

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA
ODDELEK ZA BIOLOGIJO

Matic KISOVEC

RAZŠIRJENOST PROTEINSKE DOMENE MACPF

DIPLOMSKO DELO

Univerzitetni študij

Ljubljana, 2012

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA
ODDELEK ZA BIOLOGIJO

Matic KISOVEC

RAZŠIRJENOST PROTEINSKE DOMENE MACPF

DIPLOMSKO DELO
Univerzitetni študij

DISTRIBUTION OF PROTEIN DOMAIN MACPF

GRADUATION THESIS
University studies

Ljubljana, 2012

Diplomsko delo je zaključek Univerzitetnega študija biologije. Opravljeno je bilo na Katedri za biokemijo, Oddelka za Biologijo Biotehniške fakultete.

Študijska komisija Oddelka za biologijo je za mentorja diplomskega dela imenovala prof. dr. Gregorja Anderluha in recenzenta prof. dr. Petra Mačka.

Komisija za oceno in zagovor:

Predsednica: prof. dr. Kristina SEPČIČ
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Član: prof. dr. Gregor ANDERLUH
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Član: prof. dr. Peter MAČEK
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Podpisani se strinjam z objavo svoje naloge v polnem tekstu na spletni strani Digitalne knjižnice Biotehniške fakultete. Izjavljam, da je naloga, ki sem jo oddal v elektronski verziji, identična tiskani verziji.

Naloga je rezultat lastnega raziskovalnega dela.

Datum zagovora: 8.6.2012

Matic Kisovec

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD Dn
DK 577.27:575.112:004 (043.2)=163.6
KG MACPF/MACPF domena/MACPF motiv/bioinformatika/PSI-BLAST/PFT/perforin/MAC/domenska arhitektura
AV KISOVEC, Matic
SA ANDERLUH, Gregor (mentor)/ MAČEK, Peter (recenzent)
KZ SI-1000 Ljubljana, Večna pot 100
ZA Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo
LI 2012
IN RAZŠIRJENOST PROTEINSKE DOMENE MACPF
TD Diplomsko delo (univerzitetni študij)
OP XI, 48 str., 3 pregl., 15 sl., 1 pril., 143 vir.
IJ sl
JI sl/en
AI MACPF domena oz. proteini z MACPF domeno (MACPF proteini) so udeleženi v zelo različnih procesih, kot so: napad in obramba pred drugimi organizmi, ontogenetski razvoj, uravnavanje celičnega cikla, idr. Iz pregledane literature je razvidno, da so proteini, ki vsebujejo MACPF domeno, bolj ali manj splošno razširjeni po vseh kraljestvih živega sveta, z izjemo domene Archaea in skupine virusov. Prve tridimenzionalne strukture celotnega ali dela proteina z MACPF so se pojavile pred petimi leti in do sedaj je znanih sedem struktur (Plu-MACPF, BTH-MACPF, MACPF domena iz C8 α , MACPF domena iz kompleksa C8 α +C8 γ , celotni komponenti C6 in C8 komplementnega sistema ter mišji perforin). Pred časom so odkrili strukturno podobnost MACPF domene z CDC domeno iz CDC toksinov in to je pospešilo raziskovanje MACPF domene. CDC in MACPF domeni imata kljub zelo majhni podobnosti v primarni zgradbi zelo podobno terciarno zgradbo in funkcijo. Obe spadata v razred proteinov, ki tvorijo transmembranske pore (PFT). V diplomski nalogi smo ugotavljali razširjenost proteinske MACPF domene s pomočjo bioinformatičnih orodij in podatkovnih zbirk. Iskali smo s tremi proteinskimi sondami (Plu-MACPF, Bth-MACPF, h-PFN), ki smo jih uporabili za iskanje z metodo PSI-BLAST. Izkazalo se je, da je MACPF domena resnično splošno prisotna pri organizmih iz vseh kraljestev (izjema je Archaea). Daleč največ aminokislinskih zaporedij za MACPF domeno smo našli pri skupini živali (Metazoa) in sicer 522, kar je 64,5% od vseh 809. Našli smo virusno proteinsko zaporedje z MACPF domeno. Vprašanju razširjenosti smo dodali tudi hipotezo o pogosti domenski arhitekturi proteinov z MACPF domeno. V naših podatkih nismo zasledili pogoste kombinacije domen. Preverili smo, ali je MACPF motiv res prisoten pri vseh proteinih z MACPF domeno. Iz naših podatkov smo izluščili dva glicinska (G) ostanka, ki sta izredno dobro ohranjena.

KEY WORDS DOCUMENTATION

DN Dn
DC 577.27:575.112:004 (043.2)=163.6
CX MACPF/MACPF domain/MACPF motif/bioinformatics/PSI-BLAST/PFT/
perforin/MAC/domain organization
AU KISOVEC, Matic
AA ANDERLUH, Gregor (supervisor)
PP SI-1000 Ljubljana, Večna pot 100
PB University of Ljubljana, Biotechnical Faculty, Department of Biology
PY 2012
TI Distribution of protein domain MACPF
DT Graduation Thesis (University studies)
NO XI, 48 p., 3tab., 15 fig., 1 ann., 143 ref.
LA sl
AL sl/en
AB MACPF (Membrane Attack Complex/Perforin) domain and proteins with MACPF domain (aka. MACPF proteins) play a part in different scenarios: attack and defense against other organisms, ontogenetic development, regulation of cell cycle, etc. The MACPF domain is found across the taxonomic system excluding the kingdom of Archaea and the group of viruses. The first three-dimensional structure of a MACPF protein was established about 5 years ago and since then 6 other were reported (Plu-MACPF, Bth-MACPF, MACPF domain from C8 α , MACPF domain from complex C8 α +C8 γ , entire C6 and C8 component of the complement system and murine perforin). Some time ago it became clear that MACPF proteins and CDC proteins share a common fold and also have a very similar function, however, their primary structure is very different. They both belong to a group of pore-forming proteins (PFP). Our goal was to detect as many candidates with the MACPF domain as possible using three different probes (Plu-MACPF, Bth-MACPF, h-PFN) and search through different public databases using PSI-BLAST. As expected we found the MACPF domain scattered across completely different taxonomic groups except Archaea. By far the most sequences with the MACPF domain were found in animals (Metazoa). We found one viral sequence with the MACPF domain. Our interest was also in possible conserved domain organizations. We did not find any within our data. We also examined if the MACPF motif is really conserved among MACPF-containing proteins. It turns out that the motif is (at least partially) well conserved. At least, a two-glycine motif (GX₉G) is universally present in all MACPF proteins.

KAZALO VSEBINE

KLJUČNA DOKUMENTACIJSKA INFORMACIJA	III
KEY WORDS DOCUMENTATION	IV
KAZALO VSEBINE	V
KAZALO PREGLEDNIC	VII
KAZALO SLIK	VIII
KAZALO PRILOG	IX
OKRAJŠAVE	X
1 UVOD	1
1.1 Opredelitev problema	1
1.2 Namen našega dela	2
1.3 Naši hipotezi	2
2 PREGLED OBJAV	2
2.1 Proteinska družina z MACPF domeno	3
2.1.1 Zgradba MACPF domene	4
2.1.2 Razširjenost MACPF domene	6
2.1.3 Perforin	10
2.2 Toksini, ki tvorijo pore (PFT)	13
2.2.1 Delitev PFT	14
2.2.2 Posledice delovanja PFT	14
2.3 CDC toksini, toksini odvisni od holesterola	16
2.3.1 Razširjenost CDC toksinov	19
3 METODE	20
3.1 Iskanje po podatkovnih zbirkah	21
3.2 Obdelava zadetkov	21
4 REZULTATI	23
4.1 Razširjenost proteinske MACPF domene	23
4.2 Domenska arhitektura oz. domenski vzorec	25
4.3 MACPF motiv	28
4.4 Poravnava MACPF domen	29

4.5	Dodatne lastnosti proteinov z MACPF domeno	31
5	RAZPRAVA IN SKLEPI	32
5.1	Razširjenost MACPF domene.....	32
5.2	Domenska arhitektura	33
5.3	MACPF motiv.....	34
5.4	SKLEPI.....	36
6	POVZETEK	36
7	VIRI.....	37
ZAHVALA		
PRILOGA		

KAZALO PREGLEDNIC

Preglednica 1: Pregled proteinov z MACPF domeno..	9
Preglednica 2: Število taksonomskih vrst in število različnih aminokislinskih zaporedij z MACPF domeno.	23
Preglednica 3: MACPF motivi različnih skupin.	28

KAZALO SLIK

Slika 1: Primerjava domnevnih transmembranskih regij MACPF domene.....	3
Slika 2: Filogenetska razmerja med proteini z MACPF domeno in domenska arhitektura.....	5
Slika 3: Primerjava treh proteinov z MACPF/CDC domeno.....	7
Slika 4: Oligomerizacija perforina.	11
Slika 5: Shematski prikaz tvorbe por.....	13
Slika 6: Kristalna struktura štirih MACPF/CDC proteinov.	16
Slika 7: Model perfringolizina O (PFO).	17
Slika 8: Mehanizem tvorbe CDC por.	19
Slika 9: Drevo življenja, ki prikazuje vse vrste, pri katerih smo našli aminokislinska zaporedja z MACPF domeno.	24
Slika 10: Prikaz vrst v kraljestvu Metazoa (Animalia), pri katerih smo našli aminokislinska zaporedja, ki vsebujejo MACPF domeno	26
Slika 11: Prikaz vseh vrst v taksonu Bacteria, pri katerih smo našli predstavnike z MACPF domeno.	27
Slika 12: Prikaz HMM loga MACPF proteinov v delu, kjer se nahaja MACPF motiv.....	29
Slika 13: Poravnava MACPF domene nekaterih ključnih predstavnikov.....	30
Slika 14: Filogenetsko drevo na podlagi poravnave MACPF domen.....	31
Slika 15: Model mišjega (<i>Mus musculus</i>) perforina (PDB ID: 3NSJ).	35

KAZALO PRILOG

Priloga A: Proteinske domene, ki smo jih našli poleg MACPF domene, in se vežejo na sladkorje ali na membrane. Opisi domen so v angleškem jeziku.

OKRAJŠAVE

α -vijačnica	alfa vijačnica
β -lasnica	beta lasnica (ang. β -hairpin)
β -ploskev	beta ploskev
β -trak	beta trak
ALO	antrolizin O (ang. Anthrolysin O)
ASTN1&2	astrotaktin 1&2 (ang. Astrotactin 1&2)
BCL-2	B-celični limfom 2 (ang. B Cell Lymphoma 2)
BRINP-1	nevro-specifični protein 1, ki ga inducira protein BMP/RA (ang. BMP/RA Inducible Neural-specific Protein-1)
Bth-MACPF	<i>Bacteroides thetaiotaomicron</i> -MACPF
CH	skupek α -vijačnic (ang. Cluster of α -Helices)
CDC	od holesterola-odvisni citolizini (ang. Cholesterol Dependant Cytolysins)
DBCCR-1	protein 1, ki se ne izraža v tumorskih celicah mehurja (ang. Deleted in Bladder Cancer Candidate Region-1)
DNA	deoksiribonukleinska kislina
EGF	epidermalni rastni faktor (ang. Epidermal Growth Factor)
EM	elektronski mikroskop
EMBL-EBI	Evropski laboratorij za molekularno biologijo-Evropski inštitut za bioinformatiko (ang. European Molecular Biology Laboratory-European Bioinformatics Institute)
FHL	družinska hemofagocitna limfohistiocitoza (ang. Familial Haemophagocytic Lymphohistiocytosis)
HMM	skrit Markov model (ang. Hidden Markov Model)
ILY	intermedilizin (ang. Intermedilysin)
kDa	kilodalton
LLO	listeriolizin O (ang. Listeriolysin O)
MACPF	kompleks, ki napade membrano/perforin (ang. Membrane Attack Complex/ Perforin)
MPEG-1	Protein makrofagnega gena 1 (ang. Macrophage Expressed Gene-1)
NCBI	Nacionalni center biotehnoških informacij (ang. National Center for Biotechnology Information)
ORF	odprt bralni okvir (ang. Open Reading Frame)
PDB ID	identifikacijska oznaka v podatkovni zbirki PDB (ang. Protein Database Identification)
PF	perforin
PFAM	podatkovna zbirka PFAM (ang. Protein Families)
PFN	perforin

PFP	proteini, ki tvorijo pore (ang. Pore Forming Proteins)
PFR	perforin
PFO	perfringolizin O (ang. Perfingolysin O)
PFT	toksini, ki tvorijo pore (ang. Pore Forming Toxins)
Plu-MACPF	<i>Photorhabdus luminescens</i> -MACPF
PLY	pnevmolizin (ang. Pneumolysin)
PSI-BLAST	pozicijsko specifično ponavljajoče iskanje-iskalno orodje, ki temelji na osnovnih lokalnih poravnava (ang. Position Specific Iterative-Basic Local Alignment Search Tool)
SLO	streptolizin O (ang. Streptolysin O)
SLY	suilizin (ang. Suilysin)
TgPLP1	perforinu podoben protein 1 iz vrste <i>Toxoplasma gondii</i> (ang. <i>Toxoplasma gondii</i> Perforin-Like Protein 1)
TMH	transmembranske vijačnice (ang. Transmembrane Helices)
Tsl	protein podoben proteinu Torso (ang. Torso-like protein)
TSP 1	trombospondin 1 (ang. Thrombospondin-1)
VLIG	zelo velika GTPaza, ki jo lahko inducira IFN (ang. Very Large Inducible GTPases)
VLY	vaginolizin (ang. Vaginolysin)

1 UVOD

Proteinska domena MACPF (ang. Membrane Attack Complex/ Perforin) je relativno slabo raziskana in je v zadnjih letih deležna večje pozornosti. K razumevanju njene funkcije in vloge je veliko pripomogla odkrita strukturna podobnost z bolj raziskanimi CDC toksini. Ti imajo presenetljivo drugačno primarno zgradbo od MACPF molekul. Kljub razlikam v primarnih zaporedjih je funkcija izjemno podobna – tvorba por v bioloških membranah. Proteini, ki tvorijo pore so pogosto virulentni dejavniki različnih patogenih organizmov. Na drugi strani poznamo proteine z MACPF domeno, ki sodelujejo pri imunski obrambi v višjih organizmih. Med MACPF proteini je več predstavnikov z različnimi funkcijami. Te so ali slabo raziskane ali pa sploh niso raziskane. Ta širok nabor funkcij je tudi povod za zelo veliko razširjenost MACPF domene, saj so jo našli pri skoraj vseh taksonih z redkimi izjemami (arheje in virusi). Glede na omenjeno raznolikost v primarni zgradbi in veliko razširjenosti je precej presenetljivo, da naj bi znotraj MACPF domene obstajal ohranjen MACPF motiv. V kolikor obstaja ohranjen motiv - ali obstajajo tudi ohranjene kombinacije in zaporedja domen (domenska arhitektura) znotraj proteinov?

1.1 OPREDELITEV PROBLEMA

Da bi bolje razumeli razširjenost in morebitno vlogo proteinov z MACPF domeno, smo izvedli temeljito iskanje proteinov in organizmov z MACPF domeno. Sodobna bioinformatična orodja nam omogočajo hitro in učinkovito iskanje v obširnih podatkovnih zbirkah, kjer se zbirajo nukleotidna oz. aminokislinska zaporedja. Vedeli smo, da je MACPF domena »zelo« razširjena, vendar nas je zanimala natančnejša slika o njeni pojavnosti.

Če je funkcija proteinov z MACPF domeno tako ohranjena, kot kažejo podatki, potem lahko znotraj teh proteinov pričakujemo tudi kakšen dobro ohranjen odsek. Zato nas je zanimalo ali je opisan MACPF motiv res univerzalen pri vseh predstavnikih.

Proteinske domene lahko sestavljajo t.i. domensko arhitekturo. To gradijo različne domene in zaporedje le-teh znotraj proteinskega zaporedja. Glede na to, da lahko ohranjena domenska arhitektura pri določeni skupini proteinov predstavlja povezavo med domensko arhitekturo in funkcijo, nas je zanimalo, če obstaja kakšna ohranjena domenska arhitektura pri proteinih z MACPF domeno. Primer take povezave med domensko arhitekturo in funkcijo je perforin, ki tvori pore v bioloških membranah. Vsebujejo značilno kombinacijo domene C2 (vezava na membrano) in MACPF domene (tvorba pore). Ohranjena domenska arhitektura bi lahko kazala tudi na morebiten skupen evolucijski izvor proteinov z MACPF domeno.

1.2 NAMEN NAŠEGA DELA

- Temeljito preiskati javne podatkovne zbirke in najti čim več zadetkov/proteinov z MACPF domeno.
- Ugotoviti taksonomsko razširjenost organizmov z MACPF proteini.
- Ugotoviti, ali je MACPF motiv res univerzalen pri vseh aminokislinskih zaporedjih z MACPF domeno.
- S pomočjo bioinformatičnih orodij določiti proteinom z MACPF domeno: topnost, signalna zaporedja, transmembranske regije, proteinske domene in domensko zaporedje/arhitekturo.

1.3 NAŠI HIPOTEZI

- Proteini z MACPF domeno so prisotni v vseh domenah in kraljestvih živega sveta.
- Pri različnih organizmih obstajajo identične domenske arhitekture, ki vključujejo MACPF domeno.

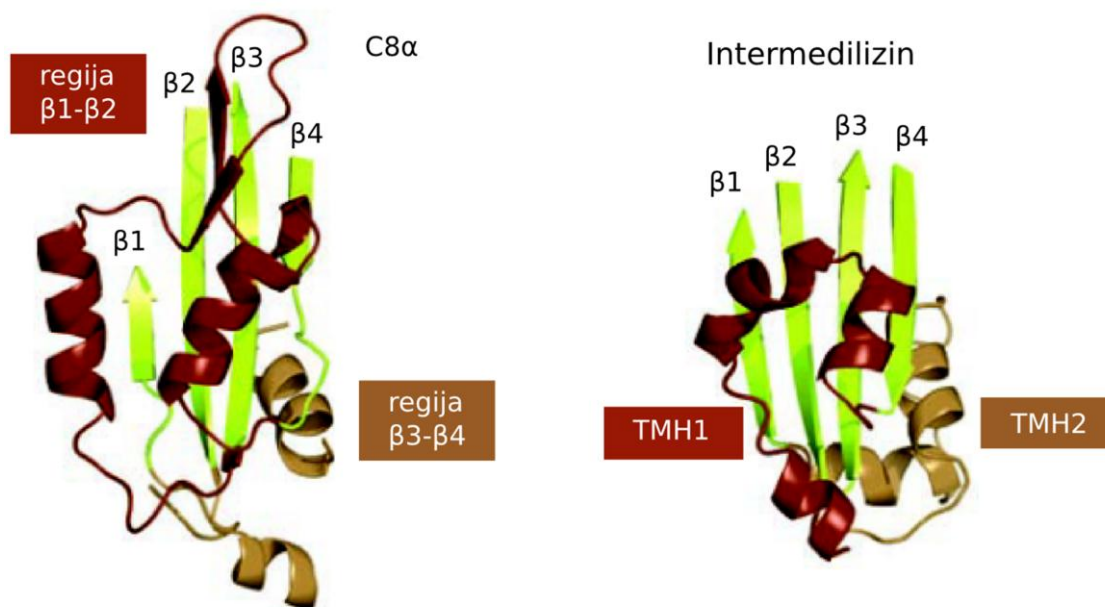
2 PREGLED OBJAV

Družini MACPF in CDC (ang. Cholesterol-Dependent Cytolysins) tvorita MACPF/CDC naddružino. Med družino MACPF in od holesterola-odvisnimi citolizini ni skoraj nikakršne podobnosti v aminokislinskih zaporedjih, zato ju včasih niso imeli za sorodni. MACPF/CDC naddružina je nastala na podlagi funkcionalne in strukturne homologije (Hadders in sod., 2007; Rosado in sod., 2007). Anderluh in Lakey (2008) opišeta tako pojav MACPF/CDC družine kot tudi pojav podobnih primerov združevanja različnih skupin PFP (ang. Pore-Forming Proteins) na podlagi sekundarnih in terciarnih struktur, ki so na voljo v različnih javnih podatkovnih zbirkah.

Za razliko od CDC, ki so večinoma litični toksini, je med MACPF proteini veliko nelitičnih predstavnikov (Rosado in sod., 2008). Vzporednice lahko potegnemo z evkariontskimi proteini BCL-2, ki so strukturno podobni kolicinom iz *Escherichia coli*. BCL-2 se vstavijo v membrano, tvorijo poro vendar ne lizirajo celic tako kot jih kolicini (Muchmore in sod., 1996). Kljub nekaterim razlikam v strukturi pa podobnosti v strukturi in predvsem funkciji kažejo na to, da imata družini MACPF in CDC proteinov skupnega prednika (Kondos in sod., 2010).

Strukturna podobnost MACPF in CDC proteinov temelji na ohranjeni upognjeni centralni β -ploskvi (MACPF/CDC domena), ki je obdana z α -vijačnicami (Slika 1). Primarni zgradbi CDC domene in MACPF domene si delita zgolj 3 ohranjene glicinske ostanke. 95% proteinov z MACPF domeno si med sabo deli 4 glicinske ostanke, vendar so le trije ostanki sorodni glicinskim ostankom iz CDC molekul. Trije ohranjeni glicinski ostanki pri CDC imajo funkcijo pregibnega tečaja. Glede na strukturno podobnost avtorji predvidevajo, da imajo 4 ohranjeni glicinski ostanki pri MACPF podobno funkcijo (Hadders in sod., 2007; Rosado in sod., 2008; Slade in sod., 2008). MACPF domena je

nujna za oligomerizacijo monomerov v predpore. Pred nastankom pore se spremeni konformacija MACPF domene, ki omogoča litično aktivnost proteinov z MACPF domeno (DiScipio in Berlin, 1999; Pipkin in Lieberman, 2007).



Slika 1: Primerjava domnevnih transmembranskih regij MACPF domene in CDC domene. Na levi je del človeškega proteina C8α in na desni del CDC toksina intermedilizina. Dve β-lasnici (ang. β-hairpin) prečita celično membrano. Ena β-lasnica nastane iz temno rdeče obarvanih α-vijačnic (β1-β2 oz. TMH1), druga iz rjavo obarvanih vijačnic (β3-β4 oz. TMH2). Centralni del MACPF domene je obarvan svetlo zeleno. Povzeto po Hadders in sod., 2007.

