

UNIVERZITA KARLOVA

Přírodovědecká fakulta

katedra parazitologie

Studijní program: Biologie

Studijní obor: Parazitologie



Bc. Vít Dohnálek

Evoluce importu proteinů do mitochondrií

Evolution of mitochondrial protein import

DIPLOMOVÁ PRÁCE

Školitel: Mgr. Pavel Doležal, Ph.D.

Praha 2020

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne

.....

Poděkování

Chtěl bych poděkovat všem z naší laboratoře, jmenovitě pak Lence Markové za její pomoc s experimentální částí této práce a mému školiteli Pavlu Doležalovi za vstřícnost, ochotu a trpělivost. Dále bych rád poděkoval Martinu Kolískovi a lidem z jeho týmu, kteří mne zasvětili do světa bioinformatiky. Díky také patří mé rodině a přátelům za jejich podporu.

Abstrakt

Přestože mitochondrie obsahují vlastní genom a ribozomy, většina mitochondriálních proteinů je kódována v jádře a translatována na cytosolických ribozomech. Musely tak vzniknout transportní komplexy umožňující import proteinů z cytosolu. Tyto komplexy a jejich fungování jsou nejlépe popsány u kvasinek. Stále častěji se však setkáváme s organismy, které řadu proteinů popsaných u kvasinek pravděpodobně postrádají. Tato práce se proto snaží přinést přehled o distribuci podjednotek jednotlivých komplexů napříč eukaryoty. Blíže se zaměřujeme na extrémní redukci těchto komplexů u parazitického prvoka *Giardia intestinalis*, u něž se zatím podařilo identifikovat podjednotek pouze několik. Přestože se i přes vysokou míru divergence podařilo identifikovat homolog porinu Tom40, kterým prochází všechny proteiny směřující do mitochondrie, homolog proteinu Sam50, jenž je zodpovědný za vkládání Tom40 do membrány, se identifikovat nepodařilo. Značná část práce se tak věnuje tomuto mezi eukaryoty ojedinělému fenoménu.

Abstract

Even though mitochondria possess their own genome and ribosomes, majority of mitochondrial proteins is encoded in the nucleus and translated by cytosolic ribosomes. Hence it was necessary to establish transport complexes allowing the import of proteins from the cytosol. These complexes are best described in yeast. However, we are encountering organisms lacking many of the subunits of these complexes with increasing frequency. Therefore, we are presenting the overview of the distribution of the subunits within eukaryotic organisms. We specifically take a closer look at parasitic protist *Giardia intestinalis* that is well known for its extreme reductions of the import complexes. There have been only few subunits identified so far. Porin Tom40, that is responsible for translocating all the incoming proteins across the outer mitochondrial membrane, has been identified despite the high divergence, while homolog of Sam50 hasn't been successfully identified yet. Sam50 is however believed to be necessary for insertion of Tom40 into the membrane. Vast part of this thesis is dedicated to this phenomenon that is highly uncommon and maybe unique among the eukaryotic organisms.

Obsah

1 Úvod.....	1
2 Literární přehled.....	2
2.1 Mitochondriální importní dráhy.....	2
2.2 TOM komplex.....	3
2.3 SAM komplex.....	4
2.4 Transmembránové β -barely.....	5
2.5 Vkládání β -barelů do membrány.....	7
2.6 Malé Tim proteiny.....	8
2.7 Redukce importních komplexů.....	9
2.8 <i>Giardia intestinalis</i>	11
2.9 Predikce 3D struktury proteinů.....	12
3 Cíle práce.....	14
4 Metodika.....	15
4.1 Sekvenční alignment.....	15
4.2 Detekce homologie.....	15
4.2.1 BLAST.....	15
4.2.2 HMMER.....	15
4.2.3 HHpred.....	15
4.2.4 CLANS.....	16
4.3 Predikce sekundární struktury.....	16
4.4 Predikce transmembránových β -barelů.....	16
4.5 Modelování 3D struktury proteinů.....	17
4.6 Fylogenetická analýza.....	17
4.7 Kultivace použitých organismů.....	17
4.7.1 <i>Escherichia coli</i>	17
4.7.2 <i>Giardia intestinalis</i>	17
4.8 Klonování a exprese v <i>G. intestinalis</i>	18
4.8.2 PCR.....	18
4.8.3 Transformace bakterií.....	20
4.8.4 Transfekce <i>G. intestinalis</i>	21
4.9 Kontrola exprese a lokalizace.....	21
4.9.1 Příprava vzorku z celkového lyzátu <i>G. intestinalis</i>	21
4.9.2 SDS-PAGE.....	22
4.9.3 Barvení Coomassie Brilliant Blue.....	22
4.9.4 Western blot.....	22
4.9.5 Fluorescenční mikroskopie.....	23

5 Výsledky.....	25
5.1 Distribuce proteinů TOM komplexu.....	25
5.2 Distribuce Sam50.....	27
5.2.1 Homolog Sam50 <i>G. intestinalis</i>	27
5.2.2 Identifikace homologu Sam50 u blízce příbuzných <i>G. intestinalis</i>	31
5.2.3 Ztráta Sam50 v dalších eukaryotických liniích.....	33
5.3 Mutace β -signálu.....	38
5.4 Funkce β -signálu v mitochondriálních organelách postrádajících Sam50.....	40
5.5 Distribuce proteinů mitochondriálních importních komplexů.....	42
5.6 Redukce importních komplexů vlivem parazitického způsobu života.....	45
5.7 Malý Tim protein <i>G. intestinalis</i>	48
6 Diskuze.....	51
6.1 Ztráta Sam50.....	51
6.2 Mutace β -signálu.....	52
6.3 Teplotní mapy distribuce proteinů mitochondriálních importních komplexů.....	53
7 Závěr.....	55
8 Seznam použitých zkratk.....	56
9 Seznam použité literatury.....	57

1. Úvod

Tato práce obsahuje několik projektů, jejichž společným cílem bylo přispět k porozumění evoluce importu proteinů do mitochondrií. Mitochondrie je organela endosymbiotického původu obalená dvěma membránami, která si až na výjimky zachovává část vlastního genomu. Většina proteinů je nicméně kódována v jádře a translatována na cytosolických ribozomech. Musely proto vzniknout importní komplexy schopné transportu proteinů z cytosolu do mitochondrie. Pro většinu projektů jsme jako modelový organismus zvolili exkavátního prvoka *Giardia intestinalis*, jenž není pouze významným parazitem člověka a dalších obratlovců, ale také organismem s natolik redukovaným importním aparátem, že vybočuje z našich představ o fungování mitochondriálních importních komplexů. V této práci se pokoušíme zjistit, zda je importní aparát *G. intestinalis* skutečně natolik odlišný nebo zda pouze nejsme schopni identifikovat homology jednotlivých proteinů vlivem divergentní evoluce. Zároveň jsme se pokusili nalézt další organismy s podobně redukovanými importními komplexy jako má *G. intestinalis*, abychom zjistili jestli je stav u *G. intestinalis* skutečně výjimečný.

Je nutné zmínit, že v současné době již známe poměrně velké množství proteinů, které se na importu podílejí, a nebylo tak možné se detailně zaměřit na všechny. Zabývali jsme se především otázkou importu transmembránových β -barelů, jelikož právě otázka importu β -barelů je u *G. intestinalis* nedořešena. Značná část práce je tak věnována hledání Sam50, ale zároveň také hledání, zda je případná ztráta Sam50 u *G. intestinalis* mezi eukaryoty skutečně unikátní záležitostí.

2. Literární přehled

2.1 Mitochondriální importní dráhy

Prekurzory (preproteiny) mitochondriálních proteinů nesou různé signální sekvence, které jim umožňují interakci s příslušnými importními komplexy. Prvotní kontakt je však vždy navázán prostřednictvím receptorů TOM komplexu (Translocase of the outer membrane) (Brix et al. 1999). Preprotein je po rozeznání receptory translokován pomocí porinu Tom40 do mezimembránového prostoru (Hill et al. 1998) (**Obr. 1**).

Signální sekvence proteinů směřujících do matrix a vnitřní membrány je tvořena 15-55 aminokyselinami na N-konci, které jsou rozpoznány receptory komplexu TIM23 (Translocase of the inner membrane) (Vögtle et al. 2009; Wiedemann et al. 2007). Tyto presekvence jsou v matrix následně odstraněny specifickými mitochondriálními proteázami (Mossmann et al. 2012). Zatímco proteinům, jejichž cílovou destinací je matrix, stačí pouze N-terminální presekvence, proteiny vnitřní membrány ještě navíc nesou hydrofobní signál, který není nijak odštěpen a zůstává i po rozpoznání součástí maturovaného proteinu (Wiedemann et al. 2007).

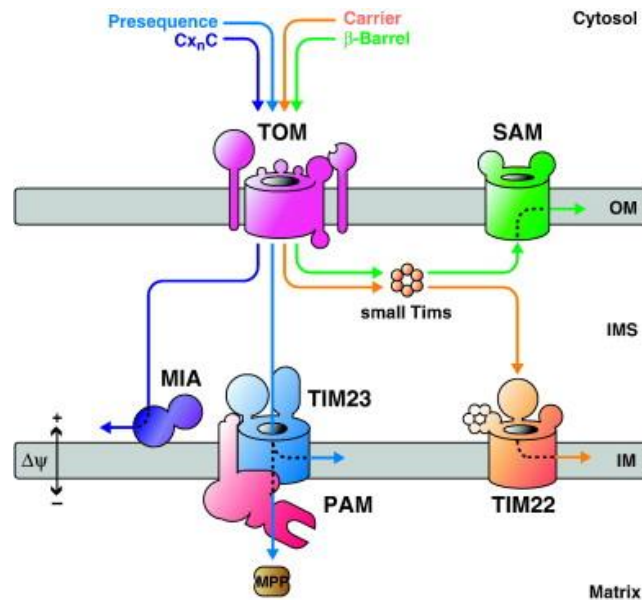
Do vnitřní membrány směřuje i specializovaná rodina mitochondriálních přenašečů (mitochondrial carrier, PF00153), které zajišťují transport metabolitů. Ty na rozdíl od ostatních proteinů nenesou štěpitelnou presekvenci a místo ní mají v sekvenci několik hydrofobních úseků, které jsou rozeznávány dalším transportním komplexem vnitřní membrány TIM22 (Sirrenberg et al. 1996).

Specifickou skupinou jsou proteiny mezimembránového prostoru, které mají ve své sekvenci úseky bohaté na cystein. Sulfhydrylová skupina cysteinu přenášeného proteinu vytváří dočasný cysteinový můstek s MIA komplexem (Mitochondrial intermembrane space assembly), jenž se stará o jejich zachycení a následné složení. Tím zabrání zpětnému úniku proteinů do cytosolu (Stojanovski et al. 2012).

Jaderně kódován je i samotný Tom40 a pro správné vložení do vnější membrány proto také potřebuje svůj specifický transportní signál. Společně s ostatními β -barely vnější mitochondriální membrány má na svém C-konci tzv. β -signál (Kutik et al. 2008). Ten je definován jako krátký aminokyselinový motiv: PoxGxxHyxHy, kde x může být libovolná aminokyselina, Po polární aminokyselina, Hy hydrofobní aminokyselina a G glycin. β -barely projdou skrz Tom40 do mezimembránového prostoru, kde jsou za pomoci chaperonů (tzv. malých Tim proteinů) přeneseny na SAM komplex (Sorting and assembly machinery), který obstará jejich vložení do membrány (Höhr et al. 2018).

Bioinformatické metody umožňují relativně snadnou detekci hlavních proteinů TOM a SAM komplexu, tedy Tom40 a Sam50, u všech eukaryotických linií. Výjimkou je *G. intestinalis* a blíže

příbuzné organismy, u nichž se nepodařilo detekovat homolog Sam50, ačkoli Tom40 je přítomen. Vzhledem k dosud zjištěné esenciální roli Sam50 by tak u těchto organismů nemělo docházet k funkčnímu skládání β -barelových proteinů do vnější mitochondriální membrány. Z evolučního hlediska se tak jedná o unikátní redukci transportních drah mitochondrie. Tato práce bude proto zaměřena především na komplexy TOM a SAM.



Obrázek 1: Schéma mitochondriálních importních drah. Preprotein je nejprve translokován pomocí TOM komplexu do mezimembránového prostoru odkud pokračuje do jednoho z dalších komplexů. Pro přenesení preproteinu do SAM nebo TIM22 komplexu jsou zapotřebí malé Tim proteiny. Převzato z (Dudek et al. 2013)

2.2 TOM komplex

TOM komplex je hlavní vstupní brána pro preproteiny importované do mitochondrie. Skládá se z několika podjednotek, jejichž počet se může značně lišit u jednotlivých eukaryotických linií (Araiso et al., 2019; Dagley et al., 2009; Heins and Schmitz, 1996). Nejlépe je TOM komplex popsán u kvasinek, kde se skládá ze sedmi podjednotek: Tom40, Tom22, Tom20, Tom7, Tom6, Tom5, Tom70 (Becker et al. 2012).

Klíčovým proteinem komplexu je β -barel Tom40, který je schopen vytvořit hydrofilní kanál, jímž dokážou preproteiny procházet do mezimembránového prostoru (Hill et al. 1998). Na rozdíl od ostatních Tom proteinů se homolog Tom40 pravděpodobně vyskytuje ve všech eukaryotických liniích, ačkoli některé homology vykazují vysokou míru divergence. U trypanosom se například nejprve předpokládalo, že namísto Tom40 využívají nepříbuzný porin, tzv. ATOM (Archaic translocase of the outer mitochondrial membrane), který je bakteriálního původu (Pusnik et al. 2011). Později ale bylo ukázáno, že ATOM je pouze velmi divergovaný Tom40 (Žarský et al. 2012).

Část přichozích preproteinů nejprve interaguje s receptorem Tom20, jehož cytosolická α -helikální doména obsahuje úsek rozpoznávající hydrofobní část N-terminální signální sekvence (Abe et al. 2000). Preproteiny s interním hydrofobním signálem jsou specificky rozeznávány receptorem Tom70, jehož C-terminální doména vytváří kapsu, kam se mohou proteiny vázat (Wu and Sha 2006). Tom70 navíc slouží i jako receptor pro protein Mfb1, který zajišťuje správnou mitochondriální morfogenezi (Kondo-Okamoto et al. 2008).

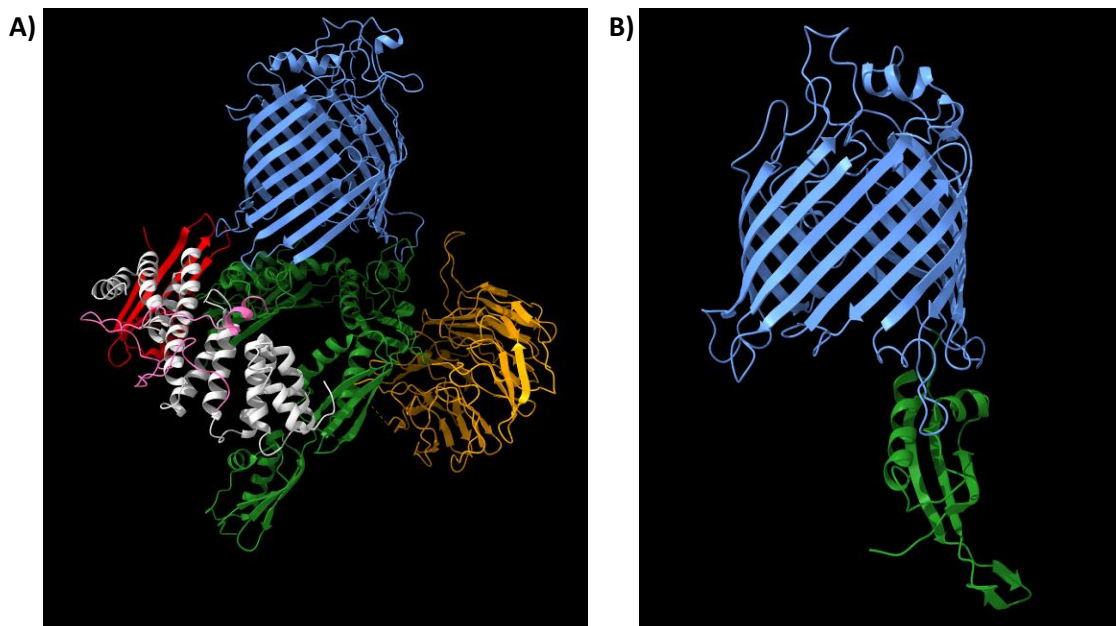
Na rozdíl od předešlých dvou proteinů, Tom22 obsahuje receptory pro signální sekvence jak v cytosolu, tak v mezimembránovém prostoru (Moczko et al. 1997). Cytosolická doména přejímá preprotein z Tom20 nebo Tom70 a posouvá ho do kanálu Tom40, zatímco doména v mezimembránovém prostoru se podílí na navázání preproteinu jakmile projde skrz kanál Tom40 (Komiya et al. 1998).

Pro správné fungování komplexu jsou také nutné tři malé α -helikální proteiny, Tom5, Tom6 a Tom7. Tom5 je v kooperaci se SAM komplexem nezbytný pro správné složení struktury Tom40 (Becker et al., 2010). Tom6 a Tom7 pravděpodobně fungují jako antagonisté, kdy Tom6 zajišťuje stabilní interakci Tom40 s receptorovými podjednotkami, zatímco Tom7 tuto stabilitu snižuje a zajišťuje tím potřebnou dynamiku celého komplexu. Při delecí Tom7 u *Saccharomyces cerevisiae* došlo ke stabilizaci interakcí mezi Tom40, Tom22 a Tom20, což vedlo k úplné inhibici proteinového transportu do mitochondrie (Dembowski et al. 2001; Hönlinger et al. 1996).

Všechny Tom proteiny jsou kódovány v jádře a translatovány na cytosolických ribozomech. Pro jejich správný import do mitochondrie je tedy zapotřebí již maturovaný TOM komplex a pro Tom40 i maturovaný SAM komplex (Becker et al. 2009).

2.3 SAM komplex

SAM komplex je zodpovědný za vkládání β -barelů do vnější mitochondriální membrány. Klíčovým proteinem SAM komplexu je Sam50, β -barel bakteriálního původu (Dolezal et al. 2006). Sam50 se skládá ze dvou domén; C-terminální domény označené jako Bac_surface_Ag (Bacterial surface antigen, PF01103) doména obsahující β -signál a N-terminální POTRA (polypeptide transport-associated, PF07244) domény (**Obr. 2**). Zatímco bakteriální homology mají POTRA domén několik, eukaryotický Sam50 má pouze jednu. Přesná role POTRA domény není známa. Předpokládalo se, že slouží jako receptor SAM komplexu, bylo však ukázáno, že i při její absenci jsou β -barelové prekurzory schopné interagovat s C terminální doménou Sam50 (Stroud et al. 2011). U plastidového proteinu Toc75, který patří do stejné rodiny jako Sam50, se POTRA domény chovají jako chaperony, je tedy možné, že zastávají stejnou funkci i v mitochondriích (O'Neil et al. 2017).



Obrázek 2: Ilustrační srovnání BAM komplexu bakterií (A) a eukaryotického Sam50 (B). Zeleně je označena POTRA doména (PF07244), modře Bac_surface_Ag doména (PF01103). BAM komplex je dále tvořen proteiny BamB (oranžová), BamC (růžová), BamD (bílá) a BamE (červená). Struktura BAM komplexu byla publikována v (Gu et al. 2016). Model Sam50 byl vytvořen pomocí homologního modelování.

Sam35 a Sam37 jsou menší proteiny SAM komplexu na cytosolické straně membrány. Sam37 interagující s Tom22 napomáhá tvorbě super komplexu TOM-SAM, který je nutný pro správnou biogenezi β -barelů (Wenz et al. 2015). Sam35 pravděpodobně úzce kooperuje se Sam50 a pomáhá mu s navázáním β -barelových prekurzorů (Chan and Lightow 2008). Zatímco Sam50 byl nalezen ve všech eukaryotických liniích, tak Sam37 u některých linií chybí a Sam35 nacházíme dokonce pouze u hub (Finn et al. 2014).

Posledním známým proteinem komplexu je Mdm10 (Mitochondrial distribution and morphology protein 10), membránový β -barel zodpovědný za mitochondriální morfologii. Na rozdíl od Sam50 a Sam35 Mdm10 interaguje s již maturovaným Tom40 a podílí se na tvorbě nového TOM komplexu (Meisinger et al. 2004).

2.4 Transmembránové β -barely

Transmembránové β -barely nalezneme nejčastěji u gram negativních bakterií. Prostřednictvím endosymbiózy však byly přeneseny i do eukaryot, kde je nalezneme v membránách mitochondrií a plastidů. Jsou definovány svou strukturou (**Obr. 3**). Jak název napovídá, jsou složeny převážně z β -listů, kdy je každý β -list navázán pomocí vodíkových vazeb na dva sousední, čímž se vytváří válcovitá struktura (Wimley 2003). β -listy mají navíc vždy sklon okolo

45° a vytvářejí pouze antiparalelní skládané listy. Většina známých β -barelů je tvořena sudým počtem β -listů, a než byla rozluštěna struktura mitochondriálního proteinu VDAC (Voltage dependent anion channel) s lichým počtem, tak se předpokládalo, že všechny β -barely musí mít počet β -listů sudý (Schulz 2000).

Celkový počet β -listů se může značně lišit. V současné době známe β -barely tvořené nejméně osmi a nejvíce dvaceti čtyřmi β -listy. Zatímco β -barely s nižším počtem β -listů mají spíše enzymatickou aktivitu, β -barely s počtem čtrnáct a více fungují většinou jako transportéry (Fairman et al. 2012). I zde však najdeme výjimky, kdy například u bakterie *Pseudomonas aeruginosa* nalezneme β -barel tvořený pouze osmi β -listy, který slouží jako kanál pro transport hydrofobních molekul skrze membránu (Touw et al. 2010). Menší β -barely ale u bakterií častěji fungují jako adhesiny či invasiny a pomáhají s napadením hostitelských buněk (Renault et al. 2009; Yamashita et al. 2011).

Dalšími funkcemi může být například schopnost vázat DNA či RNA (Bycroft et al. 1997; Lodi et al. 1995), případně mohou obsahovat aktivní místa pro enzymy (Reardon and Farber 2016) a mimo již zmíněného transportu proteinů (**viz kapitola 2.1**) se také podílejí na transportu iontů a metabolitů skrze membránu. V mitochondriích tuto funkci zastává protein VDAC, o jehož důležitosti svědčí to, že je nejhojnějším proteinem vnější mitochondriální membrány (Colombini et al. 1996; Galganska et al. 2010; Schein et al. 1976). VDAC je navíc schopný regulovat apoptózu vypuštěním proteinů cytochrom c, Smac/Diablo a AIF (apoptose-inducing factor) z mezimembránového prostoru mitochondrie do cytoplasmy (Abu-Hamad et al. 2009; Shoshan-Barmatz et al. 2010). VDAC je stejně jako Tom40 tvořen 19 β -listy a primární sekvence obou proteinů je také podobná, řadíme je proto do společné proteinové rodiny eukaryotických porinů (Bayrhuber et al. 2008; Tucker and Park 2019).

Podobně jako v mitochondriích je dnes i v plastidech většina proteinů kódována v jádře a translatována na cytosolických ribozomech (Cline and Dabney-Smith 2008). Existují zde proto funkční obdoby proteinů, které známe z mitochondrie. Tom40 je funkčně nahrazen porinem Toc75 (Perry and Keegstra 1994). Funkční obdobou Sam50 by v chloroplastu mohl být paralog Toc75 OEP80. Dostatečné důkazy pro podložení této hypotézy ovšem chybí (Day et al. 2014, 2019).



Obrázek 3: β -barel OmpA gram negativní bakterie *Klebsiella pneumoniae* složený z osmi β -listů sloužící jako adhesin (Renault et al. 2009). Zatímco část proteinu tvořená β skládanými listy (oranžově) se nachází v membráně, smyčky (modře) propojující jednotlivé β -listy jsou z větší části mimo membránu. N i C konce β -barelů se nacházejí v mezimembránovém prostoru. U bakteriálních β -barelů jsou oba konce často tvořeny dlouhou smyčkou, jako je tomu u zobrazeného proteinu (Schulz 2000). U eukaryotických β -barelů je naopak většinou alespoň jeden konec částečně tvořen α -helixem (Hiller et al. 2008; Tucker and Park 2019). Struktura KpOmpA byla publikována v (Renault et al. 2009).

Společná všem transmembránovým β -barelům je také vyšší mutační rychlost jejich primární sekvence ve srovnání se solubními proteiny. Části β -barelů nacházející se mimo membránu vykazují extrémní variabilitu a nejsou tak například vhodné pro fylogenetické studie (Schulz 2000). Tyto variabilní části proteinu nejsou tvořeny z hlediska sekundární struktury β -skládanými listy, nýbrž dlouhými smyčkami.

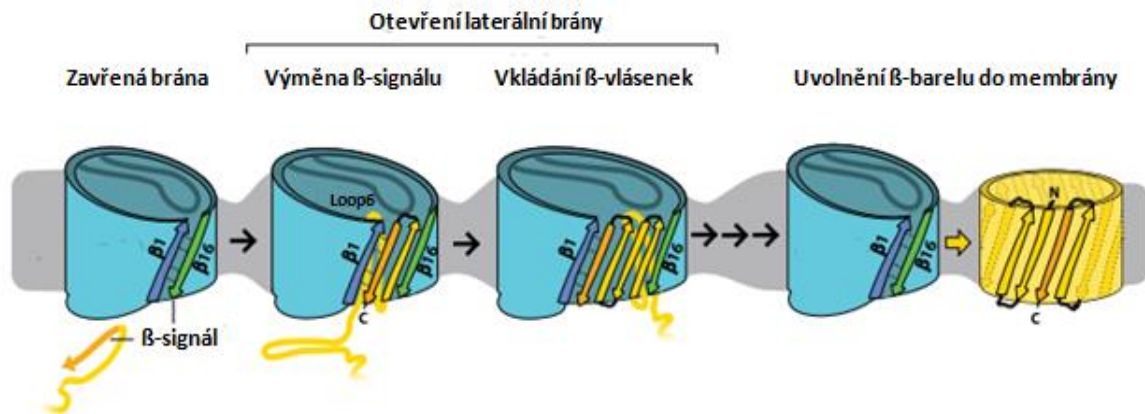
2.5 Vkládání β -barelů do membrány

V současné době existují dva odlišné modely snažící se popsat vkládání β -barelů do vnější mitochondriální membrány. První model předpokládá, že Sam50 dokáže dočasně způsobit ztenčení membrány, které umožní spontánní vstup β -barelu do membrány (Kim et al. 2012). Druhý model předpokládá, že se prekurzor β -barelu postupně tlačí do lumen Sam50 a je následně vypuštěn do membrány tzv. laterální bránou. Ta vzniká oddálením prvního a šestnáctého β -listu pouze za přítomnosti prekurzoru (Höhr et al. 2018; Noinaj et al. 2014).

