ewolucji, oraz że koewolucja oddziałujących reszt warunkuje specyficzność tych oddziaływań.

## Summary

JDP/Hsp70 systems consisting of Hsp70 chaperone and J-domain protein (JDP) are key elements of cellular proteostasis playing critical roles both under physiological conditions and under stress or pathology. JDP/Hsp70 systems function in folding of newly synthesized proteins, polypeptide trafficking across membranes, and in preventing aggregation or promoting disaggregation of protein aggregates- thus protecting the cell against detrimental effects of stress and ageing. JDP/Hsp70 functions depend on their ability to cyclically interact with substrate protein. This substrate binding cycle is controlled by binding and hydrolysis of ATP by Hsp70. JDPs are responsible for delivering substrate proteins to Hsp70 partner or for recruiting Hsp70 into specific cellular location, thus determining the system specificity. In this thesis I used phylogenetic analyses such as reconstruction of phylogenetic trees, coevolution analyses and reconstruction of ancestral sequences to investigate relatedness, protein interactions and evolution of function of proteins constituting JDP/Hsp70 systems.

JDPs are divided into three distinct classes (A, B, C) based on their similarity to bacterial JDP DnaJ; Class B JDPs have a unique ability to prevent formation of amyloid aggregates.

In the first part of my thesis, I analysed the phylogenetic relationships among class A and B JDPs from 725 prokaryotic and eukaryotic proteomes. The obtained results support an evolutionary scenario that all eukaryotic JDPs both class A and B are directly related to bacterial DnaJ (class A). Class B JDPs has emerged independently as results of class A gene duplications in bacteria, in the cytosol of eukaryotic cell and in the endoplasmic reticulum of animals. These results demonstrate that the current classification of JDPs does not reflect their evolutionary relationships and thus should be revised.

In the second part I analyzed the evolutionary origin of mitochondrial mtHsp70 involved, together with a JDP partner (Hsc20), in the process of iron-sulfur cluster (FeS) biogenesis- a prosthetic groups required for function of many proteins. My results demonstrated that mtHsp70 descended from bacterial multifunctional Hsp70 (DnaK), which in bacteria does not function in the FeS biogenesis. I also demonstrated that bacterial Hsp70 specialized in the FeS biogenesis (HscA) is not present in eukaryotic proteomes, suggesting that HscA was lost during the evolution of mitochondria, and it was replaced by the descendant of DnaK in FeS biogenesis.

In the third part, I demonstrated using phylogenetic analyses, that the specificity of J-domain of JDP interaction with Hsp70 partner is driven by the coevolution of residues involved in this interaction. The J-domain/Hsp70 interaction is a key to the stimulation of ATP hydrolysis and thus proper functioning of JDP/Hsp70 systems. Up to now, it was not clear however whether the residues involved in this interaction are evolutionary variable and whether they coevolved- such that the changes on the J-domain's side of the binding interface are compensated by changes on the Hsp70's side and vice versa. Using two well characterized JDP/Hsp70 systems DnaJ/DnaK and Hsc20/mtHsp70 I demonstrated that residues involved in the J-domain/Hsp70 interaction are evolutionary variable and that coevolution among interacting residues determines specificity of this interaction.

# 1. Wstęp

## 1.1 Systemy opiekuńcze JDP/Hsp70

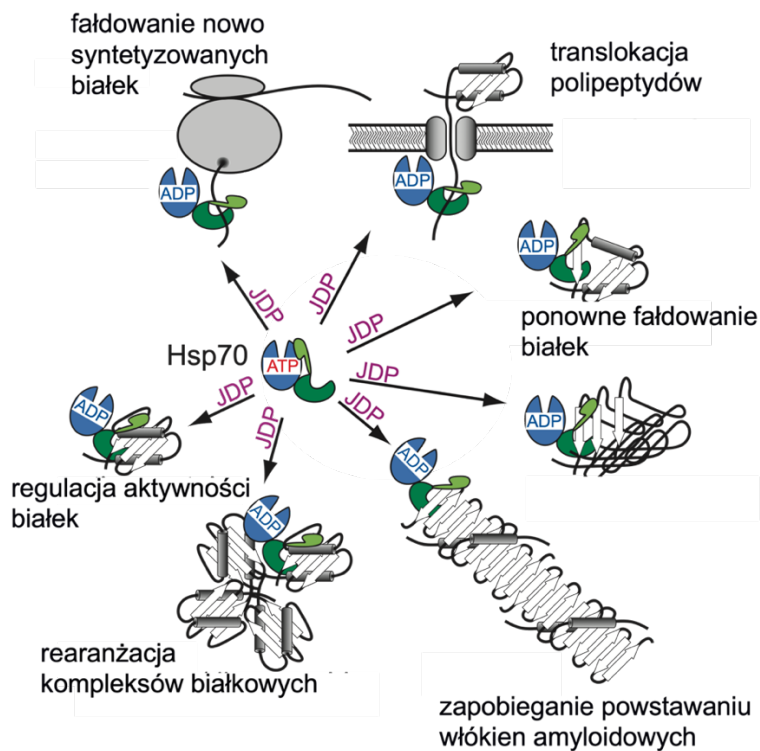
Systemy JDP/Hsp70 składające się z białka opiekuńczego Hsp70 ang. *Heat shock protein 70*) oraz białek pomocniczych posiadających domenę-J (ang. *J-domain proteins*, JDP) są podstawowymi elementami proteostazy komórkowej. Zrozumienie pochodzenia, funkcji i skomplikowanych zależności ewolucyjnych między tymi białkami ma duże znaczenie w odkrywaniu mechanizmów leżących u podstaw procesów komórkowych, w których uczestniczą oraz ich adaptacji do pełnienia zróżnicowanych funkcji. Analiza filogenetyczna stanowi użyteczne narzędzie do badania historii ewolucyjnej, procesu różnicowania lub konserwacji strukturalnej i funkcjonalnej systemów białkowych takich jak JDP/Hsp70 u różnych organizmów.

Systemy JDP/Hsp70 występują we wszystkich organizmach i są niezbędne do prawidłowego funkcjonowania komórek zarówno w warunkach fizjologicznych, jak i w warunkach stresu lub patologii [1] [2][3][4][5]. Systemy te biorą udział w wielu ważnych procesach komórkowych, w tym fałdowaniu nowo syntetyzowanych łańcuchów polipeptydowych, translokacji polipeptydów do mitochondriów, chloroplastów i retikulum endoplazmatycznego, rearanżacji kompleksów białkowych i regulacji aktywności białek [6]–[9] (Ryc. 1). Ponadto systemy JDP/Hsp70 zapobiegają agregacji i sprzyjają ponownemu fałdowaniu nieprawidłowo sfałdowanych i zdenaturowanych białek, rozpuszczają agregaty białkowe, zapobiegają powstawaniu włókien amyloidowych [[10]] oraz współpracując z systemami degradacji białek kontrolują stężenie nieprawidłowo sfałdowanych białek i agregatów białkowych [8]. W ten sposób systemy JDP/Hsp70 chronią komórki przed szkodliwymi skutkami różnorodnych czynników stresowych [11][12], stanów patofizjologicznych [13] i starzenia się organizmu [14], które powodują zaburzenia proteostazy.

W komórkach zarówno prokariotycznych, jak i eukariotycznych występuje wiele kopii Hsp70. U ludzi 13 homologów Hsp70 funkcjonuje w odrębnych przedziałach komórkowych (cytoplazma (C), retikulum endoplazmatyczne (ER) mitochondria (MT) oraz plastydy (PL)). Białka te, w zależności od rodzaju tkanki, podlegają różnej regulacji poziomu ekspresji. Chociaż Hsp70 charakteryzuje pewna redundancja funkcjonalna, członkowie tej rodziny białek mogą również wykazywać

wysoki stopień specjalizacji, poprzez wiązanie specyficznych substratów białkowych lub uczestnictwo w wyspecjalizowanych procesach komorkowych [15].

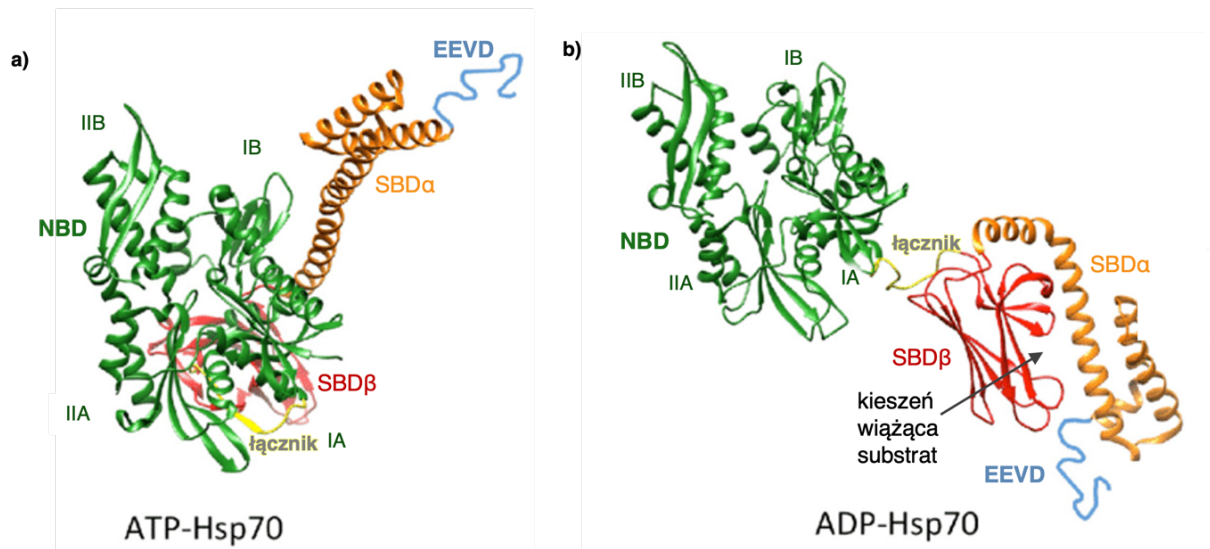
Hsp70 mogą się od siebie różnić rodzajem substratu białkowego, który rozpoznają, ponadto, ich funkcja często związana jest ze specyficzną lokalizacją komórkową. Na przykład, niektóre białka Hsp70 (Ssb i Ssz1) pełnią swoją funkcję na rybosomie, blisko tunelu wyjściowego łańcucha polipeptydowego [16][17]. Kolejną różnicą w funkcjonowaniu białek Hsp70 jest sieć oddziaływań z białkami pomocniczymi, takimi jak białka posiadające domenę J (ang. J-domain proteins, JDP)[18] i czynnikami wymiany nukleotydów (ang. *nucleotide exchange factor*, NEF), które kontrolują odpowiednio dostarczanie substratu oraz uwalnianie nukleotydu z Hsp70 [19][20]. Funkcjonalna różnorodność systemów JDP/Hsp70 jest dodatkowo zwiększana przez współpracę z innymi komórkowymi systemami opiekuńczymi, w tym małymi białkami szoku cieplnego (ang. *small heat shock proteins*, sHsp), białkami Hsp90 oraz Hsp100, jak również z systemami degradacji białek. Każda z funkcji Hsp70 zależna jest od cyklu ATPazowego polegającego na cyklicznym wiązaniu i uwalnianiu substratu białkowego kontrolowanym przez wiązanie i hydrolizę ATP [1].



**Rycina 1. Funkcje pełnione przez systemy JDP/Hsp70.** Systemy JDP/Hsp70 pełnią różnicowane funkcje wiążąc różnorodne substraty białkowe. Na podstawie [21]

## 1.2 Hsp70 – struktura i funkcja

Hsp70 zbudowane są z dwóch domen połączonych elastycznym łącznikiem: domeny wiążącej nukleotydy (ang. *nucleotide binding domain*, NBD) oraz domeny wiążącej substrat (ang. *substrate binding domain*, SBD) (Ryc. 2). NBD jest strukturalnie podobna do aktyny, składa się z dwóch „płatów” (ang. *lobe*) z głęboką szczeliną, w obrębie której wiązana jest cząsteczka ATP. Każdy z płatów jest podzielony na dwie części (IA, IIA, IB, IIB, Ryc. 2). Każda z czterech części zaangażowana jest w wiązanie ATP i niezbędna do związania przez Hsp70 substratu białkowego [22]. C-końcowa domena SBD składa się z części SBD $\beta$  złożonej z  $\beta$ -kardek tworzących kieszeń wiążącą substrat, oraz części SBD $\alpha$  stanowiącej  $\alpha$ -heliakalne wieczko (ang. *lid*), które przykrywa kieszeń wiążącą substrat stabilizując oddziaływanie Hsp70-substrat (Ryc. 2). Położenie wieczka zależy od nukleotydu związanego z Hsp70. W stanie ATP (Ryc. 2a) wieczko jest otwarte, a domeny NBD i SBD oddziałują ze sobą, hydroliza ATP do ADP indukuje rozejście się domen i zamknięcie wieczka (Ryc. 2b). [23]. Na końcu C wieczka znajduje się nieustrukturyzowany fragment, który u wielu eukariotycznych przedstawicieli Hsp70 funkcjonujących w cytoplazmie posiada ujemnie naładowany motyw Glu-Glu-Val-Asp (EEVD) [24]. Motyw ten jest dodatkowym miejscem oddziaływania niektórych Hsp70 z partnerskim JDP. Badania NMR sugerują, że w stanie związanim z ADP, NBD i SBD działają w dużej mierze niezależnie od siebie, połączone jedynie przez elastyczny łącznik (Ryc.2b) (ang. *linker*) [25].



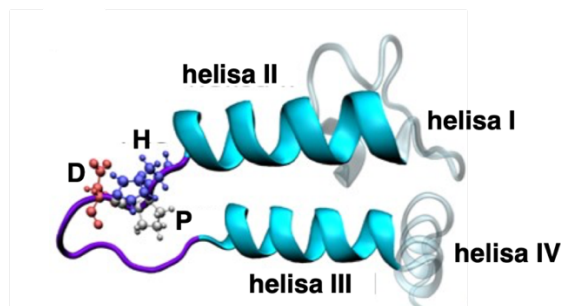
**Rycina 2. Struktura Hsp70.** Hsp70 zbudowane są z domeny wiążącej nukleotydy (NBD, zielony) połączonej za pomocą łącznika (żółty) z domeną wiążącą substrat (czerwony, żółty). Domena wiążąca substrat składa się z dwóch części SBD $\beta$  z kieszenią wiążącą substrat (czerwony) oraz SBD $\alpha$  (pomarańczowy) pełniącej funkcję „wieczka” zamykającego się na kieszeni wiążącej substrat. Na C-końcu SBD $\alpha$  znajduje się motyw EEVD (niebieski). a) Struktura Hsp70 w konformacji otwartej, związanego z ATP (DnaK, PDB: 4JNE). Domeny NBD i SBD oddziałują ze sobą, a wieczko odsłania miejsce wiązania substratu. b) Struktura Hsp70 w konformacji zamkniętej, związanego z ADP (DnaK, PDB:2KHO). Domeny NBD i SBD są od siebie oddalone, a kieszeń wiążąca substrat domeny SBD $\beta$  przykryta jest przez wieczko.

### 1.3 JDP – struktura i funkcja

Rodzina białek JDP jest bardzo zróżnicowana zarówno pod względem funkcjonalnym, jak i strukturalnym [18]. Badania sugerują, że zwiększenie liczby kopii białek JDP na drodze duplikacji kodujących je genów było ważnym czynnikiem sprzyjającym ewolucji organizmów wielokomórkowych [26]. Nazwa tej rodziny białek pochodzi od domeny-J, zidentyfikowanej po raz pierwszy w białku DnaJ z *Escherichia coli* [27]. Domena-J złożona jest z około 70 reszt aminokwasowych, które tworzą cztery  $\alpha$ -helisy tworzące strukturę spinki (Ryc. 3). Kluczowym dla funkcjonalnego oddziaływania pomiędzy JDP, a Hsp70 jest motyw His, Pro, Asp (HPD) znajdujący się na pętli łączącej dwie długie helisy: II i III [28]. Obecność domeny-J jest cechą charakterystyczną białek JDP. Poza domeną J, białka te różnią się znacznie pod względem strukturalnym [18]. Podstawową funkcją JDP jest dostarczanie substratu dla Hsp70 lub sprowadzenie Hsp70 do specyficznej lokalizacji komórkowej w pobliżu substratu oraz stymulacja aktywności ATPazowej Hsp70. [26]. Oddziaływanie domeny-J z Hsp70 stymuluje aktywność ATPazy przez co wpływa na konformacje



Hsp70 i jego zdolność do stabilnego oddziaływania z substratem białkowym. [29]  
Historia ewolucyjna białek JDP oraz wzajemne pokrewieństwo członków tej rodziny są słabo poznane [30]–[32].



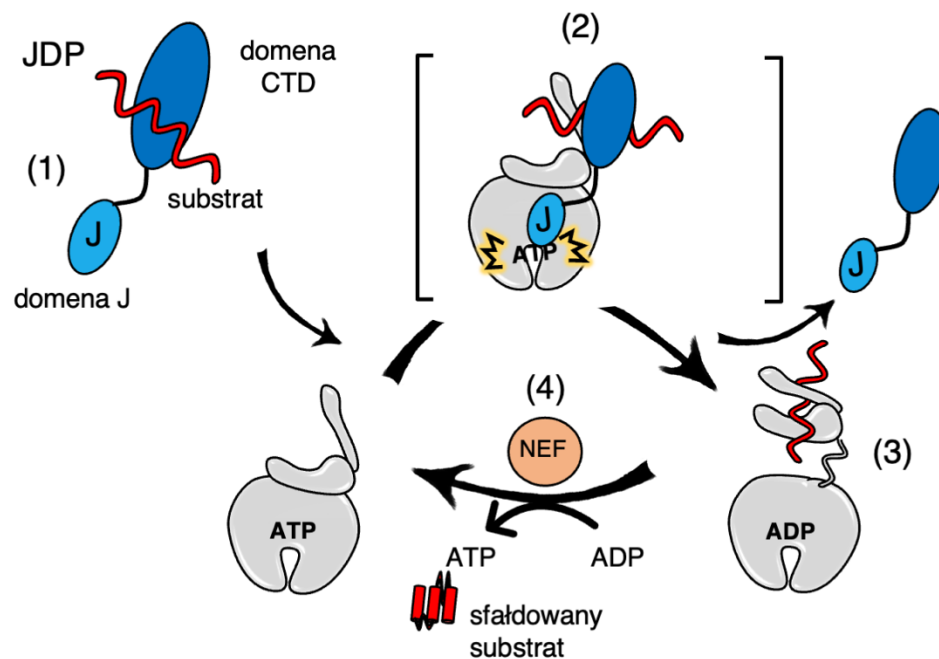
**Rycina 3. Struktura domeny-J.** Domena-J zbudowana jest z czterech helis: I, II, III, IV. Helisa II i III połączone są ruchliwą pętlą (fioletowy), na której znajduje się konserwowany u wszystkich JDP motyw HPD (histydyna-prolina-kwas asparaginowy) odpowiedzialny za stymulację aktywności ATPazowej Hsp70 (PDB: 1BQ0).

#### 1.4 Cykl wiązania substratu białkowego przez system JDP/Hsp70

Działanie systemów JDP/Hsp70 opiera się na cyklicznym wiązaniu i uwalnianiu substratu białkowego kontrolowanym przez wiązanie i hydrolizę ATP. (Ryc. 4). W trakcie cyklu Hsp70 przyjmuje dwie konformacje - otwartą lub zamkniętą (Ryc. 2). W konformacji otwartej, Hsp70 jest związane z ATP, domeny NBD i SBD oddziałują ze sobą, a  $\alpha$ -helikalne wieczko (SBD $\alpha$ ) jest przyłączone do NBD. W tej konformacji kieszeń wiążąca substrat w obrębie SBD $\beta$  jest otwarta co sprzyja szybkiemu wiązaniu i uwalnianiu substratu - powinowactwo Hsp70 do substratu jest niskie. W konformacji zamkniętej Hsp70 związane jest z ADP - domeny NBD i SBD połączone są jedynie elastycznym łącznikiem a kieszeń wiążąca substrat domeny SBD $\beta$  przykryta jest przez wieczko, co utrudnia zarówno wiązanie jak i uwalnianie substratu. W tej konformacji powinowactwo Hsp70 do substratu jest wysokie. [33][34].

Białko JDP wiąże substrat niezależnie od Hsp70 i dostarcza go do partnerskiego Hsp70 (Ryc.4, (1)) [35][36][37] w konformacji otwartej (Ryc. 2a, Hsp70-ATP) tworząc nietrwały kompleks JDP-substrat-Hsp70(ATP) (Ryc. 4, (2)) [38]. W kompleksie tym domena-J wiąże się do miejsca oddziaływania pomiędzy domenami NBD/SBD $\beta$  białka Hsp70, zaś substrat związany z JDP oddziałuje z Hsp70 w kieszeni wiążącej substrat.

Wiązanie domeny-J i substratu synergicznie stymuluje aktywność ATPazy Hsp70. Hydroliza ATP do ADP powoduje zmianę konformacji Hsp70 (Ryc. 2b) - oddzielenie domen NBD i SBD, co powoduje osłabienie oddziaływania z domeną-J, zaś zamknięcie wieczka na kieszeni wiążącej substrat stabilizuje oddziaływanie z substratem [39]. W efekcie JDP odłącza się od Hsp70, które w stanie ADP tworzy stabilny kompleks z substratem (Ryc.4, (3)). Wymiana nukleotydu związanego z Hsp70 z ADP na ATP, często stymulowana przez specyficzny czynnik wymiany nukleotydu (NEF) (Ryc.4, (4)), co skutkuje powrotem Hsp70 do konformacji otwartej (Hsp70-ATP) oraz uwolnieniu substratu [40][29][41]. Ten mechanizm wykorzystywany jest przez systemy JDP/Hsp70 do pełnienia wszystkich funkcji komórkowych - od fałdowania polipeptydów, poprzez ich transport przez błony komórkowe do ich ponownego fałdowania lub degradacji [8], [36], [42], [43]



**Rycina 4. Cykl wiązania substratu przez system JDP/Hsp70.** Na początku cyklu Hsp70 jest w konformacji otwartej, a do NBD związana jest cząsteczka ATP. Następnie kompleks JDP-substrat (1) oddziałuje z Hsp70, powstaje nietrwały kompleks potrójny Hsp70-JDP-substrat (2) dzięki czemu stymulowana jest aktywność ATPazowa Hsp70 – synergiczne oddziaływanie domeny-J i substratu jest niezbędne do efektywnej stymulacji aktywności ATPazowej. W konsekwencji hydrolizy ATP do ADP następuje zmiana konformacji Hsp70 z otwartej na zamkniętą i trwałe wiązanie substratu (3). Białko JDP opuszcza kompleks, a dzięki czynnikowi wymiany nukleotydu (NEF), ATP wymieniane jest na ADP i po uwolnieniu substratu (4), cykl może zająć od początku.

W większości przypadków pojedyncze Hsp70 współpracuje z wieloma białkami JDP.-Dzięki zróżnicowaniu domenowemu poza konserwowaną domeną-J, JDP odpowiadają za różnorodność funkcjonalną systemów JDP/Hsp70 dostarczając specyficzne substraty białkowe do partnerskiego Hsp70 [44]. Poza domeną-J, domeny występujące w białkach JDP znacząco się od siebie różnią i nie wykazują względem siebie homologii sekwencji [30]. Domeny te nie tylko wiążą specyficzne substraty, ale mogą kierować JDP do specyficznych lokalizacji komórkowych lub umożliwiać rozróżnienie prawidłowo sfałdowanych białek od ich agregatów [18]. Istnieją też wyspecjalizowane białka JDP, które dzięki wyspecjalizowanym domenom funkcjonują w procesach komórkowych takich jak składanie histonów czy biogeneza centrów żelazo-siarkowych. [18][30]. W obu przypadkach wyspecjalizowane JDP rekrutuje partnerskie Hsp70 do udziału w specyficznym procesie.

## 2. Cel pracy

Celem pracy było zastosowanie analizy filogenetycznej do zbadania zależności ewolucyjnych w rodzinie białek JDP oraz poznanie pochodzenia i różnicowania funkcjonalnego dwóch systemów JDP/Hsp70 różniących się stopniem specjalizacji.

W pierwszym rozdziale analizuję pokrewieństwo ewolucyjne członków najliczniejszej podrodziny JDP funkcjonującej zarówno u bakterii jak też we wszystkich przedziałach komórki eukariotycznej. Białka te spokrewnione są z bakteryjnym JDP DnaJ i pełnią kluczowe funkcje proteostatyczne zarówno w warunkach fizjologicznych jak też w stresie i patologii. Dotychczas nie przeprowadzono szeroko zakrojonej analizy filogenetycznej tych białek obejmującej przedstawicieli zarówno prokariotów jak też głównych grup organizmów eukariotycznych.

W drugim rozdziale analizuję pokrewieństwo ewolucyjne pomiędzy bakteryjnymi i mitochondrialnymi Hsp70 zaangażowanymi w proces biogenezy centrów żelazo-siarkowych oraz dystrybucję filogenetyczną ortologów bakteryjnych JDP oraz Hsp70 zaangażowanych w ten proces w organizmach eukariotycznych. Wyniki badań biochemicznych pokazywały, że białka te różnią się funkcjonalnie jednak ich pokrewieństwo ewolucyjne nie było dotychczas dobrze poznane.

W trzecim rozdziale analizuję dwa systemy JDP/Hsp70 różniące się stopniem specjalizacji. Wielofunkcyjny system DnaJ/DnaK oraz wyspecjalizowany w procesie biogenezy centrów żelazo-siarkowych system Hsc20/Ssq1 odpowiadając na pytanie jak pomimo znaczącego podobieństwa strukturalnego pomiędzy domenami-J możliwe jest specyficzne oddziaływanie z partnerskim Hsp70. Sprawdzam, czy koewolucja pomiędzy partnerskimi JDP/Hsp70 odpowiada za ich specyficzne oddziaływanie.

### 3. Ewolucja eukariotycznych JDP spokrewnionych z bakteryjnym DnaJ.

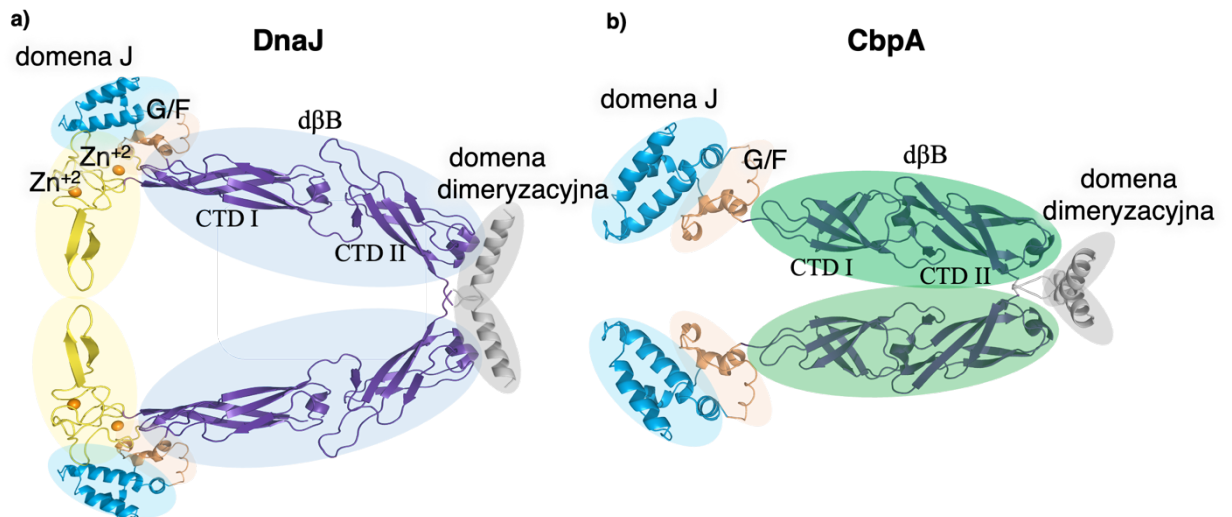
#### 3.1 Wstęp

Zmiany zachodzące w procesie ewolucji umożliwiają białkom uzyskiwanie nowych aktywności biochemicznych [45]. Mechanizmy molekularne leżące u podstawy tego procesu są jednym z głównych przedmiotów zainteresowania biologii ewolucyjnej [46]. Większość nowych aktywności białek ewoluuje z aktywności istniejących wcześniej poprzez różnego rodzaju modyfikacje, do których zaliczają się zmiany sekwencji kodującej oraz struktury białek, a także duplikacje i utraty kodujących je genów, które prowadzą do różnicowania się zarówno liczby kopii, jak i funkcji białek [46]. Kolejność zachodzenia poszczególnych modyfikacji ma kluczowe znaczenie w procesie ewolucji, jednak zależności te są trudne do ustalenia. Dlatego śledzenie historii ewolucyjnej białka od jego przodka do białka funkcjonującego współcześnie jest wyzwaniem [47]. Rodzina JDP ze względu na swoją liczebność i zróżnicowanie zarówno funkcjonalne jak też strukturalne jest interesującym modelem do badania ewolucji i specjalizacji rodzin białek [30].

Najlepiej poznane i opisane bakteryjne JDP pochodzą z *E. coli*: DnaJ oraz CbpA (Ryc. 3.1). DnaJ zbudowane jest N-końcowej domeny-J połączonej rejonem bogatym w glicynę i fenyloalaninę (rejon GF) z domeną CTD (*ang. C-terminal domain*) złożoną z dwóch części – CTD I (zawierającą rejon wiążący jony cynku) oraz CTD II (Ryc. 3.1a). Domeny CTD formują strukturę podwójnej baryłki  $\beta$ . DnaJ funkcjonuje jako dimer dzięki C-końcowej domenie dimeryzacyjnej. DnaJ bierze udział w fałdowaniu łańcucha polipeptydowego nowo syntetyzowanych białek, ponownym fałdowaniu zdenaturowanych białek oraz ich dezagregacji. Uczestniczy również w procesie replikacji DNA bakteriofaga lambda oraz kontroli odpowiedzi na szok termiczny. [29], [48]. Partnerskim Hsp70 dla DnaJ jest DnaK.

CbpA ma organizację domenową podobną do DnaJ: N-końcową domenę-J, rejon G/F, domenę CTD I, jednak pozbawioną regionu wiążącego jony cynku, domenę CTD II oraz C-końcową domenę dimeryzacyjną (Ryc. 3.1b). Cechą charakterystyczną

CbpA jest dłuższy, w porównaniu do DnaJ, region GF. Partnerem Hsp70 dla CbpA, podobnie jak w przypadku DnaJ, jest DnaK. CbpA uczestniczy w fałdowaniu łańcucha polipeptydowego białek, zapobiega ich agregacji w warunkach stresu, a także bierze udział w ochronie DNA przed uszkodzeniem. [49][50]. Analiza 1709 genomów reprezentujących główne grupy systematyczne bakterii pokazała, że podczas gdy DnaJ występuje w 99 % analizowanych genomów, CbpA stwierdzono jedynie w 51 %. [51]. W warunkach nadprodukcji CbpA może zastąpić DnaJ.



**Rycina 3.1. Modele strukturalne DnaJ i CbpA.** a) DnaJ zbudowane jest N-końcowej domeny-J (niebieski) połączonej rejonem bogatym w glicynę i fenyloalaninę (GF, pomarańczowy) z podwójną barierką  $\beta$  ( $d\beta B$ , fioletowy), tworzoną przez domeny CTD I zawierającą rejon wiążący dwa jony cynku (żółty) oraz CTD II. Na końcu C- znajduje się domena dimeryzacyjna (szary). b) CbpA składa się z N-końcowej domeny-J (niebieski), rejonu GF (pomarańczowy), podwójnej bariery  $\beta$  ( $d\beta B$ , zielony) złożonej z domen CTD I i CTD II oraz domeny dimeryzacyjnej (szary). CbpA nie posiada rejonu wiążącego jony cynku. Przygotowane na podstawie struktur o numerach PDB dla DnaJ: 1C3G, 4RWU, 6JZB dla CbpA: 1NLT, 5VSO, 6JZB.

Białka podobne strukturalnie do DnaJ i CbpA, występują we wszystkich przedziałach komórki eukariotycznej. W cytoplazmie *S.cerevisiae* strukturalnie podobne do DnaJ jest Ydj1 (Ryc. 3.2b), które współpracując z białkami Hsp70 Ssa1-Ssa4 pomaga stabilizować niesfałdowane białka, zapobiegając ich agregacji i promując ich ponowne fałdowanie [52] [53]. Pomimo ważnych funkcji Ydj1 nie jest niezbędne do wzrostu drożdży w normalnych warunkach - delecja genu kodującego Ydj1 nie prowadzi do śmierci komórki. W retikulum endoplazmatycznym oraz mitochondriach drożdży także obecne są odpowiedniki DnaJ, nazwane odpowiednio Scj1 oraz Mdj1 (Ryc. 3.2b). Scj1 w asystuje przy fałdowaniu oraz degradacji łańcuchów

polipeptydowych [54], natomiast Mdj1 poza asystowaniem przy fałdowaniu nowopowstałych białek bierze również udział w utrzymywaniu potencjału błonowego oraz integralności mitochondrialnego DNA. Mdj1 jest niezbędny dla przeżycia komórki drożdżowej [55][56].