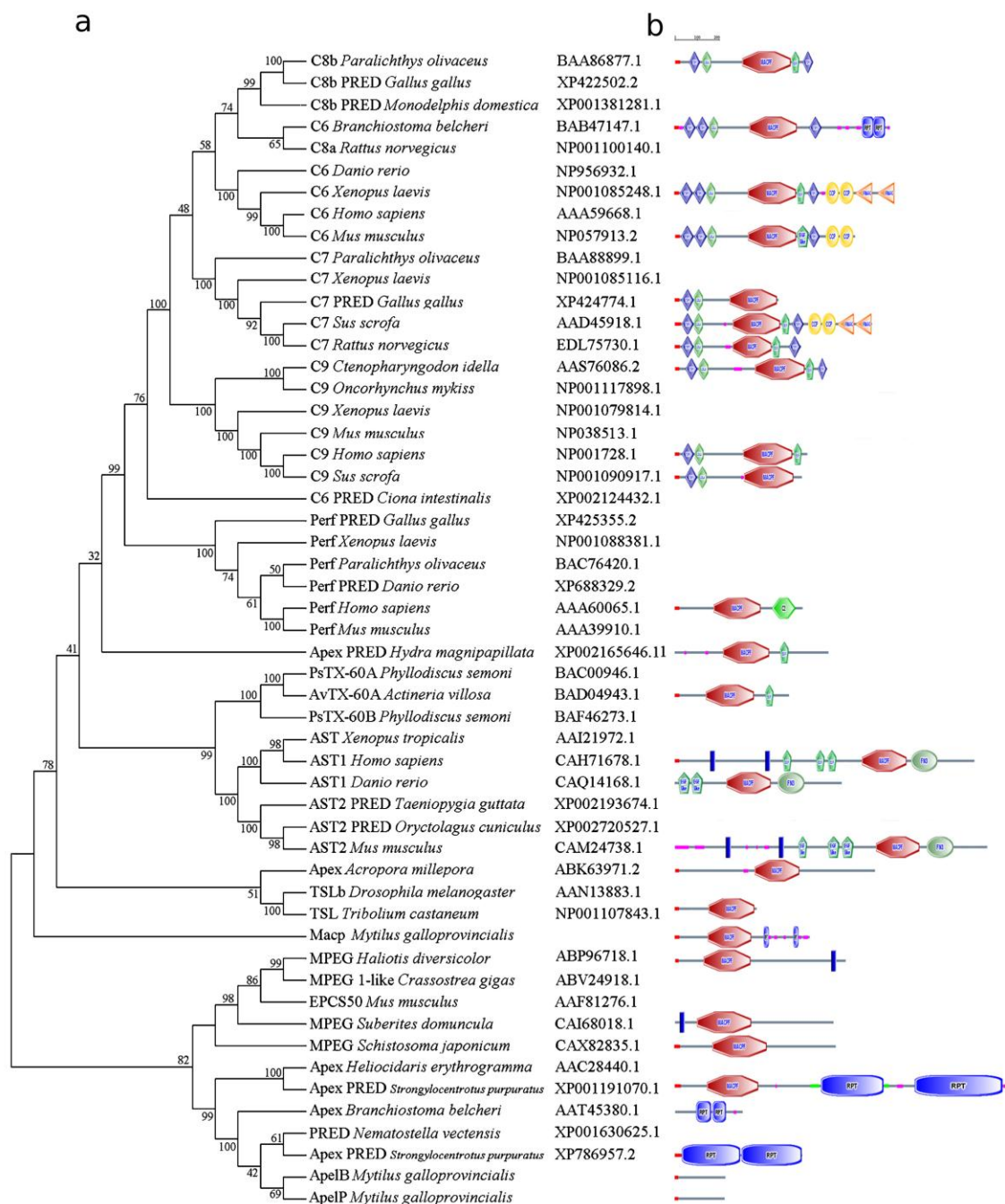
2.1 PROTEINSKA DRUŽINA Z MACPF DOMENO

MACPF družina je največja sesalska družina proteinov, ki se uvršča med proteine, ki tvorijo pore (PFP) (Rosado in sod., 2008). MACPF (ang. Membrane Attack Complex/Perforin) skupina proteinov je ena največjih skupin evkariontskih PFT (ang. Pore-Forming Toxins) molekul in je ime dobila po skupni domeni sesalskega MAC (ang. Membrane Attack Complex, MAC) in perforina (PF) (Lichtenheld in sod., 1988; Shinkai in sod., 1988; Tschopp in sod., 1986). Obe beljakovini imata ključno vlogo pri obrambi pred virusni, bakterijskimi patogeni in pri odstranjevanju tumorskih celic (Natarajan in Cowan, 1998; Smyth in sod., 2000; Voskoboinik in sod., 2006). Perforin je edina molekula, ki omogoča vstop grancima v tarčne celice (Pipkin in Lieberman, 2007). MAC je terminalno stanje komplementnega sistema, ki je del imunske obrambe (Peitsch in Tschopp, 1991).

2.1.1 Zgradba MACPF domene

Primarna zgradba MACPF proteinov je 15% do 20% identična in vsebuje dobro ohranjen motiv (Y/W-G-T/S-H-F/Y-X₆-G-G) (Ponting, 1999; Rosado in sod., 2007; Xu in sod., 2010).

S pomočjo bioinformacijskih orodij so predvideli, da membrano prečkata dve amfipatični α -vijačnici iz MACPF domene (Slika 1). Aminokislinska zaporedja teh vijačnic so najslabše ohranjena v celotni MACPF domeni (Peitsch in sod., 1990). Takrat so pripisali proteine z MACPF domeno (PF in C9) skupini α -PFT (transmembranski del toksinov so α -vijačnice). Novejši podatki kažejo, da MACPF družina najverjetneje sodi med β -PFT toksine, tako kot CDC toksini (Rosado in sod., 2007). Bioinformacijski podatki kažejo na razliko med CDC in MACPF proteini, saj vsi do sedaj opisani CDC proteini vsebujejo C-terminalno Ig domeno. MACPF proteini vsebujejo več različnih C-terminalnih domen. Npr. pri PF najdemo centralno EGF-podobno (ang. EGF-like) domeno, ki ji sledi C2 domena; C8 α vsebuje EGF-podobno domeno in domeno trombospondina 1 (TSP 1); sesalski C9 vsebuje EGF-podobno domeno in pri Plu-MACPF (protein z MACPF domeno iz *Photorhabdus luminescens*) β -prizma (ang. β -prism) domeno (Rosado in sod., 2007). Struktura Plu-MACPF je razkrila, da je β -prizma domena locirana zelo podobno kot Ig domena pri CDC proteinih in verjetno služi za vezavo na membrano. Od kalcija-odvisna C2 domena perforina zaseda podoben položaj v molekuli kot Ig domena in je nujna za interakcijo PF z membrano (Voskoboinik in sod., 2006). Rosado in sod. (2008) predvidevajo, da ima C2 domena podobne naloge kot Ig domena; torej vezava na membrano, povzročitev konformacijske spremembe in vstavitev v membrano. Jasno je, da imajo različne C-terminalne domene MACPF proteinov različne naloge. N- in C-terminalne domene pri C8 α niso potrebne za tvorbo in delovanje MAC kompleksa (Scibek in sod., 2002; Slade in sod., 2008; Hadders in sod., 2007). C-terminalni domeni MACPF proteina iz *Bacteroides thetaiotaomicron* nista podobni nobeni do sedaj znani domeni (ne v primarnem zgradbi ne v strukturi) (Xu in sod., 2010). Estévez-Calvar in sod. (2011) so prikazali nekaj različnih domenskih arhitektur pri proteinih z MACPF (Slika 2). Na isti sliki vidimo eno izmed redkih obsežnejših filogenetskih analiz proteinov z MACPF domeno.



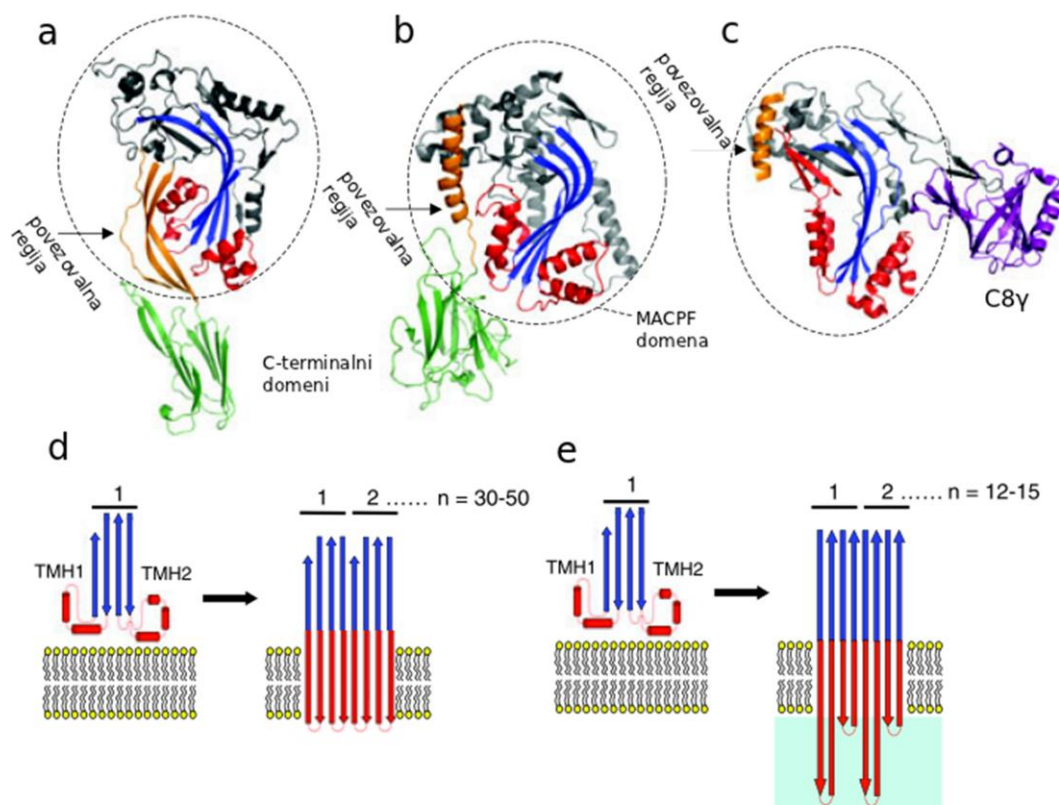
Slika 2: Filogenetska razmerja med proteini z MACPF domeno in domenska arhitektura. **a**, filogenetsko drevo izdelano po metodi Neighbour-joining. ApelB, ApelP in Macp so proteini iz *Mytilus galloprovincialis*, MPEG: gen izražen v makrofagih, EPCS50: Gen, ki sodeluje pri razvoju trofoblata, Apex: apekstrin, Tsl: protein podoben proteinu Torso, AST: astrotaktin, TX-60: toksin morskih veternic. Perf: perforin, C6, C7, C8 in C9: komponente komplekta. **b**, domenska arhitektura nekaterih predstavnikov proteinov z MACPF domeno. Legenda: rdeče območje: signalno zaporedje, rdeči šestkotnik: MACPF domena, pokončni modri štirikotnik: transmembranska regija, roza območje: območja z ponavljajočimi aminokislinami (ang. Low complexity region), zeleno območje: območje ovite vijačnice, vodoravni modri štirikotnik: notranje ponovitve, zelen krog: Fibronektin tip 3, petkotnik: EGF-podoben, šestkotnik: kalcij-vezavna domena, moder romb: TSP 1 ponovitve, zelen romb: LDLR A, rumen krog: kontrolni proteini komplekta, oranžen trikotnik: faktor 1 kompleksa, ki napade membrano (ang. factor I membrane attack complex). Prirejeno po Estévez-Calvar in sod., 2011.

Kristalne strukture (i) proteina Plu-MACPF iz *Photobacterium luminescens* (PDB ID: 2QP2) (Rosado in sod., 2007), (ii) MACPF domene iz C8 α (PDB ID: 2QQH) (Hadders in sod., 2007), (iii) MACPF domene iz kompleksa C8 α +C8 γ (PDB ID: 2RD7) (Slade in sod., 2008), (iv) proteina Bth-MACPF iz *Bacteroides thetaiotaomicron* (PDB ID: 3KK7) (Xu in sod., 2010), (v) celotna struktura C8-komponente komplementnega sistema (PDB ID: 3OJY) (Lovelace in sod., 2011), (vi) celotna struktura C6-komponente komplementnega sistema (PDB ID: 3T5O) (Aleshin in sod., 2012) in (vii) struktura mišjega perforina (PDB ID: 3NSJ) (Law in sod., 2010) so osvetlile strukturno homologijo med MACPF domeno in N-terminalnim delom CDC molekul (domene I-III). Ohranjena je predvsem osrednja β -ploskev in dva skupka α -vijačnic (TMH pri CDC = CH pri MACPF), kar je avtorje pripeljalo do sklepa, da MACPF in CDC najbrž uporabljajo analogen mehanizem za tvorbo por v membranah (Rosado in sod., 2008). Novejši podatki kažejo na to, da najbolj ohranjen del aminokislinskega zaporedja (MACPF motiv (Ponting, 1999)) MACPF molekul sodeluje pri kontroli konformacijskih sprememb in kontroli razvitja α -vijačnic v transmembranske dele, ki tvorijo β -sodček značilen za β -PFT. Dalj časa je veljalo, da ta zelo ohranjeni del sodeluje pri prečanju membrane (Baran in sod., 2009; Rosado in sod., 2008). Pomembna razlika med domeno III iz CDC in MACPF domeno je v dolžini TMH oz. CH regij (Slika 3d, e). Ena CH regija pri MACPF je daljša. Druga pomembna razlika med molekulami CDC in MACPF je različna struktura veznega člana (ang. linker) med centralno MACPF/CDC domeno in C-terminalnim koncem molekul (Slika 3a, b). Pri CDC je to dolga β -ploskev, ki se pri tvorbi pore »prelomi«. Pri MACPF molekulah je na istem mestu α -vijačnica, ki se med tvorbo pore ne spremeni (Tilley in sod., 2005; Voskoboinik in sod., 2010).

2.1.2 Razširjenost MACPF domene

Rosado in sod. (2008) navajajo, da je z uporabo PSI-BLAST (ang. Position Specific Iterative-Basic Local Alignment Search Tool) orodja (Altschul in sod., 1997) mogoče najti več kot 500 proteinov z MACPF domeno. Do sedaj opisani proteini z MACPF domeno so zbrani v Preglednici 1. Nekateri od teh izkazujejo litično aktivnost. Perforin in MAC proteini so dobro raziskani. Hemolitičen toksin, opisan pri dveh morskih vetrnicah (*Phyllodiscus semoni* in *Actinaria villosa*), vsebuje hemolitični MACPF toksin (Nagai in sod., 2002; Oshiro in sod., 2004; Satoh in sod., 2007).

Nekateri do sedaj opisani proteini z MACPF domeno nimajo litične aktivnosti. Protein apekstrin je bil najprej opisan pri morskem ježku (*Heliocidaris erythrogramma*). Raziskovalci so predvideli vlogo apekstrina kot proteina za pritrjevanje celic v embrijih (Haag in sod., 1999). Pred kratkim so pokazali, da apekstrin sodeluje pri antibakterijskemu odgovoru morskih ježkov (Dheilly in sod., 2011). Podobni proteini so bili opisani še pri hidroziju (*Hydra magnipapillata*) in koralih (*Acropora millepora*) (Miller in sod., 2007). Več kot 1000 aminokislinskih ostankov velika proteina astrotaktin 1 in 2 (ASTN1 & ASTN2), igrata pri sesalcih pomembno vlogo pri nevralnem razvoju (Adams in sod., 2002; Zheng in sod., 1996). Znana je domenska arhitektura obeh proteinov ASTN: 3 N-terminalne EGF domene, MACPF domena in C-terminalna fibronektin III (ang. fibronectin III) domena (Kondos in sod., 2010). DBCCR-1 (=DBC1) in homologni BRINP-1 naj bi delovala preko modulacije celičnega cikla kot supresorja tumorske rasti (Kawano in sod., 2004; Motomiya in sod., 2007; Kondos in sod., 2010).



Slika 3: Primerjava treh proteinov z MACPF/CDC domeno. **a**, PFO. **b**, Plu-MACPF. **c**, C8 α (MACPF)-c8 γ (vijolična). Pri vseh treh je MACPF domena obarvana modro, transmembranske regije rdeče, vezni členi oranžno. C-terminalne domene so obarvane zeleno. **d**, Molekularni mehanizem vstavitve MACPF/CDC proteina v membrano. Skupka α -vijačnic TMH 1 in 2 (rdeče) se pred vstavitvijo odvijeta v β -ploskev. β -ploskve večih MACPF/CDC proteinov tvorijo transmembranski β -sodček. **e**, Mehanizem vstavitve proteina C9 v membrano je enak (razvijte-vstavitev-tvorba β -sodčka). Razlika je v daljšem zaporedju TMH1. To zaporedje se po vstavitvi β -lasnice v membrano nahaja na drugi strani membrane (periplazma pri Gram-negativnih bakterijah, citoplazma pri evkariontih (turkizna). Povzeto po Kondos in sod., 2010).

Povzročitelji malarije *Plasmodium* spp. imajo proteine z MACPF, s katerimi lahko napadejo črevesne celice komarja ter v sesalskih jetrih omogočajo prehod preko sinusoidne plasti celic (Ishino in sod., 2004; Kadota in sod., 2004). Protein MPEG-1 vsebuje neznano C-terminalno domeno in naj bi imel vlogo pri citotoksičnosti sesalskih makrofagov. MPEG proteini naj bi imeli antibakterijsko vlogo pri spužvi *Suberites domuncula*. (Kondos in sod., 2010; Estévez-Calvar in sod., 2011). Protein MPEG so našli tudi pri ostrigi (*Crassostrea gigas*) (X. He in sod., 2011) in pri morskem ušesu (*Haliotis midae*) (Kemp in Coyne, 2011). Pri mutantih brez gena *CAD-1* so opazili preobčutljivostni odziv imunskega sistema (Morita-Yamamuro in sod., 2005). Kako pomemben je protein Tsl, pove dejstvo, da mutanti brez Tsl ne preživijo (Martin in sod., 1994; Stevens in sod., 1990). Proteini VLIG so veliki več-domenski proteini pri sesalcih. Pri človeku naj bi bil prisoten le kot psevdogen. Imel naj bi vlogo v imunskem sistemu vendar točna funkcija ni znana (Kondos in sod., 2010).

TgPLP1 iz *Toxoplasma gondii* je virulentni faktor, ki sodeluje pri razvoju pogoste parazitske bolezni – toksoplazmoze (Yan in sod., 2011). Pri rodovih *Babesia* in *Theileria* obstajajo proteinska zaporedja z dvema MACPF domenama (Kafsack in Carruthers, 2010).

Za bakterijske proteine z MACPF domeno zaenkrat še ni jasno, ali sploh tvorijo pore (Rosado in sod., 2008). Bakterijski proteini z MACPF domeno so relativno redki. V literaturi je podatek, da je v podatkovnih zbirkah okoli 40 različnih bakterijskih MACPF proteinov. Rod *Bacteroides* (*Bacteroides thetaiotaomicron* je eden glavnih predstavnikov človeškega črevesnega mikrobioma.) vsebuje približno tretjino vseh znanih bakterijskih MACPF proteinov. Najdeni so bili še pri skupini proteobakterij, cianobakterij in redu Actinomycetales (Xu in sod., 2010). Nekateri avtorji so predvideli možnost horizontalnega prenosa genov za MACPF proteine iz evkariontov v prokariote (Ponting, 1999; Wolf in sod., 1999). Novejše preliminarne filogenetske raziskave kažejo na to, da je res prišlo do neodvisnega horizontalnega prenosa gena pri rodu *Bacteroides*. Avtorji omenjajo možnost, da bakterijski MACPF proteini v človeškem črevesu igrajo pomembno vlogo v simbiotskem odnosu. Lahko bi posredovali pri kakšnem, še neznanem načinu izločanja proteinov, ali pri prehodu hranilnih snovi preko membran. Mogoče z mimikrijo varujejo simbiotsko bakterijo pred imunskim sistemom gostitelja oz. preprečujejo nastanek MAC kompleksa. MACPF protein bi lahko delovali tudi kot bakteriocini proti Gram pozitivnim bakterijam. Možno je tudi, da so bakterijski MACPF proteini virulentni faktorji pri oportunističnih patogenih v črevesju. Združevanje genov za potencialne PFT molekule kaže na možnost, da je prisoten otok patogenosti. Otok patogenosti naj bi izviral iz horizontalnega genskega prenosa (Xu in sod., 2010).

Vloga proteina Plu-MACPF iz bakterije *P. luminescens* (patogena enterobakterija iz črevesja žuželk) kljub znani 3D strukturi še ni jasna (Rosado in sod., 2008).

Patogena družina Chlamydiaceae vsebuje 13 ozko sorodnih MACPF proteinov brez znane funkcije (Ponting, 1999). Ortologe MACPF so našli v različnih vrstah.

Preglednica 1: Pregled proteinov z MACPF domeno. Povzeto in prirejeno po Rosado in sod., 2008.

Skupine MACPF proteinov	Ime proteina	Opis funkcije in izražanja
Naslednje skupine vsebujejo predstavnike z litično aktivnostjo:		
Podobni C9	C6, C7, C8 α , C8 β , C9	Tvorijo kompleks, ki napade membrano (MAC) Vloge znotraj MAC: C9 – vstavitev v membrano, tvorba pore in litična aktivnost C8 α , C7 – zasidra MAC v tarčno membrano C8 β , C6 – do sedaj niso zaznali vstavitve v membrano
Podobni perforinu	perforin	Sproščajo ga naravne celice ubijalke in citotoksični T limfociti tekom imunskega odziva.
Toksini iz morskih vetrnic	PsTX-60A PsTX-60B AvTX-60A	Hemolitični toksin morskih vetrnic, s katerim ubijejo plen. Prisoten pri vrstah <i>Phyllodiscus semoni</i> in <i>Actineria villosa</i> .
Naslednje skupine ne izkazujejo litične aktivnosti:		
Apekstrin	apekstrin	Našli so ga v sekretornih veziklih v jajčecih morskih ježkov. Povišan nivo med bakterijsko okužbo pri škrgoustki.
Astrotaktin	Astrotaktin-1, Astrotaktin-2	Nujno prisoten pri migraciji živčnih celic. Morda omogoča pritrjevanje živčnih celic.
Proteini pri rodu <i>Chlamydia</i>	-	Hipotetični proteini vrst <i>Chlamydia trachomatis</i> , <i>Chlamydophila pneumoniae</i> in <i>Chlamydophila muridarum</i> .
Protein pri cianobakteriji	-	Hipotetični protein pri <i>Trichodesmium erythracum</i> .
DBCCR/BRINP	DBCCR1 (BRINP1), DBCCR1-podoben protein 1 (BRINP3), DBCCR1-podoben protein 2 (BRINP2)	DBCCR-1 je tumor-supresorski gen, ki je pogosto okvarjen pri raku mehurja. Sodeluje pri razvoju živčevja. Skupina proteinov je znana tudi pod imenom FAM5.
EPCS50	EPCS50	Se izraža v mišjem trofoblastu med implantacijo embrija.
Glivni proteini	SpoC1-C1C	Se izraža med dozorevanjem konidijev (organ specializiran za nespolno razmnoževanje) pri vrsti <i>Emericella nidulans</i> .
Malarijski proteini	SPECT2 MAOP	SPECT2 nujen za vstop parazitov v človeška jetra. MAOP nujen za vstop parazitov v gostitelja (komar).
MPS	MPS, MPG Mpeg-1	Našli so ju v diferenciranih makrofagih. Sodeluje pri citotoksičnosti makrofagov. Najden v različnih organizmih.
Rastlinski proteini	CAD1	Sodeluje pri imunskem odzivu vrste <i>Arabidopsis thaliana</i> .
Plu-MACPF	Plu-MACPF	Protein iz bakterijske vrste <i>Photorhabdus luminescens</i> .
TgPLP1	TgPLP1	Virulentni faktor vrste <i>Toxoplasma gondii</i> .
Tsl	Tsl	Protein pri vrsti <i>Drosophila melanogaster</i> . Sodeluje pri določanju anteriorne in posteriorne strani telesa.
VLIG	VLIG	Našli so jih pri sesalcih. Vloga ni znana.

Nedavni rezultati nakazujejo, da protein CT153 (*Chlamydia* spp.) sodeluje pri patogenezi človeka in sicer naj bi bakterije s proteinom CT153 prevzemale/modificirale lipide, ki izvirajo iz gostitelja (Taylor in sod., 2010). Mitchell in sod. (2010) so predlagali zanimiv način uporabe gena za MACPF; gen iz *Chlamydia pneumoniae* bi lahko uporabili za razlikovanje izolatov bakterije iz živali oz. človeka. Izolati iz živali namreč vsebujejo razločno daljše gene za proteine z MACPF domeno kot izolati pri človeku.

V človeškem genomu je zapis za 12 proteinov z MACPF domeno in en psevdogen. Med njimi so najboljše raziskane komponente komplemnega sistema MAC in PF (Kondos in sod., 2010).

Pri nekaterih večjih skupinah kot so virusi, arheje in nematodi, MACPF proteini še niso bili najdeni (Kondos in sod., 2010).

Čeprav MACPF proteini sodelujejo pri tako različnih procesih kot je imunski odziv in razvoj tkiv, pa to ni edini znani primer take raznolikosti. Toll-podobni receptorji (ang. Toll-like receptors) imajo značilno različne vloge pri razvoju muh in sesalski imunosti (Rosado in sod., 2008).

2.1.3 Perforin

Leta 1984 so opisali litični faktor citotoksičnih T limfocitov (ang. cytotoxic T lymphocyte, CTL) in celic naravnih ubijalk (ang. natural killer cells, NK), ki spada v razred PFT beljakovin, imenovan perforin (PF=PRF=PFN) (P. A. Henkart in sod., 1984; Podack in Konigsberg, 1984). Že zgodaj so opazili funkcionalno in strukturno podobnost molekule PF s komponentami komplemnega sistema. Primarno zgradbo PF so določili leta 1988 in takrat je postalo jasno, da je molekula sorodna terminalnim komponentam komplemnega sistema (Lichtenheld in sod., 1988; Shinkai in sod., 1988).

Gen za PF (in gen za MPEG-1) naj bi nastal s podvojitvijo iz MPEG-1-podobnega predniškega gena. Gen za PF je prisoten pri organizmih iz skupine vretenčarjev s čeljustjo (Gnathostomata). Domenska arhitektura PF (MACPF, EGF, C2) ni podobna nobenem drugem proteinu, zato je izvor proteina težko ugotoviti (D'Angelo in sod., 2012).

