UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di

BIOLOGIA

CORSO DI DOTTORATO DI RICERCA IN: **BIOSCIENZE E BIOTECNOLOGIE**
CURRICOLO: **BIOTECNOLOGIE**
CICLO **XXIX**

Welcome to the jungle: insights into the innate defence peptides

Coordinatore: Ch.mo Prof. Paolo Bernardi
Supervisore: Dott.ssa Paola Venier

Dottorando: Marco Franzoi

Welcome to the jungle: insights into the innate defence peptides

PhD thesis

MARCO FRANZOI
MSc UniPD Industrial Biotechnology

Supervisor: Dott.ssa Paola Venier

2016

Acknowledgement

I would like to thank my supervisor Paola Venier for the careful scientific training as well as for her kindness and attention. Many thanks to all the people I've worked with in the Biology, Chemical Sciences and Biochemistry Departments of UniPD and in the Microbiology Departement of ETHZ (Swiss).

Many thanks to Andreas Essig for the fruitful scientific collaboration and to Prof. Markus Aebi for giving me the opportunity to work in his laboratory.

My acknowledgements to Filippo for the highly scientific discussions during the lunch breaks.

Many thanks to Ilaria, who was brave enough to become my wife less than two months ago.

"A Scout smiles and whistles under all circumstances", even a PhD.

Table of contents

Summary

Understanding innate defence mechanisms is not only an exercise of pure science, but can give the scientific community new tools in fighting pathogens. In particular, the emergence of strains multi-resistant to common antibiotics gave a rush to the identification of new bioactive molecules. Possible candidates are a variety of endogenous gene-encoded antibiotics, evolved together with other humoral effectors to protect hosts from infective agents, beyond adaptive immunity.

The main target of the thesis is to get further insights in the actual "jungle" of AMPs, in particular on their mechanism of action, with the final goal of improving their performance against so-called ESKAPE pathogens. Thus, I've developed new specific methods, starting from *in silico* predictions to high-throughput characterisation of mutants, and characterized new molecules, in particular peptides derived from bivalves, as lead compounds for anti-infective drugs. To achieve these goals, I worked with different laboratories expert in peptide synthesis, structural bioinformatics or recombinant protein production.

The first chapter gives an overview on antimicrobial peptides, with particular attention to CSαβ peptides. Such peptides display features useful for the development of new antibiotics, such as high stability and immunomodulatory properties.

The second chapter presents the work done in collaboration with the Chemical sciences Department of the University of Padua on the structural characterization of CSαβ peptides. Considering the abundance of putative AMP sequences and the scarcity of effectively produced and market-oriented peptides, we developed a cysteine-guided molecular dynamics protocol for CSαβ structure refinement. This could support the research of active fragments before the recombinant production or chemical synthesis of the whole peptide. The proposed protocol is now published in the Journal of Biomolecular Structure and Dynamics, issue of Sep 2016.

As proof of application of chapter 2 protocol, the third chapter illustrates the experimental work on myticin C, a polymorphic AMP from *M. galloprovincialis*. Such molecule displays interesting immuno-related properties. Starting from a structural model, we were able to design MytC-derived active fragments showing properties similar to the whole peptide. Part of results presented in this thesis are now published in the Journal of Agricultural and Food Chemistry, issue of Oct. 2015.

Aiming to increase the repertoire of available bioactive AMPs, we selected five putative antimicrobial peptide sequences from *M. galloprovincialis* transcriptomic data and, after preliminary trials in *E. coli*, the yeast *Pichia pastoris* was selected for a more advantageous recombinant production. This work, described in the fourth chapter, was done in collaboration with ETHZ (Swiss), namely in the laboratory of Prof. Markus Aebi, where I spent six months under the supervision of Doctor Andreas Essig. Actually two proteins have been produced, one of them showing an interesting range of antimicrobial activity.

During my staying at ETHZ, I've worked on copsin, a defensin-like peptide previously identified in the fungus *Coprinopsis cinerea* and the fifth chapter describes the high throughput production of site-directed copsin mutants and the *in silico* characterization of membrane interaction of copsin and of one copsin mutant with improved features, named k-copsin.

In the appendix are reported experimental inputs supporting other research lines carried on in the Dott. Paola Venier laboratory.

Riassunto

Capire come funziona l'immunità innata non significa solo sviluppare nuove conoscenze di base, ma può fornire alla comunità scientifica nuovi strumenti per combattere organismi patogeni. In particolare, la comparsa di ceppi multi-resistenti ai comuni antibiotici ha spinto la ricerca di nuove molecole bioattive. Tra i possibili candidati vi sono antibiotici endogeni codificati da geni ed evoluti insieme ad altre molecole effettrici per proteggere ospite da agenti infettivi indipendentemente dalla immunità di tipo adattativo.

L'obiettivo principale della tesi è stato quello di cercare nuove evidenze riguardo questa giungla di AMPs, in particolare sul loro meccanismo d'azione, con l'obiettivo di migliorare la loro performance contro il gruppo di patogeni cosiddetti ESKAPE. Per fare ciò ho sviluppato nuovi metodi di indagine specifici, a partire dalla predizione *in silico* fino alla caratterizzazione di mutanti, e caratterizzato nuove molecole, in particolare derivate da bivalvi. Per raggiungere questi risultati ho collaborato con diversi laboratori esperti nella sintesi di peptidi, in bioinformatica strutturale e nella produzione ricombinante di proteine.

I peptidi antimicrobici naturali sono strutturalmente diversi, ma presentano alcune caratteristiche comuni. Il primo capitolo offre una panoramica sui peptidi antimicrobici, con particolare attenzione ai cosiddetti peptidi CSαβ che posseggono alcune caratteristiche interessanti, come elevata stabilità e proprietà immunomodulatorie.

Il secondo capitolo presenta il lavoro fatto in collaborazione con il Dipartimento di Scienze chimiche dell'Università di Padova e riguardante la caratterizzazione strutturale di peptidi CSαβ. Considerando l'abbondanza di sequenze di possibili peptidi antimicrobici e la scarsità di peptidi efficacemente prodotti e orientati al mercato, abbiamo sviluppato un protocollo bioinformatico basato sui ponti disolfuro ed in grado di migliorare predizioni di struttura. Questo potrebbe essere utile nella ricerca di frammenti attivi senza la necessità di produrre l'intero peptide. Il protocollo è stato pubblicato nel Settembre del 2016 in Journal of Biomolecular Structure and Dynamics.

Come prova dell'utilità di tale protocollo, il terzo capitolo riporta il lavoro sperimentale effettuato sulla miticina C, un peptide antimicrobico polimorfico di *M. gallopro-vincialis* con interessanti proprietà legate all'immunità. Partendo da predizioni strutturali, siamo stati in grado di disegnare dei frammenti attivi derivanti dalla miticina C aventi proprietà simili a quelle del peptide intero. Parte di questo lavoro è già stato pubblicato nell'Ottobre del 2015 nel Journal of Agricultural and Food Chemistry.

Con l'obiettivo di aumentare il repertorio di AMP bioattivi disponibili, abbiamo selezionate cinque sequenze di possibili peptidi antimicrobici dai dati trascrittomici di *M. galloprovincialis* puntando su *Pichia pastoris*, dopo i primi tentativi in *E. coli*, per una più vantaggiosa produzione ricombinante di tali peptidi. Questa parte di lavoro, descritta nel quarto capitolo, è stata svolta in collaborazione con l'ETH di Zurigo, in particolare nel laboratorio del Prof. Markus Aebi, sotto la supervisione del Dott. Andreas Essig. Per il momento siamo stati in grado di produrre due peptidi, uno dei quali ha un interessante spettro di attività antimicrobica.

Durante il periodo all'ETHZ, ho lavorato sulla copsina, una defensina identificata nel fungo *Coprinopsis cinerea*. Nel quinto capitolo è descritta la produzione di mutanti della copsina e la caratterizzazione computazionale dell'interazione a livello di membrana di due versioni della copsina, la prima wild-tipe ed una seconda con caratteristiche migliorative. L'articolo relativo ai dati presentati sta per essere sottoposto a peer review.

In appendice sono presentati i contributi sperimentali a supporto di altre linee di lavoro presenti nel laboratorio della Dott.ssa Paola Venier.

Chapter 1

Introduction

Welcome to the jungle: insights into the innate immunity peptides

Antimicrobial Peptides:

From the discovery of penicillin, natural innate defence systems have been a source of bioactive molecules, able to significantly improve human health. Among the effector molecules, Cationic Host Defence Peptides (CHDPs) or antimicrobial peptides (AMP) have been characterized in all living kingdoms, suggesting their fundamental and ancient role in the host defence against pathogens and predators (e.g. venom toxins). Especially in organisms lacking adaptive immune systems, CHDPs play a primary role in pathogen recognition, immune modulation, enzyme inhibition, immune cells chemotaxis, tumour suppression, signalling, adaptation to abiotic stresses, wound repair and angiogenesis as well as in direct pathogen killing[1,2]. More than 16900 antimicrobial peptides sequences have been deposited in the AMP databases[3]. This is a real "jungle", where peptides differ for many features, from primary sequence to compartmentalization and mechanism of action. But all of them have a primary objective: to kill the intruder.

CHDPs with anti-infective and immunomodulatory properties show some common characteristics, despite the great variety in primary sequence, mode of action and structural organization. Most of these peptides are relatively short, between 6 and 80 amino acids, positively charged owing to arginine/lysine residues and typically have between 40% and 70% of hydrophobic residues. Most of the times, AMP are encoded as prepropeptides and then processed to the active form. In addition to the localization signal, the propeptide act as chaperone and protector against auto cytotoxicity[4–6]. Some anionic antimicrobial peptides have also been discovered, but their mechanism of action remains unclear[7,8]. AMPs are traditionally classified in four families based on the primary sequence and secondary structure: β-sheet peptides stabilized by disulphides bonds; peptides prevalently in α-helical conformation; peptides essentially unfolded and enriched in a particular amino-acid (Trp-rich, Pro-rich); cyclic peptides with one or more disulphides[9]. At the same time, each family exhibits several different mechanisms of action, most involving membrane dysfunction, inhibition of extracellular biopolymer synthesis or inhibition of intracellular functions (Figure 1)[10,11]. The next sections illustrate a particular class of antimicrobial peptides called Cysteine Stabilized AMPs and their subfamilies.

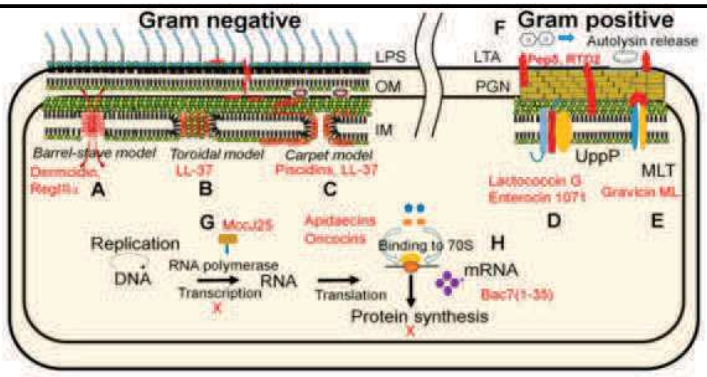


Figure 1: Membrane channel formation (A) and transmembrane pore formation are proposed for dermicidin and the C-type lectin RegIIIα, respectively. It has been proposed that human LL-37 forms a toroidal pore (B), while other antimicrobial peptides, such as piscidins, locate on the membrane surface and follow the carpet model (C). Lactococcin G and Enterocin 1071 bind to UppP, an enzyme involved in cell wall synthesis (D). Gravicin ML binds to maltose ABC receptors (E). RTD2 as well as lantibiotic Pep5 cause the release of autolysin interacting with bacterial membranes (F). Intracellularly, microcin J25 inhibits the RNA polymerase (G), while the prolin-rich apidaecins, oncocins and Bac7 block protein synthesis inhibiting ribosomal proteins (H). Abbreviations used in the figures are LPS, lipopolysaccharide; OM, outer membrane; IM, inner membrane; PGN, peptidoglycan; LTA, lipoteichoic acid; MLT, maltose transporter

Disulfide-Stabilized antimicrobial peptides

Disulfide-stabilized antimicrobial peptides are a particular class of AMPs characterized by one or more disulfide bridges conferring a stable peptide conformation. For this reason, with few exceptions, they are also known as Cysteine Stabilized αβ (CSαβ) peptides[12–14]. The widespread presence of these peptides in multicellular organisms suggests their fundamental role in the host organism[15]. In most of the cases, disulphide bridges have been demonstrated to be essential for antimicrobial activity, revealing a strict connection between tertiary structure and function, even if some peptides show the same or better activity when disulfides are reduced, such as Myticin C and hBD1[16–20]. Overall, the cysteine bonds confer proteolytic, thermal and chemical stability to peptides[13,19].

Among the CSαβ peptides we can distinguish several families according to the secondary structure motifs, disulfide array and functional properties. The biggest class of CSαβ peptides are defensins, followed by some cathelicidin, knottins, cyclotides and lipid-transfer proteins. The CSαβ classification is not unique in literature, i.e. knottins are considered either as part of the defensin family or as autonomous group, according to the applied experimental approach[13,21].

Defensins

Defensins are a large group of highly differentiated CSαβ peptides, expressed by most eukaryotes, with a broad range antimicrobial activity but also with different roles including signalling and chemotactic activity[22,23]. All known defensins are produced from a precursor with a N-terminal signal peptide directing the peptide to the endoplasmic reticulum, where the formation of disulfide bridges is guided by foldases and chaperones. After folding, defensins are directed by a propeptide to secretion or to intracellular compartments, such as phagolysosomes[24,25]. Most defensins are characterized by the so-called γ-core motif, formed by two cysteine-stabilized antiparallel β-strands, dextromeric or levomeric, and presenting a conserved amino acids pattern Gly-X-Cys or Cys-X-Gly[12]. In some cases the γ-core is sufficient for antimicrobial activity[26], or it can be stabilized by additional secondary structures[27]. Moreover, some defensins were demonstrated to form dimers and oligomers, resulting in increased antimicrobial activity[28,29]. Defensins were historically classified in two groups, fungal / plant / invertebrate defensins and vertebrate defensins. Recently Shafee, T. M. A., et al. proposed a new classification, based on conserved structural features and disulfide array[30]. This classification, largely consistent with the previous ones, demonstrated a different but convergent evolutionary pathway for the two superfamilies named *cis*- and *trans*-defensins[31]. The main difference between these two groups has been recognized in the ancestral conserved structural motif: while cis-defensins present two conserved disulfides between one of the two β-strands and an α-helix, trans-defensins have two conserved cysteines in the C-terminal β-strand, pointing in opposite direction and bonding to two different secondary elements.

Cis-defensins

Cis-defensins are present mainly in arthropods and nematodes, as well as molluscs, annelids and cnidarians. They are also present in fungi and spermatophyte plants and they have been also found in cephalochordates. Interestingly, no cis-defensin has been found so far in other *deutrostomia*, *locophotrozoa* or *viridiplantae*, a fact attributed to gene loss phenomena[32] or horizontal gene transfer[33]. Defensin-like genes have been reported also in prokaryotes, but an evolutionary link is still under investigation[34]. Nevertheless, cysteine connectivity differentiated during evolution to generate 22 different disulfide arrays (Figure 2)[31].

Plant defensins are reported to target mainly the pathogen membranes, and in some cases to inhibit enzymes. Most of these defensins were isolated in seeds and roots, but evidence of their expression was collected also in other tissues[35]. They can present eight (8C) or ten (10C) cysteines involved in disulfide bonds. One of the best characterized 8C defensin of plants is NaD1 from *Nicotiana alata*. NaD1 is active against plant pathogens *Botrytis cinerea* and *Fusarium oxysporus*[28,36]. Its mechanism of action involves oxidative killing and membrane perturbation and interaction with phosphatidylinositol 4,5 bisphosphate, as demonstrated with molecular dynamics simulations[37,38]. Other plant defensins are able to selectively inhibit insects α-amylases[39].

Fungi became part of defensins-producing organisms in 2005, with the discovery of plectasin from *Pseudoplecatania nigrella*[40]. Plectasin has been demonstrated to be active against Streptococci and Staphylococci, specifically binding Lipid II of bacterial membranes. Moreover, a modified version named NZ2114 showed potent activity against Methicillin-Resistant *Staphylococcus aureus* (MRSA) in a rabbit model[41].

Among the invertebrate AMP, Drosomycin from *Drosophila melanogaster* has been extensively characterized. First identified in 1994[42], it showed strong activity against filamentous fungi and parasites such as *Plasmodium berghei*[43,44]. The Drosomycin mechanism of action appears to involve interactions with sphingolipids and consequent membrane perturbation[45,46]. Among molluscan, *Crassostrea gigas* defensins (Cg-Defs) are constitutively expressed in the mantle to avoid pathogen colonization[47]. Cg-Defs act synergistically with other immune proteins, in order to overcome inhibiting factors such as high salt concentration and slightly basic pH typical of marine environment[48].

Trans-defensins

Trans-defensins have been found in vertebrates, molluscs, arthropods, lancelets and sea anemones. Different trans-defensins families can have different secondary structure, with different elements bonding to the conserved cysteine core. It is possible to sort trans-defensins in five different families on the basis of disulfide connectivity and structural features. Six α-defensins (HNP 1-4, human neutrophil peptides, and HD 5-6) and four β-defensins (HBD) have been identified in humans[49]. HNP 1-4 display a large copy number and high polymorphism. They are typically expressed in leukocytes and stored in primary granules after processing. HD 5-6 are produced in Paneth cells and

exhibits chemotactic proprieties[50]. The α-defensins are released in the extracellular space in response to pathogen recognition, together with other defence effectors[49].

The β-defensin genes are clustered principally on chromosome 8p23.1 and consist in two exons, the second one encoding for the mature peptide. It has been demonstrated the ability of such peptides to kill bacteria[51], viruses[52] and yeasts[20] and, even if the antimicrobial mechanism is not completely clarified, hBD3 is able to bind lipid-II and disrupt cell wall biosynthesis[53] while membrane perturbation or ion channel interactions are reported for other defensins[54,55]. Homologs of β-defensins have been found in other vertebrates, one of the most interesting being the crotamin-myotoxin in snake venom[55].

As regards θ-defensins, they are present in humans as pseudogene with a premature stop codon and have been identified in *Rhesus macaque* as cyclic peptide formed by two truncated α-defensins[56]. These molecules are currently under investigation for their anti-HIV properties[57].

Big defensins and Defensin-Like Peptide (DLP) are invertebrate trans-defensins, identified for the first time in the horseshoe crab *Tachypleus tridentatus* in 2010 and in the sea anemone *Stichodactyla helianthus* in 2016, respectively[58,59] (Figure 2).

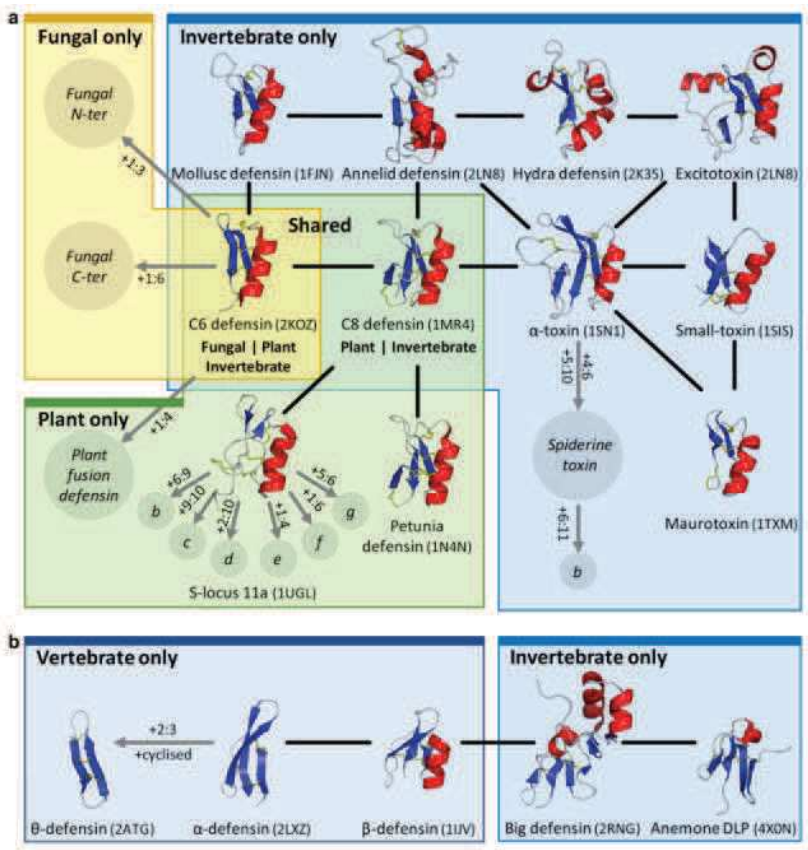


Figure 2: Relatedness within the cis- and trans-defensins. Evidence for common origin in the (a) cis-defensins and (b) trans-defensins. Structures are shown for peptides with known structures, classes with unresolved structures are represented by italicised names in circles. Putative disulphides unique to a class are denoted as x:y where x and y are the additional cysteines involved in the disulphide. Uncharacterised variants with additional disulphides are denoted by single letters (e.g. S-locus 11b, etc.). Black lines indicate homology evidence from structural similarity, grey lines indicate evidence from gene structure and organisation. The PDB codes for the proteins are given in parentheses. Structures are organised by kingdom, with a fungal representative as an example of the shared C6 defensins and a plant representative for the shared C8 defensins (Shafee et al., 2016).

Other CSαβ

Cathelicidins are a class of AMPs expressed in humans and other animal species. They possess a conserved precursor and a highly polymorphic mature peptide at the C-terminal, with structural conformations ranging from α-helix[60] to β-sheet[26] and linear[61]. Only some cathelicidins are CSαβ peptides, for example protegrins are stabilized by disulfide bonds and share structural features with θ-defensins suggesting a convergent selection of structural features for common functions[62].

Knottins are a large class of peptides with multiple functions characterized by a cysteine knot formed by three disulfides. Antimicrobial knottins have been identified in plants, fungi, insects and spiders. Plant cyclotides are a particular sub-class of this family characterized by a circular backbone[63].

Lipid-transfer-proteins (LPT) are plant AMPs of about 9 kDa with four disulfide bonds essential to the antimicrobial function. These proteins can present antimicrobial, antifungal, antiviral and antiproliferative activities[64].

AntiMicrobial Peptides applications

AMPs have been regarded for a long time as potential alternatives to classical antibiotics for fighting emerging multi-resistant strains[65]. Moreover, immunomodulatory properties of these peptides can be considered a further pharmaceutical benefit compared to common antibiotics[1]. Actually, the initial enthusiasm towards AMPs crashed on the difficulties to develop marketable molecules. In fact, despite over 2000 patents regarding AMPs, only few arrived to commercialization[66]. From a clinical point of view, potent AMPs are often cytotoxic or toxic in vivo, preventing systemic applications[67]. Examples are gramicidin S[68] and indolicidin[69]. To our knowledge, there are actually about 10 antimicrobial peptides in preclinical or clinical phase, with the most promising being Pexiganan (Locilex®), derived from African Clawed Frog magainin and actually in phase III trial as topic agent against diabetes foot infections[70]. Iseganan (IB-367), a derivate of protegrin 1, is in phase III for treatment of oral mucositis and pneumonia. An oral care product called P113+, derived from Histatin, is actually commercialized by Pacgen Biopharmaceuticals Corp. and passed clinical phase II trial for the use against oral candidiasis[71]. Omiganan, an analogue of indolicidin is in phase III as topical cream for treatment of skin infections and catheter associated bacteria[72]. Finally Brilacidin, a mimetic of defensins, is in phase III for system administration against Gram+ skin infections[73].

From a non-clinical point of view, there are some other examples of commercial AMPs. The bacteriocin nisin from *Lactococcus lactis* is used as food preservative since 1988[74] and it is currently under investigation for potential biomedical applications[75]. At the same time, many AMPs are under investigation as nutraceutics in swine and poultry production[76] and as cosmetics ingredients[77].