U człowieka sytuacja jest bardziej skomplikowana: w cytoplazmie występują trzy białka podobne do DnaJ: DNAJA1, DNAJA2 i DNAJA4 (Ryc. 3.2b). DNAJA1 pomaga w fałdowaniu białek, dojrzewaniu receptora androgenowego i spermatogenezie. [57] DNAJA2 bierze udział w utrzymaniu proteostazy komórkowej i fałdowaniu łańcuchów polipeptydowych [58]. W ludzkich mitochondriach funkcjonuje homolog Mdj1 nazywany DNAJA3, który poza asystowaniem w fałdowaniu białek, jest niezbędny do utrzymania mitochondrialnego DNA [59]

Białka JDP strukturalnie podobne do CbpA również występują w komórkach eukariotycznych (Ryc. 3.2b). U *S. cerevisiae* Sis1 strukturalnie przypomina CbpA- jego CTD I nie posiada rejonu wiążącego jony cynku, a rejon GF jest dłuższy niż u Ydj1. Sis1, uczestniczy w fałdowaniu białek *de novo* oraz ponownym fałdowaniu i dezagregacji białek, a także jest niezbędny dla utrzymania prionów drożdżowych [52]. Co ciekawe, podczas gdy brak Ydj1 w komórkach drożdży może być kompensowany funkcjonalnie poprzez nadprodukcję innych białek JDP w tym Sis1, delecja Sis1 jest letalna - jednak molekularne podstawy istotnej dla przeżycia funkcji Sis1 nie zostały dotychczas poznane. U człowieka występują trzy cytoplazmatyczne białka JDP strukturalnie przypominające CbpA i Sis1: DNAJB1, DNAJB4 oraz DNAJB5 (Ryc. 3.2b). DNAJB1 posiada zdolność zapobiegania agregacji fibryli amyloidowych [60]. W retikulum endoplazmatycznym człowieka występuje białko – DNAJB11, które posiada cechy strukturalne CbpA, jednak domena CTDI tego białka zawiera fragment rejonu wiążącego jony cynku charakterystycznego dla DnaJ i Ydj1. DNAJB11 wiąże jeden jon cynku. Nie wiadomo jakie jest pokrewieństwo ewolucyjne DNAJB11 z bakteryjnymi CbpA i DnaJ oraz z eukariotycznymi homologami tych białek.

### 3.1.1 Klasyfikacja JDP

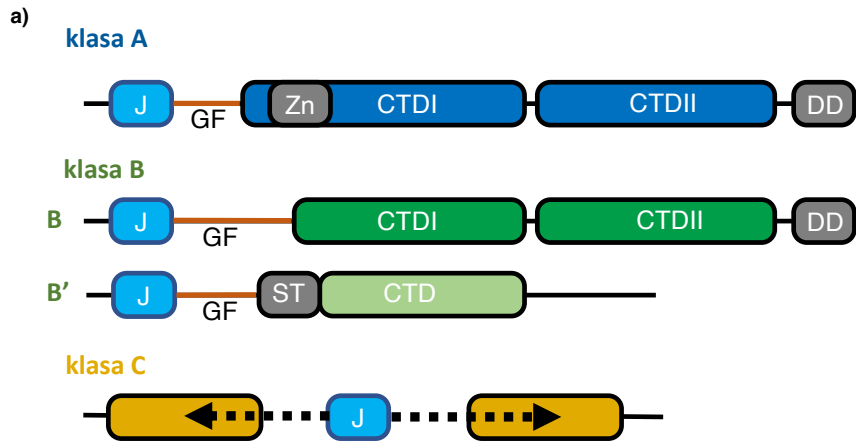
W oparciu o podobieństwo organizacji domenowej JDP podzielone zostały na trzy klasy: A, B, C. (Ryc. 3.2a) [31]. Do klasy A zalicza się białka JDP o układzie domen takim samym jak DnaJ - N-końcowa domena-J połączona rejonem GF z

domeną CTD złożoną z dwóch części CTD I (zawierającej rejon wiążący dwa jony cynku) oraz CTD II razem formujących strukturę podwójnej baryłki  $\beta$  (Ryc. 3.1a, 3.2a). Na końcu C - domena dimeryzacyjna. Do klasy B zalicza się białka JDP posiadające N-końcową domenę-J, za którą znajduje się rejon GF (Ryc. 3.2a). Większość białek należących do klasy B posiada strukturę domenową odpowiadającą bakteryjnemu CbpA. [18], [61] Wśród JDP klasy B można także wyróżnić podgrupę białek posiadającą za rejonem GF rejon bogaty w reszty seryny i treoniny (rejon ST), a w części C - końcowej (Ryc. 3.2a, CTD) domenę złożoną z  $\beta$ -katek, nie formującą jednak baryłki  $\beta$ . W dalszej części pracy białka te nazywam: klasa B'.






Reprezentanci klasy B' obecni są jedynie u tkankowców (Metazoa). U człowieka występuje trzech przedstawicieli klasy B': DNAJB2, DNAJB6 oraz DNAJB8 (Ryc.3.2b). Białka te, podobnie jak DNAJB1 należące do klasy B mają zdolność zapobiegania agregacji fibryli amyloidowych [62][63][64]. Pokrewieństwo ewolucyjne białek klasy B' z kanonicznymi przedstawicielami JDP klasy B pozostaje nieznane.

Do klasy C zalicza się wszystkie pozostałe białka JDP posiadające funkcjonalną domenę-J bez sąsiadującego rejonu GF. U przedstawicieli klasy C domena-J może znajdować się w dowolnym miejscu struktury białka; na końcu N, na końcu C lub w środku. Poza domeną-J przedstawiciele klasy C posiadają szeroką gamę domen i motywów (Ryc. 3.2a) [65]. Do klasy C należy wiele wyspecjalizowanych JDP takich jak białka uczestniczące w biogenezie centrów żelazo-siarkowych czy też białka zaangażowane w biogenezę histonów lub splatanie mRNA [18][30].





b)

		<b>Klasa A</b>	<b>Klasa B</b>	<b>Klasa B'</b>
 <i>E. coli</i>		 <b>1</b> (DnaJ)	 <b>1</b> (CbpA)	-
 <i>S. cerevisiae</i>	<b>C</b>	<b>3</b> (Ydj11, Xdj1, Apj1)	<b>1</b> (Sis1)	-
	<b>ER</b>	<b>1</b> (Scj1)	-	-
	<b>MT</b>	<b>1</b> (Mdj1)	-	-
 <i>H. sapiens</i>	<b>C</b>	<b>3</b> (DNAJA1/A2/A4)	<b>4</b> (DNAJB1/B4/B5/B13)	<b>3</b> (DNAJB2/B6/B8)
	<b>ER</b>	-	<b>1</b> (DNAJB11)	-
	<b>MT</b>	<b>1</b> (DNAJA3)	-	-

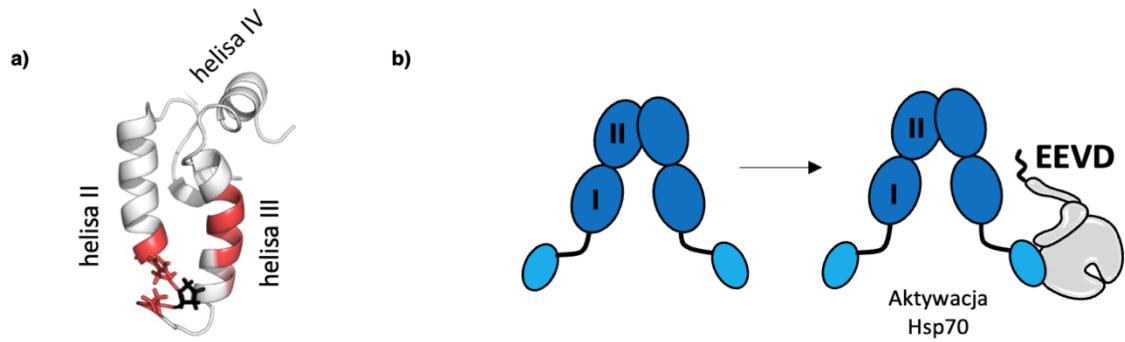
**Rycina 3.2. Klasyfikacja JDP w oparciu o strukturę domenową.** a) Klasa A - układ domen taki jak w DnaJ Klasa B- definiowana przez N-kończową domenę-J i rejon GF. Kanoniczne JDP klasy B mają układ domen taki jak w CbpA, JDP klasy B' mają N-kończową domenę-J, rejon GF, rejon bogaty w Ser i Tre (ST) oraz  $\beta$ - kartkową domenę CTD nietworzącą baryłki  $\beta$ . Klasa C- to pozostałe białka JDP. Domena-J może być na końcu C, końcu N lub w środku. Pozostałe domeny są bardzo różnorodne. b) Białka klasy A, B i B' funkcjonujące w komórkach *E. coli*, *S. cerevisiae* oraz *H. sapiens*. C- cytoplazma, ER- retikulum endoplazmatyczne, MT- mitochondria.

### 3.1.2 Oddziaływanie JDP należących do klas A i B z Hsp70

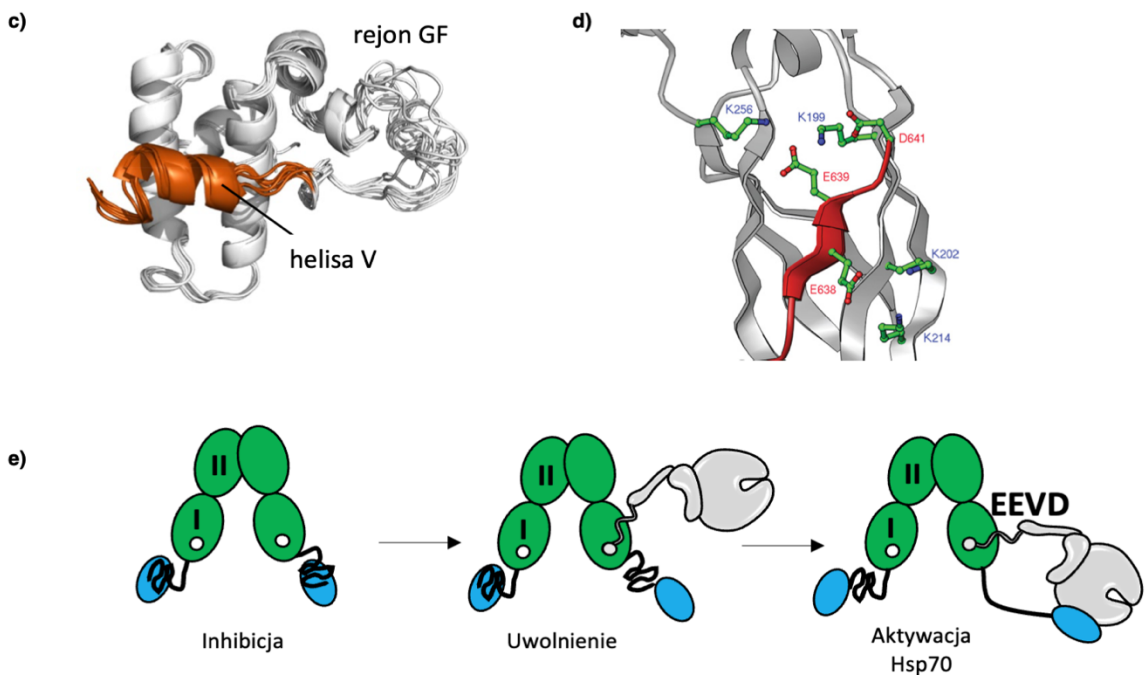
Białka JDP należące do klas A i B są podobne strukturalnie, natomiast zarówno funkcje, które wykonują, jak i mechanizm ich oddziaływania z Hsp70 są znacząco różne. Żeby w pełni zrozumieć na czym polega różnica między oddziaływaniem białek JDP klasy A i B z Hsp70 należy się najpierw przyjrzeć budowie domeny-J obu klas. JDP klasy A posiadają domenę-J zbudowaną z czterech helis: helisy I, helisy II połączonej z helisą III pętlą, na której znajduje się motywy HPD niezbędny do prawidłowego oddziaływania JDP z Hsp70 oraz krótszej od pozostałych helisy IV, która jest ustawiona do helis II i III prostopadle (Ryc. 3.3a). Białka JDP klasy A oddziałują z NBD Hsp70 jedynie poprzez domenę-J, stymulując aktywność ATPazy Hsp70 oraz indukując jego zmianę konformacyjną. Rejonem kluczowym dla oddziaływania domeny-J z Hsp70 jest helisa II i III wraz z motywem HPD. Oddziaływanie domena-J/Hsp70 jest niezbędne do przekazania sygnału allosterycznego prowadzącego do hydrolizy ATP indukującej zmiany konformacyjne Hsp70 [66] (Ryc. 3.3a, Ryc. 1.3).

JDP klasy B posiadają w obrębie rejonu GF heliakalny fragment nazywany helisą V (reszty aminokwasowe 98-106 dla DNAJB1 oraz 68-74 dla Sis1). Helisa V oddziałuje z domeną-J hamując zdolność do stymulacji aktywności ATPazowej Hsp70, co nazywane jest auto-inhibicją domeny-J. (Ryc. 3.3a). Helisa V oddziałuje z helisami II i III domeny-J blokując jej zdolność do wiązania Hsp70, uniemożliwiając tym samym powstanie aktywnego kompleksu (ryc. 3.3e) [67]. Ponadto JDP klasy B, drożdżowe białko Sis1, jak i ludzkie DNAJB1, w obrębie domeny CTDI posiadają dodatkowe miejsce oddziaływania z Hsp70 [62] [68] [69]. Trzy reszty lizyny (K199, K202, K214 w Sis1) oddziałują z motywem EEVD znajdującym się na C końcu SBD Hsp70 (Ryc. 3.3b). Oddziaływanie to prowadzi do odłączenia się helisy V rejonu GF od domeny-J, a tym samym uwolnienia domeny-J, która może stymulować aktywność ATPazową Hsp70 (Ryc. 3.3c).

## Klasa A



## Klasa B



**Rycina 3.3. Oddziaływanie JDP klasy A i B z Hsp70.** a) Struktura domeny-J JDP klasy A – na pętli łączącej helisę II i III zaznaczony jest motyw HPD. b) JDP klasy A oddziałują z partnerskim Hsp70 poprzez domenę-J. c) Struktura domeny-J (szary) wraz z fragmentem rejonu GF (pomarańczowy) u JDP klasy B. Helisa V rejonu GF oddziałuje z helisami II i III domeny-J hamując jej zdolność oddziaływania z Hsp70. d) dodatkowe miejsca oddziaływania JDP klasy B z Hsp70. Motyw EEVD Hsp70 wiąże się do CTD I JDP klasy B. Reszty K199, K202, K256 CTD I w kontakcie z EEVD (czerwony). Numery reszt podane są dla Sis1. e) JDP klasy B z Hsp70 oddziałują dwuetapowo reszty lizyny znajdujące się na CTD I (I) oddziałują z motywem EEVD Hsp70. To powoduje uwolnienie helisy V rejonu GF od domeny-J i umożliwia oddziaływanie domeny-J z Hsp70. Domeny CTD I i CTD II odpowiednio I i II, domena-J – jasnoniebieski, Hsp70 – szary.

Różny mechanizm oddziaływania JDP klasy A i B z Hsp70 sprawia, że każda z klas oddziałuje inaczej z agregatami amorficznymi białek, a JDP klasy B mają zdolność oddziaływania także z fibrylami amyloidowymi. Białka klasy A nie są w stanie zastąpić białek klasy B w procesie dezagregacji agregatów białkowych i włókien amyloidowych [60], [70][71]W przypadku białek *S. cerevisiae* zarówno Ydj1 (klasa A), jak i Sis1 (klasa B) oddziałują z amorficznymi agregatami białkowymi [32]. Jednak różny mechanizm oddziaływania z Hsp70 sprawia, że różni się zarówno ich powinowactwo do agregatów, jak i sposób działania w dezagregacji. W przypadku agregatów amorficznych takich jak denaturowana termicznie lucyferaza Ydj1 (klasa A) wiąże substrat samodzielnie, a następnie rekrutuje do agregatu partnerskie Hsp70. W przypadku Sis1 (klasa B) wiązanie agregatu białkowego wymaga jednoczesnej obecności JDP i Hsp70- prawdopodobnie ze względu na dodatkowe miejsce oddziaływania Sis1/Hsp70. Dodatkowo, Sis1 jest w stanie przyłączyć do agregatu większą liczbę cząsteczek Hsp70 niż Ydj1, [68]. Powodem, dla którego tak się dzieje może być dodatkowe miejsce wiązania Hsp70, dzięki któremu Sis1 może związać więcej cząsteczek Hsp70 niż Ydj1. Oddziaływanie z motywem EEVD Hsp70 zachodzi zarówno w obecności ATP jak też ADP związanego z Hsp70 - jest więc niezależne od tego czy Hsp70 znajduje się w konformacji zamkniętej czy otwartej. Podczas gdy domena-J wiąże się tylko do konformacji otwartej Hsp70-ATP [72].

Efekt nagromadzenia Hsp70 został pokazany także dla ludzkiego JDP klasy B – DNAJB1 oddziałującego z agregatami  $\alpha$ -synukleiny tworzącymi włókna amyloidowe [62]. Nagromadzenie Hsp70, podobnie jak w przypadku Sis1, sprzyjało destabilizacji amyloidów pozwalając na ich łatwiejszą dezagregację. W Przypadku JDP klasy B mamy więc do czynienia z mechanizmem regulacyjnym służącym do precyzyjnej kontroli kierowania Hsp70 do substratu białkowego. Mechanizm ten jest zarówno unikatowy, jak i niezbędny dla prawidłowego funkcjonowania Hsp70.

Kolejną grupą JDP oddziałującą z włóknami amyloidowymi są białka klasy B': DNAJB6 oraz DNAJB8. Za oddziaływanie JDP klasy B' z włóknami amyloidowymi odpowiedzialny jest rejon ST charakterystyczny dla tej klasy białek. DNAJB6 jest niezwykle skuteczne w hamowaniu tworzenia włókien amyloidowych powstających z peptydu A $\beta$ 42 związanych z rozwojem choroby Alzheimera [73]. DNAJB6 hamuje tworzenie włókien amyloidowych przy bardzo niskim stosunku molowym do peptydu

A $\beta$ 42. DNAJB6 oddziałuje z amyloidem  $\beta$  powodując spadek szybkości nukleacji pierwotnej, tym samym hamując tworzenie zaczątków amyloidu [74][63]. DNAJB6 oraz DNAJB8 są także najbardziej skutecznymi czynnikami antyagregacyjnymi innego substratu amyloidowego - poliQ, powiązanego z chorobą Parkinsona. Przypuszcza się, że grupy hydroksylowe (OH) reszt seryny i treoniny w rejonie ST, które posiadają zdolność do tworzenia wiązań wodorowych z substratem, konkurując z białkami tworzącymi amyloid o miejsce wiązania, uniemożliwiają w ten sposób tworzenie włókien amyloidowych [74][75][76].

Dezagregacja amyloidów z udziałem systemów JDP/Hsp70 może spowalniać rozwijanie się chorób neurodegeneracyjnych. Próby oceny potencjału terapeutycznego metod opartych na zmianie aktywności dezagregacyjnej systemów JDP/Hsp70 pokazują, że ma ona korzystny wpływ na dezagregację  $\alpha$ -synukleiny w hodowlach komórkowych oraz modelach zwierzęcych [77][78]. Badania tych systemów mogą dostarczać podstaw dla przyszłego rozwoju środków terapeutycznych, które będą specyficznym zwiększały aktywność dezagregacyjną Hsp70.

## 3.2 Metody

### 3.2.1 Przygotowanie zbioru danych do analiz filogenetycznych JDP należących do klas A i B

#### *Zestawienia sekwencji aminokwasowych JDP*

Zestawienie I obejmowało: dobrze poznane i opisane sekwencje JDP klasy A i B pochodzących z *E. coli* (DnaJ, CbpA), *S. cerevisiae* (Ydj1, Xdj1, Apj1, Scj1, Mdj1, Sis1). oraz *H. sapiens* (DNAJA1, DNAJA2, DNAJA3, DNAJA4, DNAJB1, DNAJB4, DNAJB5, DNAJB6). Sekwencje zestawiałam przy użyciu programu Clustal Omega z zastosowaniem domyślnych parametrów.

Zestawieni II obejmowało: 130 JDP będących ortologami białek: (1) należących do klasy A (DnaJ, DNAJA1, DNAJA2, Ydj1, Xdj1, Mdj1), (2) należących do klasy B (CbpA, DNAJB1, DNAJB4, DNAJB5, DNAJB9, Sis1) oraz (3) należących do klasy C (Hsc20, DNJC20, DNAJC24, DNAJB12). Sekwencje zestawiałam przy użyciu programu Clustal Omega z zastosowaniem domyślnych parametrów. Zestawienie II wykorzystałam do trenowania modeli Markowa do specyficznego rozpoznawania JDP należących do klasy A lub B.

#### *Przygotowanie profili Markowa (ang. Hidden Markov Model Profiles, HMM)*

Zestawienie I przekształciłam w ukryty Model Markowa (HMM A/B) przy użyciu programu *hmmbuild* z pakietu HAMMER [79], potrzebny do przygotowania ostatecznego zestawienia sekwencji JDP klasy A i B do analiz filogenetycznych.

Profile Markowa dla domeny-J, CTD I oraz CTD II białek JDP klasy A i B, przygotowałam przy użyciu programu *hmmbuild* z pakietu HAMMER w oparciu o fragmenty Zestawienia I odpowiadające każdej z wymienionych domen. Następnie, każdy z modeli HMM „trenowałam” i testowałam wykorzystując zestawienie II. Testowanie modeli zakończyłam, gdy bezbłędnie identyfikowały obecność lub brak, danej domeny we wszystkich sekwencjach aminokwasowych w zestawieniu II.

### *Przeszukanie bazy danych proteomów*

725 proteomów bakterii, archeonów, grzybów, zwierząt oraz roślin pobrałam z bazy danych Uniprot. Korzystając z narzędzia *hmmsearch* z pakietu HMMER [79], wykorzystując profile HMM dla domeny-J, CTD I oraz CTD II przeszukałam wszystkie zgromadzone proteomy pod względem obecności białek posiadających te domeny. W pierwszej kolejności przeszukałam proteomy na obecność domeny-J, a następnie na obecność domen CTD I oraz CTD II. Sekwencje dające pozytywny sygnał dla domeny-J oraz przynajmniej jednej z domen CTD włączyłam do zestawienia. W wyniku przeszukania otrzymałam 8446 sekwencji aminokwasowych JDP.

### *Zestawienia sekwencji aminokwasowych JDP do analiz filogenetycznych*

#### *Zestawienie A/B*

Z zestawienia sekwencji JDP, które otrzymałam po przeszukaniu proteomów usunęłam sekwencje powtarzające się (100% identyczność). Dodatkowo zestawienie zweryfikowałam manualnie usuwając sekwencje fragmentaryczne oraz wyraźnie krótsze lub dłuższe od pozostałych. Ostatecznie zestawienie obejmowało 1588 sekwencji JDP.

Zgromadzone sekwencje JDP po manualnej korekcie ponownie zestawiałam przy użyciu profilu HMM A/B korzystając z narzędzia *hmmalign* z pakietu HAMMER. Zestawienie sekwencji tą metodą pozwala otrzymać wartość prawdopodobieństwa *a posteriori* (ang. *posterior probability*, pp) dla każdej kolumny zestawienia osobno dla każdej z sekwencji. Parametr pp przyjmuje wartość od 0 do 1. Wartości te pozwalają, po uśrednieniu wyników z każdej kolumny w zestawieniu, ocenić jakość każdej pozycji zestawienia. Z zestawienia usunęłam wszystkie kolumny, których wsparcie pp było niższe niż 0.7 i tak przygotowane zestawienie użyłam do rekonstrukcji filogenezy JDP klasy A i B metodą Maksymalnej wiarygodności (ang. *Maximum Likelihood*, ML).

#### *Zestawienie A/B-B'*

Ze względu na to, że sekwencje JDP klasy B' różnią się od sekwencji kanonicznych JDP klasy B przygotowałam zestawienie sekwencji nie zawierające białek klasy B' (Zestawienie A/B-B'). Opierając się na topologii drzewa otrzymanego

metodą ML z użyciem modelu LG+I+G4 przydzieliłam każdą sekwencje Zestawienia A/B do klasy A, B lub B'. Następnie z Zestawienia A/B usunęłam wszystkie sekwencje należące do klasy B', a pozostałe sekwencje zestawiałam ponownie z użyciem programu Clustal Omega z zastosowaniem domyślnych parametrów.

#### *Zestawienie C-B'/ER*

Zestawienie to obejmowało jedynie eukariotyczne sekwencje JDP klasy A i B funkcjonujących w cytoplazmie i ER. Zestwienie to przygotowałam w oparciu o Zestawienie A/B poprzez usunięcie sekwencji-bakteryjnych, sekwencji z mitochondriów i plastydów, a także sekwencji JDP klasy B'. Sekwencje pozostałe po usunięciu zestawiałam ponownie z użyciem programu Clustal Omega z zastosowaniem domyślnych parametrów [80]. Zestawienie C-B'/ER wykorzystałam do analizy Bayesowskiej.

### 3.2.2 Rekonstrukcja drzewa filogenetycznego JDP

#### *Rekonstrukcje drzewa filogenetycznego JDP metodą ML*

Obliczenia wykonałam wykorzystując Zestawienie A/B oraz Zestawienie A/B-B' przy użyciu programu IQ-Tree stosując dwa modele ewolucji (1 i 2) oraz metodę jackknife (3). (1) model LG+I+G4; wykorzystujący macierz podstawień LG z pozycjami niezmiennymi oraz dyskretny model Gamma z 4 kategoriami szybkości ewolucji [81], wybrany na podstawie kryterium Akaike jako najlepiej dopasowany do danych. (2) model mieszany C20 - stosowany w przypadku sekwencji wysyconych zmianami [82], grupuje kolumny w zestawieniu o podobnej kompozycji aminokwasowej w 20 kategorii. Do każdej z 20 kategorii dopasowywany jest osobny model ewolucji, a funkcja wiarygodności jest liczona dla sumy wszystkich pozycji. Wsparcie statystyczne dla obu modeli weryfikowałam metodą *bootstrap* (1000 powtórzeń). (3) Metoda *jackknife* [83]: wygenerowałam 100 zestawień usuwając losowo z Zestawienia odpowiednio A/B oraz A/B-B' 40% pozycji. Na podstawie każdego zestawienia przygotowywałam drzewo ML stosując model ewolucji LG+I+G4. Topologia najwyższej zgodności (konsensus) wszystkich drzew obliczonych w ten sposób jest ostateczną topologią drzewa *jackknife*. Wsparcie dla węzłów pokazuje w ilu ze 100 prób wystąpił dany węzeł.



### Rekonstrukcje drzewa filogenetycznego metodą Bayesowską

Rekonstrukcję drzewa filogenetycznego metodą Bayesowską dla Zestawienia C-B'/ER przeprowadziłam przy użyciu programu PhyloBayes stosując cztery modele ewolucji: (1) model najlepiej dopasowany bez zmienności w obrębie pozycji (ang. *Best fit model without variation accross sites*, CAT1LG) (2) model zezwalający na heterogenność pozycji aminokwasowych. (ang. *Bayesian mixture model for across-site heterogeneities*, CATLG) [84] oraz dwa modele empiryczne, gdzie heterogenność pozycji aminokwasowych jest wyznaczana z zestawienia (ang. *Empirical profile mixture models for single gene analysis*) (3) model C60 [82] oraz (4) model W5 [85]. Dla każdego modelu obliczone były 4 łańcuchy. Zbieżność obliczeń sprawdziłam przy użyciu programu *bpcomp* po usunięciu z analizy (ang. *burnin*) pierwszych 600 drzew. Zbieżność sprawdzałam oddzielnie dla każdego z wykorzystanych modeli badając każdą możliwą parę łańcuchów. Topologię najwyższej zgodności (konsensus) przygotowaną na podstawie wszystkich drzew z pary najbardziej zbieżnych łańcuchów przedstawiam jako ostateczną topologię otrzymaną przy zastosowaniu danego modelu. W przypadku analizy Bayesowskiej wsparciem drzewa nie są losowe powtórzenia *bootstrap* (których wartości znajdują się w przedziale od 1 do 100), a wartość prawdopodobieństwo *a posteriori* (ang. *posteriori probability*), którego wartość znajduje się w przedziale od 0 do 1.

### Ukorzenie drzewa ML

Aby ukorzenie drzewo ML przetestowałam trzy grupy zewnętrzne: (1) bakteryjne sekwencje CbpA (klasa B), (2) bakteryjne sekwencje DnaJ (klasa A) oraz (3) łącznie bakteryjne sekwencje CbpA i DnaJ. Aby sprawdzić, które z miejsc ukorzenia jest najbardziej prawdopodobne zastosowałam cztery testy statystyczne: (1) porównanie logarytmu wartości funkcji wiarygodności, (2) test Kishino Hasegawa z zastosowaniem wag (ang. *Weighted KH test*, KH) [86], (3) test Shimodaira Hasegawa z zastosowaniem wag (ang. *Weighted SH test*, WSH) [87] oraz (4) test AU (ang. *Approximately Unbiased test*) [88]. Wynikiem każdego z testów jest wartość *p-value*. Im wartość ta jest bliższa 1, tym dana topologia drzewa jest bardziej prawdopodobna.

### 3.2.3 Rekonstrukcja sekwencji przodków

Rekonstrukcje przeprowadziłam przy użyciu programu FastML wykorzystując empiryczną metodę Bayesowską oraz metodę ML do rekonstrukcji insercji i delecji [89].

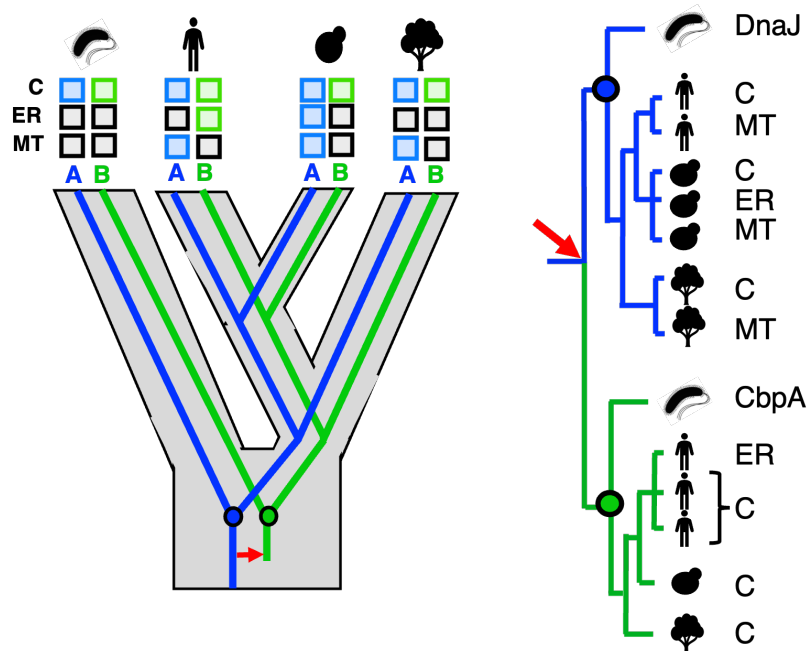
## 3.3 Wyniki

### 3.3.1 Hipotetyczny scenariusz ewolucji JDP zgodny z podziałem na klasy A i B.

Obecność przedstawicieli JDP zarówno klasy A, jak też klasy B u bakterii oraz występowanie zróżnicowanej liczby ich kopii u eukariontów (Ryc.3.2b) rodzi szereg pytań. Czy JDP klasy A występujące w komórkach eukariotycznych pochodzą od bakteryjnych przodków należących do klasy A oraz czy eukariotyczne JDP klasy B pochodzą od bakteryjnych JDP klasy B? Jakie są ewolucyjne zależności pomiędzy JDP klasy A i B występującymi w poszczególnych przedziałach komórki eukariotycznej? Czy białka JDP należące do klasy B' są ewolucyjnie spokrewnione z kanonicznymi JDP klasy B, klasy A czy też należy je zaliczyć do klasy C?

Na Rycinie 3.4 prezentuję scenariusz zgodny z hipotezą, że podział JDP na klasy A i B odzwierciedla ich powiązania ewolucyjne. Ze względu na to, że DnaJ (klasa A) występuje u 99% bakterii podczas gdy CbpA (klasa B) tylko u 51% [51] przyjął, że DnaJ reprezentuje przodka białek klasy A i B (Ryc. 3.4, linia niebieska). Według tego scenariusza u podstawy drzewa doszło do duplikacji genu kodującego JDP klasy A dając początek linii ewolucyjnej klasy B (Ryc. 3.4, linia zielona). Potomkowie tych pierwotnych linii ewolucyjnych to współczesne JDP należące do klas A i B zarówno u bakterii jak też w komórkach eukariotycznych. Kolejne duplikacje oraz utraty genów kodujących JDP klasy A i B w liniach ewolucyjnych spowodowały, że w niektórych przedziałach komórki eukariotycznej obecne są więcej niż jedna kopia lub też brak jest przedstawiciela jednej z klas. Zgodnie z tym scenariuszem wszystkie JDP klasy A są bliżej spokrewnione z DnaJ niż z CbpA (Ryc.3.4b, niebieska kropka), zaś przedstawiciele klasy B są bliżej spokrewnieni z CbpA niż z DnaJ (Ryc.3.4b, zielona kropka). Scenariusz ten sugeruje również, że bakteryjne DnaJ może być

modelem doświadczalnym eukariotycznych JDP klasy A, zaś CbpA modelem doświadczalnym JDP klasy B.