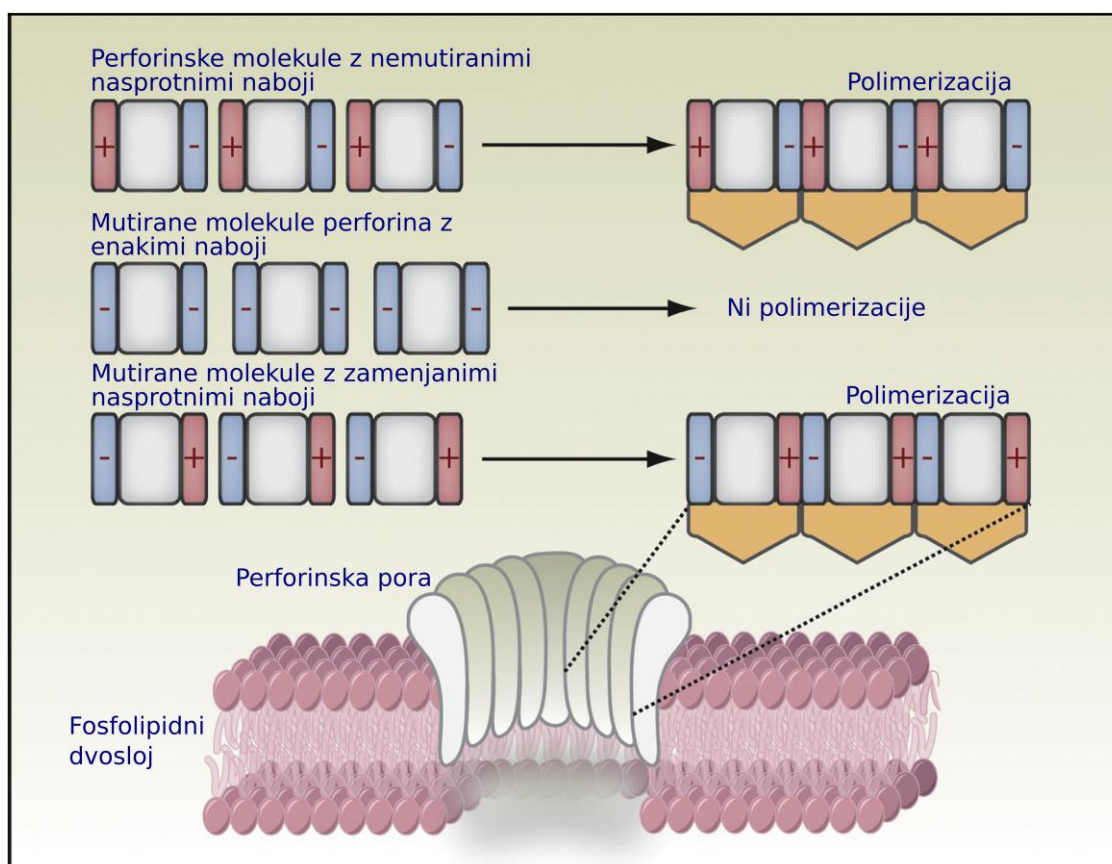
2.1.3.1 Zgradba perforina in perforinske pore

Perforin je sestavljen iz 3 domen: N-terminalne MACPF domene, C-terminalne C2 domene in vmesne centralne EGF domene. Od kalcija-odvisna domena C2 je nujna za interakcijo z membrano in je daljno sorodna z Ig-domeno CDC molekul. Nekaj mutacij povezanih z boleznijo FHL (opisana spodaj) je lociranih na območju EGF regije, ki pa nima jasno določene funkcije (Kondos in sod., 2010). Oligomerizacija in predrtje membrane naj bi poteklo podobno kot pri CDC (Voskoboinik in sod., 2010).

Na podlagi EM posnetkov, znanstveniki predvidevajo, da je orientacija MACPF domene v PF nasprotna orientaciji centralne domene v CDC toksinih (Tilley in sod., 2005; Law in sod., 2010). Vzdolžni prerez PF pore lahko shematsko predstavimo kot »()« in kot »()« (» za CDC molekule, kjer oklepaj prestavlja centralno ukrivljeno β -ploskev in prostor med njima

lumen pore. Pojavila pa se je težava, saj novejši podatki kažejo, da je orientacija pri komponenti kompleksa C8 ») («. Evolucijsko sta PF in C8 precej bližje kot PF in CDC beljakovine. Vse naštetu je znanstvenike pripeljalo do zanimive hipoteze. V opisanem primeru naj bi bila funkcija in struktura evolucijsko ohranjena, način vezave med monomeri pa se je spremenil (Lovelace in sod., 2011).

Perforinsko poro sestavlja od 12 do 18 monomer (Rosado in sod., 2007) in za razliko od MAC kompleksa ne potrebuje pomožnih proteinov za tvorbo por (Slika 4). Hadders in sod. (2007) navajajo, da imajo perforinske pore okrog 20 podenot. Pri tvorbi PF predpore je zelo pomemben naboj aminokislin na robovih PF monomer (Baran in sod., 2009). Vsak monomer ima na robovih nasprotne naboje. Notranji premer pore je približno 14 nm (Kondos in sod., 2010). Pred kratkim je bil objavljen članek, v katerem Praper in sod. (2011) pokažejo, da človeški PF tvori heterogene pore na umetnih lipidnih dvoslojih. Predlagali so tudi možnost povečevanja funkcionalne pore z dodatnimi monomeri. Pokazali so tudi obstoj predpore kot vmesni korak pri tvorbi PF pore. Poleg tvorbe por PF spodbuja tudi nastanek endocitoznih veziklov v tarčni membrani (Praper in sod., 2011).



Slika 4: Oligomerizacija perforina. Slika shematsko prikazuje vpliv nabojev na robovih PF monomer. Interakcija nabojev (modra in rdeča barva) pri nemutiranemu PF sproži oligomerizacijo in konformacijsko spremembo, ki povzroči vstavitve amfipatičnih lasnic (oranžna barva) v membrano. Z mutacijo lahko spremenimo en naboj in do oligomerizacije ne pride (srednja vrsta). Kombinacija dveh mutacij, ki spremeni naboj na obeh straneh, pa ne ovira oligomerizacije (tretja vrsta). Prirejeno po Podack, 2009.

2.1.3.2 Aktivnost perforina

PF je običajno shranjen v citoplazmi v sekretornih veziklih in se sprosti iz celice, ko ta pride v stik z virusno okuženo ali transformirano (tumorsko) celico (Voskoboinik in sod., 2006; J. A. Trapani, 1998; Keefe in sod., 2005; G. Chen in sod., 1995). pH uravnava aktivnost PF (Praper in sod., 2010). Perforin z ustvarjanjem por lizira in ubije celice preko nekroze (pogosteje *in vitro*); drug način delovanja je preko vnosa proapoptotskih proteaz (grancimov) v tarčno celico in ti sprožijo apoptozo (pogosteje *in vivo*) (Shiver in sod., 1992; Bolitho in sod., 2007; Kondos in sod., 2010). Naloga PF je predvsem ubijanje z virusi okuženih celic in nekaterih rodov bakterijskih znotrajceličnih patogenov (*Mycobacterium*, *Brucella*) (Voskoboinik in sod., 2010).

Prirojeno pomanjkanje PF povzroča družinsko hemofagocitno limfocitocitozo (FHL=HLH) (Farquhar in sod., 1958; Stepp in sod., 1999). Gre za redko avtosomalno recesivno bolezen z obolevnostjo 0,12 bolnika na 100 000 ljudi. Prvo znamenje bolezni, na celični ravni, je zmanjšana citotoksična aktivnost limfocitov (Gholam in sod., 2011). Bolezen je običajno smrtna v prvih letih življenja (Voskoboinik in sod., 2004; Voskoboinik in sod., 2005).

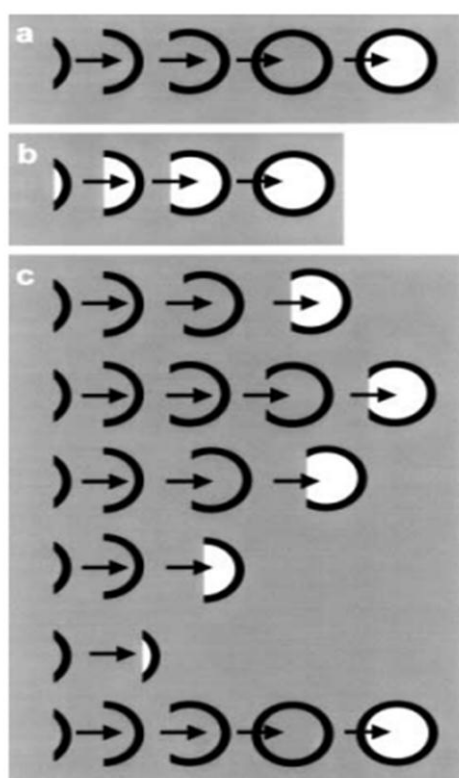
Opisali so tudi preveliko aktivnost PF. Ta privede do ubijanja jetrnih celic pri miši, ki posledično trpi za diabetesom tipa I (Kägi in sod., 1997; Voskoboinik in sod., 2010). PF sodeluje pri poškodbah možgansko-krvne prepreke tekom cerebralne malarije (Kondos in sod., 2010; Voskoboinik in sod., 2010). Opisanih je tudi nekaj kemičnih spojin, ki lahko direktno inhibirajo delovanje PF (Lyons in sod., 2011).

V PF je znanih več različnih mest, kjer se lahko pojavijo škodljive mutacije (Voskoboinik in sod., 2006). Nekatero škodljivo mutacijo se pojavljajo na najslabše ohranjenih delih MACPF domene; v primeru PF je to v transmembranskih regijah CH. To je nenavadno, saj so običajno mutacije na ohranjenih delih proteinov (aktivnih mestih) najbolj škodljive za funkcije proteinov. Rosado in sod. (2008) so predlagali naslednjo razlago: CH regija se med prehodom iz predpore v poro razvije v amfipatično β -ploskev, ki prečka membrano. Za prečenje membrane ni potrebno specifično zaporedje aminokislin, pač pa le poseben amfipatični (hidrofobni) značaj končne β -ploskve. Nekateri MACPF proteini, kot omenjeno, sploh ne ustvarjajo por v membrane. Zato lahko na tem področju opazimo relativno velike razlike med MACPF proteini, pa so ti še vedno funkcionalni, hkrati pa nekatere specifične mutacije porušijo amfipatični značaj CH regije in protein ni funkcionalen. Zanimiv je polimorfizem PF A91V s frekvenco ~3% do ~17% populacije, ki je povezan z napačnim zvijanjem proteina, zmanjšano litično aktivnostjo in poznim nastopom FHL (Trambas in sod., 2005; Voskoboinik in sod., 2007). Opisali so vsaj 50 različnih mutacij PF gena (Gholam in sod., 2011).

V genomu miši in človeka je PF zapisan samo enkrat brez dodatnih kopij. Poskusi na miših so jasno pokazali, da je PF nujen pri različnih procesih in nima funkcijskega dvojnika v celicah. Če prištejemo zraven še vse zgoraj opisane bolezni, ki so direktno povezane s PF lahko vidimo, da ima razumevanje PF potencialno vlogo pri razvoju novih terapevtskih postopkov (Voskoboinik in sod., 2010).

2.2 TOKSINI, KI TVORIJO PORE

Rosado in sod. (2008) navajajo, da je znanih več kot 80 skupin toksinov, ki tvorijo pore. Skupina PFT vsebuje kratke peptidne molekule in velike globularne proteine. PFT predstavljajo približno 1/3 vseh znanih bakterijskih toksinov (Alouf in Popoff, 2006) pri Gram negativnih in pozitivnih vrstah (Aroian in Goot, 2007; Parker in Feil, 2005). Skupina PFT proteinov je ena najbolj raziskanih izmed vseh različnih skupin proteinov, ki se vežejo na membrane. To dejstvo lahko pripišemo naslednjim razlogom: (i) bakterijski izvor, (ii) enostavni za izolacijo in čiščenje ter (iii) enostavna določitev funkcije. Veliko pozornosti pritegnejo tudi zaradi vpletenosti v bolezni (Anderluh in Lakey, 2008).



Slika 5: Shematski prikaz tvorbe por. Površina membrane je v sivi barvi, oligomeri v črni, pore v beli. **a-c**, predlagani modeli oligomerizacije in tvorbe por. **a**, Mehanizem, kjer najprej poteče oligomerizacija in nato nastane pora. **b**, Mehanizem, kjer pora nastane že kmalu po začetku oligomerizacije. Dokler so oligomerni loki nepopolni, jih dopolnjuje lipidni rob na delu, kjer še ni PFT molekul. **c**, Gilbert, 2002 predlaga kombinacijo obeh mehanizmov. Tukaj najprej nastanejo oligomerni kompleksi na površini membrane (popolnoma krožni ali nepopolni loki), nato pa poteče vstavitve in tvorba pore. Povzeto po Gilbert, 2002.

Sposobnost sinteze PFT je evolucijsko dobro ohranjena, saj jih najdemo pri amebah, nematodih, rastlinah (Gonzalez in sod., 2007) in človeku (Hadders in sod., 2007; Rosado in sod., 2007). Predstavnike PFT s podobnimi primarnimi zgradbami združujemo v družine. Zanimivo, da družine pogosto med sabo niso prav nič podobne/sorodne, razen v funkciji, torej tvorbi por v membranah (Iacovache in sod., 2008).

Posebnost vseh predstavnikov iz družine PFT je, da so po sintezi izločeni v zunajcelično vodno okolje, kasneje se vežejo na membrane tarčnih celic in se vstavijo/pogreznejo v membrane kot integralni proteini, ki tvorijo pore. Do vstavitve v membrane pride preko naslednjih stopenj: (i) difuzija do tarčne celice in vezava s specifičnimi receptorji, (ii) večina potem na površini celice oligomerizira. Nazadnje pride do (iii) strukturne preureditve monomer, kar omogoči ugreznitev v membrano in tvorbo pore (Tilley in Saibil, 2006; Iacovache in sod., 2008). Receptorji za PFT so lahko: transmembranski proteini, preko GPI-vezani proteini, lipidi in na lipide vezani sladkorji. Lipidni rafti so pogosto vpleteni in omogočijo koncentriranje monomerov na manjši površini in lažjo polimerizacijo (Bischofberger in sod., 2009; Iacovache in sod., 2010). Vezava na membrano in oligomerizacija običajno poteka v prisotnosti holesterola (ni nujno, glej Caserta in sod., 2008 ter Hotze in Tweten, 2012) oz. lipidnih raftov (Rossjohn in sod., 1997). Vmesno stopnjo oligomerizacije so dokazali pri večini do sedaj znanih PFT (Bischofberger in sod., 2009). Nekateri PFT proteini oligomerizirajo šele, ko je del proteina že zasidran v membrano. Zaenkrat še ni jasno, kakšno je zaporedje oligomerizacije in

kdaj točno se pojavi odprtina (Slika 5); ali se dodajajo posamezni monomeri ali se lahko združijo multimerni intermediati ali se dogaja oboje hkrati (Iacovache in sod., 2010). Pri večini PFT nastanejo s procesom oligomerizacije krožne strukture imenovane predpore (ang. prepore). Predpore so oligomerni kompleksi PFT molekul na površini membrane, ki so sposobni takoj vstopiti v membrano s svojim transmembranskim delom (Hotze in Tweten, 2012). Število povezanih monomer je odvisno od posameznega toksina; opisani so primeri od 7 do 50 monomer v eni predpore. Sledijo strukturne preureditve monomer vendar ta proces, razen pri nekaterih izjemah, še ni podrobno opisan. Različne pore prepuščajo različno velike molekule, od majhnih (npr. ATP) do makromolekul (npr. veliki proteini) (Bischofberger in sod., 2009).

2.2.1 Delitev PFT

Najenostavneje delimo PFT glede na tip strukture, ki tvori transmembranski del proteina. Znani sta dve skupini: α -PFT (transmembranska struktura so α -vijačnice) in β -PFT, kjer je transmembranska struktura β -sodček (Gouaux, 1997; Lesieur in sod., 1997; Feil in sod., 2010). Zanimivo je, da bi do dveh enakih skupin prišli, če bi sklepali glede na prisotnost oz. odsotnost hidrofobnih območij v primarni zgradbi PFT proteinov. Med α -PFT prištevamo npr.: kolicine iz *Escherichia coli*, aktinoporine iz morskih vetrnic, toksin davice iz *Corynebacterium diphtheriae*, eksotoksin A iz *Pseudomonas aeruginosa* in Cry toksine iz *Bacillus thuringiensis*. Večja skupina β -PFT vsebuje npr.: α -toksin in druge toksine iz *Staphylococcus aureus*, Cyt δ -endotoksin iz *Bacillus thuringiensis* in večje število od holesterola-odvisnih citolizinov (CDC) (Iacovache in sod., 2008).

Vloga PFT v organizmu je odvisna predvsem od lokacije in količine toksina. PFT lahko perforirajo skorajda vsako vrsto celice, zato so pomembni pri različnih procesih v zelo različnih organizmih. Med bakterijsko okužbo lahko PFT zmanjšajo imunski odziv gostitelja. Pore v celični membrani lahko mikroorganizmom omogočijo dostop do hranilnih snovi, ki uhajajo skozi pore. Pri mikroorganizmih brez primernih proteinskih sekrecijskih poti, pore omogočijo transport proteinov iz patogena v citosol gostiteljske celice. PFT omogočajo tudi izstop patogena iz fagosoma v citoplazmo, kar je opisano pri bakteriji *L. monocytogenes*, ki uporablja listeriolizin (LLO) za lizo fagosoma. Pri *L. monocytogenes* so ugotovili, da ima LLO tudi vlogo pri podvojevanju mikroorganizma znotraj gostiteljske celice. Pri evkariontih citotoksične T celice imunskega sistema z izločanjem PF ubijejo z virusom okužene celice in druge transformirane celice. PF omogoča vstop grancimov (ang. granzyme) v okužene celice in granicimi povzročijo apoptotično smrt celic. Hepatocite izločajo PFP protein C9, ki je del MAC kompleksa (Bischofberger in sod., 2009). Komponenta C9 komplementnega sistema tvori pore v celičnih membranah in je del prirojenega imunskega sistema (Sodetz in Plumb, 2005).

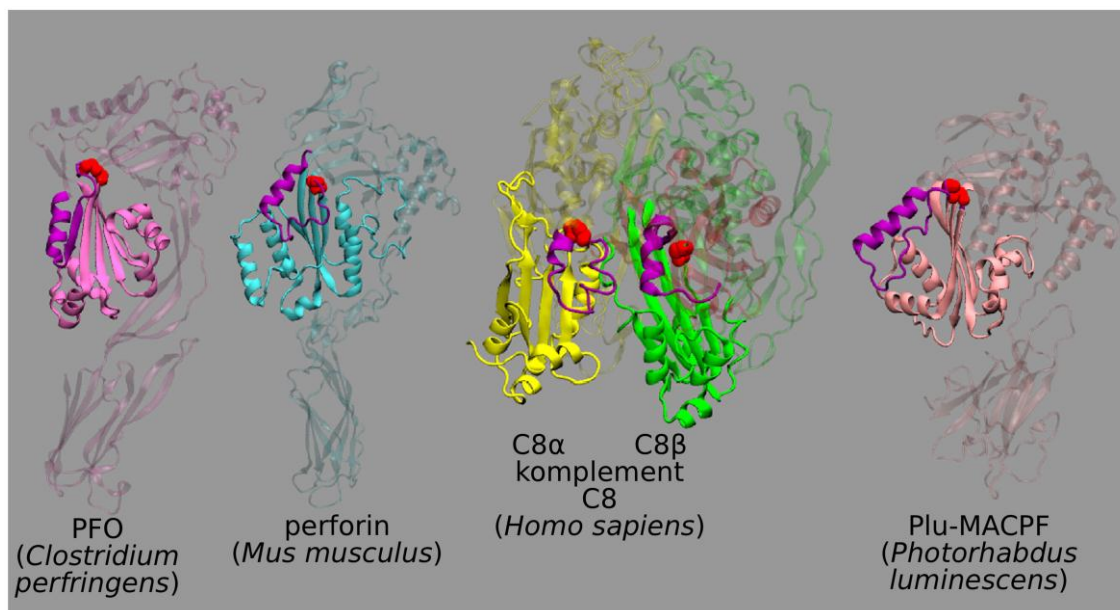
2.2.2 Posledice delovanja PFT

Posledice tvorbe neselektivnih por v membranah so odvisne predvsem od lokacije por. Preluknjanje mitohondrijskih membran poruši mitohondrijski transmembranski potencial, v citoplazmo se sprostijo različne toksične snovi in nazadnje pride do apoptoze celice.

Večina znanih PFT se vstavi v plazemsko membrano in porušijo njeno integriteto. Pore omogočijo nemoten prehod različnih ionov in tako porušijo transmembranski potencial ter spremenijo ionsko sestavo citosola. Negativen vpliv PFT na celice je predvsem posledica spremenjene ionske sestave na obeh straneh membran (Bischofberger in sod., 2009). Visoke koncentracije PFT vodijo do celične smrti (Bhakdi in Trantum-Jensen, 1991; K. L. Nelson in sod., 1999; Golstein in Kroemer, 2007; Kennedy in sod., 2009; Kepp in sod., 2010). Opisali so tudi možno povezavo med PFT proteini in nevrodegenerativnimi boleznimi, kot sta Alzheimerjeva in Parkinsonova bolezen (Bischofberger in sod., 2009). Pojavila se je tudi hipoteza, da se celice lahko odzivajo na nizke (hipolitične) koncentracije PFT (Aroian in Goot, 2007; Bischofberger in sod., 2009) in celo opozorijo druge celice na prisotnost PFT (Yarovinsky in sod., 2008). Gonzalez in sod. (2011) v preglednem članku navajajo več signalnih poti, ki so morda pod vplivom PFT, kot dokaz, da se celice odzivajo na nizke koncentracije PFT.

Rezultat okužbe s PFT ni nujno smrt okužene celice. Gonzalez in sod. (2011) so raziskovali sposobnost preživetja celic po okužbi s PFT. Aerolizin in LLO so uporabili na različnih celicah. Opazovali so njihov odziv in čas trajanja, da celice obnovijo poškodovano membrano. Pokazali so, da je (i) kalij (K^+) glavni regulator odziva na okužbo s PFT; (ii) med okužbo s PFT pride do ustavitve proteinske sinteze in (iii) PFT povzročijo nastanek lipidnih kapljic v citoplazmi. Opazili so tudi avtofagijo, ki so jo opisali že prej (Gutierrez in sod., 2007; Kloft in sod., 2010; Mestre in sod., 2010; Meyer-Morse in sod., 2010), vendar tokrat niso našli nikakršnih dokazov o vlogi avtofagije pri celicah okuženih s PFT. Iz člankov Bischofbergerja in sod. (2009) ter Gonzaleza in sod. (2011) je razvidno, da se mehanizmi obnove celične membrane razlikujejo od toksina do toksina, saj je čas, potreben za popravilo, zelo različen, od nekaj sekund pri mehanskih poškodbah (brez prisotnosti PFT) do vsaj 6 ur pri aerolizinu.

Odpornost celic proti določenem PFT je odvisna od tipa celice, tipa PFT in receptorjev na zunanji površini celične membrane (Gonzalez in sod., 2011).