References

1. Hilchie, A. L., Wuerth, K. & Hancock, R. E. W. Immune modulation by multifaceted cationic host defense (antimicrobial) peptides. *Nat. Chem. Biol.* **9,** 761–768 (2013).

2. Mansour, S. C., Pena, O. M. & Hancock, R. E. W. Host defense peptides: front–line immunomodulators. *Trends Immunol.* **35,** 443–450 (2014).

3. Aguilera–Mendoza, L. *et al.* Overlap and diversity in antimicrobial peptide databases: compiling a non–redundant set of sequences. *Bioinformatics* **31,** 2553–2559 (2015).

4. Zou, G., de Leeuw, E., Lubkowski, J. & Lu, W. Molecular Determinants for the Interaction of Human Neutrophil α Defensin 1 with its Propeptide. *J. Mol. Biol.* **381,** 1281–1291 (2008).

5. Lay, F. T. *et al.* The C–terminal propeptide of a plant defensin confers cytoprotective and subcellular targeting functions. *BMC Plant Biol.* **14,** 41 (2014).

6. Wu, Z. *et al.* Impact of pro segments on the folding and function of human neutrophil alpha–defensins. *J. Mol. Biol.* **368,** 537–549 (2007).

7. Becucci, L., Valensin, D., Innocenti, M. & Guidelli, R. Dermcidin, an anionic antimicrobial peptide: influence of lipid charge, pH and Zn2+ on its interaction with a biomimetic membrane. *Soft Matter* **10,** 616–626 (2013).

8. Phoenix, D. A., Dennison, S. R. & Harris, F. in *Antimicrobial Peptides* 83–113 (Wiley–VCH Verlag GmbH & Co. KGaA, 2013).

9. Hancock, R. E. W. & Sahl, H.–G. Antimicrobial and host–defense peptides as new anti–infective therapeutic strategies. *Nat. Biotechnol.* **24,** 1551–1557 (2006).

10. Yeaman, M. R. & Yount, N. Y. Mechanisms of Antimicrobial Peptide Action and Resistance. *Pharmacol. Rev.* **55,** 27–55 (2003).

11. Wang, G. *et al.* Antimicrobial Peptides in 2014. *Pharmaceuticals* **8,** 123–150 (2015).

12. Yount, N. Y. & Yeaman, M. R. Multidimensional signatures in antimicrobial peptides. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7363–7368 (2004).

13. Schmitt, P., Rosa, R. D. & Destoumieux–Garzón, D. An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes. *Biochim. Biophys. Acta BBA - Biomembr.* **1858,** 958–970 (2016).

14. Dias, R. de O. & Franco, O. L. Cysteine–stabilized αβ defensins: From a common fold to antibacterial activity. *Peptides* **72,** 64–72 (2015).

15. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415,** 389–395 (2002).

16. Domeneghetti, S. *et al.* Structural and Antimicrobial Features of Peptides Related to Myticin C, a Special Defense Molecule from the Mediterranean Mussel Mytilus galloprovincialis. *J. Agric. Food Chem.* **63,** 9251–9259 (2015).

17. Ma, B. *et al.* The Disulfide Bond of the Peptide Thanatin Is Dispensible for Its Antimicrobial Activity In Vivo and In Vitro. *Antimicrob. Agents Chemother.* **60,** 4283–4289 (2016).

18. Haag, A. F. *et al.* Role of cysteine residues and disulfide bonds on the activity of a legume root nodule–specific, cysteine–rich peptide. *J. Biol. Chem.* jbc.M111.311316 (2012). doi:10.1074/jbc.M111.311316

19. Essig, A. *et al.* Copsin, a novel peptide–based fungal antibiotic interfering with the peptidoglycan synthesis. *J. Biol. Chem.* **289,** 34953–34964 (2014).

20. Schroeder, B. O. *et al.* Reduction of disulphide bonds unmasks potent antimicrobial activity of human β–defensin 1. *Nature* **469,** 419–423 (2011).

21. Craik, D. J., Čemažar, M., Wang, C. K. L. & Daly, N. L. The cyclotide family of circular miniproteins: Nature's combinatorial peptide template. *Pept. Sci.* **84,** 250–266 (2006).

22. Miles, K. *et al.* Dying and Necrotic Neutrophils Are Anti-Inflammatory Secondary to the Release of α-Defensins. *J. Immunol.* **183,** 2122–2132 (2009).

23. Röhrl, J., Yang, D., Oppenheim, J. J. & Hehlgans, T. Human beta-defensin 2 and 3 and their mouse orthologs induce chemotaxis through interaction with CCR2. *J. Immunol. Baltim. Md 1950* **184,** 6688–6694 (2010).

24. Satchell, D. P. *et al.* Interactions of mouse Paneth cell alpha-defensins and alpha-defensin precursors with membranes. Prosegment inhibition of peptide association with biomimetic membranes. *J. Biol. Chem.* **278,** 13838–13846 (2003).

25. Wu, Z. *et al.* From pro defensins to defensins: synthesis and characterization of human neutrophil pro alpha-defensin-1 and its mature domain. *J. Pept. Res. Off. J. Am. Pept. Soc.* **62,** 53–62 (2003).

26. Steinberg, D. A. *et al.* Protegrin-1: a broad-spectrum, rapidly microbicidal peptide with in vivo activity. *Antimicrob. Agents Chemother.* **41,** 1738–1742 (1997).

27. Schneider, T. *et al.* Plectasin, a fungal defensin, targets the bacterial cell wall precursor Lipid II. *Science* **328,** 1168–1172 (2010).

28. Lay, F. T. *et al.* Dimerization of plant defensin NaD1 enhances its antifungal activity. *J. Biol. Chem.* **287,** 19961–19972 (2012).

29. Hoover, D. M. *et al.* The structure of human beta-defensin-2 shows evidence of higher order oligomerization. *J. Biol. Chem.* **275,** 32911–32918 (2000).

30. Shafee, T. M. A., Lay, F. T., Phan, T. K., Anderson, M. A. & Hulett, M. D. Convergent evolution of defensin sequence, structure and function. *Cell. Mol. Life Sci.* 1–20 (2016). doi:10.1007/s00018-016-2344-5

31. Shafee, T. M. A., Lay, F. T., Hulett, M. D. & Anderson, M. A. The Defensins Consist of Two Independent, Convergent Protein Superfamilies. *Mol. Biol. Evol.* **33,** 2345–2356 (2016).

32. Hughes, A. L. & Friedman, R. Shedding genomic ballast: extensive parallel loss of ancestral gene families in animals. *J. Mol. Evol.* **59,** 827–833 (2004).

33. Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A. & Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **16,** 50 (2015).

34. Zhu, S. Evidence for myxobacterial origin of eukaryotic defensins. *Immunogenetics* **59,** 949–954 (2007).

35. Tam, J. P., Wang, S., Wong, K. H. & Tan, W. L. Antimicrobial Peptides from Plants. *Pharmaceuticals* **8,** 711–757 (2015).

36. Lay, F. T., Brugliera, F. & Anderson, M. A. Isolation and Properties of Floral Defensins from Ornamental Tobacco and Petunia. *Plant Physiol.* **131,** 1283–1293 (2003).

37. Hayes, B. M. E. *et al.* Identification and Mechanism of Action of the Plant Defensin NaD1 as a New Member of the Antifungal Drug Arsenal against Candida albicans. *Antimicrob. Agents Chemother.* **57,** 3667–3675 (2013).

38. Payne, J. A. E. *et al.* The plant defensin NaD1 introduces membrane disorder through a specific interaction with the lipid, phosphatidylinositol 4,5 bisphosphate. *Biochim. Biophys. Acta BBA - Biomembr.* **1858,** 1099–1109 (2016).

39. Bloch, C. & Richardson, M. A new family of small (5 kDa) protein inhibitors of insect α-amylases from seeds or sorghum (Sorghum bicolor (L) Moench) have sequence homologies with wheat γ-purothionins. *FEBS Lett.* **279,** 101–104 (1991).

40. Mygind, P. H. *et al.* Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus. *Nature* **437,** 975–980 (2005).

41. Xiong, Y. Q. *et al.* Efficacy of NZ2114, a novel plectasin-derived cationic antimicrobial peptide antibiotic, in experimental endocarditis due to methicillin-resistant Staphylococcus aureus. *Antimicrob. Agents Chemother.* **55,** 5325–5330 (2011).

42. Fehlbaum, P. *et al.* Insect immunity. Septic injury of Drosophila induces the synthesis of a potent antifungal peptide with sequence homology to plant antifungal peptides. *J. Biol. Chem.* **269,** 33159–33163 (1994).

43. Tian, C. *et al.* Gene expression, antiparasitic activity, and functional evolution of the drosomycin family. *Mol. Immunol.* **45,** 3909–3916 (2008).

44. Zhang, Z.-T. & Zhu, S.-Y. Drosomycin, an essential component of antifungal defence in Drosophila. *Insect Mol. Biol.* **18,** 549–556 (2009).

45. Gao, B. & Zhu, S.-Y. Differential potency of drosomycin to Neurospora crassa and its mutant: implications for evolutionary relationship between defensins from insects and plants. *Insect Mol. Biol.* **17,** 405–411 (2008).

46. Ferket, K. K. A., Levery, S. B., Park, C., Cammue, B. P. A. & Thevissen, K. Isolation and characterization of Neurospora crassa mutants resistant to antifungal plant defensins. *Fungal Genet. Biol.* **40,** 176–185 (2003).

47. Gueguen, Y. *et al.* Characterization of a defensin from the oyster Crassostrea gigas. Recombinant production, folding, solution structure, antimicrobial activities, and gene expression. *J. Biol. Chem.* **281,** 313–323 (2006).

48. Schmitt, P., Lorgeril, J. de, Gueguen, Y., Destoumieux-Garzón, D. & Bachère, E. Expression, tissue localization and synergy of antimicrobial peptides and proteins in the immune response of the oyster Crassostrea gigas. *Dev. Comp. Immunol.* **37,** 363–370 (2012).

49. Lehrer, R. I. Primate defensins. *Nat. Rev. Microbiol.* **2,** 727–738 (2004).

50. Grigat, J., Soruri, A., Forssmann, U., Riggert, J. & Zwirner, J. Chemoattraction of Macrophages, T Lymphocytes, and Mast Cells Is Evolutionarily Conserved within the Human α-Defensin Family. *J. Immunol.* **179,** 3958–3965 (2007).

51. Guaní-Guerra, E., Santos-Mendoza, T., Lugo-Reyes, S. O. & Terán, L. M. Antimicrobial peptides: general overview and clinical implications in human health and disease. *Clin. Immunol. Orlando Fla* **135,** 1–11 (2010).

52. Wilson, S. S., Wiens, M. E. & Smith, J. G. Antiviral mechanisms of human defensins. *J. Mol. Biol.* **425,** 4965–4980 (2013).

53. Sass, V. *et al.* Human β-Defensin 3 Inhibits Cell Wall Biosynthesis in Staphylococci. *Infect. Immun.* **78,** 2793–2800 (2010).

54. Brogden, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* **3,** 238–250 (2005).

55. Yount, N. Y. *et al.* Selective reciprocity in antimicrobial activity versus cytotoxicity of hBD-2 and crotamine. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 14972–14977 (2009).

56. Tang, Y. Q. *et al.* A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science* **286,** 498–502 (1999).

57. Daly, N. L. *et al.* Retrocyclin-2: Structural Analysis of a Potent Anti-HIV θ-Defensin,. *Biochemistry (Mosc.)* **46,** 9920–9928 (2007).

58. Kawabata, S. Immunocompetent molecules and their response network in horseshoe crabs. *Adv. Exp. Med. Biol.* **708,** 122–136 (2010).

59. Tysoe, C. *et al.* Potent Human α-Amylase Inhibition by the β-Defensin-like Protein Helianthamide. *ACS Cent. Sci.* **2,** 154–161 (2016).

60. Wang, G. Structures of Human Host Defense Cathelicidin LL-37 and Its Smallest Antimicrobial Peptide KR-12 in Lipid Micelles. *J. Biol. Chem.* **283,** 32637–32643 (2008).

61. Rozek, A., Friedrich, C. L. & Hancock, R. E. Structure of the bovine antimicrobial peptide indolicidin bound to dodecylphosphocholine and sodium dodecyl sulfate micelles. *Biochemistry (Mosc.)* **39,** 15765–15774 (2000).

62. Knyght, I., Clifton, L., Saaka, Y., Lawrence, M. J. & Barlow, D. J. Interaction of the Antimicrobial Peptides Rhesus θ-Defensin and Porcine Protegrin-1 with Anionic Phospholipid Monolayers. *Langmuir ACS J. Surf. Colloids* **32,** 7403–7410 (2016).

63. Rosengren, K. J., Daly, N. L., Plan, M. R., Waine, C. & Craik, D. J. Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J. Biol. Chem.* **278,** 8606–8616 (2003).

64. Carvalho, A. de O. & Gomes, V. M. Role of plant lipid transfer proteins in plant cell physiology–a concise review. *Peptides* **28,** 1144–1153 (2007).

65. Malmsten, M. Antimicrobial peptides. *Ups. J. Med. Sci.* **119,** 199–204 (2014).

66. Waghu, F. H., Barai, R. S., Gurung, P. & Idicula-Thomas, S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.* **44,** D1094–1097 (2016).

67. Yount, N. Y. & Yeaman, M. R. Emerging Themes and Therapeutic Prospects for Anti-Infective Peptides. *Annu. Rev. Pharmacol. Toxicol.* **52,** 337–360 (2012).

68. Kondejewski, L. H. *et al.* Modulation of Structure and Antibacterial and Hemolytic Activity by Ring Size in Cyclic Gramicidin S Analogs. *J. Biol. Chem.* **271,** 25261–25268 (1996).

69. Ahmad, I., Perkins, W. R., Lupan, D. M., Selsted, M. E. & Janoff, A. S. Liposomal entrapment of the neutrophil-derived peptide indolicidin endows it with in vivo antifungal activity. *Biochim. Biophys. Acta* **1237,** 109–114 (1995).

70. Mangoni, M. L., McDermott, A. M. & Zasloff, M. Antimicrobial peptides and wound healing: biological and therapeutic considerations. *Exp. Dermatol.* **25,** 167–173 (2016).

71. Han, J., Jyoti, M. A., Song, H.-Y. & Jang, W. S. Antifungal Activity and Action Mechanism of Histatin 5-Halocidin Hybrid Peptides against Candida ssp. *PLoS ONE* **11,** (2016).

72. Sader, H. S., Fedler, K. A., Rennie, R. P., Stevens, S. & Jones, R. N. Omiganan Pentahydrochloride (MBI 226), a Topical 12-Amino-Acid Cationic Peptide: Spectrum of Antimicrobial Activity and Measurements of Bactericidal Activity. *Antimicrob. Agents Chemother.* **48,** 3112–3118 (2004).

73. Mensa, B., Howell, G. L., Scott, R. & DeGrado, W. F. Comparative Mechanistic Studies of Brilacidin, Daptomycin, and the Antimicrobial Peptide LL16. *Antimicrob. Agents Chemother.* **58,** 5136–5145 (2014).

74. Delves-Broughton, J., Blackburn, P., Evans, R. J. & Hugenholtz, J. Applications of the bacteriocin, nisin. *Antonie Van Leeuwenhoek* **69,** 193–202 (1996).

75. Shin, J. m. *et al.* Biomedical applications of nisin. *J. Appl. Microbiol.* **120,** 1449–1465 (2016).

76. Wang, S., Zeng, X., Yang, Q. & Qiao, S. Antimicrobial Peptides as Potential Alternatives to Antibiotics in Food Animal Industry. *Int. J. Mol. Sci.* **17,** (2016).

77. Rahnamaeian, M. & Vilcinskas, A. Short antimicrobial peptides as cosmetic ingredients to deter dermatological pathogens. *Appl. Microbiol. Biotechnol.* **99,** 8847–8855 (2015).

Chapter 2

A molecular dynamics strategy for CSαβ peptides disulfide-assisted model refinement.

Franzoi M.[1#], Sturlese M.[2], Bellanda M.[3], Mammi S.[3]

[1]Department of Biology, University of Padova , Via Ugo Bassi 58/B, 35131 Padova, Italy.
[2]Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences, University of Padova, Via Marzolo 5, 35131, Padova, Italy.
[3]Department of Chemical Sciences, University of Padova , Via Marzolo 1, 35131 Padova, Italy.

Journal of Biomolecular Structure and Dynamics. 2016 Sep 1:1-25.

Contribution:
Initial sets creation
Disulfides algorithm
Molecular dynamics set-up
Clustering algorithm
Data analysis

Abstract

Many cysteine-stabilized antimicrobial peptides from a variety of different living organisms could be good candidates for the development of anti-infective agents. In the absence of experimentally obtained structural data, peptide modelling is an essential tool for understanding structure-activity relationships and for optimizing the bioactive moieties. Focussing on cysteine-rich peptide structures, we reproduced the case of structure predictions in the so-called midnight-zone. We developed our protocol on a training set derived by clustering the available Cysteine Stabilized αβ (CSαβ) structures in nine different representative families and tested it on peptides randomly selected from each family. Starting from draft models, we tested a structure-based disulfide predictor and we used cysteine distances as constraints during molecular dynamics. Finally, we proposed an analysis for final structure selection. Accordingly, we obtained a mean Root Mean Square Deviation improvement of 21% for the test set. Our findings demonstrate that it is possible to predict the network of disulfide bridges in cysteine stabilized peptides and to use this result to improve the accuracy of structural predictions. Finally, we applied this new procedure to predict the structure of royalisin, a cysteine-rich peptide with unknown structure.

Introduction

Antimicrobial peptides (AMP) are a class of compounds useful to extend the pool of bioactive molecules to fight pathogens with increasing resistance towards common antibiotics[1]. Among the AMP variety, Cysteine-stabilized αβ (CSαβ) peptides could be good candidates for the development of new drugs because of their stability, broad antibacterial spectrum and, often, low haemolytic activity[2]. Usually, these peptides show a highly conserved cysteine array but differ substantially in primary sequence[3]. In some cases, the production of the whole peptide is difficult and not necessary to obtain an active molecule. In fact, the antimicrobial activity can be assured by only a specific fragment of the molecule, suggesting an easier and less expensive production of a minimal active peptide, instead of the whole peptide[4,5]. Since the number of fully characterized AMPs to date is only a small fraction of the potential antimicrobials identified from steadily growing genomic and transcriptomic data, reliable predictive models can be a good starting point to design possible variants and new active fragments[6].

Advances in bioinformatics have provided researchers with powerful "Single click" modelling servers based on ab initio, fold-recognition, homology, or mixed approaches[7]. On the one hand, such tools usually offer good results in terms of overall topology but, on the other hand, they can produce bias in the side chain orientation, disulfide connection and secondary structures arrangement, especially when the peptide of interest has low sequence similarity with deposited structures[8]. These errors could generate misinterpretations of structure-activity relationships and lead to arduous rational peptide design. To overcome such problems, accurate structure refinement is often necessary.

In the case of cysteine stabilized proteins, structural model improvements can be achieved using disulfide-guided protocols. In fact, especially in the case of ab initio folding tools or low homology modelling, the introduction of contact restraints can lead to better predictions[9]. Moreover, the disulfide array can support sequence homology as an index of structure similarity[10]. Many tools are now available for sequence-based disulfide prediction, often based on Support Vector Machines (SVM) methods[11]. The use of draft structures for distance-based disulfide prediction has been previously explored alone[12] or in tandem with other parameters[13]. The smart use of molecular dynamics protocols has been demonstrated to be a helpful tool to refine homology modelled structures[14] and has been effectively applied in the case of structurally uncharacterized AMPs[15].

We have previously used molecular dynamics (MD) to structurally characterize Mytilin C, an antimicrobial and antiviral peptide from *Mytilus galloprovincialis*. The obtained results lead to the identification of a region composed by highly variable amino acid positions and the minimal active peptide sequence[4].

The present study describes a distance-based disulfide prediction algorithm for CSαβ antimicrobial peptides, using initial structures obtained with I-TASSER and excluding templates with more than 20% sequence identity, to reproduce the case of midnight-

zone predictions[16]. A disulfide-guided protocol was created for the refinement and structure selection of CSαβ peptides based on molecular dynamics simulations (Figure 1). We demonstrate that this protocol is able to effectively predict disulfide bonds of selected peptides and to improve the overall quality of the model compared to the starting structure.

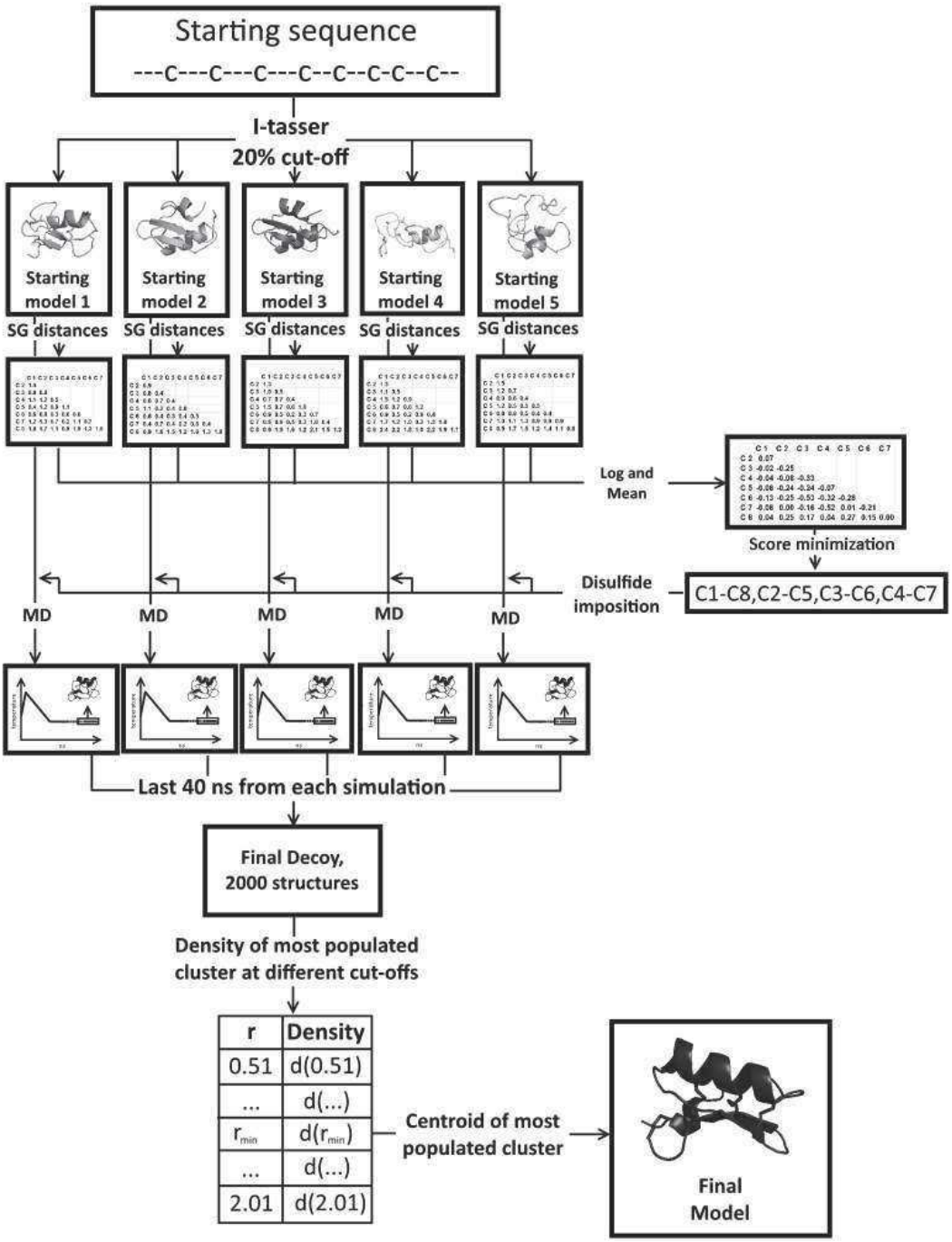


Figure 1: Refinement protocol workflow: sequences of selected peptides were used for modelling, excluding templates with more than 20% identity. Combinations of logarithmic means of cysteine distances were analysed to minimize the score. Disulphides were imposed to each starting model and sent to molecular dynamics simulations. Total 400 trajectories snapshots from each simulation were joined and clustered basing on structural similarity at different cut-offs. The centroid of the most populated cluster with the lowest density was selected as final model.