**Rycina 3.4. Hipotetyczny scenariusz ewolucji JDP zgodny z podziałem na klasy A i B.**

Scenariusz zakłada, że wszystkie białka klasy A (niebieski) są bliżej spokrewnione z bakteryjnym DnaJ, niż z CbpA oraz że wszyscy przedstawiciele klasy B (zielony) są bliżej spokrewnieni z bakteryjnym CbpA, niż z DnaJ. Czerwona strzałka oznacza duplikację pierwotnego genu kodującego JDP klasy A, co doprowadziło do powstania linii ewolucyjnej JDP klasy B. Przyjęłam założenie, że białka klasy A są pierwotne, ponieważ homologi DnaJ (klasa A) występują w 99% genomów bakterii, podczas gdy homologi CbpA (klasa B) występują jedynie w 51% proteomów. Lewy panel: schemat ewolucji JDP klas A i B na tle drzewa gatunków. Kwadraty pokazują obecność białek klasy A (niebieskie) oraz B (zielone) w komórkach bakterii, zwierząt, drożdży i roślin. Czarny kwadrat oznacza brak przedstawiciela danej klasy. Prawy panel: Kladogram przedstawiający topologię drzewa filogenetycznego JDP klas A i B zgodny ze scenariuszem ewolucyjnym przedstawionym na lewym panelu. C – cytoplazma, MT – mitochondria, ER – retikulum endoplazmatyczne. Wspólny przodek JDP klasy A – niebieska kropka, wspólny przodek JDP klasy B – zielona kropka.

**3.3.2 Analiza filogenetyczna JDP klasy A i B**

Aby przeanalizować zależności ewolucyjne pomiędzy JDP klasy A i B oraz zweryfikować hipotetyczny scenariusz ich ewolucji przedstawiony powyżej zgromadziłam 725 proteomów bakterii, archeonów, grzybów, zwierząt oraz roślin w celu zidentyfikowania wszystkich JDP klasy A i B należących do tych organizmów.

Na podstawie zestawienia dobrze poznanych i opisanych sekwencji aminokwasowych JDP klasy A i B (Zestawienie I) pochodzących z *E. coli*, *S. cerevisiae* oraz *H. sapiens* przygotowałam profile HMM zarówno dla domeny-J, jak i dla domen CTD I i CTD II, które wykorzystałam do przeszukania proteomów (patrz Metody), co pozwoliło mi zidentyfikować 8446 sekwencji JDP klasy A i B. Po ich zestawieniu usunęłam sekwencje powtarzające się, a także sekwencje znacząco dłuższe lub krótsze od pozostałych. W efekcie otrzymałam zestawienie 1588 sekwencji JDP klasy A i B (Zestawienie A/B) na którym przeprowadziłam kolejne analizy.

Drzewo filogenetyczne JDP klasy A i B zrekonstruowałam metodą maksymalnej wiarygodności (ML) z zastosowaniem modelu ewolucji LG+I+G4, który był najlepiej dopasowany do danych w zestawieniu. Aby prawidłowo ukorzeniec otrzymane drzewo przetestowałam trzy grupy zewnętrzne: (1) bakteryjne sekwencje CbpA (klasa B), (2) bakteryjne sekwencje DnaJ (klasa A) oraz (3) łącznie bakteryjne sekwencje CbpA i DnaJ stosując cztery testy statystyczne (patrz Metody). Drzewo, którego grupą zewnętrzną były sekwencje CbpA (klasa B) w każdym z testów okazało się najbardziej prawdopodobne (Tabela 1) (Ryc. 3.5a.)

**Tabela 1. Wyniki testów umiejscowienia korzenia na drzewie ML.**

Grupa zewnętrzna	logL	$\Delta L$	WKH	WSH	AU
CbpA	-259146.10	0	0.69	0.90	0.80
CbpA + DnaJ	-259146.10	$5.82 * e^{-11}$	0	0.03	0.39
DnaJ	-259146.10	$8.73 * e^{-11}$	0.04	0.15	0.51

logL – wartość logarytmu funkcji wiarygodności

$\Delta L$  – różnica wartości funkcji wiarygodności w stosunku do najlepiej dopasowanej topologii

WKH - test Kishino Hasegawa z zastosowaniem wag

WSH - test Shimodaira Hasegawa z zastosowaniem wag

AU - test AU (ang. *Approximatly Unbiased test*)

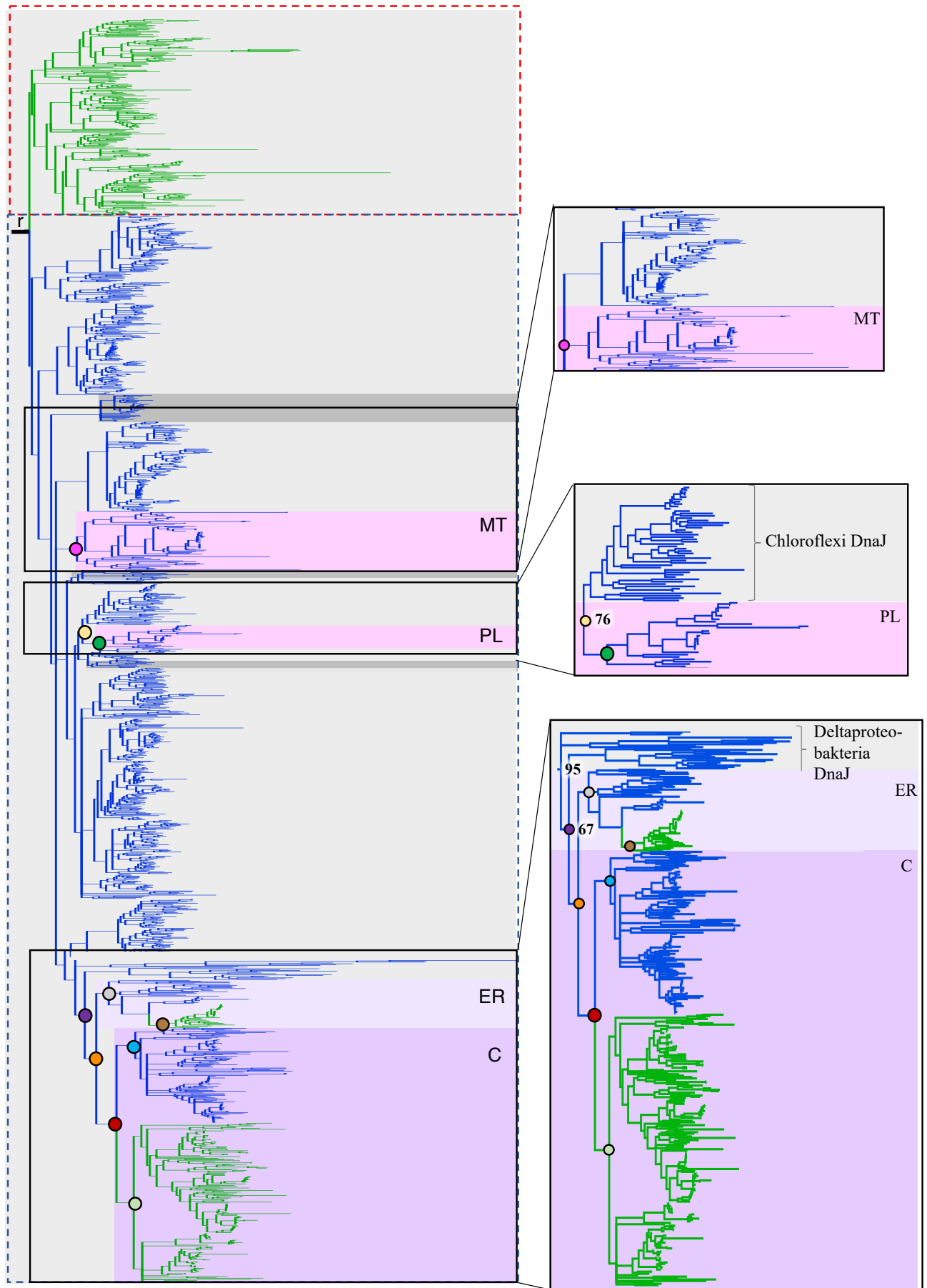
### 3.3.3 Pochodzenie ewolucyjne eukariotycznych JDP klasy A i B

Drzewo filogenetyczne JDP klasy A i B (Ryc. 3.5a) podzielone jest na dwa główne klady: jeden grupuje sekwencje bakteryjne CbpA (klasa B) (Ryc. 3.5 czerwona przerywana linia), natomiast drugi bakteryjne sekwencje DnaJ (klasa A), JDP z archeonów oraz wszystkie eukariotyczne JDP (klasa A i klasa B) (Ryc. 3.5 niebieska przerywana linia). Taka topologia drzewa sugeruje, że wszystkie eukariotyczne JDP

należące zarówno do klasy A, jak i B są bliżej spokrewnione z bakteryjnym DnaJ (klasa A), niż z CbpA (klasa B). Otrzymany wynik nie jest nie zgodny z hipotetycznym scenariuszem ewolucyjnym pokazanym na Ryc. 3.4., który zakładał, że wszystkie eukariotyczne JDP klasy A pochodzą od bakteryjnego przodka klasy A, natomiast wszystkie eukariotyczne JDP klasy B pochodzą od bakteryjnego przodka klasy B.

Sekwencje JDP z archeonów nie tworzą grupy monofiletycznej; formują trzy kłady, które są rozproszone na drzewie pomiędzy bakteryjnymi sekwencjami DnaJ. Żaden z kładów archeonów nie jest blisko spokrewnionymi z eukariotycznymi JDP. Może to oznaczać, że JDP występujące u archeonów pochodzą od białek bakteryjnych. Ich źródłem pochodzenia mógł być horyzontalny transfer genów. Ze względu na ten fakt oraz niską liczbę sekwencji JDP pochodzących od archeonów, dla uproszczenia, w dalszej części tej pracy nie wyróżniam ich i odnoszę się do nich jako do sekwencji bakteryjnych.

Eukariotyczne JDP należące do klas A i B nie tworzą na drzewie grupy monofiletycznej, co oznacza, że nie mają jednego, wspólnego przodka. Zamiast tego białka JDP klasy A występujące w mitochondriach i plastydach tworzą na drzewie oddzielne kłady (Ryc. 3.5 kropki odpowiednio różowa i zielona), każdy z nich ma wspólnego przodka z bakteryjnymi białkami DnaJ (Ryc. 3.5). Cytoplazmatyczne JDP klasy A i B tworzą z białkami występującymi w ER klasy A i B grupę monofiletyczną (Ryc.3.5, pomarańczowa kropka). Oznacza to, że eukariotyczne JDP funkcjonujące w cytoplazmie i ER mają wspólne pochodzenie, niezależnie od ich przynależności do klasy A i B.



**Rycina 3.5. Pochodzenie JDP klasy A i B.** (lewy panel) Drzewo ML (LG+I+G4) obejmujące 1588 sekwencji aminokwasowych JDP należących do klasy A i B. Grupę zewnętrzną stanowią bakteryjne sekwencje CbpA (korzeń oznaczony literą r). Czerwona przerywana linia oznacza kład CbpA (klasa B), niebieska – kład obejmujący bakteryjne DnaJ oraz eukariotyczne JDP klasy A i B. Linie niebieskie oznaczają JDP klasy A, natomiast linie zielone – JDP klasy B. Tło szare oznacza sekwencje pochodzące z organizmów prokariotycznych (jasnoszare – bakterii, ciemnoszare – archeonów), natomiast tła fioletowe – eukariotycznych pochodzących z mitochondriów (MT), plastydów (PL), retikulum endoplazmatycznego (ER) oraz cytoplazmy (C). (prawy panel) powiększone fragmenty drzewa obejmujące eukariotyczne JDP. Różowa kropka oznacza wspólnego przodka mitochondrialnych JDP; zielona kropka – wspólnego przodka plastydowych JDP, żółta kropka -wspólnego przodka plastydowych JDP i DnaJ z bakteryjnego kładu *Chloroflexi*; fioletowa kropka – wspólnego przodka JDP z ER oraz cytoplazmy i DnaJ z *Deltaproteobakterii*, pomarańczowa kropka – wspólnego przodka JDP z cytoplazmy i ER, szara kropka – wspólnego przodka JDP z ER, brązowa kropka – wspólnego przodka JDP klasy B z ER czerwona kropka – wspólnego przodka JDP cytoplazmatycznych klas A i B, jasnoniebieska kropka – wspólnego przodka cytoplazmatycznych JDP klasy A, jasnozielona kropka - wspólnego przodka cytoplazmatycznych JDP klasy B; Liczby obok węzłów oznaczają wsparcie *bootstrap*. Węzły o wsparciach *bootstrap* < 50 oznaczone są jako politomie.

W oparciu o otrzymane drzewo filogenetyczne nie można jednoznacznie określić pochodzenia mitochondrialnych JDP klasy A (Ryc. 3.5), ponieważ wsparcie węzła było zbyt niskie (*bootstrap* < 50). Kład mitochondrialnych JDP znajduje się na politomii i nie można na tej podstawie stwierdzić, z którą grupą bakteryjnych DnaJ mają one najbliższego wspólnego przodka. Jedyne, co na podstawie uzyskanego wyniku można powiedzieć, to że mitochondrialne JDP klasy A mają bliższego wspólnego przodka z DnaJ niż z CbpA. Można także stwierdzić, że mitochondrialne JDP klasy A mają różne pochodzenie od plastydowych JDP klasy A oraz od JDP klasy A i B z cytoplazmy i ER. W innym przypadku mitochondrialne JDP klasy A tworzyłyby grupę monofiletyczną z jednym z tych kładów.

Plastydowe JDP klasy A są najbliższej spokrewnione z białkami DnaJ występującymi w bakteriiach należących do grupy *Chloroflexi* (Ryc. 3.5), - wsparcie na rozgałęzieniu reprezentującym ich wspólnego przodka (Ryc.3.5, żółta kropka) jest wysokie (*bootstrap* 76). Takie pokrewieństwo plastydowych JDP klasy A jest zgodne z endosymbiotycznym pochodzeniem chloroplastów od bakterii fotosyntetyzujących.

JDP z cytoplazmy i ER tworzące na drzewie grupę monofiletyczną (Ryc. 3.5, pomarańczowa kropka) są najbliższej spokrewnione z bakteryjnymi białkami DnaJ pochodzącymi z *Deltaproteobakterii* Ryc. 3.5, fioletowa kropka). Wniosek ten wspiera

wysokie wsparcie węzła reprezentującego wspólnego przodka tych JDP (*bootstrap* 67) oraz węzła powyżej, reprezentującego wspólnego przodka DnaJ *Deltaproteobakterii* (*bootstrap* 95). Cytoplazmatyczne JDP klasy A oraz klasy B tworzą grupę monofiletyczną podzieloną na dwa siostrzane klady jeden reprezentujący białka klasy A i drugi reprezentujący białka klasy B zarówno kanoniczne jak też białka B'. (Ryc. 3.5, kropki odpowiednio jasnoniebieska oraz jasnozielona), (Ryc. 3.5, czerwona kropka). Kładem siostrzanym do cytoplazmatycznych JDP jest kład obejmujący wszystkie JDP z ER (Ryc. 3.5, szara kropka) – zarówno przedstawicieli klasy A występujących u roślin i grzybów, jak też przedstawicieli klasy B występujących u zwierząt.

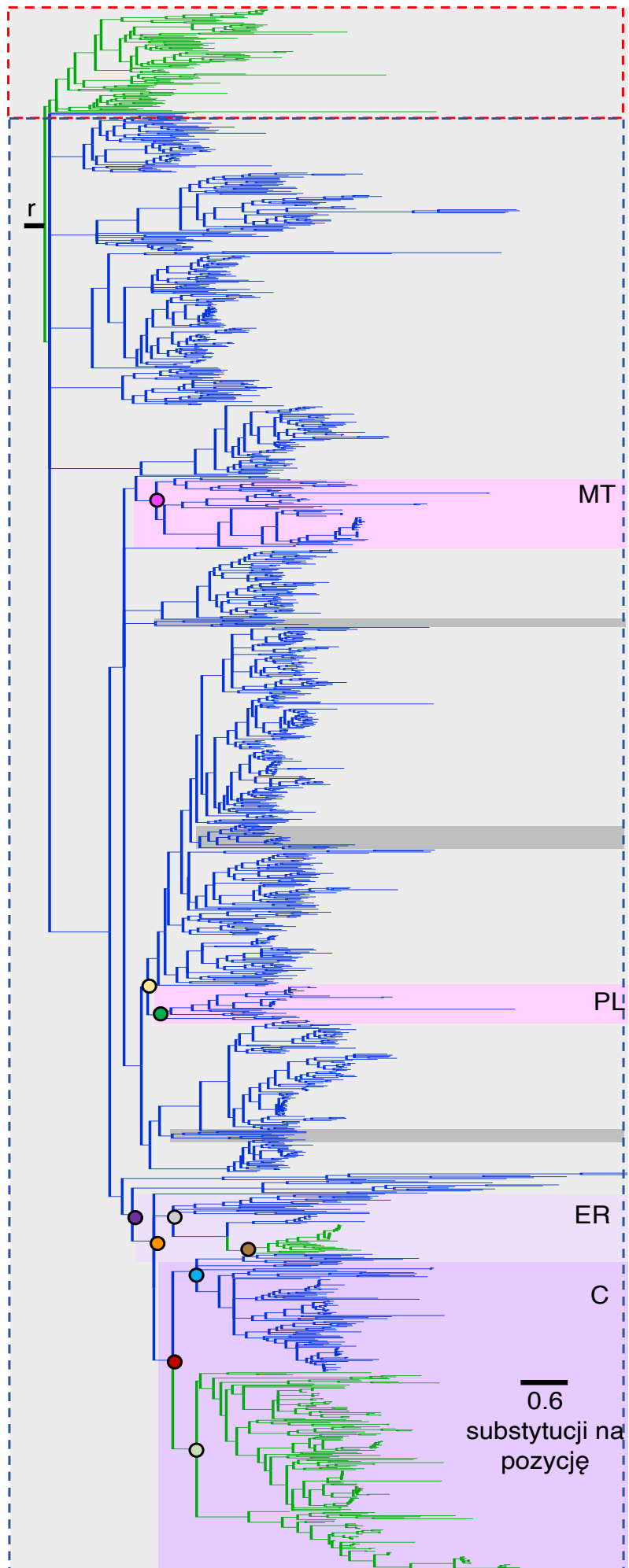
### 3.3.4 Weryfikacja topologii drzewa JDP klas A i B

#### *Analiza ML z zastosowaniem mieszanego modelu ewolucji C20*

Aby zweryfikować w jakim stopniu otrzymana przez mnie topologia drzewa JDP klas A i B jest wrażliwa na kompozycję aminokwasową poszczególnych kolumn w zestawieniu zastosowałam mieszany model ewolucji (C20). Model C20 grupuje kolumny w zestawieniu o podobnej kompozycji aminokwasowej w dwadzieścia kategorii. Do każdej z kategorii kompozycji dopasowywany jest osobny model ewolucji, a funkcja wiarygodności jest liczona dla sumy wszystkich pozycji.

Drzewo uzyskane przy użyciu modelu C20 (Ryc 3.6) wspiera topologię otrzymaną przy zastosowaniu modelu LG+I+G4 w odniesieniu do pochodzenia eukariotycznych JDP klasy A i B od bakteryjnych DnaJ - wszystkie eukariotyczne JDP tworzą kład z bakteryjnym DnaJ (Ryc. 3.6, przerywana, niebieska linia), natomiast grupę zewnętrzną stanowią bakteryjne sekwencje CbpA (Ryc. 3.6, przerywana, czerwona linia). Drzewo C20 wspiera również wspólne pochodzenie JDP z cytoplazmy i ER. – białka te tworzą grupę monofiletyczną (Ryc. 3.6, pomarańczowa kropka) oraz to, że zarówno białka plastydowe jak i mitochondrialne są bliżej spokrewnione z bakteryjnymi DnaJ niż z innymi eukariotycznymi JDP (Ryc. 3.6, kropka odpowiednio żółta i różowa). Podsumowując topologia drzewa uzyskanego z użyciem modelu C20 (Ryc. Ryc. 3.6) jest bardzo podobna do topologii uzyskanej z zastosowaniem modelu LG+I+G4.



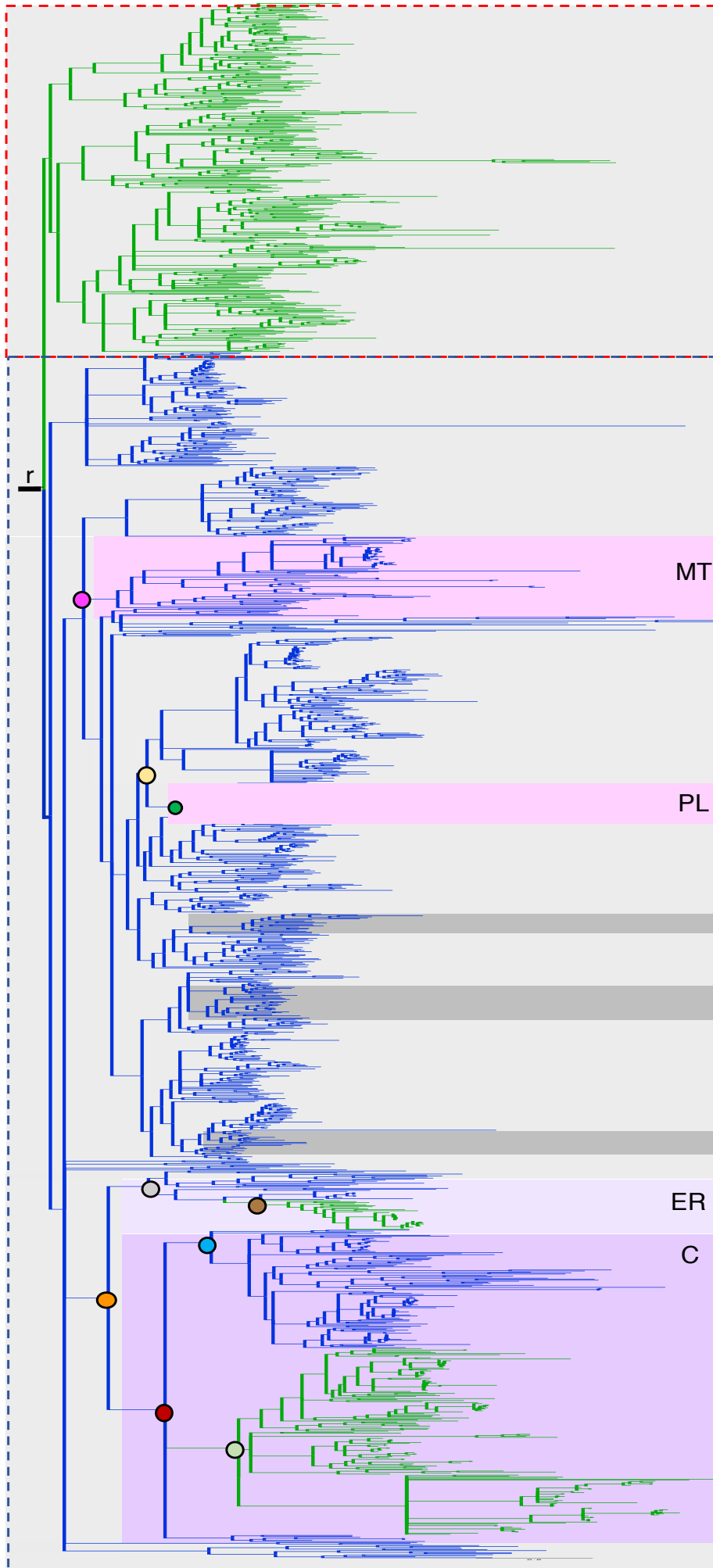


**Rycina 3.6. Drzewo ML JDP klasy A i B otrzymane z zastosowaniem modelu C20.** Grupę zewnętrzną stanowią sekwencje CbpA (korzeń oznaczony literą r). Czerwona przerywana linia oznacza kład CbpA, niebieska – kład obejmujący bakteryjne DnaJ oraz eukariotyczne JDP klasy A i B. Linie niebieskie oznaczają JDP klasy A, natomiast linie zielone – JDP klasy B. Tło szare oznacza sekwencje pochodzące z organizmów prokariotycznych (jasnoszare – bakterii, ciemnoszare – archeonów), natomiast tła fioletowe – eukariotycznych pochodzących z mitochondriów (MT), plastydów (PL), retikulum endoplazmatycznego (ER) oraz cytoplazmy (C). Różowa kropka oznacza wspólnego przodka mitochondrialnych JDP; zielona kropka – wspólnego przodka plastydowych JDP, żółta kropka -wspólnego przodka plastydowych JDP i DnaJ z bakteryjnego kładu Chloroflexi; fioletowa kropka – wspólnego przodka JDP z ER oraz cytoplazmy i DnaJ z Deltaproteobakterii, pomarańczowa kropka – wspólnego przodka JDP z cytoplazmy i ER, szara kropka – wspólnego przodka JDP z ER, brązowa kropka – wspólnego przodka JDP klasy B z ER czerwona kropka – wspólnego przodka JDP cytoplazmatycznych klas A i B, jasnoniebieska kropka – wspólnego przodka cytoplazmatycznych JDP klasy A, jasnozielona kropka - wspólnego przodka cytoplazmatycznych JDP klasy B; .Węzły o wsparciach *bootstrap* < 50 oznaczone są jako politomie.

Jako kolejny test topologii drzewa ML LG+I+G4 zastosowałam metodę *jackknife*, która odpowiada na pytanie czy różne części zestawienia wspierają tą samą topologię drzewa. Żeby otrzymać drzewo *jackknife* przygotowałam 100 zestawień, w których losowo usunęłam 40% pozycji Zestawienia A/B i na podstawie każdego z nich obliczyłam drzewo ML stosując model LG+I+G4. Konsensus otrzymanych drzew jest ostatecznym wynikiem w tej metodzie.

Drzewo *jackknife* również wspiera scenariusz ewolucyjny, w którym wszystkie eukariotyczne JDP są bliżej spokrewnione z bakteryjnymi DnaJ, niż z CbpA (Ryc. 3.7, linie przerywane) a JDP funkcjonujące w cytoplazmie i ER mają wspólnego przodka (Ryc. 3.7, pomarańczowa kropka), jednakże kład, który tworzą znajduje się na politomii, więc nie można określić z którą grupą bakteryjnych DnaJ mają najbliższego wspólnego przodka. Podobnie jak w dwóch poprzednich analizach JDP z cytoplazmy i ER oraz JDP mitochondrialne i plastydowe tworzą niezależne grupy monofiletyczne.

Wsparcia węzłów drzewa otrzymanego metodą *jackknife* są niższe niż drzew otrzymanych pozostałymi metodami, ponieważ niektóre z prób losowych posiadają większą zawartość brakujących pozycji w zestawieniu – a więc sygnał filogenetyczny jest obniżony. Efektem tego jest większa liczba politomii na drzewie *jackknife* w porównaniu z drzewami ML LG+I+G4 oraz C20. Jednak, topologia drzewa *jackknife* nie różni się znacząco od topologii drzew LG+I+G4 i C20.



### **Rycina 3.7. Drzewo ML JDP klasy A i B otrzymane z zastosowaniem metody *jackknife*.**

Grupę zewnętrzną stanowią sekwencje CbpA (korzeń oznaczony literą r). Czerwona przerywana linia oznacza kład CbpA, niebieska – kład obejmujący bakteryjne DnaJ oraz eukariotyczne JDP klasy A i B. Linie niebieskie oznaczają JDP klasy A, natomiast linie zielone – JDP klasy B. Tło szare oznacza sekwencje pochodzące z organizmów prokariotycznych (jasnoszare – bakterii, ciemnoszare – archeonów), natomiast tła fioletowe – eukariotycznych pochodzących z mitochondriów (MT), plastydów (PL), retikulum endoplazmatycznego (ER) oraz cytoplazmy (C). Różowa kropka oznacza wspólnego przodka mitochondrialnych JDP; zielona kropka – wspólnego przodka plastydowych JDP, żółta kropka – wspólnego przodka plastydowych JDP i DnaJ z bakteryjnego kładu Chloroflexi; fioletowa kropka – wspólnego przodka JDP z ER oraz cytoplazmy i DnaJ z Deltaproteobakterii, pomarańczowa kropka – wspólnego przodka JDP z cytoplazmy i ER, szara kropka – wspólnego przodka JDP z ER, brązowa kropka – wspólnego przodka JDP klasy B z ER czerwona kropka – wspólnego przodka JDP cytoplazmatycznych klas A i B, jasnoniebieska kropka – wspólnego przodka cytoplazmatycznych JDP klasy A, jasnozielona kropka – wspólnego przodka cytoplazmatycznych JDP klasy B; .Węzły o wsparciach < 50 oznaczone są jako politomie.

### *Wzajemne pokrewieństwo JDP klasy A i B funkcjonujących w cytoplazmie i ER*

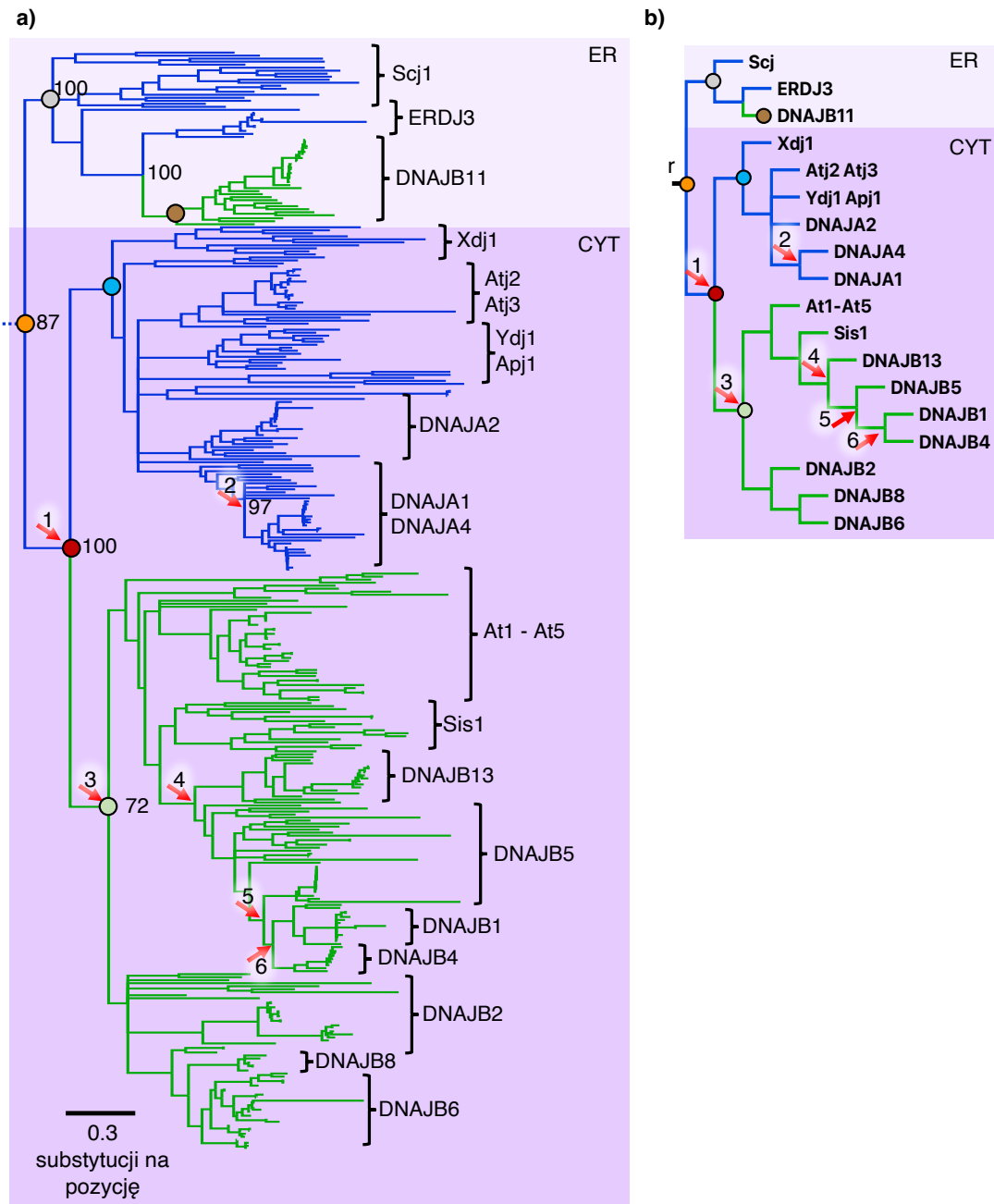
Z perspektywy moich badań najbardziej interesujący jest fragment drzewa zawierający sekwencje JDP funkcjonujących w cytoplazmie i ER, zawierający JDP klasy B oraz B'. Na ich wzajemnych powiązaniach ewolucyjnych skupię się w dalszej części rozdziału (Ryc. 3.8).

Z mojej analizy wynika, że cytoplazmatyczne białka klasy A tworzą grupę siostrzaną względem białek klasy B, co sugeruje, że klasy A i B powstały w wyniku duplikacji genu kodującego ich wspólnego przodka (Ryc. 3.8, strzałka z numerem 1). Po duplikacji białka klasy A i B ewoluowały niezależnie. Wsparcie węzła, na którym doszło do duplikacji wynosi 100, oznacza to, że wystąpił on w każdej próbie statystycznej *bootstrap*, co implikuje dużą pewność wyniku.

W kładzie cytoplazmatycznych JDP klasy A grupę zewnętrzną stanowią drożdżowe białka Xdj1, natomiast pozostałe grupy białek (roślinne Atj2/3, drożdżowe Ydj1, Apj1 oraz zwierzęce DNAJA2/1/4) tworzą politomię, ponieważ wsparcia na węzłach prowadzących do każdej z grup były niskie (*bootstrap* < 50). Z kolei topologia kładu obejmującego ludzkie sekwencje (DNAJA1/2/4) sugeruje, że białka A1 oraz A4, tworzące wysoko wsparty kład (*bootstrap* 97), powstały na skutek duplikacji genu, która zaszła późno w historii ich linii ewolucyjnej (Ryc. 3.8, czerwona strzałka z numerem 2).