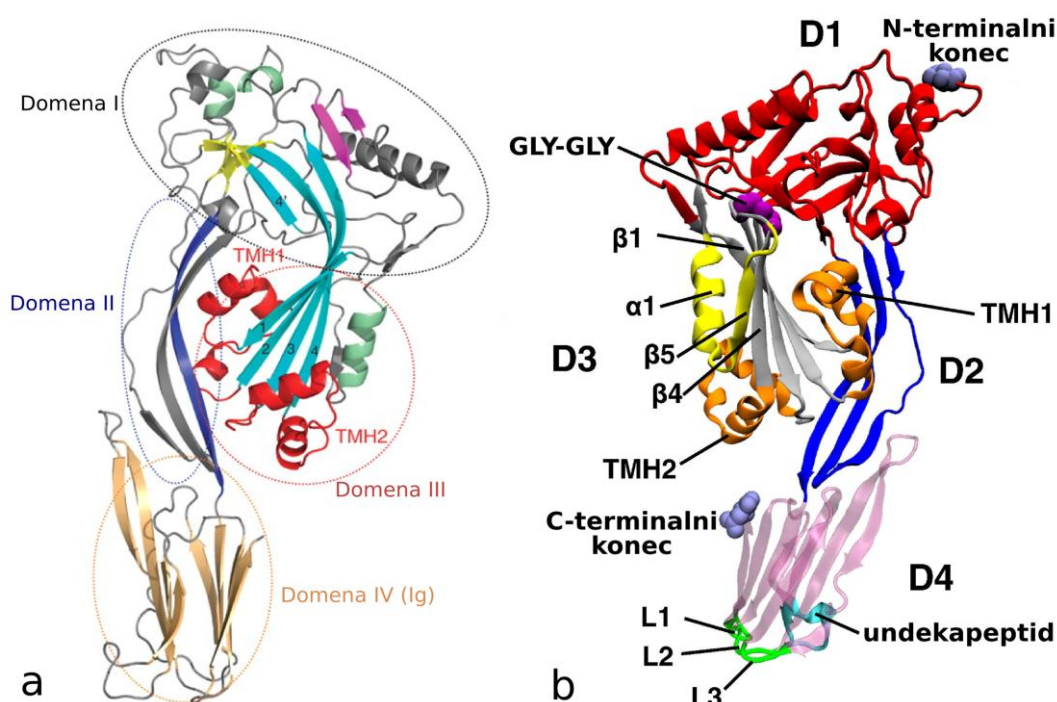
Slika 6: Kristalna struktura štirih MACPF/CDC proteinov. Od leve proti desni si sledijo perfringolizin O, perforin, komponenta komplekta C8 in Plu-MACPF. Osvetljeni deli molekul so CDC/MACPF domene. Z rdečo barvo sta označena glicinska ostanka, ki sta ohranjena pri vseh CDC molekulah. Najdemo ju tudi pri MACPF molekulah. Povzeto po Hotze in Tweten, 2011.

2.3 CDC TOKSINI

Leta 1987 so Walker in sod. objavili prvo primarno zgradbo CDC toksina – pnevmolizina (PLY). Istega leta so določili aminokislinsko zaporedje drugega CDC toksina – streptolizina O (SLO) (Kehoe in sod., 1987). Danes je znanih več kot 25 zaporedij predstavnikov CDC družine. Zaporedja CDC toksinov so med sabo od 40% do 70% identična. Toksini so veliki od 50 do 72 kDa. Centralna domena, ki tvori transmembranski del proteina, je velika približno 50 kDa. Nekateri imajo na N-terminalnem koncu še do 150 aminokislinske dolge podaljške, ki pa so relativno slabo raziskani. Različni avtorji predvidevajo, da podaljški lahko vplivajo na stabilnost in citotoksičnost toksina in v enem primeru celo pomagajo pri transportu nekaterih tretjih molekul skozi membrano. Blizu C-terminalnega konca CDC se nahaja zelo dobro ohranjen motiv enajstih aminokislinske imenovan tudi »undekapeptid« ali »triptofansko bogat motiv« (Slika 7b), ki nima dobro razjasnjene vloge. Pri ILY so Polekhina in sod. (2005) pokazali povezavo med undekapeptidom in zmožnostjo prehoda iz predpore v poro. Primarna struktura CDC toksinov je hidrofilna in nima jasno določenih hidrofobnih odsekov, ki bi lahko prečili membrano (Hotze in Tweten, 2012). Nekateri CDC in MACPF proteini potrebujejo proteolitično cepitev, da postanejo aktivni (Taylor in sod., 2010).

Rosjohn in sod. (1997) so prvi opisali 3D strukturo CDC molekule PFO (Slika 6 prikazuje PFO in še tri molekule z MACPF/CDC domeno), ki je virulentni faktor bakterije *Clostridium perfringens*. Danes so znane kristalne strukture še vsaj treh CDC toksinov: intermedilizina (ILY) iz *Streptococcus intermedius*, antrolizina O (ALO) iz *Bacillus*

anthracis in suilizina (SLY) iz *Streptococcus suis* (Hotze in Tweten, 2012). Nekoč so predvidevali, da je holesterol samo receptor za vse CDC proteine. Kasneje se je izkazalo npr. pri PFO, da je holesterol lahko hkrati receptor in signal za transformacijo iz predpore v poro (Rosado in sod., 2008). Kot smo že omenili, obstajajo vsaj štiri izjeme, kjer vezava in tvorba predpore poteče brez holesterola (SLO, ILY, VLY in LLY). Za ILY so Giddings in sod. (2004) dokazali, da ima glikoproteinski receptor (CD59) in holesterol deluje kot sprožilec prehoda iz predpore v poro. Že pred tem je bilo znano, da ILY specifično lizira samo človeške eritrocite (Hotze in Tweten, 2012).



Slika 7: Model perfringolizina O (PFO). **a**, na modelu so označene domene I-IV. Domene I-III sestavljajo CDC domeno. Domena III je tudi vezni člen med Ig domeno (domena IV) in preostankom molekule. Pri tvorbi pore se domena III "prelomi" (glej shemo na Sliki 2j). Označena sta dela, ki tvorita transmembranski β -lasnici in sicer TMH1 in 2. **b**, Na modelu PFO lahko vidimo: D1-D4, domene I-IV (enako kot pri 7a); TMH1 in 2 (oranžna), transmembranski lasnici; β 1,4,5, β -trakovi MACPF/CDC domene; α 1, α -vijačnica iz MACPF/CDC domene (β 1,4,5 in α 1 sodelujejo pri konformacijskih spremembah med tvorbo pore); zanke L1-L3 in undekapeptid sodelujejo pri interakciji s holesterolom in zasidrajo molekulo; Gly-Gly, glicinski par (ohranjen pri vseh CDC). Povzeto po Rosado in sod., 2007 ter Hotze in Tweten, 2011.

Strukturne raziskave so razkrile tri pomembne strukturne motive v CDC molekulah. Prvi motiv je par glicin-glicin med β 4 in β 5 trakovima v CDC domeni (Slika 7b). Diglicinski motiv pri CDC molekulah deluje kot tečaj (zgib), ki omogoči konformacijsko spremembo molekule in vstavitve v membrano (Rosado in sod., 2007). Zamenjava kateregakoli glicina povzroči inaktivacijo toksina. Tukaj je zanimivo to, da so našli isti motiv tudi v strukturah MACPF proteinov (Hotze in Tweten, 2012). Druga skupina avtorjev poroča tudi o treh ohranjenih glicinskih ostankih pri CDC in MACPF (Rosado in sod., 2007). Drugi motiv je par treonin-levcin, ki naj bi bil esencialen za prepoznavo in vezavo holesterola. Tretji

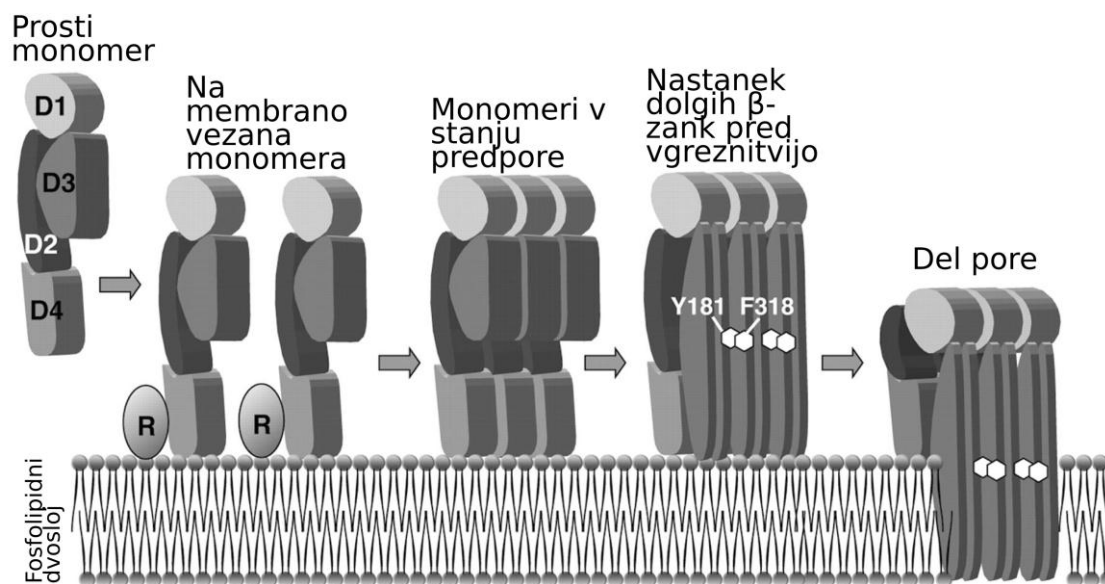
motiv (YXYX₁₄RS) so našli pri ILY in naj bi bil odgovoren za vezavo na CD59 (Hotze in Tweten, 2012).

Ploščata molekula PFO ima N-terminalni konec v obliki kocke (CDC domena = domene I-III na Sliki 7a) in na C-terminalnem koncu vsebuje imunoglobulinsko domeno (Ig domena = domena IV) (Slika 7a). Znotraj CDC domene najdemo centralno β -ploskev iz petih β -trakov, ki je na sredini upognjena za 90°. Na bazi β -ploskve sta dve α -vijačnici, imenovani TMH 1 in 2, dolgi približno 30 aminokislin. CH regiji pri MACPF sta občutno daljši in sicer 58 oziroma 62 aminokislinskih ostankov. Biofizikalne raziskave in opazovanje zamrznjenih vzorcev pod elektronskim mikroskopom (Cryo-EM) so pripeljale do spoznanja, da ti vijačnici po transformaciji tvorita transmembranski del molekule. Vijačnici, ki sami po sebi nista sposobni prečinja membrane, se odvijeta v amfipatični β -lasnici, ki lahko prečita membrano (Hotze in Tweten, 2012; Shepard in sod., 1998; Shatursky in sod., 1999; Tilley in sod., 2005). Zaporedji aminokislin obeh TMH regij sta med najslabše ohranjenimi pri CDC molekulamh. Značilno amfipatično zaporedje β -lasnice, ki je nujna za prehod membrane, je dobro ohranjeno.

Uveljavljen princip vezave, polimerizacije in vstavitve CDC v membrano (Slika 8):

- Prepoznavna in vezava CDC na membrano preko Ig domene (domena IV=D4). V Ig domeni se nahaja »undekapeptid« in zanke L1, L2 in L3. Zanke so glavni faktor pri vezavi na membrano (Hotze in Tweten, 2012). Že leta 1980 so Johnson in sod. pokazali, da je razmerje vezave med holesterolom in CDC 1:1, kar pomeni, da ima ena CDC molekula eno vezavno mesto za holesterol. Kasnejše raziskave so pokazale, da ni ves prisoten holesterol primeren za vezavo CDC. Za učinkovito vezavo CDC toksina morajo imeti receptorske membrane vsaj 30 molarnih % holesterola v membrani. To pomeni, da je samo majhen delež prisotnega holesterola primeren receptor za CDC. Količina razpoložljivega holesterola je odvisna od sestave fosfolipidnega dvosloja; fosfolipidi, ki so bolj skupaj (več nasičenih acilnih verig) preprečujejo dostop do holesterola. Tudi velikost fosfolipidne polarne glave vpliva, saj večji polarni deli sterično ovirajo vezavo CDC na holesterol (Hotze in Tweten, 2012).
- Lateralna difuzija CDC, asociacija z drugimi molekulami CDC in nastanek krožne strukture – predpore. Raziskave so pokazale, da se monomeri povežejo med sabo z robovi centralnih β -ploskev s pomočjo vodikovih vezi (Ramachandran in sod., 2005; Tilley in sod., 2005; Tweten, 2005). Povezava dveh monomer povzroči notranje strukturne spremembe, ki so kasneje nujne za vstavitev TMH regij v membrano (Hotze in Tweten, 2012; Palmer in sod., 1996). Cryo-EM slike razkrivajo mešanico obročastih predpor in nedokončanih predpor v obliki loka. Ali ti nedokončani kompleksi vstavijo svoj β -sodček v membrano, ni znano, vendar avtorji predvidevajo, da ne. Opazovani nedokončani kompleksi so lahko tudi artefakti elektronske mikroskopije. Podatki kažejo, da pore v membrani tvorijo samo največji kompleksi, torej dokončane krožne predpore. Predpore se lahko nahajajo v različnih vmesnih stanjih (Hotze in Tweten, 2012).
- Preko Ig domene se prenese »signal«, ki sproži konformacijsko spremembo CDC domene in še pomembnejše domene II, ki se »prelomi« in omogoči vstop regij TMH 1 in 2 v membrano. Vijačnici TMH 1 in 2 se pred tem razvijeta v β -

lasnici in se vstavita v membrano. Trenutna hipoteza je, da vezava na membrano in interakcije med monomeri povzročijo razvitje α -vijačnice v β -lasnico. Kakšen je signal, ki povzroči vstavitvev toksina v membrano, še ni znano. Omejujoč dejavnik tvorbe pore je oligomerizacija monomer, ko pa je predpora dokončana, se prehod v poro hitro zgodi (Hotze in Tweten, 2012).



Slika 8: Mehanizem tvorbe CDC por. Oznake D1-D4 so enake kot na Sliki 7b. Po vezavi topnih monomerov na površino membrane pride do oligomerizacije, konformacijske spremembe domene D3 in nazadnje konformacijska sprememba D2 omogoči vgreznitev β -lasnic v membrano. Prikazana sta tirozin na mestu 181 (Y181) in fenilalanin 318 (F318). Ta dva aminokislinska ostanka sodelujeta pri usklajenem delovanju monomerov med vgreznjenjem v membrano. R = receptor. Povzeto po Tweten, 2005.

2.3.1 Razširjenost CDC toksinov

Bacillus anthracis - anthrolizin O (ALO), *Clostridium perfringens* – perfringolizin O (PFO), *Streptococcus pneumoniae* – pnevmolizin (PLY), *Streptococcus pyogenes* – streptolizin O (SLO) so nekatere patogene gram-pozitivne bakterije, ki sintetizirajo CDC proteine za lažjo okužbo tkiv in celic. Večina znanih CDC proteinov vsebuje sekretorna zaporedja (Rosado in sod., 2008). Izjema je PLY brez sekretornega signalnega zaporedja, ki se izloča iz celic z neznanim mehanizmom (Balachandran in sod., 2001). Bakterije, ki izločajo CDC, so večinoma ekstracelularni patogeni človeka ali žuželk. Izjemi sta znotrajcelična patogena *Listeria monocytogenes* in *Listeria ivanovii*, ki izločata LLO in ivanolizin (ILO) v citosol okuženih celic (Rosado in sod., 2008).

CDC imajo različne vloge med bakterijsko invazijo: (i) pore povzročajo nekrozo tkiv in tako pospešijo invazijo; (ii) CDC omogočijo izstop patogenov iz fagosomov v citoplazmo okuženih celic; (iii) CDC povzročajo vnetne procese, ki še dodatno pripomorejo k obsegu poškodb tkiva, medtem ko nekateri (iv) CDC lahko povzročijo smrt nevronov, tako da preluknjajo mitohondrijske ovojnice (Rosado in sod., 2008; Cockeran in sod., 2003; Braun in sod., 2007). LLO dodajajo tudi kot proteinski adjuvans cepivom proti tuberkulozi in raku. PFO je eden najbolje raziskanih CDC toksinov in ima pomembno vlogo pri

infekcijah (gangrena). SLO je tudi dobro raziskan in se pogosto uporablja za permeabilizacijo bioloških membran. PLY je pomemben virulentni faktor, ki sodeluje pri invaziji tkiva, deluje kot mediator vnetnih procesov in je aktivator komplementnega sistema (Gilbert, 2010).

Konformacijska mobilnost in spontana oligomerizacija v različno velike komplekse otežuje raziskovanje strukture CDC molekul. Zaenkrat še niso povsem razvozlane podrobnosti molekularnih interakcij med membranskimi receptorji in CDC. Prav tako še niso bile natančno raziskane obsežne spremembe sekundarne in terciarne strukture, ki se zgodijo pri prehodu iz predpore v poro (Rosado in sod., 2008). Niti ni znano, kaj se zgodi s fosfolipidi, ki se znajdejo znotraj bodočega lumna pore, niti kaj sproži in regulira transformacijo iz predpore v poro (Hotze in Tweten, 2012).

3 METODE

3.1 ISKANJE PO PODATKOVNIH ZBIRKAH

>*Photorhabdus luminescens*_MACPF (Plu-MACPF)

```
RKLSLFD DTVIAGAEMIGTSYDVF GK YCNV GSCMNSL FDERKINASEDNFKKVTILGKTLKVPYYIDCY
SVGDLKYTNASGESIESYQSNISSKSRIKGNLYFFSASLKVDFD TDSL TDFENAFSRIQYTYDLYILKS
SAEALKEFLKESVKTALDKADTEEDMNDLFNTW GSHFLSGVVMGGCAQYSSSTNKYTSNL TNSFDVVA
ASFAGFIGLSARTGNSFMEDIKKFRSASNIKTHAIGGDL SRFD PFGGATSADQPSAEEIAAAKKAFEDW
KASVPNAPELVNFADSNPLTGIWELCSDRTQKAKLKKHFETVWAPAESA KR RVHADYIDEIIIGINNTN
TPPEGYIGLKSTKDENLNSKGNICLFMHKAKYDPNIDNKDCITELKFITVRDKSPEGDWWKIPQDIYIS
P
```

>*Bacteroides thetaiotaomicron*_MACPF (Bth-MACPF)

```
FTDATDFLGC SYAVENGTSIIGDFANAKYPVVMKLLERYPSYINPKELRTTETKALSYSDFDRLEKN
KTFTKT VKSGFSLNLGPFKFRQKTIKETFVHNTDDSEKVVHGELSIEV VNGMLNLQTAPSALRKIAAD
YLDEL FVDALYNSSMVELMQSYGEFVLTGYT TGGRASALFYGVDTNSIQFDSKEKMDVA INASYEWN
KKPTAPSDTIHSASGNLSIGTKRENSETITNKFSALSYSIKTLGGAYGYSISTPPYDITNYSIDLTPWL
QSLNDPKTHTMIDLQDGGLYPISDFILEENFKQRYNDTHMDFQYQESL
```

>*Homo sapiens*_PF_MACPF (h-PFN)

```
ARSECKRSHKFVPGAWLAGEVDVTS LRRSGSFVDTQRFLRPDGTCTLCENALQEGTLQRLPLALTNW
RAQSGCQRHVTRAKVSSTEAVARDAARSIRNDWKVGLDVT PKPTSNVHVS VAGSHSQAANFAAQKTHQ
DQYSFSTDTVECRFYSFHVVHTPPLHPDFKRALGDLPHHFNASTQPAYLR LISNYGTHFIRAVELGGRI
SALTALRTCELALEGLTDNEVEDCLTVEAQVNIGIHGSI SAEAKACEEKKKKHKMTASFHQTYRERHSE
VVGHHTSINDLLFGIQAGPEQYSAWVNSLPGSPGLVDYTTLEPLHVLLDSQDPRREALRRALSQYLTD
ARWR
```

Za PSI-BLAST sonde, s katerimi smo preiskali različne podatkovne zbirke, smo izbrali aminokislinska zaporedja MACPF domene treh proteinov z MACPF domeno (Plu-MACPF, Bth-MACPF, h-PFN).

PSI-BLAST iskanje (Altschul in sod., 1997) smo izvedli na straneh NCBI. BLAST je bioinformacijsko orodje (Altschul in sod., 1990), ki s pomočjo algoritmov omogoča primerjavo nukleotidnih zaporedij DNA ali v našem primeru aminokislinskih zaporedij. BLAST se uporablja predvsem za iskanje podobnosti med zaporedji. PSI-BLAST je izpeljanka BLAST algoritma, ki statistično pomembne poravnave združi v matriko oz. profil, s katerim potem izvaja nadaljnja iskanja (ponovitve iskanja ali ang. iterations). Tako

iskanje prinese zadetke, ki so daljno sorodni, vendar še vedno biološko pomembni. Pregledali smo vse podatkovne zbirke na NCBI (nr, refseq_protein, swissprot, pat, pdb, env_nr). Pri iskanju z PSI-BLAST smo ohranili privzete nastavitve; dodatno smo izbrali le filter »Low complexity regions«. Ta filter izloči iz iskanja odseke zaporedja aminokislin, kjer se določene aminokislinske pogosto ponavljajo. Zaradi velikega števila zadetkov smo iskanje razdelili na manjše taksonomske skupine, npr. rastline, glive, primati,... Vsako poizvedbo smo štiri do petkrat ponovili. Izbrali smo zadetke z E-vrednostjo $\leq 0,005$. Manjša kot je E-vrednost, večja je verjetnost, da je zadelek statistično pomemben in se ni pojavil po naključju. Vse zadetke smo zbrali v NCBI zbirki (ang. NCBI Collection), ki omogoča enostavno prenašanje zadetkov iz zbirke na lasten računalnik.

Pri bakterijah smo uporabili še izpeljanko BLAST algoritma imenovano tblastn (Gertz in sod., 2006). Ta algoritem omogoča iskanje nukleotidnih zaporedij s pomočjo aminokislinskega zaporedja, ki se na začetku prevede v vseh šest možnih bralnih okvirjev. Najdene DNA-zadetke smo prevedli v aminokislinsko zaporedje s pomočjo ExPasy Translate Tool (<http://web.expasy.org/translate>). Izbrali smo najdaljši odprti bralni okvir (ang. open reading frame, ORF).

Poleg NCBI podatkovnih zbirk smo z orodjem PSI-BLAST pregledali še UniProt-TrEMBL podatkovno zbirko (Apweiler in sod., 2004) na EMBL-EBI (<http://www.ebi.ac.uk/>). Nazadnje smo ročno pregledali SMART (Schultz in sod., 1998; Letunic in sod., 2011) in PFAM (Punta in sod., 2011) zbirko MACPF proteinov. Naša zbirka vsebuje proteine, ki so bili v naštetih podatkovnih zbirkah dne 24.4.2011.

3.2 OBDELAVA ZADETKOV

Vsa aminokislinska zaporedja smo prenesli v preglednico in s pomočjo poravnave zaporedij določili identična ali zelo podobna zaporedja in jih izločili iz nadaljnje obdelave. Za zelo podobna zaporedja smo smatrali zaporedja, ki so se ujemala v več kot 95% mest. V primerih, kjer smo imeli različno dolga zaporedja in so se krajša zaporedja sto odstotno ujemala z najdaljšim, smo izbrali le najdaljše zaporedje.

Za določitev proteinskih domen smo uporabili spletni strani InterProScan (Zdobnov in Apweiler, 2001) in PFAM. InterProScan smo uporabili predvsem za zaznavanje proteinskih domen. Opisi domen v Prilogi A izhajajo iz strani PFAM in NCBI. Nekatere domene so bile anotirane že v osnovnih zaporedjih, prenesenih iz strežnika NCBI. Transmembranske dele aminokislinskih zaporedij smo iskali z orodji: MPEx (Snider in sod., 2009), SOSUI (Hirokawa in sod., 1998; Mitaku in Hirokawa, 1999; Mitaku in sod., 2002), Tmpred (Hofmann in Stoffel, 1993), ΔG predictor (Hessa in sod., 2007). Ta orodja za iskanje transmembranskih delov iščejo hidrofobne dele zaporedij, ki so običajen znak za prisotnost transmembranskih regij. Za določanje topnosti smo uporabili PROSO II (Smialowski in sod., 2007). PROSO II določi topnost oz. netopnost aminokislinskega zaporedja na podlagi dostopnih eksperimentalnih podatkov v podatkovnih zbirkah. Za določitev signalnih zaporedij smo uporabili SignalP (Dyrløv Bendtsen in sod., 2004; H Nielsen in sod., 1997) in TargetP (Emanuelsson in sod., 2007). SignalP in TargetP, delujeta na podlagi eksperimentalnih podatkov. Raziskovalci so ročno pregledali določen vzorec proteinov in podatke o vsebnosti signalnih zaporedij. Na podlagi tega vzorca so izdelali algoritme, ki lahko predvidevajo ali se v določenih zaporedjih pojavijo signalna zaporedja.

Posebnost je, da so algoritmi sposobni učenja in se prilagajajo novim eksperimentalnim podatkom, ki jih razvijalci vnesejo.

Poravnave celotnih aminokislinskih zaporedij in BLOCK poravnave za iskanje motivov smo naredili s programsko opremo Vector NTI (G. Lu in Moriyama, 2004). Poravnave celotnih zaporedij smo uporabili za začetno izločitev identičnih in zelo podobnih zaporedij. BLOCK poravnave uporabljamo za iskanje kratkih neprekinjenih in evolucijsko dobro ohranjenih zaporedij, ki se ponovijo v več aminokislinskih zaporedjih.