Material and methods

Training and test sets

From the Antimicrobial Peptide Database 2 (APD2)[17], 55 entries were selected to form the initial data set. The so-called Combined Helix and Beta (packed) peptides in the APD2 with known structures deposited in the Protein Data Bank (PDB) were filtered to exclude chains longer than 80 amino acids or with less than four cysteines[18]. Protein structures were clustered according to their paired TM-scores[19] using ProCKSI-Server[20]. Clustering cut-off was set to 0.3, leading to nine different clusters. The training set, used for parameter optimization, was obtained by selecting the cluster centroids, defined as the structure with the highest average TM-score among the models of the same family. The test set included one randomly selected structure from each cluster and was used to validate the refinement protocol proposed in this study.

Starting model generation

The sequences of all selected peptides were retrieved from the PDB and the initial models were computed using the I-TASSER suite (v. 4.3) without imposing disulfide bonds and excluding templates with more than 20% identity[21]. The obtained ensemble of five models for each peptide was submitted to the refinement protocol. The disulfide connectivity was predicted using a home-made script. In brief, the logarithm of the distance between each pair of sulphur atoms was calculated. Then, the means of these pairwise distances among the five starting models were used to generate a score matrix and scanned for the optimal combination of disulfides, defined as the minimal sum of the scores for all the possible combinations of bonds. Accuracy was evaluated calculating Q_C for cysteines and Q_P for proteins, defined as C_C/T_C and C_P/T_P, respectively, where C_C is the number of correctly assigned cysteines and T_C the total number of cysteines in the starting sequences, C_P the number of proteins with all the disulfides correctly assigned and T_P the total number of proteins in the set.

Molecular dynamics simulations

The molecular dynamics simulations were designed to evaluate the possibility of increasing the accuracy of the predicted structures. In order to perform such calculation we take advantage of accessing Brutus cluster of ETH in Zurich. The Gromacs 4.6[22] package was used as the MD engine. The correct disulfide connectivity was imposed during GROMACS model generation through the pdb2gmx command. All the simulations were performed using 2 fs time steps, in a triclinic box with edge distance from the molecule

set to 1.2 nm and filled with water molecules described by the SPC/E water model. Initially, 20000 steps of energy minimization at 300 K were performed using a steepest descent algorithm. For the simulation, the long-range Coulomb interactions were computed with the Particle Mesh Ewald (PME) method[23]. The V-rescale algorithm[24] for temperature coupling (300 K) and the Parrinello-Rahman algorithm for pressure coupling at 1.0 atm[25] were imposed. For neighbour searching, a Verlet scheme with a cut-off of 1.2 was set and the grip update method was imposed. Then, 5000 steps of solvent equilibration were performed using an Isothermal-isobaric (NPT) ensemble and applying a positional constraint on the peptide. The leap-frog integrator was used to integrate Newton's equations of motion. A non-bonded cut-off distance of 9 Å was imposed. The next step was a 10 ns Simulated Annealing (SA). The initial temperature was set at 300 K, then it was increased, in 1 ns, to 365 K and 340 K for protein and water, respectively, and then decreased for 8 ns to reach 300 K, were it was maintained for 1 ns. The Coulomb type for the SA was switched to Reaction-field and the non-bonded cut-off to 1.2 nm. Finally, a free dynamics of 100 ns at 300 K was performed maintaining the same parameters as for the SA. All the simulations were performed using the CHARMM27 force field[26].

Clustering

For each peptide of the training and test sets, structures from the last 40 ns of the five independent simulations were sampled every 0.1 ns, and then joined to form a 2000 structures decoy finally clustered using the GROMOS algorithm from GROMACS[27]. For final cluster selection, several clusterings were performed, changing the cut-off parameter. Specifically, the ratio *r* between the cut-off and the number of amino acids of the target peptide was ranged from 0.51 to 2.01 in 0.01 steps following the equation:

$$\text{Cutoff } (r) = N_{AA} * r$$

Thus, only the most populated cluster was extracted for each r value and, within this group of selected clusters, the one with the lowest density was kept and its centroid selected as representative structure.

Density is defined as:

$$D = N / ([RMSD] / N_{tot})$$

were N is the number of structures in the cluster, [RMSD] is the mean RMSD of the centroid compared to the other structures in the cluster, N_{tot} is the total number of structures in the decoy[28]. Final centroids were further refined with a 100 step minimization using the parameters described above.

Results and discussion

Generation of the training and test sets

Clustering of the cysteine-packed αβ peptides is consistent with a similar analysis recently published[29] (Supplementary Figure 1). Both training and test sets are representative of different families of AMPs arising from several organism phyla. Protein 2rng, a big defensin from *Tachypleus tridentatus*, showed a remarkable difference from all the other selected AMPs, clustering by itself. To maintain the same number of proteins in the training set and in the test set, two proteins were selected from the most populated cluster, i.e. cluster number 4 (Table 1).

	Cluster	PDB	Organism	Name	Disulfides array	Length
Data set:	1	2ln4	*Copris tripartitus*	Coprisin	1-4,2-5,3-6	44
	2	2b68	*Crassostrea gigas*	CgDefensin1	1-5,2-6,3-7,4-8	43
	3	1ozz	*Archaeopreponademophon*	ARD1	1-4,2-5,3-6	44
	4	2gl1	*Vigna radiata*	VrDefensin2	1-8,2-5,3-6,4-7	47
	5	1bhp	*Triticum aestivum*	β-purothionein	1-8,2-7,3-6,4-5	45
	6	1nbj	*Viola odorata*	Cycloviolacin	1-4,2-5,3-6	30
	7	1e4t	*Mus musculus*	MmβDefensin1	1-5,2-4,3-6	37
	8	1p9z	*Eucommia ulmoides*	Antifungal Peptide 2	1-5,2-9,3-6,4-7,8-10	41
	9	2rng	*Tachypleus tridentatus*	Big defensin	1-5,2-4,3-6	79
Test set:	1	2ln8	*Hirudo medicinalis*	Theromacin	1-7,2-6,3-8,4-9,5-10	75
	2	2eem	*Mytilus edulis*	MytilinB	1-5,2-6,3-7,4-8	34
	3	2k35	*Hydra magnipapillata*	Hydramacin1	1-6,2-5,3-7,4-8	60
	4	2ksk	*Saccharum officinarum*	SoDefensin1	1-8,2-5,3-6,4-7	71
	4	1jkz	*Pisum Sativum*	Psd1	1-8,2-5,3-6,4-7	46
	5	1jmn	*Viscum album*	ViscotoxinA2	1-6,2-5,3-4	46
	6	1r1f	*Palicourea condensata*	Cyclotide Palicourein	1-4,2-5,3-6	37
	7	1z99	*Crotalus durissus*	Crotamine	1-5,2-4,3-6	42
	8	2l1q	*Homo sapiens*	Liver antimicrobial peptide 2	1-3,2-4	40

Table 1: list of the proteins used in the training set and in the test set with the corresponding cluster, the PDB entry, the organism of origin, name, disulfide bond connectivity and lenght.

Starting draft structures

The starting models were generated using the I-TASSER suite imposing a maximal homology of 20%, to mimic the case of so-called "midnight-zone" structure prediction[16]. As previously reported, the starting model with the highest C-score, i.e., the I-TASSER model quality estimation score, is not always the most similar one to the structure deposited in the PDB[30] (Figure 2). Thus, to increase the possibility to obtain a good prediction from the simulation, the MD protocol was applied to all the generated models.

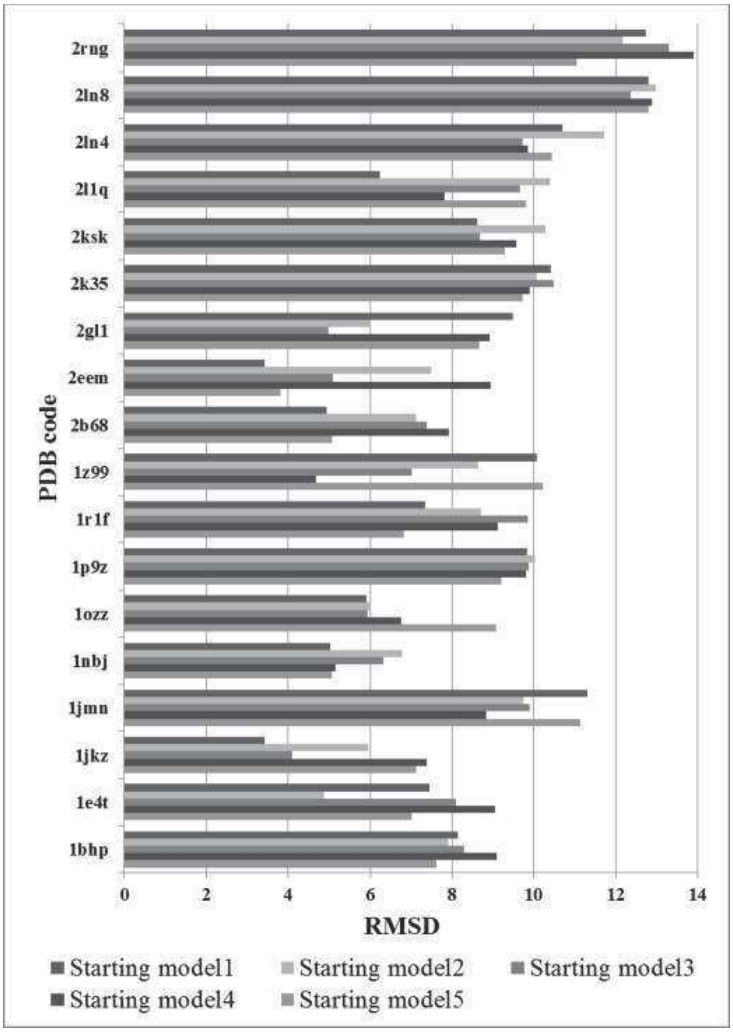


Figure 2: RMSD between the starting models and the experimental structure in PDB calculated using the g_rms tool in the GROMACS package.

Disulfides

The disulfide prediction algorithm was based on sulphur atom distances in order to set up a protocol suitable for disulfide prediction starting from draft structures only. As an improvement for the prediction, the assigned scores to pairs of cysteines were based

not on the simple distance but on a logarithmic function. In this way, the sensitivity for neighbour cysteines conserved among the draft structures was increased. Appling the disulfide prediction algorithm, the Q_P of the nine structures in the training and the test set was 0.78 and 0.66, respectively. The Q_C was 0.81 and 0.78 for the training and the test set, respectively. To our knowledge, homology models were used in SVM protocols for disulfide predictions[13,31], but this is the first attempt to validate a protocol for disulfide pattern prediction involving only draft structures imposing a low identity cut-off with known structures. The Q_C of available sequence based disulfide bond predictors is between 0.60 for DISLOCATE[32] and 0.73 for Cyscon[11] when imposing an identity lower than 20%.

The present protocol looks promising, but the sequence selection for the data set is limited to CSαβ peptides and therefore the Q_C and Q_P values cannot be directly compared to those obtained with the SVM algorithms, which are based on a much larger number of sequences. Furthermore, the SVM algorithms are a combination of different predictive terms. Nevertheless, the present approach could be integrated in disulfide predictors to further improve their performance or used to confirm a hypothetical cysteine connectivity.

Structure refinement

Simulation

A multi-step protocol was designed to refine the raw models generated by I-TASSER and to enhance conformational sampling. The protocol is based on the force field CHARMM27, which was proven to be particularly suitable for rapid folding and stabilization of secondary structure of the peptides at middle-high temperatures, as desirable for the simulated annealing step during the structure refinement[33,34].

First, the system is placed in a fully explicit solvation box and subjected to an energy minimization stage to optimize the model geometries after the introduction of the disulfide bonds and to reduce possible steric clashes originated in the placement of the water molecules. After a brief solvent equilibration step, a classical SA protocol[35] was applied, heating the peptide to 365 K to overcome potential energy barriers followed by slow cooling to reach a low energy minimum. The maximum temperature was set to explore different conformations without undergoing global unfolding[36]. Finally, the system was maintained at 300 K for 100 ns by a classical MD. The MD simulation was conducted for 100 ns in light of the observation that the vast majority of conformations in the training set required few tens of nanoseconds to reach stability. In both the simulation steps (SA and MD) RMSD values between the initial and the final conformations are lower than 1 nm (Supplementary Figure 2), except in one case, as expected for conformational changes more than folding/unfolding phenomena[37]. Nevertheless, the structures subjected to the simulation showed substantial variations in term of RMSD compared to the

corresponding starting structure, consequent to the rearrangement of the secondary elements.

Cut-off vs RMSD

In the training set, the correspondence between the centroid from the most populated cluster and the experimental structure was analysed, and different cut-offs were imposed to find the best cut-off to apply. From this analysis, the cut-off corresponding to the most populated cluster with the lowest density was selected. This is in contrast with other analyses that claim that higher density corresponds to better structures; this is the case when comparing simulations of different peptides (Figure 3), but not when comparing different clusters from the same simulation. Comparison of the RMSD between the final structure and the experimental model with density values is consistent with similar analyses previously performed [28,30].

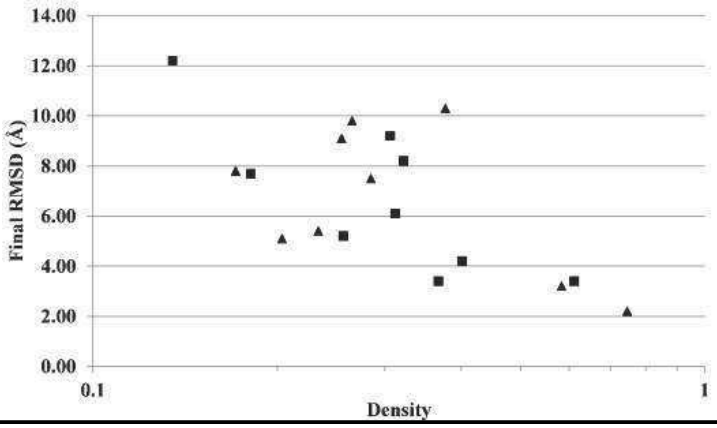


Figure 3: Density of the final cluster compared to the RMSD between the final model and the corresponding structure deposited in the PDB. Squares: proteins in the training set; Triangles: proteins in the test set.

Interestingly, for all the analysed decoys, the selected cluster is representative of structures from at least two different simulations, suggesting a convergence between the models. After the final minimization, the simulation protocol returned high quality final conformations as confirmed by the Ramachandran plot where the number of outliers (19/943) is comparable to that of the PDB structures (21/943) (Supplementary Figure 3).

Comparing the experimental peptide structures with the starting and the final ones resulting from MD simulations, the mean RMSD improvement (Δ RMSD) for the training and test set was 1.58 Å and 1.64 Å, respectively, corresponding to -19.5% and -21.1% in RMSD value. New models were also computed with I-TASSER imposing the correct sulphur atoms distances between bonded cysteines as an alternative disulfide-assisted modelling method. RMSDs with experimental structures are reported in Table 2 as RMSD-T.

In the case of good (RMSD < 5 Å) starting models, the main part of the final RMSD arises from conformational changes in terminal parts and loops of the protein. Thus, cysteine based refinement of the core of the protein is partially or completely lost in the RMSD calculation by the rearrangements of flexible parts as demonstrated by 2b68 and 2eem in the training and test set, respectively. In fact, the flexible loop of 2b68 (15 AA,

37% of the sequence) contributes for 70% of the RMSD. For 2eem, loop contribution to RMSD is 56% (11 AA, 25.5% of the sequence). Notably, for some of the analysed peptides, structures are present in the decoy with much better RMSD than that of the final structure (Table 2). This evidence confirms the quality of the MD protocol but suggests the need to further develop the clustering algorithms.

	PDB	Starting RMSD	B-RMSD	Final RMSD	Δ RMSD	RMSD-T
Training set	2ln4	10.8	7.7	9.2	-1.6	9.8
	2b68	4.9	4.7	5.2	0.3	6.6
	1ozz	6.0	5.1	4.2	-1.8	10.0
	2gl1	9.6	4.1	3.4	-6.2	5.6
	1bhp	8.1	7.7	7.7	-0.4	7.9
	1nbj	5.2	3.8	3.4	-1.8	9.4
	1e4t	7.7	4.6	6.1	-1.6	8.8
	1p9z	9.7	8.0	8.2	-1.5	1.7
	2rng	11.8	10.9	12.2	0.4	14.3
Test set	2ln8	12.8	12.4	10.3	-2.5	12.0
	2eem	3.5	3.1	2.2	-1.3	6.1
	2k35	10.4	9.2	9.8	-0.6	9.6
	2ksk	8.5	7.0	7.5	-1.0	8.7
	1jkz	4.1	3.3	3.2	-0.9	5.1
	1jmn	11.6	7.9	7.8	-3.8	9.4
	1r1f	7.8	5.7	5.4	-2.4	9.3
	1z99	10.1	4.9	9.1	-1.0	10.4
	2l1q	6.4	5.6	5.1	-1.3	7.7

Table 2: set of RMSD, in Å, compared to the experimental structure: starting model 1 (Starting RMSD); RMSD of the best structure in the entire decoy generated from the five simulations (B-RMSD); RMSD of the final structure selected from the clustering (Final RMSD) after minimization; difference between initial and final RMSD (Δ RMSD); RMSD of model 1 obtained with I-TASSER imposing experimental disulfides.

Case studies

Cycloviolacin

This cyclotide from *Viola odorata* was identified and characterized for the first time in 1999 (PDB entry 1df6)[38] and its structure was solved at higher resolution in 2003 (PDB entry 1nbj, used as reference structure)[39]. The two structures appear similar except for the stretch CTVTALLGC, which is a 3_{10}-helix in the first case and α-helix in the second one, suggesting a possible flexibility in that region. The draft structures were calculated without imposing cyclization, assuming that only the primary sequence is known. The model with the highest C-score from I-TASSER (model 1) exhibits two β-strands, as in the experimentally determined structure 1nbj, while the helix is completely missing. Its predicted disulfide array (1-4, 2-6, 3-5, referring to 1nbj) is different from the experimental one (1-4, 2-5, 3-6) although its RMSD is the best among the calculated models. On the other hand, our algorithm correctly predicted the natural bonds. This is a clear example where, having more than one draft structure, our algorithm can improve the prediction of correct cysteine connectivity.

The best structure obtained from the molecular dynamics protocol derived from starting model 2 (RMSD = 3.8 Å), but the selected centroid came from model 5 MD with a computed RMSD of 4.0 Å, thus demonstrating the efficacy of the protocol even without

imposing cyclization of the peptide. Even though our final structure does not contain amino acids folded in a helix, analysing the entire model 5 dynamics, the transient folding of the protein in a 3_{10}-helix between amino acids 11-13, as expected from the corresponding region of the deposited structure 1df6, is evident (Figure 4)[40].

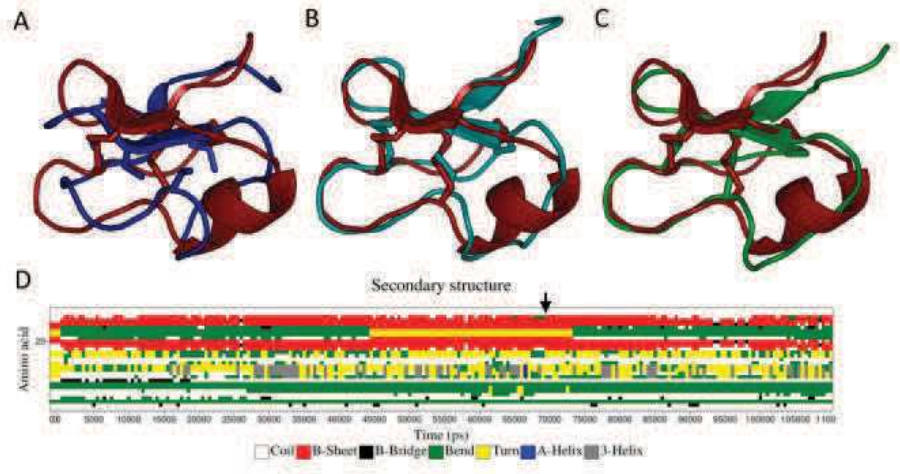


Figure 4: In red, the structure of 1nbj from the PDB (A, B and C). A: in blue, starting model 1; B: in cyan, the best structure in the decoy; C: in green, the final structure from the MD simulation. D: Secondary structure analysis of the final model simulation using DSSP (Kabsch & Sander, 1983). The black arrow corresponds to the final structure snapshot.

Royalisin

Figure 5: Superimposition of royalisin starting model 1 (black) and the final structure (gray).

As proof of concept, the proposed methods were applied to royalisin, an antimicrobial peptide identified in *Apis mellifera* royal jelly in 1990[41] and effectively produced in *E. coli*[42,43]. Despite its early discovery, no experimental structure is present in the PDB. Royalisin displays 43% identity with Coprisin (PDB identifier: 2ln4), and can be considered a typical case of "twilight zone" homology modelling[8].

Our analysis confirmed in the royalisin's a typical defensin fold. Based on the density values reported in this study, the final density of 0.609 suggests an RMSD of about 4 Å from an hypothetical experimental structure, and an RMSD between 1 Å and 6 Å when taking in consideration a more extensive density analysis[30]. Data mining the PDB using DALI[44] with the final model as query, the first hit was lucifensin from *Lucilia sericata*. Compared to this and other defensins, royalisin presents a longer flexible loop between cysteines 3 and 17 and a C-terminal tail (Figure 5). The relative high mobility of this tail, its absence in similar peptides[45] and its marginal involvement in antibacterial activity[46] suggest a different role for this portion of the peptide.

Concluding remarks

Our molecular dynamics protocol was shown to improve the structural prediction of CSαβ peptides in the so called midnight homology region (< 20%), with better performance compared to directly imposing disulfides to the I-TASSER tool (Table 2). Our algorithm for disulfide connectivity prediction, starting from I-TASSER draft structures, can be a useful tool when data about cysteine array are missing. The possibility to use the algorithm along with other tools will be explored. Even though the molecular dynamics protocol requires a considerable amount of computational power, the analysis of the simulation is also useful to reveal intrinsic flexible parts of the structure (Figure 4). Density calculation effectively correlates with final RMSD, as extensively demonstrated for other consolidated protocols, and it is a good parameter to evaluate the quality of the final model (Figure 3). The proposed protocol could be extended to predict disulfide connectivity and structural models of cysteine stabilized proteins.

References

1. Fjell, C. D., Hiss, J. A., Hancock, R. E. W. & Schneider, G. Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.* **11,** 37–51 (2012).

2. Mojsoska, B. & Jenssen, H. Peptides and Peptidomimetics for Antimicrobial Drug Design. *Pharmaceuticals* **8,** 366–415 (2015).

3. Dias, R. de O. & Franco, O. L. Cysteine-stabilized αβ defensins: From a common fold to antibacterial activity. *Peptides* **72,** 64–72 (2015).

4. Domeneghetti, S. *et al.* Structural and Antimicrobial Features of Peptides Related to Myticin C, a Special Defense Molecule from the Mediterranean Mussel Mytilus galloprovincialis. *J. Agric. Food Chem.* **63,** 9251–9259 (2015).

5. Romestand, B., Molina, F., Richard, V., Roch, P. & Granier, C. Key role of the loop connecting the two beta strands of mussel defensin in its antimicrobial activity. *Eur. J. Biochem. FEBS* **270,** 2805–2813 (2003).

6. Huang, X., Gao, C., Zhao, Q. & Li, C. Antimicrobial Characterization of Site-Directed Mutagenesis of Porcine Beta Defensin 2. *PLoS ONE* **10,** (2015).

7. Schwede, T. Protein Modelling: What Happened to the 'Protein Structure Gap'? *Struct. Lond. Engl. 1993* **21,** (2013).