W kladzie cytoplazmatycznych białek klasy B grupę zewnętrzną stanowią białka klasy B' – DNAJB2, DNAJB6 oraz DNAJB8 tworzące grupę monofiletyczną, co sugeruje, że gen kodujący wspólnego przodka cytoplazmatycznych JDP klasy B uległ duplikacji, po której jedna z kopii różnicowała się tworząc klasę B' (Ryc. 3.8, czerwona strzałka z numerem 3). Wsparcie węzła reprezentującego wspólnego przodka kanonicznych B oraz B' jest satysfakcjonujące (*bootstrap* 72). Wewnątrz kladu B' grupę zewnętrzną stanowią DNAJB2, z których najprawdopodobniej wyewoluowały białka DNAJB6 oraz DNAJB8. Otrzymany wynik jest zgodny z poprzednimi analizami pokazującymi przynależność białek B6 i B8 do jednego kladu [75]. W kladzie skupiającym białka kanonicznej klasy B grupę zewnętrzną stanowią roślinne At1-At5. Zaś drożdżowe białka Sis1 mają wspólnego przodka z białkami ludzkimi (DNAJB1, DNAJB4, DNAJB5, DNAJB13), co oznacza, że są z nimi najbliższej spokrewnione. Wynik ten jest zgodny z drzewem życia, na którym królestwa zwierząt oraz grzybów formują kład tworząc supergrupę Opisthokonta. Topologia kladu obejmująca ludzkie JDP klasy B sugeruje, że w ich linii ewolucyjnej doszło do trzech duplikacji genu prowadzących do powstania czterech ludzkich białek klasy B (Ryc. 3.8, czerwone strzałki 4,5,6). Pierwsza duplikacja (numer 4) dała początek DNAJB13, kolejna duplikacja (numer 5) - DNAJB5, ostatnia duplikacja (numer 6) spowodowała powstanie DNAJB1 oraz DNAJB4.

JDP klasy A i B funkcjonujące w ER są najbliższej spokrewnione z JDP pochodzącymi z cytoplazmy. Węzeł reprezentujący ich wspólnego przodka jest wysoko wsparty (*bootstrap* 87) (Ryc. 3.8, czerwona kropka). W obrębie kladu JDP z ER grupę zewnętrzną stanowią sekwencje pochodzące z grzybów (Scj1 – klasa A) (*bootstrap* 100). Sekwencje ludzkie DNAJB13 (klasa B) - posiadające częściowy rejon wiążący jony cynku oraz roślinne ERDJ3 (klasa A) tworzą siostrzane grupy monofiletyczne, a węzeł, który do nich prowadzi jest bardzo wysoko wsparty (*bootstrap* 100).



**Rycina 3.8. Zależności filogenetyczne pomiędzy JDP z cytoplazmy i ER.** a) Fragment drzewa ML LG+I+G4 przedstawiający białka funkcjonujące w cytoplazmie i ER. Liczby znajdujące się obok węzłów drzewa oznaczają wsparcie *bootstrap*. Linie niebieskie prowadzą do białek klasy A, natomiast zielone – klasy B. Jasne tło oznacza sekwencje pochodzące z ER, natomiast ciemne – z cytoplazmy. Klady podpisane są nazwami białek ludzkich, drożdżowych oraz roślinnych, które do nich należą. Strzałkami zaznaczone są węzły drzewa, gdzie doszło do duplikacji prowadzącej do powstania 1 – cytoplazmatycznych JDP klas A i B, 2- DNAJA1 oraz DNAJA4, 3 – cytoplazmatycznych JDP klas B i B', 4 – DNAJB13, 5 – DNAJB5, 6 – DNAJB1 oraz DNAJB4. b) kladogram będący odwzorowaniem topologii drzewa z panelu a). pomarańczowa kropka oznacza wspólnego przodka JDP z cytoplazmy i ER, szara kropka – wspólnego przodka JDP z ER, brązowa kropka – wspólnego przodka JDP klasy B z ER, czerwona kropka – wspólnego przodka JDP cytoplazmatycznych klas A i B, jasnoniebieska kropka – wspólnego przodka cytoplazmatycznych JDP klasy A, jasnozielona kropka – wspólnego przodka cytoplazmatycznych JDP klasy B;

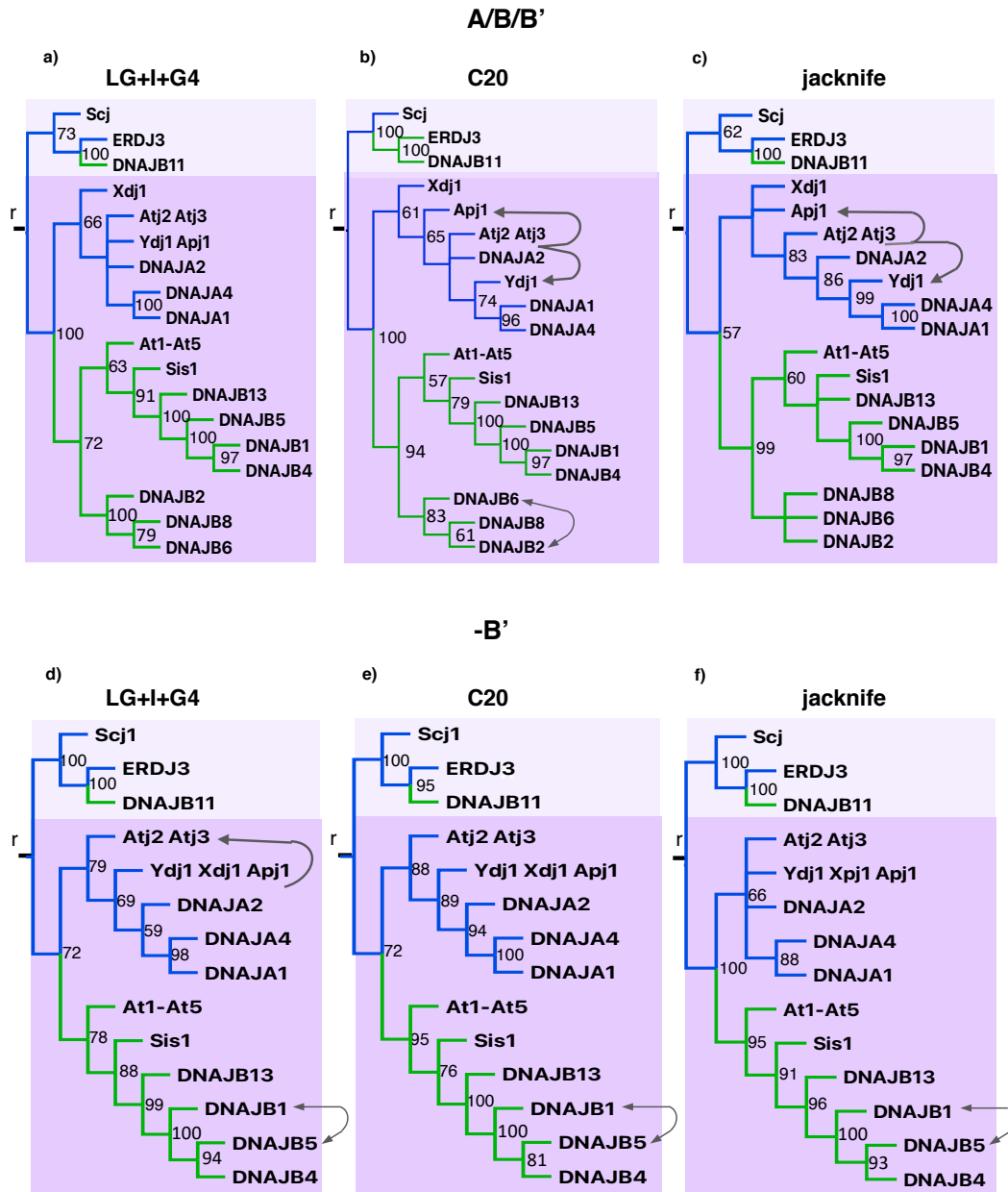
W dalszej części rozdziału analizuję podobieństwa i różnice pomiędzy topologią fragmentu drzewa ML LG+I+G4 obejmującą JDP z cytoplazmy i ER, a drzewami otrzymanymi za pomocą modelu C20 i metody *jackknife*. Dla łatwiejszego porównania topologii drzew otrzymanych za pomocą różnych metod posługuję się kladogramami (Ryc. 3.8b), które przedstawiają wzajemne pokrewieństwa między białkami, jednocześnie ignorując długości gałęzi drzewa. Ponieważ topologia drzewa ML LG+I+G4 (Ryc. 3.5) pokrywa się w znacznym stopniu z topologiami, które otrzymałam wykorzystując model C20 (Ryc. 3.6), oraz metodę *jackknife* (Ryc. 3.7), porównując je skupiam się na różnicach.

Pomiędzy drzewami LG+I+G4 (Ryc. 3.9a), a C20 (Ryc. 3.9b) występują jedynie dwie różnice: (1) zmieniła się pozycja białek pochodzących z grzybów – Ydj1 oraz Apj1 znalazły się na osobnych gałęziach drzewa, a nie politonii, jak w drzewie LG. Ydj1 na drzewie C20 jest bliżej spokrewniony z ludzkimi białkami DNAJA1 oraz DNAJA4, niż z pozostałymi białkami pochodzącymi z grzybów. (2) druga różnica widoczna jest wśród białek B', gdzie DNAJB2, a nie DNAJB8 tworzy grupę siostrzana z DNAJB6. Podsumowując, różnice między drzewami nie są znaczące i występują jedynie dla białek charakteryzujących się szybkim tempem ewolucji (długie gałęzie na drzewie LG+I+G4).

Topologia drzewa *jackknife* (Ryc. 3.9c) jest zgodna z topologią drzewa LG+I+G4 w głównych węzłach, jednak wsparcie węzła rozdzielającego cytoplazmatyczne JDP na klasę A oraz B jest niższe (wsparcie 57) niż w przypadku dwóch pozostałych metod gdzie powtórzył się on we wszystkich próbach statystycznych (*bootstrap* 100), co może być spowodowane faktem, że niektóre z prób *jackknife* posiadają większą zawartość brakujących pozycji w zestawieniu (ang. *gaps, missing data*), co skutkuje obniżeniem sygnału filogenetycznego. Różnice widoczne są również w kladzie białek należących do klasy A – białka pochodzące z grzybów zmieniły pozycję. Apj1 oraz Xdj1 tworzą politomię z kladem utworzonym przez pozostałe białka klasy A, natomiast Ydj1 znalazł się w kladzie z białkami ludzkimi, pomiędzy DNAJA2, a grupą monofiletyczną utworzoną przez DNAJA1 oraz DNAJA4. Wszystkie cytoplazmatyczne białka klasy B tworzą klad, natomiast wsparcie węzła prowadzącego do DNAJB13 jest niskie, więc pozostaje on nierozwiązany. Nierozwiązane pozostają także zależności w kladzie B'. Podsumowując powyższe wyniki można zauważyć, że topologie drzew otrzymanych

różnymi metodami nie są znacząco różne – wspierają ten sam scenariusz ewolucyjny: JDP cytoplazmatyczne tworzą siostrzane grupy monofiletyczne obejmujące klasy A oraz B, a JDP klasy B' są najbliższej spokrewnione z JDP klasy B. W przypadku JDP klasy A i B z ER topologie wszystkich analizowanych drzew są identyczne. W tym przypadku zastanawiające jest jednak, że ludzkie białko DNAJB11 (klasa B) jest siostrzane dla roślinnego ERDJ3 (klasa A) a nie dla drożdżowego Scj1 (klasa A). Tak więc topologia drzewa JDP z ER nie jest zgodna z filogenezą gatunków, gdzie zwierzęta i grzyby tworzą kład Opisthokonta a rośliny stanowią grupę zewnętrzną względem tego kladu.





**Rycina 3.9. Porównanie topologii drzew ML dla fragmentu drzewa obejmującego JDP z cytoplazmy i ER.** Kladogramy odpowiadające topologiom drzew filogenetycznych otrzymanych metodą maksymalnej wiarygodności: Panel górny - dla pełnego zestawienia JDP z wykorzystaniem: a) modelu LG+I+G4 b) z wykorzystaniem modelu mieszanego C20 – oraz c) metodą *jackknife*. Panel dolny - dla zestawienia JDP -B' z wykorzystaniem: d) modelu LG+I+G4 e) z wykorzystaniem modelu mieszanego C20 oraz f) metodą *jackknife*. Różnice względem drzewa LG+I+G4 dla Zestawienia A/B/B' zaznaczone są za pomocą szarych strzałek. Linie niebieskie prowadzą do białek klasy A, linie zielone – klasy B. Tło jaśniejsze oznacza sekwencje pochodzące z ER, natomiast tło ciemniejsze – cytoplazmy. Klady podpisane są nazwami białek ludzkich, drożdżowych oraz roślinnych, które do nich należą.

### 3.3.5 Analiza filogenetyczna ML JDP bez białek klasy B'

Kolejnym krokiem, który podjęłam w celu testowania topologii drzewa JDP klas A i B było wykluczenie białek B' z Zestawienia A/B (Zestawienie A/B-B'). Z powodu braku kanonicznych domen CTD I oraz CTD II oraz szybkiego tempa ewolucji sekwencji tych białek zestawiały się one najslabiej. Do analizy Zestawienia A/B-B' zastosowałam takie same metody jak w przypadku Zestawienia A/B – model LG+I+G4, model mieszany C20 oraz metodę *jackknife*.

W pierwszej kolejności obliczyłam filogenezę ML używając modelu LG+I+G4. Uzyskana topologia pokrywała się z poprzednimi, jeżeli chodzi o główne węzły drzewa: białka cytoplazmatyczne oraz pochodzące z ER miały wspólnego przodka, a w obrębie kladu białek cytoplazmatycznych były dwie grupy monofiletyczne; jedna skupiająca białka klasy A i druga skupiająca białka klasy B. W kladzie ER zależności między białkami pozostały niezmiennione (Ryc. 3.9d).

W przypadku drzewa LG+I+G4 obliczonego w oparciu o zestawienie A/B-B' w kladzie cytoplazmatycznych białek A nie występują politomie – grupę zewnętrzną stanowią białka roślin, a prowadzący do nich węzeł jest satysfakcjonująco wsparty (*bootstrap* 72), białka grzybów (Ydj1, Xdj1, Apj1) tym razem tworzą grupę monofiletyczną, co jest zgodne z poprzednio publikowanymi wynikami [90]. Białka ludzkie ponownie tworzą kład będący grupą siostrzaną do białek grzybów, w obrębie którego DNAJA1 oraz DNAJA4 są spokrewnione ze sobą bliżej niż z DNAJA2. W kladzie cytoplazmatycznych JDP klasy B grupę zewnętrzną ponownie tworzą białka roślinne, a białka grzybów (Sis1) są w relacji siostrzanej do białek zwierzęcych: DNAJB13, DNAJB1, DNAJB5 oraz DNAJB4. Jedyna różnica, którą można zaobserwować w kladzie JDP klasy B w porównaniu do analizy z zastosowaniem Zestawienia A/B to zamiana miejsc DNAJB1 oraz DNAJB5 – w uzyskanym drzewie to DNAJB5, a nie DNAJB1 tworzy grupę monofiletyczną z DNAJB4.

Topologia drzewa uzyskanego z zastosowanie modelu C20 (Ryc. 3.9e) jest identyczna jak w przypadku modelu LG+I+G4. Wsparcia *bootstrap* wszystkich węzłów drzewa są wyższe niż w przypadku pełnego zestawienia (Zestawienie A/B). W przypadku analizy *jackknife* topologia drzewa filogenetycznego dla cytoplazmatycznych JDP klasy B oraz dla JDP klasy B z ER pokrywają się z topologią uzyskaną przy użyciu modeli LG+I+G4 i C20 (Ryc. 3.9f). W przypadku cytoplazmatycznych białek

klasy A zależności ewolucyjne nie zostały rozwiązane z wysokim wsparciem statystycznym, co jest zaznaczone w postaci politomi (Ryc. 3.9f). Jedynie białka DNAJA1 oraz DNAJA4 tworzą kład, co wspierają wszystkie uzyskane przeze mnie wyniki. Pozostałe białka klasy A tworzą politomię. Podsumowując, usunięcie sekwencji B' zwiększyło statystyczne wsparcie drzewa JDP nie zmieniając w istotny sposób jego topologii.

### 3.3.6 Analiza Bayesowska

Kolejnym krokiem mającym na celu weryfikację zależności ewolucyjnych pomiędzy białkami JDP z cytoplazmy i ER była analiza z zastosowaniem metody Bayesowskiej. Ze względu na ograniczenia techniczne (czas analizy oraz moc obliczeniowa, którą dysponowałam) zdecydowałam się na ograniczenie analizy Bayesowskiej do zestawienia obejmującego sekwencje białek JDP z cytoplazmy i ER bez sekwencji klasy B' (Zastawienie C-B'/ER). Analizę przeprowadziłam stosując cztery modele ewolucji: (1) CAT1LG, (2) CATLG, (3) C60 oraz (4) W5 (patrz Metody)

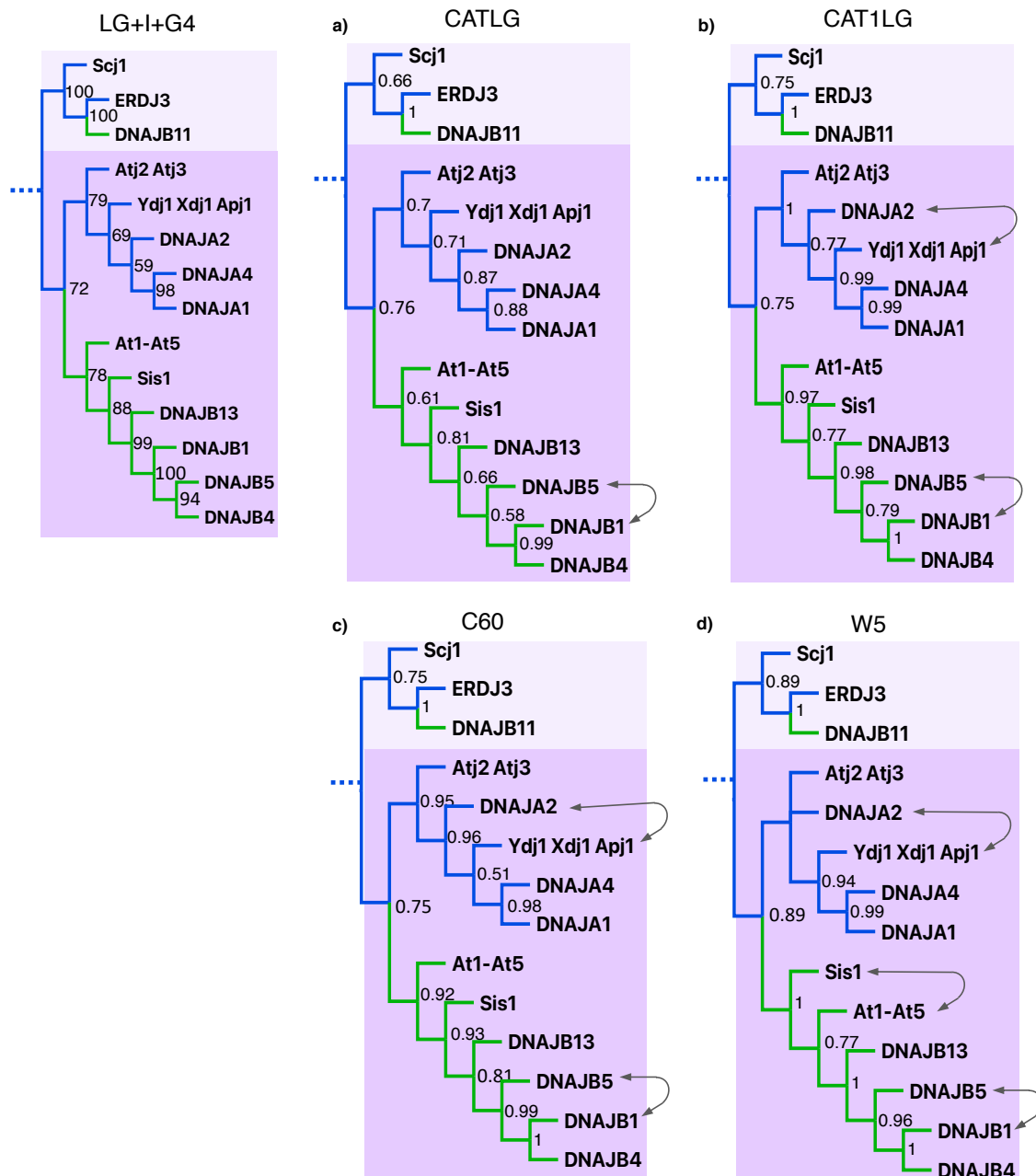
Topologie drzew otrzymanych metodą Bayesowską porównałam z najlepiej wspartym statystycznie drzewem otrzymanym metodą ML z zastosowaniem modelu LG+I+G4 dla zestawienia niezawierającego białek B' (Zestawienie A/B-B') (Rys. 3.9a). W przypadku analizy Bayesowskiej wsparciem drzewa nie są losowe powtórzenia *bootstrap* (których wartości znajdują się w przedziale od 1 do 100), a wartość prawdopodobieństwo *a posteriori* (którego wartość znajduje się w przedziale od 0 do 1).

Topologia drzewa, które uzyskałam stosując model (2) zezwalający na heterogenność pozycji aminokwasowych (CATLG, Ryc. 3.10a) w analizie Bayesowskiej była najbardziej podobna do topologii ML uzyskanej stosując model LG+I+G4. Jedyną różnicą jest położenie grupy DNAJB1 względem DNAJB4, według analizy Bayesowskiej CATLG białka DNAJB1 są bliżej spokrewnione z DNAJB4 niż DNAJB5.

Topologia drzewa, które uzyskałam stosując model 1 – CAT1LG (Ryc. 3.10b) oraz model 3 - C60 (Ryc. 3.10c) w analizie Bayesowskiej była identyczna. Względem topologii ML-LG+I+G4 zaobserwowałam jedynie dwie różnice: pierwszą jest zmiana

miejsce DNAJB4 i DNAJB1, drugą jest zmiana pozycji cytoplazmatycznych JDP klasy A pochodzących z grzybów (Ydj1, Xdj1, Apj1) względem ludzkiego DNAJA2 – tworzą one kład z pozostałymi białkami ludzkimi – DNAJA4 i DNAJA1.

Topologia, którą otrzymałam stosując model (4 - W5) odpowiada pozostałym otrzymanym przeze mnie topologiom drzew Bayesowskich, jeżeli chodzi o układ głównych węzłów drzewa, lecz różni się od pozostałych drzew (CATLG, CAT1LG, C60) położeniem kładu drożdżowych białek Sis1, który w przypadku tej analizy stanowią grupę zewnętrzną kładu cytoplazmatycznych białek klasy B (Ryc. 3.10d). Wewnątrz kładu cytoplazmatycznych JDP klasy B widoczna jest jeszcze jedna różnica - ponownie DNAJB4 jest bliżej spokrewnione z DNAJB1 niż z DNAJB5. W kładzie cytoplazmatycznych JDP klasy A białka pochodzące z grzybów (Ydj1, Xdj1, Apj1) tworzą kład z DNAJA1 oraz DNAJA4, które są ze sobą najbliższej spokrewnione, natomiast pozostałe węzły w klasie cytoplazmatycznych białek A pozostają nierozwiązane. Podsumowując różnice w topologii drzew Bayesowskich względem drzewa ML LG+I+G4 nie są znaczące. Nie dotyczą kluczowych dla tej analizy węzłów drzewa wskazujących na niezależne pochodzenie JDP klasy B w cytoplazmie i ER od białek klasy A.



**Rycina 3.10. Porównanie topologii drzew obliczonych metodą Bayesowską.** Kladogramy odpowiadające czterem topologiom drzew filogenetycznych otrzymanych metodą Bayesowską z wykorzystaniem modelu: a) CATLG b) CAT1LG c) C60 oraz d) W5. Różnice w topologii przedstawione są w porównaniu do kladogramu reprezentującego drzewo ML (LG+I+G4, lewy górny róg). Różnice te zaznaczone są za pomocą szarych strzałek. Linie niebieskie prowadzą do białek klasy A, natomiast zielone – klasy B. Tło jaśniejsze oznacza sekwencje pochodzące z ER, natomiast tło ciemniejsze – cytoplazmy. Klady podpisane są nazwami białek ludzkich, drożdżowych oraz roślinnych, które do nich należą.

### 3.3.7 Ewolucja reszt aminokwasowych zaangażowanych w oddziaływanie JDP klasy B z sekwencją EEVD Hsp70.

Unikatową cechą zarówno ludzkie jak też drożdżowych JDP klasy B funkcjonujących w cytoplazmie jest zdolność oddziaływania z sekwencją EEVD znajdująca się na C-końcu partnerskiego Hsp70. W przypadku drożdżowego JDP klasy B - Sis1 trzy reszty lizyny (K199, K202, K214) zlokalizowane w obrębie domeny CTDI oddziałują z motywem EEVD białka Hsp70 Ssa1 [64]. Oddziaływanie to, jest niezbędne do aktywacji domeny-J Sis1 poprzez uwolnienie helisy V rejonu GF hamującej aktywność domeny-J. Mechanizm wiązania EEVD oraz hamowania domeny J przez helisę V nie występuje u przedstawicieli JDP klasy A. Postanowiłam więc sprawdzić, na jakim etapie ewolucji białka klasy B uzyskały dodatkowe miejsce wiązania Hsp70.

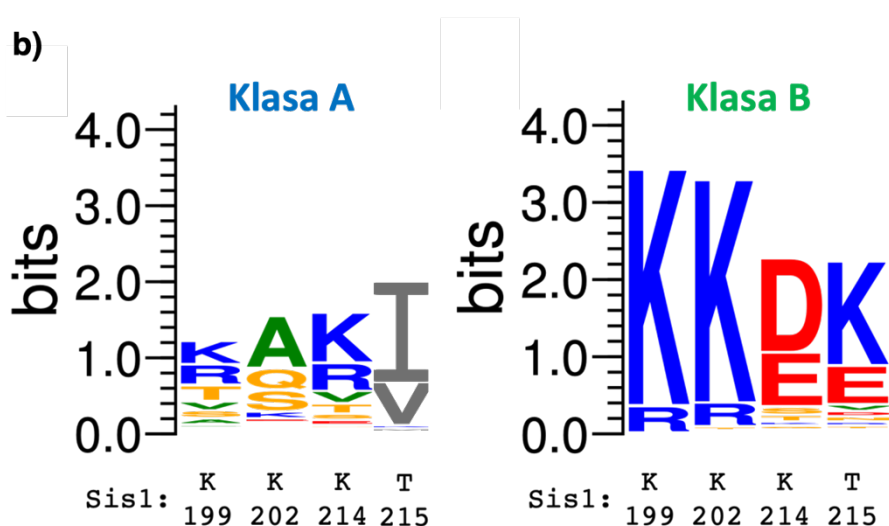
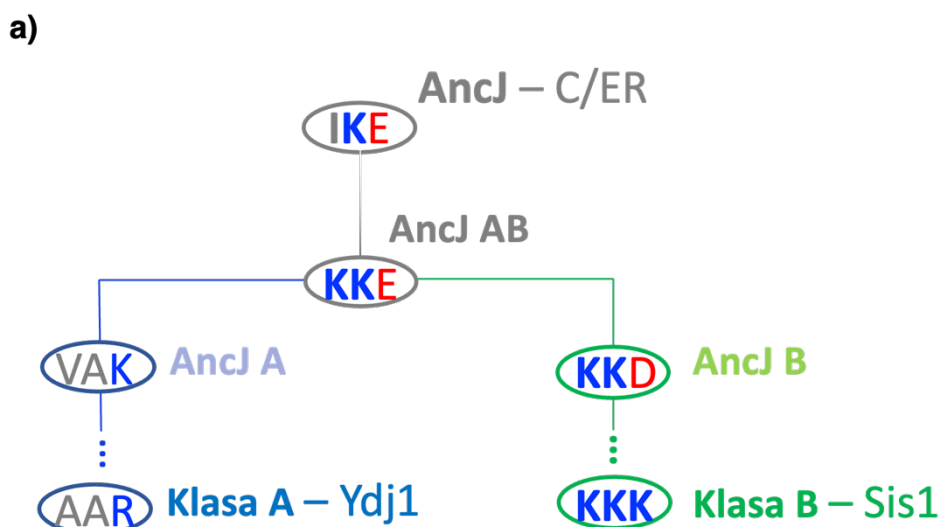
Aby odpowiedzieć na to pytanie przeprowadziłam rekonstrukcję sekwencji przodków cytoplazmatycznych JDP klasy A i B opierając się o zestawienie sekwencji zawierające cytoplazmatyczne oraz retikularne JDP. Do rekonstrukcji sekwencji przodków użyłam programu FastML wykorzystując empiryczną metodę Bayesowską oraz metodę ML do rekonstrukcji insercji i delecji. Do analizy wykorzystałam fragment drzewa ML oraz zestawienia obejmujący JDP z cytoplazmy i ER (Zestawienie C/ER), ponieważ wyniki moich analiz filogenetycznych wykazały, że białka te mają wspólne pochodzenie, niezależne od pozostałych eukariotycznych JDP.

Aby prześledzić ewolucję miejsca wiązania analizowałam pozycje homologiczne do K199, K202 oraz K214 Sis1 (Ryc. 3.11a) u wspólnego przodka białek JDP klasy A i B z cytoplazmy i ER (AncJ), wspólnego przodka JDP z cytoplazmy (AncJAB) oraz wspólnego przodka cytoplazmatycznych JDP klasy A (AncA) oraz klasy B (AncJB). Jako sekwencje referencyjne, reprezentujące białka współczesne, wykorzystałam przedstawicieli klas A i B u *S. cerevisiae* – odpowiednio Ydj1 oraz Sis1.

Jedna z lizyn zaangażowanych w wiązanie EEVD była obecna już u wspólnego przodka JDP z cytozolu i ER (AncJ). Druga lizyna (K199) pojawiła się na gałęzi łączącej AncJ z AncJ AB. Wspólny przodek cytoplazmatycznych JDP (AncJAB) posiadał dwie z trzech reszt lizyny - K202 oraz K199. Taki stan utrzymał się także u wspólnego przodka JDP klasy B (AncJB) natomiast u wspólnego przodka

cytoplazmatycznych JDP klasy A obie lizyny zastąpione zostały aminokwasami nieposiadającymi ładunku. Pojawiła się jednak reszta lizyny (K214), która nie występuje u przodków JDP klasy B. Współczesne Sis1 (klasa B) posiada trzy reszty lizyny: (K199, K202 oraz K214) zaangażowane w wiązanie EEVD, natomiast współczesne Ydj1 (klasa A) nie posiada pozytywnie naładowanych reszt w pozycjach 199 i 202 ma jednak resztę argininy (R214). Może to oznaczać, że wspólny przodek cytoplazmatycznych JDP klasy A i B posiadał reszty lizyny, które umożliwiły wiązanie EEVD a po duplikacji kodującego go genu cecha ta uległa zróżnicowaniu; zachowały ją białka klasy B zaś utraciły białka klasy A.

Aby zweryfikować ten scenariusz ewolucyjny postanowiłam sprawdzić czy homologiczne reszty lizyny są konserwowane u innych przedstawicieli JDP klasy B oraz jak wygląda homologiczny rejon w białkach klasy A. Do analizy użyłam zestawienia sekwencji, które wykorzystałam do analiz filogenetycznych (Zestawienie A/B). Konserwację na każdej z pozycji przedstawiłam w formie logo na rycinie 3.11b, gdzie wysokość słupka jest proporcjonalna do zakonserwowania danej pozycji zestawienia.



**Rycina 3.11. Rekonstrukcja miejsca oddziaływania EEVD z CTDI u przodków JDP z cytoplazmy i ER.** a) Rekonstrukcja miejsc homologicznych do K199, K202 oraz K214 (numeracja dla Sis1) - miejsca oddziaływania Sis1 (klasa B) z sekwencją EEVD Hsp70. Wspólny przodek JDP z cytoplazmy i ER – AncJ, wspólny przodek JDP z cytoplazmy - AncJ, wspólny przodek cytozoolowych JDP – AncJAB oraz wspólny przodek JDP klasy B - AncJ B. b) Konserwacja reszt lizyny w pozycjach homologicznych do K199, K202, K214, T215 Sis1 przedstawiona jako logo, gdzie wysokość słupka odzwierciedla poziom konserwacji na danej pozycji.