Za poravnavo MACPF motivov smo s pomočjo funkcije Najdi motiv (ang. Find motif) v programu Vector NTI poiskali 19 MACPF motivov pri različnih organizmih. Dodatno smo vključili 50 aminokislinskih ostankov pred in za motivom. Poravnavo MACPF smo izvedli na spletni strani M-COFFEE (<http://tcoffee.crg.cat/apps/tcoffee/do:mcoffee>) (Di Tommaso in sod., 2011; Moretti in sod., 2007; Notredame in sod., 2000; Wallace in sod., 2006). Poravnave zaporedij z MACPF motivom smo izvozili iz programa GeneDoc (Nicholas in sod., 1997). Filogenetsko drevo smo izdelali z orodjem »protpars« po načelu varčnosti (ang. parsimony) na spletni strani Mobylye (<http://mobylye.pasteur.fr>) (Néron in sod., 2009; Han in Zmasek, 2009). Uporabili smo privzete nastavitve in dodatno izbrali test samovzorčenja (ang. Bootstrap). Sheme z rezultati (Slika 9, 10, 11) in Slika 15 smo narisali s programom Inkscape Vector Graphics Editor.

Podatke za HMM logo smo prenesli na spletne strani podatkovne zbirke PFAM (<http://pfam.sanger.ac.uk/family/PF01823#tabview=tab6>). HMM logo je bil izdelan s pomočjo spletnega orodja LogoMat-M (Schuster-Böckler in sod., 2004).

Model molekule mišjega perforina smo izdelali s pomočjo orodja Visual molecular dynamics (VMD) (Humphrey in sod., 1996). Aminokislinsko zaporedje (PDB ID: 3NSJ) smo pridobili iz podatkovne zbirke PDB (Berman in sod., 2000).

4 REZULTATI

4.1 RAZŠIRJENOST PROTEINSKE MACPF DOMENE

Preglednica 2: Število taksonomskih vrst in število različnih aminokislinskih zaporedij z MACPF domeno.

	Število taksonomskih vrst ($\Sigma=192$)	Število različnih aminokislinskih zaporedij ($\Sigma=809$)
Eukarya	142 (74%)	741 (91,6%)
Metazoa	87 (45,3%)	522 (64,5%)
Chordata	54 (28,1%)	453 (56%)
Vertebrata	49 (25,5%)	399 (49,3%)
Mammalia	28 (14,6%)	259 (32%)
Eutheria	26 (13,5%)	236 (29,1%)
Osteichthyes	14 (7,3%)	97 (12%)
Ostali Vertebrata	7 (3,6%)	43 (5,3%)
Ostali Metazoa	33 (17,2%)	69 (8,5%)
Fungi	24 (12,5%)	36 (4,4%)
Plantae	16 (8,3%)	82 (10,1%)
Ostali Eukarya	15 (7,8%)	101 (12,5%)
Bacteria	49 (25,5%)	67 (8,3%)
Bacteroidia	30 (15,6%)	48 (5,9%)
Virus	1 (0,5%)	1 (0,1%)

Našli smo 809 različnih proteinskih zaporedij z MACPF domeno pri 192 taksonomskih vrstah (Preglednica 2). Med evkarionti je 142 vrst (74% vseh) s 741 proteinskimi zaporedji (91,6% vseh) in med prokarionti 49 vrst (25,5%) z 68 proteinskimi zaporedji (8,3%). Našli smo en virusni protein z MACPF domeno pri Anguillid herpesvirus 1.

Daleč največ vrst in unikatnih zaporedij prihaja iz živalskega kraljestva (Metazoa). 26 vrst prištevamo v skupino placentalnih sesalcev (Eutheria) in pri tej skupini je bilo najdenih 236 zaporedij (29,1%). V kraljestvu rastlin (Plantae) smo našli 16 vrst z 82 unikatnimi proteini. Večinoma gre za rastline v klasičnem pomenu z izjemo ene vrste mahu, ene zelene alge in praproti podobne »drežice«. MACPF domena se pojavi še pri skupinah migetalkarjev (Ciliophora) in različnih praživalih (Apicomplexa, Amoebozoa, Heterolobosea).

Pri prokariontih smo našli manj zadetkov in sicer 67 zadetkov pri 49 vrstah. Večina mikroorganizmov spada v debli Chlamydiae in Proteobacteria ter razred Bacteroidia. Med bakterijskimi rodovi z MACPF najdemo: organizme črevesnega mikrobioma človeka (*Alistipes*, *Bacteroides*), organizme iz morja (*Dokdonia*, *Leeuwenhoekiella*, *Plesiocystis*, *Zunongwangia*), organizem iz kopenskih vod (*Beggiatoa*), organizem izoliran v močvirju (*Mucilaginibacter*), rastlinska patogena (*Clavibacter*, *Ralstonia*), bioluminiscentnega patogena insektov (*Photorhabdus*), človeške patogene (*Chlamydia*, *Chlamydophila*, *Chryseobacterium*, *Filifactor*, *Porphyrromonas*), cianobakterijo (*Trichodesmium*), zeleno žveplovo bakterijo (*Chlorobium*) idr.

Mediana dolžine proteinskih zaporedij v naši zbirki je 590 aminokislinskih ostankov. Najdaljše zaporedje z MACPF domeno je dolgo 2374 oziroma najkrajše 75 aminokislinskih ostankov.

4.2 DOMENSKA ARHITEKTURA OZ. DOMENSKI VZOREC

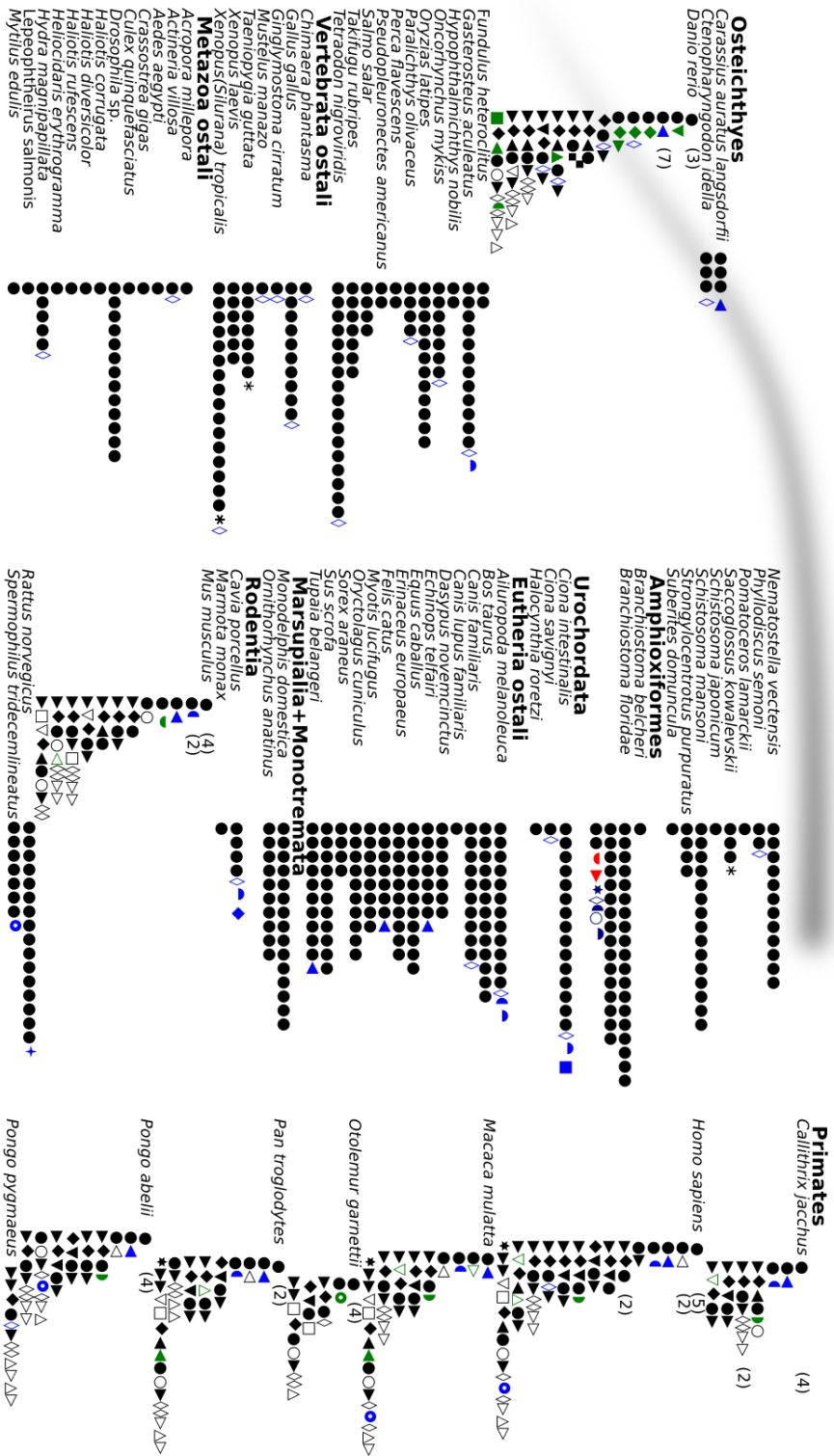
Domensko arhitekturo smo določili za red Primates (7 taksonomskih vrst), vrsti *Mus musculus* (hišna miš) in *Danio rerio* (navadna cebrica) iz kraljestva Metazoa. Pri omenjenih podrobneje obdelanih vrstah (109 aminokislinskih zaporedij) nobena domenska arhitektura ni prisotna pri vseh. Največkrat se ponovi samostojna MACPF domena (26 zaporedij, 23,9%) in pa kombinacija MACPF domene in C2 domene (13 zaporedij, 11,9%). Oba primera se pojavita pri sedmih vrstah. Domena TSP 1 je pri podrobneje obdelanih proteinih pri skupini Metazoa največkrat na prvem mestu (37 primerov od 109, 33,9%). Pri skupini Metazoa smo našli 39 različnih domen (Slika 10) poleg MACPF domene.

Na veliko raznolikost domenske arhitekture kažejo podatki podatkovne zbirke PFAM, kjer je zabeleženih 48 različnih arhitektur. Na strani podatkovne zbirke PFAM močno prevladujejo predstavniki s samostojno MACPF domeno (486 od skupno 720 zaporedij, 67,5%). Pri naših rezultatih je delež veliko manjši, kar gre verjetno na račun veliko manjšega vzorca podrobno obdelanih predstavnikov. Naše subjektivno opažanje je tudi, da je PFAM algoritem bolj specifičen in prepozna le domene z majhno E-vrednostjo. Z uporabo drugih orodij (InterProScan) smo namreč pogosto dobili še dodatne domene.

Različne najdene domene so v literaturi boljše ali slabše opisane, večina pa ni podrobno raziskanih. Našli smo 61 različnih proteinskih domen. Domene, ki se vežejo na biološke membrane in domene, ki vežejo sladkorje smo zbrali v Prilogi A.

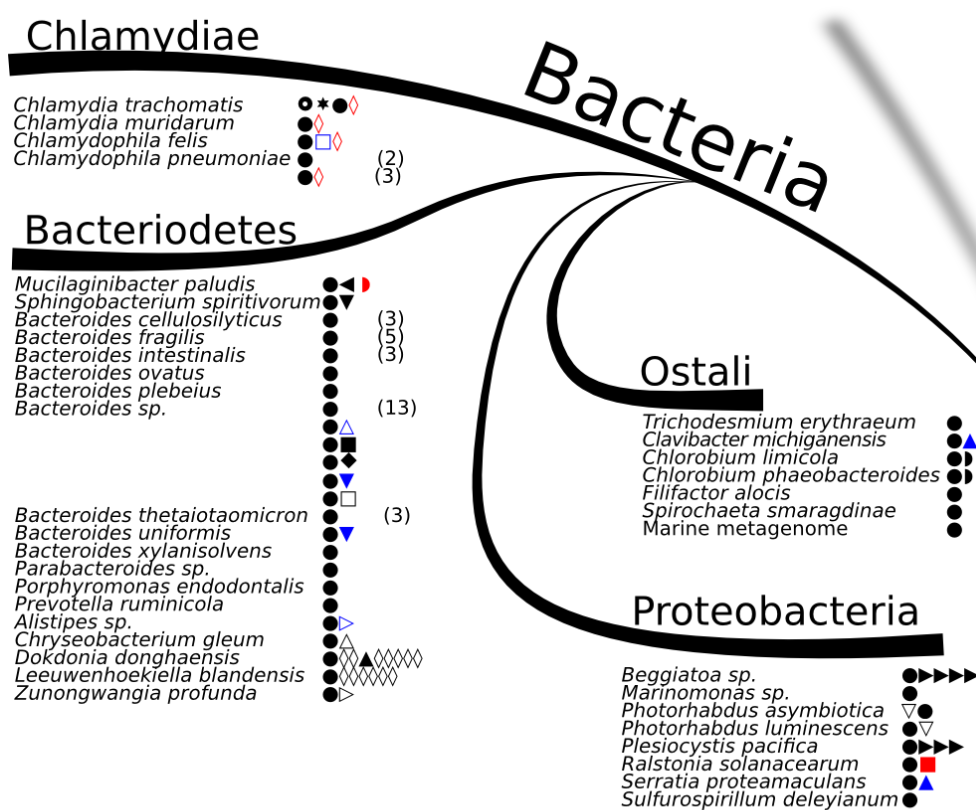
Določili smo tudi domensko arhitekturo pri vseh predstavnikih kraljestva Bacteria z najdeno MACPF domeno (67 zaporedij; 8,3%). Domenska arhitektura bakterijskih proteinov se razlikuje od ostalih predvsem v vrsti, številu in zaporedju domen v proteinih (Slika 11). Pri skupini Bacteria smo našli popolnoma druge domene kot pri skupini Metazoa. Število različnih domen v posameznem proteinu je očitno manjše pri skupini Bacteria. Pri skupini Bacteria smo našli 22 različnih domen poleg MACPF. Zanimivo je, da je pri bakterijah MACPF domena skoraj vedno tista, ki je najbližja N-terminalnem koncu proteinskega zaporedja (65 primerov, 97%). Pri živalih je MACPF 59-krat (54,1%) na prvem mestu.

Metazoa



Slika 10: Prikaz vrst v kraljestvu Metazoa (Animalia), pri katerih smo našli aminokislinska zaporedja, ki vsebujejo MACPF domeno. Posebej razčlenjen je red Primates in vrsti *Danio rerio* in *Mus musculus*. Pri nerazčlenjenih skupinah znaki ● prikazujejo koliko različnih aminokislinskih zaporedij, z MACPF domeno, smo našli pri posamezni vrsti. Števila v oklepajih povejo, kolikokrat se določena domenska arhitektura ponovi pri dotični vrsti. Pri nerazčlenjenih vrstah so poleg MACPF domene označene še domene, ki lahko interagirajo z sladkorji oz. membrani. Te so razporejene naključno. Legenda: ●-najdeno aminokislinsko zaporedje, ki vsebuje MACPF domeno, *-primer zaporedja z dvema MACPF domenama, ◊-končasti rahlovirus glikoprotein, ◀-Ca vezavni protein, □-EIB modul, ▷-faktor I MAC, △-Kazal, ▽-Keratin B2, ○-laminin/EGF naddružina, ◀-NOL1 NOP2 SUN, podvrste EGF, ◀-Giardia-specifični površinski protein, ○-receptor rastnega faktorja, ◀-metalionin, ■-bakteriocin II, ◀-C2 perforinska domena, ○-domena hormonskega receptorja, ◀-končasti rahlovirus glikoprotein, ◀-Ca vezavni protein, □-EIB modul, ▷-faktor I MAC, △-Kazal, ▽-Keratin B2, ○-laminin/EGF naddružina, ◀-NOL1 NOP2 SUN, ◀-LDLR A, ★-PTN MK heparin vezavna domena, ▼-Squash, ◊-SUSHI, ▶-TSP 1, ◀-ADP-ribozil ciklaza, ▲-Kollagen, ▼-disintegrin, ◀-EEFV, ▶-eIF5 eIF2B, ▷-fibronektin 3, △-GATaza 2, ○-MATH, ◀-območje interakcij receptor-ligand, ■-reverzna transkriptaza, ▶-RGS, ○-spektrin, □-transketolaza, ▼-VSP, ▽-zf-BED

Očitna razlika med živalskimi in bakterijskimi predstavniki je v številu MACPF zaporedij, pri posamezni vrsti. Za razliko od živali, je pri bakterijah običajno prisotno le eno zaporedje z MACPF domeno pri posamezni vrsti. Edina izjema je vrsta *Chlamydomphila pneumoniae* z dvema različnima zaporedjema. Število zaporedij z MACPF pri podrobneje obdelanih živalih je od 1 do 20 (povprečno ~10). Različno je tudi število domen, ki so prisotne poleg MACPF. Število dodatnih domen pri bakterijah je največ štiri. Med 67 bakterijskimi aminokislinskimi zaporedji obstajata dve izjemi, ki imata več kot 4 domene (6 oz. 8 dodatnih domen).



Slika 11: Prikaz vrst v taksonu Bacteria, pri katerih smo našli predstavnike z MACPF domeno. Poleg vrst je, s simboli, prikazana domenska arhitektura za posamezno vrsto. V kolikor je več različnih arhitektur pri eni vrsti so različne arhitekture prikazane vsaka v svoji vrstici. Število v oklepaju pove kolikokrat se taka domenska arhitektura pri tej vrsti pojavi. ●-najdeno aminokislinsko zaporedje, ki vsebuje MACPF domeno, vezava na sladkorje (◇-MIR domena, ■-"jacalin"-podobna lektinska domena, ▶-B-lektin), vezava na membrane (▼-integrin alfa N-terminalna domena, △-GRAM, □-Xlink, ▲-MABP, ▷-ENV poliprotein), ▶-podobna kalcij-vezavni domeni, ●-FolB, ★-fosducin, ◀-koprogen oksidaza, ▼-BH4, ■-Tub, ◆-RepB, □-SHMT, △-NAQ-podobna domena, ◇-Ankirinske ponovitve, ▲-Rep, ▷-SecA, ▶-Hemopeksin, ▽-GA

4.3 MACPF MOTIV

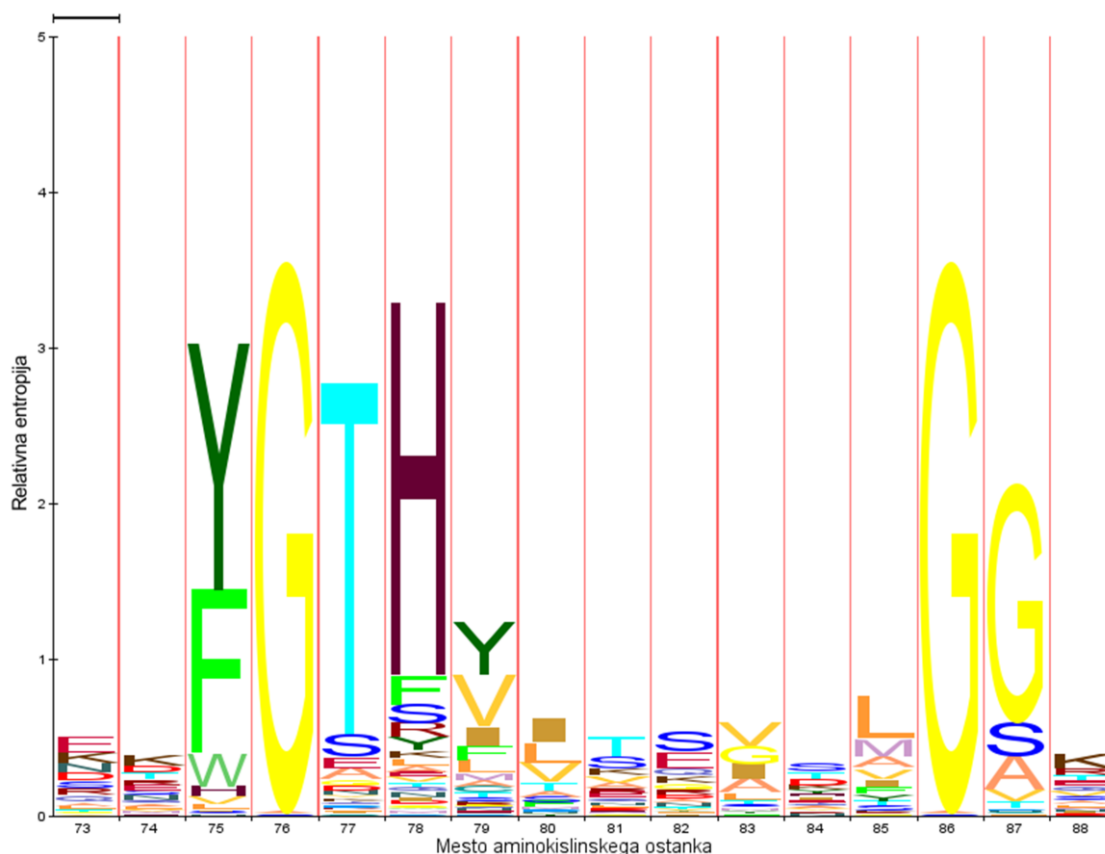
Iz vzorca, približno 410 zaporedij z MACPF domeno, smo izluščili motive za posamezne skupine organizmov (Preglednica 3). Naši podatki kažejo dva izjemno ohranjena glicinska ostanka na mestih 2 in 12 v MACPF motivu. Ostanke sta ohranjena pri prav vseh organizmih v vseh kraljestvih, kjer je bila najdena MACPF domena. Mesta 3, 4 in 13 so tudi dokaj dobro ohranjena, druga mesta v MACPF motivu pa slabše. Pri določenih skupinah organizmov, npr. mesto 4 pri skupini Metazoa in Plantae ter mesto 13 pri skupinah Plantae in Bacteria, pa so zelo dobro ohranjena.

Preglednica 3: MACPF motivi različnih skupin. Podčrtani aminokislini se ponovita pri vseh skupinah. Vrstica MACPF motiv predstavlja MACPF motiv, kot je zapisan v literaturi. Simbol / pomeni »ali«. Simbol X predstavlja katerokoli aminokislino. Legenda barv: Metazoa (zelena), oker (Plantae), turkizna (Bacteria), rdeča (Virus).

MACPF motiv (Ponting, 1999)	Položaj znotraj MACPF motiva	1	2	3	4	5	6	7	8	9	10	11	12	13
	Y/W-G-T/S-H-F/Y-X6-G-G	Y/W	<u>G</u>	T	H	F/Y	X	X	X	X	X	X	<u>G</u>	G
Apicomplexa	X-G-T-H-X7-G-G	X	<u>G</u>	T	H	X	X	X	X	X	X	X	<u>G</u>	G
Ciliophora	F/W-G-T-H-X7-G-G/S	F/W	<u>G</u>	T	H	X	X	X	X	X	X	X	<u>G</u>	G/S
Amphioxioformes	X-G-T-H-F/Y-X6-G-G/R	X	<u>G</u>	T	H	F/Y	X	X	X	X	X	X	<u>G</u>	G/R
Urochordata	X-G-T-H-F/Y-X6-G-G	X	<u>G</u>	T	H	F/Y	X	X	X	X	X	X	<u>G</u>	G
Osteichthyes	Y/F-G-T-H-F/Y-X6-G-G	F/Y	<u>G</u>	T	H	F/Y	X	X	X	X	X	X	<u>G</u>	G
Marsupialia+ Monotremata	F/Y-G-T-H-X7-G-G/A	F/Y	<u>G</u>	T	H	X	X	X	X	X	X	X	<u>G</u>	G/A
Primates	F/Y-G-T/S-H-X7-G-G/S	F/Y	<u>G</u>	T/S	H	X	X	X	X	X	X	X	<u>G</u>	G/S
Rodentia	F/Y-G-T-H-X7-GG	F/Y	<u>G</u>	T	H	X	X	X	X	X	X	X	<u>G</u>	G
Ostali Eutheria	F/Y-G-T-H-X6-L-GG	F/Y	<u>G</u>	T	H	X	X	X	X	X	X	L	<u>G</u>	G
Ostali Vertebrata	F/Y-G-T/S-H-F/Y-X6-G-G/S	F/Y	<u>G</u>	T	H	F/Y	X	X	X	X	X	X	<u>G</u>	G
Ostali Metazoa	X-G-T-H-X7-G-X	X	<u>G</u>	T	H	X	X	X	X	X	X	X	<u>G</u>	X
Amoebozoa	Y-G-X-X-Y-X6-G-G-X-L/I	Y	<u>G</u>	X	X	Y	X	X	X	X	X	X	<u>G</u>	G
Ascomycota+ Basidiomycota	F/Y-G-X8-L/I	F/Y	<u>G</u>	X	X	X	X	X	X	X	X	L/I	<u>G</u>	G
Heterolobosea	F/Y-G-X9-G-G-X-V/I	F/Y	<u>G</u>	X	X	X	X	X	X	X	X	X	<u>G</u>	G
Eudicotyledonae	F/Y-G-T-H-V/I-V/I-X5-G-G	F/Y	<u>G</u>	T	H	V/I	V/I	X	X	X	X	X	<u>G</u>	G
Monocotyledonae	F/Y-G-T-H-V/I-V/I-X5-G-G-K/Q-D	F/Y	<u>G</u>	T	H	V/I	V/I	X	X	X	X	X	<u>G</u>	G
Ostali Plantae	F/Y-G-T-H-X7-G-G	F/Y	<u>G</u>	T	H	X	X	X	X	X	X	X	<u>G</u>	G
Chlamydiaceae	F/V-G-T-H-Y-V/I-A/T-S-V/A-T-Y/F-G-G	F/V	<u>G</u>	T	H	Y	V/I	A/T	S	V/A	T	Y/F	<u>G</u>	G
Proteobacteria	Y/W-G-X-H/Y-X7-G-G	Y/W	<u>G</u>	X	H/Y	X	X	X	X	X	X	X	<u>G</u>	G
Bacteriales	Y/W-G-X-H/Y-X7-G-G	Y/W	<u>G</u>	X	H/Y	X	X	X	X	X	X	X	<u>G</u>	G
Virus	WGVGF(VAHVEM)GG	W	<u>G</u>	V	G	F	V	A	H	V	E	M	<u>G</u>	G

Na podlagi podatkov, ki so na voljo na strani podatkovne zbirke PFAM, smo izdelali HMM logo MACPF motiva. MACPF motiv se na Sliki 12 nahaja med položajema 75 in 87.