8. Khor, B. Y., Tye, G. J., Lim, T. S. & Choong, Y. S. General overview on structure prediction of twilight-zone proteins. *Theor. Biol. Med. Model.* **12,** (2015).

9. Kinch, L. N., Li, W., Monastyrskyy, B., Kryshtafovych, A. & Grishin, N. V. Evaluation of free modeling targets in CASP11 and ROLL. *Proteins Struct. Funct. Bioinforma.* n/a-n/a (2016). doi:10.1002/prot.24973

10. Chuang, C.-C., Chen, C.-Y., Yang, J.-M., Lyu, P.-C. & Hwang, J.-K. Relationship between protein structures and disulfide-bonding patterns. *Proteins Struct. Funct. Bioinforma.* **53,** 1–5 (2003).

11. Yang, J., He, B.-J., Jang, R., Zhang, Y. & Shen, H.-B. Accurate disulfide-bonding network predictions improve ab initio structure prediction of cysteine-rich proteins. *Bioinforma. Oxf. Engl.* **31,** 3773–3781 (2015).

12. Lin, H.-H. & Tseng, L.-Y. DBCP: a web server for disulfide bonding connectivity pattern prediction without the prior knowledge of the bonding state of cysteines. *Nucleic Acids Res.* **38,** W503–W507 (2010).

13. Yu, D.-J. *et al.* Disulfide Connectivity Prediction Based on Modelled Protein 3D Structural Information and Random Forest Regression. *IEEEACM Trans. Comput. Biol. Bioinforma. IEEE ACM* **12,** 611–621 (2015).

14. Fischer, A. W. *et al.* CASP11 – An Evaluation of a Modular BCL::Fold-Based Protein Structure Prediction Pipeline. *PLoS ONE* **11,** (2016).

15. Porto, W. F., Fensterseifer, G. M. & Franco, O. L. In silico identification, structural characterization, and phylogenetic analysis of MdesDEF-2: a novel defensin from the Hessian fly, Mayetiola destructor. *J. Mol. Model.* **20,** 1–8 (2014).

16. Meier, A. & Söding, J. Context similarity scoring improves protein sequence alignments in the midnight zone. *Bioinforma. Oxf. Engl.* **31,** 674–681 (2015).

17. Wang, G., Li, X. & Wang, Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* **37,** D933–D937 (2009).

18. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

19. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57,** 702–710 (2004).

20. Barthel, D., Hirst, J. D., Błazewicz, J., Burke, E. K. & Krasnogor, N. ProCKSI: a decision support system for Protein (structure) Comparison, Knowledge, Similarity and Information. *BMC Bioinformatics* **8,** 416 (2007).

21. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738 (2010).

22. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinforma. Oxf. Engl.* **29,** 845–854 (2013).

23. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

24. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 14101 (2007).

25. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).

26. Freddolino, P. L. & Schulten, K. Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophys. J.* **97,** 2338–2347 (2009).

27. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **38,** 236–240 (1999).

28. Zhang, Y., Kolinski, A. & Skolnick, J. TOUCHSTONE II: A New Approach to Ab Initio Protein Structure Prediction. *Biophys. J.* **85,** 1145–1164 (2003).

29. Lee, H.-T. *et al.* A Large-Scale Structural Classification of Antimicrobial Peptides, A Large-Scale Structural Classification of Antimicrobial Peptides. *BioMed Res. Int. BioMed Res. Int.* **2015, 2015,** e475062 (2015).

30. Zhang, Y. & Skolnick, J. SPICKER: a clustering approach to identify near-native protein folds. *J. Comput. Chem.* **25,** 865–871 (2004).

31. Lin, H. H., Hsu, J. C. & Chen, Y. F. Disulfide Bonding Pattern Prediction Server Based on Normalized Pair Distance by MODELLER. in *2012 International Symposium on Computer, Consumer and Control (IS3C)* 581–584 (2012). doi:10.1109/IS3C.2012.152

32. Savojardo, C. *et al.* Improving the prediction of disulfide bonds in Eukaryotes with machine learning methods and protein subcellular localization. *Bioinformatics* **27,** 2224–2230 (2011).

33. Cino, E. A., Choy, W.-Y. & Karttunen, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **8,** 2725–2740 (2012).

34. Lindorff-Larsen, K. *et al.* Systematic Validation of Protein Force Fields against Experimental Data. *PLoS ONE* **7,** (2012).

35. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220,** 671–680 (1983).

36. Day, R., Bennion, B. J., Ham, S. & Daggett, V. Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* **322,** 189–203 (2002).

37. Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol. Biol. Clifton NJ* **857,** 231–257 (2012).

38. Craik, D. J., Daly, N. L., Bond, T. & Waine, C. Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* **294,** 1327–1336 (1999).

39. Rosengren, K. J., Daly, N. L., Plan, M. R., Waine, C. & Craik, D. J. Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J. Biol. Chem.* **278,** 8606–8616 (2003).

40. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).

41. Fujiwara, S. *et al.* A potent antibacterial protein in royal jelly. Purification and determination of the primary structure of royalisin. *J. Biol. Chem.* **265,** 11333–11337 (1990).

42. Shen, L. *et al.* Expression of Acc-Royalisin gene from royal jelly of Chinese honeybee in Escherichia coli and its antibacterial activity. *J. Agric. Food Chem.* **58,** 2266–2273 (2010).

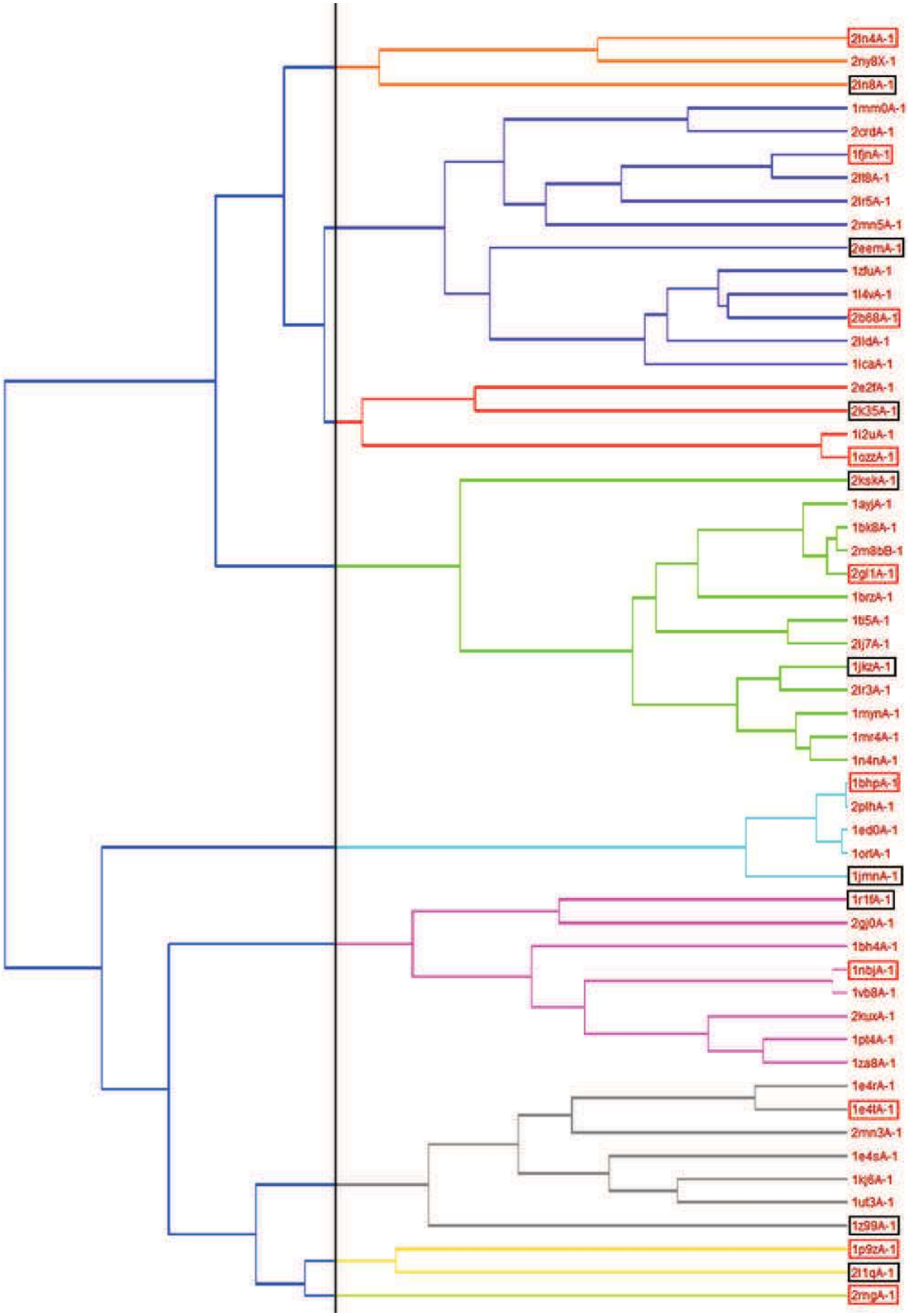
43. Tseng, J.-M. *et al.* Facilitative production of an antimicrobial peptide royalisin and its antibody via an artificial oil-body system. *Biotechnol. Prog.* **27,** 153–161 (2011).

44. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549 (2010).

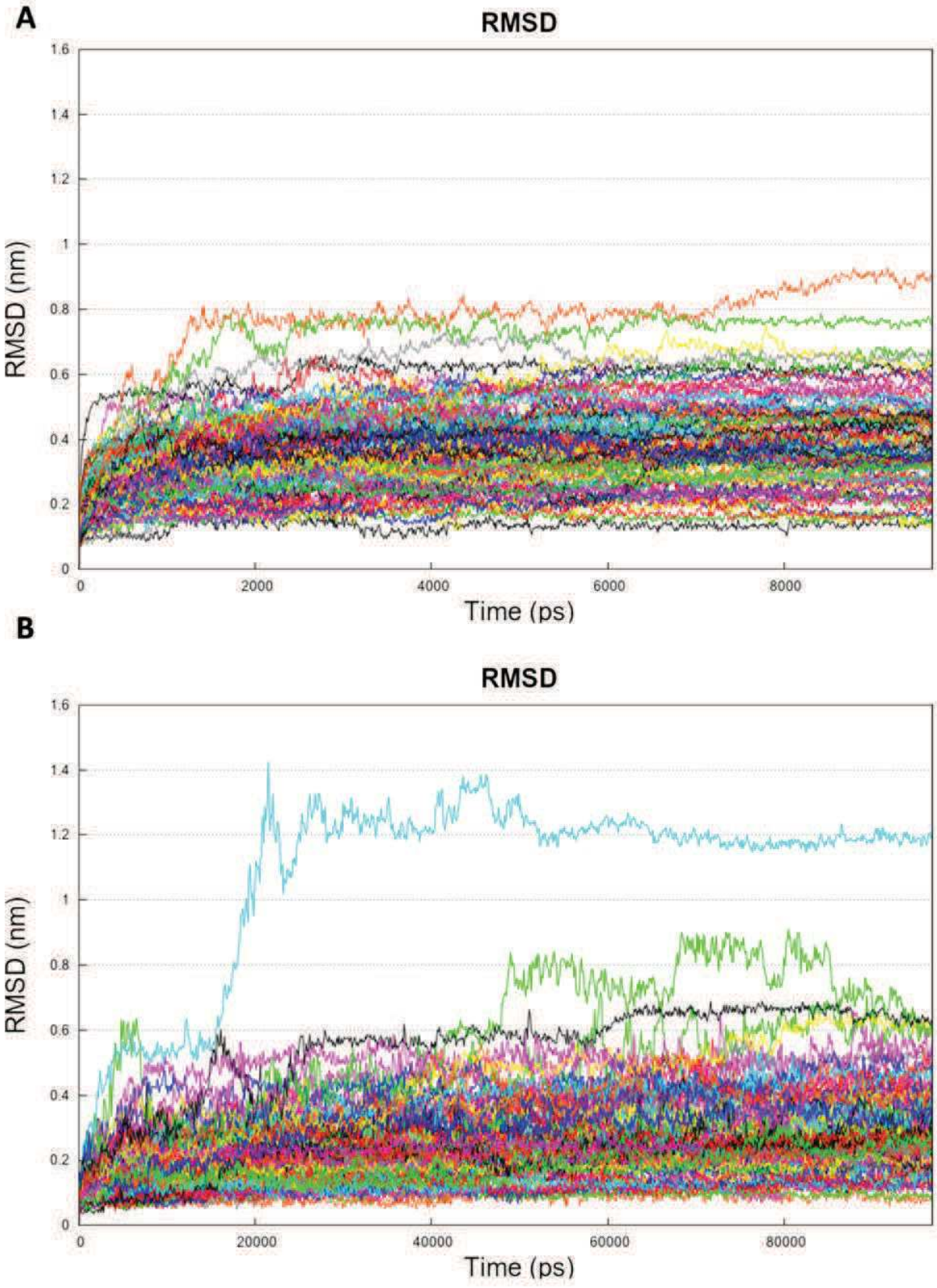
45. Klaudiny, J., Albert, Š., Bachanová, K., Kopernický, J. & Šimúth, J. Two structurally different defensin genes, one of them encoding a novel defensin isoform, are expressed in honeybee Apis mellifera. *Insect Biochem. Mol. Biol.* **35,** 11–22 (2005).

46. Bílikova, K., Huang, S.-C., Lin, I.-P., Šimuth, J. & Peng, C.-C. Structure and antimicrobial activity relationship of royalisin, an antimicrobial peptide from royal jelly of Apis mellifera. *Peptides* **68,** 190–196 (2015).

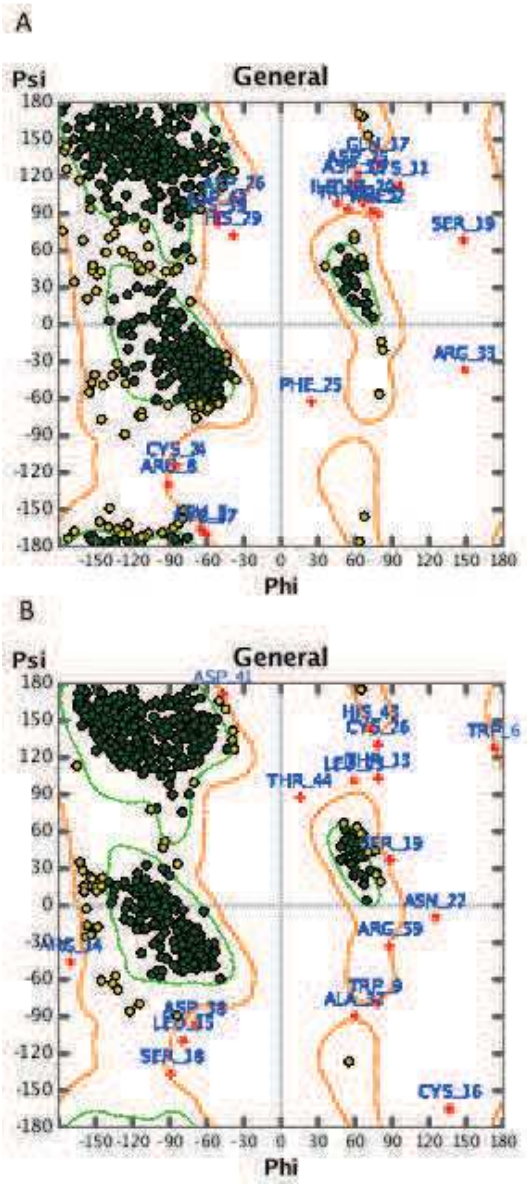
Supplementary material



Supplementary Figure 1: clustering of cysteine-packed αβ peptides from APD2 based on the pairwise TM-score. Cut-off was set to 0.3. Centroids are circled in red, test set elements in black.



Supplementary Figure 2: RMSD of simulated annealing (panel A) and molecular dynamics (panel B) for all the runs.



Supplementary figure 3: Ramachandran plot of all the analysed proteins structures in PDB (panel A) and final structures from simulations (panel B).

Chapter 3

Structural and Antimicrobial Features of Peptides Related to Myticin C

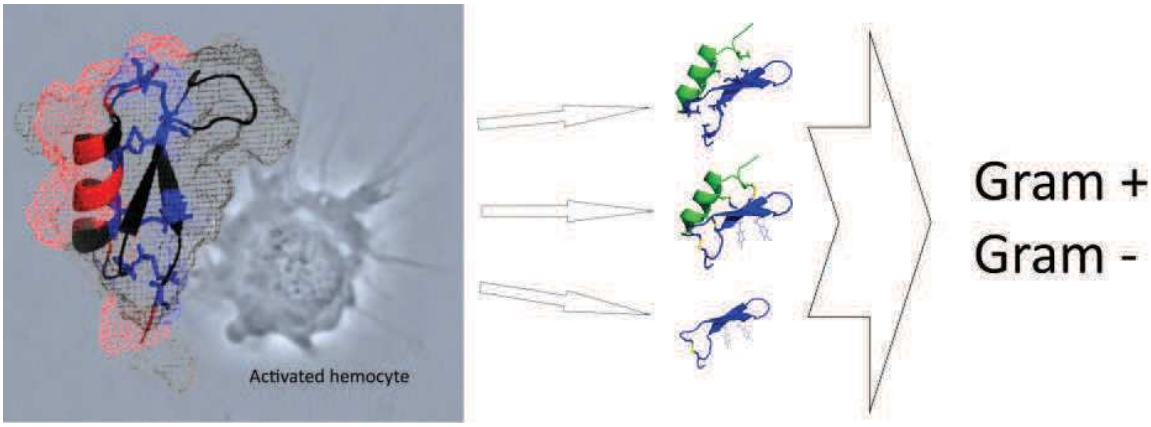
Franzoi M., Domeneghetti S., Damiano N., Sandre M., Cavion F., Grigoletto R., Tavano R., Marin O., Venier P.

Part of this chapter is published in:
J Agric Food Chem. 2015 Oct 28;63(42):9251-9. doi: 10.1021/acs.jafc.5b03491. Epub 2015 Oct 16.

Contribution:
Structural prediction
Microbial testing
Mechanism of action

Abstract

Mussels in general, and Mytilus spp. in particular, possess a large number of cysteine-stabilized αβ peptides. Among them, myticin C (MytC) is expressed in some hundreds of transcript variants following *in vivo* stimulation of mussels with live or dead bacteria. Starting from two sequences found expressed in Italian mussels and named MytC-a and MytC-b, we tested the antimicrobial activity of the whole mature peptides of 40 amino acidic residues and related peptide fragments designed according to structural information. In general, MytC and related peptides are active at pH 5 against Gram-positive and Gram-negative bacteria. We also investigated the mechanisms of action of such peptides, revealing different behaviours against different bacterial strains. Overall, the experimental assays suggested a membrane permeabilization activity at doses lower than MIC. This evidence could be related to a double role of MytC: at low concentrations it could act as adjuvant for other immunity effectors, eventually directed inside pathogen cells, while at high doses it can exert a manifest antimicrobial effect.



Introduction

Bivalves are known to express several families of antimicrobial peptides (AMP) and mainly cysteine-stabilized AMPs such as mytilins, defensins and myticins have been identified in *Mytilus* spp.[1–4]. Even if the AMP abundance in circulating haemocytes of adult mussels has been extensively proved, confirming such cells as main effectors of the innate immunity, also mucosa-associated cells likely play a defensive role and produce AMPs[5–7]. Myticins were originally discovered with a proteomic approach, by fractionating haemolymph protein extracts obtained from immunostimulated mussels[5]. Initially, only two isoforms were identified and named myticin A (MytA) and myticin B (MytB). cDNA cloning of such peptides revealed eight cysteines, found to be all oxidized from mass spectroscopy data[5]. Both MytA and MytB are translated as prepropeptides, with a conserved 20 AA signal peptide, an anionic C-terminal extension of 36 AA already found in similar AMPs and probably involved in cytoprotection and folding[8,9], and a central, 40 AA long mature peptide including 8 conserved cysteines, similarly to the cysteine-stabilized αβ (CSαβ) peptides mytilins and defensins. The molecular hallmarks of myticin prepro peptides are described in the NCBI Conserved Domain Database (pfam10690;http://www.ncbi.nlm.nih.gov/ Structure/cdd/cdd.shtml).

Considering their pH-related activity[10], the mature myticin peptides have been proposed to be active in pathogen-containing phagolysosomes as well as in the haemolymph and, after massive cell degranulation, in the intercellular space[3]. The Myticin C (MytC) transcript cluster was discovered in the transcriptome of haemocytes from immunostimulated mussels, as independent and highly variable family compared to the already known myticins[11]. The specific expression of MytC in granular haemocytes has been confirmed with immunohistochemistry (IHC) and *in situ* hybridization (ISH)[12]. MytC differs from MytA and MytB for a longer C-terminal 40 AA propeptide. More than 300 transcript variants of MytC are deposited in GenBank, expected to result in mature peptides with a molecular weight between 4.35 and 4.6 kDa and isoelectric point between 8.05 and 8.86. In general, the molecular variability characterizing MytC and other immune effectors has been correlated with the need of host organisms to counteract the adaptive changes of rapidly replicating pathogens[13–15]. Increasing the repertoire of AMPs able to act synergically helps the host to overcome specific resistance mechanism against single effectors of the innate immune system[16–18]. Other myticin variants has been identified in *M. coruscus*[19].

Functional assays proved the antibacterial and antiviral properties of myticin C and, at the same time, its immunomodulatory and chemotactic effects[10,20,21]. In fact, Chinook salmon embryo cells transfected with GFP-MytC plasmids significantly increased their resistance to VHSV (a virus causing haemorrhagic septicaemia in fishes), modulated the expression levels of immuno-related genes, and their lysate was able to chemo-attract haemocytes, compared to non-transfected cells[20]. In addition, a chemically-synthesized

MytC variant in reduced form was able to decrease the VHSV infectivity and to induce mussel haemocyte chemotaxis. The same (reduced) MytC was able to lessen the *Escherichia coli* growth, at acidic pH and at 250 μM concentration, to aggregate anionic phospholipids and to perturb negatively charged membranes[10]. In oyster haemocytes, such reduced MytC showed antiviral activity against OsHV-1 whereas Vero cells treated with Tat-MytC resulted to be more resistant to the human viruses HSV-1/2[21].

Material and methods

Sequence selection and structural model generation

Precursor transcript variants of MytC were downloaded as nucleotide sequences from GenBank and Sequence Read Archive (http://www.ncbi.nlm.nih.gov). The AA conservation profile of the mature peptide only was achieved with the freely available tool WebLogo v3.4 using the sample SRR286640 (hemocyte sequence reads from mussels farmed in the Venice lagoon area)[22] for MytC-a. MytC-b sequence corresponds to an highly charged sequence found in Haemocytes sample Mg_Hae03_05A11 (NCBI: CAM56810)[11]. The 3D model of the mature peptide was computed with I-tasser[23] and then refined with GROMACS v. 4.6.5 (www.gromacs.org)[24].

Based on the Cα distances, disulfide bonds fully compatible with the cysteine positions (5–24, 10–33, 14–35, 19–38) were imposed. For all simulations, the Charmm force field was used in a 5.188 × 5.188 × 3.630 nm cell filled with SPC/E water model, in isothermal–isobaric ensemble (NPT).

A 1.2 nm distance cutoff was imposed both for the Lennard-Jones van der Waals (vdW) potential and for the electrostatic interactions, using the particle mesh Ewald (PME) method, with Fourier spacing 0.12 and PME order 4. Berendsen and Parrinello-Rahman's temperature-pressure coupling was chosen for solvent equilibration and simulated annealing, respectively, with isotropic pressure coupling. After a minimization of 2000 steps, the linear constraint solver (LINCS) method for bond constraints was chosen. For solvent equilibration, a 2 ps molecular dynamics simulation was performed at 300 K, with constrained protein atom positions. Finally, simulated annealing of 10 ns was performed with annealing times of 0, 2, and 8 ns, at temperatures of 300, 365, and 280 K for protein and 300, 320, and 280 K for water, respectively. For the 100 ns simulation, the same experimental protocol was used, except that the simulated annealing was replaced with free dynamics at 300 K. Final clustering was performed using the GROMOS algorithm[25].