První a šestnáctý β -list Sam50 jsou k sobě navázány pomocí β -signálu. Když je do lumen transportován β -barelový prekurzor, tak je tato vazba přerušena a první β -list Sam50 je navázán na β -list s β -signálem nově přichozího proteinu (**Obr. 4**). Zatímco C-konec prekurzoru s β -signálem zajišťuje navázání na Sam50, N-konec umožňuje dynamické interakce mezi oběma proteiny (Höhr et al. 2018). K těmto interakcím je nezbytná přítomnost další důležité součásti

Sam50, a to krátkého velmi konzervovaného aminokyselinového motivu, tzv. loop6 (Leonard-Rivera and Misra 2012).

Po spárování prvního β -listu Sam50 s C-koncem prekursoru, jsou do laterální brány po dvojicích postupně vkládány další β -listy prekursoru. Tyto dvojice antiparalelních skládaných listů označujeme jako β -vlásečky. Jakmile jsou v bráně přítomny všechny β skládané listy prekursoru, je protein vypuštěn z laterální brány do membrány (Höhr et al. 2018).



Obrázek 4: Model vkládání β -barelových prekursorů do membrány prostřednictvím laterální brány, která vzniká mezi 1 a 16 β -listem Sam50 po výměně β -signálu mezi Sam50 a přichozím β -barelem. Upraveno dle (Höhr et al. 2018).

2.6 Malé Tim proteiny

Jak bylo zmíněno v kapitole 2.1, pro správnou biogenezi β -barelů jsou klíčové tzv. malé Tim proteiny. Jedná se o rodinu cca 10 kDa velkých proteinů konzervovanou napříč eukaryoty (Gentle et al. 2007). Charakteristický je pro ně konzervovaný cysteinový „zinc-finger“ (zinkový prst) motiv $CX_3CX_{11-16}CX_3C$, umožňující vytvoření dvou disulfidických můstků, které pomáhají se stabilizací proteinu (Allen et al. 2003).

Malé Tim proteiny jsou stejně jako většina mitochondriálních proteinů syntetizovány na cytosolických ribozomech a do mitochondrií se dostávají skrz kanál tvořený Tom40. Při jejich biogenezi však navíc hrají důležitou roli změny redoxního potenciálu výše zmíněných cysteinů (Herrmann and Riemer 2012). V cytosolu a při průchodu Tom40 zůstávají cysteiny v redukovaném stavu, k oxidaci dochází až v mezimembránovém prostoru za pomoci MIA komplexu (Durigon et al. 2012; Lutz et al. 2003; Milenkovic et al. 2007). Na regulaci redoxního potenciálu se podílí více faktorů. V cytosolu jsou pro udržení redukčního stavu nejdůležitější ionty zinku (Lu and Woodburn 2005). V mezimembránovém prostoru je pro oxidaci pravděpodobně zásadní kooperace dvou proteinů. Thiol oxidáza Erv1 funguje jako oxidační činidlo, zatímco Hot13 působí jako chelátor zinku, který by v nenavázané formě inhiboval Erv1 (Mesecke et al. 2008; Morgan et al. 2009).

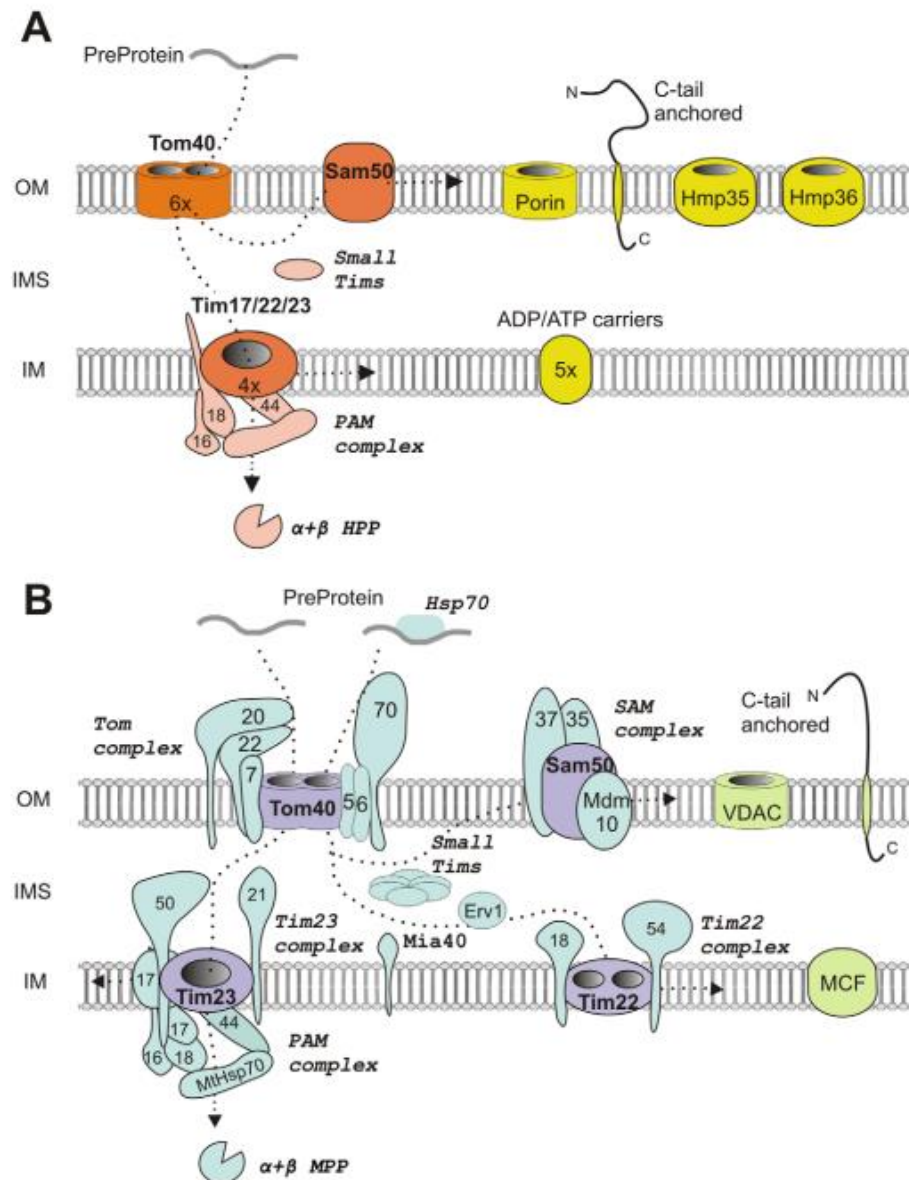
Po oxidaci a částečném složení proteinu dochází k tvorbě hexamerových komplexů, které fungují jako chaperony a pomáhají prekurzorům β -barelů a některých dalších proteinům s pohybem skrze mezimembránový prostor (Ivanova et al. 2008).

Zajímavý byl objev dvou homologů malých Tim proteinů u *Trichomonas vaginalis*. Na rozdíl od ostatních homologů neobsahují charakteristický cysteinový motiv a ve své sekvenci mají zachovaný pouze jeden cystein. Zároveň se u *T. vaginalis* nepodařilo identifikovat ani jednu z komponent MIA dráhy (Mia40, Erv1, Hot13). Předpokládá se, že by tyto modifikace mohli být způsobeny vlivem anaerobního prostředí, v němž *T. vaginalis* žije (Rada et al. 2011). U *G. intestinalis* dosud nebyly objeveny homology malých Tim proteinů, které by umožňovaly pohyb prekurzorů β -barelových proteinů v mezimembránovém prostoru. To společně s absencí Sam50 vyvolává řadu funkčních otázek o celkovém mechanismu membránového skládání β -barelů v takto modifikovaném biologickém systému.

2.7 Redukce importních komplexů

Výše uvedené popisy importních komplexů byly většinou získány studii využívajícími jako modelový organismus kvasinky. Napříč eukaryoty se ale výskyt některých podjednotek může značně lišit. Pomocí sekvenčních analýz se takto například ukázalo, že pouze tři ze sedmi podjednotek kvasinkového TOM komplexu lze běžně identifikovat u většiny eukaryot, a to Tom40, Tom22 a Tom7 (Mačašev et al. 2004). Příčinou může být nedostatečná citlivost našich metod pro detekci vzdálených homologů, avšak je možné, že se u různých linií eukaryot konvergentně vyvinuly jejich vlastní podjednotky, které bychom zase naopak nebyli schopni nalézt u kvasinek. Příkladem konvergentní evoluce může být rostlinný Tom20, který má zřejmě stejnou funkci jako kvasinkový Tom20, přestože jejich primární sekvence jsou zcela odlišné (Perry et al. 2006).

Existují i skupiny, kde máme s použitím současných metod potíže s nalezením i oněch tří základních podjednotek TOM komplexu. Například u skupiny Metamonada, zahrnující významné parazitické organismy (např.: *T. vaginalis*, *G. intestinalis*), i volně žijící organismy (např.: *Carpediemonas membranifera*), nám současné metody umožňují u většiny zástupců nalezení pouze homologu Tom40 (Leger et al. 2017). Na vině může být vysoká míra divergence primárních sekvencí, která byla ukázána například u Tom40 *G. intestinalis* pomocí klastrovací analýzy (Dagley et al. 2009). Na druhou stranu u zmíněné *T. vaginalis* se podařilo identifikovat dva β -barely vnější hydrogenosomální membrány pojmenované Hmp35 a Hmp36 (**Obr. 5**), které nevykazují zřejmou homologii s žádnými dosud známými proteiny, ale na základě predikce sekundární struktury se zdá, že by mohli mít podobnou strukturu jako proteiny VDAC (Dyall et al. 2003; Rada et al. 2011).

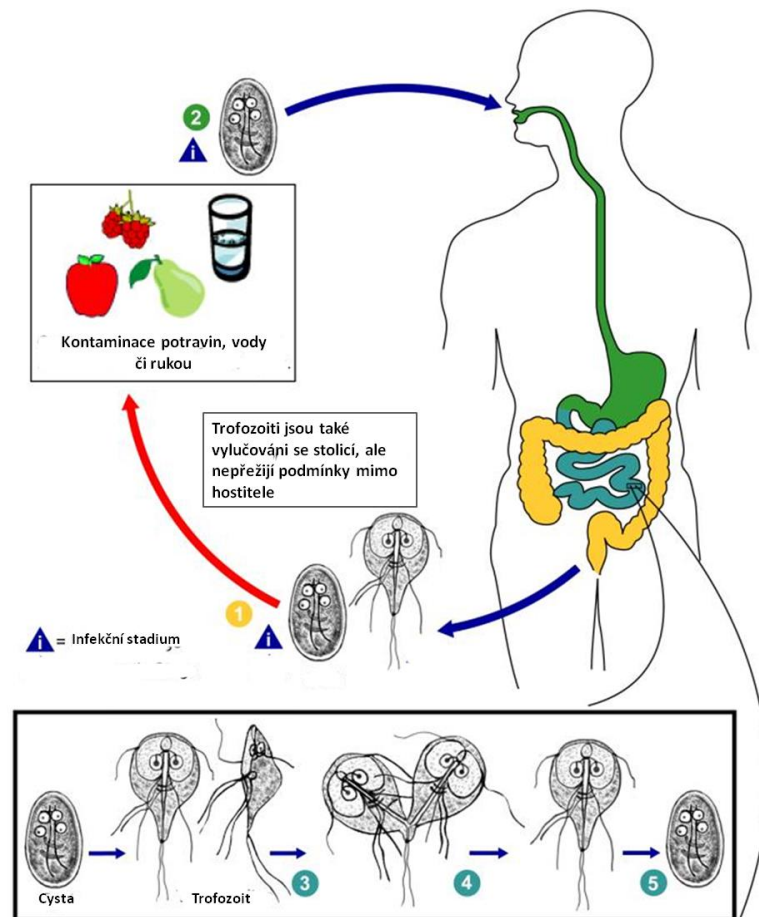


Obrázek 5: Srovnání proteinových komplexů vnitřní a vnější hydrogenosomální membrány *T. vaginalis* (A) s komplexy vnitřní a vnější mitochondriální membrány *S. cerevisiae* (B). Proteiny společné oběma organismům (Tom40, Sam50, Tim17/22/23, proteiny PAM komplexu, malé Tim proteiny) vykazují pouze nízkou sekvenční identitu, kvůli čemuž je nemožné říct, zda proteiny z rodiny Tim17/22/23 vytvářejí dva funkčně oddělené komplexy jako je tomu u *S. cerevisiae*, nebo zda vytváří pouze jeden druh multifunkčního komplexu. Homology většiny kvasinkových proteinů se u *T. vaginalis* nepodařilo identifikovat vůbec, nicméně β -barely Hmp35 a Hmp36 nemají naopak žádný homolog u *S. cerevisiae*. Ne všechny chybějící homology u *T. vaginalis* se připisují redukcí proteinových komplexů, například protein Mia40 se nepodařilo identifikovat v žádných exkavátních organismech a předpokládá se tak, že se exkavátní linie oddělila ještě před vznikem tohoto proteinu. Upraveno dle (Rada et al. 2011)

2.8 *Giardia intestinalis*

Jako modelový organismus jsme si zvolili obligátně parazitický organismus *G. intestinalis*. Jedná se o exkavátního prvoka ze skupiny Diplomonadida způsobujícího onemocnění giardiózu. Toto onemocnění může probíhat zcela bezpříznakově, u většiny infikovaných se ale projevuje průjmy a celkovou vyčerpaností. Závažnější infekce mohou způsobovat i horečku a zvracení. Přestože symptomatickou giardiózu každoročně prodělávají desítky milionů lidí, je jí kvůli nízkému zdravotnímu riziku věnována poměrně malá pozornost (Minetti et al. 2016).

G. intestinalis prochází během životního cyklu dvěma vývojovými stádii (**Obr. 6**). Infekční je stádium cysty, které z těla infikovaného odchází společně se stolicí. Když je cysta pozřena dostává se skrze trávicí trakt až do oblasti duodena případně jejunu, kde dochází k excystaci. Vzniklé stádium trofozoita se množí asexuálně binárním dělením a napadá střevní epitel. Když se trofozoit dostane do dolní části trávicího traktu, je vlivem vyšší koncentrace žluči a zvýšením pH vyvolán proces encystace (Gillin et al. 1988; Jex et al. 2020)

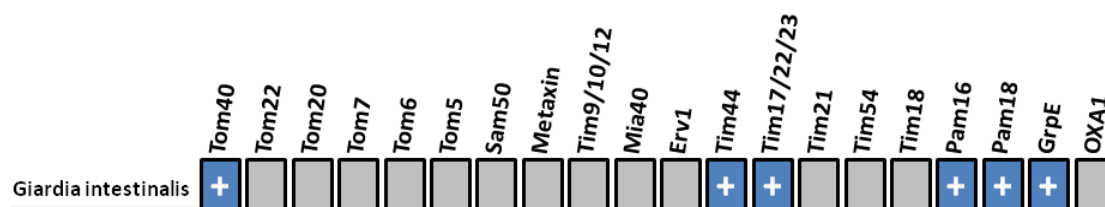


Obrázek 6: Životní cyklus *G. intestinalis*. Zeleně je označena část trávicího traktu, kterou prochází cysty před excystací, v modré části se již vyskytují trofozoiti parazitující na střevním epitelu, ve žluté části dochází k encystaci a odchodu cyst do volného prostředí společně se stolicí. Upraveno dle (<http://www.cdc.gov>)

U *G. intestinalis* došlo k četným redukcím na buněčné úrovni, chybí zde Golgiho aparát, peroxisomy a klasická mitochondrie je nahrazena mitosomy, které neobsahují genetickou informaci a nejsou schopné zajišťovat průběh oxidativní fosforylace. Jediná doposud známá zachovaná funkce u mitosomů *G. intestinalis* je podíl na syntéze železo-sirných klastrů (Tovar et al. 2003). Na druhou stranu u *G. intestinalis* došlo k duplikaci jader, vytvoření osmi bičků a speciálního adhezivního disku, pomocí něhož se přichycuje na buňky epitelu tenkého střeva, kde získává živiny prostřednictvím pinocytózy.

G. intestinalis jsme zvolili jako modelový organismus pro její extrémní redukce mitosomálních importních komplexů, zejména absenci SAM komplexu (**Tab. 1**). Výhodou je navíc i relativně snadná kultivace.

Tabulka 1: Přehled podjednotek importních mitochondriálních komplexů a jejich výskyt u *G. intestinalis* na základě literatury (Leger et al. 2017; Pyrihová et al. 2018).



2.9 Predikce 3D struktury proteinů

Pro část naší práce bylo zásadní, co možná nejlépe predikovat 3D strukturu proteinů pouze na základě aminokyselinových sekvencí. V současné době je k dispozici několik metod, které můžeme rozdělit do dvou skupin. První skupina k predikci využívá 3D struktury proteinů získané experimentálními metodami (homologní modelování), metody z druhé skupiny se naopak snaží o predikci nezávisle na použití již známé struktury (modelování ab initio). Oba přístupy mají svá specifika.

Pro homologní modelování je klíčové zvolit správnou strukturu, která bude použita jako templát. Nejlepších výsledků dosahují templáty zvolené na základě homologie (Deng et al. 2018). Přestože počet známých struktur v databázi PDB (Berman et al. 2002) každoročně roste, nalézt blízký homolog pro námi studovaný protein není vždy možné. V takovém případě je možné najít alespoň proteiny s částečně podobnou strukturou pomocí tzv. fold recognition metod (Bowie et al. 1991; Jones 1999a). Tyto metody jsou založeny na teorii, že struktura proteinů je v evoluci mnohem lépe konzervována než primární sekvence a že počet unikátních možností sbalení (foldu) proteinů je značně omezen (Illergård et al. 2009). Výhodou modelování dle templátu je vysoká přesnost v případě, že jsme schopni v databázi najít blízký homolog. To je však zároveň i limitací, jelikož homologní struktura může v databázi chybět a

přestože metody založené na fold recognition jsou neustále vylepšovány, nemusejí být vždy dostatečné (Yang et al. 2011).

Metody ab initio predikují strukturu proteinů pouze na základě fyzikálních zákonů, přesto i tak často vycházejí z již známých 3D struktur. Hlavním rozdílem je tedy absence templátu během predikce (Deng et al. 2018). Predikční programy s ab initio přístupem nejčastěji vstupní sekvenci rozdělí na menší části a nejprve vypočítají strukturu těchto fragmentů. Následně jednotlivé části skládají do finální podoby (Kim et al. 2004). Hlavní limitace tohoto přístupu souvisí s velikostí proteinů. Čím větší protein, tím je model nepřesnější a zároveň se zvyšuje celková doba predikce. Modely proteinů skládajících se z víc jak 150 aminokyselin jsou tak už velmi nepřesné a zároveň je jejich predikce velmi výpočetně náročná (Deng et al. 2018).

3. Cíle práce

1. Nalezení homologu Sam50 *G. intestinalis* či případné vysvětlení jeho absence
2. Nalezení jiných organismů bez zjevného homologu Sam50
3. Potvrzení či vyvrácení univerzality β -signálu
4. Vytvoření přehledu distribuce podjednotek importních komplexů napříč eukaryoty

4. Metodika

4.1 Sekvenční alignment

Pro alignment sekvencí byl používán program MUSCLE (Madeira et al. 2019), dostupný na webovém serveru EBI (The European Bioinformatics Institute, ebi.ac.uk/Tools/msa/muscle) nebo program MAFFT verze 7 (Rozewicki et al. 2019), který CBRC (Computational Biology Research Consortium) volně poskytuje ke stažení. Pro automatickou úpravu alignmentů, tzn. odstranění úseků s nízkou sekvenční podobností apod., byl použit program trimAl verze 1.2 (Capella-Gutiérrez et al. 2009) dostupný na webových stránkách (<http://trimal.cgenomics.org/trimal>).

4.2 Detekce homologie

Pro detekci homologních sekvencí bylo během práce užito několik následujících nástrojů.

4.2.1 BLAST

BLAST je program používaný pro detekci homologie na základě srovnávání primárních sekvencí (Altschul et al. 1997). Během práce byly používány různé varianty tohoto programu, například blastp pro srovnávání aminokyselinových sekvencí, tblastn pro srovnávání aminokyselinových sekvencí s nukleotidovými atd. Byla také používána speciální iterativní verze, tzv. PSI-BLAST, která je vhodná pro hledání vzdálenějších homologů. Program byl využíván buď pomocí webového serveru NCBI (The National Center for Integrative and Biomedical Informatics, blast.ncbi.nlm.nih.gov/Blast.cgi) nebo lokálně, kdy byla použita verze 2.10.0, kterou NCBI volně poskytuje ke stažení.

4.2.2 HMMER

HMMER je sada nástrojů pro detekci homologie využívající skrytý Markovův model (HMM). Vstupní informací může být buď samostatná sekvence nebo profil vytvořený na základě HMM z alignmentu aminokyselinových sekvencí. Tvorba HMM profilů je ve webové verzi automatizována, zatímco lokálně je možné využít program hmbuild, který je součástí standardní sady (Finn et al. 2011). Program byl používán buď na webovém serveru EBI (ebi.ac.uk/Tools/hmmer) nebo lokálně, kdy byla použita verze 3.1b2, která je EBI volně poskytnuta ke stažení.

4.2.3 HHpred

HHpred slouží k detekci vzdálených homologů. Stejně jako HMMER využívá HMM, na rozdíl od HMMER ale nesrovnává vstupní informace pouze s databází jednotlivých sekvencí, ale umožňuje hledání v databázi vytvořené z HMM profilů (např.: Pfam (Finn et al. 2014)), což

umožňuje nalezení velmi vzdálených homologů, ale počet použitelných databází je tím značně omezen (Zimmermann et al. 2018). Při použití webové verze (toolkit.tuebingen.mpg.de/tools/hhpred) je z alignmentu aminokyselinových sekvencí nejprve vytvořen HMM profil. Ten je poté iterativně srovnáván pomocí programu HHblits s UniRef30 databází. Z výsledků je vytvořen nový alignment a HMM profil, který je poté použit jako vstupní informace pro program HHsearch, který již hledá proti námi zvolené databázi.

Při lokálním užití je možné použít HMM profily vytvořené pomocí HMMER, ale rychlost vyhledávání je tím snížena. Pro tvorbu HMM profilů pro lokální použití HHpred byl proto použit program hmake, který je součástí sady nástrojů HH-suite (Söding et al. 2005). Během práce byla využívána webová verze i HH-suite verze 3, která je volně dostupná na github.

4.2.4 CLANS

Pro klastrovací analýzy byl použit program CLANS (Frickey and Lupas 2004). Tento program nejprve porovná pomocí BLAST všechny sekvence mezi sebou a ze získaných p-value spočítá vzájemnou příbuznost sekvencí. Příbuznost sekvencí je poté vynesena do interaktivního 3D grafu. Využívali jsme volně dostupnou webovou verzi programu (<https://toolkit.tuebingen.mpg.de/tools/clans>).

4.3 Predikce sekundární struktury

Pro predikci sekundární struktury byl použit program PSIPRED (Buchan and Jones 2019). Program nejprve s využitím PSI-BLAST vytvoří profil z homologních sekvencí. Následně je prostřednictvím neurálních sítí vypočítána pravděpodobnost s jakou se daná aminokyselina podílí na tvorbě α -helixu, β -listu či smyčky (Jones 1999b). Webová verze programu srovnává vstupní sekvenci s databází UniRef90. Při lokálním použití jsme z důvodu snížení časové náročnosti zvolili databázi Swiss-Prot (Boutet et al. 2007). Během práce byla využívána verze ne webovém serveru UCL (London's Global University) i verze volně poskytnutá ke stažení (bioinf.cs.ucl.ac.uk/psipred).

4.4 Predikce transmembránových β -barelů

Pro predikci převážně velmi divergentních homologů Tom40 a Sam50 jsme otestovali sensitivitu následujících tří programů: Boctopus2 (Hayat et al. 2016), PRED-TMBB (Bagos et al. 2004) a PRED-TMBB2 (Tsirigos et al. 2016). Na predikci transmembránových β -barelů u námi studovaných organismů se jako nejlepší ukázal PRED-TMBB2, jehož webová verze navíc umožňuje testování až 20 000 sekvencí najednou (compngen.org/tools/PRED-TMBB2).

4.5 Modelování 3D struktury proteinů

Pro homologní modelování 3D struktury proteinů byl použit program Modeller verze 9.23 (Webb and Sali 2017). Aktivační klíč jsme obdrželi od autorů programu. Při práci byly převzaty

některé skripty z webových stránek autorů (salilab.org/modeller) a upraveny dle potřeby. Pro modelování metodou ab initio jsme používali webovou verzi programu QUARK (Xu and Zhang 2012, 2013) dostupnou na <https://zhanglab.ccmb.med.umich.edu/QUARK/>. Jako jeden ze způsobů hodnocení modelů jsme využívali Ramachandranovy diagramy generované programem RAMPAGE (mordred.bioc.cam.ac.uk/~rapper/rampage.php). Pro vizualizaci modelů byl použit program UCSF-ChimeraX (Goddard et al. 2018).

4.6 Fylogenetická analýza

Pro fylogenetickou analýzu jsme využili sadu programů PhyloBayes verze 2.3 dostupnou na webových stránkách (<http://www.atgc-montpellier.fr/>) spravovaných IFB (Institut Français de Bioinformatique), která ze sekvenčních alignmentů za pomoci metody Monte Carlo Markov Chain (MCMC) dokáže vytvořit fylogenetické stromy (Lartillot and Philippe 2004, 2006; Lartillot et al. 2007). Pro vizualizaci stromů jsme použili program FigTree verze 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>), který je volně dostupný na github.