W grupie JDP klasy B dwie pierwsze lizyny – K199 oraz K202 są konserwowane u większości gatunków (Ryc. 3.11b), a w sekwencjach, w których na homologicznej pozycji nie występuje lizyna obecna jest arginina więc dodatni ładunek jest zachowany. Na pozycji homologicznej do K214 u większości przedstawicieli klasy B, podobnie jak u wspólnego przodka JDP klasy B (AncJB) znajduje się aminokwas



ujemnie naładowany – kwas asparaginowy lub kwas glutaminowy. Wyniki badań biochemicznych [69], że trzy reszty lizyny są niezbędne dla wiązania EEVD i aktywacji domeny J Sis1. Przyjrzałam się więc pozycji sąsiadującej z K214 w moim zestawieniu. W Sis1 na pozycji sąsiadującej z K214 występuje treonina, natomiast u ponad połowy pozostałych białek klasy B na tej pozycji znajduje się lizyna. Nie wiadomo czy u innych gatunków trzy lizyny są niezbędne dla prawidłowego oddziaływania z EEVD. Interesującym byłoby sprawdzenie czy mutacja lizyny w pozycji homologicznej do T215 wpływa na wiązanie EEVD i aktywację JDP klasy B u tych gatunków. U przedstawicieli JDP klasy A nie widać konserwacji na żadnej z pozycji homologicznych dla reszt lizyny zaangażowanych w wiązanie EEVD, a także na pozycji homologicznej do T215. Na trzech homologicznych pozycjach zmienność sekwencji jest duża, co nie jest zaskakujące, ponieważ JDP klasy A nie wiążą EEVD. Podsumowując, analiza sekwencji przodków oraz konserwacji ewolucyjnej miejsc zaangażowanych w wiązanie EEVD przez JDP klasy B pokazała, że dwie reszty lizyny zaangażowane w wiązanie EEVD występowały u wspólnego przodka cytoplazmatycznych JDP klasy A i B (AncJ AB) a po duplikacji genu zostały utrwalone w linii ewolucyjnej JDP klasy B. Ponad to wyniki mojej analizy pokazują, że udział trzech reszt lizyny w wiązaniu EEVD może być cechą unikatową dla drożdżowego Sis1- u przedstawicieli JDP klasy B innych gatunków zakonserwowane są tylko dwie reszty lizyny. Alternatywnie, JDP klasy B niemające lizyny w pozycji 214 mogą wykorzystywać do wiązania EEVD lizynę w pozycji 215. Hipoteza ta wymaga jednak doświadczalnej weryfikacji.

### 3.4 Podsumowanie i dyskusja

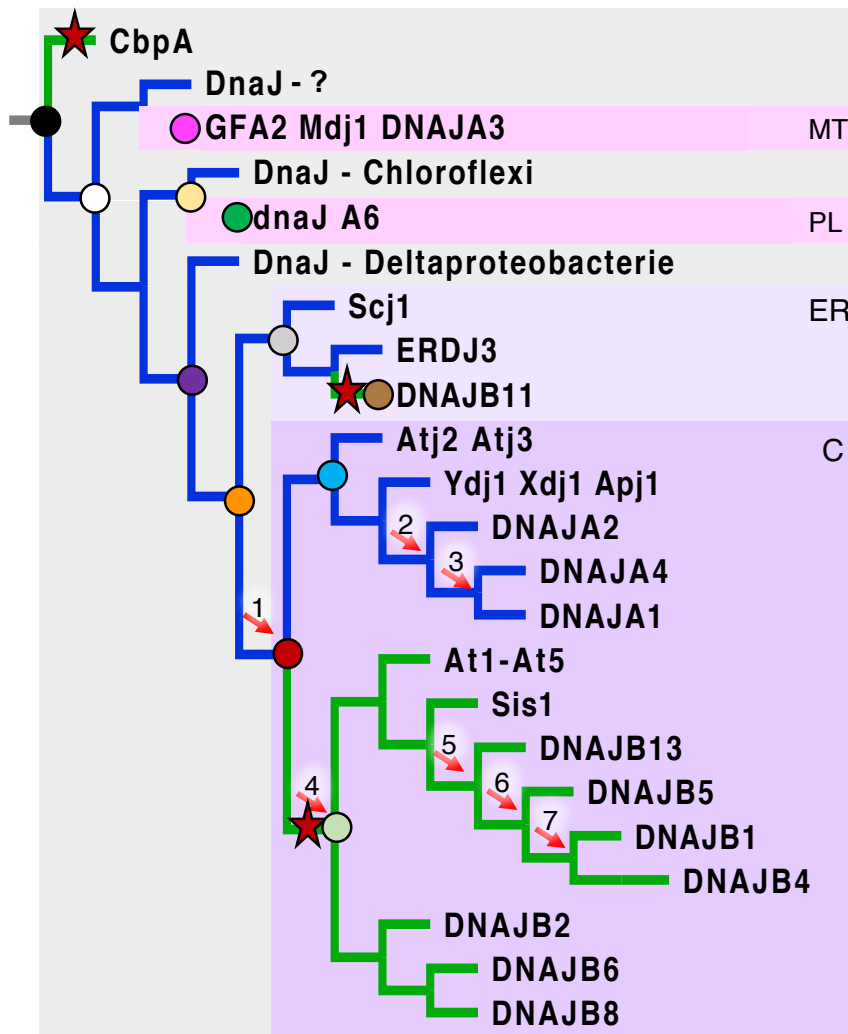
Opierając się na uzyskanych wynikach można wywnioskować, że JPD klasy A i B ewoluowały według scenariusza przedstawionego na Rycinie 3.12. Grupę zewnętrzną drzewa stanowią sekwencje bakteryjnych białek CbpA (klasa B) (Ryc. 3.12, czarna kropka). Otrzymana topologia drzewa JDP nie jest więc zgodna z hipotezą, że podział na klasy A i B odzwierciedla historię ewolucyjną tych białek – to znaczy, że eukariotyczne białka klasy B pochodzą od bakteryjnych przodków należących do klasy B, a eukariotyczne białka klasy A od bakteryjnych białek klasy A. Zamiast tego, otrzymane wyniki sugerują, że eukariotyczne JDP klasy B powstały w wyniku duplikacji genów kodujących JDP klasy A, spokrewnionych z bakteryjnymi DnaJ (klasa A). Otrzymane wyniki zmieniają nasze spojrzenie na temat pokrewieństwa ewolucyjnego między białkami należącymi do klas A i B co oznacza, że obecnie stosowana klasyfikacja JDP [31] nie odzwierciedla ich zależności ewolucyjnych, a bakteryjne JDP klasy B mogą nie być dobrym modelem doświadczalnym dla drożdżowych i ludzkich białek klasy B. Należy rozważyć opracowanie nowej klasyfikacji białek JDP, która lepiej odzwierciedlałaby historię ich ewolucji.

JDP funkcjonujące w mitochondriach, plastydach oraz w cytoplazmie/ER tworzą na drzewie grupy monofiletyczne co oznacza, że ewoluowały one niezależnie od białek funkcjonujących w innych przedziałach komórkowych. Ponadto topologia drzewa zgodna jest ze scenariuszem, że każda z tych grup jest bliżej spokrewniona z bakteryjnym DnaJ (klasa A) niż z pozostałymi eukariotycznymi JDP. W przypadku każdej z tych grup szereg duplikacji oraz utraty genów kodujących białka JDP doprowadził do zróżnicowania liczby białek JDP funkcjonujących w poszczególnych przedziałach współczesnych komórek eukariotycznych.

Białka JDP z cytoplazmy oraz ER mają wspólne pochodzenie – węzeł reprezentujący ich wspólnego przodka występował na wszystkich drzewach filogenetycznych niezależnie od zastosowanej metody obliczeniowej oraz modelu ewolucji (Ryc. 3.12, pomarańczowa kropka). Białka te są najbliżzej spokrewnione z bakteryjnymi DnaJ pochodzącymi z Deltaproteobakterii, ponieważ z nimi tworzą grupę monofiletyczną (Ryc. 3.12, fioletowa kropka).

JDP z plastydów mają najbliższego wspólnego przodka z bakteryjnymi DnaJ z *Chloroflexia* (Ryc. 3.12, żółta kropka), czyli tak zwanych zielonych bakterii, które mogą prowadzić reakcję fotosyntezy. Teoria endosymbiozy mówi o pochodzeniu plastydów od Cyanobakterii, natomiast są publikacje pokazujące, że niektóre geny plastydowe pochodzą od innych grup bakterii między innymi z klasy *Chloroflexia* [91][92]. Taki scenariusz ewolucyjny zgodny jest z drzewem filogenetycznym JDP otrzymanym w tej pracy. Pochodzenie JDP mitochondrialnych pozostaje nierozwiązane. Topologia drzewa sugeruje, że nie są one bezpośredni spokrewnione z JDP z innych przedziałów komórki eukariotyczne- mitochondrialne JDP tworzą niezależny kład (Ryc. 3.12, zielona kropka).

Cytoplazmatyczne JDP należące do klas A i B stanowią na drzewie grupy siostrzane (Ryc. 3.12, czerwona kropka), można więc wnioskować, że cytoplazmatyczne JDP klasy B powstały na skutek duplikacji genu kodującego cytoplazmatyczne JDP klasy A (Ryc. 3.12, strzałka z numerem 1). Wewnątrz kładu białek klasy A grupę zewnętrzną stanowią JDP roślinne, następny węzeł drzewa prowadzi do kładu białek grzybów – Ydj1, Xdj1 i Apj1 oraz kładu białek ludzkich – DNAJA2, DNAJA1 i DNAJA4, ten węzeł występował zarówno w analizach ML, jak i analizach Bayesowskich i był wysoko wsparty. Topologia drzewa cytoplazmatycznych JDP klasy A odzwierciedla drzewo gatunków - grzyby i zwierzęta tworzą kład. Wśród ludzkich JDP klasy A DNAJA2 jest najbardziej zewnętrzne, wyróżnicowało się po pierwszej w ludzkiej linii ewolucyjnej duplikacji genu kodującego JDP klasy A (Ryc. 3.12, strzałka z numerem 2), natomiast DNAJA1 oraz DNAJA4 powstały na skutek kolejnej duplikacji (Ryc. 3.12, strzałka z numerem 3).



**Rycina 3.12. Scenariusz ewolucji JDP należących do klas A i B zgodny z wynikami analizy filogenetycznej zaprezentowanej w tej rozprawie.** Kladogram reprezentuje zależności ewolucyjne pomiędzy bakteryjnymi i eukariotycznymi JDP klasy A i B. Linie niebieskie oznaczają JDP klasy A, natomiast linie zielone – JDP klasy B. Tło szare oznacza sekwencje pochodzące z bakterii, natomiast tła fioletowe – eukariotycznych pochodzących z mitochondriów (MT), plastydów (PL), retikulum endoplazmatycznego (ER) oraz cytoplazmy (C). Czerwone gwiazdki pokazują gałęzie drzewa, gdzie wyewoluowały JDP klasy B. Czarna kropka oznacza wspólnego przodka wszystkich JDP klas A i B bakteryjnych i eukariotycznych, biała kropka – wspólnego przodka JDP klas A i B eukariotycznych oraz DnaJ, różowa - wspólnego przodka mitochondrialnych JDP; zielona kropka – wspólnego przodka plastydowych JDP, żółta kropka -wspólnego przodka plastydowych JDP i DnaJ z bakteryjnego kladu Chloroflexi; fioletowa kropka – wspólnego przodka JDP z ER oraz cytoplazmy i DnaJ z Deltaproteobakterii, pomarańczowa kropka – wspólnego przodka JDP z cytoplazmy i ER, szara kropka – wspólnego przodka JDP z ER, brązowa kropka – wspólnego przodka JDP klasy B z ER, czerwona kropka – wspólnego przodka JDP cytoplazmatycznych klas A i B, jasnoniebieska kropka – wspólnego przodka cytoplazmatycznych JDP klasy A, jasnozielona kropka - wspólnego przodka cytoplazmatycznych JDP klasy B. Strzałkami oznaczone są węzły drzewa, gdzie doszło do duplikacji prowadzącej do powstania 1 – dwóch klas JDP w cytoplazmie, 2 – DNAJA2, 3 – DNAJA1 oraz DNAJA4, 4 – cytoplazmatycznych JDP klas B oraz B', 5 – DNAJB13, 6 – DNAJB5, 7 – DNAJB1 oraz DNAJB4.

W kladzie cytoplazmatycznych białek B białka B' stanowią odrębną grupę monofiletyczną. Oznacza to, że według scenariusza ewolucyjnego wspartego przez otrzymaną przeze mnie filogenezę JDP białka B' powstały na skutek duplikacji genu będącego wspólnym przodkiem wszystkich JDP klasy B (Ryc. 3.12, jasno zielona kropka, strzałka z numerem 4). W kladzie JDP klasy B' DNAJB2 jest zewnętrzne, a DNAJB6 oraz DNAJB8 są ze sobą najbliższym spokrewnione. Taka topologia w kladzie B' była wsparta mocniej niż topologie alternatywne. Ponieważ białka klasy B' występują tylko u twardzikowców pozycja ich wspólnego przodka z kanonicznymi białkami B jest inna niż spodziewana. Można było zakładać, że duplikacja genu, która doprowadziła do ich powstania będzie umiejscowiona w kladzie zwierzęcych białek B. Obserwowane umiejscowienie wspólnego przodka kanonicznych białek B i białek B' może być wynikiem tak zwanego "przyciągania się długich gałęzi" (ang. *long branch attraction*) [93]. Sekwencje różniące się znacząco od pozostałych sekwencji na drzewie, ale bardzo podobne względem siebie nawzajem, mogą pojawiać się na drzewie filogenetycznym poza kładem, do którego powinny należeć.

Wśród cytoplazmatycznych JDP należących do kanonicznej klasy B roślinne białka stanowią grupę zewnętrzną. Na węźle poniżej nich odgałęziają się białka grzybów – Sis1, natomiast białka ludzkie tworzą grupę monofiletyczną siostrzaną do Sis1. DNAJB13 jest najbardziej zewnętrzną wśród ludzkich białek B - wyewoluowało po pierwszej duplikacji wewnątrz tego kladu (Ryc. 3.12, strzałka 5), następnie na skutek kolejnej duplikacji pojawiło się DNAJB5 (Ryc. 3.12, strzałka 6), natomiast ostatnia duplikacja, która zaszła w tym kladzie dała początek (Ryc. 3.12, strzałka 7) DNAJB1 oraz DNAJB4, które w ostatecznej topologii drzewa stanowią grupę monofiletyczną. Wyniki te są zgodne z wcześniej publikowanymi danymi [75].

W kladzie białek z ER grupę zewnętrzną stanowią ortologi drożdżowego białka Scj1, natomiast ortologi roślinnego białka klasy A - ERDJ3 oraz ludzkiego białka klasy B -DNAJB11 są ze sobą najbliższym spokrewnione. Wynik ten nie jest zgodny z filogenezą gatunków- grzyby i zwierzęta powinny być siostrzane względem siebie. Różnicę tą można wyjaśnić zakładając, że u wspólnego przodka roślin, grzybów i zwierząt doszło do duplikacji genu kodującego JDP a później w liniach ewolucyjnych zwierząt, grzybów i roślin doszło do utraty jednej z kopii tak, że w efekcie kopie, które pozostały u roślin i zwierząt są ze sobą bliżej spokrewnione niż białko z grzybów. Co

ciekawe, DNAJB11 pomimo zaklasyfikowania do klasy B, posiada połowiczny rejon wiążący jony cynku występujący u JDP klasy A. Oznacza to, że na drodze ewolucji od przodka klasy A nie utraciło ono, w przeciwieństwie do pozostałych przedstawicieli klasy B, całego rejonu wiążącego jony cynku. Przykład ten wspiera scenariusz ewolucyjny, zgodny z topologią drzewa filogenetycznego białek JDP, według którego białka klasy B ewoluowały niezależnie od siebie z białek klasy A u bakterii oraz u eukariontów. U bakterii białko CbpA ewoluowało w wyniku duplikacji genu kodującego JDP klasy A. W cytoplazmie kanoniczne białka klasy B oraz pochodzące od nich białka klasy B' powstały w wyniku duplikacji i wtórnego różnicowania przodka będącego białkiem klasy A. Zaś w przypadku ER do powstania białka klasy B doszło tylko w linii ewolucyjnej zwierząt. Warto zaznaczyć, że w przypadku ER nie mamy do czynienia ze współistnieniem białek klasy A i B w tym samym przedziale komórkowym. U grzybów i roślin mamy JDP klasy A, zaś u zwierząt klasy B.

Warto również podkreślić, że eukariotyczne białka JDP nie są blisko spokrewnione z JDP archeonów. To nie jest zaskakujące w przypadku organelli komórkowych - mitochondria i plastydy pochodzą od bakteryjnych przodków. Jednak w przypadku cytoplazmy i ER, które zgodnie z endosymbiotyczną hipotezą powstania komórki eukariotycznej pochodzą od archeonów, ten brak pokrewieństwa może być zaskakujący. Należy jednak podkreślić, że moja analiza obejmowała stosunkowo niewiele sekwencji JDP pochodzących od archeonów. Analiza obejmująca większą liczbę sekwencji przedstawicieli archeonów jest niezbędna, aby mieć pewność, że JDP funkcjonujące w cytoplazmie i ER pochodzą od bakteryjnych przodków.

## 4. Ewolucja systemów JDP/Hsp70 wyspecjalizowanych w biogenezie centrów FeS

### 4.1 Wstęp

#### 4.1.1 Centra żelazo-siarkowe

Centra żelazo-siarkowe (FeS) to grupy prostetyczne kluczowe dla aktywności wielu białek niezbędnych dla prawidłowego funkcjonowania komórki, obecnych we wszystkich przedziałach komórkowych: cytoplazma/jądro, mitochondria, retikulum endoplazmatyczne. Białka zawierające centra FeS funkcjonują w procesach takich jak oddychanie tlenowe, fotosynteza, naprawa/replikacja DNA oraz asymilacja azotu. Wielobiałkowe maszynie zaangażowane w proces biogenezy FeS obecne są we wszystkich domenach życia. U eukariontów mitochondria, które odziedziczyły większość białek zaangażowanych w biogenezę FeS od swoich bakteryjnych przodków, odgrywają centralną rolę w tym procesie [94][95]. Zarówno bakteryjne, jak i mitochondrialne procesy biogenezy FeS składają się z dwóch głównych etapów: pierwszy to synteza centrum na białku rusztowaniu (IscU/Isu bakterie/mitochondria), a drugi to przeniesienie centrum z białka rusztowania na białka akceptorowe. W mitochondriach oraz u większości bakterii, wyspecjalizowane systemy białek opiekuńczych JDP/Hsp70 odgrywają kluczową rolę na etapie przeniesienia centrów FeS na białka akceptorowe [96][97][98][99].

#### 4.1.2 Systemy JDP/Hsp70 w biogenezie centrów FeS

Funkcja JDP/Hsp70 w biogenezie centrów FeS zależy, podobnie jak inne funkcje tych systemów opiekuńczych, od cyklicznego oddziaływania z substratem białkowym – w tym przypadku białkiem rusztowaniem IscU/Isu [100]. W biogenezie centrów FeS funkcjonuje wyspecjalizowane białko JDP (HscB/Hsc20 bakterie/mitochondria), które oddziałuje z białkiem rusztowaniem i dostarcza je do partnerskiego Hsp70 [101]. Oddziaływanie Hsp70-białko rusztowanie jest kluczowe dla przeniesienia centrów FeS na białko akceptorowe [102][103][104].

Funkcjonujące w mitochondriach JDP (Hsc20) oraz białko rusztowanie Isu są bezpośrednimi potomkami białek bakteryjnych - odpowiednio HscB i IscU. Tymczasem białka Hsp70 zaangażowane w biogenezę centrów FeS mają złożoną historię ewolucyjną [105]–[107]. W przypadku *Escherichia coli*, podobnie jak u innych

bakterii, w biogenezie FeS funkcjonuje wyspecjalizowane białko Hsp70 (HscA). Białko to wiąże tylko jeden substrat - białko rusztowanie IscU [108], [109]. Ponadto, w przeciwieństwie do innych Hsp70, w tym głównego bakteryjnego Hsp70 - DnaK, cykl wiązania IscU przez HscA nie wymaga udziału czynnika wymiany nukleotydów do wymiany ADP na ATP [99], [109].

Mitochondria większości organizmów, w tym człowieka, nie posiadają Hsp70 wyspecjalizowanego w biogenezie FeS [106]. Zamiast tego, wyspecjalizowane JDP (Hsc20) współpracuje z wielofunkcyjnym mitochondrialnym Hsp70 (mtHsp70), które bierze również udział w innych istotnych procesach, takich jak fałdowanie nowopowstających łańcuchów polipeptydowych, ponowne fałdowanie białek po narażeniu komórki na stres czy import białek przez błony mitochondrium [7]. mtHsp70 pełniąc te funkcje współpracuje z kilkoma różnymi JDP, w tym z Hsc20 w przypadku biogenezy centrów FeS. MtHsp70, podobnie jak bakteryjne DnaK, wymaga czynnika wymiany nukleotydu do wymiany ADP na ATP podczas cyklu wiązania i uwalniania Icu [110].

W przeciwieństwie do większości organizmów eukariotycznych, mitochondria *S. cerevisiae* i blisko spokrewnionych gatunków drożdży zawierają drugie, poza mtHsp70 białko Hsp70 - Ssq1. Ssq1 jest wyspecjalizowane w biogenezie centrów FeS i powstało w wyniku duplikacji genu kodującego mtHsp70, która miała miejsce u wspólnego przodka *S. cerevisiae* i *Candida albicans* [106]. Ssq1 współpracuje tylko z jednym JDP - Hsc20 i wiąże cyklicznie tylko jeden substrat - Icu. Podczas gdy paralog Ssq1, nazwany Ssc1, zachował wielofunkcyjność charakterystyczną dla mtHsp70 [42], [111]. Zarówno Ssq1, jak i Ssc1 wymagają czynnika wymiany nukleotydu do wymiany ADP na ATP [112], podobnie jak ich przodek - mtHsp70. Ze względu na obecność wyspecjalizowanego systemu białek opiekuńczych Hsc20/Ssq1 *S. cerevisiae* jest organizmem modelowym do badania molekularnych podstaw procesu biogenezy centrów FeS w mitochondriach [94], [96].

Specjalizacja białek Hsp70 funkcjonujących w procesie biogenezy centrów FeS u bakterii (HscA) oraz u drożdży (Ssq1) jest przykładem konwergentnej ewolucji na poziomie molekularnym. Pomimo tego, że białka te są dobrze scharakteryzowane biochemicznie niewiele wiadomo na temat ich powiązań ewolucyjnych. W szczególności czy białko HscA zostało utracone w czasie ewolucji mitochondriów czy



też jego homologi obecne są w komórkach eukariotycznych oraz czy Ssq1 jest rzeczywiście produktem duplikacji genu kodującego mtHsp70 [105], [106], Wykorzystałam analizy filogenetyczne, aby odpowiedzieć na te pytania.

## 4.2 Metody

### 4.2.1 Rekonstrukcja drzewa filogenetycznego Hsp70 zaangażowanych w biogenezę centrów FeS metodą ML

228 sekwencji Hsp70 ortologicznych do mtHsp70, DnaK oraz HscA pobrałam z bazy danych OMA (Zestawienie I). Następnie zestawiałam je przy użyciu Clustal Omega v1.2.2 z wykorzystaniem domyślnych parametrów [80]. Zestawienie przekształciłam w ukryty model Markowa przy użyciu programu *hmmbuild* z pakietu HMMER [113]. Zestawienie sekwencji tą metodą daje wartości prawdopodobieństwa *a posteriori* (ang. *posterior probability* - pp) dla każdej pozycji zestawienia, dla każdej z sekwencji. Usunęłam z zestawienia wszystkie pozycje, których wsparcie pp było niższe niż 0.7 [114] i tak przygotowane zestawienie wykorzystałam do rekonstrukcji filogenezy Hsp70 metodą maksymalnej wiarygodności (ML). Filogenezę ML obliczyłam przy użyciu programu IQ-Tree [115] z 1000 powtórzeń *bootstrap*, stosując model substytucji aminokwasów LG i model heterogeniczności GAMMA z czterema odrębnymi kategoriami szybkości ewolucji (LG + I + G4), który był najlepiej dopasowany do danych na podstawie kryterium Akaike.

### 4.2.2 Dystrybucja filogenetyczna bakteryjnych białek HscA oraz HscB

Z bazy danych UniProt pobrałam proteomy 339 gatunków zwierząt, 661 gatunków grzybów, 113 gatunków roślin i 38 gatunków pierwotniaków (protistita) [116]. Z bazy danych OMA pobrałam sekwencje aminokwasowe białek ortologicznych do HscA oraz HscB i zestawiałam za pomocą Clustal Omega v1.2.2 przy użyciu domyślnych parametrów. Zestawienie dla każdego z białek przekształciłam w profil Markowa przy użyciu programu *hmmbuild*. Oba modele Markowa wytrenowałam i przetestowałam: pierwszy w rozpoznawaniu HscA, ale nie innych Hsp70, przy użyciu Zestawienia I jako zestawu danych referencyjnych, jak opisano w [117], drugi w rozpoznawaniu HscB, ale nie innych JDP, używając 135 bakteryjnych białek JDP jako zestawu danych referencyjnych. Żeby poznać dystrybucję filogenetyczną białek HscA i HscB w zebranych proteomach, przeszukałam każdy z nich na obecność/brak zarówno HscA, jak i HscB korzystając z narzędzia *hmmsearch* z pakietu HMMER. Sekwencje

rozpoznane jako potencjalne ortologi HscA zestawiałam z referencyjnymi sekwencjami Hsp70 oraz umiejscowiłam je na referencyjnym drzewie filogenetycznym przy użyciu algorytmu ulokowania opartego na funkcji wiarygodności (ang. *likelihood-based placement algorithm*) z programu RaxML [118]. Potencjalne ortologi HscA zweryfikowałam dodatkowo za pomocą zwrrotnych przeszukiwań BLAST do proteomu *E. coli*.

## 4.3 Wyniki

### 4.3.1 Analiza filogenetyczna Hsp70 zaangażowanych w proces biogenezy centrów FeS

Pomimo wielu lat badań, zależności ewolucyjne pomiędzy systemami HscA/HscB i Ssq1/Hsc20 nie zostały całkowicie wyjaśnione [105], [106]. Nie jest także w pełni jasne, czy HscA występuje jedynie u bakterii, czy też jego ortologi są obecne także w komórkach eukariotycznych. Aby dokładniej przeanalizować ewolucyjne związki między HscA i Ssq1, a także określić pochodzenie mtHsp70 pobrałam z bazy danych OMA 228 sekwencji DnaK, HscA, mtHsp70, Ssc1 i Ssq1 oraz ich ortologów. W oparciu o zestawienie sekwencji zrekonstruowałam drzewo filogenetyczne metodą maksymalnej wiarygodności z zastosowaniem modelu ewolucji LG+I+G4, który zgodnie z kryterium Akaike był najlepiej dopasowany do danych w zestawieniu.

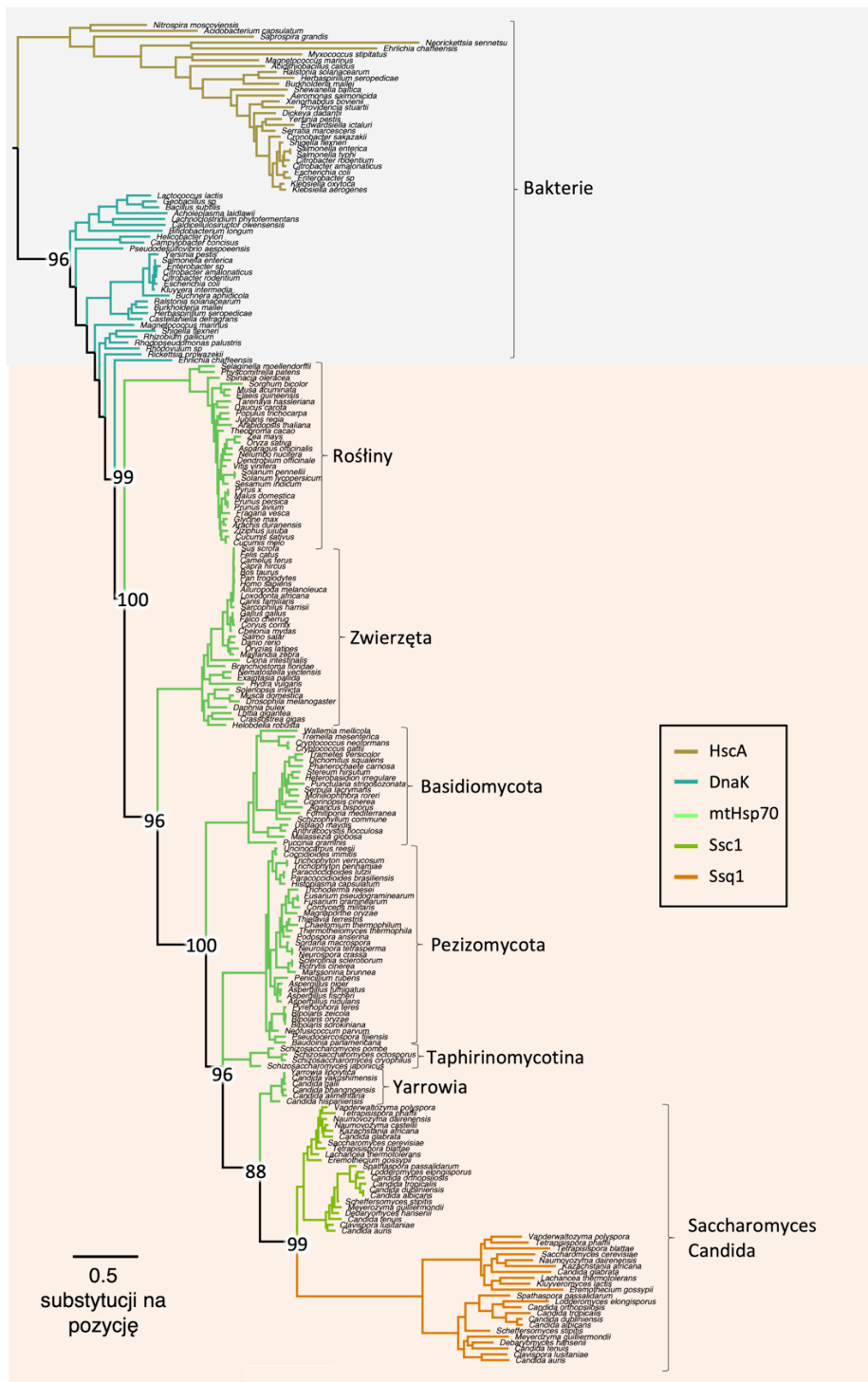
Uzyskane drzewo ML ukorzeniłam na węźle reprezentującym wspólnego przodka wszystkich białek bakteryjnych (Ryc. 4.1). Topologia drzewa wskazuje, że wszystkie mitochondrialne Hsp70 są bliżej spokrewnione z bakteryjnym DnaK niż z HscA, ponieważ eukariotyczne mtHsp70 wraz z paralogami Ssq1 oraz Ssc1 tworzą na drzewie grupę monofiletyczną z sekwencjami DnaK, a węzeł reprezentujący ich wspólnego przodka ma wysokie wsparcie statystyczne (*bootstrap* 96). Sekwencje HscA stanowią grupę zewnętrzną otrzymanego drzewa.

Układ kładów grupujących mitochondrialne Hsp70 jest zgodny z drzewem życia. Pośród sekwencji białek eukariotycznych roślinne stanowią silnie wspartą grupę zewnętrzną (*bootstrap* 100), natomiast klady zawierające sekwencje grzybów oraz zwierząt stanowią grupy siostrzane. Węzeł reprezentujący wspólnego przodka białek zwierząt i grzybów również jest dobrze wsparty (*bootstrap* 96). Wśród sekwencji

grzybów *Basidiomycota* stanowią grupę najbardziej zewnętrzną (*bootstrap* 100), następny węzeł drzewa prowadzi do kladu tworzonego przez sekwencje *Pezizomycota* wraz z *Taphirinomycotina* (*bootstrap* 96), a sekwencje mtHsp70 grzybów rodzaju *Yarrowia* tworzą grupę monofiletyczną z kladem obejmującym sekwencje z *Saccharomyces* + *Candida* (*bootstrap* 88).

Kład białek eukariotycznych zawiera zarówno wielofunkcyjne białka mtHsp70, jak i wyspecjalizowane Ssq1. Sekwencje Ssq1 wraz z Ssc1 tworzą na drzewie grupę monofiletyczną (*bootstrap* 99), co jest zgodne z wcześniej opublikowanymi danymi [105], [106] wskazującymi, że Ssq1 powstało na skutek duplikacji genu kodującego mtHsp70, która miała miejsce u wspólnego przodka *Saccharomyces* i *Candida*. Gałąź drzewa prowadząca do kladu zawierającego sekwencje Ssq1 jest wyraźnie dłuższa od pozostałych, co oznacza, że białko to po duplikacji ewoluowało szybciej od paralogu Ssc1.

Na podstawie drzewa maksymalnej wiarygodności, można stwierdzić, że pomimo iż HscA oraz Ssq1 są wyspecjalizowane do pełnienia tej samej funkcji nie są ze sobą blisko spokrewnione. Ssq1 nie jest potomkiem HscA - powstało niezależnie od HscA na skutek duplikacji genu kodującego mtHsp70, która miała miejsce u wspólnego przodka drożdży z kladów *Saccharomyces* i *Candida*. Funkcjonalna specjalizacja Ssq1 jest więc przykładem ewolucji konwergentnej na poziomie biochemicznym.



**Rycina 4.1. Filogeneza Hsp70 zaangażowanych w biogenezę centrów FeS.** Filogeneza ML oparta na 228 sekwencjach aminokwasowych Hsp70. Drzewo ukorzenione na wspólnym przodku białek bakteryjnych HscA i DnaK. Wsparcia bootstrap są pokazane dla głównych rozgałęzień drzewa. Rozwiązane są tylko węzły, dla których wartość *bootstrap* była wyższa niż 70. Na szarym tle znajdują się sekwencje bakteryjne, na tle pomarańczowym – eukariotyczne.