The superpose function of PyMOL (www.pymol.org) was used to overlay ribbon peptide structures and the homologues search in the PDB was performed via DALI server[26].

The resulting β-stranded portion of the peptides as well as the complete peptides were synthesized and oxidatively folded in the CRIBI peptide laboratory of the University of Padua (Prof. O. Marin). Complete sequences are reported in Table 1.

Peptide	Sequence	Lenght (aa)	Molecular weight (Da)	Isoelectric point	Net charge at pH 7.4	Net charge at pH 5
MytC_a-red	QSVACTSYYCSKFCGSAGCSLYGCYLLHPGKICYCLHCSR	40	4372.88	7.81	+1	+5
MytC_a-ox	QSVACTSYYCSKFCGSAGCSLYGCYLLHPGKICYCLHCSR	40	4364.88	9.42	+2	+5
MytC_a[19-40]-ox	------------------CSLYGSYLLHPGKISYSLHCSR	22	2481.30	9.37	+2	+4
MytC_b-red	QSVACTSYYCSKFCGSAGCSLYGCYKLHPGKICYCLHCRR	40	4457.95	8.24	+3	+7
MytC_b-ox	QSVACTSYYCSKFCGSAGCSLYGCYKLHPGKICYCLHCRR	40	4449.00	9.82	+4	+7
MytC_b[19-40]-ox	------------------CSLYGSYKLHPGKISYSLHCRR	22	2565.25	9.99	+4	+6

Table 1: peptides tested in this work, with sequences and physicochemical properties.

Antimicrobial testing

Minimal inhibitory concentration (MIC) was determined on selected bacterial strains (*Escherichia coli* ATCC 25922, *Pseudomonas aeruginosa* ATCC 27853, *Staphylococcus aureus* ATCC 29213, *Enterococcus faecalis* ATCC 29212, *Bacillus subtilis* ATCC 6051) using the broth microdilution susceptibility test[27]. Serial 2-fold dilutions of each peptide (0.5-32 μM, final concentration) were prepared in 50 μL volume, using 96-well microtiter plates (Sarstedt, Germany) and Muller-Hinton broth (MHB; Difco Laboratories, Detroit, MI, USA) or 50% MHB, with pH adjusted to 5 or 7. Fifty microliters of bacterial culture (~5 × 10^5 CFU/mL in MHB or 50% MHB) were then added to each well. The MIC value was defined as the lowest peptide concentration that prevented visible bacterial growth after incubation for 24 h at 37 °C. The minimal bactericidal concentration (MBC) corresponds to the lowest peptide concentration showing <1 CFU/ml after plating 50 μL from the wells onto nutrient agar plates. Each experimental condition was tested at least in two biological replicates, each one in three technical replicatesfor.

The growth inhibition kinetics of the peptides of interest was comparatively evaluated on *E. coli* and *E. faecalis* at pH 5, using a 10^7 CFU/mL bacterial inoculum. The optical density was monitored every 30 min at 620 nm for 24 h with the microplate reader Infinite 200 Pro (Tecan, Grödig, Austria).

Haemolytic essay

Fresh erythrocytes were resuspended in PBS and treated with peptide concentrations ranging from 8 to 128 μM, for 2 and 24 hours at 37 °C, and centrifuged at 2700 g for 5 minutes. Supernatant absorbance was read in microplates at 450 nm to check for the presence of haemoglobin[28]. Test were repeated in four technical replicates for each of the two biological replicates. This assay was done by collaboration with Dr. Regina Tavano (Dept. Biomedical sciences, UniPD).

Effects on bacterial membranes

The effects of mitycin C peptides on the inner and outer membrane of *E. coli* have been evaluated by using the ML-35 pYC strain, kindly obtained by the AMP-research group of UniTs, as already reported except for some modifications[29,30]. Briefly, serial peptide dilutions in SPB (7.5 mM dibasic sodium phosphate, 2.5 mM monobasic sodium phosphate, pH 7.5) were incubated with 2*10^7 *E. coli* cells resuspended in SPB and 0.15 mM CENTA or 1.5 mM ONPG in 96-well microplates. Absorbance values at 405 nm were recorded every 10 minutes for 2 hours.

The membrane permeabilization was confirmed using the SITOX Green probe, according to the selective permeation of the dye into membrane-damaged bacteria. Bacteria were grown at 37 °C in MH broth (pH 7) to mid-logarithmic phase, washed, and re-

suspended at 10^7 CFU/mL in 10 mM sodium phosphate buffer, pH 7.4. The cells were subsequently incubated for 15 min in the dark with 2 μM SYTOX Green Nucleic Acid Stain (Life Technologies). Then, the peptide to be tested was added at the MIC value, and the fluorescence increase due to the binding of cationic dye to intracellular DNA was monitored every minute for 1 h (excitation and emission wavelengths of 485 and 520 nm, respectively). Melittin was used as positive control.

B. subtilis cells at OD 1, treated with 32 μM of peptides, were observed at optical microscope (DM2000, Leica) to verify the clustering of bacterial cells after 30 minutes from the treatment. A DMSO volume equivalent to that of the peptide solution has been used as negative control[31].

Binding to PNG and LPS

200 μg LPS F3665 (Sigma) or *B. subtilis* derived PNG (Sigma) were dissolved in AB (0.1 M acetic acid and 0.1 M sodium acetate, pH 5), mixed with 10 μg of peptide and then incubated under agitation for 30 minutes. Mixture was centrifuged at 15000 g for 15 minutes, then both pellet and lyophilized supernatant were resuspended in 15 μl ultrapure water. The PNG-containing pellet was further treated with Tris-HCl 50 mM, 5% SDS at 4 °C for 15 minutes and centrifuged. The resulting fractions were incubated at 95 °C with Laemli buffer and loaded in 16 % acrylamide gel. BSA and Lysozyme (Lys) were used as controls[32,33]. Band intensities were calculated with ImageJ and the binding percentage was expressed as control band intensity minus supernatant band intensity. When possible, the pellet band intensity has been used as reference.

The results of the above functional assays have been reported and discussed in one Magistral thesis exam (Federica Cavion, UniPD, Oct 2016).

Results and discussion

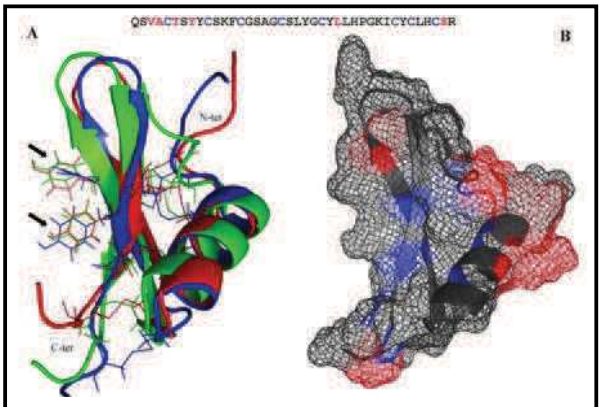


Figure 1. (A) Superimposed ribbon structures of three converging models of the mature MytC obtained by I-tasser(18) and refined with GROMACS (19). Arrows indicate two stacking tyrosines in the two antiparallel β-strands; N- and C-termini are also indicated. (B) Mesh representation of MytC highlighting cysteine (blue) and variable (red) residues. Most of the latter are clustered in the predicted helix region.

Structural prediction

The five structures resulting from the MytC-a sequence, submitted to I-tasser prediction, showed unequivocal pairing of eight conserved cysteines, also confirmed by applying the protocol described in chapter 2. Connectivities between cysteines 5 and 24, 10 and 33, 14 and 35, and 19 and 38 were set as molecular constraints for model refinement by simulated annealing with GROMACS (Figure 1A)[24,34]. Following a further simulation at 300 K for 100 ns, we selected three convergent MytC structures that clearly reproduced a compact CSαβ scaffold and a γ-core motif common to various AMPs and venom components[35]. These models showed the most variable AA residues clustered mainly on the surface in the α-helical region (Figure 1B). Both the invariant cysteine array and tolerated sequence changes are essential for the biological function while minimizing the risk of pathogen resistance[13] and are, hence, useful for nature-guided drug design[36]. The γ-core motif of MytC does not present the typical G×C element and it is likely stabilized by aromatic stacking interactions between two highly conserved tyrosines, as previously observed in tachyplesin I, a β-hairpin AMP of 17 AA isolated from horseshoe crab[37]. Once superimposed, the structural subdomains of MytC appeared closely similar to those of other CSαβ peptides, such as defensin 1 (MGD1)[38] and mytilin B[39] from M. galloprovincialis and butantoxin, a scorpion venom peptide stabilized by three disulfide bridges[40](Figure 2). MytC (PF10690) and PDB accession codes of the other peptides reported in the figure 2 are available in public protein databases[41,42].

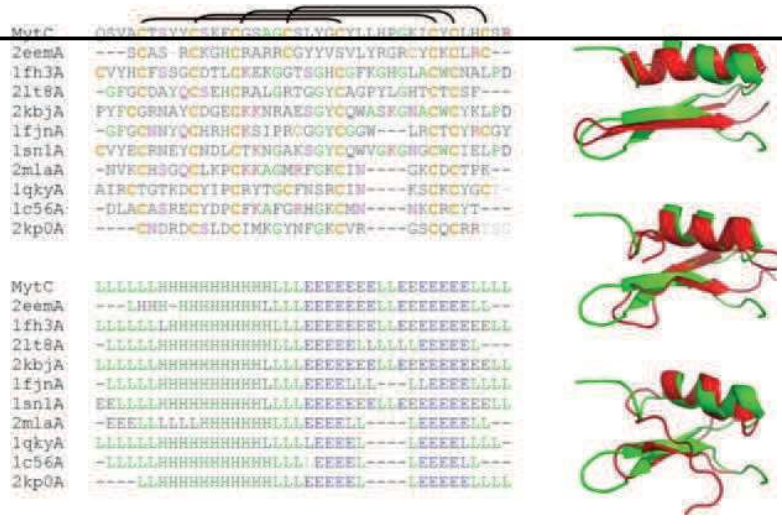


Figure 2. (Left) Multiple alignments illustrating the primary sequence (top) and secondary structure elements (bottom) of MytC and selected CSαβ peptides. (Right) Superimposed structures of MytC (green) with mytilin B (2eemA, top), MGD1 (1fjnA, center), and butantoxin (1c56A, bottom). Despite general similarities in the molecular organization and cysteine array, the superimposition reveals topological differences in terms of loop length and β-sheet twist (the γ-core region).

Antimicrobial and haemolytic activity

The tested peptides were inactive against Gram- bacteria when treated in standard conditions (MHB, pH 7). Thus, we performed antibacterial tests also at slightly acidic environment (pH 5), lower ionic strength (50% MHB) or both. Activity against Gram- bacteria was detected only at 50% MHB pH 5, with the best performance of MytC_a[19-40]-ox fragment against *E. coli* and MytC_b[19-40]-ox against *P. aeruginosa*. Such pHdependent activity was already reported for other AMPs[43] and could be relevant in subcellular locations, for example in lysosomes[3], or in a "turning on" occurring for example in tissues acidified by bacterial infections[44-46]. Curiously, both the above mentioned MytC peptides showed no activity against the other Gram- bacterium. This fact suggests a sequence-related specificity of the '19-40' fragments, owing to the higher positive charge of MytC-b version and to an higher fraction of negative charged lipids on the inner membrane of *P. aeruginosa*, compared to *E. coli* (Supp. Table 1). Hence, the greater positive charge of MytC-b could lead to a better membrane interaction with *P. aeruginosa* and, at the same time, could impair membrane interactions with *E. Coli*, or possibly trigger the activation of a stronger counter defence in the bacterium[47].

	Gram -								Gram +											
Bacteria	*E. coli* (ATCC 25922)				*P. aeruginosa* (ATCC 27853)				*E. faecalis* (ATCC 29212)				*S. aureus* (ATCC 29213)				*B. subtilis* (ATCC 6051)			
Broth	MHB		50 % MHB		MHB		50 % MHB		MHB		50 % MHB		MHB		50 % MHB		MHB		50 % MHB	
pH	7	5	7	5	7	5	7	5	7	5	7	5	7	5	7	5	7	5	7	5
MytC_a-red	>32	>32	>32	16 (16)	>32	>32	>32	>32	32	16	16	2	>32	>32	>32	>32	nd	nd	nd	4 (4)
MytC_a-ox	>32	>32	>32	32	>32	>32	>32	>32	16	8	8	4	>32	>32	>32	32	nd	nd	nd	4 (8)
MytC_a[19-40]-ox	>32	>32	>32	8 (32)	>32	>32	>32	>32	32	32	16	16	>32	>32	>32	32	>32	32 (32)	16 (16)	8 (8)
MytC_b-red	>32	>32	>32	16 (32)	>32	>32	>32	>32	16	32	8	4	>32	>32	>32	32	>32	nd	nd	4 (16)
MytC_b-ox	>32	>32	>32	16 (32)	>32	>32	>32	>32	32	32	8	8	>32	>32	>32	>32	>32	nd	nd	16
MytC_b[19-40]-ox	>32	>32	>32	>32	>32	>32	>32	16 (16)	>32	>32	32	>32	>32	>32	>32	>32	>32	nd	8 (8)	4 (32)

Table 2: MIC values of the peptides tested against different bacteria in the stated conditions (medium concentration and pH). In brackets MBC values (where absent, MBC >32). Nd: condition not determined.

About Gram+ bacteria, the MytC-related peptides impaired the growth of *E. faecalis*, even if none of the tested peptide concentrations could completely kill the bacterium. Even if not being a spore forming bacterium, *E. faecalis* is able to switch to a latent phase, capable to higher resistance to antibiotics and possibly explaining the discrepancy between MIC and MBC data[48]. At the same time, MytC-related peptides showed no activity against *S. aureus* but good MIC and MBC values against *B. subtilis* (both Gram+ strains), thus revealing a relatively good specificity and, at the same time, leaving open the question if the tested bacteria represent the main target *in vivo*. In fact, other defensin-like peptides showed significant differences in MIC values not only between bacteria of the same Gram group, but also between strains of the same specie[49].

All the synthesized peptides showed haemolytic activity <1% when tested at the maximum concentration of 128 μM, both after 2 and after 24 hours of incubation.

Interaction with bacterial membranes

E. coli membrane permeabilization

E. coli ML-35 pYC was used to detect the membrane permeabilization possibly induced by MytC-related peptides. Mellitin was used as positive control for its ability to induce pore formation on both inner and outer *E. coli* membranes (All the experimental data can be found in Supp. Figure 1 and 2).

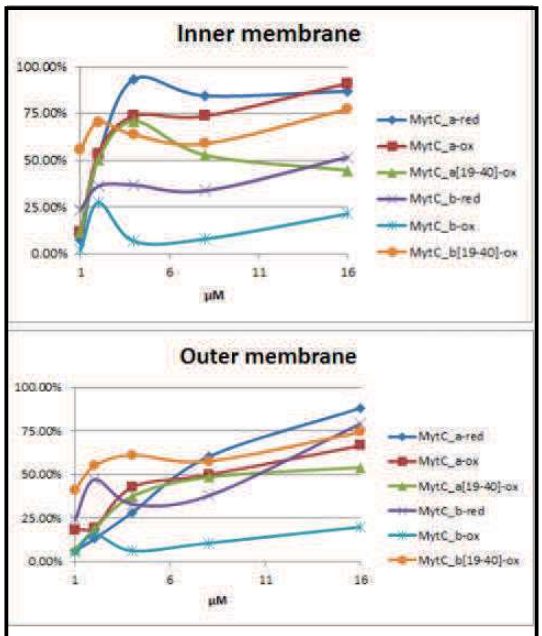


Figure 3: Permeabilizing effects on *E.coli* ML-35 after 1 hour of incubation with different concentrations of selected MytC peptides. Values are expressed as percentage of permeabilization compared to 2 μM Melittin.

Interestingly, some peptides such as MytC_b[19-40]-ox, MytC_a-red and MytC_a-ox showed permeabilizing effects on both the inner and outer *E. coli* membranes at concentrations lower than the corresponding MIC values (Figure 3). This particular feature differs from other peptides having a MIC lower than the concentration needed for the induction of membrane perturbation but, at the same time, peptides exist able to form transient pores on bacterial surfaces or vesicles[50-52]. In particular, the latter peptides create a burst of leakage in vesicles just after treatment, stopping in minutes and with dose-dependent effect[53]. Even if not showing antimicrobial activity at the tested concentration, MytC_b[19-40]-ox displayed a higher permeabilizing effect compared to other tested peptides with detectable MIC (Figure 3). Similar evidences has been obtained using SYTOX Green as reporter of membrane permeabilization. Interestingly, SYTOX signal is often stronger at pH 7 than pH 5 and increasing peptide concentration over the MIC value showed a decrease in fluorescence (Supp. Figure 3). This fact supports the hypothesis MytC-derived peptides have permeabilizing effects at low concentrations, without displaying growth inhibition, and switching to a different effect when exceeding the MIC.

Binding to bacterial components

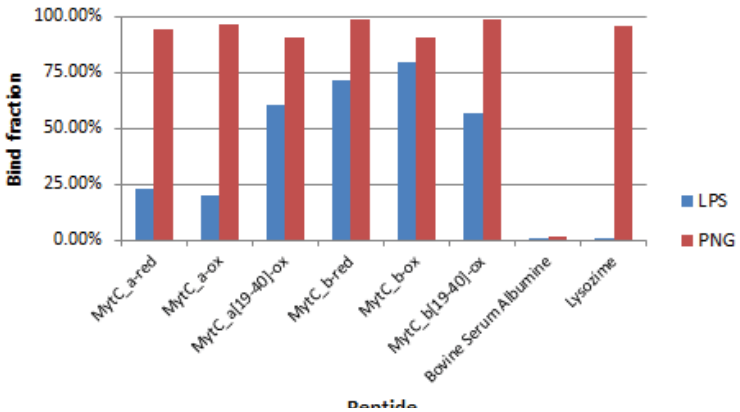


Figure 4: Following 30 minutes incubation and centrifugation at , the percentages of mytC-related peptide recovered in the pellet represent the peptide fraction bound to the bacterial components LPS and PNG in the applied experimental conditions.

All the tested peptides were able to bind LPS with different affinities, in general at higher degree for the MytC-b-related peptides (Figure 4). The interaction with LPS has been reported for other AMPs known to interfere with bacterial membranes[54]. In addition, all the tested peptides strongly interacted with the *B. subtilis* PNG (Figure 4). This finding is quite unexpected considering that such interaction could obstacle MytC-related peptides in reaching the bacterial membrane and, in this way, finally impairing their antimicrobial effect. Theoretically, a bacterium coated with peptides should trigger the haemocyte recruitment[10], and, at the same time, should facilitate bacterial agglutination and opsonisation[55].

To test such hypothesis, we observed *B. Subtilis* at the optical microscope after incubation with MytC peptides. The analysis showed the presence of agglomerates following treatment of the bacterial cells with MytC_b and its 19-40 fragment. The treatment with the mature peptides caused bigger agglomerates, suggesting the involvement of the α-helix region in the agglutination (Figure 5). Moreover, the presence of elongated bacteria could be indicative of an effect on cellular division mechanism as

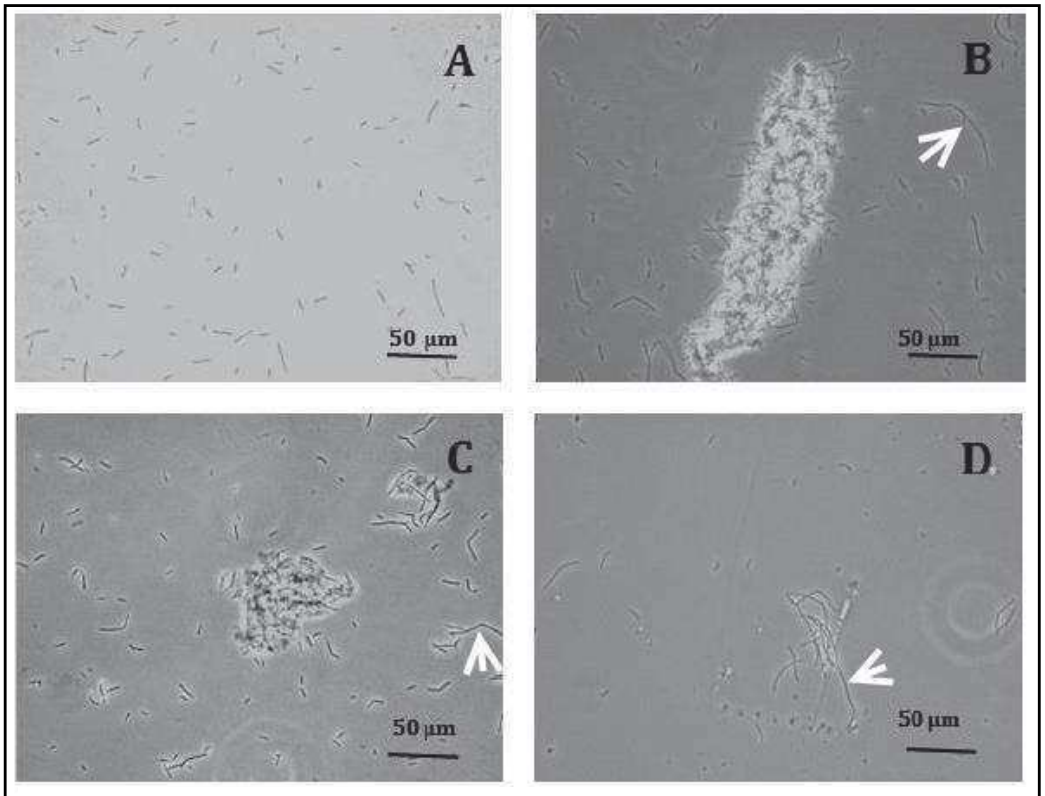


Figure 5: *B. subtilis* agglomerates after incubation of 30 minutes with MytC_b derived peptides tested at 32 µM concentration. Arrows indicates elongated bacteria. A: negative control (DMSO), B: MytC_b-red, C: MytC_b-ox, D: MytC_b[19-40]-ox

for other AMPs[56,57].

Concluding remarks

MytC is an interesting case of study both for its high gene transcript polymorphics and high expression levels in haemocytes from immune-stimulated mussels (*M. galloprovincialis*) and for its pleiotropic effects, from the chemotactic to the antiviral properties. The experimental work described in this chapter demonstrates how MytC can be effective on the bacterial surface, causing membrane permeabilization at non inhibitory concentrations, and interacting with cell wall components of both Gram- and Gram+ strains. Overall, Myticin C could support the action of other immune effectors by facilitating their translocation into a bacterial cell, as well as of immune cells, through bacterial opsonisation. At the same time, the possibility that Myticin C can impair the bacterial cell division is particularly intriguing and should be further investigated.

References

1. Gerdol, M. & Venier, P. An updated molecular basis for mussel immunity. *Fish Shellfish Immunol.* **46,** 17–38 (2015).

2. Charlet, M. *et al.* Innate immunity. Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, Mytilus edulis. *J. Biol. Chem.* **271,** 21808–21813 (1996).

3. Mitta, G., Vandenbulcke, F. & Roch, P. Original involvement of antimicrobial peptides in mussel innate immunity. *FEBS Lett.* **486,** 185–190 (2000).

4. Mitta, G., Vandenbulcke, F., Hubert, F., Salzet, M. & Roch, P. Involvement of mytilins in mussel antimicrobial defense. *J. Biol. Chem.* **275,** 12954–12962 (2000).

5. Mitta, G., Hubert, F., Noël, T. & Roch, P. Myticin, a novel cysteine-rich antimicrobial peptide isolated from haemocytes and plasma of the mussel Mytilus galloprovincialis. *Eur. J. Biochem.* **265,** 71–78 (1999).

6. Allam, B. & Raftos, D. Immune responses to infectious diseases in bivalves. *J. Invertebr. Pathol.* **131,** 121–136 (2015).

7. Allam, B. & Pales Espinosa, E. Bivalve immunity and response to infections: Are we looking at the right place? *Fish Shellfish Immunol.* **53,** 4–12 (2016).