4.7 Kultivace použitých organismů

4.7.1 *Escherichia coli*

Pro namnožení plazmidů byly použity kompetentní bakterie *E. coli* kmene TOP 10. Buňky byly transformovány pomocí tepelného šoku (42 °C, 45 s) a kultivovány na třepačce v LB médiu s přidaným selekčním antibiotikem ampicilinem (Sigma, koncentrace 100 µg/ml) při teplotě 37 °C. Pro selektování klonů transformovaných bakterií byly buňky kultivovány na agarových plotnách při teplotě 37 °C. Buňky byly dlouhodobě uchovány zamražením s 20% glycerolem.

4.7.2 *Giardia intestinalis*

Pro experimenty byly použiti trofozoiti *G. intestinalis* kmene WB1 (ATCC 30957). Buňky byly kultivovány při teplotě 37 °C v upraveném médiu TYI-S-33 s přidaným 10% hovězím sérem (Gibco) a 0,1% hovězí žlučí (Keister 1983). Po transfekci buněk byly linie selektovány pomocí antibiotik puromycin (PAA laboratorie, koncentrace 54 µg/ml) a G418 (Gold biotechnology, koncentrace 0,6 mg/ml). Buňky byly přeočkovávány do nového média každé 3 – 4 dny. Buňky byly dlouhodobě uchovány zamražením v dusíku s 0,25% dimethylsulfoxidem.

TYI-S-33	1000 ml
dH ₂ O	870 ml
10% inaktivované hovězí sérum	100 ml
BBL - Trypticase Pepton	20 g
Extrakt z kvasinek	10 g
Glukóza	10 g
Penicilin-Streptomycin	10 ml
NaCl	2 g
L-cystein	2 g
K ₂ HPO ₄	1 g
0,1% hovězí žluč	1 ml
KH ₂ PO ₄	0,6 g
Kyselina askorbová	0,2 g
Citrát železito-amonný	0,0228 g

Optimálního pH (6,8) je dosaženo přidáním NaOH. Médium je nutné sterilizovat pomocí filtrace (Millipore, Filter Steritop 0,22 µm) z důvodu citlivosti některých složek na teplo.

4.8 Klonování a exprese v *G. intestinalis*

4.8.2 PCR

Amplifikovány byly dvě varianty genu pro Tom40 *G. intestinalis* (**Tab. 2**). Standardní a poté varianta, kdy je z genu odebrána část kódující β-signal. Odebrání části sekvence bylo dosaženo pomocí použití alternativního reversního primeru.

Tabulka 2: Použité primery s vyznačenými restrikčními místy. Forward primer byl stejný pro oba konstrukty. První reverse primer nasedá na konci genu Tom40 a umožní tak amplifikaci celého genu, zatímco druhý reverse primer nasedá před úsekem kódujícím β-signal.

GiTom40 - Forward	AseI	CATG GATTA ATATGCCCTTTCCTGGTCTA
GiTom40 - Reverse	PstI	CATG CTGCAG GCGGCCTGAGAAAATGTC
GiTom40; no β-signal - Reverse	PstI	CATG CTGCAG CTTGATTGCTCGAAGA

Geny byly amplifikovány metodou PCR (Polymerase chain reaction) s použitím genomové DNA *G. intestinalis* získané izolací z buněk. Pro izolaci byl použit kit Genomic DNA mini kit (Genaid). Reakci jsme namíchali dle následujícího protokolu.

PCR	25 µl
H ₂ O	13,25 µl
5x Q5 reakční pufr	5 µl
Q5 Enhancer	5 µl
10 mM dNTPs	0,5 µl
10 µM Primer F	0,25 µl
10 µM Primer R	0,25 µl
gDNA	0,5 µl
Q5 High–Fidelity DNA Polymerase	0,25 µl

PCR program:

1. 98 °C	30 s
2. 98 °C	5 – 10 s
3. 55 °C	10 –30 s
4. 72 °C	45 s
30 cyklů 2 –4	
5. 72 °C	2 min
6. 4 °C	∞

PCR produkt byl analyzován na 1,5% agarózovém gelu za pomoci horizontální gelové elektroforézy. Díky přidanému interkalačnímu barvivu SYBR Safe (Invitrogen) bylo možné detekovat DNA prostřednictvím UV transiluminátoru.

Pomocí High Pure PCR Product Purification Kit (Roche) byla z gelu izolována DNA. Izolovaná DNA byla následně společně s plazmidem pTG (Lauwaet et al. 2007) naštěpena přes noc restrikčními enzymy dle následujícího protokolu.

Restrikční reakce (pTG)	100 µl
H ₂ O	83 µl
Plazmid/inzert	5 µl (1µg DNA)
PstI	1 µl
NdeI	1 µl
Cutsmart pufr	10 µl

Produkty restrikčních reakcí byly analyzovány pomocí 1,5% agarózového gelu a opět vyčištěny za pomoci výše zmíněného kitu. Vložení inzertu do plazmidu bylo dosaženo pomocí následující ligační reakce, která probíhala přes noc při 16 °C. V reakci jsme použili T4 ligázu (Thermo Fisher Scientific) a k ní příslušný pufr.

Ligační reakce	15 µl
Plazmid	2 µl (1µg DNA)
Inzert	5 µl (1µg DNA)
T4 ligáza	1 µl
T4 pufr	2 µl
H ₂ O	5 µl

Pro expresi proteinů v *G. intestinalis* byly využity dva plazmidy. Pro protein Tom40 byl použit modifikovaný plazmid pTG (Jerlström-Hultqvist et al. 2012), pro malý Tim plazmid pONDRA, vzniklý modifikací plazmidu pRAN-neoGDHluc (Dolezal et al. 2005).

4.8.3 Transformace bakterií

Produkt ligační reakce byl přidán ke kompetentním bakteriím *E. coli* TOP10. Buňky se následně inkubovaly 15 minut na ledu. Dalším krokem byl tepelný šok, kdy byly buňky po dobu 1 minuty vystaveny teplotě 42 °C. Následovala opětovná inkubace na ledu po dobu 2 minut. K bakteriím bylo nakonec přidáno 500 µl LB média a při teplotě 37 °C byly inkubovány na třepačce (220 RPM) po dobu 30 minut. Narostlé buňky byly přeneseny na agarovou plotnu s LB médiem a přidáním antibiotikem. Kultivace na plotnách probíhala přes noc při teplotě 37 °C.

Z každé plotny bylo vybráno pět kolonií, které byly otestovány, zda mají plazmid s příslušným genem. Kolonie byly nejprve resuspendovány v 50 µl H₂O, z nichž bylo pro testování na přítomnost plazmidu odebráno 10 µl. Střídáním teplot bylo dosaženo lyze buněk a uvolnění DNA, která byla použita v kontrolní PCR. PCR produkt byl analyzován pomocí horizontální gelové elektroforézy. Kolonie, které se díky kontrolní PCR ukázaly jako pozitivní, byly kultivovány na třepačce (220 RPM) v 5 ml LB média s přidáním antibiotikem při 37 °C.

Plazmid byl z narostlých buněk izolován pomocí High-Speed Plasmid Mini Kit (Geneaid) do 50 µl H₂O. Výsledná koncentrace DNA ve vzorku byla stanovena za pomoci přístroje NanoDrop 2000 (Thermo Fisher Scientific). Přítomnost námi vkládaného genu byla ověřena pomocí kontrolní restrikční reakce. Produkt reakce byl analyzován pomocí horizontální gelové elektroforézy. Pozitivní výsledky byly následně zaslány do sekvenční laboratoře k osekvenování pro nalezení případných mutací.

Jelikož pro následnou transfekci trofozoitů *G. intestinalis* bylo nutné větší množství DNA, byly bakterie kultivovány v 75ml LB média s příslušným antibiotikem přes noc na třepačce (220 RPM) při teplotě 37 °C. Plazmidy byly druhý den izolovány pomocí Wizard® Plus

Midipreps DNA purification System (Promega). Koncentraci DNA ve výsledném vzorku jsme stanovili na NanoDrop 2000.

4.8.4 Transfekce *G. intestinalis*

Do trofozoitů *G. intestinalis* byly vneseny dva námi vytvořené konstrukty (GiTom40 + pTG, GiTom40 (x β -signal) + pTG) a jeden konstrukt poskytnutý Mgr. Alžbětou Motyčkovou, který se skládal z plazmidu pONDRA a malého Tim proteinu *G. intestinalis*. Trofozoiti byli kultivováni v 50ml kultivačních lahvích v upraveném TYI-S-33 médiu. Před transfekcí bylo toto médium nahrazeno vychlazeným PBS. Buňky byly inkubovány na ledu po dobu 20 minut, následně byly přeneseny do centrifugačních zkumavek a při otáčkách 1200 \times g za teploty 4 °C stáčený po dobu 10 minut. Pelety byly resuspendovány v 900 μ l TYI-S-33 média a rozděleny rovnoměrně do tří 4 mm elektroporačních kyvet. Do kyvet bylo následně přidáno 50 μ g konstruktů a pomocí pipety byly vzorky promíchány. Vzorky se poté inkubovaly 10 minut na ledu. Pro elektroporaci byl použit přístroj Gene Pulser Xcell (Biorad) a program Exponential Protocol (350 V, 1000 μ F, 750 Ω). Elektroporované vzorky byly následně inkubovány na ledu po dobu 5 minut. Vzorky byly nakonec přidány do 10 ml kultivačních zkumavek s TYI-S-33 kultivačním médiem a nechány kultivovat přes noc při teplotě 37 °C. Druhý den bylo vyměněno médium a přidána selekční antibiotika puromycin pro buňky nesoucí pTG a G418 pro buňky nesoucí pONDRA. Stav kultur byl kontrolován pomocí mikroskopu a každý 3–4 den bylo měněno médium s příslušným antibiotikem. Když došlo k vytvoření souvislé vrstvy buněk na stěnách zkumavky, byla kultura přeočkovávána.

4.9 Kontrola exprese a lokalizace

4.9.1 Příprava vzorků z celkového lyzátu *G. intestinalis*

Z kultivačních zkumavek byly Pasteurovou pipetou odstraněny mrtvé buňky. Zbytek kultury byl poté inkubován po dobu 15 minut na ledu a následně stáčen při otáčkách 1200 \times g za teploty 4 °C po dobu 10 minut. Pelety byly resuspendovány v 50 μ l 1 \times koncentrovaného vzorkového pufru. Vzorky byly inkubovány při teplotě 95 °C po dobu 5 minut.

Vzorkový pufr

1M Tris	2,4ml
10% SDS	2ml
Glycerol	2,5ml
dH ₂ O	1,6ml
0,2% bromfenolová modř	1ml
β -mercaptoethanol	0,5ml

4.9.2 SDS-PAGE

Při přípravě gelů pro vertikální elektroforézu byla brána v potaz velikost námi hledaných proteinů. Pro Tom40 tak byl používán 12% polyakrylamidový gel, zatímco pro malý Tim byl použit 15%. Jako pufr byl použit TGS (Tris-Glycin-SDS, Biorad). Napětí elektroforézy bylo regulováno během separace proteinů, v zaostřovací části gelu bylo použito napětí 90 V, po přechodu vzorků do rozdělovací části gelu bylo napětí zvýšeno na 180 V.

4.9.3 Barvení Coomassie Brilliant Blue

Hladina exprese byla zjišťována pomocí Coomassie Brilliant Blue barvení gelů z vertikální elektroforézy. Gely byly barveny přes noc na rotátoru při 15 RPM. Druhý den byly gely promývány odbarvovacím roztokem, dokud nebylo odstraněno barvivo nenavázané na proteiny.

Coomassie brilliant blue obarvovací roztok

CBB (Serva)	200mg
Denaturovaný etanol	225ml
H ₂ O	225ml
Kyselina octová	50ml

Odbarvovací roztok

Denaturovaný etanol	250ml
dH ₂ O	650ml
Kyselina octová	100ml

4.9.4 Western blot

Pro detekci specifických proteinů byla využita metoda western blot. Nejprve byly připraveny filtrační papíry a nitrocelulózkové membrány velikosti odpovídající gelu. Filtrační papíry a membrány byly promyty v blotovacím pufru. Do blotovacího přístroje byl poté vložena vrstva filtračního papíru, vrstva tvořená membránou, vrstva tvořená gelem a poté opět vrstva tvořená filtračním papírem. Proteiny byly následně pomocí blotovacího přístroje přeneseny na membránu (1,5 mA/cm², 1 hodina). Pro kontrolu byly membrány nejprve nabarveny pomocí Ponceau (0,5% Ponceau S (Merck), 1% kyselina octová).

Membrána byla poté blokována v 5% roztoku nízkotučného mléka v PBS s 0,1% Tween 20 (Sigma). Blokování probíhalo přes noc při teplotě 4 °C. Druhý den byla membrána inkubována ve stejném roztoku s přidanou primární protilátkou po dobu 1 hodiny na třepačce. Následně byla membrána třikrát promývána roztokem po dobu 10 minut. Následně byla membrána přesunuta do roztoku se sekundární protilátkou. Inkubace opět probíhala po dobu 1 hodiny a

následovalo třikrát promývání v PBS po dobu 10 minut. Výsledky byly detekovány pomocí přístroje Amersham imager 600 (GE Life Sciences).

Všechny námi použité konstrukty obsahovaly HA-tag. Jako primární protilátka tak byla použita krysí monoklonální anti-HA v doporučeném ředění 1:1000. Jako sekundární protilátka byla použita anti-krysí protilátka konjugovaná s křenovou peroxidázou v doporučeném ředění 1:2000.

Blotovací pufr

Tris	3g
Glycine	14,4g
SDS	1g
Methanol	200ml
H ₂ O	700ml

4.9.5 Fluorescenční mikroskopie

Kultivační zkumavka s kulturou byla inkubována na ledě po dobu 15 minut. Po odsednutí buněk ze stěn zkumavky bylo odebráno 1,5 ml kultury a prostřednictvím injekční jehly byla odebraná část kultury vstříknuta do připravené anaerobní komůrky. Komůrka skládající se ze dvou podložních skel byla kultivována při teplotě 37 °C po dobu 15 minut. Následně byla komůrka otočena a ponechána k inkubaci na dalších 10 minut. Komůrka byla poté rozložena a podložní skla se nechala částečně oschnout. Preparáty byly fixovány po dobu 5 minut v metanolu s teplotou -20 °C a následně permeabilizovány po dobu 5 minut v acetonu s teplotou -20 °C. Preparáty se opět nechaly oschnout.

Následně byly preparáty přendány do vlhké komůrky. Na každý preparát bylo rovnoměrně nanášeno 700 µl blokovacího roztoku. Po uplynutí 1 hodiny byl blokovací roztok slit a bylo přidáno 300 µl blokovacího roztoku společně s primární protilátkou. Inkubace opět probíhala po dobu 1 hodiny. Roztok byl poté opět slit a preparáty jsme třikrát promývaly nanášením 500 µl PBS pufru na dobu 10 minut. Na preparáty byl poté přidán 1 ml blokovacího roztoku s přidanou sekundární protilátkou. Preparáty byly inkubovány po dobu 1 hodiny bez přístupu světla, aby nedošlo k znehodnocení protilátky. Preparáty byly po inkubaci opět třikrát promývány PBS puftrem pokaždé po dobu 10 minut a opět s minimálním kontaktem se světlem. Na závěr je preparát přidáno montovací médium VectaShield s přidanou barvou DAPI, které bylo překryto krycím sklíčkem. Sklíčko bylo zafixováno na podložním skle lakem na nehty.

Jako primární protilátky pro značení námi použitých konstruktů byla použita myší anti-HA protilátka v doporučeném ředění 1:1000. Pro značení mitosomů byla použita králíčí polyklonální protilátka proti proteinu GL50803_9296 vytvořená v naší laboratoři z králíka

v doporučeném ředění 1:2000. Jako sekundární protilátky byly použity protilátky proti myším IgG konjugované s Alexa 488 v doporučeném ředění 1:1000 a protilátka proti králičím IgG konjugovaná s Alexa 594 se stejným ředěním.

Blokovací pufr

Želatina (Sigma)	250mg
PBS	30ml
BSA (Bovine serum albumine, Sigma)	250mg
Tween 20 (Sigma)	50 μ l

5. Výsledky

5.1 Distribuce proteinů TOM komplexu

Jak bylo zmíněno v kapitole 2.6, u TOM komplexu došlo v průběhu evoluce k četným redukcím/modifikacím. Rozhodli jsme se proto pro ilustrativní analýzu vybraných zástupců eukaryotických skupin, u nichž je k dispozici kompletní proteom. Zatímco některé homology bylo možné identifikovat pomocí BLAST, pro jiné bylo nutné vytvořit alignmenty a využít citlivější metody HMMER.

Vycházeli jsme z podjednotek TOM komplexu známých pro *S. cerevisiae*, jejichž homology jsou velmi dobře popsány i u člověka (**Tab. 3**). Identifikovat se u všech devatenácti zástupců podařilo pouze Tom40, kdy až na výjimky stačilo použít BLAST. U šesti organismů pak byl Tom40 jedinou identifikovanou podjednotkou. Nejvíce homologů jsme identifikovali v superskupině Opisthokonta, kde jsme výraznější ztráty zaznamenali pouze u mikrosporidie *Encephalitozoon cuniculi*. V dalších skupinách byl výskyt mozaikovitý. V superskupinách Amoebozoa, SAR a Archaeplastida se nám u některých zástupců podařilo navíc identifikovat Tom7. V případě Tom22 jsme jeho homology identifikovali kromě opisthokont také u SAR a Archaeplastida. U superskupiny Excavata jsme byli schopni identifikovat pouze homology Tom40. Receptorovou podjednotku Tom20 a Tom6 se nám podařilo identifikovat pouze u opisthokontních organismů, kdežto Tom5 jsme identifikovali i u *Arabidopsis thaliana*. Tom70 se opět vyskytoval v superskupině Opisthokonta, ale navíc se nám homolog podařilo identifikovat i u *Blastocystis hominis*. Samotná identifikace Tom70 se však ukázala jako problematická. Obsahuje totiž tetratrikopeptidové (TPR) domény (PF00515 a PF13181), které jsou obsaženy i v řadě dalších proteinů, což vedlo k velkému množství falešně pozitivních výsledků zejména při použití HMMER. Z těchto důvodů bylo obtížné určit velmi divergentní homology Tom70.

Naše výsledky jsme porovnávali s literaturou a narazili jsme tak na publikované homology pro Tom20 (Q54Y27) a Tom70 (Q75JW7) identifikované u *Dictyostelium discoideum* (Buczek et al. 2016), které se neobjevily v našich výsledcích. Naše opětovná analýza pomocí BLAST nicméně ukázala, že Q54Y27 vykazuje spíše homologii s peptidylprolyl izomerázou (sekvenční podobnost 71.3% s peptidyl izomerázou *Dictyostelium purpureum*), přičemž navíc v prvních pěti stech hitech není v současné době žádný anotovaný homolog pro Tom20. U proteinu Q75JW7 je situace komplikovanější. Obsahuje TPR domény stejně jako Tom70, ale to samo o sobě nemůžeme považovat za dostatečný důkaz pro homologii. Hity s nejlepším skóre jsou při použití BLAST neanotované proteiny, které také obsahují TRP domény. Prvním anotovaným hitem je poté protein fosfatidylethanolamín N-methyltransferáza améby *Cavendishia fasciculata* se sekvenční podobností 46.4%. Jelikož v prvních pěti stech hitech není žádný anotovaný homolog Tom70, usuzujeme, že se o homolog nejedná.

Patrný je rozdíl v přítomnosti/absenci jednotlivých komponent mezi volně žijícími a parazitickými organismy (**více rozvinuto v kapitole 5.5**). Největší kontrast vidíme při srovnání *E. cuniculi* s ostatními organismy ze superskupiny Opisthokonta. Zároveň pět z šesti organismů, u nichž se nám podařilo identifikovat pouze Tom40, je parazitických. Avšak naopak u parazitické *B. hominis* se nám v rámci superskupiny SAR podařilo identifikovat nejvíce podjednotek.

Tabulka 3: Tabulka ukazuje zastoupení podjednotek TOM komplexu u 19 vybraných zástupců eukaryot. Červeně jsou označeny parazitické organismy. Jediná podjednotka společná všem vybraným eukaryotům je Tom40. Dále se nejčastěji vyskytují podjednotky Tom22 a Tom7, což odpovídá závěrům předchozích studií (Mačašev et al. 2004).

		Tom40	Tom22	Tom20	Tom7	Tom6	Tom5	Tom70
Opisthokonta	Homo sapiens	+	+	+	+	+	+	+
	Drosophila melanogaster	+	+	+	+	+	+	+
	Caenorhabditis elegans	+	+	+	+			+
	Saccharomyces cerevisiae	+	+	+	+	+	+	+
	Neurospora crassa	+	+	+	+	+		+
Archaeplastida	<i>Encephalitozoon cuniculi</i>	+						+
	Arabidopsis thaliana	+	+		+		+	
	Cyanidoschyzon merolae	+	+					
Amoebozoa	Chlamydomonas reinhardtii	+			+			
	Dictyostelium discoideum	+			+			
SAR	<i>Entamoeba histolytica</i>	+						
	<i>Blastocystis hominis</i>	+	+		+			+
	Fistulifera solaris	+	+					
	<i>Plasmodium falciparum</i>	+	+		+			
Excavata	<i>Cryptosporidium parvum</i>	+						
	<i>Giardia intestinalis</i>	+						
	<i>Trichomonas vaginalis</i>	+						
	<i>Trypanosoma brucei</i>	+						
	<i>Naegleria gruberi</i>	+						

5.2 Distribuce Sam50

Dle současného modelu biogeneze mitochondriálních membránových β -barelů je pro jejich vložení do membrány nutná funkce Sam50. Jelikož z předchozí tabulky vyplývá, že Tom40 je společný všem eukaryotům, měli bychom u všech zákonitě nalézt i Sam50. To se skutečně podařilo u všech organismů zahrnutých v našem datasetu s výjimkou *G. intestinalis*. Navíc jsme v případě *G. intestinalis* otestovali různě sestavené HMM profily pro zvýšení citlivosti vyhledávací analýzy. Zkoušeli jsme jednak vytvořit vyvážený alignment, který by zachytil konzervované úseky, ale i alignment čistě ze sekvencí Sam50 exkavátních organismů, který by umožnil vyhledávat i podle liniově specifických oblastí. Žádný se však neukázal jako dostatečný pro detekci homologu Sam50 u *G. intestinalis*. Jako další možnost jsme proto jednotlivě analyzovali všechny proteiny *G. intestinalis* z predikovaného proteomu pomocí HHpred proti Pfam databázi. Následně jsme vytvořili skript procházející jednotlivé výsledky, žádný pozitivní ale nebyl zaznamenán.

5.2.1 Homolog Sam50 *G. intestinalis*

Rozhodli jsme se také využít specializovaných programů pro predikci transmembránových β -barelů. Otestovali jsme tři z nich (**Tab. 4**). Pro testování jsme použili extrémně divergované homology Sam50 a Tom40. Jako nejcitlivější se ukázal program PRED-TMBB2, kterým jsme provedli predikci pro všechny proteiny *G. intestinalis*.

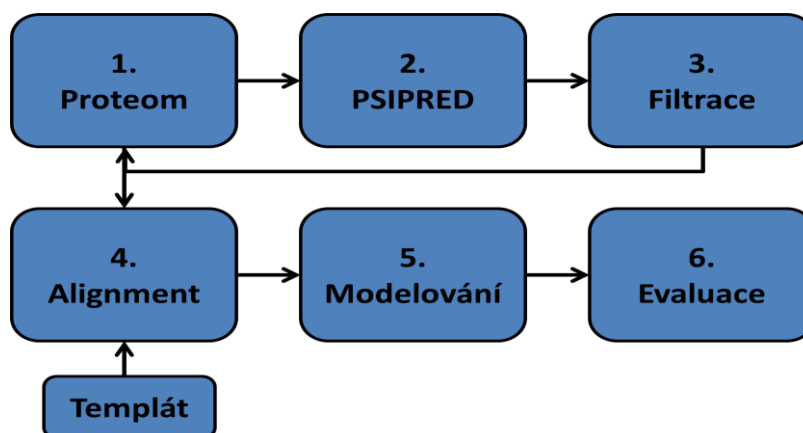
Tabulka 4: Srovnání predikčních programů na homologích Sam50 *Toxoplasma gondii* a *Trichomonas vaginalis* a Tom40 *Giardia muris*, *Plasmodium falciparum* a *Trypanosoma brucei*. Číslo udává počet predikovaných transmembránových β -listů. Předpokládá se, že Sam50 se skládá ze 16 a Tom40 z 19 β -listů (Höhr et al. 2018; Lackey et al. 2014)

	tgSam50	tvSam50	gmTom40	pfTom40	tbTom40
Boctopus2	14	0	4	8	0
PRED-TMBB	2	6	4	0	0
PRED-TMBB2	16	14	14	14	10

Celkem bylo predikováno 104 proteinů s 10-16 transmembránovými β -listy. Porovnání s našimi výsledky z HHpred však odhalilo, že se s výjimkou Tom40 pravděpodobně jedná o falešně pozitivní výsledky. Program tak patrně není optimalizován pro predikce u exkavátních organismů.

3D struktura proteinů je obecně lépe konzervována než jejich primární sekvence (Illergård et al. 2009). Na základě tohoto předpokladu a díky selhání použitých metod pracujících s primárními sekvencemi, jsme se rozhodli zaměřit na strukturní data a nalézt v proteomu *G.*

intestinalis takový protein, který by svou strukturou odpovídal Sam50. Vytvořili jsme proto následující bioinformatickou pipeline (**Obr. 7**), kdy vstupní informací je soubor ve formátu FASTA obsahující primární sekvenci a výstupem je skóre hodnotící její strukturní model.



Obrázek 7: Zjednodušená pipeline pro hledání struktury odpovídající Sam50 u *G. intestinalis*.

Proteom je nejprve rozdělen na jednotlivé sekvence (1), PSIPRED predikuje jejich sekundární strukturu (2), výsledky jsou podrobeny filtraci (3), čímž získáme seznam identifikátorů proteinů, podle nichž jsou poté z proteomu vybrány sekvence k alignování s templátem (4), na základě alignmentů jsou vytvořeny 3D modely (5), jež jsou na závěr podrobeny evaluačnímu procesu (6).