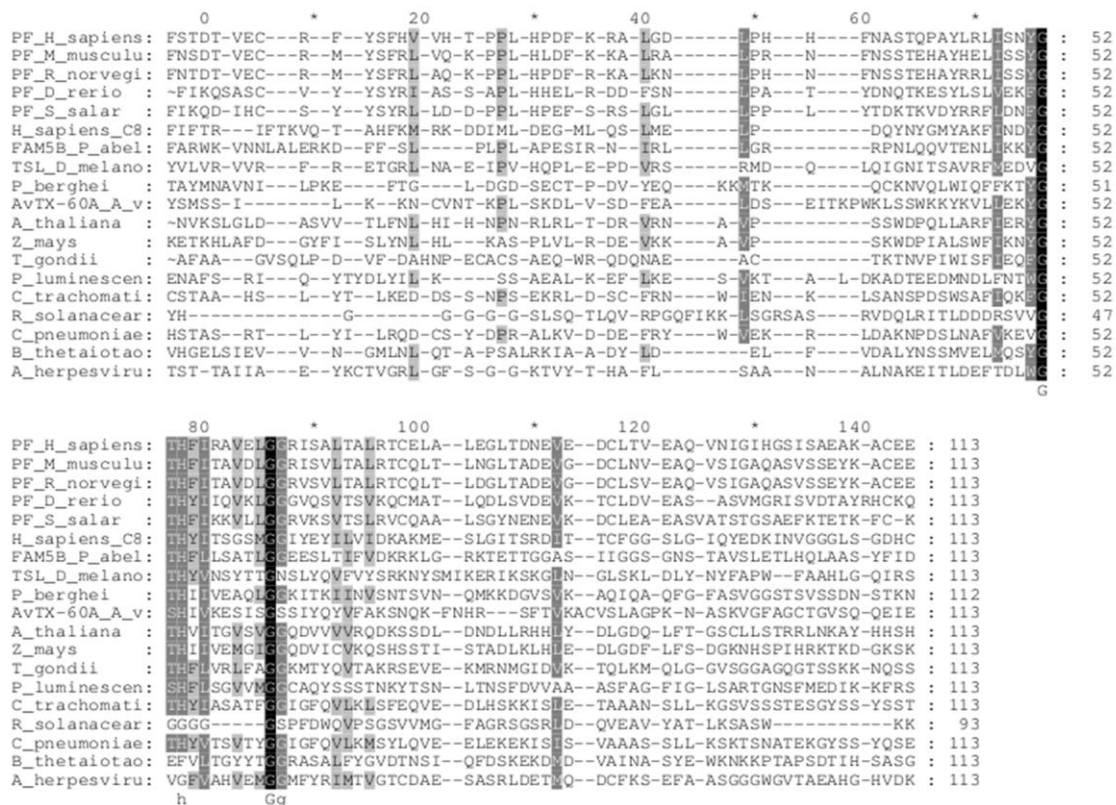
Razvidno je, da se na mestih 76 in 86 (mesti 2 in 12 v MACPF motivu) nahajata dva izjemno dobro ohranjena glicinska ostanka. Ti rezultati se ujemajo z našimi iz Preglednice 3.



Slika 12: Prikaz HMM loga MACPF proteinov v delu, kjer se nahaja MACPF motiv. Višina stolpca prikazuje pogostost določene aminokislinske oz. aminokislinske na tem mestu.

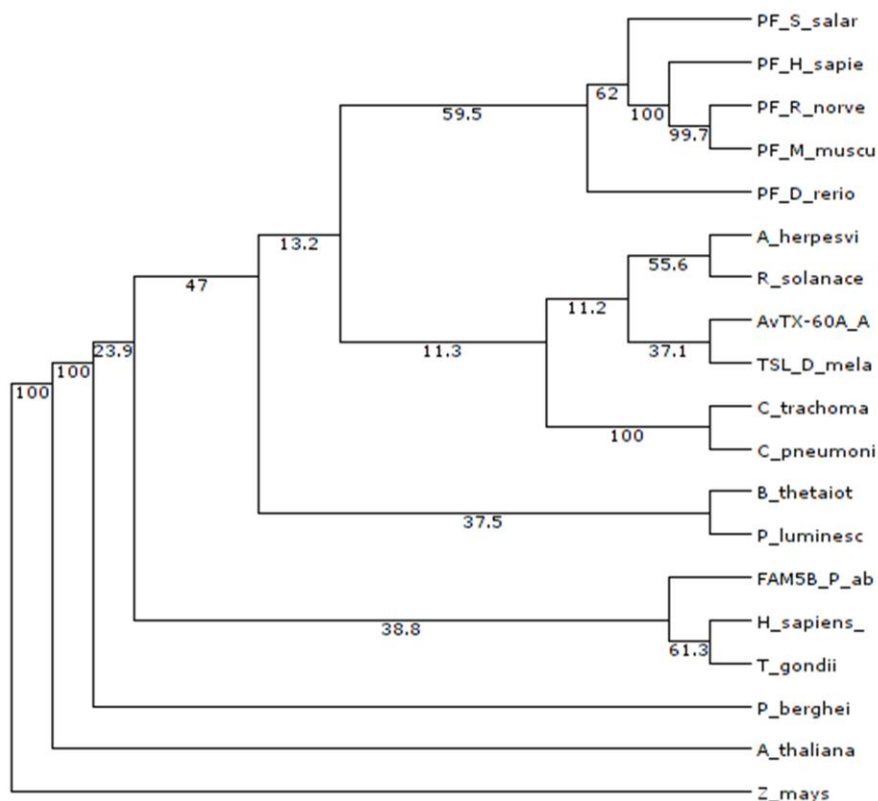
4.4 PORAVNAVA MACPF DOMEN

Poravnava nekaterih ključnih predstavnikov z MACPF domeno (Slika 13) kaže na dokaj slabo homologijo v aminokislinski zgradbi. Ohranjena glicinska ostanka se tu nahajata na mestih 76 in 86. V poravnavo smo vključili različne organizme: človeka, druge vretenčarje, nevretenčarje, rastline, bakterije in edini virus, ki smo ga našli.



Slika 13: Poravnava MACPF domene nekaterih ključnih predstavnikov. Črna ozadja pomeni 100% ohranjenost, temno siva 80% in svetlo siva 60% ohranjenost aminokislinskega ostanka. Na sliki so obarvane štiri stopnje ohranjenosti: črna (100% ohranjenost aminokislinskih ostankov), temno siva (več kot 80% ohranjenost), svetlo siva (več kot 60% ohranjenost), belo ozadje (manj kot 60% ohranjenost). Taksonomske vrste si sledijo od zgoraj navzdol v naslednjem vrstnem redu: *Homo sapiens* (PF=perforin), *Mus musculus* (PF), *Rattus norvegicus* (PF), *Danio rerio* (PF), *Salmo salar* (PF), *Homo sapiens* (C8a), *Pongo abelii*, *Drosophila melanogaster*, *Plasmodium berghei*, *Actineria villosa*, *Arabidopsis thaliana*, *Zea mays*, *Toxoplasma gondii*, *Photorhabdus luminescens*, *Chlamydia trachomatis*, *Ralstonia solanacearum*, *Chlamydomonas reinhardtii*, *Bacteroides thetaiotaomicron*, *Anguillid herpesvirus*

Na podlagi poravnave (Slika 13) smo izdelali filogenetsko drevo (Slika 14). Za test samovzorčenja (Bootstrap) smo izbrali vrednost 100. Številke pod vejami filogenetskega drevesa predstavljajo rezultate testa samovzorčenja.



Slika 14: Filogenetsko drevo na podlagi poravnave MACPF domen iz Slike 13. Okrajšave taksonomskih imen vrst so enake kot pri Sliki 13.

4.5 DODATNE LASTNOSTI PROTEINOV Z MACPF DOMENO

V vzorcu 410 podrobneje obdelanih proteinskih zaporedij, 191 (46,6%) zaporedij vsebuje zgolj samostojno MACPF domeno. Vzorec je bil izbran tako, da vsaki taksonomski vrsti pripada vsaj eno podrobno obdelano zaporedje. V primeru podobnih zaporedij smo izbrali samo najdaljše zaporedje. Upoštevali smo podobnost aminokislinskih zaporedij, dolžin zaporedij, imen zaporedij in podobnost razporeditve domen. Primer: dve zaporedji, kjer je podobnost manjša od 95%, vendar sta zaporedji podobno dolgi, imata enako zaporedje domen in izvirata iz iste vrste. V takem primeru smo izbrali samo eno zaporedje (običajno tisto, ki je daljše).

Izstopajo primeri z dvema oz. tremi ponovitvami MACPF domene znotraj enega aminokislinskega zaporedja. Proteinska zaporedja z dvema MACPF domenama smo našli pri različnih skupinah (Slika 9) od dvoživk do ameb. Našli smo 5 primerov (1,2%) z dvema MACPF domenama znotraj enega proteinskega zaporedja. Proteinska zaporedja z dvema MACPF domenama, smo našli pri vrstah: *Branchiostoma floridae* (Amphioxiformes), *Taeniopygia guttata* (Aves), *Xenopus (Silurana) tropicalis* (Amphibia), *Saccoglossus kowalevskii* (Hemichordata), *Naegleria gruberi* (Heterolobosea). Iz opisa proteina v NCBI zbirki lahko sklepamo, da gre pri vseh omenjenih primerih za predvideni protein (ang. predicted/hypothetical protein). V dveh primerih so v aminokislinskih

zaporedjih prisotne še druge domene, v treh pa sta prisotni le dve samostojni MACPF domeni.

Našli smo tudi 3 primere (0,7%) zaporedij s tremi ponovitvami MACPF domene znotraj enega proteinskega zaporedja (Slika 9). Taki primeri so pri skupini Apicomplexa, natančneje pri vrstah: *Babesia bovis*, *Plasmodium vivax* in *Theileria annulata*. Tukaj je MACPF dvakrat brez prisotnosti drugih domen, enkrat pa je prisotna še dodatna domena.

5 RAZPRAVA IN SKLEPI

5.1 RAZŠIRJENOST MACPF DOMENE

Končno število različnih proteinov z MACPF domeno na dan 24.4.2011 je bilo 809 pri 192 različnih taksonomskih vrstah. Rosado in sod. (2008) so zapisali, da je bilo leta 2008 mogoče z BLAST orodji najti več kot 500 zadetkov z MACPF, kar smo potrdili.

Naši rezultati kažejo, da so proteini, ki vsebujejo MACPF domeno, splošno razširjeni. Našli smo jih tako pri prokariotih kot pri evkariotih in sicer v domeni Bacteria in 4 kraljestvih (Protista, Plantae, Fungi in Metazoa). Največ vrst z MACPF proteini spada v skupino Metazoa (Animalia) in tu je tudi največ različnih zadetkov z MACPF domeno. Vzrokov za prevlado živali je lahko več: (i) mogoče smo tu našli največ predstavnikov, ker je kraljestvo živali z vidika genomskih podatkov relativno dobro raziskano; (ii) sonda, s katero smo našli večino ne-bakterijskih predstavnikov, je bila živalska (iz človeškega PF) ali pa (iii) so se MACPF proteini pojavili pozno tekom evolucije.

Izstopa škrgoustka (*Branchiostoma floridae*), pri kateri smo našli kar 37 različnih proteinskih zaporedij z MACPF domeno. Ta vrsta škrgoustke naj bi imela 29 genov za proteine z MACPF domeno (Huang in sod., 2008). Glede na to, da so mnogi MACPF proteini pri škrgoustki podobni komponentam MAC kompleksa pri človeku, lahko sklepamo, da imajo tudi tu obrambno vlogo. *B. floridae* je žival, pri kateri je okrog 10% vseh genov povezanih z obrambo organizma. Očitno je ta poudarek na obrambi prinesel tudi relativno veliko število proteinov z MACPF domeno. Kozmopolitsko razširjenost in 500 milijonov let obstoja škrgoustk najbrž lahko povežemo prav s to široko paleto genov, ki kodirajo različne komponente imunske obrambe.

Bakterije so zastopane z 49 vrstami, kar je relativno malo glede na ogromno raznolikost bakterijskih vrst. Omeniti je potrebno, da so nekateri avtorji (Ponting, 1999; Wolf in sod., 1999) predvideli horizontalen prenos gena z zapisom za MACPF domeno iz evkariotov v prokariote. Ta teorija delno pojasni slabo zastopanost predstavnikov bakterij. Zanimiva je velika raznolikost bakterij z MACPF. Našli smo tako globokomorsko vrsto, vrsto iz močvirja, patogene človeka, rastlin in insektov. Vsi vsebujejo zapise za proteine z MACPF domeno. Pri vrstah, ki so v takem ali drugačnem patogenem odnosu, si lažje predstavljamo vlogo MACPF domene. Lahko jim pomaga pri hitrejši invaziji ali pa pomaga pri obrambi proti imunskemu sistemu, npr. z mimikrijo (Xu in sod., 2010). Težje pa je ugibati o nalogi MACPF domene pri globokomorski vrsti mikroorganizma.

V literaturi (Kondos in sod., 2010) zasledimo podatek, da MACPF domena pri virusih ni zastopana. Mi smo našli primer virusa z MACPF domeno. Gre za Anguillid herpesvirus 1,

ki ima znan celoten genom in povzroča okužbe sladkovodnih jegulj (van Beurden in sod., 2010). Vloga potencialnega proteina z MACPF domeno je neznana.

Zanimivo je tudi, da so zaporedja z MACPF domeno veliko daljša od mediane dolžine tako prokariotskih (267 aminokislin) kot evkariontskih (361 aminokislin) proteinov (Brocchieri in Karlin, 2005).

Na tej točki je potrebno poudariti nekaj pomembnih dejstev. (i) Večina zadetkov izvira iz različnih modelnih organizmov, ki imajo običajno razvozan celoten genom. Sklepamo lahko, če bi bilo znanih še več celotnih genomov, bi logično našli še več zadetkov z MACPF domeno. Nekatero skupino so relativno slabo zastopane, npr. rastline z le 16 vrstami. Glede na to, da smo našli proteine z MACPF domeno v dokaj različnih rodovih, morda lahko sklepamo, da mnoge vrste rastlin vsebujejo proteine z MACPF domeno. (ii) Mi smo se z PSI-BLAST iskanjem osredotočili na iskanje proteinov z MACPF domeno, na podlagi aminokislinske sestave. Dandanes obstajajo napredni algoritmi, ki upoštevajo tudi sekundarne in terciarne zgradbe proteinov. Orodja s takimi algoritmi bi morda prinesla še več rezultatov, ker je aminokislinska zgradba MACPF proteinov zelo heterogena med različnimi predstavniki (glej predel z MACPF motivom na Sliki 13). (iii) Opozoriti je potrebno tudi na možnost, da vsa ta aminokislinska zaporedja v naravi ne obstajajo. Med zadetki je bilo namreč kar nekaj primerov predvidenih proteinov (ang. predicted protein, hypothetical protein). Običajno gre za proteinska zaporedja, ki so nastala s translacijo nukleotidnih zaporedij. Taka zaporedja običajno še niso bila izolirana iz organskega materiala.

5.2 DOMENSKA ARHITEKTURA

V literaturi ni veliko primerov analize domenske arhitekture pri proteinih z MACPF (na primer Estévez-Calvar in sod., 2011). Pri proteinih z znano 3D strukturo je domenska sestava znana (PF, Plu-MACPF, C8, C9). Tu so bile opisane prisotne domene in tudi bolj ali manj raziskane funkcije teh domen (Rosado in sod., 2007; Hadders in sod., 2007; Slade in sod., 2008; Lovelace in sod., 2011).

Pri naših rezultatih se je izkazalo, da se v aminokislinskih zaporedjih najpogosteje pojavi samostojna MACPF domena. Takih primerov je lahko še več, vendar nismo pregledali vseh 809 zaporedij. Gre za relativno veliko število proteinov s samostojno MACPF domeno. Zlasti, če upoštevamo, da v literaturi še ni podrobneje opisanih primerov proteinov s samostojno MACPF domeno. Opozoriti je potrebno tudi na možnost, da dodatnih domen nismo našli, ker le-te še niso odkrite oz. znane. Zanimivi so tudi primeri, kjer sta prisotni dve ali celo tri MACPF domene znotraj enega zaporedja. V literaturi (Kafsack in Carruthers, 2010) so omenjeni primeri z več MACPF domenami, vendar je njihova vloga še popolnoma neznana.

Pri devetih evkariontskih vrstah, kjer smo podrobneje obdelali domensko arhitekturo, hitro opazimo veliko raznolikost tako v številu kot razporeditvi domen. Kot smo že omenili, se najpogosteje pojavi samostojna MACPF domena. Če te primere izpustimo, je raznolikost še toliko bolj očitna. Prav noben domenski vzorec se ne ponovi pri vseh 9 relativno ozko sorodnih vrstah (vse so iz skupine vretenčarjev). Omenili smo tudi že, da so pri živalih do sedaj našli največ različnih MACPF proteinov in 9 izbranih vrst potrjuje to dejstvo. Pri

človeku smo med 20 zaporedji namreč našli kar 14 različnih domenskih vzorcev. Iz naših rezultatov bi lahko sklepali, da v naravi niso pogoste zveze MACPF domene z drugo domeno oz. domenami. To potrjujejo tudi podatki podatkovne zbirke PFAM. Kombinacija LDL receptorja A in MACPF domene se pojavi 55-krat med 720 zaporedji (7,6%) in ta kombinacija je najpogostejša.

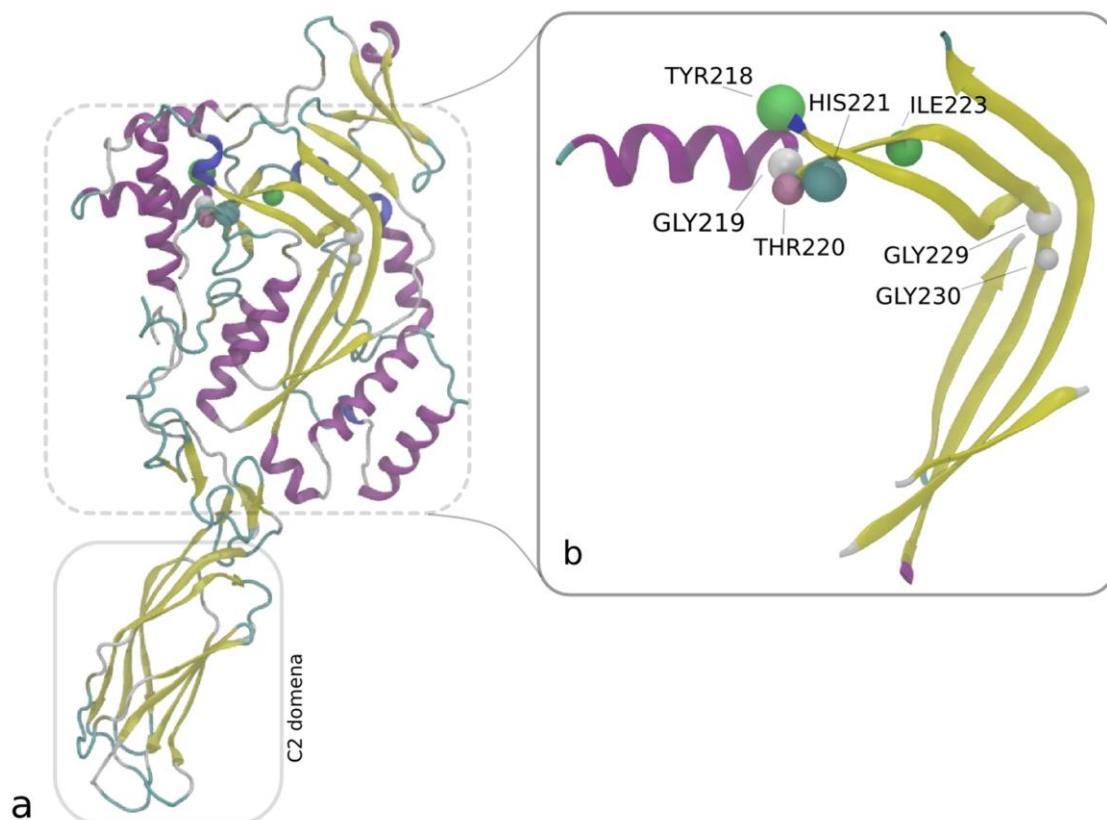
Opazili smo tudi zanimivo razliko v vrstnem redu domen pri živalih in bakterijah. Na N-terminalnem koncu je pri bakterijah najpogosteje MACPF domena (97%). Pri živalih sicer na N-terminalnem koncu tudi prevladuje MACPF, vendar veliko manj pogosto (54,1%).

Znano je, da imajo bakterijska zaporedja običajno manj domen od evkariontskih (Wang in sod., 2011), kar prikazujejo tudi naši podatki.

5.3 MACPF MOTIV

MACPF motiv je prvič opisal Ponting (1999). S poravnavo proteinskih zaporedij rodu *Chlamydiae* je prvi pokazal, da imajo proteini z MACPF domeno skupen, bolj ali manj ohranjen motiv. Osem let kasneje so Rosado in sod. (2007) zapisali motiv v značilni obliki (Y/W-G-T/S-H-F/Y-X₆-GG) z enočrkovno aminokislinsko kodo¹. Naši podatki se delno ujemajo z motivom, ki smo ga zasledili v literaturi. Dva glicinska (G) ostanka se popolnoma ujemata, ostali ostanki pa so bolj variabilni, kot napoveduje motiv. Če za primerjavo vzamemo HMM logo s spletne strani PFAM vidimo, da je na določenih mestih variabilnost večja. Na splošno je variabilnost znotraj MACPF domene velika. Zaporedje MACPF motiva, je vsaj delno ohranjeno pri vseh predstavnikih. Lahko predvidevamo, da ima ta del pomembno funkcijo, katera zahteva dokaj specifično primarno zgradbo. Ohranjena glicinska ostanka ležita v osrednji β -ploskvi oz. v njeni bližini (Slika 15).

¹ Alanin, A; Asparagin, N; Aspartat, D; Arginin, R; Cistein, C; Fenilalanin, P; Glicin, G; Glutamat, E; Glutamin, Q; Histidin, H; Izoleucin, I; Leucin, L; Lizin, K; Metionin, M; Prolin, P; Serin, S; Tirozin, Y; Treonin, T; Triptofan, W; Valin, V



Slika 15: Model mišjega (*Mus musculus*) perforina (PDB ID: 3NSJ). **a**, Molekulo sestavlja C2 domena (spodaj) in centralna MACPF domena. α -vijačnice so obarvane vijolično, β -ploskve rumeno. **b**, Na povečanem odseku so prikazani dobro ohranjeni aminokislinski ostanki MACPF motiva. Nahajajo se znotraj oz. v neposredni bližini centralne »prelomljene« β -ploskve MACPF domene. Najbolje sta ohranjena GLY 219 in GLY229.