### 4.3.2 Dystrybucja filogenetyczna HscA oraz HscB

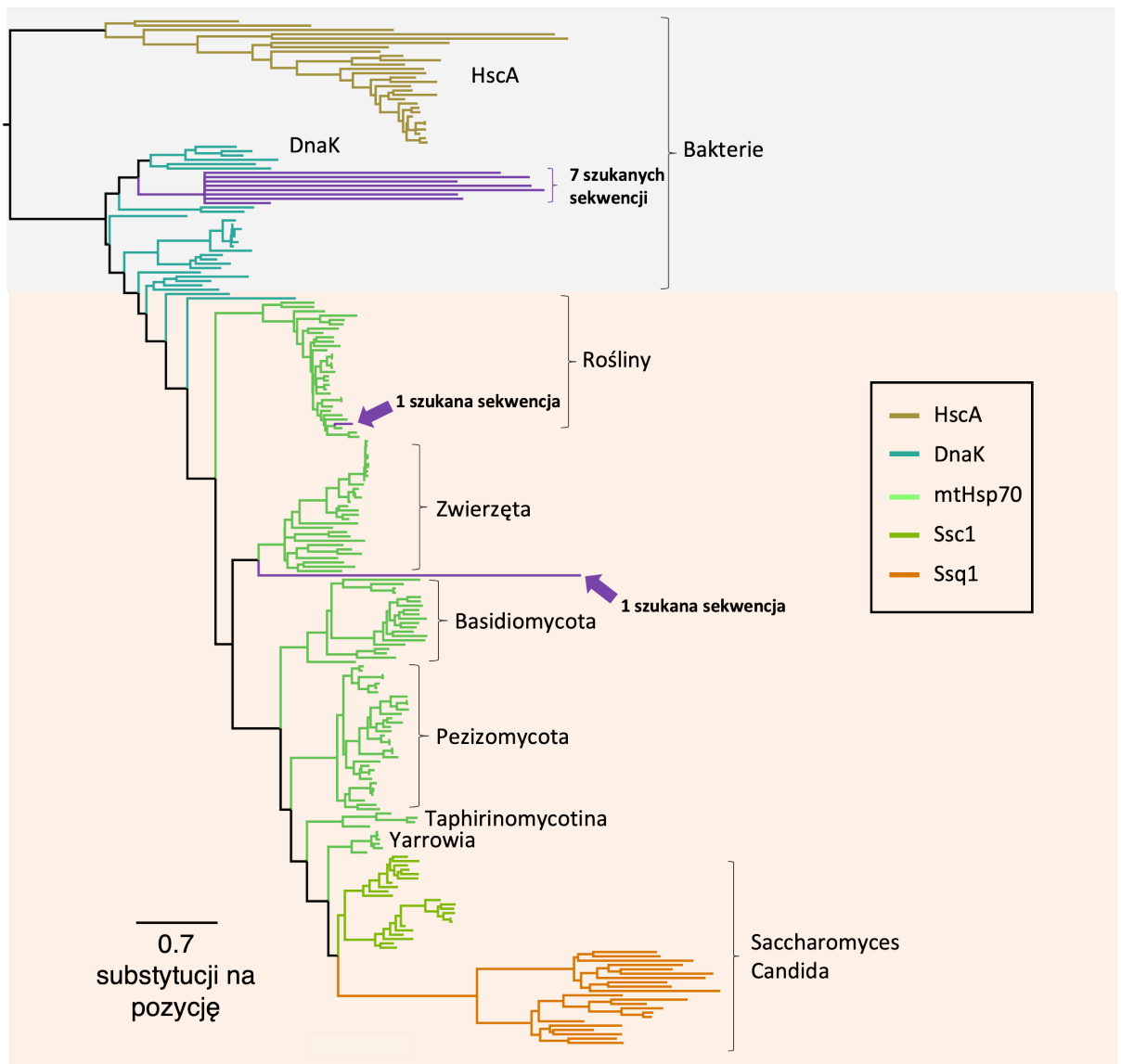
W celu analizy dystrybucji filogenetycznej ortologów HscA oraz HscB u eukariontów zgromadziłam proteomy 339 zwierząt, 661 grzybów, 113 roślin i 38 pierwotniaków (protista), które następnie przeszukałam używając profili HMM przygotowanych dla każdego z białek w oparciu o zestawienia sekwencji ortologicznych (patrz Metody).

**Tabela 2. Identyfikacja HscA i HscB w proteomach eukariotycznych.**

Przeszukane proteomy	Szukane białko	
	HscA	HscB
Grzyby	0/661	661/661
Zwierzęta	2/339	339/339
Rośliny	7/113	113/113
Protisty	0/38	38/38

Wyniki przeszukiwania (Tabela 2) wykazały obecność ortologów HscB we wszystkich 1151 zbadanych proteomach organizmów eukariotycznych, co jest zgodne z jego pochodzeniem od bakteryjnych przodków oraz kluczową rolą ortologów HscB w procesie mitochondrialnej biogenezy FeS. W przeciwieństwie do HscB, potencjalne ortologi HscA zidentyfikowałam jedynie w 9 spośród przeszukiwanych proteomów. Aby sprawdzić czy zidentyfikowane sekwencje faktycznie są ortologami HscA wykonałam dwa dodatkowe testy.

Pierwszym z nich było dołączenie sekwencji potencjalnych orthologów HscA do drzewa filogenetycznego prezentowanego na Ryc 4.1 używając algorytmu ulokowania opartego na funkcji wiarygodności (Ryc. 4.2) [118]. Algorytm ten pozostawia topologię referencyjnego drzewa bez zmian, a jedynie dołącza do niej zadane sekwencje w oparciu o zestawienie. Żadna z sekwencji dołączonych do drzewa nie grupowała się z kładem HscA; 7 dołączyło do kładu DnaK, 1 do kładu mtHsp70 pochodzących z roślin oraz 1 do mtHps70 zwierząt. Wynik ten sugeruje, że profil HMM błędnie zidentyfikował te sekwencje jako ortologi HscA.



**Rycina 4.2. Potencjalne ortologi HscA dołączone do drzewa ML Hsp70.** Topologia drzewa Hsp70 jest dokładnym odwzorowaniem topologii drzewa przedstawionego na Ryc. 4.1 z dołączonymi potencjalnymi ortologami HscA. Na szarym tle znajdują się sekwencje bakteryjne, na tle pomarańczowym – eukariotyczne. Potencjalne ortologi HscA zaznaczone są kolorem fioletowym oraz strzałkami.

Drugim testem był zwrotny BLAST do proteomu *E. coli*. Zwrotny BLAST polega na znalezieniu sekwencji najbardziej podobnej do zadanej w proteomie organizmu referencyjnego. Otrzymane wyniki pokazały, że w każdym przypadku zwrotny BLAST zidentyfikował sekwencje DnaK jako najbardziej podobną do potencjalnego ortloga HscA. Dla wszystkich przeszukań BLAST wartość *e-value* będące miarą istotności statystycznej wyniku jest bliska lub równa 0, co oznacza, że każde z przeszukań jest istotne statystycznie. Otrzymane wyniki pokazują, że każda z sekwencji zidentyfikowanych jako potencjalny eukariotyczny ortolog HscA jest

bardziej podobna do bakteryjnego DnaK niż do HscA (Tabela 3). Wyniki przeszukiwania proteomów, a także oba sposoby weryfikacji potencjalnych ortologów HscA: zarówno algorytm ulokowania oparty na funkcji wiarygodności, jak i zwrotny BLAST wykazały, że analizowane w tej pracy proteomy eukariotyczne nie posiadają ortologa HscA.

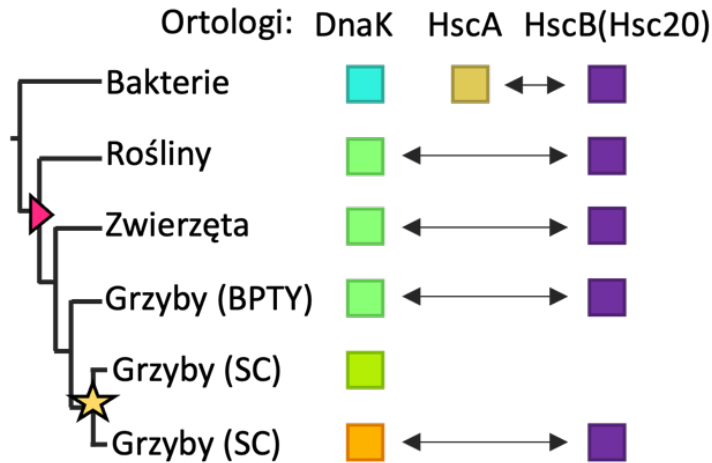
**Tabela 3. Wyniki zwrotnego BLAST do proteomu *E. coli* dla potencjalnych ortologów HscA.**

Numer identyfikacyjny*	Najlepsze trafienie BLAST	Wynik	E-value	Procent identyczności
A0A2P6SKD8	Molecular chaperone DnaK	1309	0.0	62.62%
A0A2P6VE39	Molecular chaperone DnaK	1105	0.0	73.08%
A0A443NWH4	Molecular chaperone DnaK	1091	0.0	61.08%
A0A175YM19	Molecular chaperone DnaK	1257	0.0	63.60%
A0A388M9J0	Molecular chaperone DnaK	736	5e-121	41.37%
A0A4P1QS49	Molecular chaperone DnaK	1377	0.0	94%
A0A251VA77	Molecular chaperone DnaK	1624	0.0	67.32%
A0A402FMW4	Molecular chaperone DnaK	1087	0.0	81.71%
A0A074ZTS7	Molecular chaperone DnaK	1008	0.0	72.31%

\* Numery identyfikacyjne w bazie danych Uniprot dla sekwencji potencjalnych ortologów HscA

#### 4.4 Dyskusja

Historia ewolucyjna białek Hsp70 zaangażowanych w proces biogenezy FeS jest skomplikowana. Wyniki moich analiz wspierają scenariusz, zgodnie z którym bakteryjne HscA wyspecjalizowane w biogenezie FeS zostało utracone u wspólnego przodka eukariontów (Ryc. 4.3, różowy trójkąt) w trakcie ewolucji mitochondriów. Jego rolę przejęło mtHsp70, potomek bakteryjnego wielofunkcyjnego Hsp70 DnaK, współpracującego z wieloma JDP, w tym z Hsc20 w procesie biogenezy centrów FeS. Ponadto filogeneza ML potwierdza, że wyspecjalizowane Ssq1, współpracujące jedynie z Hsc20, nie zostało odziedziczone od bakteryjnych przodków mitochondriów. Zamiast tego, białko to powstało w u podstawy u wspólnego przodka siostrzanych kładów *Saccharomyces* oraz *Candida* na drodze duplikacji genu kodującego wielofunkcyjne mtHsp70 (Ryc. 4.3, gwiazdka).



**Rycina 4.3. Historia ewolucyjna systemów JDP/Hsp70 zaangażowanych w biogenezę centrów FeS.** Kladogram reprezentuje zależności ewolucyjne Hsp70 zaangażowanych w biogenezę FeS. Kolorowe kwadraty pokazują obecność lub brak ortologów danego białka (DnaK, HscA, HscB(Hsc20)) w każdej z grup taksonomicznych. BPTY – kład obejmujący grzyby Basidiomycota, Pezizomycotina, Taphiromycotina, Yarrowia. S.C. – kład obejmujący grzyby Saccharomycotina + Candida. Duplikacja genu kodującego mtHsp70 w kladzie SC oznaczona jest gwiazdką, utrata wyspecjalizowanego w biogenezie FeS Hsp70 - różowy trójkąt.

Ponadto Ssq1 ze względu na specjalizację funkcjonalną ewoluowało szybciej niż siostrzane do niego Ssc1, co na drzewie ML widoczne jest w postaci długiej gałęzi prowadzącej do kładu Ssq1. Ten scenariusz ewolucyjny został zaproponowany już wcześniej [105], [111], [119], ale analizy oparte były na ograniczonych zbiorach danych dostępnych w tamtym czasie, a nie na dużym zbiorze analizowanym w tej rozprawie. W oparciu o uzyskane wyniki mogę stwierdzić, że mitochondrialne Ssq1 i bakteryjne HscA nie są blisko spokrewnione, a zatem ich specjalizacja polegająca na współpracy z jednym białkiem JDP wyspecjalizowanym w biogenezie FeS oraz ich specyficzność substratowa względem białka rusztowania Isu/IscU stanowią przykład ewolucji konwergentnej na poziomie biochemicznym [120]–[122].

Historia ewolucji JDP zaangażowanych w biogenezę centrów FeS jest mniej skomplikowana. Mitochondria odziedziczyły białko JDP od swoich bakteryjnych przodków, co sugeruje, że homologiczny proces biogenezy centrów FeS był obecny u bakteryjnych endosymbiontów, które dały początek mitochondriom w komórkach eukariotycznych. Białko JDP w mitochondriach odgrywa podobną rolę jak jego



bakteryjny odpowiednik – dostarcza substrat białkowy (Isu) do partnerskiego Hsp70 podczas biogenezy centrów FeS.

## 5. Ewolucja miejsca oddziaływania domeny-J JDP z partnerskim Hsp70

### 5.1 Wstęp

Sercem mechanizmu kontroli allosterycznej stojącej za zmianami konformacyjnymi Hsp70 jest sieć oddziaływań pomiędzy domeną NBD, SBD $\beta$  i łącznikiem, która ma miejsce, gdy Hsp70 jest w konformacji związanej z ATP-konformacja otwarta [123]–[128]. Łącznik odgrywa ważną rolę, ponieważ jego oddziaływanie z NBD jest niezbędne do stymulacji aktywności ATPazy Hsp70 [25], [129]–[131]. W obecności substratu peptydowego zaobserwowano przejściowy, allosterycznie aktywny, stan pośredni Hsp70-ATP w kompleksie z peptydem gotowy do hydrolizy ATP [125], [132].

Jednak pełna stymulacja ATPazy Hsp70 wymaga oddziaływania z domeną-J partnerskiego JDP [29]. Wczesne prace nad oddziaływaniem JDP DnaJ z Hsp70 DnaK z *Escherichia coli* pokazały, że arginina R167 znajdująca się w domenie NBD DnaK, jest niezbędna dla wydajnej stymulacji aktywności ATPazy Hsp70 przez domenę-J DnaJ [129], [133], [134]. Struktura krystalograficzna kompleksu domena-J DnaJ z DnaK [135] pokazuje, że domena-J wiąże się do DnaK w miejscu utworzonym przez oddziałujące ze sobą domeny NBD, SBD $\beta$  oraz łącznik. Reszty aminokwasowe helisy II i III domeny-J oraz łącząca je pętla zawierająca motyw HPD tworzą polarne i hydrofobowe oddziaływania z NBD/SBD $\beta$ /łącznikiem. W oddziaływaniu domeny-J DnaJ/DnaK bezpośrednio uczestniczy 18 reszt aminokwasowych znajdujących się na domenie J oraz 38 reszt DnaK [37]. Podobna analiza została przeprowadzona przez nasz zespół dla mitochondrialnego systemu JDP/Hsp70 zaangażowanego w proces biogenezy centrów żelazo-siarkowych (FeS). W tym przypadku nie dysponowaliśmy strukturą krystalograficzną. Strukturę kompleksu Hsc20/Ssq1 modelowaliśmy stosując dokowanie molekularne i symulacje dynamiki molekularnej [66]. W otrzymanym kompleksie Hsc20/Ssq1-ATP domena-J oddziaływała z Ssq1 w bardzo podobny sposób jak miało to miejsce w przypadku domena-J DnaJ/DnaK-ATP [66][37]. Zidentyfikowaliśmy 17 reszt domeny-J Hsc20 kontaktujących się z 18 resztami Ssq1. Przy czym największy wkład do energii kompleksu wykazywało 8 naładowanych reszt: 3 dodatnio naładowane reszty Hsc20 zlokalizowane na helisie II (R37, K38, R41), tworzące sieć oddziaływań elektrostatycznych z 4 ujemnie naładowanymi resztami

Ssq1 znajdującymi się na NBD (D246, E248, D249, E253) oraz jedna reszta na helisie III domeny-J (K70). Pomimo dużego podobieństwa strukturalnego kompleksów Hsc20/Ssq1 i DnaJ/DnaK oraz bardzo podobnego mechanizmu oddziaływania domeny-J DnaJ i Hsc20 z partnerskimi Hsp70 reszty zaangażowane w te oddziaływania nie były takie same. Sugeruje to, że pomimo zachowanego mechanizmu oddziaływania domena-J/Hsp70 specyficzność tego oddziaływania może się zmieniać poprzez zaangażowanie różnych reszt aminokwasowych. Aby zweryfikować tą hipotezę analizowałam konserwację miejsca oddziaływania domena-J/Hsp70 dla dwóch systemów: ortologów DnaJ/DnaK z bakterii oraz mitochondriów jak również ortologów Hsc20/mtHsp70 z mitochondriów.

## 5.2 Metody

### 5.2.1 Rekonstrukcja filogenezy Hsp70

Aby zrekonstruować filogenezę białek Hsp70 (ortologi mtHsp70, Ssc1, Ssq1 oraz DnaK) zestawiałam 289 sekwencji aminokwasowych pochodzących z bakterii, roślin, zwierząt oraz grzybów przy użyciu Clustal Omega v1.2.2 z wykorzystaniem domyślnych parametrów (Zestawienie Hsp70).

Zestawienie przekształciłam w ukryty model Markowa [113] przy użyciu programu hmmbuild z pakietu HMMER [79]. Z zestawienia usunęłam wszystkie kolumny, których wsparcie prawdopodobieństwa *a posteriori* pp było niższe niż 0.7 i tak przygotowane zestawienie wykorzystałam do rekonstrukcji filogenezy Hsp70 metodą Maksymalnej wiarygodności.

Filogenezę ML obliczyłam przy użyciu programu RAxML v8.2.10 [118] z 1000 szybkich powtórzeń *bootstrap* (ang. *ultrafast bootstrap*) stosując model substytucji aminokwasów LG i tempo podstawień aminokwasowych o rozkładzie GAMMA z czterema odrębnymi kategoriami szybkości ewolucji (LG + I + G) [136], który na podstawie kryterium Akaike w programie ProtTest v3.2 [137] był najlepiej dopasowany do danych. Otrzymaną topologię drzewa filogenetycznego Hsp70 wykorzystywałam we wszystkich późniejszych obliczeniach.

### 5.2.2 Analiza koewolucji reszt aminokwasowych zaangażowanych w oddziaływanie domena-J/Hsp70

W celu analizy koewolucji reszt aminokwasowych zaangażowanych w oddziaływanie domena-J/Hsp70 wykorzystałam program Coev oparty o funkcję maksymalnej wiarygodności [138]. Obliczenia wykonałam osobno dla dwóch systemów: Hsc20/mtHsp70 oraz DnaJ/DnaK w każdym przypadku do obliczeń wykorzystałam topologię drzewa filogenetycznego Hsp70. Dla pary Hsc20/mtHsp70 wykorzystałam sekwencje mtHsp70 oraz topologię fragmentu drzewa grupującego mtHsp70 z *Saccharomycotina*, *Pezizomycotina* oraz *Taphrinomycotina*, natomiast dla pary DnaJ/DnaK wykorzystałam sekwencje DnaK oraz topologię fragmentu drzewa grupującego DnaK z *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria*, *Acidobacteria*, *Epsilonproteobacteria*, *Chloroflexi*, *Actinobacteria*, grupa FCB, grupa PVC, *Negativicutes*, *Tenericutes*, *Bacilli* oraz *Clostridia*. Dla każdej sekwencji mtHsp70 oraz DnaK dobrałam sekwencje odpowiednio Hsc20 lub DnaJ pochodzące z tych samych gatunków.

Przygotowałam następujące zestawienia sekwencji: Hsc20 (Zestawienie I) i mtHsp70 (Zestawienie II) oraz DnaJ (Zestawienie III) i DnaK (Zestawienie IV), które zestawiałam używając Clustal Omega z zastosowaniem domyślnych parametrów. Następnie w każdym zestawieniu pozostawiłam jedynie pozycje, które w oparciu o analizę struktury kompleksów domena-J DnaJ/DnaK z *E. coli* oraz Hsc20/Ssq1 z *S. cerevisiae* zidentyfikowano jako zaangażowane w oddziaływanie domena-J/Hsp70.

Dla każdej pary zestawień zastosowałam dwa modele: model zakładający koewolucję (Coev) oraz model ewolucji niezależnej (M0). Wynikiem tej analizy jest wartość AIC (kryterium informacyjne Akaike oparte na maksimum funkcji wiarygodności), która obliczona jest dla wszystkich możliwych kombinacji par aminokwasów oddziałujących par białek. Dopasowanie modelu Coev w porównaniu z modelem M0 przedstawiłam za pomocą różnicy kryterium informacyjnego  $\Delta$ Akaike ( $\Delta$ AIC =  $AIC_{M0} - AIC_{Coev}$ ). Aby określić, które pary reszt aminokwasowych koewoluują niezbędne było określenie progu odcięcia wartości  $\Delta$ AIC- przekroczenie wartości progu wskazuje na koewolucję pary reszt aminokwasowych [138]. Próg odcięcia wyznaczyłam symulując 1000 zestawień sekwencji ewoluujących niezależnie,

opierając się na topologii tych samych fragmentów drzewa filogenetycznego, które zastosowałam do analizy par DnaJ/DnaK oraz Hsc20/Ssq1. Do symulacji wykorzystałam oprogramowanie *evolver* z pakietu PAML 4 [139]. Dla symulowanych danych zastosowałam program *Coev* i dwa modele ewolucji: *Coev* oraz *M0*. Dla każdej pary reszt aminokwasowych, dla każdego zestawienia obliczyłam  $\Delta AIC$ , a następnie przygotowałam rozkłady otrzymanych wartości dla systemu DnaJ/DnaK oraz dla Hsc20/Ssq1. Wyznaczyłam 99 percentyl każdego z rozkładów  $\Delta AIC$ , co pozwoliło mi wyznaczyć wartość progu odcięcia obserwowanego  $\Delta AIC$ , od którego uznawałam daną parę aminokwasów za koewoluującą z poziomem ufności  $p < 0.01$ . Oznacza to, że w przypadku kiedy sekwencje aminokwasowe ewoluują niezależnie jedynie dla 1% z nich przekroczona zostaje dana wartość  $\Delta AIC$ , przyjęta jako próg odcięcia. Co z kolei oznacza, że jeżeli analizowane przeze mnie pary reszt dają wartości  $\Delta AIC$  równą lub wyższą od progu odcięcia to prawdopodobieństwo, że są pod wpływem koewolucji wynosi 99%.

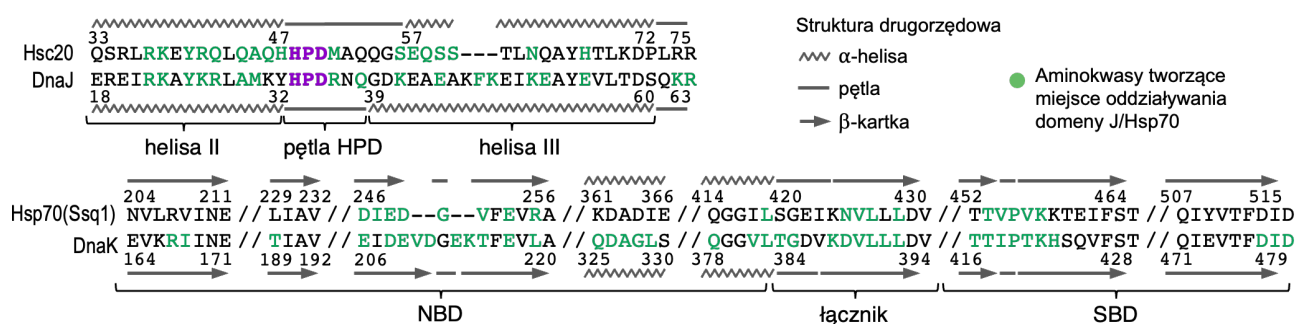
### 5.2.3 Konserwacja ewolucyjna aminokwasów zaangażowanych w oddziaływanie domena-J/Hsp70

Aby określić konserwację ewolucyjną reszt zaangażowanych w oddziaływanie domena-J/Hsp70 zidentyfikowałam reszty homologiczne do reszt zaangażowanych w oddziaływanie domena-J DnaJ z DnaK na podstawie (1) Zestawienia bakteryjnych i mitochondrialnych ortologów DnaK oraz DnaJ (2) Zestawienia bakteryjnych ortologów DnaK oraz DnaJ, natomiast dla reszt zaangażowanych w oddziaływanie domena-J Hsc20/Ssq1 na podstawie (1) Zestawienia eukariotycznych ortologów mtHsp70 oraz Hsc20 i (2) dla ortologów mtHsp70 oraz Hsc20 z gromady Ascomycota. W każdym przypadku zestawienie obejmowało jedynie pozycje homologiczne do zaangażowanych w oddziaływanie domena-J/Hsp70. Następnie, przygotowałam logo obrazujące konserwację tych pozycji w parach domena-J DnaJ/DnaK oraz domena-J Hsc20/Ssq1. Loga wygenerowałam przy użyciu serwera WebLogo [140].

## 5.3 Wyniki

### 5.3.1 Ewolucja miejsca oddziaływania domena-J /Hsp70

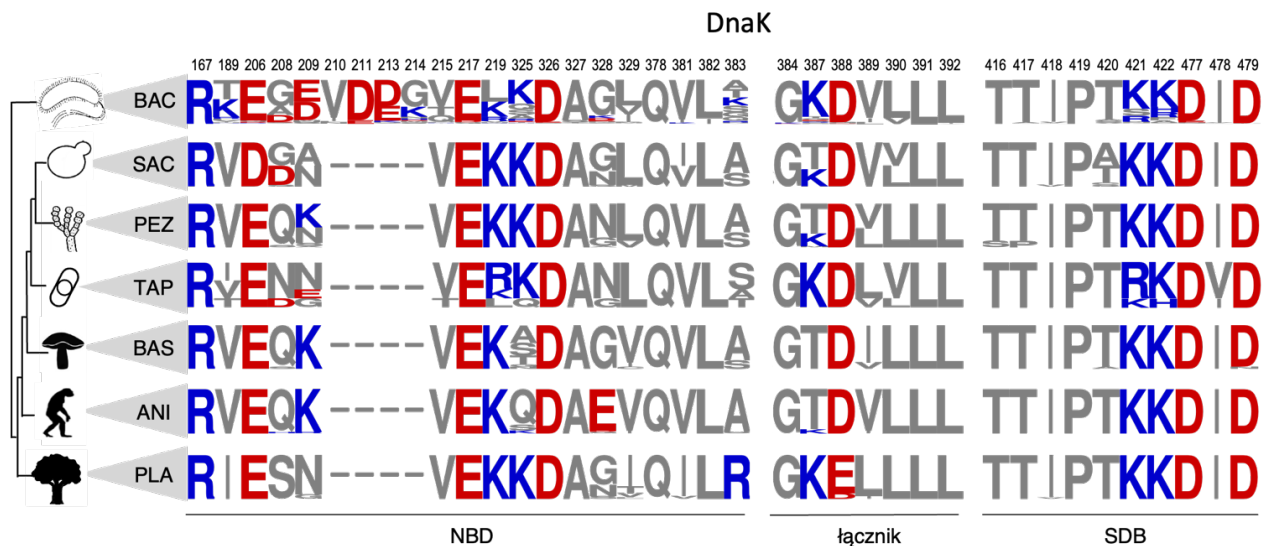
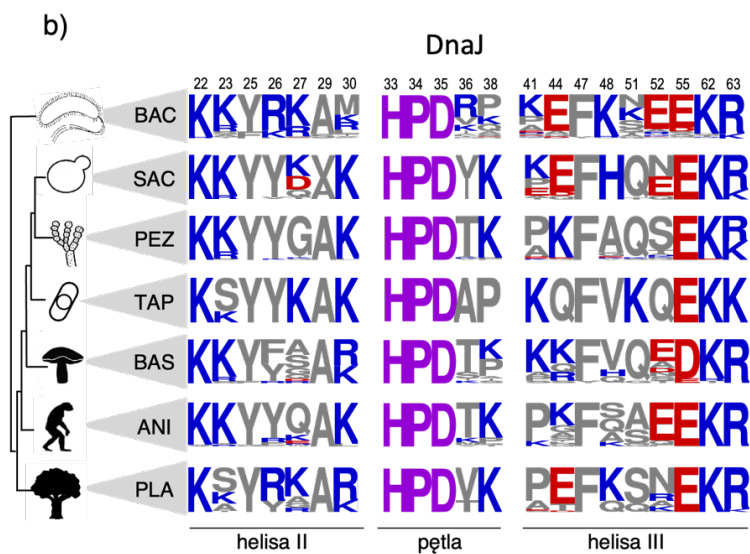
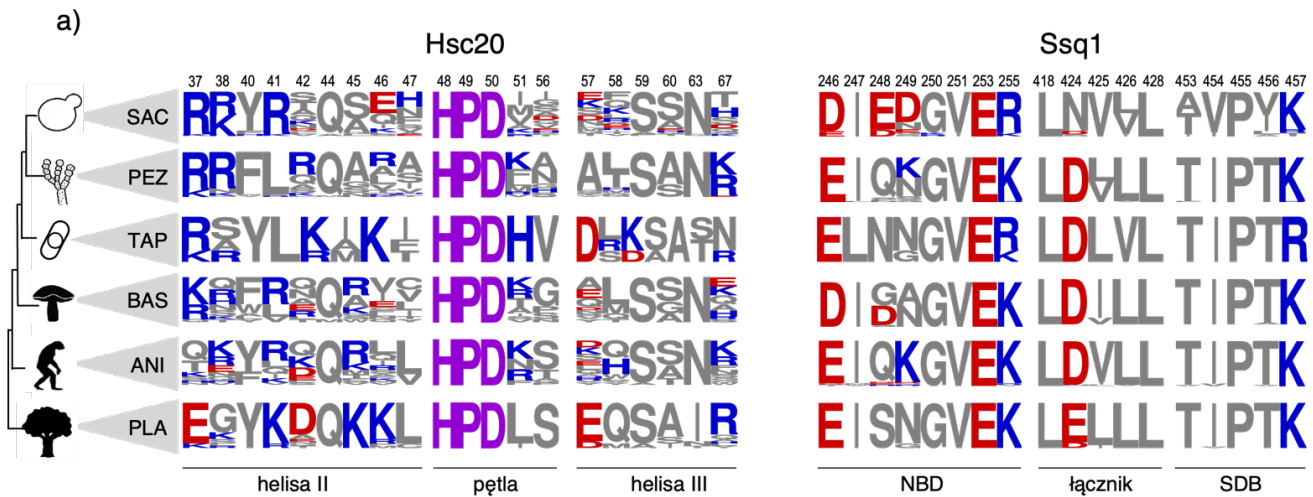
Pomimo strukturalnego podobieństwa miejsca oddziaływania domeny-J/Hsp70 dla dwóch systemów: DnaJ/DnaK i Hsc20/Ssq1 okazało się, że reszty aminokwasowe zaangażowane w to oddziaływanie znacząco się różnią. W przypadku domeny-J Hsc20 — tylko 12 z 17 pozycji tworzących kontakty z Ssq1 jest wspólnych z pozycjami tworzącymi kontakty pomiędzy domeną-J DnaJ i DnaK. Ponadto tylko w 3 z tych 12 pozycji występują w obu przypadkach te same aminokwasy. Po stronie Hsp70 spośród 38 pozycji DnaK oddziałujących z domeną-J DnaJ tylko 16 jest wspólnych dla pary domena-J Hsc20/Ssq1, z czego tylko w połowie występują te same aminokwasy (Ryc.5.1).



**Rycina 5.1. Porównanie miejsc oddziaływania domena-J/Hsp70 systemów Hsc20/Ssq1 oraz DnaJ/DnaK.** Zestawienie fragmentów sekwencji domeny-J Hsc20 i DnaJ (górny panel) oraz fragmentów sekwencji zaangażowanych w oddziaływanie z domeną-J Ssq1 i DnaK (dolny panel). Reszty aminokwasowe każdego białka, które są zaangażowane w tworzenie miejsca wiązania Hsc20/Ssq1 i DnaJ/DnaK, są zaznaczone na zielono. Motywy HPD zaznaczony jest na fioletowo.

Ponieważ reszty aminokwasowe zaangażowane w oddziaływanie domena-J/Hsp70 dla dwóch analizowanych systemów okazały się znacząco różne postanowiłam sprawdzić, jak wygląda ich konserwacja w szerszej grupie gatunków.