U všech sekvencí PSIPRED nejprve predikuje sekundární strukturu. Pro každou pozici je pak ve výstupu predikce zaznamenána jedna ze tří hodnot (H pro α -helix, E pro β -list nebo C pro ostatní), a pravděpodobnost s jakou pozice této hodnoty skutečně nabývá. Údaje o sekundární struktuře jsou následně extrahovány do souboru, který obsahuje pouze řetězec hodnot označujících sekundární strukturu a v názvu identifikátor proteinu.

Tyto soubory procházejí filtrací, zda obsahují dostatečné množství β -listů. Filtr má dva parametry, jednak minimální počet β -listů a také minimální počet nepřerušovaných pozic nabývajících hodnoty E, které tvoří jeden β -list. Počítali jsme s tím, že případný homolog Sam50 by byl jistě velmi divergentní a přesnost PSIPRED by nemusela být optimální, proto jsme parametry filtru nastavili poměrně volně, minimální počet pozic v řadě na pět a minimální počet skládaných listů na 10.

Primární sekvence prošlých proteinů jsou z FASTA formátu konvertovány na tzv. PIR formát, který je využíván programem Modeller. Jednotlivé sekvence jsou poté alignovány s templátem pro homologní modelování. Jako templát jsme zvolili BamA protein ze *Salmonella typhimurium* (PDB identifikátor: 5OR1), jehož struktura je dostupná v databázi PDB. Tento model neobsahuje POTRA domény, ale pouze Bac_surface_Ag doménu, jelikož modelování POTRA domény se ukázalo jako problematické. Jak bylo zmíněno v kapitole 2.2 bakteriální homology, mají POTRA domén oproti eukaryotům více, což vedlo k predikci modelů s nízkou

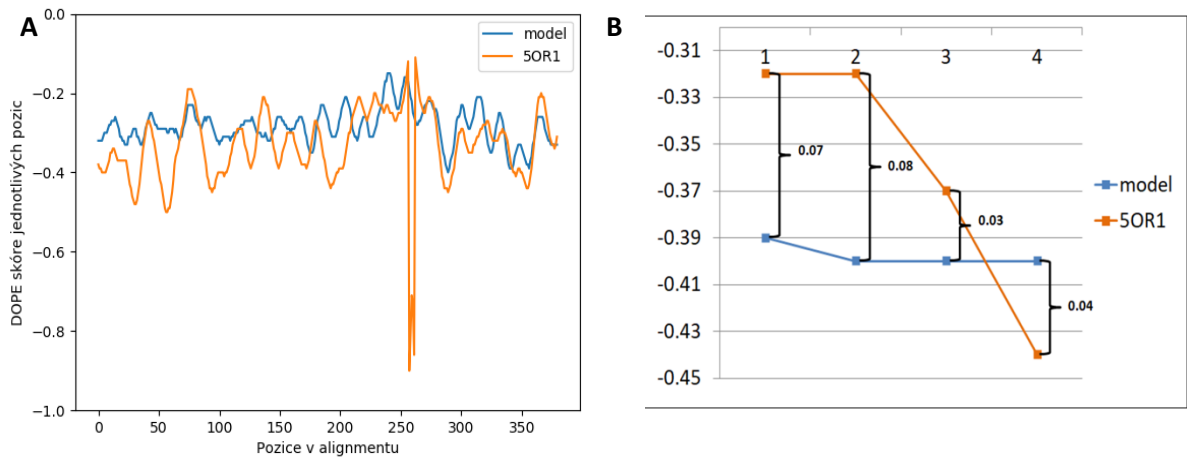
věrohodností. Doména Bac_surface_Ag je navíc více konzervována a pro naši práci tedy vhodnější.

Pro alignment jsme využili funkci Modelleru align2d(), ta na rozdíl od běžných algoritmů pro tvorbu sekvenčního alignmentu bere v potaz i strukturní údaje. Na základě alignmentu jsme poté pomocí Modelleru vytvořili homologní 3D modely ve formátu pdb.

Posledním krokem bylo hodnocení kvality modelů. Zde jsme používali kombinaci více metod. Standardní metodou pro hodnocení modelů je tzv. DOPE metoda (Discrete optimized protein energy), jejímž výstupem je DOPE skóre. Tato metoda je však vhodná pro srovnávání modelů vzniklých na základě stejného templátu. Dle Modelleru je na komparativní hodnocení modelů vzniklých na základě různých sekvencí doporučeno použít Z-skóre, které lze z DOPE skóre vypočítat (Sali and Shen 2006).

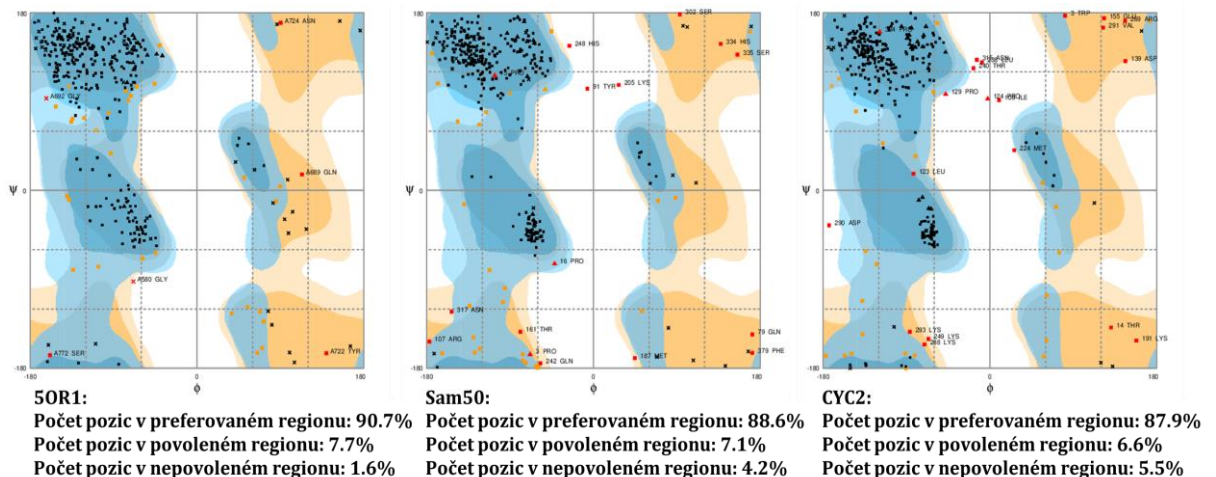
Modeller má v sobě implementovanou funkci, která výpočet Z-skóre umožňuje, zjistili jsme však, že tato metoda dává výrazně lepší skóre modelům využívající jako templát malý protein a naopak velké proteiny jsou značně znevýhodněny, použití pouze této metody hodnocení by tedy nebylo optimální.

V rámci Modelleru je také možné získat DOPE skóre pro každou aminokyselinovou pozici. Následně je možné vynést dvě křivky, jednu pro model a druhou pro templát, čím více se křivky překrývají, tím je model věrohodnější. Úseky, kde se křivky naopak výrazně rozcházejí, pravděpodobně neodpovídají reálnému stavu. Jelikož by kontrola velkého množství takovýchto grafů byla komplikovaná, vyvinuli jsme na základě DOPE skóre vlastní skórovací metodu, tzv. profilové skóre. Pro každou pozici máme dva body dané hodnotami DOPE skóre templátu a modelu, vzdálenost mezi nimi nám dá dílčí profilové skóre (**Obr. 8**). Pro každou pozici je takto změřena odchylka modelu od templátu a nakonec jsou všechny tyto odchylky sečteny ve výsledné skóre. Tato metoda se ukázala lepší než Z-skóre pro proteiny podobné velikosti jako je templát, u velikostně výrazně odlišných proteinů je však nevhodná.



Obrázek 8: Vykreslení DOPE skóre pro jednotlivé pozice modelu a templátu. Ideální by byl překryv obou křivek, čím větší je v jednotlivých bodech odchylka křivky modelu od křivky templátu, tím je model v daném úseku nevěrohodnější. Graf A ukazuje DOPE skóre pro všechny pozice, graf B ukazuje detail s označenými odchylkami, které v součtu dávají profilové skóre.

Na základě výše uvedených metod jsou poté všechny modely seřazeny od nejlepšího k nejhoršímu. Modely jsme poté ještě posuzovali podle kvality alignmentu na jehož základě byl model vybudován nebo podle Ramachandranových diagramů (**Obr. 9**).

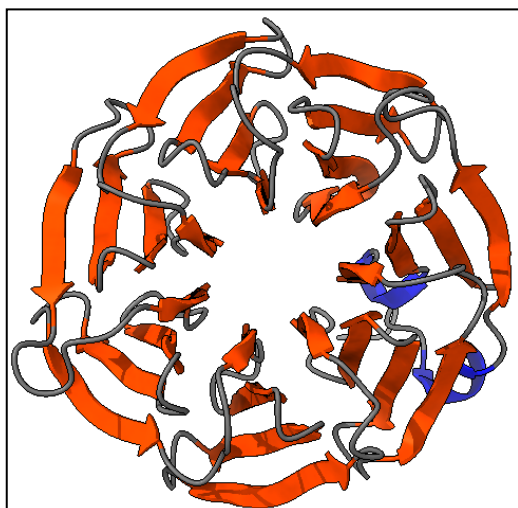


Obrázek 9: Ukázka hodnocení pomocí Ramachandranových diagramů. Tmavě modře je označen preferovaný region, světle modře povolené regiony a oranžově výjimečně povolené regiony. Vygenerovali jsme Ramachandranovy diagramy pro použitý templát, model kvasinkového Sam50 a model kvasinkového CYC2, který dosáhl lepšího Z-skóre než Sam50. Celkový počet aminokyselin nacházejících se v nepovolené oblasti je ale u CYC2 vyšší, což nám ukazuje, že model je nižší kvality.

Celý postup byl nejprve otestován na proteomu *S. cerevisiae* (UP000002311) s různými templáty, kde se původně výhradně používané Z-skóre ukázalo jako nedostatečné. Sam50 skončil z celkových 6049 proteinů nejlépe na dvacáté čtvrté pozici při použití templátu 5OR1. Zatímco C-konec proteinu obsahující Bac_surface_Ag doménu se nám dařilo modelovat poměrně úspěšně, namodelovat málo konzervovaný N-konec navíc obsahující POTRA doménu se ukázalo být problematické. Proto jsme se rozhodli, že budeme modelovat pouze úsek

primární sekvence z C-konce délkou odpovídající délce templátu. Výsledné Z-skóre se příliš nezlepšilo, umístilo Sam50 na osmnáctou pozici, zatímco nově použité profilové skóre umístilo Sam50 na pozici čtvrtou. Falešně pozitivní výsledky byly ve všech případech proteiny s podobným počtem β -listů jako má Sam50. Často se jednalo o proteiny obsahující doménu WD40 (PF00400), která je bohatá na β -listy (**Obr. 10**).

Postup byl následně aplikován na proteomu *G. intestinalis* (UP000018320). Ze vstupních 5137 proteinů filtrací sekundární struktury prošlo 400 proteinů, pro něž jsme vytvořili 3D modely. Primární sekvence proteinů s nejvěrohodnějšími modely jsme následně otestovali pomocí BLAST a HHpred. BLAST ve většině případů našel pouze homolog u jiné assembláže *G. intestinalis* případně *G. muris*. HHpred odhalil, že stejně jako u *S. cerevisiae*, většina nejlepších výsledků obsahuje WD40 doménu. Ostatní proteiny s nejlepšími modely obsahovaly různé domény s výrazným podílem β -listů. Žádný z modelů však nebyl dostatečně věrohodný, abychom mohli tvrdit, že se jedná o homolog Sam50.



Obrázek 10: 3D model WD40 domény. Oranžově jsou označeny β -listy, modře α -helixy. Struktura 4LG8 byla publikována v (Zhang et al. 2017)

5.2.2 Identifikace homologu Sam50 u blízce příbuzných *G. intestinalis*

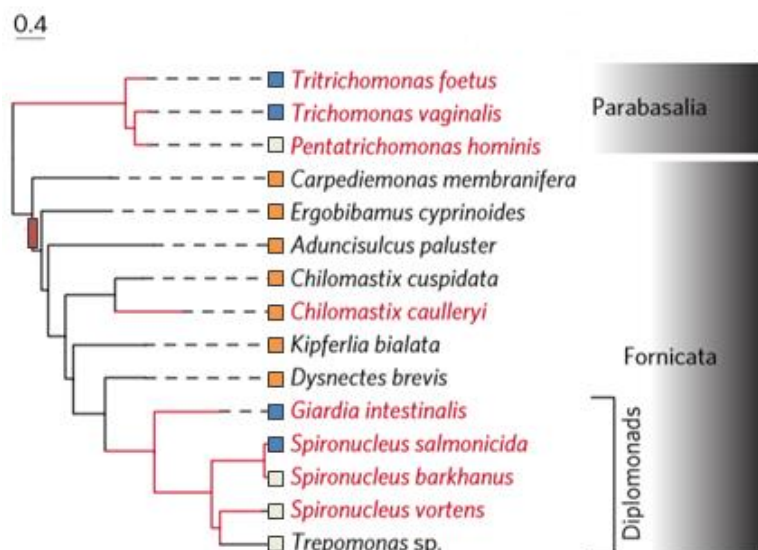
Při snaze vytvořit vhodný alignment pro identifikaci homologu Sam50 *G. intestinalis* jsme se mimo jiné pokoušeli najít homolog Sam50 v evolučně blízkých organismech. Pro skupinu Metamonada bylo dosud publikováno 12 kompletních proteomů. Homolog Sam50 se nám podařilo identifikovat pouze v proteomu *T. vaginalis* (UP000001542) a *Tritrichomonas foetus* (UP000179807), přičemž oba dva homology vykazují vysokou míru divergence oproti ostatním Sam50 (**Tab. 5**).

Tabulka 5: Tabulka obsahuje sekvenční podobnost homologů Sam50 *T. vaginalis* a *T. foetus* s homology nalezenými v rámci zjišťování distribuce Sam50 napříč eukaryoty. Přestože je sekvenční identita velmi nízká, HHpred identifikoval v obou případech s vysokou pravděpodobností Bac_surface_Ag i POTRA doménu.

	<i>T. vaginalis</i>	<i>T. foetus</i>	<i>H. sapiens</i>	<i>D. melanogaster</i>	<i>C. elegans</i>	<i>S. cerevisiae</i>	<i>N. crassa</i>	<i>E. cuniculi</i>	<i>A. thaliana</i>	<i>C. merolae</i>	<i>C. reinhardtii</i>	<i>D. discoideum</i>	<i>E. histolytica</i>	<i>B. hominis</i>	<i>F. solaris</i>	<i>P. falciparum</i>	<i>C. parvum</i>	<i>T. brucei</i>	<i>N. gruberi</i>
Trichomonas vaginalis	100	30.8	16.9	17.2	18.0	14.8	15.1	11.4	14.2	16.1	17.7	14.0	14.4	8.6	15.3	13.4	13.2	11.8	15.9
Trichomonas foetus	30.8	100	18.5	12.6	13.4	14.8	16.1	12.2	14.5	14.0	14.0	13.7	12.8	7.0	13.7	15.3	16.1	14.5	12.9

K dispozici jsme také měli transkriptomická data sedmi organismů (**Obr. 11**) ze skupiny Fornicata do níž se řadí i *G. intestinalis*. Pomocí BLAST a HMMER jsme v nich hledali možný homolog Sam50.

Žádný pozitivní výsledek jsme nezaznamenali pouze u *Chilomastix caulleryi* a *Kipferlia bialata*, nicméně BLAST pozitivních výsledků od ostatních organismů prokázal, že se pravděpodobně ve většině případů jedná o bakteriální kontaminace. S výjimkou jednoho proteinu od *Carpodomonas membranifera* se ve výsledcích hledání BLAST nacházely pouze bakteriální homology. Pro protein 2387 z datasetu *C. membranifera* však BLAST našel i řadu eukaryotických homologů. HHpred u tohoto proteinu identifikoval POTRA a Bac_surface_Ag doménu a C-konec proteinu navíc obsahuje β -signál. Na základě těchto zjištění považujeme tento protein za homolog Sam50.



Obrázek 11: Fylogenetický strom obsahující organismy ze skupiny fornicata a parabasalia. Červeně jsou označeny parazitické organismy. Barevné čtverce označují, jaká data jsme měli pro daný organismus k dispozici, pro modré jsme měli proteom, pro oranžové transkriptom a pro bílé jsme data neměli. Červený obdélník strom rozděluje na část, kde jsme schopni identifikace homologu Sam50 a na část kde už nikoli. Získáváme tak poměrně velké množství blíže příbuzných organismů s absencí homologu Sam50. Upraveno dle (Leger et al. 2017).

5.2.3 Ztráta Sam50 v dalších eukaryotických liniích

Ztráta Sam50 u *G. intestinalis* bývá v literatuře označována za unikátní (Jedelský et al. 2011; Kay et al. 2013). Rozhodli jsme se tedy zjistit, zda se nám podaří nalézt další eukaryotické organismy či celé linie bez Sam50.

Pro hledání jsme zvolili databázi UniRef100, která obsahuje všechny sekvence z databází UniProtKB a vybrané sekvence z databáze UniParc (Suzek et al. 2015). Každá sekvence má ve své anotaci identifikační číslo proteinu společně s názvem organismu, což jsme využili v naší bioinformatické pipeline. Databáze zahrnuje i sekvence pro organismy s nekompletním proteomem. Abychom částečně omezili velké množství výsledků, kdy Sam50 chybí z důvodu nekompletnosti dat pro daný organismus, hledali jsme organismy s homologem Tom40 ale chybějícím Sam50. Sekvence prokaryotických organismů, které jsou v databázi také přítomny, byly odfiltrovány díky nepřítomnosti homologu Tom40.

Vstupní informací tak jsou dva sekvenční alignmenty z nichž jsou vytvořeny HMM profily pro HMMER. Otestovali jsme několik alignmentů pro oba proteiny. Jako nejlepší se nakonec ukázaly relativně malé alignmenty vycházející z dat získaných při ilustrativní analýze distribuce podjednotek TOM a SAM komplexu (**viz kapitoly 5.1 a 5.2**), které ale obsahují rovnoměrné množství sekvencí ze všech superskupin.

V první verzi pipeline byly z výsledků HMMER přečteny identifikátory všech proteinů, ty byly následně srovnány s databází UniRef100 pro taxonomické zařazení jednotlivých eukaryotických druhů. Takto vznikly dva seznamy taxonů, z nichž se vypsal ty nacházející se pouze v seznamu pro Tom40 pozitivní.

Tato pipeline se ukázala jako funkční, ale v důsledku rozsáhlosti databáze UniRef100 jako příliš pomalá pro reálné využití. Rozhodli jsme se proto rozdělit databázi na deset přibližně stejně velkých celků. Tím se výrazně urychlilo hledání pomocí HMMER a zároveň nám menší velikost databáze umožnila vytvořit v rámci Pythonu slovník, kde je pod každým identifikátorem proteinu uloženo jméno taxonu. Program tedy nemusí hledat pro každý protein v rozsáhlém textovém souboru řádek s příslušným identifikátorem, ale hledá v interním slovníku, což dobu celého procesu snižuje o desítky hodin a umožnilo nám to testovat více vstupních alignmentů. Zbytek pipeline zůstal stejný, tedy vytvoří se dvě množiny taxonů a hledáme rozdíl mezi nimi.

Celkem jsme takto získali 363 taxonů bez homologu Sam50. Ve většině případů se jedná o falešně pozitivní výsledky způsobené nekompletností proteomických dat pro daný organismus. Pro příklad se nám tak ve výsledcích objevilo čtrnáct různých druhů z rodu *Drosophila*. Zároveň se nám vyskytlo několik případů, kdy byla při anotaci Tom40 použita pouze nějaká z vyšších taxonomických kategorií, například se tak mezi výsledky dostala podtřída Aves. Jako falešně pozitivní jsme tedy brali taxony, kde známe alespoň jednoho zástupce spadajícího do podřazeného taxonu s homologem Sam50 nebo v případě, že je výsledkem druh, kde nalezneme v rámci rodu alespoň jeden organismus s homologem Sam50. Tímto způsobem se nám podařilo vyloučit 288 taxonů.

Pro dalších 61 taxonů nemáme dostatečné množství sekvencí, abychom je mohli považovat za pozitivní výsledky. Ve většině případů se jedná o druhy, pro něž je v databázi pouze malé množství proteinů, a zároveň jsou jedinými zástupci daného rodu vyskytující se v databázi.

Zbylo tak 14 taxonů, které můžeme označit za pozitivní výsledky. Čtyři z výsledků jsou různé asambláže *G. intestinalis*, dále pak příbuzné diplomonády *G. muris* a *S. salmonicida*. Mezi výsledky se také objevila bakterie *Macroccoccus caseolyticus*, při BLASTU sekvence této bakterie označené jako homolog Tom40 se však ukázalo, že se pravděpodobně jedná o falešně pozitivní výsledek a chybnou anotaci v databázi, jelikož sekvence vykazuje více jak 55% identitu s několika sekvencemi organismů ze skupiny Ascomycota.

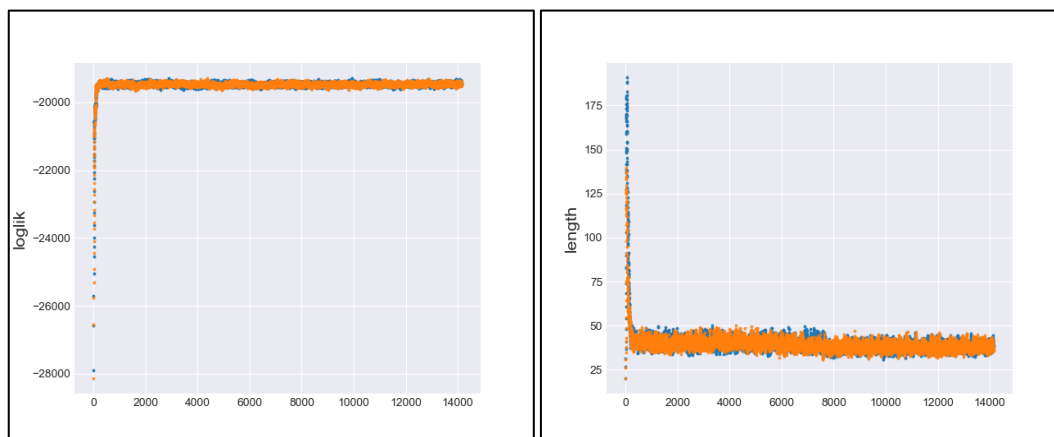
Zbývající organismy tvoří dvě mikrosporidie (*Edhazardia aedis*, *Nematocida parisii*), dva nálevníci (*Ichthyophthirius multifiliis*, *Tetrahymena thermophila*) a tři askomycetní houby (*Taphrina deformans*, *Neolecta irregularis*, *Magnaportheopsis poae*). Pro všech sedm organismů jsou k dispozici kompletní proteomy, které jsme otestovali pomocí HHpred proti databázi Pfam. Pro snadnou orientaci ve výsledcích z HHpred jsme vytvořili jednoduchý vyhledávací nástroj, kdy zadáme jméno či Pfam identifikátor domény a jsou nám vypsaný

veškeré pozitivní výsledky a jejich E-value. Žádný protein obsahující POTRA či Bac_surface_Ag doménu jsme však nezaznamenali.

Pomocí fylogenetické analýzy jsme se pokusili zjistit, zda jsou si výše zmíněné organismy v rámci své skupiny blízce příbuzné, jako je tomu u *G. intestinalis* a jí příbuzných ze skupiny Fornicata (**Obr. 13, 14, 15**). Pro analýzu jsme použili čtyři geny, které se ukázaly jako snadno identifikovatelné ve všech použitých organismech, a to 18S rRNA, α -tubulin, elongační faktor 1 a elongační faktor 2.

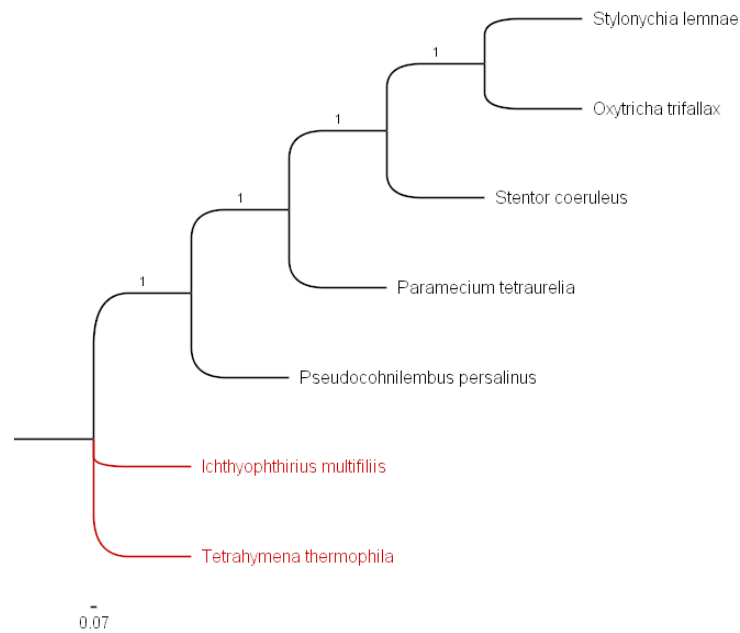
Pro celý proces jsme vytvořili jednoduchou bioinformatickou pipeline využívající programy Mafft, trimAl a BLAST. Nejprve pomocí BLAST najdeme pro každý organismus příslušné proteiny, jejich sekvence jsou následně alignovány pomocí Mafft a problematické úseky obsahující mezery ve více než 20% případů jsou odstraněny programem trimAl. Všechny čtyři alignmenty pro každý organismus jsou nakonec složeny do jedné sekvence, tzv. super genu. Odlišný přístup může být tvorba fylogenetického stromu pro každý gen zvlášť a následné vytvoření konsensus stromu, tato metoda však ve většině případů dosahuje horších výsledků (Gadagkar et al. 2005).