Na Sliki 13 vidimo relativno slabo ohranjenost aminokislinskih zaporedij z MACPF domeno. Ohranjen je del med mesti 75 in 87, kjer se nahaja MACPF motiv. Izven tega območja je aminokislinsko zaporedje slabo ohranjeno. V literaturi smo zasledili podatke o slabi ohranjenosti aminokislinskih zaporedij pri proteinih z MACPF domeno. Očitno je vloga MACPF domene taka, da so možne velike variacije v primarni zgradbi. Določeni deli, kjer prihaja do pregibanj (MACPF motiv) oz. do interakcij med monomeri, so bolj ohranjeni, medtem ko transmembranski deli nimajo ohranjene primarne zgradbe, ker to ni potrebno. Ohranjen mora biti amfipatični značaj transmembranskih β -lasnic. Variabilnost je še večja, če upoštevamo še dodatne domene, ki so poleg MACPF.

Poskus izdelave filogenetskega drevesa (Slika 14) se nam je bolj ali manj izjalovil, saj je ohranjenost aminokislinskih zaporedij tako slaba, da sorodstveni odnosi niso realni. Združeni so perforini (PF), druga razmerja pa so vprašljiva.

5.4 SKLEPI

- Proteini z MACPF domeno so splošno razširjeni – z izjemo skupine Archaea.
- Ni domene oz. domen, ki bi se zelo pogosto pojavljale v povezavi z MACPF domeno.
- Ni posebne proteinske domenske arhitekture, ki bi vključevala MACPF domeno.

6 POVZETEK

Za nalogo smo si zastavili opraviti čim bolj splošno in široko iskanje MACPF domene v največjih prosto dostopnih podatkovnih zbirkah. Pričakovano smo dobili veliko število zadetkov, in sicer preko tisoč.

Bioinformatika v zadnjem desetletju dobiva vedno večjo vlogo pri raziskovalnem delu. Bioinformatika omogoča učinkovit pregled že znanih dejstev in podatkov, dobro zasnovano bodočega poskusa ter analizo dobljenih podatkov. Dobljene rezultate shranimo v podatkovnih zbirkah tako, da so na voljo vsem, ki so zainteresirani. Nam je predvsem koristila zmožnost iskanja proteinskih zaporedij s pomočjo kratkih aminokislinskih zaporedij (BLAST) znotraj velikih podatkovnih zbirk.

Zgradba proteinov z MACPF domeno včasih spominja na sestavljanke; glavni funkcionalni del (MACPF) obdajajo različne dodatne domene. Te imajo ponekod znano vlogo, npr. vezava na membrane, vezava na sladkorje, vezava na sosednje monomere. Drugje je vloga še neznana. Te dodatne domene se lahko nahajajo pred ali za MACPF domeno, se enkrat ali večkrat ponovijo, lahko je prisotnih več dodatnih domen ali pa jih ni. Protein lahko vsebuje več (2 ali 3) MACPF domen. V naših podatkih nismo zaznali posebne ponavljajoče domenske arhitekture, kar kaže na veliko raznolikost proteinov z MACPF domeno.

Izkazalo se je, da sta vedno prisotna dva glicinska ostanka (pri mišjem perforinu sta to G219 in G229).

Proteini z MACPF domeno so lahko topni ali netopni, dolgi ali kratki in lahko vsebujejo predvidene transmembranske dele.

Proteini z MACPF domeno so prisotni v vseh kraljestvih živega sveta (izjema Archaea) in so izjemno raznoliki. Pri skupini živali (Metazoa) smo našli največ zaporedij z MACPF domeno in sicer 522, kar je 64.5% od vseh 809 zaporedij z MACPF.

Za človeka so običajno najbolj zanimivi in tudi najbolj raziskani MACPF proteini iz patogenih vrst. Veliko je primerov, kjer MACPF domena sodeluje pri obrambi pred patogenimi organizmi. Dobro je raziskan tudi del imunskega sistema, kjer sodeluje MACPF domena (MAC). Presenetljivo je število proteinov z MACPF domeno, ki ne spadajo v nobeno izmed prejšnjih skupin. Nekateri sodelujejo pri razvoju organizmov, nekateri so popolnoma neraziskani. Z vidika raziskovanja in medicine, je MACPF domena zelo zanimiva in pomembna. Raziskovanje vodi do boljšega razumevanja in prispeva k novim oblikam zdravljenja prirojjenih in akutnih bolezenskih stanj.

7 VIRI

- Adams N.C., Tomoda T., Cooper M., Dietz G. in Hatten M.E. 2002. Mice that lack astrotactin have slowed neuronal migration. *Development*, 129, 4: 965–972
- Aleshin A.E., Schraufstatter I.U., Stec B., Bankston L.A., Liddington R.C. in DiScipio R.G. 2012. Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of membrane attack complex (MAC). *The Journal of Biological Chemistry*, 287, 13: 10210–10222
- Alouf J.E. in Popoff M.R. eds. 2006. *The comprehensive sourcebook of bacterial protein toxins*. Tretja izd., New York: Academic Press.
- Altschul S.F., Gish W., Miller W, Myers E.W. in Lipman D J 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215, 3: 403–410
- Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller Webb in Lipman David J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 17: 3389–3402
- Anderluh G. in Lakey J.H. 2008. Disparate proteins use similar architectures to damage membranes. *Trends in Biochemical Sciences*, 33, 10: 482–490
- Apweiler R., Bairoch A. in Wu C.H. 2004. Protein sequence databases. *Current Opinion in Chemical Biology*, 8, 1: 76–80
- Aroian R. in Goot F.G. 2007. Pore-forming toxins and cellular non-immune defenses (CNIDs). *Current Opinion in Microbiology*, 10, 1: 57–61
- Balachandran P., Hollingshead S.K., Paton J.C. in Briles D.E. 2001. The Autolytic Enzyme LytA of *Streptococcus pneumoniae* Is Not Responsible for Releasing Pneumolysin. *Journal of Bacteriology*, 183, 10: 3108–3116
- Baran K., Dunstone M., Chia J., Ciccone A., Browne Kylie A., Clarke C.J.P., Lukoyanova N., Saibil H.R., Whisstock James C., Voskoboinik I. in Trapani J.A. 2009. The Molecular Basis for Perforin Oligomerization and Transmembrane Pore Assembly. *Immunity*, 30, 5: 684–695
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. in Bourne P.E. 2000. The Protein Data Bank. *Nucleic Acids Research*, 28, 1: 235–242
- Beurden S.J. van, Bossers A., Voorbergen-Laarman M.H.A., Haenen O.L.M., Peters S., Abma-Henkens M.H.C., Peeters B.P.H., Rottier P.J.M. in Engelsma M.Y. 2010. Complete genome sequence and taxonomic position of anguillid herpesvirus 1. *Journal of General Virology*, 91, 4: 880–887
- Bhakdi S. in Tranum-Jensen J. 1991. Alpha-toxin of *Staphylococcus aureus*. *Microbiological Reviews*, 55, 4: 733–751

- Bischofberger M., Gonzalez M.R. in Goot F.G. 2009. Membrane injury by pore-forming proteins. *Current Opinion in Cell Biology*, 21, 4: 589–595
- Bolitho P., Voskoboinik I., Trapani J.A. in Smyth M.J. 2007. Apoptosis induced by the lymphocyte effector molecule perforin. *Current Opinion in Immunology*, 19, 3: 339–347
- Braun J.S., Hoffmann O., Schickhaus M., Freyer D., Dagand E., Bermpohl D., Mitchell T.J., Bechmann I. in Weber J.R. 2007. Pneumolysin Causes Neuronal Cell Death through Mitochondrial Damage. *Infection and Immunity*, 75, 9: 4245–4254
- Brocchieri L. in Karlin S. 2005. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Research*, 33, 10: 3390–3400
- Caserta J.A., Hale M.L., Popoff M.R., Stiles B.G. in McClane B.A. 2008. Evidence that Membrane Rafts Are Not Required for the Action of *Clostridium perfringens* Enterotoxin. *Infection and Immunity*, 76, 12: 5677–5685
- Chen G., Shi L, Litchfield D.W. in Greenberg A.H. 1995. Rescue from granzyme B-induced apoptosis by Wee1 kinase. *The Journal of Experimental Medicine*, 181, 6: 2295–2300
- Cockeran R., Anderson R. in Feldman C. 2003. Pneumolysin in the immunopathogenesis and treatment of pneumococcal disease. *Expert Review of Anti-Infective Therapy*, 1, 2: 231–239
- D'Angelo M.E., Dunstone M. A., Whisstock James C, Trapani J.A. in Bird Phillip I 2012. Perforin evolved from a gene duplication of MPEG1, followed by a complex pattern of gene gain and loss within Euteleostomi. *BMC Evolutionary Biology*, 12, 1: 59
- Dheilly N.M., Haynes P.A., Bove U., Nair S.V. in Raftos D.A. 2011. Comparative proteomic analysis of a sea urchin (*Heliocidaris erythrogramma*) antibacterial response revealed the involvement of apextrin and calreticulin. *Journal of Invertebrate Pathology*, 106, 2: 223–229
- DiScipio R.G. in Berlin C. 1999. The architectural transition of human complement component C9 to poly(C9). *Molecular Immunology*, 36, 9: 575–585
- Dyrløv Bendtsen J., Nielsen Henrik, von Heijne Gunnar in Brunak Søren 2004. Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology*, 340, 4: 783–795
- Emanuelsson O., Brunak Soren, von Heijne Gunnar in Nielsen Henrik 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protocols*, 2, 4: 953–971
- Estévez-Calvar N., Romero A., Figueras A. in Novoa B. 2011. Involvement of pore-forming molecules in immune defense and development of the Mediterranean mussel (*Mytilus galloprovincialis*). *Developmental & Comparative Immunology*, 35: 1017–1031

- Farquhar J.W., Macgregor A.R. in Richmond J. 1958. Familial Haemophagocytic Reticulosis. *British Medical Journal*, 2, 5112: 1561–1564
- Feil S.C., Polekhina G., Gorman M.A. in Parker M.W. 2010. Introduction. G. Anderluh in J. Lakey (ured.). *Proteins Membrane Binding and Pore Formation*. New York, NY: Springer New York: 1–13 pp.
- Gertz E., Yu Y.-K., Agarwala R., Schäffer A. in Altschul S. 2006. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biology*, 4, 1: 1–14
- Gholam C., Grigoriadou S., Gilmour K.C. in Gaspar H.B. 2011. Familial haemophagocytic lymphohistiocytosis: advances in the genetic basis, diagnosis and management. *Clinical & Experimental Immunology*, 163: 271–283
- Giddings K.S., Zhao J., Sims P.J. in Tweten R.K. 2004. Human CD59 is a receptor for the cholesterol-dependent cytolysin intermedilysin. *Nat Struct Mol Biol*, 11, 12: 1173–1178
- Gilbert R.J.C. 2010. Cholesterol-Dependent Cytolysins. G. Anderluh in J. Lakey (ured.). *Proteins Membrane Binding and Pore Formation*. New York, NY: Springer New York: 56–66 pp. .
- Gilbert R.J.C. 2002. Pore-forming toxins. *Cellular and Molecular Life Sciences (CMLS)*, 59: 832–844
- Golstein P. in Kroemer G. 2007. Cell death by necrosis: towards a molecular definition. *Trends in Biochemical Sciences*, 32, 1: 37–43
- Gonzalez M.R., Bischofberger M., Frêche B., Ho S., Parton R.G. in Goot F.G. 2011. Pore-forming toxins induce multiple cellular responses promoting survival. *Cellular Microbiology*, 13, 7: 1026–1043
- Gonzalez M.R., Bischofberger M., Pernot L., Goot F.G. in Frêche B. 2007. Bacterial pore-forming toxins: The (w)hole story? *Cellular and Molecular Life Sciences*, 65: 493–507
- Gouaux E. 1997. Channel-forming toxins: tales of transformation. *Current Opinion in Structural Biology*, 7, 4: 566–573
- Gutierrez M.G., Saka H.A., Chinen I., Zoppino F.C.M., Yoshimori T., Bocco J.L. in Colombo M.I. 2007. Protective role of autophagy against *Vibrio cholerae* cytolysin, a pore-forming toxin from *V. cholerae*. *Proceedings of the National Academy of Sciences*, 104, 6: 1829–1834
- Haag E.S., Sly B.J., Andrews M.E. in Raff R.A. 1999. Apextrin, a Novel Extracellular Protein Associated with Larval Ectoderm Evolution in *Heliocidaris erythrogramma*. *Developmental Biology*, 211, 1: 77–87
- Hadders M.A., Beringer D.X. in Gros P. 2007. Structure of C8 -MACPF Reveals Mechanism of Membrane Attack in Complement Immune Defense. *Science*, 317: 1552–1554

- Han M.V. in Zmasek C.M. 2009. phyloXML: XML for evolutionary biology and comparative genomics. *BMC bioinformatics*, 10, 1: 356
- Henkart P.A., Millard P.J., Reynolds C.W. in Henkart M.P. 1984. Cytolytic activity of purified cytoplasmic granules from cytotoxic rat large granular lymphocyte tumors. *The Journal of Experimental Medicine*, 160, 1: 75–93
- Hessa T., Meindl-Beinker N.M., Bernsel A., Kim H., Sato Y., Lerch-Bader M., Nilsson I., White S.H. in von Heijne Gunnar 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, 450, 7172: 1026–1030
- He X., Zhang Y. in Yu Z. 2011. An Mpeg (macrophage expressed gene) from the Pacific oyster *Crassostrea gigas*: Molecular characterization and gene expression. *Fish & Shellfish Immunology*, 30, 3: 870–876
- Hirokawa T., Boon-Chieng S. in Mitaku S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14, 4: 378–379
- Hofmann in Stoffel W. 1993. TMBASE - a database of membrane spanning protein segments. *Biol Chem HoppeSeyler*, 374, 374: 166
- Hotze E.M. in Tweten R.K. 2012. Membrane assembly of the cholesterol-dependent cytolysin pore complex. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818, 4: 1028–1038
- Huang S., Yuan S., Guo L., Yu Yanhong, Li J., Wu T., Liu T., Yang M., Wu K., Liu H., Ge J., Yu Yingcai, Huang H., Dong M., Yu C., Chen S. in Xu A. 2008. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Research*, 18, 7: 1112–1126
- Humphrey W., Dalke A. in Schulten K. 1996. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14: 33–38
- Iacovache I., Bischofberger M. in Goot F.G. 2010. Structure and assembly of pore-forming proteins. *Current Opinion in Structural Biology*, 20, 2: 241–246
- Iacovache I., Goot F.G. in Pernot Lucile 2008. Pore formation: An ancient yet complex form of attack. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1778, 7-8: 1611–1623
- Ishino T., Yano K., Chinzei Y. in Yuda M. 2004. Cell-Passage Activity Is Required for the Malarial Parasite to Cross the Liver Sinusoidal Cell Layer. *PLoS Biol*, 2, 1: e4
- Johnson M.K., Geoffroy C. in Alouf J.E. 1980. Binding of cholesterol by sulfhydryl-activated cytolysins. *Infection and Immunity*, 27, 1: 97–101
- Kadota K., Ishino T., Matsuyama T., Chinzei Y. in Yuda M. 2004. Essential role of membrane-attack protein in malarial transmission to mosquito host. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 46: 16310–16315
- Kafsack B.F.C. in Carruthers V.B. 2010. Apicomplexan perforin-like proteins. *Communicative & Integrative Biology*, 3: 18–23

- Kägi D., Odermatt B., Seiler P., Zinkernagel R.M., Mak T.W. in Hengartner H. 1997. Reduced Incidence and Delayed Onset of Diabetes in Perforin-deficient Nonobese Diabetic Mice. *The Journal of Experimental Medicine*, 186, 7: 989–997
- Kawano H., Nakatani T., Mori T., Ueno S., Fukaya M., Abe A., Kobayashi M., Toda F., Watanabe M. in Matsuoka I. 2004. Identification and characterization of novel developmentally regulated neural-specific proteins, BRINP family. *Molecular Brain Research*, 125, 1-2: 60–75
- Keefe D., Shi Lianfa, Feske S., Massol R., Navarro F., Kirchhausen T. in Lieberman J. 2005. Perforin Triggers a Plasma Membrane-Repair Response that Facilitates CTL Induction of Apoptosis. *Immunity*, 23, 3: 249–262
- Kehoe M.A., Miller L., Walker J.A. in Boulnois G.J. 1987. Nucleotide sequence of the streptolysin O (SLO) gene: structural homologies between SLO and other membrane-damaging, thiol-activated toxins. *Infection and Immunity*, 55, 12: 3228–3232
- Kemp I.K. in Coyne V.E. 2011. Identification and characterisation of the Mpeg1 homologue in the South African abalone, *Haliotis midae*. *Fish & Shellfish Immunology*, 31, 6: 754–764
- Kennedy C.L., Smith D.J., Lyras D., Chakravorty A. in Rood J.I. 2009. Programmed Cellular Necrosis Mediated by the Pore-Forming α -Toxin from *Clostridium septicum*. *PLoS Pathog*, 5, 7: e1000516
- Kepp O., Galluzzi L., Zitvogel L. in Kroemer G. 2010. Pyroptosis – a cell death modality of its kind? *European Journal of Immunology*, 40, 3: 627–630
- Kloft N., Neukirch C., Bobkiewicz W., Veerachato G., Busch T., Hoven G., Boller K. in Husmann M. 2010. Pro-autophagic signal induction by bacterial pore-forming toxins. *Medical Microbiology and Immunology*, 199: 299–309
- Kondos S.C., Hatfaludi T., Voskoboinik I., Trapani J.A., Law R.H.P., Whisstock J. C. in Dunstone M. A. 2010. The structure and function of mammalian membrane-attack complex/perforin-like proteins. *Tissue Antigens*, 76: 341–351
- Law R.H.P., Lukyanova Natalya, Voskoboinik I., Caradoc-Davies T.T., Baran K., Dunstone Michelle A., D'Angelo M.E., Orlova E.V., Coulibaly F., Verschoor S., Browne Kylie A., Ciccone A., Kuiper Michael J., Bird Phillip I., Trapani J.A., Saibil H.R. in Whisstock James C. 2010. The structural basis for membrane binding and pore formation by lymphocyte perforin. *Nature*, 468: 447–451
- Lesieur C., Vécsey-Semjén B., Abrami L., Fivaz M. in Goot F.G. 1997. Membrane insertion: The strategies of toxins (review). *Molecular Membrane Biology*, 14, 2: 45–64
- Letunic I., Doerks T. in Bork P. 2011. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research*

- Lichtenheld M.G., Olsen K.J., Lu P., Lowrey D.M., Hameed A., Hengartner H. in Podack E.R. 1988. Structure and function of human perforin. *Nature*, 335: 448–451
- Lovelace L.L., Cooper C.L., Sodetz J.M. in Lebioda L. 2011. Structure of Human C8 Protein Provides Mechanistic Insight into Membrane Pore Formation by Complement. *Journal of Biological Chemistry*, 286: 17585–17592
- Lu G. in Moriyama E.N. 2004. Vector NTI, a balanced all-in-one sequence analysis suite. *Briefings in Bioinformatics*, 5, 4: 378–388
- Lyons D.M., Huttunen K.M., Browne Kylie A., Ciccone A., Trapani J.A., Denny W.A. in Spicer J.A. 2011. Inhibition of the cellular function of perforin by 1-amino-2,4-dicyanopyrido[1,2-a]benzimidazoles. *Bioorganic & Medicinal Chemistry*, 19, 13: 4091–4100
- Martin J.-R., Raibaud A. in Ollo R. 1994. Terminal pattern elements in *Drosophila* embryo induced by the torso-like protein. *Nature*, 367, 6465: 741–745
- Mestre M.B., Fader C.M., Sola C. in Colombo M.I. 2010. α -hemolysin is required for the activation of the autophagic pathway in *Staphylococcus aureus* infected cells. *Autophagy*, 6: 110–125
- Meyer-Morse N., Robbins J.R., Rae C.S., Mochegova S.N., Swanson M.S., Zhao Z., Virgin H.W. in Portnoy D. 2010. Listeriolysin O Is Necessary and Sufficient to Induce Autophagy during *Listeria monocytogenes* Infection. *PLoS ONE*, 5, 1: e8610
- Miller D.J., Hemmrich G., Ball E.E., Hayward D.C., Khalturin K., Funayama N., Agata K. in Bosch T.C. 2007. The innate immune repertoire in Cnidaria - ancestral complexity and stochastic gene loss. *Genome Biology*, 8: R59
- Mitaku S. in Hirokawa T. 1999. Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. *Protein Engineering*, 12, 11: 953–957
- Mitaku S., Hirokawa T. in Tsuji T. 2002. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. *Bioinformatics*, 18, 4: 608–616
- Mitchell C.M., Hovis K.M., Bavoil P.M., Myers G.S., Carrasco J.A. in Timms P. 2010. Comparison of koala LPCoLN and human strains of *Chlamydia pneumoniae* highlights extended genetic diversity in the species. *BMC Genomics*, 11: 442
- Moretti S., Armougom F., Wallace I.M., Higgins Desmond G, Jongeneel C.V. in Notredame Cedric 2007. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Research*, 35, Web Server issue: W645–648
- Morita-Yamamuro C., Tsutsui T., Sato M., Yoshioka H., Tamaoki M., Ogawa D., Matsuura H., Yoshihara T., Ikeda A., Uyeda I. in Yamaguchi J. 2005. The Arabidopsis Gene CAD1 Controls Programmed Cell Death in the Plant Immune System and Encodes a Protein Containing a MACPF Domain. *Plant and Cell Physiology*, 46, 6: 902–912

- Motomiya M., Kobayashi Miwako, Iwasaki N., Minami A. in Matsuoka Ichiro 2007. Activity-dependent regulation of BRINP family genes. *Biochemical and Biophysical Research Communications*, 352, 3: 623–629
- Muchmore S.W., Sattler M., Liang H., Meadows R.P., Harlan J.E., Yoon H.S., Nettlesheim D., Chang B.S., Thompson C.B., Wong S.-L., Ng S.-C. in Fesik S.W. 1996. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature*, 381, 6580: 335–341
- Nagai H., Oshiro N., Takuwa-Kuroda K., Iwanaga S., Nozaki M. in Nakajima T. 2002. Novel proteinaceous toxins from the nematocyst venom of the Okinawan sea anemone *Phyllodiscus semoni* Kwietniewski. *Biochemical and Biophysical Research Communications*, 294, 4: 760–763
- Natarajan K. in Cowan J.A. 1998. Solution structure of a synthetic lytic peptide: the perforin amino terminus. *Chemistry & Biology*, 5, 3: 147–154
- Nelson K.L., Brodsky R.A. in Buckley J.T. 1999. Channels formed by subnanomolar concentrations of the toxin aerolysin trigger apoptosis of T lymphomas. *Cellular Microbiology*, 1, 1: 69–74
- Néron B., Ménager H., Maufrais C., Joly N., Maupetit J., Letort S., Carrere S., Tuffery P. in Letondal C. 2009. Mobylye: A New Full Web Bioinformatics Framework. *Bioinformatics*, 25, 22: 3005–3011
- Nicholas K.B., Nicholas H.B. in Deerfield D.W. 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNEW. NEWS*, 4: 14
- Nielsen H, Engelbrecht J., Brunak S in von Heijne G 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*, 10, 1: 1–6
- Notredame C, Higgins D G in Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302, 1: 205–217
- Oshiro N., Kobayashi C., Iwanaga S., Nozaki M., Namikoshi M., Spring J. in Nagai H. 2004. A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone *Actinaria villosa*. *Toxicon*, 43: 225–228
- Palmer M., Saweljew P., Vulicevic I., Valeva A., Kehoe M.A. in Bhakdi S. 1996. Membrane-penetrating Domain of Streptolysin O Identified by Cysteine Scanning Mutagenesis. *Journal of Biological Chemistry*, 271, 43: 26664–26667
- Parker M.W. in Feil S.C. 2005. Pore-forming protein toxins: from structure to function. *Progress in Biophysics and Molecular Biology*, 88, 1: 91–142
- Peitsch M.C., Amiguet P., Guy R., Brunner J., Maizel Jr J.V. in Tschopp J. 1990. Localization and molecular modelling of the membrane-inserted domain of the ninth component of human complement and perforin. *Molecular Immunology*, 27, 7: 589–602