8. Lay, F. T. *et al.* The C-terminal propeptide of a plant defensin confers cytoprotective and subcellular targeting functions. *BMC Plant Biol.* **14,** 41 (2014).

9. Zou, G., de Leeuw, E., Lubkowski, J. & Lu, W. Molecular Determinants for the Interaction of Human Neutrophil α Defensin 1 with its Propeptide. *J. Mol. Biol.* **381,** 1281–1291 (2008).

10. Martinez-Lopez, A. *et al.* pH-Dependent Solution Structure and Activity of a Reduced Form of the Host-Defense Peptide Myticin C (Myt C) from the Mussel Mytilus galloprovincialis. *Mar. Drugs* **11,** 2328–2346 (2013).

11. Pallavicini, A. *et al.* High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions. *Dev. Comp. Immunol.* **32,** 213–226 (2008).

12. Balseiro, P., Moreira, R., Chamorro, R., Figueras, A. & Novoa, B. Immune responses during the larval stages of Mytilus galloprovincialis: Metamorphosis alters immunocompetence, body shape and behavior. *Fish Shellfish Immunol.* **35,** 438–447 (2013).

13. Padhi, A. & Verghese, B. Molecular diversity and evolution of myticin-C antimicrobial peptide variants in the Mediterranean mussel, Mytilus galloprovincialis. *Peptides* **29,** 1094–1101 (2008).

14. Vera, M., Martínez, P., Poisa-Beiro, L., Figueras, A. & Novoa, B. Genomic Organization, Molecular Diversification, and Evolution of Antimicrobial Peptide Myticin-C Genes in the Mussel (Mytilus galloprovincialis). *PLOS ONE* **6,** e24041 (2011).

15. Tassanakajon, A., Somboonwiwat, K. & Amparyup, P. Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Dev. Comp. Immunol.* **48,** 324–341 (2015).

16. Cassone, M. & Jr, L. O. Synergy among antibacterial peptides and between peptides and small-molecule antibiotics. *Expert Rev. Anti Infect. Ther.* **8,** 703–716 (2010).

17. Schmitt, P., Lorgeril, J. de, Gueguen, Y., Destoumieux-Garzón, D. & Bachère, E. Expression, tissue localization and synergy of antimicrobial peptides and proteins in the immune response of the oyster Crassostrea gigas. *Dev. Comp. Immunol.* **37,** 363–370 (2012).

18. Yan, H. & Hancock, R. E. W. Synergistic Interactions between Mammalian Antimicrobial Defense Peptides. *Antimicrob. Agents Chemother.* **45,** 1558–1560 (2001).

19. Mei, W. *et al.* A preliminary analysis on the construction and expression of eukaryotic expression vectors of mytilin and myticin from Mytilus coruscus. *Agric. Sci. Technol.* (2010).

20. Balseiro, P. *et al.* Mytilus galloprovincialis Myticin C: A Chemotactic Molecule with Antiviral Activity and Immunoregulatory Properties. *PLOS ONE* **6,** e23140 (2011).

21. Novoa, B. *et al.* Antiviral Activity of Myticin C Peptide from Mussel: an Ancient Defense against Herpesviruses. *J. Virol.* **90,** 7692–7702 (2016).

22. Rosani, U. *et al.* Massively Parallel Amplicon Sequencing Reveals Isotype-Specific Variability of Antimicrobial Peptide Transcripts in Mytilus galloprovincialis. *PLOS ONE* **6,** e26680 (2011).

23. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738 (2010).

24. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinforma. Oxf. Engl.* **29,** 845–854 (2013).

25. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **38,** 236–240 (1999).

26. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–549 (2010).

27. Wiegand, I., Hilpert, K. & Hancock, R. E. W. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3,** 163–175 (2008).

28. Zhang, J., Peng, S., Cheng, X. & Wang, H. Functional analysis of hybrid peptide CAMA-Syn: expression in mammalian cells and antimicrobial potential. *Protein Pept. Lett.* **19,** 1076–1081 (2012).

29. Tossi, A., Tarantino, C. & Romeo, D. Design of synthetic antimicrobial peptides based on sequence analogy and amphipathicity. *Eur. J. Biochem. FEBS* **250,** 549–558 (1997).

30. Epand, R. F., Pollard, J. E., Wright, J. O., Savage, P. B. & Epand, R. M. Depolarization, Bacterial Membrane Composition, and the Antimicrobial Action of Ceragenins. *Antimicrob. Agents Chemother.* **54,** 3708–3713 (2010).

31. Pulido, D. *et al.* Antimicrobial action and cell agglutination by the eosinophil cationic protein are modulated by the cell wall lipopolysaccharide structure. *Antimicrob. Agents Chemother.* **56,** 2378–2385 (2012).

32. Park, S.-C. *et al.* Amphipathic alpha-helical peptide, HP (2-20), and its analogues derived from Helicobacter pylori: pore formation mechanism in various lipid compositions. *Biochim. Biophys. Acta* **1778,** 229–241 (2008).

33. Wehbi, H. *et al.* The Peptidoglycan-Binding Protein FimV Promotes Assembly of the Pseudomonas aeruginosa Type IV Pilus Secretin. *J. Bacteriol.* **193,** 540–550 (2011).

34. Franzoi, M., Sturlese, M., Bellanda, M. & Mammi, S. A molecular dynamics strategy for CSαβ peptides disulfide-assisted model refinement. *J. Biomol. Struct. Dyn.* **0,** 1–25 (2016).

35. Yeaman, M. R. & Yount, N. Y. Unifying themes in host defence effector polypeptides. *Nat. Rev. Microbiol.* **5,** 727–740 (2007).

36. Zhu, S., Gao, B. & Tytgat, J. Phylogenetic distribution, functional epitopes and evolution of the CSalphabeta superfamily. *Cell. Mol. Life Sci. CMLS* **62,** 2257–2269 (2005).

37. Laederach, A., Andreotti, A. H. & Fulton, D. B. Solution and Micelle-Bound Structures of Tachyplesin I and Its Active Aromatic Linear Derivatives,. *Biochemistry (Mosc.)* **41,** 12359–12368 (2002).

38. Gueguen, Y. *et al.* Characterization of a defensin from the oyster Crassostrea gigas. Recombinant production, folding, solution structure, antimicrobial activities, and gene expression. *J. Biol. Chem.* **281,** 313–323 (2006).

39. Roch, P., Yang, Y., Toubiana, M. & Aumelas, A. NMR structure of mussel mytilin, and antiviral-antibacterial activities of derived synthetic peptides. *Dev. Comp. Immunol.* **32,** 227–238 (2008).

40. Holaday, S. K., Martin, B. M., Fletcher, P. L. & Krishna, N. R. NMR Solution Structure of Butantoxin. *Arch. Biochem. Biophys.* **379,** 18–27 (2000).

41. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

42. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–D230 (2014).

43. Becucci, L., Valensin, D., Innocenti, M. & Guidelli, R. Dermcidin, an anionic antimicrobial peptide: influence of lipid charge, pH and Zn2+ on its interaction with a biomimetic membrane. *Soft Matter* **10,** 616–626 (2013).

44. Hirsch, J. G. ANTIMICROBIAL FACTORS IN TISSUES AND PHAGOCYTIC CELLS. *Bacteriol. Rev.* **24,** 133–140 (1960).

45. Rolland, S. *et al.* pH controls both transcription and post-translational processing of the protease BcACP1 in the phytopathogenic fungus Botrytis cinerea. *Microbiol. Read. Engl.* **155,** 2097–2105 (2009).

46. Hernroth, B. *et al.* Impact of ocean acidification on antimicrobial activity in gills of the blue mussel (Mytilus edulis). *Fish Shellfish Immunol.* **55,** 452–459 (2016).

47. Andersson, D. I., Hughes, D. & Kubicek-Sutherland, J. Z. Mechanisms and consequences of bacterial resistance to antimicrobial peptides. *Drug Resist. Updat.* **26,** 43–57 (2016).

48. Rodríguez-Niklitschek, C. & Oporto V, G. H. Clinical implications of Enterococcus faecalis microbial contamination in root canals of devitalized teeth: Literature review. *Rev. Odontológica Mex.* **19,** e177–e182 (2015).

49. Ericksen, B., Wu, Z., Lu, W. & Lehrer, R. I. Antibacterial Activity and Specificity of the Six Human α-Defensins. *Antimicrob. Agents Chemother.* **49,** 269–275 (2005).

50. Santo, K. P., Irudayam, S. J. & Berkowitz, M. L. Melittin Creates Transient Pores in a Lipid Bilayer: Results from Computer Simulations. *J. Phys. Chem. B* **117,** 5031–5042 (2013).

51. Matsuzaki, K., Murase, O., Fujii, N. & Miyajima, K. Translocation of a Channel-Forming Antimicrobial Peptide, Magainin 2, across Lipid Bilayers by Forming a Pore. *Biochemistry (Mosc.)* **34,** 6521–6526 (1995).

52. Lee, M.-T., Sun, T.-L., Hung, W.-C. & Huang, H. W. Process of inducing pores in membranes by melittin. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 14243–14248 (2013).

53. Wimley, W. C. Describing the Mechanism of Antimicrobial Peptide Action with the Interfacial Activity Model. *ACS Chem. Biol.* **5,** 905–917 (2010).

54. Giuliani, A., Pirri, G. & Rinaldi, A. in *Antimicrobial Peptides* (eds. Giuliani, A. & Rinaldi, A. C.) 137–154 (Humana Press, 2010).

55. Robert, É. *et al.* Mimicking and Understanding the Agglutination Effect of the Antimicrobial Peptide Thanatin Using Model Phospholipid Vesicles. *Biochemistry (Mosc.)* **54,** 3932–3941 (2015).

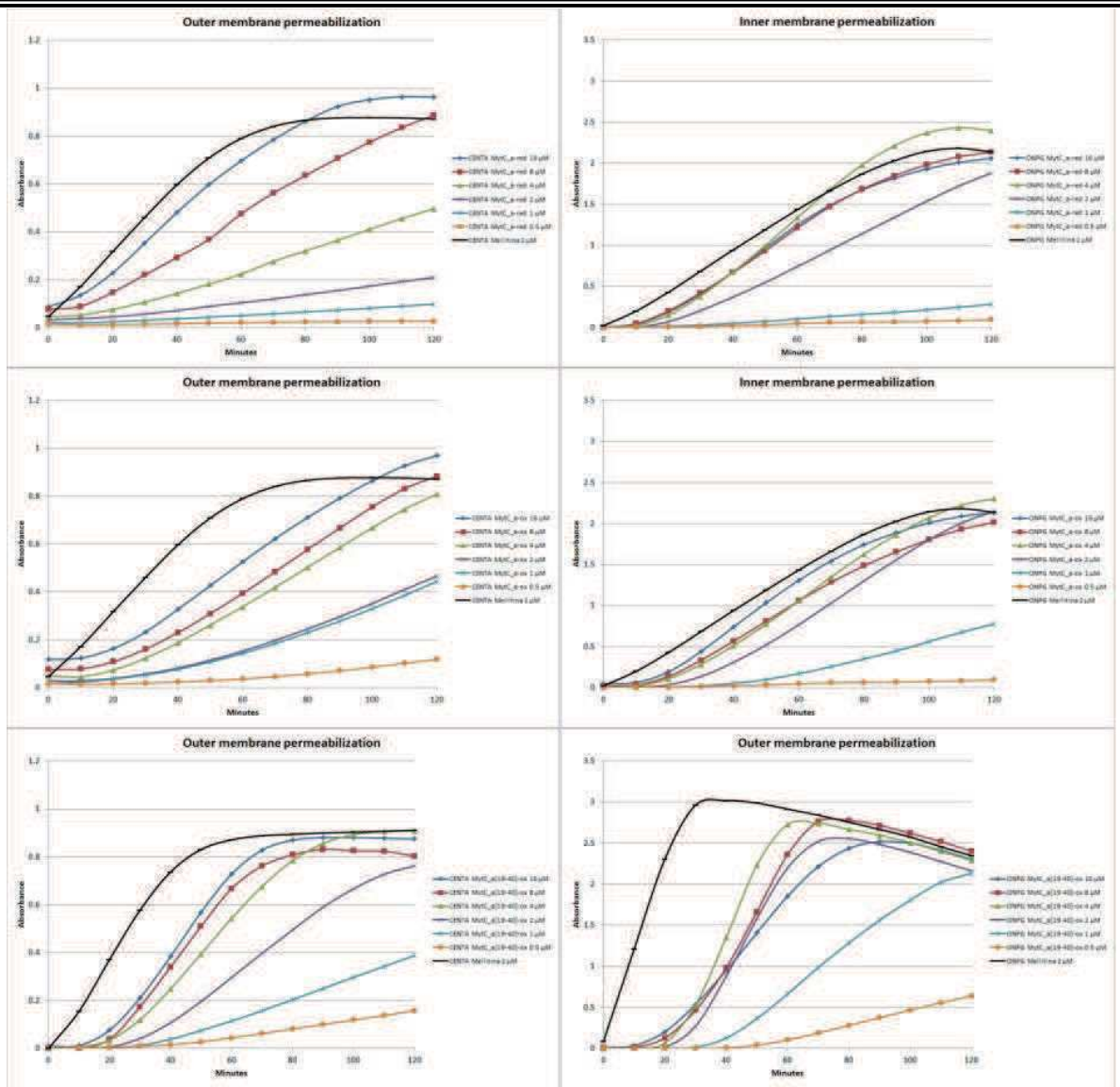
56. Ray, S., Dhaked, H. P. S. & Panda, D. Antimicrobial peptide CRAMP (16–33) stalls bacterial cytokinesis by inhibiting FtsZ assembly. *Biochemistry (Mosc.)* **53,** 6426–6429 (2014).

57. Yadavalli, S. S. *et al.* Antimicrobial peptides trigger a division block in Escherichia coli through stimulation of a signalling system. *Nat. Commun.* **7,** 12340 (2016).

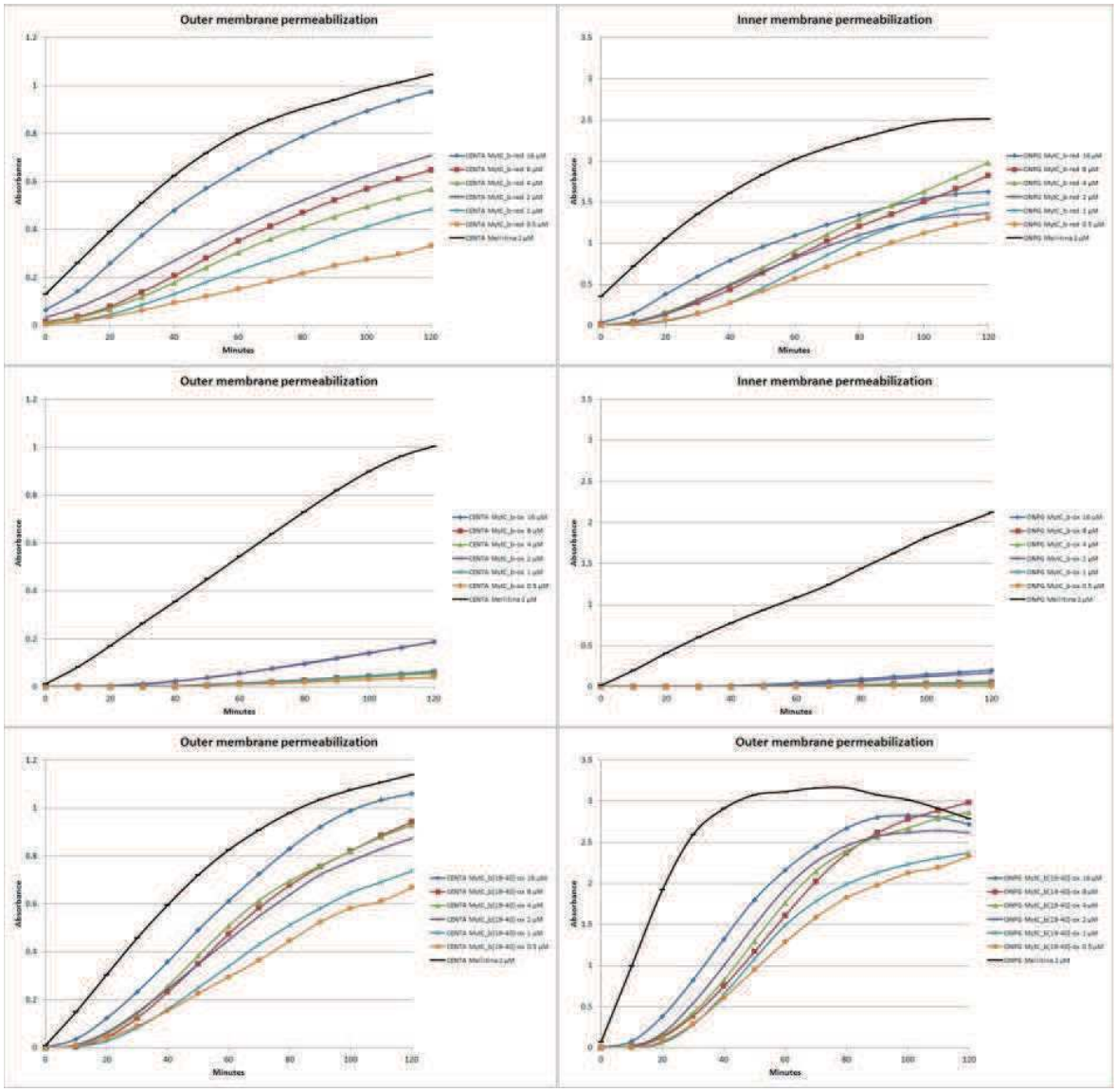
Supplementary material

	Charges	*E. coli*	*P. aeruginosa*	*E. faecalis*	*S. aureus*	*B. subtilis*
PE	+-	80%	60%			12%
PG	-	15%	21%	19%	57%	70%
CL	--	5%	11%	27%	19%	4%
DGDG					24%	
L-PG	++-			20%		
Other			8%	34%		26%

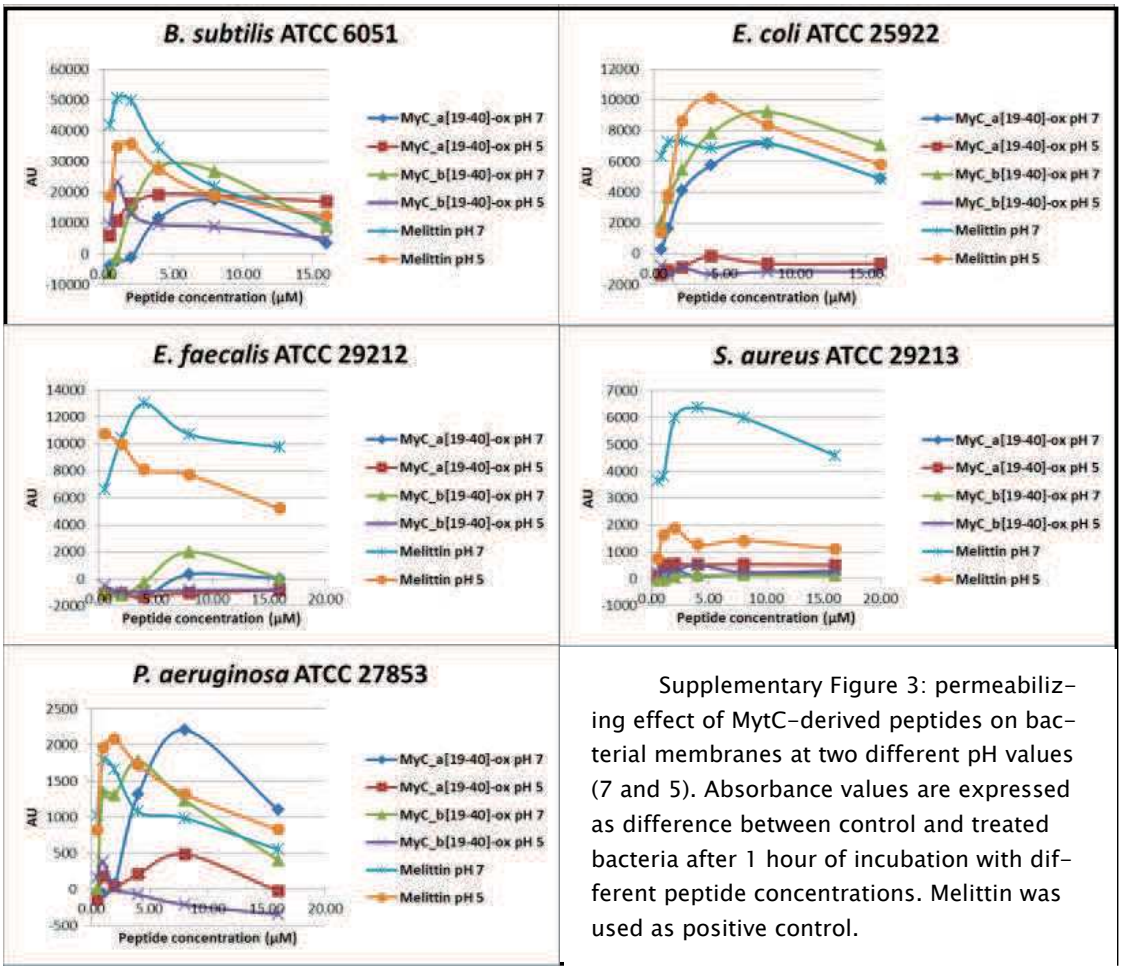
Supplementary Table 1: lipid composition of bacterial plasma membranes, expressed in percentage. PE: Phosphatidylethanolamine; PG Phosphatidylglycerol; CL: Cardiolipin; DGDG: digalactosyldiacylglycerol; L-PG: lysyl-phosphatidylglycerol.



Supplementary figure 1: kinetic measurement of the outer (left) and inner (right) membrane permeabilization induced in E. coli ML-35 pYC by MytC-derived peptides. Data were collected every 10 minutes for 2 hours. Melittin was used as positive control. Data exemplify one representative biological replicate and are reported as mean of three technical replicates. From top to bottom: MytC_a-red, MytC_a-ox, MytC_a[19-40]-ox.



Supplementary Figure 2: kinetic measurement of the outer (left) and inner (right) membrane permeabilization induced in E. coli ML-35 pYC by MytC-derived peptides. Data were collected every 10 minutes for 2 hours. Melittin has been used as positive control. Data exemplify one representative biological replicate and are reported as mean of three technical replicates. From top to bottom: MytC_b-red, MytC_b-ox, MytC_b[19-40]-ox.



Supplementary Figure 3: permeabilizing effect of MytC-derived peptides on bacterial membranes at two different pH values (7 and 5). Absorbance values are expressed as difference between control and treated bacteria after 1 hour of incubation with different peptide concentrations. Melittin was used as positive control.

Chapter 4

Recombinant production of *Mytilus galloprovin-cialis* AMPs in *Pichia pastoris*.

Franzoi M., Essig A., Venier P.

Contribution:
Structural models
Cloning
Expression
Purification
Antimicrobial testing

Chapter 4

Abstract

The innate immunity of *Mytilus galloprovincialis* has been extensively investigated, but, in recent years, new interesting evidences emerge from transcriptomic and genomic data. However, despite several families of antimicrobial peptides (AMP) have been identified, only few of them have been efficiently produced and characterized. Aiming to characterize selected peptides expressed in *Mytilus galloprovincialis*, we undertaken the heterologous expression of five putative mussel AMP belonging to different families in the methylotrophic yeast *Pichia pastoris*. As a result of the work done so far, we were able to effectively produce an arthropod defensin-like (ADL) peptide and myticusin. Despite its negative net charge at pH 7, ADL was demonstrated to be active against both Gram+ and Gram- bacteria.

Introduction

In recent years, marine organisms have been extensively investigated as possible sources of new antimicrobial peptides[1]. Marine environment is characterized by high salt concentration and slightly alkaline pH, conditions usually limiting the activity of known AMPs[2,3]. Moreover, marine viruses and bacteria commonly occur in high concentration, about 10^9 and 10^6 count/ml of sea water, respectively[4]. To defend themselves from such competitors, marine organisms express a great variety of AMPs, with the most promising and abundant families being those of peptides forming hairpin-like β-sheet and α-helical/β-sheet mixed motifs stabilized by intramolecular disulphide bonds, namely the CSαβ peptides[4].