Pro každou skupinu jsme ve PhyloBayes nechali běžet čtyři nezávislé MCMC a pomocí programu bpcomp, který je součástí PhyloBayes sady, jsme kontrolovali, zda se nějaký řetězec nezasekl v lokálním maximu. PhyloBayes navíc produkuje soubor obsahující statistické hodnoty, z nichž se po vynesení do grafu dá usoudit, zda se MCMC již ustálil, případně jaký nastavit burn-in při tvorbě výsledného stromu (**Obr. 12**).

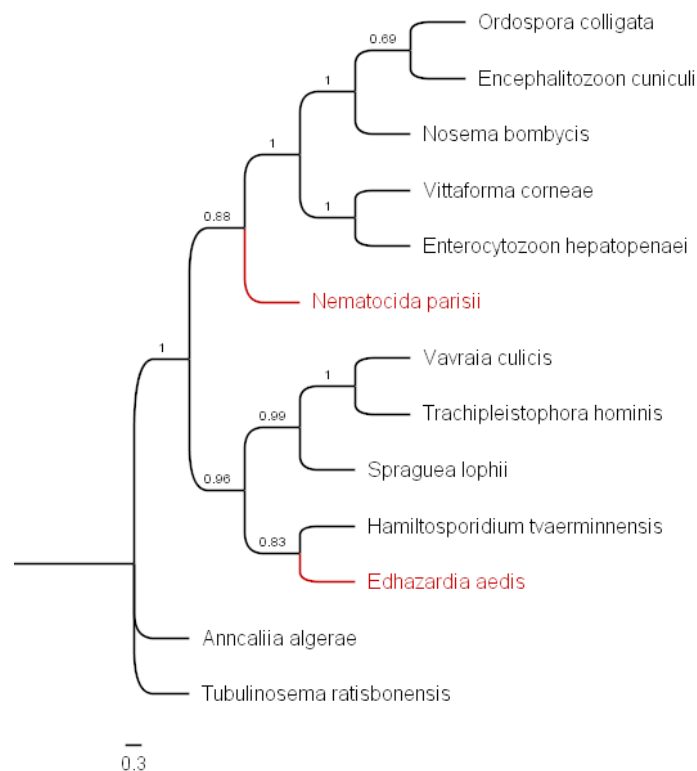


Obrázek 12: Ukázka dvou statistických hodnot vynesných do grafu pro dva nezávisle běžící MCMC ve PhyloBayes. Můžeme vidět, že se MCMC ustálí poměrně rychle a není proto potřeba nastavovat vysoký burn-in.

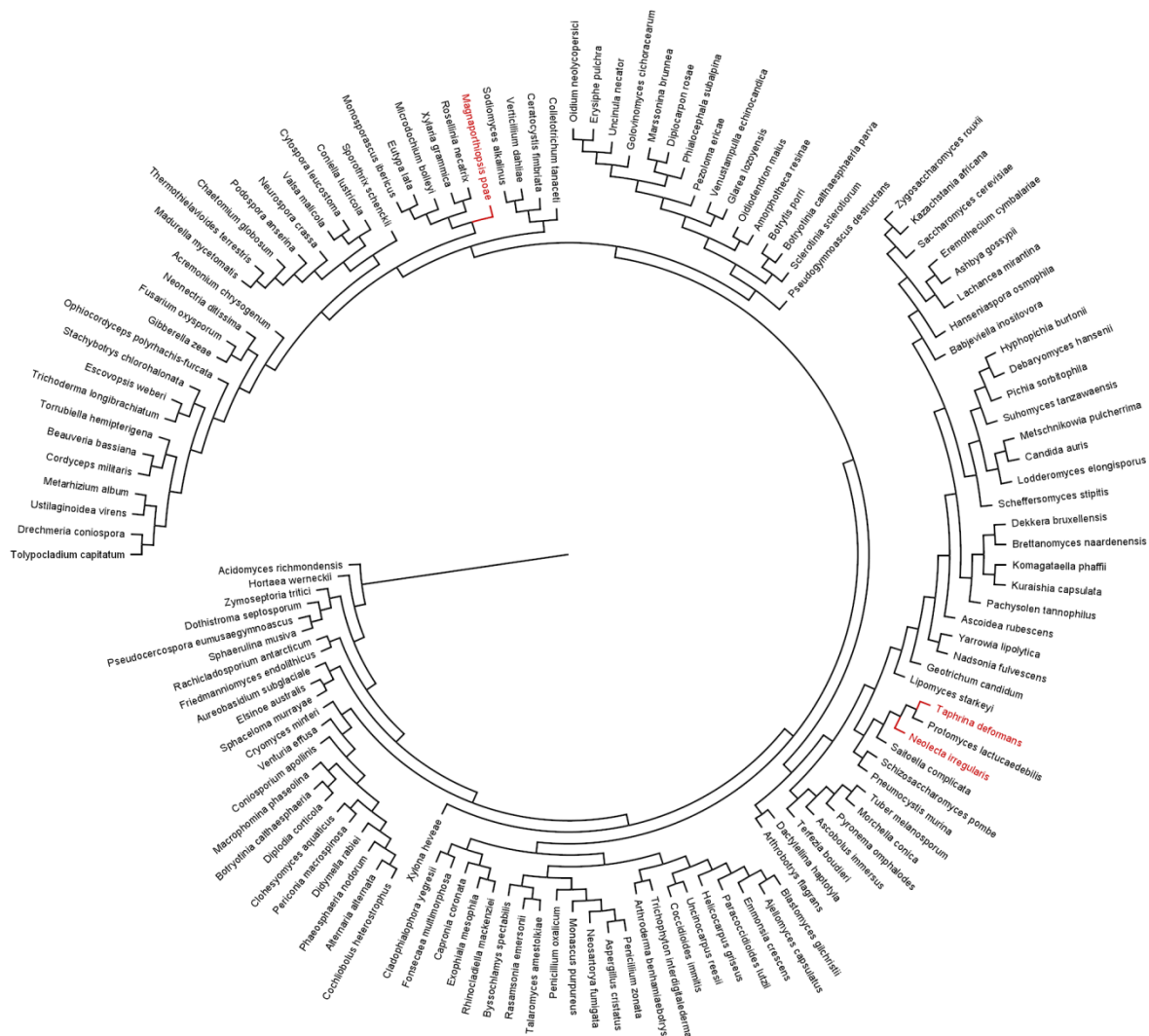
Pro Ascomycota jsme proces museli dvakrát opakovat, jelikož kontroly programem bpcomp naznačovaly zaseknutí řetězce v lokálním maximu. Při tvorbě výsledných stromů jsme nastavili burnin na 1000 a počítali jsme s každým desátým stromem. Pro vizualizaci stromů jsme použili program FigTree.



Obrázek 13: Fylogenetický strom skupiny Ciliophora skládající se ze sedmi druhů. Červeně označené druhy postrádají Sam50. Vidíme, že oba druhy tvoří v rámci skupiny oddělenou větev, nicméně poměrně nízký počet použitých druhů může zkreslovat výsledek. Obsažená čísla udávají posteriorní pravděpodobnost.



Obrázek 14: Fylogenetický strom skupiny Mikrosporidia skládající se z třinácti druhů. Červeně označené druhy postrádají Sam50. Oproti skupině Ciliophora zde druhy netvoří jednu oddělenou větev.



Obrázek 15: Fylogenetický strom skupiny Ascomycota skládající se ze 135 druhů. Červeně označené druhy postrádají Sam50. Zatímco *M. poae* se nachází v zcela odlišné části stromu, *N. irregularis* a *T. deformans* vytváří společnou větev. Jejich součástí je ale i druh *Protomyces lactucaedebilis* u něhož jsme homolog Sam50 dokázali identifikovat. Hodnoty posteriorní pravděpodobnosti jsme v rámci zvýšení přehlednosti nezobrazovali.

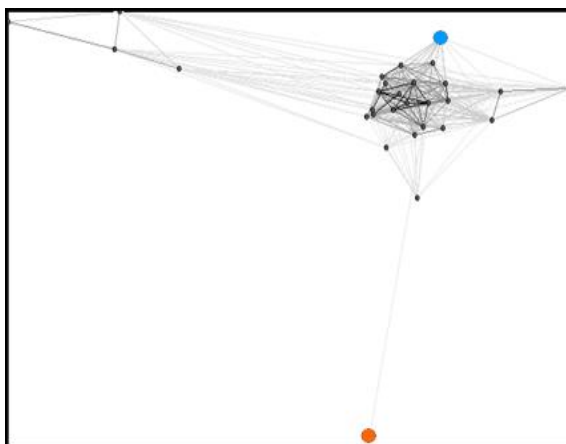
Z výsledků vyplývá, že jedinými dvěma skupinami eukaryot bez homologu Sam50 jsou výše zmíněné organismy ze skupiny Metamonada a nálevníci z řádu Hymenostomatida. Z důvodu nízkého počtu proteomických dat pro nálevníky je výsledek diskutabilní.

V ostatních případech se jednalo o z hlediska fylogenetiky osamocené organismy. U všech tří askomycetních hub se nám podařilo identifikovat alespoň některé z proteinů SAM komplexu, konkrétně homology Sam37 a Mdm10, které ukazují na přítomnost SAM komplexu.

Mikrosporidie jsou známé extrémními redukcemi genomu a zároveň mají stejně jako *G. intestinalis* mitochondrie redukované na mitosomy (Nakjang et al. 2013; Williams et al. 2002). Případná ztráta Sam50 by tak u nich byla věrohodnější než u askomycetních hub.

Dle našich zjištění také homology Sam50 vykazují u mikrosporidií vysokou míru divergence a jsou oproti většině známých homologů o cca 150 aminokyselin kratší. Všechny mikrosporidiální homology Sam50 jsme klasifikovali pomocí HHpred, který i přes vysokou míru divergence poskytoval výsledky s velmi dobrými E-value. Pro obě mikrosporidie, u nichž se nám nepodařilo identifikovat homolog Sam50, jsme pomocí HHpred otestovali celý jejich proteom, ale žádné sekvence s kýženými doménami se mezi výsledky neobjevily.

Je třeba zmínit, že samotný Tom40 *N. parisii* je velice divergentní a liší se i od ostatních mikrosporidiálních homologů Tom40 (**Obr. 16**). Tom40 *E. aedis* naproti tomu nevykazuje výraznou odlišnost od ostatních mikrosporidií. Pokud je Tom40 takto odlišný, lze předpokládat, že by i případný homolog Sam50 vykazoval vysokou míru divergence a je tedy možné, že použité metody pro identifikaci homologie nejsou dostatečně citlivé.



Obrázek 16: Klusterovací analýza třiceti mikrosporidiálních homologů Tom40. Oranžově je označen Tom40 *N. parisii*, modře Tom40 *E. aedis*. Sekvenční identita NpTom40 s ostatními homology se pohybuje ve většině případů okolo 15%. Nejvyšší sekvenční identitu (20.15%) má s homologem Tom40 *Anncaliia algerae*.

5.3 Mutace β -signálu

β -signál β -barelových membránových proteinů je definován poměrně volně (**viz kapitola 2.1**). Zajímalo nás, zda se nám i přesto podaří najít nějaké mitochondriální β -barely vybočující z této definice, případně zda bude nějaká pozice náchylnější k mutacím než ostatní. Pomocí HMMER jsme z databáze UniProtKB získali celkem 13 473 homologů Sam50, Tom40 a VDAC. Vzájemná podobnost sekvencí VDAC a Tom40 vyžadovala pro rozlišení klusterovací analýzu. Díky této analýze jsme byli schopni následně vytvořit tři datasety. Nejméně sekvencí jsme identifikovali pro Sam50, celkem 3671. Pro Tom40 jsme identifikovali 4074 sekvencí a pro VDAC 5728. Nepoměr počtu sekvencí je způsoben poměrně častým výskytem více isoformů Tom40 a VDAC u jednoho organismu.

Pro hledání β -signálu v datasetech jsme vytvořili jednoduchý program využívající regulární výraz, tedy jeden řetězec popisující množinu řetězců. V našem případě tedy všechny možné kombinace aminokyselin, které mohou tvořit β -signál. β -signál se nemusí nacházet na úplném

konci proteinu, ale může být následován ještě poměrně dlouhým úsekem aminokyselin. U Tom40 *S.cerevisiae* je to například 28 aminokyselin. Rozhodli jsme se proto prohledávat posledních 40 aminokyselin C-konce proteinu.

Všechny sekvence z datasetů pro Sam50 a VDAC obsahovaly standardní β -signál. Pro Tom40 jsme však identifikovali několik homologů s mutacemi v β -signálu (**Tab. 6**). Ve všech případech se jedná o jednobodové mutace, avšak výsledků je příliš málo, abychom mohli s jistotou tvrdit, že je nějaká pozice výrazně náchylnější k mutacím než ostatní. Jediná pozice, která se zdá být společná skutečně všem β -barelům je pozice číslo tři tvořená glycinem. Mezi výsledky se objevily dva organismy ze skupiny Metazoa, konkrétně *Sus strofa* a *Acyrtosiphon pisum*. Oba organismy však mají i druhou isoformu Tom40 se standardním β -signálem. Bodová mutace se vyskytla i u homologu Tom40 mikrosporidie *Nosema bombycis*, u níž se nám isoformu Tom40 se standardním β -signálem nalézt nepodařilo. Posledním nalezeným organismem s mutací v β -signálu u Tom40 je *T. vaginalis*, kde se nám podařilo identifikovat celkem šest homologů Tom40. Pouze jeden z nich obsahoval bodovou mutaci. Musíme zmínit, že přestože klastrovací analýza zařadila všech šest proteinů do datasetu Tom40, nemůžeme s jistotou tvrdit, že se ve všech případech skutečně jedná o homolog Tom40. Při otestování sekvencí pomocí HHpred s použitím databáze PDB jsme u jedné sekvence zaznamenali nejlepší hit pro VDAC a nikoli Tom40. U ostatních byly sice Tom40 na nejlepší pozici, ale hity pro VDAC proteiny neměly ve většině případů výrazně horší skóre. Dle literatury (Rada et al. 2011) se však jedná výhradně o homology Tom40 a rozhodli jsme se je proto v datasetu ponechat.

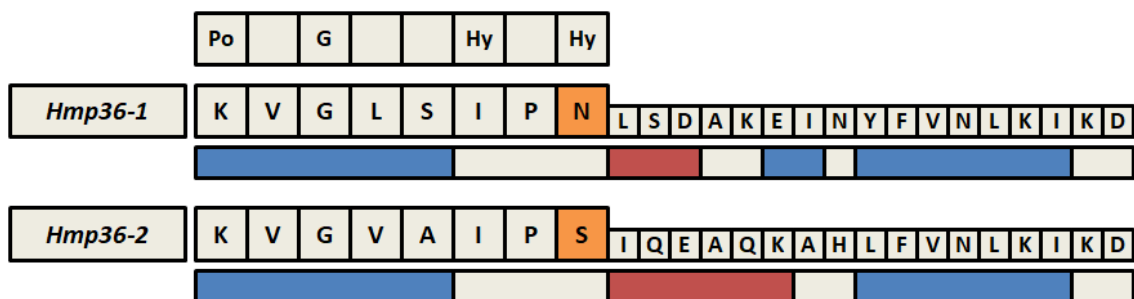
Tabulka 6: přehled nalezených nestandardních β -signálů u homologů Tom40. Oranžově jsou označeny aminokyseliny vymykající se definici β -signálu. Jelikož pouze čtyři sekvence z 13 473 testovaných nesly mutaci, navíc ve všech případech pouze jednobodovou, tak můžeme říct, že β -signál je skutečně až na naprosté výjimky konzervovaný napříč eukaryoty.

	Po		G			Hy		Hy
<i>Sus scrofa</i>	R	R	G	R	I	T	P	V
<i>Acyrtosiphon pisum</i>	M	M	G	A	G	L	T	I
<i>Nosema bombycis</i>	I	Y	G	Y	G	L	N	Y
<i>Trichomonas vaginalis</i>	K	L	G	L	G	F	Y	S

Zajímavým příkladem β -barelů jsou proteiny *T. vaginalis* Hmp35 a Hmp36 zmíněné v kapitole 2.6, které se nacházejí ve vnější hydrogenosomální membráně. V databázi UniProtKB jsme našli celkem 3 sekvence pro Hmp35 a dvě pro Hmp36. Identifikovali jsme také jeden

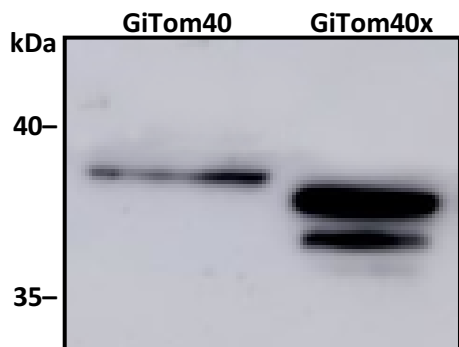
pravděpodobný homolog Hmp35 u *Tritrichomonas foetus*. Zatímco Hmp35 proteiny včetně homologu *T. foetus* mají standardní β -signál, u Hmp36 nalezneme β -signály s jednobodovou mutací následované v obou případech 17 aminokyselinami (**Tab. 7**). Při predikci sekundární struktury poloha posledního β -listu neodpovídá poloze β -signálu (Rada et al. 2011). Přesnost predikce sekundární struktury u proteinu, u něž neznáme žádné vzdálenější homology, však nemusí být optimální. Přesnost u membránových proteinů tvořených převážně β -listy je navíc obvykle nižší než u membránových α -helikálních proteinů či solubilních proteinů (Yan et al. 2015).

Tabulka 7: přehled nalezených nestandardních β -signálů u Hmp36 proteinů *T. vaginalis*. Proteiny mají 47,35% sekvenční identitu. Pod sekvencí jsou vyobrazeny výsledky predikce sekundární struktury, kdy modře jsou vyznačeny β -listy a červeně α -helixy. Vidíme, že se β -signál nenachází na posledním β -listu a zároveň se tři poslední aminokyseliny motivu ani nepodílejí na tvorbě β -listu.



5.3 Funkce β -signálu v mitochondriálních organelách postrádajících Sam50

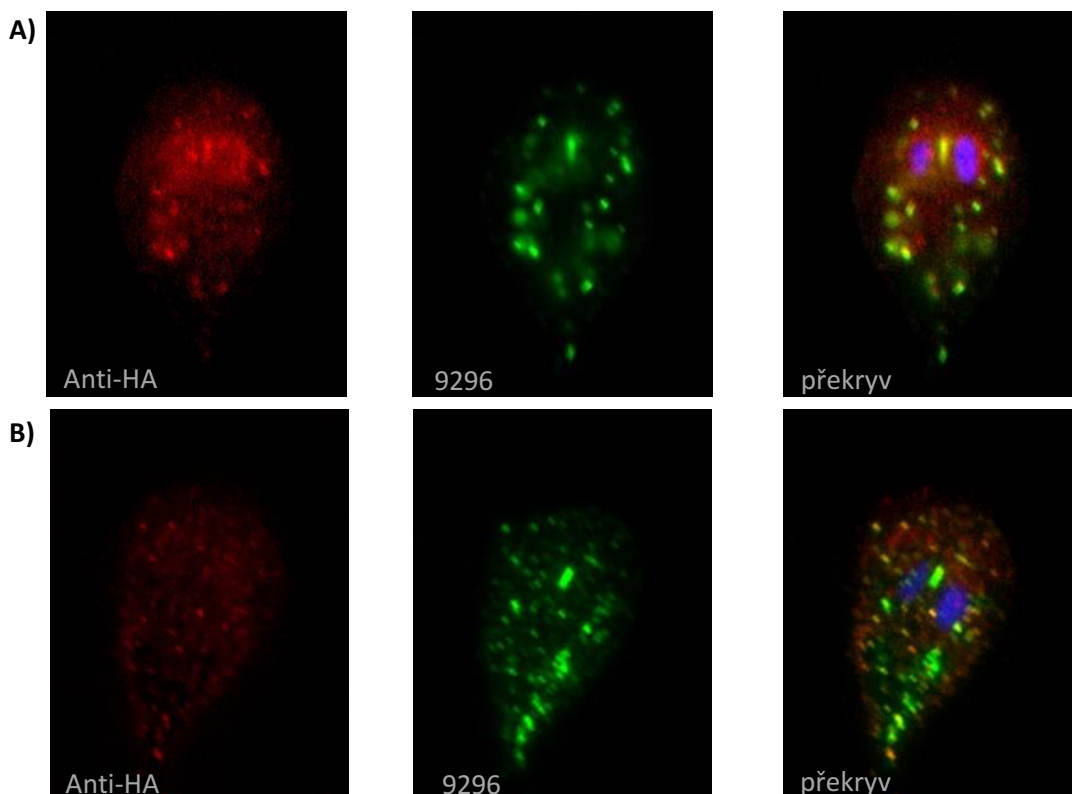
Získané výsledky pro *G. intestinalis* (absence Sam50 a přítomnost Tom40 s β -signálem) vytvářejí unikátní experimentální systém pro testování funkce signálu během směřování a skládání β -barelu v membráně. Z toho důvodu jsme vytvořili dva konstrukty nesoucí N-terminální HA-tagovaný Tom40 *G.intestinalis* s a bez β -signálu. Konstrukty jsme poté vnesli do trofozoitů *G. intestinalis* a vyseletovali stabilní transformované linie. Expresi proteinů jsme ověřili western blotem z celkového lyzátu buněk (**Obr. 17**).



Obrázek 17: Western blot z celkového lyzátu buněk. Srovnání bandů standardního Tom40 s Tom40 s odstraněným β -signálem. Po odstranění β -signálu detekujeme tři proteiny označené HA-tagem. GiTom40 zhruba odpovídá predikované molekulové hmotnosti 37,8 kDa. Predikovaná molekulová hmotnost po odstranění β -signálu je 35,4 kDa. Spodní ze tří bandů zhruba odpovídá této velikosti.

Predikovaná molekulová hmotnost standardního Tom40 na western blotu odpovídala. Po odstranění β -signálu jsme pozorovali rozpadnutí na tři menší proteiny. Avšak predikovaná molekulová hmotnost po odstranění β -signálu zhruba odpovídala nejmenšímu z bandů, nikoli největšímu, jak by šlo spíše předpokládat.

Expresce HA-tagovaných proteinů ukázala v obou případech částečnou mitosomální lokalizaci, jak je patrné z překryvu se signálem pro mitosomální marker GL50803_9296 (**Obr. 18**). V případě Tom40 bez β -signálu byla však mitosomální lokalizace méně častá a ovlivněna byla i buněčné distribuce mitosomálního markeru GL5080_9296.



Obrázek 18: Lokalizace standardního Tom40 (A) a Tom40 bez β -signálu (B) pomocí fluorescenční mikroskopie u *G. intestinalis*. Červeně je značen náš konstrukt pomocí anti-HA protilátky, mitosomy jsou značeny zeleně pomocí anti-GL5080_9296. Jádra jsou pomocí DAPI barvena modře.

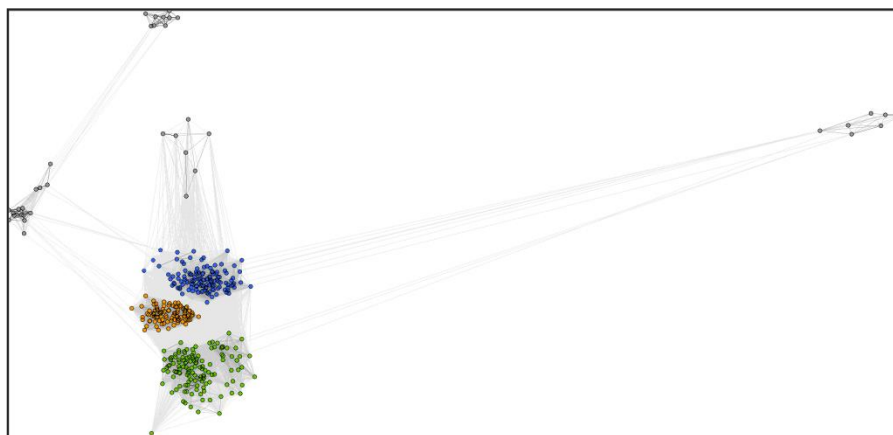
5.5 Distribuce proteinů mitochondriálních importních komplexů

Po detailní analýze distribuce komponent TOM komplexu napříč eukaryoty (**viz kapitola 5.1**), jsme se rozhodli pro vytvoření podobné analýzy ve větším měřítku s méně detailním přístupem. Namísto devatenácti reprezentativních zástupců jsme se rozhodli pro analýzu použít organismy s dostupným proteomem. Z naší analýzy i z dalších pozorování vyplývá, že jediným proteinem TOM komplexu společným všem eukaryotickým superskupinám je Tom40 (Žarský et al. 2012), proto jsme pomocí HMMER nejprve všechny proteomy otestovali na přítomnost homologu Tom40. Celkem jsme takto otestovali 2598 proteomů, z toho 1398 bylo v databázi UniProt označeno jako tzv. referenční proteom (manuálně i automaticky ověřený proteom). Mezi referenčními proteomy jsme zaznamenali pouze dva bez zjevného homologu Tom40 (*Rhizophagus irregularis* ze skupiny Mucoromycota a *Sprague lorii* ze skupiny Mikrosporidia). Z 1200 nereferenčních proteomů však 141 postrádalo zjevný homolog Tom40, přičemž řada z těchto proteomů náležela i organismům ze skupiny Metazoa, kde považujeme ztrátu Tom40 za nepravděpodobnou. Na základě tohoto testování jsme se proto rozhodli pracovat téměř výhradně s referenčními proteomy. Některé nereferenční proteomy jsme se rozhodli zařadit do analýzy z důvodu nízkého počtu dat pro danou skupinu eukaryot. Naopak některé referenční proteomy byly vyřazeny z důvodu nadbytečnosti (k dispozici jsou například tři datasety pro *G. intestinalis*). K dispozici jsme také měli dataset pro *Andalucia godoyi* z řádu *Jakobida* (Gray et al. 2020). Nakonec tak bylo pro analýzu použito celkem 797 organismů, které jsme rozdělili do 24 fylogenetických skupin.