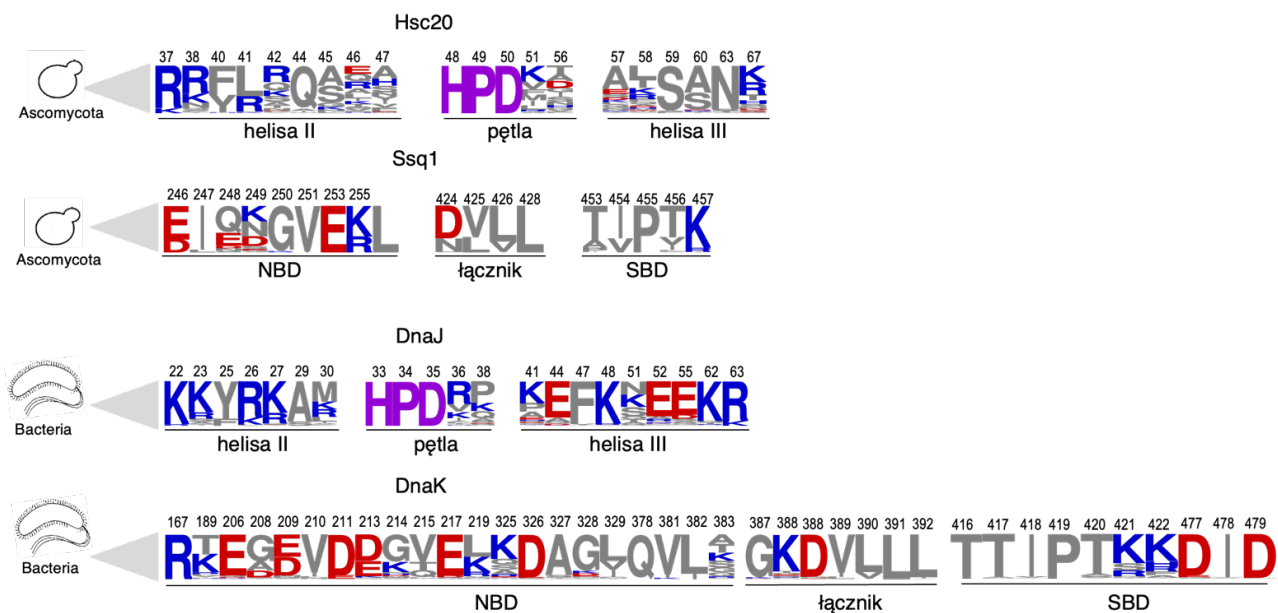
Przeanalizowałam zmienność reszt aminokwasowych homologicznych do reszt uczestniczących w oddziaływaniu domena-J/Hsp70 dla systemów Hsc20/Ssq1 (Ryc. 5.2a) oraz DnaJ/DnaK (Ryc. 5.2b). Dla systemu Hsc20/Ssq1 analizowałam ortologi mtHsp70 pochodzące z roślin, zwierząt i grzybów. W przypadku pary DnaJ/DnaK analizowałam ortologii tych białek z bakterii oraz mitochondriów. Konserwację przedstawiłam w formie logo, gdzie wielkość liter odzwierciedla częstość występowania aminokwasu w danej pozycji zestawienia.



**Rycina 5.2. Zmienność sekwencji pozycji, które tworzą miejsca oddziaływania domena-J/Hsp70.** Zmienność pokazana jest dla pozycji zaznaczonych na zielono na Ryc. 5.2 dla a) mitochondrialnych ortologów Hsc20 (panel lewy) i Ssq1 (panel prawy) oraz b) bakteryjnych i mitochondrialnych ortologów DnaJ (panel górny) oraz DnaK (panel dolny) z następujących grup taksonomicznych: Saccharomycotina (SAC), Pezizomycotina (PEZ), Taphiromycotina (TAP), Basidiomycota (BAS), Zwierzęta (ANI) i Rośliny (PLA) oraz Bakterie (BAC) dla b). Reszty aminokwasowe pokolorowane są zależnie od posiadani/braku ładunku: naładowane dodatnio (niebieski), naładowane ujemnie (czerwony) i nienaładowane (szary). Motyw HPD (fioletowy). Numeracja pozycji dla a) *S. cerevisiae* b) *E. coli*. Relacje filogenetyczne między grupami taksonomicznymi przedstawia kladogram po lewej stronie każdego panelu.

Analiza pokazała, że reszty aminokwasowe odpowiedzialne za interakcje domena-J/Hsp70 w przypadku obu analizowanych systemów (Hsc20/Ssq1 oraz DnaJ/DnaK) nie tylko różnią się znacząco między sobą, ale także wykazują dużą zmienność wśród przedstawicieli innych grup taksonomicznych. Szczególnie dotyczy to domeny-J. Żadna z reszt zaangażowanych w oddziaływanie z Hsp70 od strony domeny-J Hsc20 nie jest w pełni zakonserwowana w pozostałych grupach taksonomicznych, natomiast w przypadku domeny-J DnaJ jedynie 4 reszty spośród 18 są zakonserwowane. Przyjrzałam się więc jak konserwacja tych reszt wygląda wśród gatunków blisko spokrewnionych z *S. cerevisiae* oraz *E. coli*. Zmienność sekwencji jest ewidentna dla domen J zarówno w przypadku Hsc20 (Ryc. 5.3, górny panel), jak też DnaJ (Ryc. 5.3, dolny panel) nawet wśród gatunków blisko spokrewnionych z *S. cerevisiae* i *E. coli* – odpowiednio Ascomycota (67 sekwencji, Ryc.5.5, lewy panel) i bakterii (117 sekwencji, Ryc. 5.5, prawy panel), co sugeruje, że nawet blisko spokrewnione systemy JDP/Hsp70 wykorzystują różne reszty aminokwasowe w oddziaływaniu domena-J/Hsp70.





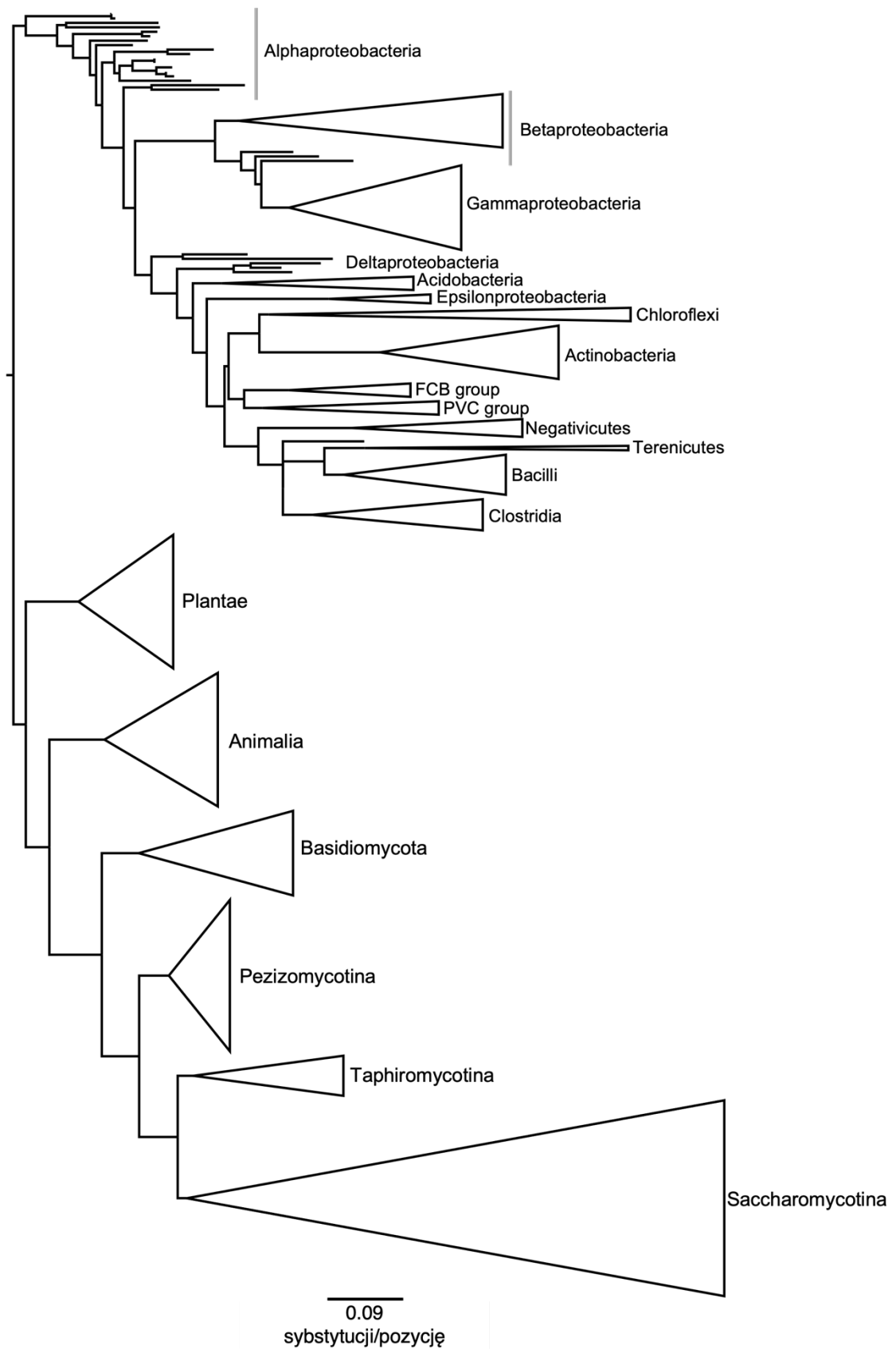
**Rycina 5.3. Zmienność sekwencji zaangażowanych w oddziaływanie domena-J/Hsp70 u gatunków blisko spokrewnionych z *S. cerevisiae* i *E. coli*.** Zmienność pokazana jest dla pozycji zaznaczonych na zielono na Ryc. 5.2 dla ortologów Hsc20 (panel górny – Hsc20) i ortologów Ssq1 (panel górny – Ssq1) oraz ortologów DnaJ (panel dolny – DnaJ) oraz DnaK (panel dolny – DnaK) z Ascomycota dla kompleksu Hsc20/Ssq1 oraz bakterii dla DnaJ/DnaK. Reszty aminokwasowe pokolorowane są zależnie od posiadani/braku ładunku: naładowane dodatnio (niebieski), naładowane ujemnie (czerwony) i nienaładowane (szary). Motyw HPD (fioletowy). Numeracja pozycji dla a) *S. cerevisiae* b) *E. coli*.

### 5.3.2 Koewolucja oddziaływania domena-J/Hsp70

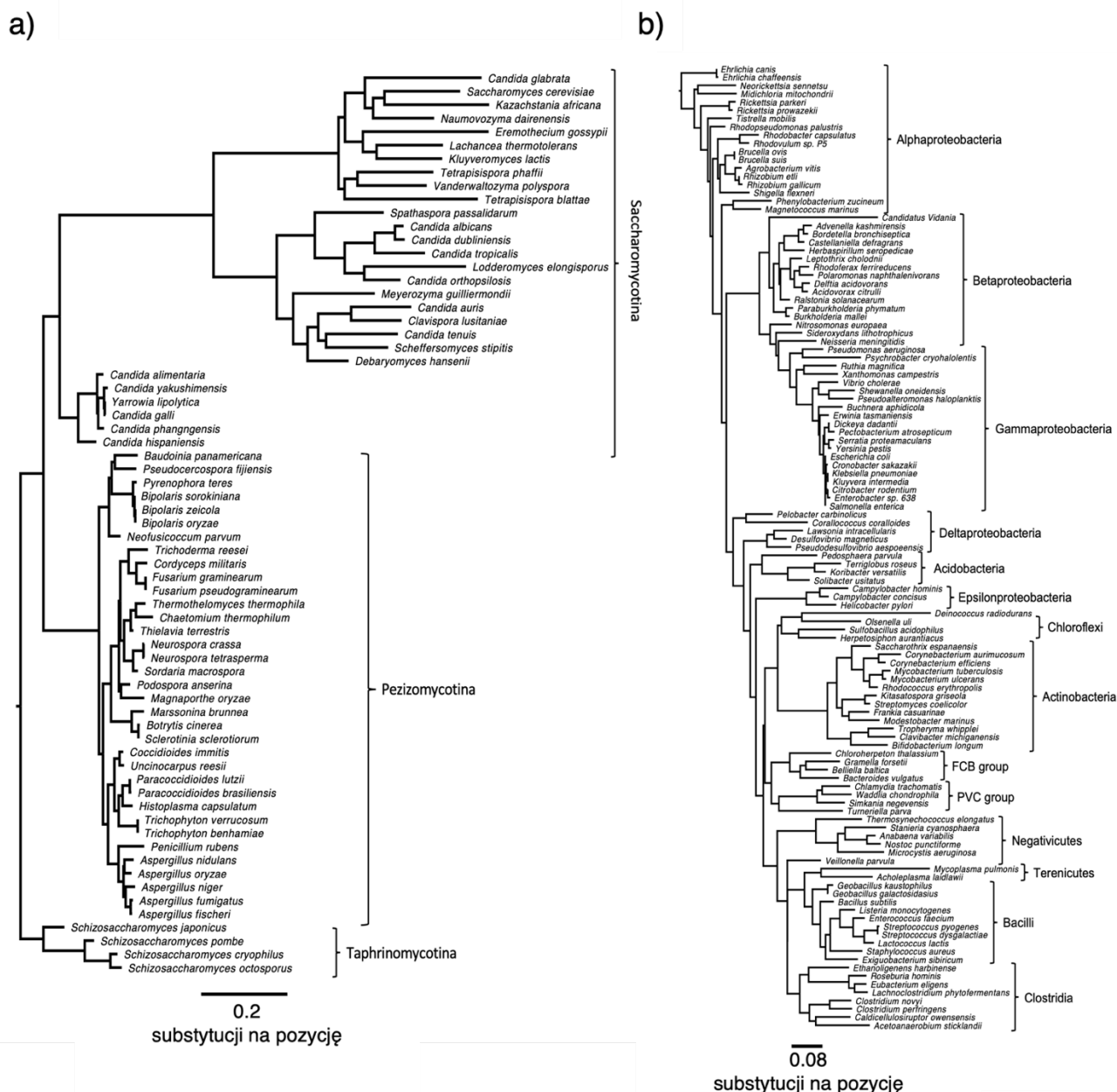
Obserwacja, że reszty aminokwasowe zaangażowane w oddziaływanie domena-J/Hsp70 w systemach Hsc20/Ssq1 oraz DnaJ/DnaK różnią się od siebie, a także ich zmienność u blisko spokrewnionych gatunków skłoniła mnie do zadania pytania jak to możliwe, że poszczególne pary JDP/Hsp70 są w stanie się rozpoznawać i tworzyć funkcjonalne kompleksy. Postawiłam hipotezę, że koewolucja (mechanizm ewolucyjny, gdzie zmiana aminokwasu w sekwencji jednego z oddziałujących białek powodują kompensującą zmianę w sekwencji drugiego białka [141]) może być odpowiedzią na to pytanie.

Aby sprawdzić czy reszty zaangażowane w oddziaływanie domena-J/Hsp70 faktycznie koewoluują zrekonstruowałam drzewo filogenetyczne Hsp70 – DnaK oraz mtHsp70 metodą maksymalnej wiarygodności stosując model ewolucji LG + I + G, który został określony jako najlepiej dopasowany do danych. Topologia drzewa jest

zgodna z zależnościami filogenetycznymi między grupami taksonomicznymi w drzewie życia [142] (Ryc. 5.4). Do analiz wybrałam fragmenty drzewa reprezentujące sekwencje pochodzące z 67 gatunków grzybów i 117 gatunków bakterii blisko spokrewnionych z *S. cerevisiae* (Ryc. 5.5, lewy panel) oraz z *E. coli* (Ryc. 5.5, prawy panel) Dla każdej z sekwencji Hsp70 wyszukałam sekwencję oddziałującego z nim ortologa Hsc20 lub DnaJ. Po zebraniu danych przygotowałam cztery zestawienia sekwencji: Ssq1, Hsc20, DnaJ oraz DnaK i w każdym z nich pozostawiłam jedynie pozycje (kolumny) homologiczne do reszt zaangażowanych w oddziaływanie domena-*J/Hsp70* odpowiednio dla systemów Hsc20/Ssq1 lub DnaJ/DnaK (Ryc. 5.1, zielone reszty aminokwasowe).

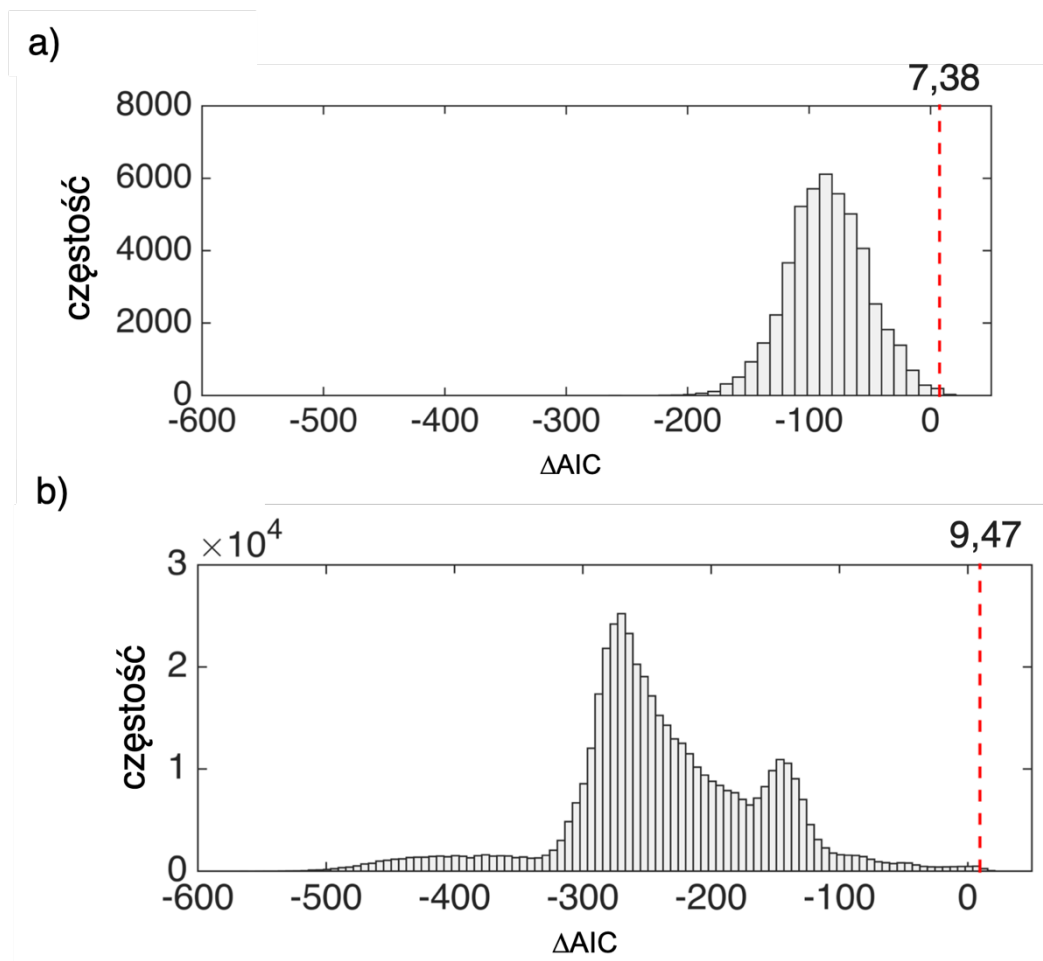


**Rycina 5.4. Filogeneza DnaK i mtHsp70.** Drzewo filogenetyczne zrekonstruowane metodą ML dla 299 sekwencji bakteryjnych i eukariotycznych ortologów DnaK i mtHsp70. Klady podpisane są nazwami grup taksonomicznych, które reprezentują.



**Rycina 5.5. Fragmenty drzewa filogenetycznego Hsp70 wykorzystane do analizy koewolucji domena-J/Hsp70. a) fragment drzewa Hsp70 zawierający 67 sekwencji Ssq1 i mtHsp70 pochodzących z Ascomycota, wykorzystanych do analizy konserwacji miejsca oddziaływania oraz koewolucji Hsc20/Ssq1 b) fragment drzewa zawierający 117 sekwencji DnaK z bakterii wykorzystanych do analizy konserwacji miejsca oddziaływania oraz koewolucji DnaJ/DnaK.**

Następnie dla każdej pary białek przetestowałam dwa modele ewolucji sekwencji: model Coev zakładający koewolucję par reszt aminokwasowych należących do tych sekwencji oraz model M0 zakładający niezależną ewolucję sekwencji. Na podstawie różnicy wartości kryterium Akaike dla modelu M0 oraz modelu Coev ( $\Delta AIC = AIC_{M0} - AIC_{Coev}$ ) oceniłam dla każdej pary sekwencji, który model jest lepiej dopasowany do danych. Pozwoliło mi to określić, które z par reszt aminokwasowych koewoluowały. Próg odcięcia wartości  $\Delta AIC$  świadczący o koewolucji danej pary wyznaczyłam na podstawie symulacji ewolucji niezależnej sekwencji (patrz Metody). Dla pary Hsc20/Ssq1 próg ten wyniósł 7,38 (Ryc. 5.6a), natomiast dla pary DnaJ/DnaK 9,47 (Ryc. 5.6b).



**Rycina 5.6. Rozkłady  $\Delta AIC$  obliczone dla danych symulowanych w niezależnym modelu ewolucji.** a) dla każdej pary pozycji pochodzących z symulacji sekwencji aminokwasowej dla fragmentu filogenezy Hsp70 (Ryc. 5.5, lewy panel). Próg odcięcia dla poziomu ufności  $p < 0,01 = 7,38$  (czerwona przerywana linia) b) dla każdej pary pozycji pochodzących z sekwencji aminokwasowej fragmentu filogenezy Hsp70 (Ryc 5.5, prawy panel). Próg odcięcia dla poziomu ufności  $p < 0,01 = 9,47$  (czerwona przerywana linia).

Zidentyfikowałam 16 par koewoluujących reszt aminokwasowych dla systemu Hsc20/Ssq1 (Tabela 4, Ryc. 5.7a, górny panel) oraz 20 par koewoluujących reszt aminokwasowych dla systemu DnaJ/DnaK (Tabela 5, Ryc. 5.7a, dolny panel). Otrzymane wyniki są zgodne z hipotezą, że koewolucja reszt zaangażowanych w oddziaływanie domena-J/Hsp70 może wyjaśniać, w jaki sposób funkcjonalne oddziaływania domeny-J/Hsp70 są zachowane pomimo zmienności sekwencji zarówno domen J, jak i Hsp70.

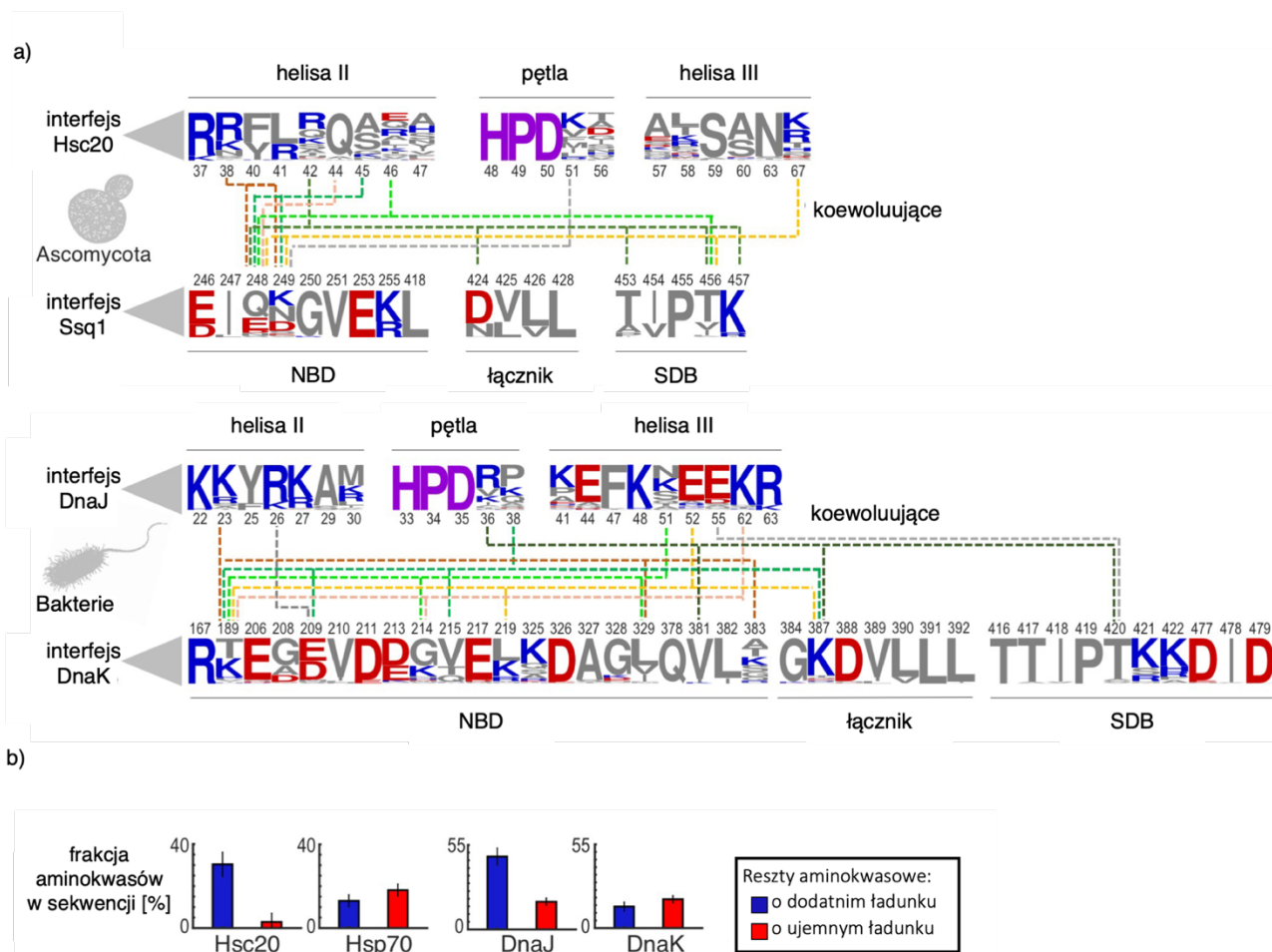
Pomimo zmienności sekwencji zaangażowanych z oddziaływanie domena-J/Hsp70 widoczny jest trend, zgodnie z którym sekwencje po stronie domeny-J bogate są w aminokwasy naładowane dodatnio podczas gdy po stronie Hsp70 miejsce wiązania zwiera aminokwasy o ładunku ujemnym (Ryc. 5.7b). To sugeruje, że komplementarne ładunki mają znaczenie w oddziaływaniu obu białek. Warto wspomnieć, że oddziaływania elektrostatyczne mają większy zasięg niż oddziaływania van der Waalsa oraz oddziaływania o charakterze hydrofobowym.

**Tabela 4. Kryterium Akaike ( $\Delta AIC$ ) dla koewoluujących par reszt tworzących miejsce wiązania Hsc20/Ssq1.**

Pozycje koewoluujące Hsc20-Hsp70	AIC Coev	AIC M0	$\Delta AIC$
38-248	240.33	270.97	30.64
67-249	318.42	346.36	27.94
45-249	329.69	351.30	21.61
42-456	347.12	367.27	20.15
42-248	336.41	354.48	18.07
51-249	351.36	368.23	16.88
38-249	335.58	352.18	16.60
67-456	261.98	277.66	15.67
46-248	342.87	357.31	14.44
42-457	333.25	346.90	13.65
46-456	356.64	370.10	13.45
45-248	257.69	269.81	12.12
42-424	332.37	343.90	11.52
42-453	326.50	337.62	11.12
44-248	138.17	148.26	10.09
67-248	255.09	264.87	9.78

**Tabela 5. Kryterium Akaike ( $\Delta AIC$ ) dla koewoluujących par reszt tworzących miejsce wiązania DnaJ/DnaK.**

Pozycje koewoluujące DnaJ-DnaK	AIC Coev	AIC M0	$\Delta AIC$
51-329	567.33	596.53	29.19
38-209	456.05	483.87	27.83
26-209	230.09	250.86	20.77
23-383	658.30	677.39	19.09
38-189	587.70	605.57	17.87
23-189	534.22	551.87	17.65
36-387	455.19	472.41	17.21
36-381	384.13	400.77	16.64
62-189	385.02	401.40	16.38
38-215	435.32	451.36	16.04
38-387	485.28	500.13	14.85
51-189	583.08	597.93	14.85
52-219	415.49	429.86	14.37
55-420	363.08	377.27	14.18
23-329	536.32	550.47	14.15
51-214	507.30	520.54	13.25
36-420	428.11	439.55	11.45
52-387	334.05	345.00	10.95
52-189	440.03	450.44	10.41
62-214	314.56	324.01	9.45



**Rycina 5.7. Koewolucja miejsca oddziaływania domena-J/Hsp70.** a) Logo przedstawione jak na Ryc. 5.4. Pary reszt będące pod wpływem koewolucji dla oddziaływania domena-J Hsc20/Ssq1 (górny panel) oraz domena-J DnaJ/DnK (dolny panel) ze wsparciem statystycznym  $p < 0,01$  połączone są przerywanymi liniami. Numeracja pozycji jest taka sama jak dla sekwencji *S. cerevisiae* (górny panel) i *E. coli* (dolny panel). (b) Frakcja reszt aminokwasowych domeny-J Hsc20 i mtHsp70 (po lewej) oraz domeny-J DnaJ oraz DnaK (po prawej), które są naładowane dodatnio (niebieski) i ujemnie (czerwony). Wąsy oznaczają odchylenie standardowe.

## 5.4 Dyskusja

Na podstawie otrzymanych wyników można wnioskować, że w oddziaływanie domeny-J JDP z Hsp70 zaangażowane są reszty aminokwasowe charakteryzujące się znaczną zmiennością, które nie są zakonserwowane nawet wśród gatunków blisko spokrewnionych. Na pierwszy rzut oka zmienność sekwencji zaangażowanych w oddziaływanie kluczowe dla funkcjonowania systemów JDP/Hsp70 może wydawać się zaskakująca. Pojawia się pytanie, w jaki sposób ta różnorodność ewoluowała bez utraty funkcji. Wyniki przeprowadzonych przeze mnie analiz sugerują, że było to możliwe



zarówno dzięki skomplikowanej sieci oddziaływań pomiędzy resztami aminokwasowymi domeny-J i Hsp70 jak też koewolucji sekwencji zaangażowanych w to oddziaływanie. W przypadku analizowanych systemów (Hsc20/Ssq1 oraz DnaJ/DnaK) kluczowe reszty aminokwasowe po jednej stronie oddziaływania kontaktują się z więcej niż jedną resztą po stronie przeciwnej. Taka wielopłaszczyznowa sieć oddziaływań jest bardzo elastyczna. Z upływem czasu reszty aminokwasowe mogą być zamieniane, dodawane lub usuwane z miejsca oddziaływania bez utraty zdolności do utrzymania funkcjonalnego oddziaływania pomiędzy koewoluującymi partnerami, umożliwiając zmiany zarówno specyficzności, jak i siły ich oddziaływania. Koewolucja miejsca oddziaływania może również wyjaśnić, w jaki sposób JDP może zmienić partnerskie Hsp70, tak jak miało to miejsce w mitochondriach drożdży – kiedy Hsc20 zmieniło partnera z wielofunkcyjnego mtHsp70 na wyspecjalizowane Ssq1 [106]. W pierwszej kolejności mutacje w Hsc20 mogły umożliwić oddziaływanie z Ssq1 bez utraty zdolności oddziaływania z mtHsp70, następnie koewolucja Hsc20/Ssq1 mogła wzmocnić nowe oddziaływanie jednocześnie osłabiając oddziaływanie z mtHsp70 [119].

Ewolucyjna zmienność miejsca oddziaływania domena-J/Hsp70 może mieć również wpływ na funkcjonowanie systemów JDP/Hsp70 w szerszym kontekście, umożliwiając powstanie skomplikowanych sieci oddziaływań pomiędzy partnerskimi białkami. W sytuacji, gdy pojedyncze Hsp70 oddziałuje z wieloma JDP koewolucja pozwalałaby dostosować siłę oddziaływania między konkretnymi parami JDP/Hsp70 do potrzeb komórki. Koewolucja może również wyjaśnić, dlaczego niektóre JDP mają większe powinowactwo do partnerskich Hsp70 niż inne. W tym przypadku znaczenie może mieć ładunek dodatni aminokwasów zaangażowanych w oddziaływanie od strony domeny-J. Domeny-J o silniejszym ładunku będą oddziaływać z partnerskim Hsp70 bardziej efektywnie. W sytuacji odwrotnej, gdy w jednym przedziale komórkowym funkcjonuje więcej niż jedno Hsp70 koewolucja mogłaby sprawić, że każde z nich oddziałuje tylko z niektórymi z dostępnych JDP. Przykładem mogą być cytoplazmatyczne Hsp70 Ssa oraz Ssb drożdży *S. cerevisiae*, które pomimo pochodzenia od wspólnego przodka oddziałują z różnymi JDP [143], [144].