Among marine organisms, mussels have been identified as promising candidates for the discovery of new active molecules. In fact, as filter feeders they come in contact with many microorganisms and, lacking adaptive immunity, they must have evolved efficient innate systems to fight potential pathogens[5]. Moreover, adult mussels are sessile and cannot move from highly contaminated areas. For this reason they need to react immediately and strongly to avoid diseases and mortality. We pointed our attention on *Mytilus galloprovincialis*, a species present in the Mediterranean Sea, introduced in South Africa and in northern Pacific Ocean coasts for breeding[6] and currently colonizing the South Australian coasts. . The Mediterranean mussel is reported more resistant than other bivalves to seasonally occurring infective syndromes, and it is economically relevant for many geographical regions in the world.

Several families of CSαβ peptides have been identified in *Mytilus galloprovincialis* in recent years. The defensin MGD1 has been extensively characterized and showed antimicrobial activity mainly against Gram+ bacteria[7,8]. Actually, the MGD family counts at least five peptides[9]. In addition, two sequences have been identified as putative arthropod-like defensins[5]. Mytilins count several members and also divergent sequences of so-called pseudomytilins. Mytilin B shows a typical defensin fold[10] and was demonstrated to be active against both bacteria and viruses[10,11]. Myticins are presented in chapter 3. Mytimycins have been so far identified as antifungal peptides, active against *Neurospora crassa* and *Fusarium* spp.[12], but the involvement of mytimycin variants in other functional roles cannot be excluded[5]. Mytimacins are expressed in tissues other than haemocytes, probably secreted to sustain defensive reactions at mucosal level[13]. Big-defensins are one of the few examples of trans-defensins in invertebrates[14]. Such peptides have been proposed as ancestors of vertebrates defensins[15]. Finally, a large and highly variable family of Cysteine Rich Peptides (MgCRP) has been identified but their functional roles, and their possible involvement in defence mechanisms, still requires investigation[16].

In order to investigate their antimicrobial activity, we focused our attention on five putative antimicrobial peptides representing different families, four of them identified

only as transcript sequences: pseudomytilin-1 (AJQ21532.1), myticin C (CAM56810.1), Arthropod Defensin-Like 1 (AJQ21502.1), mytimycin “short” and myticusin, both the latter from transcriptomic data produced in our laboratory and not yet published.

Material and methods

Chemicals and strains

Escherichia coli (strain DH5α) and *Pichia pastoris* (strain NRRLY11430) were used for all cloning and expression studies. All chemicals, if not otherwise mentioned, were bought at the highest available purity from Sigma-Aldrich.

Structure modelling

The myticin C (MyC) structure has been taken from Domeneghetti et al., 2015[17]. Arthropod Defensin-Like 1 (ADL) and pseudomytilin-1 (Ps) models derived from the application of the method described in chapter 2[18]. Myticusin (Myt) and Mytimycin "short" (MyS) structures have not been refined due to the difficulty in defining their disulfide array.

Cloning

DNA sequences for the prepropeptides were purchased from GenScript (Piscataway, NJ 08854, USA), codon-optimized for the expression in *Pichia pastoris* and with flanking EcoRI and SalI restriction sites, useful for cloning in pPICZ A vector (Invitrogen) (Supplementary Table 1). After restriction of insert and plasmid (O/N 37 °C), products were purified with PCR Clean-Up kit (Macherey-Nagel) and ligation was performed with 30 ng of insert, 50 ng of vector, 1 µl of T4 Dna-ligase (NEB), 1µl of ligase buffer (NEB) and water up to 10 µl, at RT for 4 hours. The entire solution was then used for transformation into chemo-competent *E. coli* cells. The resulting plasmid was isolated, verified by sequencing (Microsynth, Balgach, Switzerland), linearized with SacI, and transformed into *P. pastoris* as previously described[19,20]. Transformants were confirmed by colony PCR using the following primers: 5'-AOX1, 5'-CCTCCAATACGGTCCCTACCCTC-3', and 3'-AOX1, 5'-CAGGCAAATGGCATTCTGACATCC-3'[21] and sequenced[22].

αfactor-AMPs were generated from prepropeptides sequences cloning only the propeptide in frame with *Pichia* αfactor in the pPICZα vector (Invitrogen). Oligo-nucleotides were designed to flank the propeptide sequence carrying a tag for XhoI and NotI restriction enzymes (Supplementary Table 2). A touch down PCR was performed by using the pPICZA-AMP plasmid as template[23]. After the PCR run, Clean-Up and restriction reactions were performed in XhoI (NEB), NotI (NEB) and NEB 3.1 buffer. Subsequent steps were performed as previously described.

Plasmids encoding for αfactor-AMP-His^6 were created starting from previous plasmids and using the experimental setup derived from the Q5 protocol of New England

BioLabs (NEB, Ipswich, MA). In particular, two oligo-nucleotides were designed to exclude the propeptide (where present), the stop codon and the other undesired nucleotides between the mature peptide and the 6xHis tag on the plasmid. In brief, the initial PCR reaction was executed in a total volume of 25 µl with the Phusion high-fidelity DNA polymerase (Thermofisher), 3 ng of plasmid template, and the corresponding forward and reverse primers at 1 µM concentration (Supplementary Table 2). Circularization of the PCR product and removal of the template plasmid was performed in a total volume of 22 µl T4 ligase buffer (NEB) with 2 µl of the PCR reaction and the following enzymes: 0.3 µl FD-DpnI (Thermo), 0.3 U/µl T4 PNK (Thermo), 9 U/µl T4 ligase (NEB). The reaction mixture was incubated at room temperature for 45 min and subsequently at 37 °C for 30 min before stopping it on ice. The entire solution was then used to transform chemo-competent *E. coli* cells. Schemes of all constructs and peptide abbreviations are reported in Figure 1.

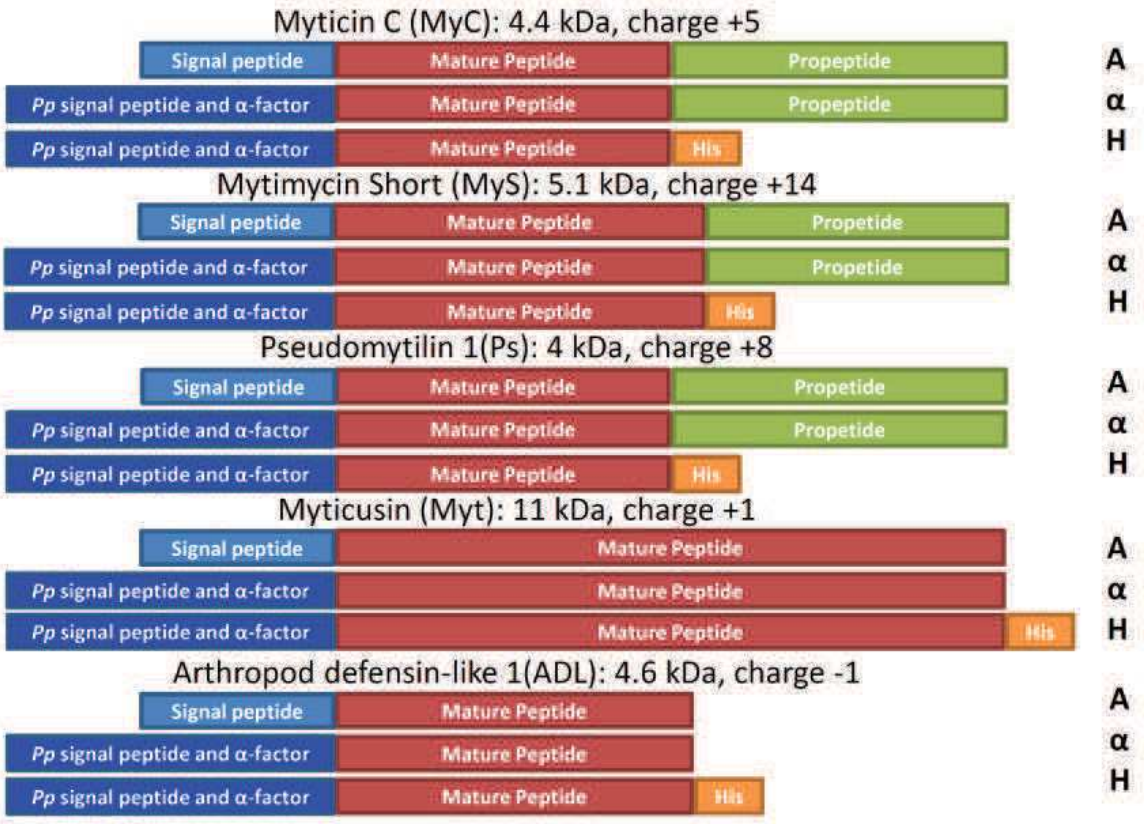


Figure 1: schematic representation and abbreviation of studied constructs. Mature peptides molecular weight and calculated isoelectric point are also reported.

P. pastoris cultivation in 96-well microtiter plates

Positive *P. pastoris* transformants were inoculated in 200 µl BMDY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.3% (w/v) YNB without amino acids (BD Bioscience), 100 mM potassium phosphate buffer pH 6, 1% (w/v) D-glucose) in a 96-half-deepwell plate (Enzyscreen, Haarlem, The Netherlands) and cultured for 24 h (28 °C, 225 rpm). After

centrifugation (3000 x g, 20 °C, 10 min), supernatant were discarded and the cell pellets were resuspended in 190 µl BMMY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.3% (w/v) YNB without amino acids (BD Bioscience), 100 mM potassium phosphate buffer pH 6, 1% (v/v) methanol) and cultured for 72 h (28 °C, 225 rpm). Methanol was added to 1% (v/v) in a time interval of 12 h. Supernatant was harvested by centrifugation (2000 x g, 20 °C, 10 min) and stored at 4 °C.

Peptide expression and purification

A single colony of recombinant *Pichia pastoris* was inoculated in 10 ml of BMDY medium in a 100 ml baffled flask for 8 hours at 28 °C, than diluted in 90 ml of BMDY in a 1 l baffled flask ON at 28 °C. Cells were collected by centrifugation (2000 x g, 20 °C, 5 min) and resuspended in 100 ml minimal medium (1.34% YNB, 0.1 M potassium phosphate buffer (pH 6.0), 0.4 mg biotin, 0.5% Ammonium chloride, 10 ml MeOH, deionized water to 1 l). The culture broth was incubated for 72 h (28 °C, 180 rpm) and methanol was added to 1% (v/v) in a time interval of 12 h. Expression was confirmed with SDS-PAGE and silver staining or, when possible, in western blot using antiHistag-HRP-conjugated antibody.

Antimicrobial testing

Agar diffusion tests were performed on LB agar plates. Plates were overlaid with 8 ml (round plates) or 15 ml (square plates) of water agar inoculated with 10^7 cells/ml of the tested bacterial strains. Antibiotic assays disks (Whatman, UK) were wetted two times with 20 µl of the tested solution. Results were checked after 24 hours at 37 °C.

Broth microdilution susceptibility test was performed as previously described in chapter 3.

Results and discussion

Structural modelling

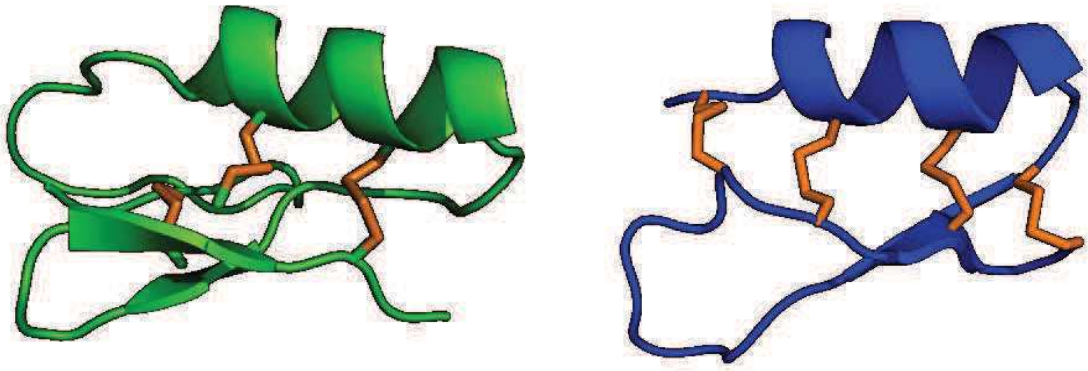


Figure 2: cartoon representation of ADL (green) and pseudomytilin (blue) structures derived from the refining protocol described in chapter 2. Disulfides are represented in orange.

The ADL sequence presents a signal peptide and no propeptide. Even if the signal peptide predictors suggested the cleavage to occur at Ala21, the comparison with homologous sequences indicates the mature peptide to start with Leu25 and proteolysis to occur after the two Arg23-24, a typical proteolytic site already reported for similar sequences[24,25]. Interestingly ADL shows -1 net charge at pH7, being so far the only negatively charged defensin identified in mussels. The predicted ADL structure is typical of defensins, even lacking the Gly-X-Cys or Cys-X-Gly motif, usually found in such peptides[26]. The α-helix is connected to two β-strands by two disulfide bonds. ADL exhibits a 16 AA N-ter extension in addition to the common CSαβ fold, connected to one β-strand through a disulfide bond, as already found for other defensins, e.g. insect defensin A (pdb entry: 1ica)[27], with a calculated RMSD between the two of 3.0 Å.

The pseudomytilin fold is similar to already characterised mytilin-A[10], with the same disulfide array and a RMSD of 2.8 Å (pdb entry: 2eem). Compared to mytilin-A, pseudomytilin presents an highly charged N-terminal helical domain, with six positively charged amino acids in the first 13 positions.

Arthropod Defensin-Like expression and activity

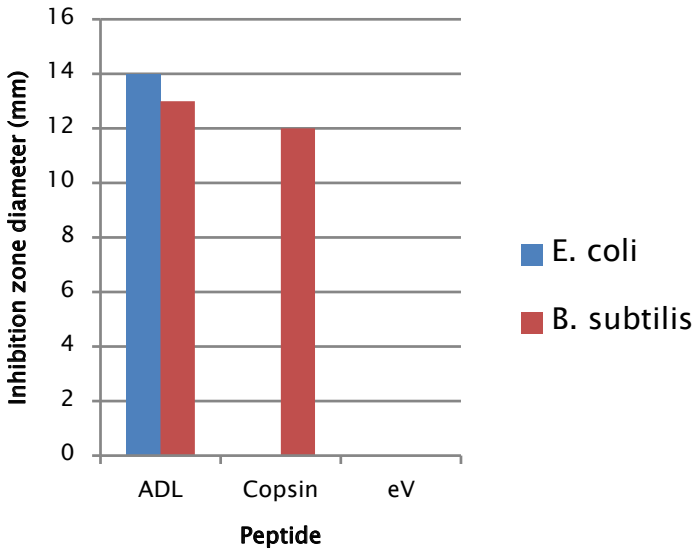


Figure 3: Inhibition zones from ADL and Copsin. LB plates were overlaid with water agar containing the bacterial inoculum. Total 40 µl of *P.pastoris* supernatant was used to wet filter paper disks for the assay.

ADL_A was expressed in 100 ml BMMY for 72 hours at 28 °C. Aliquots of 40 µl of supernatant were used for diffusion agar assays against *E. coli* and *B. subtilis*, considered as reporter strains for Gram- and Gram+ bacteria, respectively. Copsin was used as positive control against *B. subtilis* and an empty vector mutant (eV) was used as negative control. The assay produced an inhibition halo against both *E. coli* and *B. subtilis* (Figure 3). According to the APD database[28], the antimicrobial activity of an anionic mussel defensin it is demonstrated here for the first time. When we replaced the ADL signal peptide with the *Pichia pastoris* α-factor, the supernatant of the expression batch was active not only against *E. coli* and *B. subtilis*, but also it produced a 9 mm inhibition zone against *P. aeruginosa*. Such activity, not previously noticed, is probably connected to an higher yield in ADL production. In order to purify the active peptide, an expression assay of ADL with a His-tag fused at the C-terminal was performed in minimal medium and the peptide purification was carried on a Nickel affinity column (Figure 4). The resulting fractions showed significant antimicrobial activity against *B. subtilis*, compared to His-tagged Copsin as positive control and eV as negative control (Supp. Figure 1). Such activity was completely lost after concentration, dialysis or desalting in C18 column (data not shown). Actually we are working on a different purification protocol with a gel filtration column optimized for such small peptides.

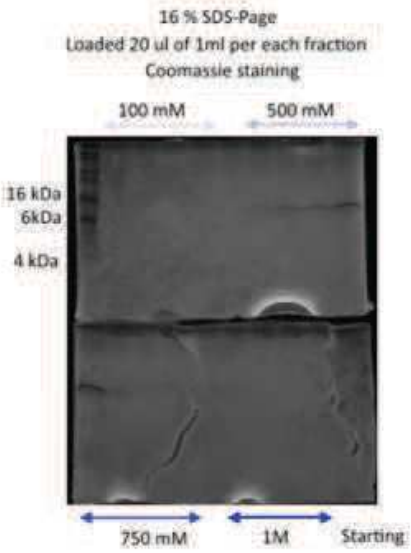


Figure 4: SDS-page of Ni^{+}-chromatography fractions. Three elution steps (500 mM, 750 mM, and 1 M imidazole) were performed, Bands corresponding to the ADL_H m.w. (5.5 kDa) are visible between 500 mM and 750 mM imidazole.

Myticusin expression

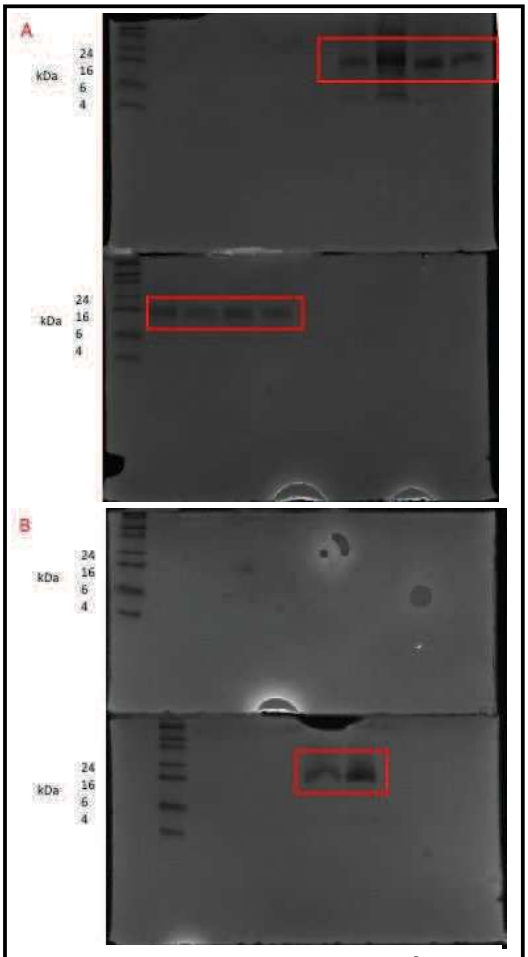


Figure 5: SDS-page of Ni⁺-chromatography fractions (panel A) and gel filtration fractions (panel B). Positive fractions, corresponding to the Myt_H m.w. (11.8 kDa) , are highlighted in red.

The expression of Myt_α and Myt_H in full and minimal medium were confirmed with silver staining (Myt_α), dot blot and western blot (Myt_H). Myt_H was purified using Ni⁺-affinity column and eluted with increasing concentrations of imidazole. Concentrated positive fractions were further purified in gel filtration column. Myt_H is efficiently purified in Ni⁺-affinity with few impurities eventually deriving from partial degradation or column contamination (Figure 5), panel A). No impurities were detected after gel filtration, allowing the collection of the pure peptide for further testing (Figure 5 panel B). The total yield from 100 ml of culture resulted to be about 10 mg calculated with BCA assay (Thermo). By now, no antimicrobial activity was detected for Myt, but it has been possible to test its MIC in the broth microdilution susceptibility test only up to 20 μM and only against *B. subtilis*, *S.aureus*, *P. aeruginosa, E. faecalis* and *E. coli*. The Myt_H expression needs further improvements to obtain an adequate amount of peptide in order to confirm the putative antimicrobial activity[29].

Other peptides

The other expressed peptides (MyC, Ps and MyS) did not show activity in no one of the tested conditions. From anti His-tag dot blot and western blot of AMP_H versions we propose some consideration. Based on dot blot, MyC seems to be expressed but no band was detectable in western blot. Thus, we think that the terminal Arg-Arg motif of the MyC mature peptide could be recognized by *P. pastoris* proteases and, as consequence, the His-tag could be cleaved off.

The supernatant of *P. pastoris* expressing Ps_H shows a His-tag positive band at around 15 kDa, corresponding to Ps mature peptide plus the α-factor, suggesting that the mature peptide part prevents proteolytic cleavage. In particular, the starting Arginine residue can prevent Kex2 protease activity[30].

Finally MyS is expressed at very low levels. This is not surprising considering Myticusins are considered strictly antifungal and, for this reason, can have detrimental effect on the selected recombinant organism[31].

Concluding remarks

Post-translational modification can impair the retrival of functional proteins. This is particularly true in the case of chemically synthesized cysteine-rich peptides, because the development of effective refolding protocols is often expensive and time consuming. The use of eukaryotic cells for recombinant production can lead to better yields. Nevertheless, in the native organism the folding of proteins can be linked to highly specific processes. Moreover, host-specific characteristics, such as proteases or grow conditions, can be detrimental for recombinant production.

Starting from five peptide sequences, each in three different constructs, we were able to effectively produce just two of them and, actually, only one showed antimicrobial activity. We were able, for the first time, to demonstrate the antimicrobial activity of a mussel anionic defensin and further work will be done in order to effectively purify and completely characterize such molecule.

References

1. Rajanbabu, V., Chen, J.-Y. & Wu, J.-L. in *Springer Handbook of Marine Biotechnology* (ed. Prof, S.-K. K.) 747–758 (Springer Berlin Heidelberg, 2015).

2. Goldman, M. J. *et al.* Human β-Defensin-1 Is a Salt-Sensitive Antibiotic in Lung That Is Inactivated in Cystic Fibrosis. *Cell* **88,** 553–560 (1997).

3. Wu, G. *et al.* Effects of Cations and PH on Antimicrobial Activity of Thanatin and s-Thanatin Against Escherichia coli ATCC25922 and B. subtilis ATCC 21332. *Curr. Microbiol.* **57,** 552–557 (2008).

4. Falanga, A. *et al.* Marine Antimicrobial Peptides: Nature Provides Templates for the Design of Novel Compounds against Pathogenic Bacteria. *Int. J. Mol. Sci.* **17,** (2016).

5. Gerdol, M. & Venier, P. An updated molecular basis for mussel immunity. *Fish Shellfish Immunol.* **46,** 17–38 (2015).

6. Springer, S. A. & Crespi, B. J. Adaptive gamete-recognition divergence in a hybridizing Mytilus population. *Evol. Int. J. Org. Evol.* **61,** 772–783 (2007).

7. Hubert, F., Noël, T. & Roch, P. A Member of the Arthropod Defensin Family from Edible Mediterranean Mussels (Mytilus galloprovincialis). *Eur. J. Biochem.* **240,** 302–306 (1996).

8. Yang, Y. S. *et al.* Solution structure and activity of the synthetic four-disulfide bond Mediterranean mussel defensin (MGD-1). *Biochemistry (Mosc.)* **39,** 14436–14447 (2000).

9. Venier, P. *et al.* Insights into the innate immunity of the Mediterranean mussel Mytilus galloprovincialis. *BMC Genomics* **12,** 69 (2011).

10. Roch, P., Yang, Y., Toubiana, M. & Aumelas, A. NMR structure of mussel mytilin, and antiviral-antibacterial activities of derived synthetic peptides. *Dev. Comp. Immunol.* **32,** 227–238 (2008).