V rámci této analýzy nás nezajímalo pouze stav TOM komplexu, ale i komplexů SAM, TIM22, TIM23, PAM, MIA a MIM. Celkem jsme tak ověřili přítomnost 31 proteinů u námi vybraných organismů, respektive jsme zjišťovali procentuální zastoupení organismů s přítomným homologem daného proteinu v rámci fylogenetické skupiny. Pro tyto účely jsme si rovněž vytvořili bioinformatickou pipeline. Pro každý protein bylo nejprve nutné vytvořit alignment, který obsahoval sekvence, co nejlépe pokrývající spektrum eukaryot. Na základě alignmentu byl vytvořen HMM profil, který HMMER použije pro hledání v jednotlivých proteomech. Proteiny získanými v prvním kole HMMER hledání je obohacen původní alignment a zbudován nový HMM profil specifický pro každou fylogenetickou skupinu. V první verzi pipeline jsme počítali ještě s dalšími koly HMMER, dokud nedošlo k ustálení počtu sekvencí v alignmentu. Avšak v případě, že se do alignmentu dostaly chybně identifikované sekvence, v každém dalším kroku se jejich počet znásoboval, což ve finále vedlo k velkému množství falešně pozitivních výsledků. Jako optimální se ukázalo jedno opakování HMMER. Po druhém HMMER hledání jsou spočteny počty druhů s homologem daného proteinu a celkové počty druhů v každé fylogenetické skupině. Z těchto hodnot je nakonec spočítáno procentuální zastoupení pozitivních organismů v každé skupině. Pro přehlednější zobrazení výsledků jsme procentuální hodnoty převedli na RGB hodnoty, které udávají barvu příslušného pole. Získáme tak teplotní mapu distribuce proteinů mitochondriálních importních komplexů (**Obr. 20**).

Během práce se vyskytlo několik problémů. Přestože výše zmíněné omezení iterativních kroků zamezilo velkému množství falešně pozitivních výsledků, u některých proteinů se úplná eliminace nepodařila. Tom70 se ukázal jako problematický při použití HMMER už při ilustrativní analýze (**viz kapitola 5.1**) a pro tuto analýzu musel být zcela vyřazen. S obdobnými problémy jsme se potýkali u proteinů Mdm10, Tim50, Sam37 a Tim54. Falešně pozitivní výsledky jsme se nejprve pokusili eliminovat pomocí klastrovací analýzy, kdy jsme dataset reprezentativních sekvencí pro každý protein obohatili sekvencemi z naší analýzy. Tato metoda se však neukázala jako vhodná, jelikož řada výsledků byla hraniční, a nedokázali jsme posoudit, zda se jedná o pozitivní či falešně pozitivní výsledek. Místo klastrovací analýzy jsme tak zvolili zpětnou kontrolu pomocí HHpred. Zatímco Sam37 a Tim54 mohli být díky tomuto postupu zařazeny do studie, u Tim50 a Mdm10 se nepodařilo přesvědčivě eliminovat falešně pozitivní výsledky. Tim50 obsahuje tzv. NIF doménu (NLI interacting factor-like phosphatase, PF03031), která se nachází i v řadě jiných proteinů. I přes nízkou podobnost primárních sekvencí se pro Mdm10 se do výsledků často dostávaly homology Tom40 a nešlo je efektivně eliminovat. Podobnost proteinů však byla ukázána již v předchozích pracích např. (Flinner et al. 2013).

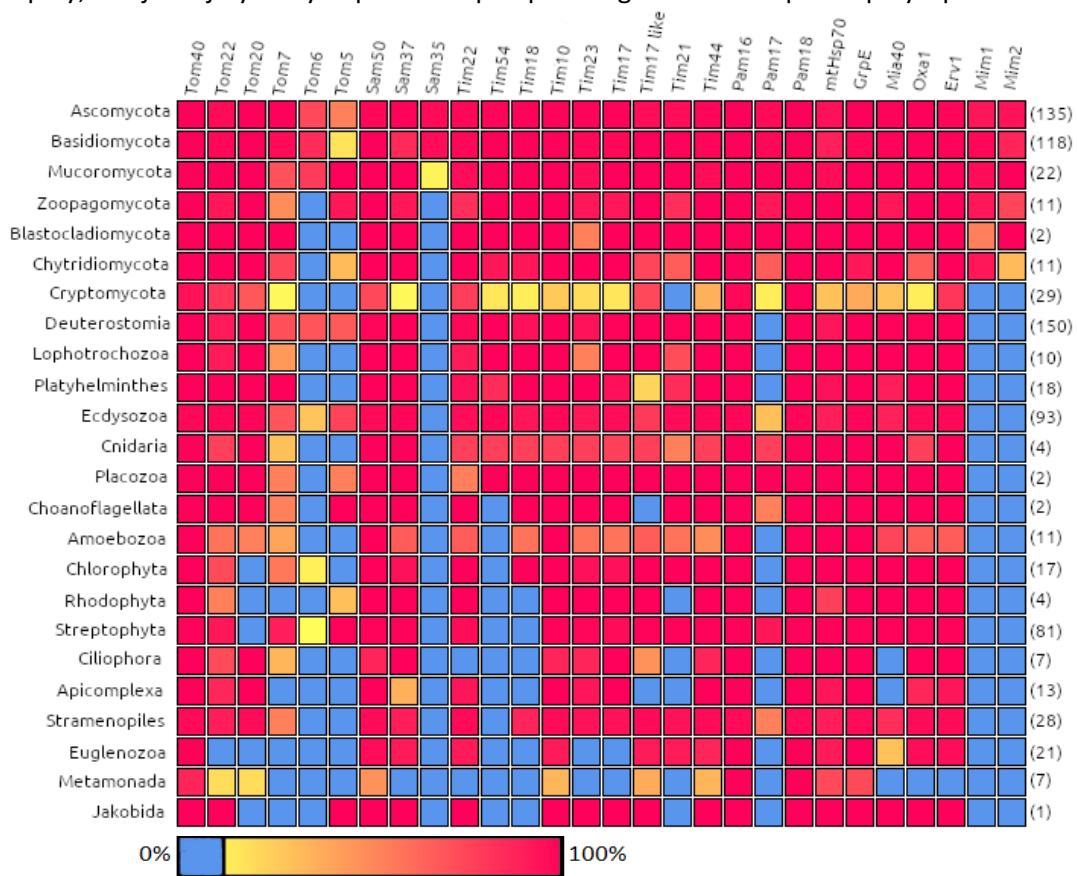
Jako problematické se také ukázaly proteiny z rodiny Tim17 (PF02466), tedy Tim17, Tim22 a Tim23. Tyto proteiny sdílí poměrně vysokou sekvenci identitu a snadno může dojít k jejich záměně. Pro naši analýzu jsme tak vytvořili pouze jeden alignment sestávající se ze sekvencí všech tří proteinů a výsledky jsme poté podrobili klastrovací analýze (**Obr. 19**). Pro klastrovací analýzu jsme vytvořili dataset ze sekvencí Tim17, Tim22 a Tim23, které vytvářejí tři jasně odlišitelné klastry. Dataset jsme následně obohatili našimi sekvencemi a sledovali, k jakému ze tří klastrů se přidají. Zde je nutné říct, že proteinová rodina Tim17 neobsahuje pouze tři zmíněné proteiny, ve skutečnosti obsahuje minimálně deset odlišných proteinů (Žárský and Doležal 2016). Pro zjednodušení jsme všechny tyto ostatní proteiny zařadili do jedné kategorie Tim17-like proteinů.



Obrázek 19: Klastrovací analýza proteinů z rodiny Tim17 pro organismy ze skupiny Chytridiomycota. Zeleně jsou označeny Tim23, oranžově Tim17 a modře Tim22. Šedivě jsou označeny všechny ostatní proteiny z rodiny Tim17, které jsme dále nesespecifikovali. Stejná analýza byla provedena pro všech 24 zkoumaných fylogenetických skupin.

Obdobně jako rodina Tim17 sdružuje více podrodin blíže příbuzných proteinů, rodina Tim10 (PF02953) sdružuje různé malé Tim proteiny. Rozdělení na základě klastrovací analýzy se ale ukázalo jako méně efektivní než v předchozím případě, rozhodli jsme se proto sekvence nerozdělovat do podrodin a ponechali jsme je v souhrnné rodině označené Tim10.

Během automatizovaného procesu analýzy jsou ukládány mezivýsledky, ke kterým jsme se mohli v případě podezření na chybné výsledky vrátit. Podařilo se nám díky nim například odhalit chybnou identifikaci několika sekvencí jako Tom20 u skupin Streptophyta, Chlorophyta a Amoebozoa. I přes naši kontrolu bylo výsledků příliš velké množství, než abychom je všechny mohli manuálně ověřit, a nedá se tak vyloučit přítomnost falešně pozitivních výsledků. Zároveň je při čtení výsledků také nutné brát v potaz počet dostupných dat pro jednotlivé fylogenetické skupiny, kde je zřejmý velký nepoměr ve prospěch organismů ze superskupiny Opisthokonta.



Obrázek 20: Teplotní mapa distribuce proteinů importních mitochondriálních komplexů. V závorce jsou uvedeny počty testovaných druhů pro danou fylogenetickou skupinu. Pořadí komplexů je následující: TOM (Tom40-Tom5), SAM (Sam50-Sam35), TIM22 (Tim22-Tim10), TIM23 (Tim23-Tim21), PAM (Tim44-GrpE), MIA (Mia40-Erv1), MIM (Mim1, Mim2). První řádek mapy obsahuje Ascomycota, do nichž spadá i *S. cerevisiae*, u níž byla velká část z proteinů použitých v této studii poprvé popsána. Na opačné straně mapy leží Jakobida, jejichž mitochondrie mají mitochondriální genom s největším počtem protein kódujících genů (Burger et al. 2013). Do skupiny Cryptomycota byly zahrnuty mikrosporidie společně s *Rozella allomycis* a *Paramicrosporidium saccamoebae* na základě literatury (Letcher et al. 2013; Quandt et al. 2017). Většina z proteinů v této skupině byla nalezena právě u těchto dvou druhů, případně u primitivní mikrosporidie *Mitosporidium daphniae*.

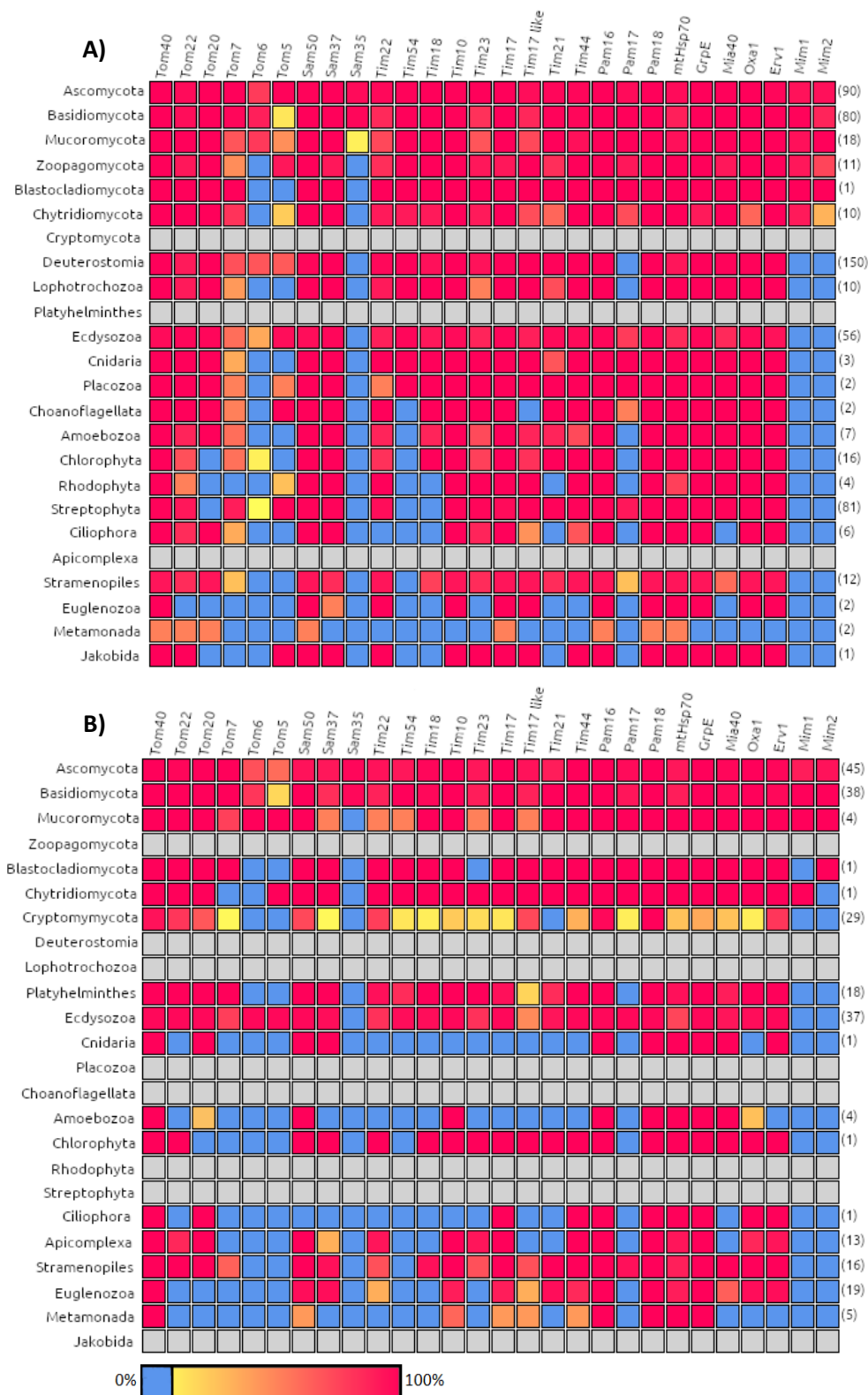
Z analýzy vyplývá, že proteiny zachovalými ve všech liniích jsou Tom40, malé Tim proteiny, Tim44, mtHsp70, GrpE, Pam16 a Pam18. S výjimkou Pam17, který jako jediný chybí i ve skupině Jakobida, se zdají být všechny proteiny PAM komplexu poměrně dobře konzervovány až na skupiny Cryptomycota a Metamonada, kde je zastoupení menší. U těchto dvou skupin se nám obecně dařilo identifikovat menší množství homologů. Je nutné zmínit, že většina zástupců z obou skupin má redukované formy mitochondrií (hydrogenosomy, mitosomy), což by mohlo být zdůvodněním redukce samotných importních komplexů. Komponenty TIM23 translokázy jsou na rozdíl od motorového PAM komplexu méně konzervovány. Nejčastěji z tohoto komplexu chybí Tim21, Tim23 a Tim17 se ve většině organismů vyskytují buď oba, nebo žádný. TIM22 komplex se zdá být konzervovaný ještě o poznání méně s výjimkou malých Tim proteinů, které ale nemusí být nutně vázány pouze na TIM22 (**viz kapitola 2.1**). Tim54 nalezneme pouze ve skupinách Metazoa, Fungi a Cryptomycota, z čehož lze odvozovat, že se vyvinul až u posledního společného předka těchto skupin a u většiny Cryptomycot byl následně ztracen. U SAM komplexu se nám potvrdila ztráta Sam50 u několika organismů specifikovaných v kapitole 5.2.3, kde jsme se problematikou blíže zabývali. Zatímco výskyt Sam37 až na výjimky kopíruje výskyt Sam50, Sam35 se pravděpodobně vyvinul až u odvozených hub. Podobně tak se pravděpodobně i celý MIM komplex vyvinul až u hub. Nejpestřejší je situace u TOM komplexu, kde je výskyt jednotlivých komponent poměrně mozaikovitý, což ukazuje spíše na druhotné ztráty proteinů. Tom7 se například vyskytuje téměř ve všech liniích, ale málokdy jsme ho byli schopni identifikovat pro všechny druhy dané linie.

5.6 Redukce importních komplexů vlivem parazitického způsobu života

S použitím stejné bioinformatické pipeline jako v předchozí kapitole jsme se rozhodli zjistit, zda existuje korelace mezi parazitickým způsobem života a redukcemi mitochondriálních importních komplexů. Z předchozí analýzy je například na první pohled patrné, že organismy ze skupiny Cryptomycota, která je blízce příbuzná houbám, ztratily celou řadu proteinů. Avšak předchozí analýza umožňuje pouze srovnávání celých linií. Pro srovnání organismů v rámci jednotlivých linií bylo nejprve nutné z původních datasetů vyčlenit všechny parazitické organismy a vytvořit z nich nové databáze. Ty byly následně použity pro výpočet procentuálního zastoupení homologů stejně jako původní datasety při zjišťování celkové distribuce proteinů mitochondriálních importních komplexů. Procentuální zastoupení u neparazitických zástupců bylo poté vypočteno srovnáním zastoupení u parazitických zástupců se zastoupením u všech zástupců. Na základě výsledků jsme vykreslily dvě teplotní mapy (**Obr. 21**).

Z výsledků je na první pohled patrné, že se nám u parazitických organismů podařilo identifikovat menší množství homologů. Zejména u skupin Amoebozoa, Cnidaria a Ciliophora jsou rozdíly markantní, avšak počet dostupných dat nebyl ideální. Na druhou stranu i u

parazitických hub, pro něž jsme měli více dat, jsme až na výjimky zaznamenali menší množství homologů. Ve skupinách Metamonada a Euglenozoa, kde jsme měli naopak méně dat pro neparazitické zástupce, se spíše zdá, že parazitický způsob života nemá zásadní vliv. U obou skupin je však identifikace homologů často složitá a je pravděpodobné, že se nám ne všechny podařilo nalézt.



Obrázek 21: Teplotní mapa distribuce proteinů importních mitochondriálních komplexů u neparazitických organismů (A) a parazitických organismů (B). Šedivě jsou vyznačeny linie bez příslušných zástupců. V závorce jsou uvedeny počty testovaných druhů pro danou fylogenetickou skupinu. Pořadí komplexů je následující: TOM (Tom40-Tom5), SAM (Sam50-Sam35), TIM22 (Tim22-Tim10), TIM23 (Tim23-Tim21), PAM (Tim44-GrpE), MIA (Mia40-Erv1), MIM (Mim1, Mim2). Obecně se dá říci, že se nám u parazitických zástupců podařilo identifikovat méně homologů, pro většinu linií jsme však měli k dispozici menší počet parazitických než neparazitických zástupců.

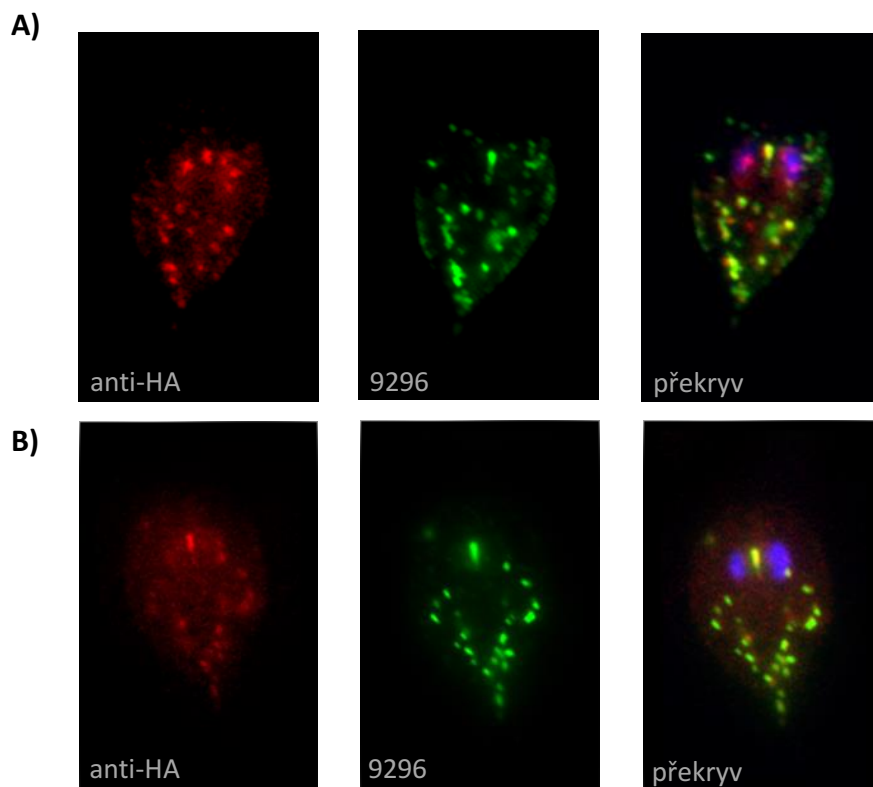
5.7 Malý Tim protein *G. intestinalis*

Jedním z vedlejších výstupů našich bioinformatických analýz byla identifikace možného homologu pro malý Tim protein v genomu *G. intestinalis* (V6TL65, GL50803_2173). Podobně jako malé Tim proteiny u *T. vaginalis* tento protein postrádá charakteristický cysteinový motiv a jeho sekvence obsahuje pouze jeden cystein. Pravděpodobnostní skóre (probability score) navíc nebylo ideální (36,75%). BLAST sekvence proti databázi UniProtKB nám kvůli nízké sekvenční identitě s ostatními homology malých Tim proteinů s ověřením správnosti klasifikace nepomohl. Při použití BLAST proti transkriptomickým databázím použitým v kapitole 5.2.2 jsme však byli schopni identifikovat několik homologů. Z nalezených sekvencí jsme vytvořili alignment a HMM profil, který jsme následně použili v HMMER hledání v transkriptomech. Postupným obohacováním alignmentu o nově nalezené sekvence jsme byli schopni identifikovat poměrně velké množství sekvencí (**Tab. 8**). HHpred finální alignment sestavený z 23 sekvencí klasifikoval jako rodinu malých Tim proteinů s pravděpodobnostním skóre 99,77%. Některé sekvence testované jednotlivě přitom dosahovali hraničního skóre podobně jako sekvence *G. intestinalis*, většina sekvencí od *C. membranifera* naopak dosahovala velmi dobrých skóre. Většina sekvencí také obsahovala čtyři cysteiny, avšak vzdálenost mezi nimi se značně odlišovala od standardního motivu (CX₃CX₁₁₋₁₆CX₃C). S pomocí finálního HMM profilu jsme také hledali v proteomu *S. salmonicida*, avšak nebyli jsme schopni identifikovat žádný homolog.

<i>G. intestinalis</i>	1	1	36,75
<i>C. membranifera</i>	15	4	92,99
<i>E. cyprinoides</i>	1	4	81,42
<i>A. paluster</i>	1	5	92,99
<i>C. cuspidata</i>	1	4	58,97
<i>C. caulleryi</i>	0		
<i>K. bialata</i>	2	2	96,85
<i>D. brevis</i>	2	3	90,98
<i>S. salmonicida</i>	0		

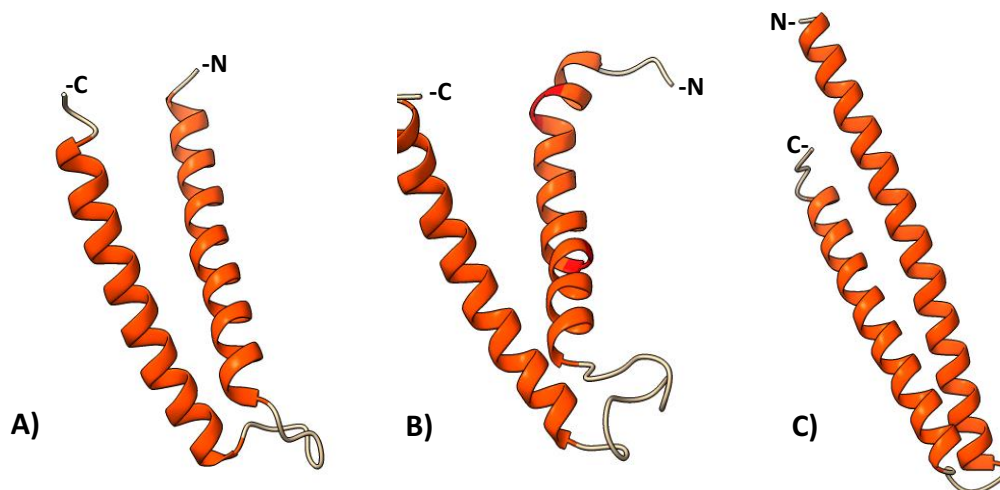
Tabulka 8: Počet homologů malých Tim proteinů identifikovaných u *G. intestinalis* a jí blízké příbuzných organismů. Hodnoty obsažené v prvním sloupci udávají počet nalezených homologů, hodnoty ve druhém udávají počet cysteinů v sekvenci s nejlepším Probability skóre pro daný druh. Poslední sloupec uvádí tyto Probability skóre. Celkem se nám podařilo identifikovat 23 homologů. U *C. caulleryi* jsme dostali jeden výsledek blízko prahu signifikance, protein obsahoval „zinc finger“ motiv, avšak s největší pravděpodobností se nejednalo o malý Tim.

Expresí HA-tagovaného proteinu *G. intestinalis* ukázala jeho specifickou mitosomální lokalizaci, jak je patrné z překryvu se signálem pro mitosomální marker GL50803_9296 (**Obr. 22**). Expresi proteinu se nám však zřejmě z důvodu nízké hladiny exprese nepodařilo ověřit pomocí western blotu.



Obrázek 22: Lokalizace malého Tim proteinu pomocí fluorescenční mikroskopie u *G. intestinalis*. Červeně je značen náš protein pomocí anti-HA protilátky, mitosomy jsou značeny zeleně pomocí anti-GL5080_9296. Jádra jsou pomocí DAPI barvena modře. V některých buňkách jsme pozorovali signál nejen v mitosomech, ale i v oblasti jader (A), v naprosté většině buněk však lokalizace našeho konstruktů odpovídala mitosomům (B).

Jako další kontrolu jsme zvolili predikci 3D struktury proteinu (**Obr. 23**). V tomto případě bylo díky malé velikosti proteinu možné použít přístup ab initio. Zároveň jsme použili i homologní modelování, jelikož jsme měli k dispozici vhodný templát, strukturu Tim9 *S. cerevisiae*, která byla publikována v (Baker et al. 2009).



Obrázek 23: Predikce 3D struktury malého Tim proteinu *G. intestinalis*. A) Tim9 *S. cerevisiae* (PDB identifikátor: 3dxr, řetězec A), B) model vytvořený pomocí homologního modelování, C) model vytvořený metodou ab initio. Oranžově jsou vyznačeny α -helixy, červeně jsou vyznačeny místa přerušení α -helixu. N a C označují konce proteinů.