## 6. Bibliografia

- [1] B. Bukau and A. L. Horwich, 'The Hsp70 and Hsp60 Review Chaperone Machines', 1998.
- [2] D. Balchin, M. Hayer-Hartl, and F. U. Hartl, 'In vivo aspects of protein folding and quality control', *Science*, vol. 353, no. 6294. American Association for the Advancement of Science, Jul. 01, 2016. doi: 10.1126/science.aac4354.
- [3] M. P. Mayer and B. Bukau, 'Hsp70 chaperones: Cellular functions and molecular mechanism', *Cellular and Molecular Life Sciences*, vol. 62, no. 6. pp. 670–684, Mar. 2005. doi: 10.1007/s00018-004-4464-6.
- [4] Q. Liu, C. Liang, and L. Zhou, 'Structural and functional analysis of the Hsp70/Hsp40 chaperone system', *Protein Science*, vol. 29, no. 2. Blackwell Publishing Ltd, pp. 378–390, Feb. 01, 2020. doi: 10.1002/pro.3725.
- [5] Q. Liu and E. A. Craig, 'Molecular biology: Mature proteins braced by a chaperone', *Nature*, vol. 539, no. 7629. Nature Publishing Group, pp. 361–362, 2016. doi: 10.1038/nature20470.
- [6] S. Preissler and E. Deuerling, 'Ribosome-associated chaperones as key players in proteostasis', *Trends in Biochemical Sciences*, vol. 37, no. 7. pp. 274–283, Jul. 2012. doi: 10.1016/j.tibs.2012.03.002.
- [7] W. Neupert and M. Brunner, 'The protein import motor of mitochondria', *Nature Reviews Molecular Cell Biology*, vol. 3, no. 8. pp. 555–565, 2002. doi: 10.1038/nrm878.
- [8] N. B. Nillegoda and B. Bukau, 'Metazoan Hsp70-based protein disaggregases: Emergence and mechanisms', *Frontiers in Molecular Biosciences*, vol. 2, no. OCT. Frontiers Media S.A., Oct. 09, 2015. doi: 10.3389/fmolb.2015.00057.
- [9] R. Sousa and E. M. Lafer, 'The role of molecular chaperones in clathrin mediated vesicular trafficking', *Frontiers in Molecular Biosciences*, vol. 2, no. MAY. Frontiers Media S.A., May 19, 2015. doi: 10.3389/fmolb.2015.00026.
- [10] J. Gillis *et al.*, 'The DNAJB6 and DNAJB8 protein chaperones prevent intracellular aggregation of polyglutamine peptides', *Journal of Biological Chemistry*, vol. 288, no. 24, pp. 17225–17237, Jun. 2013, doi: 10.1074/jbc.M112.421685.
- [11] D. Balchin, M. Hayer-Hartl, and F. U. Hartl, 'In vivo aspects of protein folding and quality control', *Science*, vol. 353, no. 6294. American Association for the Advancement of Science, Jul. 01, 2016. doi: 10.1126/science.aac4354.
- [12] A. Mogk, B. Bukau, and H. H. Kampinga, 'Cellular Handling of Protein Aggregates by Disaggregation Machines', *Molecular Cell*, vol. 69, no. 2. Cell Press, pp. 214–226, Jan. 18, 2018. doi: 10.1016/j.molcel.2018.01.004.
- [13] B. Qu, Y. Jia, Y. Liu, H. Wang, G. Ren, and H. Wang, 'The detection and role of heat shock protein 70 in various nondisease conditions and disease conditions: a literature review', *Cell Stress and Chaperones*, vol. 20, no. 6. Cell Stress and Chaperones, pp. 885–892, Nov. 01, 2015. doi: 10.1007/s12192-015-0618-8.

- [14] D. M. Walther *et al.*, ‘Widespread proteome remodeling and aggregation in aging *C. elegans*’, *Cell*, vol. 161, no. 4, pp. 919–932, May 2015, doi: 10.1016/j.cell.2015.03.032.
- [15] R. Rosenzweig, N. B. Nillegoda, M. P. Mayer, and B. Bukau, ‘The Hsp70 chaperone network’, *Nature Reviews Molecular Cell Biology*, vol. 20, no. 11. Nature Publishing Group, pp. 665–680, Nov. 01, 2019. doi: 10.1038/s41580-019-0133-3.
- [16] K. Peisker, M. Chiabudini, and S. Rospert, ‘The ribosome-bound Hsp70 homolog Ssb of *Saccharomyces cerevisiae*’, *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1803, no. 6. pp. 662–672, Jun. 2010. doi: 10.1016/j.bbamcr.2010.03.005.
- [17] P. Huang, M. Gautschi, W. Walter, S. Rospert, and E. A. Craig, ‘The Hsp70 Ssz1 modulates the function of the ribosome-associated J-protein Zuol’, *Nat Struct Mol Biol*, vol. 12, no. 6, pp. 497–504, Jun. 2005, doi: 10.1038/nsmb942.
- [18] H. H. Kampinga and E. A. Craig, ‘The HSP70 chaperone machinery: J proteins as drivers of functional specificity’, *Nature Reviews Molecular Cell Biology*, vol. 11, no. 8. pp. 579–592, Aug. 2010. doi: 10.1038/nrm2941.
- [19] J. L. Abrams, J. Verghese, P. A. Gibney, and K. A. Morano, ‘Hierarchical functional specificity of cytosolic heat shock protein 70 (Hsp70) nucleotide exchange factors in yeast’, *Journal of Biological Chemistry*, vol. 289, no. 19, pp. 13155–13167, 2014, doi: 10.1074/jbc.M113.530014.
- [20] A. Bracher and J. Verghese, ‘The nucleotide exchange factors of Hsp70 molecular chaperones’, *Frontiers in Molecular Biosciences*, vol. 2, no. APR. Frontiers Media S.A., Apr. 07, 2015. doi: 10.3389/fmolb.2015.00010.
- [21] M. P. Mayer and L. M. Gierasch, ‘Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones’, *Journal of Biological Chemistry*, vol. 294, no. 6. American Society for Biochemistry and Molecular Biology Inc., pp. 2085–2097, Feb. 08, 2019. doi: 10.1074/jbc.REV118.002810.
- [22] ‘Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein’.
- [23] E. B. Bertelsen, H. Zhou, D. F. Lowry, G. C. Flynn, and F. W. Dahlquist, ‘Topology and dynamics of the 10 kDa C-terminal domain of DnaK in solution’, 1999.
- [24] E. R. P. Zuiderweg, L. E. Hightower, and J. E. Gestwicki, ‘The remarkable multivalency of the Hsp70 chaperones’, *Cell Stress and Chaperones*, vol. 22, no. 2. Cell Stress and Chaperones, pp. 173–189, Mar. 01, 2017. doi: 10.1007/s12192-017-0776-y.
- [25] C. A. English, W. Sherman, W. Meng, and L. M. Gierasch, ‘The Hsp70 interdomain linker is a dynamic switch that enables allosteric communication between two structured domains’, *Journal of Biological Chemistry*, vol. 292, no. 36, pp. 14765–14774, Sep. 2017, doi: 10.1074/jbc.M117.789313.
- [26] R. Zhang, D. Malinverni, D. M. Cyr, P. D. L. Rios, and N. B. Nillegoda, ‘J-domain protein chaperone circuits in proteostasis and disease’, *Trends in Cell*

- Biology*, vol. 33, no. 1. Elsevier Ltd, pp. 30–47, Jan. 01, 2023. doi: 10.1016/j.tcb.2022.05.004.
- [27] D. Walls, M. Zyliczll, and C. Georgopoulosss, ‘THE JOURNAL OF B I O ~ I C A L CHEMISTRY The NH<sub>2</sub>-terminal 108 Amino Acids of the Escherichia coli DnaJ Protein Stimulate the ATPase Activity of DnaK and Are Sufficient for A Replication\*’, 1994.
- [28] ‘A conserved HPD sequence of the J-domain is necessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding’.
- [29] K. Liberek, J. Marszalek, D. Angt, C. Georgopoulostt, M. Zylicz, and A. M. Campbell, ‘Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK Communicated by’, 1991.
- [30] H. H. Kampinga *et al.*, ‘Function, evolution, and structure of J-domain proteins’, *Cell Stress and Chaperones*, vol. 24, no. 1. Cell Stress and Chaperones, pp. 7–15, Jan. 01, 2019. doi: 10.1007/s12192-018-0948-4.
- [31] ‘JDP classification Structure, function and evolution of DnaJ- conservation and adaptation of chaperone function’.
- [32] N. B. Nillegoda *et al.*, ‘Evolution of an intricate J-protein network driving protein disaggregation in eukaryotes’, doi: 10.7554/eLife.24560.001.
- [33] P. De Los Rios and A. Barducci, ‘Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption’, *Elife*, vol. 3, May 2014, doi: 10.7554/elife.02218.
- [34] M. P. Mayer, ‘Intra-molecular pathways of allosteric control in Hsp70s’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1749. Royal Society Publishing, Jun. 19, 2018. doi: 10.1098/rstb.2017.0183.
- [35] J. Jiang *et al.*, ‘Structural Basis of J Cochaperone Binding and Regulation of Hsp70’, *Mol Cell*, vol. 28, no. 3, pp. 422–433, Nov. 2007, doi: 10.1016/j.molcel.2007.08.022.
- [36] D. Cochaperones *et al.*, ‘Mechanism of Regulation of Hsp70 Chaperones’, 1999.
- [37] R. Kityk, J. Kopp, and M. P. Mayer, ‘Molecular Mechanism of J-Domain-Triggered ATP Hydrolysis by Hsp70 Chaperones’, *Mol Cell*, vol. 69, no. 2, pp. 227–237.e4, Jan. 2018, doi: 10.1016/j.molcel.2017.12.003.
- [38] Y. Jiang, P. Rossi, and C. G. Kalodimos, ‘PROTEIN FOLDING Structural basis for client recognition and activity of Hsp40 chaperones’. [Online]. Available: <http://science.sciencemag.org/>
- [39] D. Schmid, A. Baici, H. Gehring, and P. Christen, ‘Kinetics of Molecular Chaperone Action’, 1994.
- [40] M. Kabani, J.-M. Beckerich, and J. L. Brodsky, ‘Nucleotide Exchange Factor for the Yeast Hsp70 Molecular Chaperone Ssa1p’, *Mol Cell Biol*, vol. 22, no. 13, pp. 4677–4689, Jul. 2002, doi: 10.1128/mcb.22.13.4677-4689.2002.
- [41] Z. Dragovic, S. A. Broadley, Y. Shomura, A. Bracher, and F. U. Hartl, ‘Molecular chaperones of the Hsp110 family act as nucleotide exchange factors of Hsp70s’, *EMBO Journal*, vol. 25, no. 11, pp. 2519–2528, Jul. 2006, doi: 10.1038/sj.emboj.7601138.

- [42] C. Voisine, B. Schilke, M. Ohlson, H. Beinert, J. Marszalek, and E. A. Craig, 'Role of the Mitochondrial Hsp70s, Ssc1 and Ssq1, in the Maturation of Yfh1', 2000.
- [43] G. Sehlenstedt, S. Harris, B. Risse, R. Lill, and P. A. Silver, 'A Yeast DnaJ Homologue, Scjlp, Can Function in the Endoplasmic Reticulum with BiP/Kar2p Via a Conserved Domain That Specifies Interactions with Hsp70s'.
- [44] E. A. Craig and J. Marszalek, 'How Do J-Proteins Get Hsp70 to Do So Many Different Things?', *Trends in Biochemical Sciences*, vol. 42, no. 5. Elsevier Ltd, pp. 355–368, May 01, 2017. doi: 10.1016/j.tibs.2017.02.007.
- [45] D. I. Andersson, J. Jerlström-Hultqvist, and J. Näsvall, 'Evolution of new functions de novo and from preexisting genes', *Cold Spring Harb Perspect Med*, vol. 5, no. 7, Jul. 2015, doi: 10.1101/cshperspect.a017996.
- [46] M. W. Giorgianni, N. L. Dowell, S. Griffin, V. A. Kassner, J. E. Selegue, and S. B. Carroll, 'The origin and diversification of a novel protein family in venomous snakes', doi: 10.1073/pnas.1920011117/-/DCSupplemental.
- [47] A. S. Pillai *et al.*, 'Origin of complexity in haemoglobin evolution', *Nature*, vol. 581, no. 7809, pp. 480–485, May 2020, doi: 10.1038/s41586-020-2292-y.
- [48] S. Sharma *et al.*, 'Monitoring Protein Conformation along the Pathway of Chaperonin-Assisted Folding', *Cell*, vol. 133, no. 1, pp. 142–153, Apr. 2008, doi: 10.1016/j.cell.2008.01.048.
- [49] J. G. Bird, S. Sharma, S. C. Roshwalb, J. R. Hoskins, and S. Wickner, 'Functional analysis of CbpA, a DnaJ homolog and nucleoid-associated DNA-binding protein', *Journal of Biological Chemistry*, vol. 281, no. 45, pp. 34349–34356, Nov. 2006, doi: 10.1074/jbc.M603365200.
- [50] S. Cosgriff *et al.*, 'Dimerization and DNA-dependent aggregation of the Escherichia coli nucleoid protein and chaperone CbpA', *Mol Microbiol*, vol. 77, no. 5, pp. 1289–1300, Sep. 2010, doi: 10.1111/j.1365-2958.2010.07292.x.
- [51] R. Barriot, J. Latour, M. P. Castanié-Cornet, G. Fichant, and P. Genevaux, 'J-Domain Proteins in Bacteria and Their Viruses', *Journal of Molecular Biology*, vol. 432, no. 13. Academic Press, pp. 3771–3789, Jun. 12, 2020. doi: 10.1016/j.jmb.2020.04.014.
- [52] Z. Lu and D. M. Cyr, 'Protein Folding Activity of Hsp70 Is Modified Differentially by the Hsp40 Co-chaperones Sis1 and Ydj1\*', 1998.
- [53] A. K. Mandal, N. B. Nillegoda, J. A. Chen, and A. J. Caplan, 'Ydj1 Protects Nascent Protein Kinases from Degradation and Controls the Rate of Their Maturation', *Mol Cell Biol*, vol. 28, no. 13, pp. 4434–4444, Jul. 2008, doi: 10.1128/mcb.00543-08.
- [54] S. Silberstein, G. Sehlenstedt, P. A. Silver, and R. Gilmore, 'A Role for the DnaJ Homologue Scjlp in Protein Folding in the Yeast Endoplasmic Reticulum', 1998. [Online]. Available: <http://www.jcb.org>
- [55] N. Rowley *et al.*, 'Mdjlp, a Novel Chaperone of the DnaJ Family, Is Involved in Mitochondrial Biogenesis and Protein Folding', 1994.
- [56] G. L. Ciesielski *et al.*, 'Nucleoid localization of Hsp40 Mdj1 is important for its function in maintenance of mitochondrial DNA', *Biochim Biophys Acta Mol Cell*

- Res*, vol. 1833, no. 10, pp. 2233–2243, Oct. 2013, doi: 10.1016/j.bbamcr.2013.05.012.
- [57] K. Terada *et al.*, ‘A type I DnaJ homolog, DjA1, regulates androgen receptor signaling and spermatogenesis’, *EMBO Journal*, vol. 24, no. 3, pp. 611–622, Feb. 2005, doi: 10.1038/sj.emboj.7600549.
- [58] I. Baaklini, M. J. H. Wong, C. Hantouche, Y. Patel, A. Shrier, and J. C. Young, ‘The DNAJA2 substrate release mechanism is essential for chaperone-mediated folding’, *Journal of Biological Chemistry*, vol. 287, no. 50, pp. 41939–41954, Dec. 2012, doi: 10.1074/jbc.M112.413278.
- [59] S. F. Wang *et al.*, ‘Dnaja3/tid1 is required for mitochondrial dna maintenance and regulates migration and invasion of human gastric cancer cells’, *Cancers (Basel)*, vol. 12, no. 11, pp. 1–23, Nov. 2020, doi: 10.3390/cancers12113463.
- [60] X. Gao *et al.*, ‘Human Hsp70 Disaggregase Reverses Parkinson’s-Linked  $\alpha$ -Synuclein Amyloid Fibrils’, *Mol Cell*, vol. 59, no. 5, pp. 781–793, Sep. 2015, doi: 10.1016/j.molcel.2015.07.012.
- [61] ‘Structure, function and evolution of DnaJ- conservation and adaptation of chaperone function’.
- [62] A. S. Wentink *et al.*, ‘Molecular dissection of amyloid disaggregation by human HSP70’, *Nature*, vol. 587, no. 7834, pp. 483–488, Nov. 2020, doi: 10.1038/s41586-020-2904-6.
- [63] N. Österlund, M. Lundqvist, L. L. Ilag, A. Gräslund, and C. Emanuelsson, ‘Amyloid- $\beta$  oligomers are captured by the DNAJB6 chaperone: Direct detection of interactions that can prevent primary nucleation’, *Journal of Biological Chemistry*, vol. 295, no. 24, pp. 8135–8144, Jun. 2020, doi: 10.1074/jbc.RA120.013459.
- [64] S. McMahan, S. Bergink, H. H. Kampinga, and H. Ecroyd, ‘DNAJB chaperones suppress destabilised protein aggregation via a region distinct from that used to inhibit amyloidogenesis’, *J Cell Sci*, vol. 134, no. 7, Apr. 2021, doi: 10.1242/jcs.255596.
- [65] P. Walsh, D. Bursac, Y. C. Law, D. Cyr, and T. Lithgow, ‘The J-protein family: Modulating protein assembly, disassembly and translocation’, *EMBO Rep*, vol. 5, no. 6, pp. 567–571, Jun. 2004, doi: 10.1038/sj.embor.7400172.
- [66] B. Tomiczek *et al.*, ‘Two-step mechanism of j-domain action in driving hsp70 function’, *PLoS Comput Biol*, vol. 16, no. 6, Jun. 2020, doi: 10.1371/journal.pcbi.1007913.
- [67] O. Faust *et al.*, ‘HSP40 proteins use class-specific regulation to drive HSP70 functional diversity’, *Nature*, vol. 587, no. 7834, pp. 489–494, Nov. 2020, doi: 10.1038/s41586-020-2906-4.
- [68] H. Wyszowski *et al.*, ‘Class-specific interactions between Sis1 J-domain protein and Hsp70 chaperone potentiate disaggregation of misfolded proteins’, doi: 10.1073/pnas.2108163118/-/DCSupplemental.
- [69] H. Y. Yu *et al.*, ‘Roles of intramolecular and intermolecular interactions in functional regulation of the Hsp70 J-protein co-chaperone Sis1’, *J Mol Biol*, vol. 427, no. 7, pp. 1632–1643, 2015, doi: 10.1016/j.jmb.2015.02.007.

- [70] H. Rampelt *et al.*, ‘Metazoan Hsp70 machines use Hsp110 to power protein disaggregation’, *EMBO Journal*, vol. 31, no. 21, pp. 4221–4235, Oct. 2012, doi: 10.1038/emboj.2012.264.
- [71] A. Scior *et al.*, ‘Complete suppression of Htt fibrilization and disaggregation of Htt fibrils by a trimeric chaperone complex’, *EMBO J*, vol. 37, no. 2, pp. 282–299, Jan. 2018, doi: 10.15252/emboj.201797212.
- [72] P. Goloubinoff and P. D. L. Rios, ‘The mechanism of Hsp70 chaperones: (entropic) pulling the models together’, *Trends in Biochemical Sciences*, vol. 32, no. 8, pp. 372–380, Aug. 2007. doi: 10.1016/j.tibs.2007.06.008.
- [73] D. M. Walsh and D. J. Selkoe, ‘A $\beta$  oligomers - A decade of discovery’, *Journal of Neurochemistry*, vol. 101, no. 5, pp. 1172–1184, Jun. 2007. doi: 10.1111/j.1471-4159.2006.04426.x.
- [74] C. Månsson *et al.*, ‘Conserved S/T Residues of the Human Chaperone DNAJB6 Are Required for Effective Inhibition of A $\beta$ 42 Amyloid Fibril Formation’, *Biochemistry*, vol. 57, no. 32, pp. 4891–4892, Aug. 2018, doi: 10.1021/acs.biochem.8b00353.
- [75] J. Hageman *et al.*, ‘A DNAJB Chaperone Subfamily with HDAC-Dependent Activities Suppresses Toxic Protein Aggregation’, *Mol Cell*, vol. 37, no. 3, pp. 355–369, Feb. 2010, doi: 10.1016/j.molcel.2010.01.001.
- [76] V. Kakkar *et al.*, ‘The S/T-Rich Motif in the DNAJB6 Chaperone Delays Polyglutamine Aggregation and the Onset of Disease in a Mouse Model’, *Mol Cell*, vol. 62, no. 2, pp. 272–283, Apr. 2016, doi: 10.1016/j.molcel.2016.03.017.
- [77] Y. V. Taguchi *et al.*, ‘Hsp110 mitigates  $\alpha$ -synuclein pathology in vivo’, *Proc Natl Acad Sci U S A*, vol. 116, no. 48, pp. 24310–24316, Nov. 2019, doi: 10.1073/pnas.1903268116.
- [78] J. Tittelmeier *et al.*, ‘The HSP110/HSP70 disaggregation system generates spreading-competent toxic  $\alpha$ -synuclein species’, *EMBO J*, vol. 39, no. 13, Jul. 2020, doi: 10.15252/emboj.2019103954.
- [79] S. R. Eddy, ‘Accelerated profile HMM searches’, *PLoS Comput Biol*, vol. 7, no. 10, 2011, doi: 10.1371/journal.pcbi.1002195.
- [80] F. Sievers *et al.*, ‘Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega’, *Mol Syst Biol*, vol. 7, 2011, doi: 10.1038/msb.2011.75.
- [81] Z. Yang, R. Nielsen, and M. Hasegawa, ‘Models of Amino Acid Substitution and Applications to Mitochondrial Protein Evolution’, *Mol. Biol. Evol*, vol. 15, no. 12, pp. 1600–1611, 1998, [Online]. Available: <https://academic.oup.com/mbe/article/15/12/1600/963095>
- [82] H. C. Wang, B. Q. Minh, E. Susko, and A. J. Roger, ‘Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation’, *Syst Biol*, vol. 67, no. 2, pp. 216–235, Mar. 2018, doi: 10.1093/sysbio/syx068.
- [83] J. Shi, Y. Zhang, H. Luo, and J. Tang, ‘Using jackknife to assess the quality of gene order phylogenies’, 2010. [Online]. Available: <http://www.biomedcentral.com/1471-2105/11/168>

- [84] L. S. Quang, O. Gascuel, and N. Lartillot, ‘Empirical profile mixture models for phylogenetic reconstruction’, *Bioinformatics*, vol. 24, no. 20, pp. 2317–2323, Oct. 2008, doi: 10.1093/bioinformatics/btn445.
- [85] H. C. Wang, K. Li, E. Susko, and A. J. Roger, ‘A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny’, *BMC Evol Biol*, vol. 8, no. 1, 2008, doi: 10.1186/1471-2148-8-331.
- [86] H. Kishino and M. Hasegawa, ‘Journal of Molecular Evolution Evaluation of the Maximum Likelihood Estimate of the Evolutionary Tree Topologies from DNA Sequence Data, and the Branching Order in Hominoidea’, 1989.
- [87] H. Shimodaira and M. Hasegawa, ‘Letter to the Editor Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference’, *Mol. Biol. Evol*, vol. 16, no. 8, pp. 1114–1116, 1999, [Online]. Available: <https://academic.oup.com/mbe/article/16/8/1114/2925508>
- [88] H. Shimodaira, ‘An approximately unbiased test of phylogenetic tree selection’, *Syst Biol*, vol. 51, no. 3, pp. 492–508, Jun. 2002, doi: 10.1080/10635150290069913.
- [89] H. Ashkenazy *et al.*, ‘FastML: A web server for probabilistic reconstruction of ancestral sequences’, *Nucleic Acids Res*, vol. 40, no. W1, Jul. 2012, doi: 10.1093/nar/gks498.
- [90] C. Sahi *et al.*, ‘Sequential duplications of an ancient member of the DnaJ-family expanded the functional chaperone network in the eukaryotic cytosol’, *Mol Biol Evol*, vol. 30, no. 5, pp. 985–998, May 2013, doi: 10.1093/molbev/mst008.
- [91] A. Reyes-Prieto and A. Moustafa, ‘Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes’, *Sci Rep*, vol. 2, 2012, doi: 10.1038/srep00955.
- [92] K. Suzuki and S. Y. Miyagishima, ‘Eukaryotic and Eubacterial Contributions to the Establishment of Plastid Proteome Estimated by Large-Scale Phylogenetic Analyses’, *Mol Biol Evol*, vol. 27, no. 3, pp. 581–590, Mar. 2010, doi: 10.1093/molbev/msp273.
- [93] E. Susko and A. J. Roger, ‘Long Branch Attraction Biases in Phylogenetics’, *Syst Biol*, vol. 70, no. 4, pp. 838–843, Jun. 2021, doi: 10.1093/sysbio/syab001.
- [94] R. Lill *et al.*, ‘The role of mitochondria and the CIA machinery in the maturation of cytosolic and nuclear iron-sulfur proteins’, *European Journal of Cell Biology*, vol. 94, no. 7–9. Elsevier GmbH, pp. 280–291, Jul. 01, 2015. doi: 10.1016/j.ejcb.2015.05.002.
- [95] B. Roche, L. Aussel, B. Ezraty, P. Mandin, B. Py, and F. Barras, ‘Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity’, *Biochimica et Biophysica Acta - Bioenergetics*, vol. 1827, no. 3. pp. 455–469, Mar. 2013. doi: 10.1016/j.bbabi.2012.12.010.
- [96] R. Dutkiewicz and M. Nowak, ‘Molecular chaperones involved in mitochondrial iron–sulfur protein biogenesis’, *Journal of Biological Inorganic Chemistry*, vol. 23, no. 4. Springer Verlag, pp. 569–579, Jun. 01, 2018. doi: 10.1007/s00775-017-1504-x.





- System of *Escherichia coli*’, *Source*, vol. 97, no. 14, pp. 7790–7795, 2000, doi: 10.1073/pnas.1302.
- [109] J. J. Silberg, K. G. Hoff, and L. E. Vickery, ‘The Hsc66-Hsc20 Chaperone System in *Escherichia coli*: Chaperone Activity and Interactions with the DnaK-DnaJ-GrpE System’, 1998.
- [110] R. Dutkiewicz, B. Schilke, H. Knieszner, W. Walter, E. A. Craig, and J. Marszalek, ‘Ssq1, a mitochondrial Hsp70 involved in iron-sulfur (Fe/S) center biogenesis: Similarities to and differences from its bacterial counterpart’, *Journal of Biological Chemistry*, vol. 278, no. 32, pp. 29719–29727, Aug. 2003, doi: 10.1074/jbc.M303527200.
- [111] B. Schilke *et al.*, ‘The Cold Sensitivity of a Mutant of *Saccharomyces cerevisiae* Lacking a Mitochondrial Heat Shock Protein 70 Is Suppressed by Loss of Mitochondrial DNA’.
- [112] S. Laloraya, P. J. T. Dekker, W. Voos, E. A. Craig, and N. Pfanner, ‘Mitochondrial GrpE Modulates the Function of Matrix Hsp70 in Translocation and Maturation of Preproteins’, 1995.
- [113] S. R. Eddy, ‘Profile hidden Markov models’.
- [114] G. K. Sarangi, F. Romagné, and S. Castellano, ‘Distinct patterns of selection in selenium-dependent genes between land and aquatic vertebrates’, doi: 10.1093/molbev/msy070/4970564.
- [115] B. Q. Minh *et al.*, ‘IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era’, *Mol Biol Evol*, vol. 37, no. 5, pp. 1530–1534, May 2020, doi: 10.1093/molbev/msaa015.
- [116] A. Bateman, ‘UniProt: A worldwide hub of protein knowledge’, *Nucleic Acids Res*, vol. 47, no. D1, pp. D506–D515, Jan. 2019, doi: 10.1093/nar/gky1049.
- [117] G. Parra, K. Bradnam, and I. Korf, ‘CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes’, *Bioinformatics*, vol. 23, no. 9, pp. 1061–1067, 2007, doi: 10.1093/bioinformatics/btm071.
- [118] A. Stamatakis, ‘RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies’, *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, May 2014, doi: 10.1093/bioinformatics/btu033.
- [119] S. Puksza *et al.*, ‘Co-evolution-driven switch of J-protein specificity towards an Hsp70 partner’, *EMBO Rep*, vol. 11, no. 5, pp. 360–365, May 2010, doi: 10.1038/embor.2010.29.
- [120] P. A. Christin, D. M. Weinreich, and G. Besnard, ‘Causes and evolutionary significance of genetic convergence’, *Trends in Genetics*, vol. 26, no. 9, pp. 400–405, Sep. 2010, doi: 10.1016/j.tig.2010.06.005.
- [121] M. Y. Galperin and E. V. Koonin, ‘Divergence and convergence in enzyme evolution’, *Journal of Biological Chemistry*, vol. 287, no. 1, pp. 21–28, Jan. 02, 2012. doi: 10.1074/jbc.R111.241976.
- [122] J. F. Storz, ‘Causes of molecular convergence and parallelism in protein evolution’, *Nature Reviews Genetics*, vol. 17, no. 4. Nature Publishing Group, pp. 239–250, Apr. 01, 2016. doi: 10.1038/nrg.2016.11.

- [123] R. Kityk, J. Kopp, I. Sinning, and M. P. Mayer, ‘Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones’, *Mol Cell*, vol. 48, no. 6, pp. 863–874, Dec. 2012, doi: 10.1016/j.molcel.2012.09.023.
- [124] R. Qi *et al.*, ‘Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP’, *Nat Struct Mol Biol*, vol. 20, no. 7, pp. 900–907, Jul. 2013, doi: 10.1038/nsmb.2583.
- [125] A. Zhuravleva, E. M. Clerico, and L. M. Gierasch, ‘An interdomain energetic tug-of-war creates the allosterically active state in Hsp70 molecular chaperones’, *Cell*, vol. 151, no. 6, pp. 1296–1307, Dec. 2012, doi: 10.1016/j.cell.2012.11.002.
- [126] R. Kityk, M. Vogel, R. Schlecht, B. Bukau, and M. P. Mayer, ‘Pathways of allosteric regulation in Hsp70 chaperones’, *Nat Commun*, vol. 6, Sep. 2015, doi: 10.1038/ncomms9308.
- [127] M. Vogel, B. Bukau, and M. P. Mayer, ‘Allosteric regulation of Hsp70 chaperones by a proline switch’, *Mol Cell*, vol. 21, no. 3, pp. 359–367, Feb. 2006, doi: 10.1016/j.molcel.2005.12.017.
- [128] I. J. General, Y. Liu, M. E. Blackburn, W. Mao, L. M. Gierasch, and I. Bahar, ‘ATPase Subdomain IA Is a Mediator of Interdomain Allostery in Hsp70 Molecular Chaperones’, *PLoS Comput Biol*, vol. 10, no. 5, 2014, doi: 10.1371/journal.pcbi.1003624.
- [129] M. Vogel, M. P. Mayer, and B. Bukau, ‘Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker’, *Journal of Biological Chemistry*, vol. 281, no. 50, pp. 38705–38711, Dec. 2006, doi: 10.1074/jbc.M609020200.
- [130] J. F. Swain, G. Dinler, R. Sivendran, D. L. Montgomery, M. Stotz, and L. M. Gierasch, ‘Hsp70 Chaperone Ligands Control Domain Association via an Allosteric Mechanism Mediated by the Interdomain Linker’, *Mol Cell*, vol. 26, no. 1, pp. 27–39, Apr. 2007, doi: 10.1016/j.molcel.2007.02.020.
- [131] T. R. Alderson, J. H. Kim, K. Cai, R. O. Frederick, M. Tonelli, and J. L. Markley, ‘The specialized Hsp70 (HscA) interdomain linker binds to its nucleotide-binding domain and stimulates ATP hydrolysis in both cis and trans configurations’, *Biochemistry*, vol. 53, no. 46, pp. 7148–7159, Nov. 2014, doi: 10.1021/bi5010552.
- [132] A. L. Lai *et al.*, ‘Key features of an Hsp70 chaperone allosteric landscape revealed by ion-mobility native mass spectrometry and double electron-electron resonance’, *Journal of Biological Chemistry*, vol. 292, no. 21, pp. 8773–8785, May 2017, doi: 10.1074/jbc.M116.770404.
- [133] C. Dnaj *et al.*, ‘Interaction of the Hsp70 Molecular Chaperone, DnaK, with Its’, 1998.
- [134] C. S. Gassler *et al.*, ‘Mutations in the DnaK chaperone affecting interaction with the DnaJ cochaperone (Hsp70 heat shock proteins protein folding)’, 1998. [Online]. Available: [www.pnas.org](http://www.pnas.org).
- [135] D. Malinverni, A. Jost Lopez, P. De, L. Rios, G. Hummer, and A. Barducci, ‘Modeling Hsp70/Hsp40 interaction by multi-scale molecular simulations and coevolutionary sequence analysis’, doi: 10.7554/eLife.23471.001.

- [136] S. Q. Le and O. Gascuel, ‘An improved general amino acid replacement matrix’, *Mol Biol Evol*, vol. 25, no. 7, pp. 1307–1320, Jul. 2008, doi: 10.1093/molbev/msn067.
- [137] D. Darriba, G. L. Taboada, R. Doallo, and D. Posada, ‘ProtTest 3: Fast selection of best-fit models of protein evolution’, *Bioinformatics*, vol. 27, no. 8, pp. 1164–1165, Apr. 2011, doi: 10.1093/bioinformatics/btr088.
- [138] L. Dib, D. Silvestro, and N. Salamin, ‘Evolutionary footprint of coevolving positions in genes’, *Bioinformatics*, vol. 30, no. 9, pp. 1241–1249, May 2014, doi: 10.1093/bioinformatics/btu012.
- [139] Z. Yang, ‘PAML: a program package for phylogenetic analysis by maximum likelihood Downloaded from’, 1997. [Online]. Available: <http://mw511.biol.berkeley.edu/ziheng/paml.html>.
- [140] G. E. Crooks, G. Hon, J. M. Chandonia, and S. E. Brenner, ‘WebLogo: A sequence logo generator’, *Genome Res*, vol. 14, no. 6, pp. 1188–1190, Jun. 2004, doi: 10.1101/gr.849004.
- [141] D. De Juan, F. Pazos, and A. Valencia, ‘Emerging methods in protein co-evolution’, *Nature Reviews Genetics*, vol. 14, no. 4, pp. 249–261, Apr. 2013. doi: 10.1038/nrg3414.
- [142] L. A. Hug *et al.*, ‘A new view of the tree of life’, *Nat Microbiol*, vol. 1, no. 5, Apr. 2016, doi: 10.1038/nmicrobiol.2016.48.
- [143] A. E. Meyer, N.-J. Hung, P. Yang, A. W. Johnson, and E. A. Craig, ‘The specialized cytosolic J-protein, Jjj1, functions in 60S ribosomal subunit biogenesis’, 2007. [Online]. Available: [www.pnas.org/cgi/doi/10.1073/pnas.0610704104](http://www.pnas.org/cgi/doi/10.1073/pnas.0610704104)
- [144] C. Sahi and E. A. Craig, ‘Network of general and specialty J protein chaperones of the yeast cytosol’, 2007.