- Peitsch M.C. in Tschopp J. 1991. Assembly of macromolecular pores by immune defense systems. *Current Opinion in Cell Biology*, 3, 4: 710–716
- Pipkin M.E. in Lieberman J. 2007. Delivering the kiss of death: progress on understanding how perforin works. *Current Opinion in Immunology*, 19, 3: 301–308
- Podack E.R. 2009. How to Polymerize in Order to Survive. *Immunity*, 30, 5: 668–670
- Podack E.R. in Konigsberg P.J. 1984. Cytolytic T cell granules. Isolation, structural, biochemical, and functional characterization. *The Journal of Experimental Medicine*, 160, 3: 695–710
- Polekhina G., Giddings K.S., Tweten R.K. in Parker M.W. 2005. Insights into the action of the superfamily of cholesterol-dependent cytolysins from studies of intermedilysin. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 3: 600–605
- Ponting C.P. 1999. Chlamydial homologues of the MACPF (MAC/perforin) domain. *Current Biology*, 9, 24: R911–R913
- Praper T., Beseničar M.P., Istinič H., Podlesek Z., Metkar S.S., Froelich C.J. in Anderluh G. 2010. Human perforin permeabilizing activity, but not binding to lipid membranes, is affected by pH. *Molecular Immunology*, 47, 15: 2492–2504
- Praper T., Sonnen A.F.-P., Kladnik A., Andrighetti A.O., Viero G., Morris K.J., Volpi E., Lunelli L., Serra M.D., Froelich C.J., Gilbert R.J.C. in Anderluh G. 2011. Perforin Activity at Membranes Leads to Invaginations and Vesicle Formation. *Proceedings of the National Academy of Sciences*, 108, 52: 21016–21021
- Praper T., Sonnen A., Viero G., Kladnik A., Froelich C.J., Anderluh G., Dalla Serra M. in Gilbert R.J.C. 2011. Human Perforin Employs Different Avenues to Damage Membranes. *Journal of Biological Chemistry*, 286, 4: 2946–2955
- Punta M., Coghill P.C., Eberhardt R.Y., Mistry J., Tate J., Boursnell C., Pang N., Forslund K., Ceric G., Clements J., Heger A., Holm L., Sonnhammer E.L.L., Eddy S.R., Bateman A. in Finn R.D. 2011. The Pfam protein families database. *Nucleic Acids Research*, 40, D1: D290–D301
- Ramachandran R., Tweten R.K. in Johnson A.E. 2005. The domains of a cholesterol-dependent cytolysin undergo a major FRET-detected rearrangement during pore formation. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 20: 7139–7144
- Rosado C.J., Buckle A. M., Law R.H.P., Butcher R.E., Kan W.-T., Bird C.H., Ung K., Browne K. A., Baran K., Bashtannyk-Puhlovich T.A., Faux N.G., Wong W., Porter C.J., Pike R.N., Ellisdon A.M., Pearce M.C., Bottomley S.P., Emsley J., Smith A.I., Rossjohn J., Hartland E.L., Voskoboinik I., Trapani J.A., Bird P. I., Dunstone M. A. in Whisstock J. C. 2007. A Common Fold Mediates Vertebrate Defense and Bacterial Attack. *Science*, 317: 1548–1551
- Rosado C.J., Kondos S.C., Bull T.E., Kuiper M. J., Law R.H.P., Buckle Ashley M., Voskoboinik I., Bird Phillip I., Trapani J.A., Whisstock James C. in Dunstone

- Michelle A. 2008. The MACPF/CDC family of pore-forming toxins. *Cellular Microbiology*, 10, 9: 1765–1774
- Rossjohn J., Feil S.C., McKinstry W.J., Tweten R.K. in Parker M.W. 1997. Structure of a Cholesterol-Binding, Thiol-Activated Cytolysin and a Model of Its Membrane Form. *Cell*, 89, 5: 685–692
- Satoh H., Oshiro N., Iwanaga S., Namikoshi M. in Nagai H. 2007. Characterization of PsTX-60B, a new membrane-attack complex/perforin (MACPF) family toxin, from the venomous sea anemone *Phyllodiscus semoni*. *Toxicon*, 49, 8: 1208–1210
- Schultz J., Milpetz F., Bork Peer in Ponting C.P. 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences*, 95, 11: 5857–5864
- Schuster-Böckler B., Schultz J. in Rahmann S. 2004. HMM Logos for visualization of protein families. *BMC Bioinformatics*, 5, 1: 7
- Scibek J.J., Plumb M.E. in Sodetz J.M. 2002. Binding of Human Complement C8 to C9: Role of the N-Terminal Modules in the C8 α Subunit. *Biochemistry*, 41, 49: 14546–14551
- Shatursky O., Heuck A.P., Shepard L.A., Rossjohn J., Parker M.W., Johnson A.E. in Tweten R.K. 1999. The Mechanism of Membrane Insertion for a Cholesterol-Dependent Cytolysin: A Novel Paradigm for Pore-Forming Toxins. *Cell*, 99, 3: 293–299
- Shepard L.A., Heuck A.P., Hamman B.D., Rossjohn J., Parker M.W., Ryan K.R., Johnson A.E. in Tweten R.K. 1998. Identification of a Membrane-Spanning Domain of the Thiol-Activated Pore-Forming Toxin *Clostridium perfringens* Perfringolysin O: An α -Helical to β -Sheet Transition Identified by Fluorescence Spectroscopy. *Biochemistry*, 37, 41: 14563–14574
- Shinkai Y., Takio K. in Okumura K. 1988. Homology of perforin to the ninth component of complement (C9). *Nature*, 334: 525–527
- Shiver J.W., Su L. in Henkart P.A. 1992. Cytotoxicity with target DNA breakdown by rat basophilic leukemia cells expressing both cytolysin and granzyme A. *Cell*, 71, 2: 315–322
- Slade D.J., Lovelace L.L., Chruszcz M., Minor W., Lebioda Lukasz in Sodetz J.M. 2008. Crystal Structure of the MACPF Domain of Human Complement Protein C8 α in Complex with the C8 γ Subunit. *Journal of Molecular Biology*, 379: 331–342
- Smialowski P., Martin-Galiano A.J., Mikolajka A., Girschick T., Holak T.A. in Frishman D. 2007. Protein solubility: sequence based prediction and experimental verification. *Bioinformatics*, 23, 19: 2536–2542
- Smyth M.J., Thia K.Y.T., Street S.E.A., MacGregor D., Godfrey D.I. in Trapani J.A. 2000. Perforin-Mediated Cytotoxicity Is Critical for Surveillance of Spontaneous Lymphoma. *The Journal of Experimental Medicine*, 192, 5: 755–760

- Snider C., Jayasinghe S., Hristova K. in White S.H. 2009. MPEX: A tool for exploring membrane proteins. *Protein Science*, 18, 12: 2624–2628
- Sodetz J.M. in Plumb M.E. 2005. Complement: Terminal Pathway. John Wiley & Sons, Ltd. *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons
- Stepp S.E., Dufourcq-Lagelouse R., Deist F.L., Bhawan S., Certain S., Mathew P.A., Henter J.-I., Bennett M., Fischer A., Basile G. de S. in Kumar V. 1999. Perforin Gene Defects in Familial Hemophagocytic Lymphohistiocytosis. *Science*, 286, 5446: 1957–1959
- Stevens L.M., Frohnhof H.G., Klingler M. in Nusslein-Volhard C. 1990. Localized requirement for torso-like expression in follicle cells for development of terminal anlagen of the *Drosophila* embryo. *Nature*, 346, 6285: 660–663
- Taylor L.D., Nelson D.E., Dorward D.W., Whitmire W.M. in Caldwell H.D. 2010. Biological Characterization of *Chlamydia trachomatis* Plasticity Zone MACPF Domain Family Protein CT153. *Infection and Immunity*, 78: 2691–2699
- Tilley S.J., Orlova E.V., Gilbert R.J.C., Andrew P.W. in Saibil H.R. 2005. Structural Basis of Pore Formation by the Bacterial Toxin Pneumolysin. *Cell*, 121, 2: 247–256
- Tilley S.J. in Saibil H.R. 2006. The mechanism of pore formation by bacterial toxins. *Current Opinion in Structural Biology*, 16, 2: 230–236
- Di Tommaso P., Moretti S., Xenarios I., Orobiteg M., Montanyola A., Chang J.-M., Taly J.-F. in Notredame Cedric 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, 39, Web Server issue: W13–17
- Trambas C., Gallo F., Pende D., Marcenaro S., Moretta L., De Fusco C., Santoro A., Notarangelo L., Arico M. in Griffiths G.M. 2005. A single amino acid change, A91V, leads to conformational changes that can impair processing to the active form of perforin. *Blood*, 106, 3: 932–937
- Trapani J.A. 1998. Dual Mechanisms of Apoptosis Induction by Cytotoxic Lymphocytes. Kwang W. Jeon (ured.). *International Review of Cytology*. Academic Press: 111–192 pp.
- Tschopp J., Masson D. in Stanley K.K. 1986. Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytotoxicity. *Nature*, 322, 6082: 831–834
- Tweten R.K. 2005. Cholesterol-Dependent Cytolysins, a Family of Versatile Pore-Forming Toxins. *Infection and Immunity*, 73, 10: 6199–6209
- Voskoboinik I., Dunstone Michelle A., Baran K., Whisstock James C. in Trapani J.A. 2010. Perforin: structure, function, and role in human immunopathology. *Immunological Reviews*, 235, 1: 35–54
- Voskoboinik I., Smyth M.J. in Trapani J.A. 2006. Perforin-mediated target-cell death and immune homeostasis. *Nature Reviews Immunology*, 6: 940–952

- Voskoboinik I., Sutton V.R., Ciccone A., House C.M., Chia J., Darcy P.K., Yagita H. in Trapani J.A. 2007. Perforin activity and immune homeostasis: the common A91V polymorphism in perforin results in both presynaptic and postsynaptic defects in function. *Blood*, 110, 4: 1184–1190
- Voskoboinik I., Thia M.-C., De Bono A., Browne K., Cretney E., Jackson J.T., Darcy P.K., Jane S.M., Smyth M.J. in Trapani J.A. 2004. The Functional Basis for Hemophagocytic Lymphohistiocytosis in a Patient with Co-inherited Missense Mutations in the Perforin (PFN1) Gene. *The Journal of Experimental Medicine*, 200, 6: 811–816
- Voskoboinik I., Thia M.-C. in Trapani J.A. 2005. A functional analysis of the putative polymorphisms A91V and N252S and 22 missense perforin mutations associated with familial hemophagocytic lymphohistiocytosis. *Blood*, 105, 12: 4700–4706
- Walker J.A., Allen R.L., Falmagne P., Johnson M.K. in Boulnois G.J. 1987. Molecular cloning, characterization, and complete nucleotide sequence of the gene for pneumolysin, the sulfhydryl-activated toxin of *Streptococcus pneumoniae*. *Infection and Immunity*, 55, 5: 1184–1189
- Wallace I.M., O'Sullivan O., Higgins Desmond G. in Notredame Cedric 2006. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Research*, 34, 6: 1692–1699
- Wang M., Kurland C.G. in Caetano-Anollés G. 2011. Reductive evolution of proteomes and protein structures. *Proceedings of the National Academy of Sciences*, 108, 29: 11954–11958
- Wolf Y.I., Aravind L. in Koonin E.V. 1999. Rickettsiae and Chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends in Genetics*, 15, 5: 173–175
- Xu Q., Abdubek P., Astakhova T., Axelrod H.L., Bakolitsa C., Cai X., Carlton D., Chen C., Chiu H.-J., Clayton T., Das D., Deller M.C., Duan L., Ellrott K., Farr C.L., Feuerhelm J., Grant J.C., Grzechnik A., Han G.W., Jaroszewski L., Jin K.K., Klock H.E., Knuth M.W., Kozbial P., Krishna S.S., Kumar A., Lam W.W., Marciano D., Miller M.D., Morse A.T., Nigoghossian E., Nopakun A., Okach L., Puckett C., Reyes R., Tien H.J., Trame C.B., van den Bedem H., Weekes D., Wooten T., Yeh A., Zhou J., Hodgson K.O., Wooley J., Elsliger M.-A., Deacon A.M., Godzik A., Lesley S.A. in Wilson I.A. 2010. Structure of a membrane-attack complex/perforin (MACPF) family protein from the human gut symbiont *Bacteroides thetaiotaomicron*. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 66: 1297–1305
- Yan H.-K., Yuan Z.-G., Petersen E., Zhang X.-X., Zhou D.-H., Liu Q., He Y., Lin R.-Q., Xu M.-J., Chen X.-L., Zhong X.-L. in Zhu X.-Q. 2011. *Toxoplasma gondii*: Protective immunity against experimental toxoplasmosis induced by a DNA vaccine encoding the perforin-like protein 1. *Experimental Parasitology*, 128, 1: 38–43
- Yarovinsky T.O., Monick M.M., Husmann M. in Hunninghake G.W. 2008. Interferons Increase Cell Resistance to Staphylococcal Alpha-Toxin. *Infection and Immunity*, 76, 2: 571–577

Kisovec M. Razširjenost proteinske domene MACPF.

Dipl. delo. Ljubljana, Univ. v Ljubljani, Biotehniška fakulteta, Odd. za biologijo, 2012

Zdobnov E.M. in Apweiler R. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 17, 9: 847–848

Zheng C., Heintz N. in Hatten M.E. 1996. CNS Gene Encoding Astrotactin, Which Supports Neuronal Migration Along Glial Fibers. *Science*, 272, 5260: 417–419

ZAHVALA

Na tem mestu bi se rad zahvalil vsem, ki so kakorkoli pripomogli k nastanku tega diplomskega dela.

Najprej bi se rad zahvalil mentorju, prof. dr. Gregorju Anderluhu, ki je poskrbel, da je proces diplomskega dela ves čas tekkel v pravo smer. Hvala za vse napotke, pripombe, komentarje in dragocen čas.

Posebej bi se rad zahvalil tudi Mihi Miklju, ki mi je pomagal na vseh stopnjah izdelave diplomskega dela.

Zahvaljujem se prof. dr. Petru Mačku za hiter in temeljit pregled diplomskega dela. Hvala tudi predsednici prof. dr. Kristini Sepčič za pregled in napotke pred zagovorom.

Hvala bratoma in staršem, ki so mi tekom študija stali ob strani.

Zahvalil bi se še vsem sošolkam in sošolcem za izjemno lepa študentska leta.

Tjaša, hvala za pomoč, potrpežljivost in srečo!

PRILOGA

Priloga A

Proteinske domene, ki smo jih našli poleg MACPF domene, ki se vežejo na sladkorje (rumeno obarvano ozadje) ali na membrane (modro ozadje). Opisi domen so v angleškem jeziku.

Bakterijske domene

MIR Domain in ryanodine and inositol trisphosphate receptors and protein O-mannosyl transferases.

Jacalin-like lectin domain Proteins containing this domain are lectins. It is found in 1 to 6 copies in these proteins. The domain is also found in the animal prostatic spermine-binding protein. This entry represents a mannose-binding lectin domain with a beta-prism fold consisting of three 4-stranded beta-sheets, with an internal pseudo 3-fold symmetry. Some lectins in this group stimulate distinct T- and B- cell functions, such as Jacalin, which binds to the T-antigen and acts as an agglutinin. This domain is found in 1 to 6 copies in lectins. The domain is also found in the salt-stress induced protein from rice and an animal prostatic spermine-binding protein.

B-lectin (D-mannose binding lectin): These proteins include mannose-specific lectins from plant as well as bacteriocins from bacteria.

Integrin alpha N-terminal domain Integrins are the major metazoan receptors for cell adhesion to extracellular matrix proteins and, in vertebrates, also play important roles in certain cell-cell adhesions, make transmembrane connections to the cytoskeleton and activate many intracellular signalling pathways. The integrin receptors are composed of alpha and beta subunit heterodimers. Each subunit crosses the membrane once, with most of the polypeptide residing in the extracellular space, and has two short cytoplasmic domains. Some members of this family have EGF repeats at the C terminus and also have a vWA domain inserted within the integrin domain at the N terminus.

GRAM The GRAM domain is found in glucosyltransferases, myotubularins and other putative membrane-associated proteins. It is normally about 70 amino acids in length. It is thought to be an intracellular protein-binding or lipid-binding signalling domain, which has an important function in membrane-associated processes.

Xlink The link domain is a hyaluronan(HA)-binding region found in proteins of vertebrates that are involved in the assembly of extracellular matrix, cell adhesion, and migration.

MAPB

It has been suggested that the MABP domain has a membrane-associated function, perhaps even specific interactions with membrane components. It is plausible that the eukaryotic MABP domains are adaptators that help linking other associated domains found in the same polypeptide to vesicular membranes. The MABP domain has an internal repeat structure of three homologous segments. Consistent with this, the structurally characterized representative *Photobacterium* plu1415, showed that this region precisely corresponds to a type-I β -prism domain with an internal three fold symmetry (see PDB:2QP2). Each of the three sub-domains of the β -prism structure is a distinctive three-stranded β -sheet. The MABP domain shares a triradial symmetry with β -sheets parallel to the prism axis. The β -prism fold is associated with membrane interaction. The majority of the eukaryotic MABP domain versions contain a conserved cysteine in the first and third subdomain of the β -prism.

ENV polyprotein

Enveloped viruses such as Human immunodeficiency virus 1, influenza virus, and Ebola virus sp. express a surface glycoprotein that mediates both cell attachment and fusion of viral and cellular membranes.

BIG2

Members of this family are found in bacterial and phage surface proteins such as intimins.

MACPF

MAC/Perforin domain. The membrane-attack complex (MAC) of the complement system forms transmembrane channels. These channels disrupt the phospholipid bilayer of target cells, leading to cell lysis and death. A number of proteins participate in the assembly of the MAC. Freshly activated C5b binds to C6 to form a C5b-6 complex, then to C7 forming the C5b-7 complex. The C5b-7 complex binds to C8, which is composed of three chains (alpha, beta, and gamma), thus forming the C5b-8 complex. C5b-8 subsequently binds to C9 and acts as a catalyst in the polymerisation of C9. Active MAC has a subunit composition of C5b-C6-C7-C8-C9. Perforin is a protein found in cytolytic T-cell and killer cells. In the presence of calcium, perforin polymerises into transmembrane tubules and is capable of lysing, non-specifically, a variety of target cells. There are a number of regions of similarity in the sequences of complement components C6, C7, C8-alpha, C8-beta, C9 and perforin.

Glivna domena

Ricin B lectin

Polysaccharide metabolism and transport. Ricin is a legume lectin from the seeds of the castor bean plant, *Ricinus communis*. The seeds are poisonous to people, animals and insects and just one milligram of ricin can kill an adult. Primary structure analysis has shown the presence of a similar domain in many carbohydrate-recognition proteins like plant and bacterial AB-toxins, glycosidases or proteases. This domain, known as the ricin B lectin domain, can be present in one or more copies and has been shown in some instance to bind simple sugars, such as galactose or lactose.

Evkariontske domene

Coagulation factor 5/8

Coagulation factor 5/8 C-terminal domain, discoidin domain; Cell surface-attached carbohydrate-binding domain, present in eukaryotes and assumed to have horizontally

transferred to eubacterial genomes

Lectin C domain

A C-type lectin (CLEC) is a type of carbohydrate-binding protein domain known as a lectin. The C-type designation is from their requirement for calcium for binding. Proteins that contain C-type lectin domains have a diverse range of functions including cell-cell adhesion, immune response to pathogens and apoptosis.

Agglutinin

Agglutinins are sugar-specific lectins that can agglutinate erythrocytes and other cell types. Lectins occur widely in plants, as well as some microorganisms and animal. Agglutinin from *Amaranthus caudatus* (amaranthin) is a lectin from the ancient South American crop, amaranth grain. Although its biological function is unknown, it can agglutinate A, B and O red blood cells, and has a carbohydrate-binding site that is specific for the methyl-glycoside of the T-antigen found linked to serine or threonine residues of cell surface glycoproteins .

Mannose-specific B-lectin

Members of this domain are plant lectins. Curculin is a sweet-tasting and taste-modifying protein from the fruits of *Curculigo latifolia* (Lumbah). The three mannose-binding sites are devoid of mannose-binding activity. Other members of this domain are mannose specific and have diverse functions. The lectin of the saffron crocus (*Crocus sativus*) (Saffron) specifically interacts with a yeast mannan and is a major corm protein specifically expressed in this organ

Defensin beta

Beta defensins are a family of mammalian defensins. The beta defensins are antimicrobial peptides implicated in the resistance of epithelial surfaces to microbial colonization. Defensins are 2-6 kDa, cationic, microbicidal peptides active against many Gram-negative and Gram-positive bacteria, fungi, and enveloped viruses, containing three pairs of intramolecular disulfide bonds. On the basis of their size and pattern of disulfide bonding, mammalian defensins are classified into alpha, beta and theta categories. Every mammalian species explored thus far has beta-defensins.

EGF & EGF 3

EGF-like domain is an evolutionary conserved protein domain, of about thirty to forty amino-acid residues long, which was found in a large number of mostly animal proteins. All these repeats are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (except in prostaglandin H/G synthase). The EGF domain includes six cysteine residues which have been shown (in EGF) to be involved in three disulfide bonds. The main structure is a two-stranded beta-sheet followed by a loop to a C-terminal short two-stranded sheet. This family includes a variety of EGF-like domain homologues. This family includes the C-terminal domain of the malaria parasite MSP1 protein.

Giardia variant-specific surface protein=VSP

During infection, the intestinal protozoan parasite *Giardia lamblia* virus undergoes continuous antigenic variation which is determined by diversification of the parasite's major surface antigen, named VSP (variant surface protein).

- Growth factor receptor** Signal transduction
- Metallothionein** Metallothionein (MT) is a family of cysteine-rich, low molecular weight (MW ranging from 500 to 14000 Da) proteins. They are localized to the membrane of the Golgi apparatus. MTs have the capacity to bind both physiological (such as zinc, copper, selenium) and xenobiotic (such as cadmium, mercury, silver, arsenic) heavy metals through the thiol group of its cysteine residues, which represents nearly the 30% of its amino acidic residues.
- Bacteriocin II** Class II bacteriocins are a class of small peptides that inhibit the growth of various bacteria. Many Gram-positive bacteria produce ribosomally synthesized antimicrobial peptides, termed bacteriocins. Bacteriocins for which disulfide bonds are the only modification to the peptide are Class II bacteriocins.
- C2 domain of perforin** A C2 domain is a protein structural domain involved in targeting proteins to cell membranes. It is a beta-sandwich composed of 8 α -strands that co-ordinate two or three calcium ions, which bind in a cavity formed by the first and final loops of the domain, on the membrane binding face.
- HRM** This extracellular domain contains four conserved cysteines that probably for disulphide bridges. The domain is found in a variety of hormone receptors. It may be a ligand binding domain.
- Rhabdovirus spike glycoprotein** Frequently abbreviated to G protein. The glycoprotein spike is made up of a trimer of G proteins. Channel formed by glycoprotein spike is thought to function in a similar manner to Influenza virus M2 protein channel, thus allowing a signal to pass across the viral membrane to signal for viral uncoating.
- Major facilitator superfamily** The Major Facilitator Superfamily (MFS) is a large and diverse group of secondary transporters that includes uniporters, symporters, and antiporters. MFS proteins facilitate the transport across cytoplasmic or internal membranes of a variety of substrates including ions, sugar phosphates, drugs, neurotransmitters, nucleosides, amino acids, and peptides.
- 7-transmembrane receptor** This family is known as Family B, the secretin-receptor family or family 2 of the G-protein-coupled receptors (GPCRs). They have been described in many animal species, but not in plants, fungi or prokaryotes. Three distinct sub-families are recognised. Subfamily B1 contains classical hormone receptors, such as receptors for secretin and glucagon, that are all involved in cAMP-mediated signalling pathways. Subfamily B2 contains receptors with long extracellular N-termini, such as the leukocyte cell-surface antigen CD97 (P48960); calcium-independent receptors for latrotoxin (such as O94910), and brain-specific angiogenesis inhibitors (such as O14514) amongst others. Subfamily B3 includes Methuselah and other Drosophila proteins (e.g. P83119). Other than the typical seven-transmembrane region, characteristic structural features include

an amino-terminal extracellular domain involved in ligand binding, and an intracellular loop (IC3) required for specific G-protein coupling.

Sodium:solute symporter family

Sodium/substrate transporters are grouped in different families based on sequence similarities. One of these families, known as the sodium:solute symporter family (SSSF), contains over a hundred members of pro- and eukaryotic origin. The average hydropathy plot for SSSF proteins predicts 11 to 15 putative transmembrane domains (TMs) in alpha-helical conformation. A secondary structure model of PutP from *Escherichia coli* suggests the protein contains 13 TMs with the N-terminus located on the periplasmic side of the membrane and the C-terminus facing the cytoplasm. The results support the idea of a common topological motif for members of the SSSF. Transporters with a C-terminal extension are proposed to have an additional 14th TM.