Predikované modely se liší především ve velikosti N-terminálního α -helixu. U modelu získaného homologním modelováním je kratší (aminokyseliny 5-35 u homologního, 2-45 u ab initio) a obsahuje dva úseky, kdy je α -helix přerušen. C-terminální α -helixy se naopak liší pouze v jedné aminokyselině. Pomocí Ramachandronových grafů jsme se pokusili odhalit problematické úseky. U homologního modelu problematické úseky sedí s přerušením N-terminálního α -helixu, u ab initio modelu jsme v této části proteinu žádný problém nezaznamenali. Oba dva modely však ukazují, že se jedná o strukturu, která odpovídá malým Tim proteinům.

6. Diskuze

6.1 Ztráta Sam50

Jedním z hlavních cílů práce bylo objasnit, zda *G. intestinalis* skutečně ztratila Sam50 či nikoli. Pro ztrátu Sam50 přinášíme dva důkazy. Prvním je neschopnost identifikovat Sam50 i u blízce příbuzných *G. intestinalis*, kdy nejbližší příbuzný organismus se Sam50 je *C. membranifera*. Velká část metamonád tak Sam50 postrádá. Pro *G. intestinalis* máme navíc k dispozici několik teoretických proteomů pro různé asambláže, a považujeme tedy za nepravděpodobné, že by byl Sam50 ztracen během chybné predikce proteinů. Možným vysvětlením je, že se Sam50 stal u těchto organismů příliš divergentním na to, aby jej současné metody pro hledání homologů dokázaly detekovat. Připomeňme například identifikaci Tom40 u *G. intestinalis*, pro níž bylo zapotřebí použít HMMER, jelikož BLAST nebyl dostatečně citlivý (Dagley et al. 2009). Je možné, že se octneme v obdobné situaci, kdy s příchodem nových citlivějších metod bude možné Sam50 snadno identifikovat i u *G. intestinalis*.

V našem dalším úsilí jsme proto upravili naši strategii na hledání strukturně podobných proteinů. V tomto případě bylo klíčovým faktorem správně vyhodnotit věrohodnost 3D modelů, což se ukázalo jako poměrně problematické. Přestože jsme kombinovali více přístupů, je prakticky nemožné zamezit nadhodnocení některých modelů. Zejména proteiny s doménami bohatými na β -listy získávaly vysoké skóre, ačkoli se nejednalo o β -barely. Seřazení modelů dle věrohodnosti je tak spíše indikativní a nikoliv absolutní. Abychom však nevěnovali více času optimalizaci skórovacích metod než samotnému hledání Sam50, rozhodli jsme se, že bude časově přijatelnější používat nedokonalé metody a manuálně pak ověřit nejlepší výsledky. Navíc žádný z modelů proteinů *G. intestinalis* nedosáhl skóre, které bychom mohli považovat za přijatelné a nebylo tak nutné řešit přesné pořadí modelů. Přestože uznáváme, že naše pipeline pro hledání Sam50 na základě 3D struktury není dokonalá a zejména v oblasti hodnocení modelů by ještě potřebovala vylepšit, považujeme výsledek za signifikantní. V případě, že by Sam50 skutečně chyběl, vyvstává otázka, jak jsou u *G. intestinalis* vkládány β -barely do membrány. Jednou z hypotéz může být, že Tom40 převzal roli Sam50 a dokáže sám vkládat ostatní β -barely do membrány. Podobný případ možná existuje. Jak již bylo zmíněno v kapitole 2.3, v chloroplastech došlo k duplikaci proteinu Toc75, který funguje obdobně jako Tom40. O funkci paralogu OEP80 se spekuluje, ale jednou z možností je, že zastává stejnou funkci jako Sam50 (Day et al. 2019). U *G. intestinalis* se nám ovšem nepodařilo identifikovat více kopií Tom40 a jeden protein by tak musel zastávat obě funkce, což z jiných organismů neznáme.

V kapitole 5.2.3 jsme uvedli další organismy bez homologu Sam50. Pomocí fylogenetických analýz jsme odhalili, že se na rozdíl od *G. intestinalis* jedná o ztráty na úrovni druhů. S výjimkou nálevníků, kde jsme však měli k dispozici pouze velmi omezený dataset, tvoří

organismy bez Sam50 oddělenou větev jako je tomu ve skupině Metamonada. Tom40 má z naší zkušenosti obecně větší sklon k divergenci než Sam50, a někdy tak může být složité homolog identifikovat (ATOM u trypanosom, Tom40 *N.parisii*), zatímco Sam50 je podstatně konzervovanější. O tom svědčí i fakt, že jsme schopni identifikovat bez větších obtíží i bakteriální homology, což se u Tom40 dosud nepodařilo. Jelikož se nám podařilo identifikovat homolog Tom40, je nepravděpodobné, že by proteom obsahoval Sam50 a my jsme ho nebyli schopni identifikovat. Na druhou stranu zejména u askomycetních hub považujeme ztrátu za krajně nepravděpodobnou i s ohledem na to, že jsme u nich identifikovali některé jiné komponenty SAM komplexu. Možným vysvětlením může být i nekompletnost proteomu. U mikrosporidií by se dala ztráta očekávat více. Stejně jako *G. intestinalis* mají místo klasických mitochondrií mitosomy a redukce genomu je u nich častým jevem.

Otázka vkládání Tom40 do vnější mitosomální membrány u *G. intestinalis* zůstává otevřena. Zajímavým nálezem je v tomto směru identifikace malého Tim proteinu. U eukaryot se Sam50 jsou malé Tim proteiny nezbytné pro přenesení Tom40 do SAM komplexu. Přítomnost malého Timu tak nasvědčuje, že Tom40 přichází do membrány z mezimembránového prostoru. Jediným známým β -barelem vnější mitosomální membrány u *G. intestinalis* je pak samotný Tom40, který by tak teoreticky mohl přebírat funkci Sam50.

6.2 Mutace β -signálu

V kapitole 5.3 jsme se zabývali β -signálem mitochondriálních β -barelů a z výsledků analýzy vyplývá, že až na naprosté výjimky je β -signál zachován. V případě výjimek se jedná o jednobodové mutace, přičemž většinu výjimek nalezneme u *T. vaginalis*. Bylo by zajímavé udělat podobnou analýzu zahrnující nejen mitochondriální β -barely, ale i bakteriální a chloroplastové β -barely, na což bohužel již nezbyl čas. Dva z šesti případů mutace β -signálu se vyskytují u proteinů z rodiny Hmp35, která je pravděpodobně specifická pro trichomonády. Mohlo by tak být přínosné zkusit najít další mitochondriální β -barely specifické pro určité eukaryotické linie a zjistit jak vypadá jejich β -signál. Přestože programů pro predikci transmembránových β -barelů je více (**Tab. 4**), jejich spolehlivost je diskutabilní. Stejně tak programy predikující buněčnou lokalizaci nejsou vždy spolehlivé a pro nalezení nových eukaryotických β -barelů by bylo zapotřebí vytvořit bioinformatickou pipeline využívající součinnost více predikčních programů. Jelikož je β -signál poměrně dobře konzervován, dal by se pro hledání také dobře použít.

U proteinů Hmp36 jsme navíc narazili na další zvláštnost. Dle PSIPRED poloha β -signálu neodpovídá poloze posledního β -listu. Přesnost PSIPREDU u proteinů bez více známých homologů však nemusí být optimální. Navíc jak bylo zmíněno v kapitole 5.3, přesnost predikce u transmembránových proteinů je obecně nižší než u solubilních. Přesto by jistě stálo zato

provést analýzu, v rámci které bychom predikovali sekundární strukturu všech β -barelů použitých v kapitole 5.3 a zjistili v kolika procentech případů β -signál skutečně odpovídá poslednímu predikovanému β -listu. Predikce sekundární struktury je však i při lokálním použití PSIPRED časově náročná a v rámci této diplomové práce pro takovouto analýzu již nezbyl čas. PSIPRED sice umožňuje vynechání PSI-BLAST kroku, čímž se rychlost predikce výrazně zvýší, z naší zkušenosti je však dopad na přesnost a značný a pro takovouto analýzu by použití této verze mohlo vést k chybným závěrům.

6.3 Teplotní mapy distribuce proteinů mitochondriálních importních komplexů

V kapitolách 5.5 a 5.6 jsme vytvořili celkem tři teplotní mapy vyobrazující distribuci proteinů mitochondriálních importních komplexů. Analýza ukázala, že za univerzálně konzervované můžeme považovat proteiny Tom40, Pam18, Pam16, mtHsp70, GrpE a malé Tim proteiny. Jedná se o proteiny z různých komplexů. Zatímco u MIM komplexu je patrné, že se vyvinul až u hub, ztráty komplexů TIM22 a TIM23 považujeme za sekundární. Pravděpodobně není náhodou, že se nacházejí zejména u anaerobních organismů. Na absenci těchto komplexů se tak mohla významně podílet ztráta membránového potenciálu, který by mohl být nezbytný pro jejich funkci. U SAM komplexu nám teplotní mapy potvrdily výsledky analýzy z kapitoly 5.2.3 (Ztráta Sam50 v dalších eukaryotických liniích).

Původní snaha o vytvoření zcela automatizovaného procesu se ukázala jako nereálná. V první verzi programu jsme používali postupné obohacování alignmentu o nově nalezené homology specifické pro danou eukaryotickou linii a opakované HMMER hledání. Pro některé proteiny se tato metoda ukázala jako vhodná, ve většině případů s sebou však nesla pouze velké množství falešně pozitivních výsledků. Množství chyb jsme se pokoušeli omezit vhodnějším nastavením minimální E-value pro signifikantní výsledek. Snížení minimální E-value ale přineslo jiné problémy, divergentní homology, například Tom40 ve skupině Euglenozoa, nebyly vůbec identifikovány. Abychom nemuseli zvlášť připravovat specifický postup pro každý protein v každé skupině, rozhodli jsme se, že počet iterativních kroků omezíme na jeden. Tím se výrazně snížil počet falešně pozitivních výsledků, ale nepochybně jsme tím i ztratily některé vzdálenější homology. I přes snahu o automatizaci procesu bylo nutné výsledky manuálně kontrolovat. Kvůli velkému množství dat jsme nebyli schopni projít všechny výsledky, ale snažili jsme se o odstranění, co největšího množství nepřesností. Pro některé fylogenetické skupiny jsme měli poměrně málo dat, což jistě zkresluje výsledek. Výsledek je také ovlivněn tím, že většina těchto proteinů byla poprvé popsána u *S. cerevisiae*. U MIM komplexu či Sam35 je jasně vidět, že se jedná o proteiny specifické pro houby. Kolik podobných specifických proteinů mají ostatní skupiny, je pochopitelně těžší zjistit. Skupina Metamonada z analýzy vyšla jako skupina s největším počtem redukcí importních komplexů, pro *T. vaginalis* ale známe specifické

mitochondriální β -barely (Hmp35, Hmp36) a je tak otázkou kolik unikátních proteinů organismy z této i ostatních skupiny mají. Pro lepší pochopení evoluce mitochondriálních importních komplexů by také bylo vhodné mít více proteomů či transkriptomů pro Jakobida, jejichž mitochondrie se zachovávají nejvíce mitochondriální DNA ze všech známých eukaryot (Burger et al. 2013). Jak moc je stav importních komplexů *A. godoyi* typický pro Jakobida nedokážeme říct, avšak nepředpokládáme, že by zde došlo k výraznějším sekundárním ztrátám proteinů.

Teplotní mapy srovnávající parazitické a neparazitické organismy nám ukázaly, že ve většině případů mají parazité komplexy redukovanější. Parazitické organismy ale také často žijí v anaerobním prostředí a mají redukované formy mitochondrií. Forma mitochondrie není v naší analýze nijak zohledněna, ale jistě by stálo zato vytvořit obdobné teplotní mapy reflektující stav mitochondrie.

7. Závěr

Cílem této práce bylo získat nové poznatky o evoluci importu proteinů do mitochondrií. Důraz byl kladen především na proteinové komplexy vnější mitochondriální membrány, tedy komplexy TOM a SAM, které až na výjimky najdeme ve všech eukaryotických organismech. Právě na tyto výjimky jsme se zaměřili. Tato práce tak přináší důkazy o absenci homologu Sam50 u *G. intestinalis* a jí blízce příbuzných organismů. Zároveň jsme ukázali, že takováto ztráta nemusí být mezi eukaryoty zcela ojedinělá, avšak u ostatních nalezených organismů bez zjevného homologu Sam50 jsou důkazy pro absenci slabší. Jelikož jsou hlavními proteiny TOM i SAM komplexu β -barely, byla provedena analýza sekvence pomocí níž spolu β -barely interagují, tzv. β -signálu. Dokázali jsme nalézt pouze šest mitochondriálních β -barelů s mutací v β -signálu, přičemž tři z případů se vyskytovaly u *T. vaginalis*. Že jsou importní komplexy organismů ze skupiny Metamonada, kam patří právě *G. intestinalis* a *T. vaginalis*, neobvyklé bylo prokázáno i naší analýzou distribuce proteinů importních mitochondriálních komplexů, do níž bylo zahrnuto téměř osm set organismů. V ní se také ukázalo, že parazitické organismy mají komplexy ve srovnání se svými volně žijícími příbuznými často redukované. Na druhou stranu jsme se v rámci této práce často setkávali s velmi divergentními homology, které nebylo vždy možné jednoduše identifikovat, a to zejména ve skupinách, kde nacházíme největší množství redukcí. Považujeme tak za pravděpodobné, že některé homology není možné identifikovat, jelikož současné metody pro detekci homologie nejsou dostatečně citlivé.

8. Seznam použitých zkratk

ATOM	„Archaic translocase of the outer mitochondrial membrane“
Bac_surface_Ag	„Bacterial surface antigen“
BAM	„ β -barrel assembly machinery“
CYC2	„Cytochrome c mitochondrial import factor 2“
DAPI	4,6-diamid-2-fenylindol
DNA	deoxyribonukleová kyselina
dNTPs	deoxynukleotidy
G418	geneticin
HA-tag	hemagglutininový tag
HMM	skrytý Markovův model
LB	růstové médium pro bakterie
MCMC	Monte Carlo Markov Chain
Mdm10	„Mitochondrial distribution and morphology protein 10“
MIA	„Mitochondrial intermembrane space assembly“
MIM	„Machinery of the inner membrane“
NIF	fosfatázová doména
PAM	„Mitochondrial protein import motor“
PBS	fosfátový pufr
PCR	polymerázová řetězová reakce
POTRA	„Polypeptide transport-associated domain“
RNA	ribonukleová kyselina
SDS	dodecylsírán sodný
SAM	„Sorting and assembly machinery“
TGS	„tris-glycine-SDS“ pufr
TIM	„Translocase of the inner membrane“
TOM	„Translocase of the outer membrane“
Tween	polyoxyethylensorbitanmonolaurát
VDAC	„Voltage dependent anion channel“

9. Seznam použité literatury

- Abe Yoshito, Toshihiro Shodai, Takanori Muto, Katsuyoshi Mihara, Hisayoshi Torii, Shuh ichi Nishikawa, Toshiya Endo, & Daisuke Kohda.** 2000. "Structural Basis of Presequence Recognition by the Mitochondrial Protein Import Receptor Tom20." *Cell* **100**(5):551–60.
- Abu-Hamad Salah, Nir Arbel, Doron Calo, Laetitia Arzoine, Adrian Israelson, Nurit Keinan, Ronit Ben-Romano, Orr Friedman, & Varda Shoshan-Barmatz.** 2009. "The VDAC1 N-Terminus Is Essential Both for Apoptosis and the Protective Effect of Anti-Apoptotic Proteins." *Journal of Cell Science* **122**(11):1906–16.
- Allen Scott, Hui Lu, David Thornton, & Kostas Tokatlidis.** 2003. "Juxtaposition of the Two Distal CX3C Motifs via Intrachain Disulfide Bonding Is Essential for the Folding of Tim10." *Journal of Biological Chemistry* **278**(40):38505–13.
- Altschul S., T. Madden, A. Schäffer, J. Zhang, Z. Zhang, W. Miller, & D. Lipman.** 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* **25**(17):3389–3402.
- Araiso Yuhei, Akihisa Tsutsumi, Jian Qiu, Kenichiro Imai, Takuya Shiota, Jiyao Song, Caroline Lindau, Lena Sophie Wenz, Haruka Sakaue, Kaori Yunoki, Shin Kawano, Junko Suzuki, Marilena Wischnewski, Conny Schütze, Hirotaka Ariyama, Toshio Ando, Thomas Becker, Trevor Lithgow, Nils Wiedemann, Nikolaus Pfanner, Masahide Kikkawa, & Toshiya Endo.** 2019. "Structure of the Mitochondrial Import Gate Reveals Distinct Preprotein Paths." *Nature* **575**(7782):395–401.
- Bagos Pantelis G., Theodore D. Liakopoulos, Ioannis C. Spyropoulos, & Stavros J. Hamodrakas.** 2004. "A Hidden Markov Model Method, Capable of Predicting and Discriminating β -Barrel Outer Membrane Proteins." *BMC Bioinformatics* **5**:1–13.
- Baker J. Michael, T. Chaille Webb, A. David Stroud, S. Catherine Palmer, E. Ann Frazier, Bernard Guiard, Agnieszka Chacinska, M. Jacqueline Gulbis, & T. Michael Ryan.** 2009. "Structural and Functional Requirements for Activity of the Tim9–Tim10 Complex in Mitochondrial Protein Import." *Molecular Biology of the Cell* **20**:769–77.
- Bayrhuber Monika, Thomas Meins, Michael Habeck, Stefan Becker, Karin Giller, Saskia Villinger, Clemens Vornhein, Christian Griesinger, Markus Zweckstetter, & Kornelius Zeth.** 2008. "Structure of the Human Voltage-Dependent Anion Channel." *Proceedings of the National Academy of Sciences of the United States of America* **105**(40):15370–75.
- Becker Thomas, Guiard Bernard, Thornton Nicolas, Stroud A. David Wiedemann Nils.** 2010. "Assembly of the Mitochondrial Protein Import Channel Role of Tom5 in Two-Stage Interaction of Tom40 with the SAM Complex." *Molecular Biology of the Cell* **21**(24):3106–3113.
- Becker Thomas, Lena Böttinger, & Nikolaus Pfanner.** 2012. "Mitochondrial Protein Import: From Transport Pathways to an Integrated Network." *Trends in Biochemical Sciences* **37**(3):85–91.
- Becker Thomas, Michael Gebert, Nikolaus Pfanner, & Martin van der Laan.** 2009. "Biogenesis of Mitochondrial Membrane Proteins." *Current Opinion in Cell Biology* **21**(4):484–93.

- Berman Helen M., Tammy Battistuz, T. N. Bhat, Wolfgang F. Bluhm, Philip E. Bourne, Kyle Burkhardt, Zukang Feng, Gary L. Gilliland, Lisa Iype, Shri Jain, Phoebe Fagan, Jessica Marvin, David Padilla, Veerasamy Ravichandran, Bohdan Schneider, Narmada Thanki, Helge Weissig, John D. Westbrook, & Christine Zardecki.** 2002. "The Protein Data Bank." *Acta Crystallographica Section D: Biological Crystallography* **58**(6 1):899–907.
- Boutet Emmanuel, Damien Lieberherr, Michael Tognolli, Michel Schneider, & Amos Bairoch.** 2007. "UniProtKB/Swiss-Prot: The Manually Annotated Section of the UniProt KnowledgeBase." *Methods in Molecular Biology* **406**:89–112.
- Bowie James U., Roland Lütthy, & David Eisenberg.** 1991. "A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure." *Science* **253**:164–70.
- Brix Jan, Stefan Rüdiger, Bernd Bukau, Jens Schneider-Mergener, & Nikolaus Pfanner.** 1999. "Distribution of Binding Sequences for the Mitochondrial Import Receptors Tom20, Tom22, and Tom70 in a Presequence-Carrying Preprotein and a Non-Cleavable Preprotein." *Journal of Biological Chemistry* **274**(23):16522–30.
- Buchan Daniel W. A., & David T. Jones.** 2019. "The PSIPRED Protein Analysis Workbench: 20 Years On." *Nucleic Acids Research* **47**(W1):W402–7.
- Buczek Dorota, Malgorzata Wojtkowska, Yutaka Suzuki, Seiji Sonobe, Yukinori Nishigami, Monika Antoniewicz, Hanna Kmita, & Wojciech Makalowski.** 2016. "Protein Import Complexes in the Mitochondrial Outer Membrane of Amoebozoa Representatives." *BMC Genomics* **17**(1).
- Burger Gertraud, Michael W. Gray, Lise Forget, & B. Franz Lang.** 2013. "Strikingly Bacteria-like and Gene-Rich Mitochondrial Genomes throughout Jakobid Protists." *Genome Biology and Evolution* **5**(2):418–38.
- Bycroft Mark, Tim J. P. Hubbard, Mark Proctor, Stefan M. V. Freund, & Alexey G. Murzin.** 1997. "The Solution Structure of the S1 RNA Binding Domain: A Member of an Ancient Nucleic Acid-Binding Fold." *Cell* **88**(2):235–42.
- Capella-Gutiérrez Salvador, José M. Silla-Martínez, & Toni Gabaldón.** 2009. "TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* **25**(15):1972–73.
- Chan Nickie C., & Trevor Lightow.** 2008. "The Peripheral Membrane Subunits of the SAM Complex Function Codependently in Mitochondrial Outer Membrane Biogenesis." *Molecular Biology of the Cell* **19**(1):126–36.
- Cline Kenneth, & Carole Dabney-Smith.** 2008. "Plastid Protein Import and Sorting: Different Paths to the Same Compartments." *Current Opinion in Plant Biology* **11**(6):585–92.
- Colombini M., E. Blachly-Dyson, & M. Forte.** 1996. "VDAC, a Channel in the Outer Mitochondrial Membrane." *Ion Channels* **4**:169–202.
- Dagley Michael J., Pavel Dolezal, Vladimir A. Likić, Ondrej Smid, Anthony W. Purcell, Susan K. Buchanan, Jan Tachezy, & Trevor Lithgow.** 2009. "The Protein Import Channel in the Outer Mitochondrial Membrane of *Giardia Intestinalis*." *Molecular Biology and Evolution* **26**(9):1941–47.

- Day Philip M., Kentaro Inoue, & Steven M. Theg.** 2019. "Chloroplast Outer Membrane β -Barrel Proteins Use Components of the General Import Apparatus." *Plant Cell* **31**(8):1845–55.
- Day Philip M., Daniel Potter, & Kentaro Inoue.** 2014. "Evolution and Targeting of Omp85 Homologs in the Chloroplast Outer Envelope Membrane." *Frontiers in Plant Science* **5**(10):1–18.
- Dembowski Markus, Klaus Peter Künkele, Frank E. Nargang, Walter Neupert, & Doron Rapaport.** 2001. "Assembly of Tom6 and Tom7 into the TOM Core Complex of *Neurospora Crassa*." *Journal of Biological Chemistry* **276**(18):17679–85.
- Deng Haiyou, Ya Jia, & Yang Zhang.** 2018. "Protein Structure Prediction." *International Journal of Modern Physics B* **32**(18):1–18.
- Dolezal Pavel, Vladimir Likic, Jan Tachezy, & Trevor Lithgow.** 2006. "Evolution of the Molecular Machines for Protein Import into Mitochondria." *Science* **313**:314–18.
- Dolezal Pavel, Ondrej Smid, Petr Rada, Zuzana Zubáková, Dejan Bursać, Robert Suták, Jana Nebesárová, Trevor Lithgow, & Jan Tachezy.** 2005. "Giardia Mitosomes and Trichomonad Hydrogenosomes Share a Common Mode of Protein Targeting." *Proceedings of the National Academy of Sciences of the United States of America* **102**(31):10924–29.
- Dudek Jan, Peter Rehling, & Martin van der Laan.** 2013. "Mitochondrial Protein Import: Common Principles and Physiological Networks." *Biochimica et Biophysica Acta - Molecular Cell Research* **1833**(2):274–85.
- Durigon Romina, Qi Wang, Efrain Ceh Pavia, Chris M. Grant, & Hui Lu.** 2012. "Cytosolic Thioredoxin System Facilitates the Import of Mitochondrial Small Tim Proteins." *EMBO Reports* **13**(10):916–22.
- Dyall Sabrina D., Dianna C. Lester, Rachel E. Schneider, Maria G. Delgadillo-Correa, Evelyn Plümper, Ariadna Martinez, Carla M. Koehler, & Patricia J. Johnson.** 2003. "Trichomonas Vaginalis Hmp35, a Putative Pore-Forming Hydrogenosomal Membrane Protein, Can Form a Complex in Yeast Mitochondria." *Journal of Biological Chemistry* **278**(33):30548–61.
- Fairman James W., Nicholas Noinaj, & Susan K. Buchanan.** 2012. "The Structural Biology of β -Barrel Membrane Proteins: A Summary of Recent Reports." *Current Opinion in Structural Biology* **21**(4):523–31.
- Finn Robert D., Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, Kirstie Hetherington, Liisa Holm, Jaina Mistry, Erik L. L. Sonnhammer, John Tate, & Marco Punta.** 2014. "Pfam: The Protein Families Database." *Nucleic Acids Research* **42**(1):222–30.
- Finn Robert D., Jody Clements, & Sean R. Eddy.** 2011. "HMMER Web Server: Interactive Sequence Similarity Searching." *Nucleic Acids Research* **39**(2):29–37.
- Flinner Nadine, Lars Ellenrieder, Sebastian B. Stiller, Thomas Becker, Enrico Schleiff, & Oliver Mirus.** 2013. "Mdm10 Is an Ancient Eukaryotic Porin Co-Occurring with the ERMES Complex." *Biochimica et Biophysica Acta - Molecular Cell Research* **1833**(12):3314–25.

- Frickey Tancred, & Andrei Lupas.** 2004. "CLANS: A Java Application for Visualizing Protein Families Based on Pairwise Similarity." *Bioinformatics* **20**(18):3702–4.
- Gadagkar Sudhindra R., Michael S. Rosenberg, & Sudhir Kumar.** 2005. "Inferring Species Phylogenies from Multiple Genes: Concatenated Sequence Tree versus Consensus Gene Tree." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **304**(1):64–74.
- Galganska Hanna, Andonis Karachitos, Malgorzata Wojtkowska, Olgierd Stobienia, Malgorzata Budzinska, & Hanna Kmita.** 2010. "Communication between Mitochondria and Nucleus: Putative Role for VDAC in Reduction/Oxidation Mechanism." *Biochimica et Biophysica Acta - Bioenergetics* **1797**(6):1276–80.
- Gentle Ian E., Andrew J. Perry, Felicity H. Alcock, Vladimir A. Likić, Pavel Dolezal, Ee Ting Ng, Anthony W. Purcell, Malcolm McConville, Thomas Naderer, Anne Laure Chanez, Fabien Charrière, Caroline Aschinger, André Schneider, Kostas Tokatlidis, & Trevor Lithgow.** 2007. "Conserved Motifs Reveal Details of Ancestry and Structure in the Small TIM Chaperones of the Mitochondrial Intermembrane Space." *Molecular Biology and Evolution* **24**(5):1149–60.
- Gillin F. D., D. S. Reiner, & S. E. Boucher.** 1988. "Small-Intestinal Factors Promote Encystation of Giardia Lamblia in Vitro." *Infection and Immunity* **56**(3):705–7.
- Goddard Thomas D., Conrad C. Huang, Elaine C. Meng, Eric F. Pettersen, Gregory S. Couch, John H. Morris, & Thomas E. Ferrin.** 2018. "UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis." *Protein Science* **27**(1):14–25.
- Gray Michael W., Gertraud Burger, Romain Derelle, Vladimír Klimeš, Michelle M. Leger, Matt Sarrasin, Čestmír Vlček, Andrew J. Roger, Marek Eliáš, & B. Franz Lang.** 2020. *The Draft Nuclear Genome Sequence and Predicted Mitochondrial Proteome of Andalusia Godoyi, a Protist with the Most Gene-Rich and Bacteria-like Mitochondrial Genome*. Vol. 18.
- Gu Yinghong, Huanyu Li, Haohao Dong, Yi Zeng, Zhengyu Zhang, Neil G. Paterson, Phillip J. Stansfeld, Zhongshan Wang, Yizheng Zhang, Wenjian Wang, & Changjiang Dong.** 2016. "Structural Basis of Outer Membrane Protein Insertion by the BAM Complex." *Nature* **531**(7592):64–69.
- Hayat Sikander, Christoph Peters, Nanjiang Shu, Konstantinos D. Tsirigos, & Arne Elofsson.** 2016. "Inclusion of Dyad-Repeat Pattern Improves Topology Prediction of Transmembrane β -Barrel Proteins." *Bioinformatics* **32**(10):1571–73.
- Heins Lisa, & Udo Klaus Schmitz.** 1996. "A Receptor for Protein Import into Potato Mitochondria." *Plant Journal* **9**(6):829–39.
- Herrmann Johannes M., & Jan Riemer.** 2012. "Mitochondrial Disulfide Relay: Redox-Regulated Protein Import into the Intermembrane Space." *Journal of Biological Chemistry* **287**(7):4426–33.
- Hill Kerstin, Kirstin Model, Michael T. Ryan, Klaus Dietmeier, Falk Martint, Richard Wagner, & Nikolaus Pfanner.** 1998. "Tom40 Forms the Hydrophilic Channel of the Mitochondrial Import Pore for Preproteins." *Nature* **395**(6701):516–21.
- Hiller Sebastian, Robert G. Garces, Thomas J. Malia, Vladislav Y. Orekhov, Marco Colombini,**

- & Gerhard Wagner.** 2008. "Solution Structure of the Integral Human Membrane Protein VDAC-1 in Detergent Micelles." *Science* **321**(5893):1206–10.
- Höhr Alexandra I. C., Caroline Lindau, Christophe Wirth, Jian Qiu, David A. Stroud, Stephan Kutik, Bernard Guiard, Carola Hunte, Thomas Becker, Nikolaus Pfanner, & Nils Wiedemann.** 2018. "Membrane Protein Insertion through a Mitochondrial β -Barrel Gate." *Science* **359**(6373).
- Hönlinger A., U. Bömer, A. Alconada, C. Eckerskorn, F. Lottspeich, K. Dietmeier, & N. Pfanner.** 1996. "Tom7 Modulates the Dynamics of the Mitochondrial Outer Membrane Translocase and Plays a Pathway-Related Role in Protein Import." *The EMBO Journal* **15**(9):2125–37.
- Illergård Kristoffer, David H. Ardell, & Arne Elofsson.** 2009. "Structure Is Three to Ten Times More Conserved than Sequence - A Study of Structural Response in Protein Cores." *Proteins: Structure, Function and Bioinformatics* **77**(3):499–508.
- Ivanova Ekaterina, Thomas A. Jowitt, & Hui Lu.** 2008. "Assembly of the Mitochondrial Tim9-Tim10 Complex: A Multi-Step Reaction with Novel Intermediates." *Journal of Molecular Biology* **375**(1):229–39.
- Jedelský Petr L., Pavel Doležal, Petr Rada, Jan Pyrih, Ondřej Šmíd, Ivan Hrdý, Miroslava Šedinová, Michaela Marcinčíková, Lubomír Voleman, Andrew J. Perry, Neritza Campo Beltrán, Trevor Lithgow, & Jan Tachezy.** 2011. "The Minimal Proteome in the Reduced Mitochondrion of the Parasitic Protist *Giardia Intestinalis*." *PLoS ONE* **6**(2):15–21.
- Jerlström-Hultqvist Jon, Britta Stadelmann, Sandra Birkestedt, Ulf Hellman, & Staffan G. Svärd.** 2012. "Plasmid Vectors for Proteomic Analyses in *Giardia*: Purification of Virulence Factors and Analysis of the Proteasome." *Eukaryotic Cell* **11**(7):864–73.
- Jex Aaron R., Staffan Svärd, Kari D. Hagen, Hannah Starcevich, Samantha J. Emery-Corbin, Balu Balan, Chris Nosala, & Scott C. Dawson.** 2020. "Recent Advances in Functional Research in *Giardia Intestinalis*." *Advances in Parasitology* **107**:97–137.
- Jones David T.** 1999a. "GenTHREADER: An Efficient and Reliable Protein Fold Recognition Method for Genomic Sequences." *Journal of Molecular Biology* **287**(4):797–815.
- Jones David T.** 1999b. "Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices." *Journal of Molecular Biology* **292**(2):195–202.
- Kay Christopher J., Karen Lawler, & Ian D. Kerr.** 2013. "Analysis of the Sam50 Translocase of Excavate Organisms Supports Evolution of Divergent Organelles from a Common Endosymbiotic Event." *Bioscience Reports*.
- Keister David B.** 1983. "Axenic Culture of *Giardia Lamblia* in TYI-S-33 Medium Supplemented with Bile." *Transactions of the Royal Society of Tropical Medicine and Hygiene* **77**(4):487–88.
- Kim David E., Dylan Chivian, & David Baker.** 2004. "Protein Structure Prediction and Analysis Using the Robetta Server." *Nucleic Acids Research* **32**(Web Server issue):W526-31.
- Kim Kelly H., Suraaj Aulakh, & Mark Paetzel.** 2012. "The Bacterial Outer Membrane β -Barrel Assembly Machinery." *Protein Science* **21**(6):751–68.

- Komiya Tohru, Sabine Rospert, Caria Koehler, Renate Looser, Gottfried Schatz, & Katsuyoshi Mihara.** 1998. "Interaction of Mitochondrial Targeting Signals with Acidic Receptor Domains along the Protein Import Pathway: Evidence for the 'acid Chain' Hypothesis." *EMBO Journal* **17**(14):3886–98.
- Kondo-Okamoto Noriko, Janet M. Shaw, & Koji Okamoto.** 2008. "Tetratricopeptide Repeat Proteins Tom70 and Tom71 Mediate Yeast Mitochondrial Morphogenesis." *EMBO Reports* **9**(1):63–69.
- Kutik Stephan, Diana Stojanovski, Lars Becker, Thomas Becker, Michael Meinecke, Vivien Krüger, Claudia Prinz, Chris Meisinger, Bernard Guiard, Richard Wagner, Nikolaus Pfanner, & Nils Wiedemann.** 2008. "Dissecting Membrane Insertion of Mitochondrial β -Barrel Proteins." *Cell* **132**(6):1011–24.
- Lackey Sebastian W. K., Rebecca D. Taylor, Nancy E. Go, Annie Wong, E. Laura Sherman, & Frank E. Nargang.** 2014. "Evidence Supporting the 19 β -Strand Model for Tom40 from Cysteine Scanning and Protease Site Accessibility Studies." *Journal of Biological Chemistry* **289**(31):21640–50.
- Lartillot Nicolas, Henner Brinkmann, & Hervé Philippe.** 2007. "Suppression of Long-Branch Attraction Artefacts in the Animal Phylogeny Using a Site-Heterogeneous Model." *BMC Evolutionary Biology* **7**(SUPPL. 1):1–26.
- Lartillot Nicolas, & Hervé Philippe.** 2004. "A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process." *Molecular Biology and Evolution* **21**(6):1095–1109.
- Lartillot Nicolas, & Hervé Philippe.** 2006. "Computing Bayes Factors Using Thermodynamic Integration." *Systematic Biology* **55**(2):195–207.
- Lauwaet Tineke, Barbara J. Davids, Ascención Torres-Escobar, Shanda R. Birkeland, Michael J. Cipriano, Sarah P. Preheim, Daniel Palm, Staffan G. Svärd, Andre G. McArthur, & Frances D. Gillin.** 2007. "Protein Phosphatase 2A Plays Crucial Role in Differentiation." *Mol Biochem Parasitol.* **152**(1):80–89.
- Leger Michelle M., Martin Kolisko, Ryoma Kamikawa, Courtney W. Stairs, Keitaro Kume, Ivan Čepička, Jeffrey D. Silberman, Jan O. Andersson, Feifei Xu, Akinori Yabuki, Laura Eme, Qianqian Zhang, Kiyotaka Takishita, Yuji Inagaki, Alastair G. B. Simpson, Tetsuo Hashimoto, & Andrew J. Roger.** 2017. "Organelles That Illuminate the Origins of Trichomonas Hydrogenosomes and Giardia Mitosomes." *Nature Ecology and Evolution* **1**(4):1–7.
- Leonard-Rivera Margaret, & Rajeev Misra.** 2012. "Conserved Residues of the Putative L6 Loop of Escherichia Coli BamA Play a Critical Role in the Assembly of β -Barrel Outer Membrane Proteins, Including That of BamA Itself." *Journal of Bacteriology* **194**(17):4662–68.
- Letcher Peter M., Salvador Lopez, Robert Schmieder, Philip A. Lee, Craig Behnke, Martha J. Powell, & Robert C. McBride.** 2013. "Characterization of Amoebophilidium Protococcarum, an Algal Parasite New to the Cryptomycota Isolated from an Outdoor Algal Pond Used for the Production of Biofuel." *PLoS ONE* **8**(2).
- Lodi Patricia J., James A. Ernst, John Kuszewski, Alison B. Hickman, Alan Engelman, Robert Craigie, G. Marius Clore, & Angela M. Gronenbom.** 1995. "Solution Structure of the DNA

- Binding Domain of HIV-1 Integrase." *Biochemistry* **34**(31):9826–33.
- Lu Hui, & Joanna Woodburn.** 2005. "Zinc Binding Stabilizes Mitochondrial Tim10 in a Reduced and Import-Competent State Kinetically." *Journal of Molecular Biology* **353**(4):897–910.
- Lutz Thomas, Walter Neupert, & Johannes M. Herrmann.** 2003. "Import of Small Tim Proteins into the Mitochondrial Intermembrane Space." *EMBO Journal* **22**(17):4400–4408.
- Mačašev Diana, James Whelan, Ed Newbigin, Marcio C. Silva-Filho, Terrence D. Mulhern, & Trevor Lithgow.** 2004. "Tom22', an 8-KDa Trans-Site Receptor in Plants and Protozoans, Is a Conserved Feature of the TOM Complex That Appeared Early in the Evolution of Eukaryotes." *Molecular Biology and Evolution*.
- Madeira Fábio, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, Adrian R. N. Tivey, Simon C. Potter, Robert D. Finn, & Rodrigo Lopez.** 2019. "The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019." *Nucleic Acids Research* **47**(W1):W636–41.
- Meisinger Chris, Michael Rissler, Agnieszka Chacinska, Luiza K. Sanjuán Szklarz, Dusanka Milenkovic, Vera Kozjak, Birgit Schönfisch, Christiane Lohaus, Helmut E. Meyer, Michael P. Yaffe, Bernard Guiard, Nils Wiedemann, & Nikolaus Pfanner.** 2004. "The Mitochondrial Morphology Protein Mdm10 Functions in Assembly of the Preprotein Translocase of the Outer Membrane." *Developmental Cell* **7**(1):61–71.
- Mesecke Nikola, Karl Bihlmaier, Barbara Grumbt, Sebastian Longen, Nadia Terziyska, Kai Hell, & Johannes M. Herrmann.** 2008. "The Zinc-Binding Protein Hot13 Promotes Oxidation of the Mitochondrial Import Receptor Mia40." *EMBO Reports* **9**(11):1107–13.
- Milenkovic Dusanka, Kipros Gabriel, Bernard Guiard, Agnes Schulze-Specking, Nikolaus Pfanner, & Agnieszka Chacinska.** 2007. "Biogenesis of the Essential Tim9-Tim10 Chaperone Complex of Mitochondria: Site-Specific Recognition of Cysteine Residues by the Intermembrane Space Receptor Mia40." *Journal of Biological Chemistry* **282**(31):22472–80.
- Minetti Corrado, Rachel M. Chalmers, Nick J. Beeching, Chris Probert, & Kenneth Lamden.** 2016. "Giardiasis." *BMJ (Online)* **355**(October):1–9.
- Moczko M., U. Bömer, M. Kübrich, N. Zufall, A. Hönlinger, & N. Pfanner.** 1997. "The Intermembrane Space Domain of Mitochondrial Tom22 Functions as a Trans Binding Site for Preproteins with N-Terminal Targeting Sequences." *Molecular and Cellular Biology* **17**(11):6574–84.
- Morgan Bruce, Swee Kim, Guanhua Yan, & Hui Lu.** 2009. "Zinc Can Play Chaperone-like and Inhibitor Roles during Import of Mitochondrial Small Tim Proteins." *Journal of Biological Chemistry* **284**(11):6818–25.
- Mossmann Dirk, Chris Meisinger, & F. Nora Vögtle.** 2012. "Processing of Mitochondrial Presequences." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1819**(9–10):1098–1106.
- Nakjang Sirintra, Tom A. Williams, Eva Heinz, Andrew K. Watson, Peter G. Foster, Kacper M. Sendra, Sarah E. Heaps, Robert P. Hirt, & T. Martin Embley.** 2013. "Reduction and Expansion Inmicrosporidian Genome Evolution: New Insights from Comparative

- Genomics." *Genome Biology and Evolution* **5**(12):2285–2303.
- Noinaj Nicholas, Adam J. Kuszak, Curtis Balusek, James C. Gumbart, & Susan K. Buchanan.** 2014. "Lateral Opening and Exit Pore Formation Are Required for BamA Function." *Structure* **22**(7):1055–62.
- O'Neil Patrick K., Lynn G. L. Richardson, Yamuna D. Paila, Grzegorz Piszczek, Srinivas Chakravarthy, Nicholas Noinaj, & Danny Schnell.** 2017. "The POTRA Domains of Toc75 Exhibit Chaperone-like Function to Facilitate Import into Chloroplasts." *Proceedings of the National Academy of Sciences of the United States of America* **114**(24):E4868–76.
- Perry Andrew J., Joanne M. Hulett, Vladimir A. Likić, Trevor Lithgow, & Paul R. Gooley.** 2006. "Convergent Evolution of Receptors for Protein Import into Mitochondria." *Current Biology* **16**(3):221–29.
- Perry S. E., & K. Keegstra.** 1994. "Envelope Membrane Proteins That Interact with Chloroplastic Precursor Proteins." *Plant Cell* **6**(1):93–105.
- Pusnik Mascha, Oliver Schmidt, Andrew J. Perry, Silke Oeljeklaus, Moritz Niemann, Bettina Warscheid, Trevor Lithgow, Chris Meisinger, & André Schneider.** 2011. "Mitochondrial Preprotein Translocase of Trypanosomatids Has a Bacterial Origin." *Current Biology* **21**(20):1738–43.
- Pyrihová Eva, Alžběta Motyčková, Luboš Voleman, Natalia Wandyszewska, Radovan Fišer, Gabriela Seydlová, Andrew Roger, Martin Kolísko, & Pavel Doležal.** 2018. "A Single Tim Translocase in Themitosomes of *Giardia Intestinalis* Illustrates Convergence of Protein Import Machines in Anaerobic Eukaryotes." *Genome Biology and Evolution* **10**(10):2813–22.
- Quandt C. Alisha, Denis Beaudet, Daniele Corsaro, Julia Walochnik, Rolf Michel, Nicolas Corradi, & Timothy Y. James.** 2017. "The Genome of an Intranuclear Parasite, *Paramicrosporidium Saccamoebae*, Reveals Alternative Adaptations to Obligate Intracellular Parasitism." *ELife* **6**:29594.
- Rada Petr, Pavel Doležal, Petr L. Jedelský, Dejan Bursac, Andrew J. Perry, Miroslava Šedinová, Kateřina Smíšková, Marian Novotný, Neritza Campo Beltrán, Ivan Hrdý, Trevor Lithgow, & Jan Tachezy.** 2011. "The Core Components of Organelle Biogenesis and Membrane Transport in the Hydrogenosomes of *Trichomonas Vaginalis*." *PLoS ONE* **6**(9):1–16.
- Reardon Deirdre, & Gregory K. Farber.** 2016. "The Structure of α/β Barrel Proteins." *The FASEB Journal* **9**(7):497–503.
- Renault Marie, Olivier Saurel, Jerzy Czaplicki, Pascal Demange, Virginie Gervais, Frank Löhr, Valérie Réat, Martial Piotto, & Alain Milon.** 2009. "Solution State NMR Structure and Dynamics of KpOmpA, a 210 Residue Transmembrane Domain Possessing a High Potential for Immunological Applications." *Journal of Molecular Biology* **385**(1):117–30.
- Rozewicki John, Songling Li, Karlou Mar Amada, Daron M. Standley, & Kazutaka Katoh.** 2019. "MAFFT-DASH: Integrated Protein Sequence and Structural Alignment." *Nucleic Acids Research* **47**(W1):W5–10.
- Sali Andrej, & Min-Yi Shen.** 2006. "Statistical Potential for Assessment and Prediction of

- Protein Structures." *Protein Science* **15**:2507–24.
- Schein Stanley J., Marco Colombini, & Alan Finkelstein.** 1976. "Reconstitution in Planar Lipid Bilayers of a Voltage-Dependent Anion-Selective Channel Obtained from Paramecium Mitochondria." *The Journal of Membrane Biology* **30**(1):99–120.
- Schulz Georg E.** 2000. "B-Barrel Membrane Proteins." *Current Opinion in Structural Biology* **10**(4):443–47.
- Shoshan-Barmatz Varda, Nurit Keinan, Salah Abu-Hamad, Dalia Tyomkin, & Lior Aram.** 2010. "Apoptosis Is Regulated by the VDAC1 N-Terminal Region and by VDAC Oligomerization: Release of Cytochrome c, AIF and Smac/Diablo." *Biochimica et Biophysica Acta - Bioenergetics* **1797**(6–7):1281–91.
- Sirrenberg Christian, Matthias F. Bauer, Bernard Guiard, Walter Neupert, & Michael Brunner.** 1996. "Import of Carrier Proteins into Mitochondrial Inner Membrane." *Nature* **384**:236–42.
- Söding Johannes, Andreas Biegert, & Andrei N. Lupas.** 2005. "The HHpred Interactive Server for Protein Homology Detection and Structure Prediction." *Nucleic Acids Research* **33**(SUPPL. 2):244–48.
- Stojanovski Diana, Piotr Bragoszewski, & Agnieszka Chacinska.** 2012. "The MIA Pathway: A Tight Bond between Protein Transport and Oxidative Folding in Mitochondria." *Biochimica et Biophysica Acta - Molecular Cell Research* **1823**(7):1142–50.
- Stroud David A., Thomas Becker, Jian Qiu, Diana Stojanovski, Sylvia Pfannschmidt, Christophe Wirth, Carola Hunte, Bernard Guiard, Chris Meisinger, Nikolaus Pfanner, & Nils Wiedemann.** 2011. "Biogenesis of Mitochondrial β -Barrel Proteins: The POTRA Domain Is Involved in Precursor Release from the SAM Complex." *Molecular Biology of the Cell* **22**(16):2823–33.
- Suzek Baris E., Yuqi Wang, Hongzhan Huang, Peter B. McGarvey, & Cathy H. Wu.** 2015. "UniRef Clusters: A Comprehensive and Scalable Alternative for Improving Sequence Similarity Searches." *Bioinformatics* **31**(6):926–32.
- Touw Debra S., Dimki R. Patel, & Bert van den Berg.** 2010. "The Crystal Structure of OprG from *Pseudomonas Aeruginosa*, a Potential Channel for Transport of Hydrophobic Molecules across the Outer Membrane." *PLoS ONE* **5**(11).
- Tovar Jorge, Lidya B. Sa, Robert Sutak, & Gloria Leo.** 2003. "Mitochondrial Remnant Organelles of *Giardia* Function in Iron-Sulphur Protein Maturation." **426**(November):1–5.
- Tsirigos Konstantinos D., Arne Elofsson, & Pantelis G. Bagos.** 2016. "PRED-TMBB2: Improved Topology Prediction and Detection of Beta-Barrel Outer Membrane Proteins." *Bioinformatics* **32**(17):i665–71.
- Tucker Kyle, & Eunyong Park.** 2019. "Cryo-EM Structure of the Mitochondrial Protein-Import Channel TOM Complex at near-Atomic Resolution." *Nature Structural & Molecular Biology* 1–9.
- Vögtle F. Nora, Stefanie Wortelkamp, René P. Zahedi, Dorothea Becker, Claudia Leidhold, Kris Gevaert, Josef Kellermann, Wolfgang Voos, Albert Sickmann, Nikolaus Pfanner, &**

- Chris Meisinger.** 2009. "Global Analysis of the Mitochondrial N-Proteome Identifies a Processing Peptidase Critical for Protein Stability." *Cell* **139**(2):428–39.
- Webb B., & A. Sali.** 2017. "Comparative Protein Structure Modeling Using MODELLER." *Curr Protoc Bioinformatics* **54**(5):1–37.
- Wenz Lena Sophie, Lars Ellenrieder, Jian Qiu, Maria Bohnert, Nicole Zufall, Martin van der Laan, Nikolaus Pfanner, Nils Wiedemann, & Thomas Becker.** 2015. "Sam37 Is Crucial for Formation of the Mitochondrial TOM-SAM Supercomplex, Thereby Promoting β -Barrel Biogenesis." *Journal of Cell Biology* **210**(7):1047–54.
- Wiedemann Nils, Martin Van Der Laan, Dana P. Hutu, Peter Rehling, & Nikolaus Pfanner.** 2007. "Sorting Switch of Mitochondrial Presequence Translocase Involves Coupling of Motor Module to Respiratory Chain." *Journal of Cell Biology* **179**(6):1115–22.
- Williams Bryony A. P., Robert P. Hirt, John M. Lucocq, & T. Martin Embley.** 2002. "A Mitochondrial Remnant in the Microsporidian *Trachipleistophora Hominis*." *Nature* **418**(6900):865–69.
- Wimley William C.** 2003. "The Versatile β -Barrel Membrane Protein." *Current Opinion in Structural Biology* **13**(4):404–11.
- Wu Yunkun, & Bingdong Sha.** 2006. "Crystal Structure of Yeast Mitochondrial Outer Membrane Translocon Member Tom70p." *Nature Structural and Molecular Biology* **13**(7):589–93.
- Xu Dong, & Yang Zhang.** 2012. "Ab Initio Protein Structure Assembly Using Continuous Structure Fragments and Optimized Knowledge-Based Force Field." *Proteins* **80**(7):1715–35.
- Xu Dong, & Yang Zhang.** 2013. "Toward Optimal Fragment Generations for Ab Initio Protein Structure Assembly." *Proteins: Structure, Function and Bioinformatics* **81**(2):229–39.
- Yamashita Satoshi, Petra Lukacik, Travis J. Barnard, Nicholas Noinaj, Suleyman Felek, Tiffany M. Tsang, Eric S. Krukoni, B. Joseph Hinnebusch, & Susan K. Buchanan.** 2011. "Structural Insights into Ail-Mediated Adhesion in *Yersinia Pestis*." *Structure* **19**(11):1672–82.
- Yan Renxiang, Xiaofeng Wang, Lanqing Huang, Feidi Yan, Xiaoyu Xue, & Weiwen Cai.** 2015. "Prediction of Structural Features and Application to Outer Membrane Protein Identification." *Scientific Reports* **5**:1–13.
- Yang Yuedong, Eshel Faraggi, Huiying Zhao, & Yaoqi Zhou.** 2011. "Improving Protein Fold Recognition and Template-Based Modeling by Employing Probabilistic-Based Matching between Predicted One-Dimensional Structural Properties of Query and Corresponding Native Properties of Templates." *Bioinformatics* **27**(15):2076–82.
- Žárský Vojtěch, & Pavel Doležal.** 2016. "Evolution of the Tim17 Protein Family." *Biology Direct* **11**(1):1–13.
- Žárský Vojtěch, Jan Tachezy, & Pavel Doležal.** 2012. "Tom40 Is Likely Common to All Mitochondria." *Current Biology* **22**(12).

Zhang Yuzhe, Yue Li, Xiao Liang, Zhongliang Zhu, Hongbin Sun, Hao He, Jinrong Min, Shanhui Liao, & Yanli Liu. 2017. "Crystal Structure of the WD40 Domain of Human PRPF19." *Biochemical and Biophysical Research Communications* **493**(3):1250–53.

Zimmermann Lukas, Andrew Stephens, Seung Zin Nam, David Rau, Jonas Kübler, Marko Lozajic, Felix Gabler, Johannes Söding, Andrei N. Lupas, & Vikram Alva. 2018. "A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at Its Core." *Journal of Molecular Biology* **430**(15):2237–43.