11. Mitta, G., Vandenbulcke, F., Hubert, F., Salzet, M. & Roch, P. Involvement of mytilins in mussel antimicrobial defense. *J. Biol. Chem.* **275,** 12954–12962 (2000).

12. Charlet, M. *et al.* Innate immunity. Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, Mytilus edulis. *J. Biol. Chem.* **271,** 21808–21813 (1996).

13. Gerdol, M., De Moro, G., Manfrin, C., Venier, P. & Pallavicini, A. Big defensins and mytimacins, new AMP families of the Mediterranean mussel Mytilus galloprovincialis. *Dev. Comp. Immunol.* **36,** 390–399 (2012).

14. Shafee, T. M. A., Robinson, A. J., Weerden, N. van der & Anderson, M. A. Structural homology guided alignment of cysteine rich proteins. *SpringerPlus* **5,** 1–7 (2016).

15. Zhu, S. & Gao, B. Evolutionary origin of β-defensins. *Dev. Comp. Immunol.* **39,** 79–84 (2013).

16. Gerdol, M. *et al.* Identification and Characterization of a Novel Family of Cysteine-Rich Peptides (MgCRP-I) from Mytilus galloprovincialis. *Genome Biol. Evol.* **7,** 2203–2219 (2015).

17. Domeneghetti, S. *et al.* Structural and Antimicrobial Features of Peptides Related to Myticin C, a Special Defense Molecule from the Mediterranean Mussel Mytilus galloprovincialis. *J. Agric. Food Chem.* **63,** 9251–9259 (2015).

18. Franzoi, M., Sturlese, M., Bellanda, M. & Mammi, S. A molecular dynamics strategy for CSαβ peptides disulfide-assisted model refinement. *J. Biomol. Struct. Dyn.* **0,** 1–25 (2016).

19. Essig, A. *et al.* Copsin, a Novel Peptide-based Fungal Antibiotic Interfering with the Peptidoglycan Synthesis. *J. Biol. Chem.* **289,** 34953–34964 (2014).

20. Wu, S. & Letchworth, G. J. High efficiency transformation by electroporation of Pichia pastoris pretreated with lithium acetate and dithiothreitol. *BioTechniques* **36,** 152–154 (2004).

21. Kombrink, A. Heterologous production of fungal effectors in Pichia pastoris. *Methods Mol. Biol. Clifton NJ* **835,** 209–217 (2012).

22. Lõoke, M., Kristjuhan, K. & Kristjuhan, A. Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques* **50,** 325–328 (2011).

23. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19,** 4008 (1991).

24. Richman, A. M. *et al.* Inducible immune factors of the vector mosquito Anopheles gambiae: biochemical purification of a defensin antibacterial peptide and molecular cloning of preprodefensin cDNA. *Insect Mol. Biol.* **5,** 203–210 (1996).

25. Viljakainen, L. & Pamilo, P. Identification and molecular characterization of defensin gene from the ant Formica aquilonia. *Insect Mol. Biol.* **14,** 335–338 (2005).

26. Yeaman, M. R. & Yount, N. Y. Unifying themes in host defence effector polypeptides. *Nat. Rev. Microbiol.* **5,** 727–740 (2007).

27. Cornet, B. *et al.* Refined three-dimensional solution structure of insect defensin A. *Structure* **3,** 435–448 (1995).

28. Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* gkv1278 (2015). doi:10.1093/nar/gkv1278

29. Liao, Z. *et al.* Molecular characterization of a novel antimicrobial peptide from Mytilus coruscus. *Fish Shellfish Immunol.* **34,** 610–616 (2013).

30. Bader, O., Krauke, Y. & Hube, B. Processing of predicted substrates of fungal Kex2 proteinases from Candida albicans, C. glabrata, Saccharomyces cerevisiae and Pichia pastoris. *BMC Microbiol.* **8,** 116 (2008).

31. Mitta, G., Vandenbulcke, F. & Roch, P. Original involvement of antimicrobial peptides in mussel innate immunity. *FEBS Lett.* **486,** 185–190 (2000).

Supplementary material

Supplementary Table 1: sequences of AMPs presented in this study, codon optimized for production in *Pichia pastoris*.

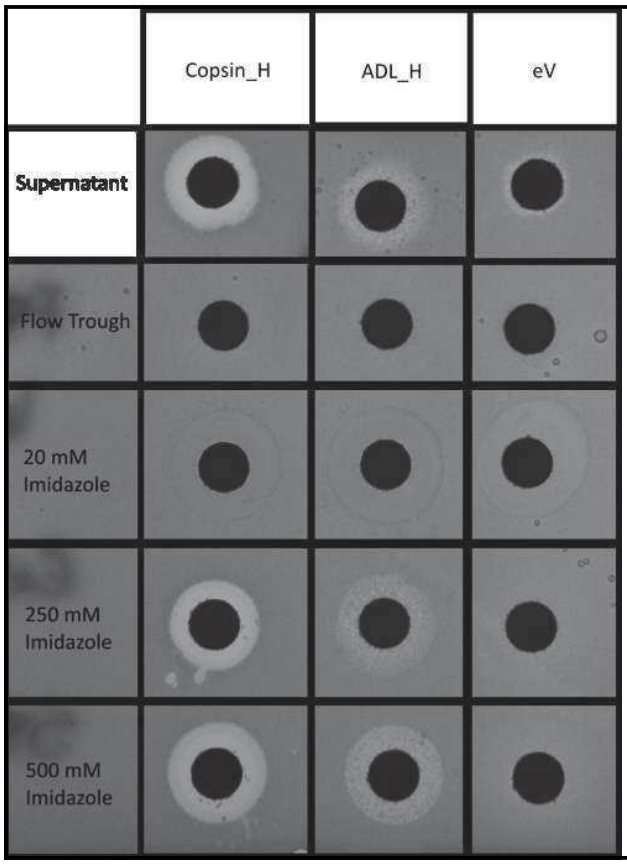
Peptide	Sequence
Myticin C (MyC)	**GAATTC**ATGAAGGCTACTATTTTGTTGGCTGTCGTTGTCGCTGTTATTGTCGGAGTCC AGGAGGCTCAGTCTGTCGCTTGTACCAGTTATTACTGTTCTAAGTTTTGTGGTTCTGC TGGTTGTTCTTTGTACGGTTGTTACAAGTTGCATCCAGGTAAAATCTGTTACTGTTTG CACTGTAGAAGAGCTGAATCTCCTTTGGCTTTGTCTGGTTCTGCTAGAAACGTTAACG ATAAGAACAACGAGATGGACAATTCCCCTGTTATGAACGAGGTTGAGAACTTGGACC AGGAGATGGATATGTTTTAG**GTCGAC**
Pseudomytilin (Ps)	**GAATTC**ATGAAAACCGCTATCTTCTTGGCAATCACCGTTTATGTCTTGTTCGCTGTCC AGGATGCTGACGCTAGATGTAGAACCAGATGTAGATTGAAGTGTTTTGGTAGAGGTT GTGGTGCTTACTTTGCTGCTGAAAGAGGTCCATTCTGTTTGTGTAAGTGTTACAGATG TGCTAACGAATCTATCACTAAGATCGGTGTTATTCAAACTGTTGAGCAATCTTCTATG TTCCACCAGCAGATTGATAATTTCGCAGAAGAGCAGGCACAGTTCATTCAGTAG**GTC GAC**
Mytitusin (Myt)	**GAATTC**ATGGAGAGTGTTGCTTTGTACTGTTTGTTGTTTTGTTTCGTTGCTTTGGGTA ATATCCAGGTTTCCGTCGGTTCAGATCACCAGATGGCTCAATCTGCTTGTATGGGTG TTGCTCAAGATGCTGCTTACGCTTCTGCTATTCCAAGAACTTGTTCTGGTGGTAAAAC TTGTGAATCTATTTGTGCTGATGCTACTTCTACTATGAACAAATACTCTACTGCTGGT GGTCCTTTGTCTACTGCTAGATGTGTTGATGCTTTTCATATCTACTCTAGAAGAGGTC AAGAGGATGTTCCATACCAACCTTTCGTTGTTTCTTGGAAGTATGGAGTCAAGGGTT GCTTTCAAACTAACTGCGGACCTAACTTCTGTTGCTGTATTGTCTAA**GTCGAC**
Mytimicin Short (MyS)	**GAATTC**ATGTCATTGCTTTTGAGATTGACTCTTTTGTTCGTTATCTGTAGTGTCATCA TGGAAAACGCTAATGGTGCCTGTTGCAGAAGAAAGAGAAGAGGTAGATGTGGAGACT GCACAAAGCCAACTCCTTGTTGCGGTTACAAAAGATGTAACATTTTCTGTTGCGGAT GTAATTGCAGAAAGGGTGGATGTGGTGGAAGAAGAAAAAGACAACTTTCTCATGACA CTACCGATAGATCCACCTATATGACATTCATTGGTTTGGACATCAACAAGAATGGAA TGATGGAACAGTTTGAGTTCTCTAAAGCTCTTGAGCAAATGCAGATTACTGATAACG TTACCGTCTTGCATCACTGGTCCGTTATGGATGAAGACAAGGATGGACTTATCACTA TGGAAGAGTTCGATAGAGAGAAATTGTAA**GTCGAC**
Arthropod-like defensins (ALD)	**GAATTC**ATGTTTAAGTTGGGTTTCCTTATTGTTGGATTGGTCTTTCTTATCTCTGGTA CAGAAGGAGCCCCACAAAGAAGATTGACTTGTGAGGCTGGTTCCGCCAACATTTTTG GTGTTCAGTTGGGAACTACCGCATGTCTTACCCATTGCTTGTTCCTTGGTCACTTGGG AGCTTACTGCGATTCACATAATGTCTGTCATTGCAGACACTAA**GTCGAC**

Supplementary Table 2: primers used for construction of plasmids containing *Mytilus galloprovincialis* derived peptides with *P. pastoris* α-factor.

MyC	For	TCTCTCGAGAAAAGACAGTCTGTCGCTTGTACCAG
	Rev	CTGGCGGCCGCCGACTAAAACATATCCATCTCCTGG
Ps	For	TCTCTCGAGAAAAGAAGATGTAGAACCAGATGTAGATTG
	Rev	CTGGCGGCCGCCGACTACTGAATGAACTGTGCCTG
Myt	For	TCTCTCGAGAAAAGAGTTTCCGTCGGTTCAGATC
	Rev	CTGGCGGCCGCCGACTTAGACAATACAGCAACAGAAG
MyS	For	TCTCTCGAGAAAAGAGCCTGTTGCAGAAGAAAGAG
	Rev	CTGGCGGCCGCCGACTTACAATTTCTCTCTATCGAAC
ADL	For	TCTCTCGAGAAAAGATTGACTTGTGAGGCTGGTTC
	Rev	CTGGCGGCCGCCGACTTAGTGTCTGCAATGACAG

Supplementary Table 3: primers used for construction of plasmids containing *Mytilus galloprovincialis* derived peptides with a His-tag at the C-terminal.

MyC	For	ACATCTGTAACACTTACACAAACAGAATGGAC
	Rev	GGTTCTGGACATCATCATCATCATCATTGAGTTTGTAG
Ps	For	GACAATACAGCAACAGAAGTTAGGTCCGC
	Rev	GGTTCTGGACATCATCATCATCATCATTGAGTTTGTAGC
Myt	For	TCTTCTACAGTGCAAACAGTAACAGATTTTACC
	Rev	GGTTCTGGACATCATCATCATCATCATTGAGTTTGTAG
MyS	For	TTTTCTTCTTCCACCACATCCAC
	Rev	GGTTCTGGACATCATCATCATCATCATTGAGTTTGTAG
ADL	For	GTGTCTGCAATGACAGACATTATGTG
	Rev	GGTTCTGGACATCATCATCATCATCATTGAGTTTGTAG



Supplementary Figure 1: purification of *P. pastoris* expressing ADL_H supernatant in Nickel affinity column. His tagged Copsin has been used as positive control, empty vector as negative control. 40 µl of the resulting fractions were spotted in agar diffusion test against *B. subtilis*. Eluting imidazole concentration are reported.

Chapter 5

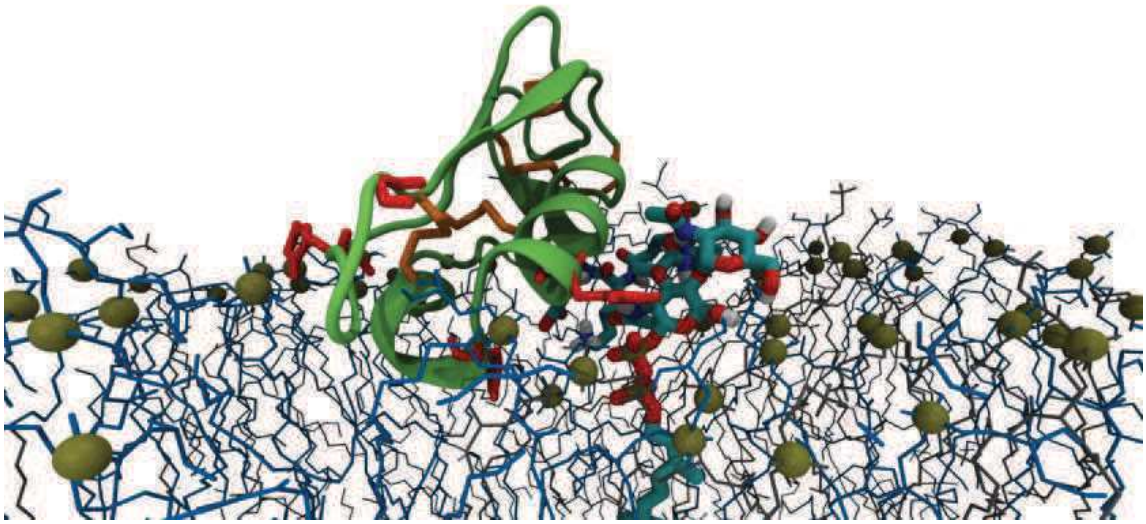
High throughput generation and computational characterization of copsin improved mutants

M. Franzoi, Y. van Heuvel, S. Thomann, N. Schürch, P. Kallio, M. Aebi, P. Venier, A. Essig

Contribution:
Mutant library
Expression
Antimicrobial testing
Molecular dynamics simulations

Abstract

Defensins are a class of cysteine-rich antimicrobial peptides, expressed by virtually all eukaryotes as part of their innate immune response. Due to their unique mode of action and rapid killing of pathogenic microbes, defensins are considered as promising alternatives to clinically applied antibiotics. Copsin is a defensin-like peptide, previously identified in the mushroom *Coprinopsis cinerea*. It is active against a range of Gram positive bacteria through binding to the peptidoglycan precursor lipid II and prevention of a proper cell wall formation. In this study, we present a new workflow for the generation, production and activity-driven selection of copsin derivatives, based on the expression in *Pichia pastoris*. The generation of 152 single amino acid mutants, and combinations thereof, allowed the identification of k-copsin, a peptide variant exhibiting a significantly improved activity against *Bacillus subtilis* and *Staphylococcus aureus*. Furthermore, we characterized *in silico* the membrane interactions of both copsin and k-copsin, in the presence and absence of lipid II. The molecular dynamics data highlighted a variable binding to lipid II, with a clear preference for the MurNAc moiety involved in 47% and 35% of the total contacts with copsin and k-copsin, respectively. The mutated amino acids were located in loop-regions of k-copsin and appeared to be crucial for the perturbation of the bacterial membrane. Taken together, these structural data on the interaction between copsin and its bacterial target provide a better understanding of how natural defensin-like peptides can be further developed for antibacterial therapies.



Improved copsin derivative (green ribbon structure) bound to Lipid II and bacterial membrane at the endpoint of a 300 ns molecular dynamics simulation.

Chapter 5

Introduction

Defensins are a major family of antimicrobial peptides (AMPs), conserved across the eukaryotic kingdom[1,2]. They are defined as relatively short (10–80 amino acids), cysteine-rich peptides with a net positive charge at pH 7 and a high proportion of hydrophobic residues. As key effectors of innate immunity, defensins rapidly kill invading microbes, such as bacteria and fungi. A current classification of defensins is based on conserved structural elements and disulfide patterns. Vertebrate defensins, for instance the mammalian α-, β- and θ-defensins, consist of a β-sheet structure stabilized by a specific connectivity of six cysteine residues[3,4]. Consistent with their evolutionary diversification, a number of vertebrate defensins exert a broad spectrum activity against microbes (Gram positive/negative bacteria, fungi, as well as viruses) and possess immunomodulatory properties[3,5,6]. Invertebrates, fungi and plants predominantly express the so-called cysteine-stabilized αβ-defensins (CSαβ), which are characterized by one helix and two anti-parallel β-strands interconnected by three disulfide bonds[4,7]. This core structure can be further extended by a β-strand and additional disulfide bonds, which have been identified in plant and fungal defensins[8–10]. In comparison to their counterparts in vertebrates, CSαβ defensins exhibit higher specific activity against bacterial or fungal subgroups[9,11–13].

The cell membrane and cell wall biosynthesis of bacteria are primary targets of many defensins. Based on the interaction of the cationic, amphipathic AMPs with bacterial membranes, the proposed action modes included peptide binding to negatively charged phospholipid head-groups, further interference with the lipid bilayer and pore formation or unspecific membrane perturbation[14,15]. However, recent studies described more targeted killing mechanisms, for example, peptide binding to membrane-bound undecaprenyl pyrophosphoryl-MurNAc-pentapeptide-GlcNAc, termed lipid II, and interference with the peptidoglycan biosynthesis. Lipid II is a unique, conserved structure among Eubacteria and it is essential for building the cell wall, therefore representing an ideal target for antibiotics. In particular, CSαβ defensins gain specificity by binding to lipid II [13,17,18]. Modelling of structural NMR-data revealed that the fungal defensin plectasin binds specifically to the pyrophosphate moiety of lipid II and to the D-glutamic acid of the pentapeptide. Additionally, the hydrophobic part of plectasin locates at the membrane interface, but there was no membrane permeabilization detectable[17]. A lower affinity for lipid II was described for mammalian defensins, such as the human beta-defensin 3 (hBD3)[13,19].

Multiple and divergent defensins evolved as effective antimicrobial molecules over millions of years without physiological evidence of bacterial resistance, as observed for clinically used antibiotics[1]. Evolutionary distant antibacterial CSαβ scaffolds may lead to a more specific therapeutic action in humans[20,21]. Hence, understanding the mode of action of fungal CSαβ peptides effective against relevant pathogenic bacteria could provide interesting hints for the development of next generation of antibiotics[17]. Copsin, which

was discovered in the mushroom *Coprinopsis cinerea,* is composed of a CSαβ core fold interconnected by six disulfide bonds and an N-terminal pyroglutamate[22]. It shows a bactericidal activity against Gram positive bacteria, such as Listeria, Enterococci, or Bacilli. Thorough biochemical studies revealed a stoichiometric interaction with lipid II with involvement of the third position of the pentapeptide. In this study we describe a new strategy, based on Golden Gate cloning and expression in *Pichia pastoris*, to generate copsin derivatives and to measure their antibacterial activity in high throughput manner. In order to highlight the lipid II binding mechanism, we performed molecular dynamics simulations in membranes of Gram positive bacteria with native copsin and one mutant being significantly more effective against *Bacillus subtilis* and *Staphylococcus aureus*.

Chapter 5

Material and methods

Chemicals and strains

Escherichia coli (strain DH5α) and *Pichia pastoris* (strain NRRLY11430) were used for all cloning and expression studies. All chemicals, if not otherwise mentioned, were at the highest available purity and were purchased from Sigma-Aldrich.

Antimicrobial activity – Bacillus spp.

The following strains of the *Bacillus cereus* group were tested: *B. cereus* ATCC 10876, ATCC 11778, BWB; *B. anthracis* NCTC 8234 (Sterne), NCTC 10340 (Vollum), Pasteur, Ba11 (clinical isolate, CH 2014). *B. anthracis* strains were processed under BSL-3 conditions and *B. subtilis* 168 was used as the control strain in all experiments. MIC values were determined by using the microdilution broth method according to the CLSI guidelines with minor modifications[23]. In brief, bacteria were cultivated overnight on Columbia blood agar with 5% sheep blood (Biomérieux, Switzerland) and colonies were homogenized in Mueller-Hinton broth (MHB; BD Diagnostics, NJ) to McF 1 (corresponding to 10^7 cfu/ml for the *B. cereus* group and 10^8 cfu/ml for *B. subtilis* respectively). The bacterial suspensions were diluted in cation-adjusted MHB (CAMHB; BD Diagnostics) to a concentration of 10^6 cfu/ml and added to two-fold dilution series of copsin (0.25 -128 µg/ml) in 96-well polypropylene microtiter plates. The plates were incubated at 37°C for 20 h. The MIC is defined as the lowest concentration of copsin without visible growth and was determined by applying the four-eye-principle. The cfu/ml was determined by plating dilutions of the growth control on tryptic soy agar plates. Each strain was tested at least in two separate experiments.

Fermentation and purification of recombinant copsin

The codon-optimized prepro-copsin insert (**Figure S1**) was cloned into the pPICZA plasmid using the restriction enzymes EcoRI/SalI (pPICZA-optCopsin; GenScript, NJ). The pPICZA-optCopsin construct was linearized with SacI and transformed into *P. pastoris* by electroporation as described previously[22,24]. The standard methanol-limited fed-batch procedure was performed according to the *Pichia* fermentation process guidelines of Invitrogen (www.thermofisher.com). In brief, three to five colonies of pPICZA-optCopsin transformants were cultured in 40 ml of BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.3% (w/v) YNB without amino acids (BD Diagnostics), 100 mM potassium phosphate buffer pH 6, 1% (w/v) glycerol) containing 100 µg/ml Zeocin (LabForce, Switzerland) in a baffled shake-flask overnight at 28 °C. The entire culture broth was used to

inoculate a 3.6 l Labfors 5 bioreactor (Infors, Switzerland) filled with 1.2 l of basal salt medium (4% (w/v) glycerol) supplemented with PTM_1 trace salts. The cultivation temperature was kept at 28 °C and the pH was maintained at 5.0 by adding ammonium hydroxide. The aeration rate was kept constant at 1.2 l/min. The batch mode of fermentation process was performed, first, using an agitation speed of 750 rpm for 24-30 h, followed by a limited glycerol feed at 15.5 ml/h for 4 h. After depletion of glycerol, indicated by a rise of dissolved oxygen (DO) to >80%, a methanol pulse (5 ml/l) was conducted. After a DO spike, the methanol fed-batch phase was started and maintained for 72 h. The DO level was maintained at 10-20% of oxygen saturation by regulating the methanol feed rate (5.7-9.5 ml/h per liter of the initial volume) and agitation speed (1000-1400 rpm).

The fermentation broth was centrifuged (8000 x g, 20 °C, 20 min) and the resulting supernatant was diluted in ddH_2O to a conductivity of 9-11 mS/cm. The pH of the supernatant was adjusted to 7 by adding ammonium hydroxide, vacuum filtered (0.22 µm rapid-Filtermax 500; Techno Plastic Products TPP, Switzerland) and loaded on a self-made SP Sephadex cation exchange column pre-equilibrated with binding buffer A (20 mM sodium phosphate pH 7, 10 mM NaCl). The column was washed with 17% buffer B (20 mM sodium phosphate pH 7, 1 M NaCl) and copsin was eluted with 60% buffer B. The elution was monitored at UV absorbance of 210 nm and 280 nm. Fractions containing copsin were pooled and concentrated in a 2 kDa Spectra/Por dialysis membrane (Spectrum Laboratories, CA) as previously described[22]. The concentrated eluate was dialyzed against buffer C (2 mM sodium phosphate buffer pH 7.3, 70 mM NaCl) at 4°C for 24 h. Further copsin concentration and purification was achieved by using a 6-8 kDa Spectra/Por dialysis membrane (Spectrum Laboratories) and by dialyzing against buffer C. The protein concentration was determined by a BCA protein assay, according to the manufacturer's instructions (Thermo Scientific, MA).

Site-directed mutagenesis - Whole-plasmid PCR

The experimental setup was derived from the Q5 site directed mutagenesis protocol of New England BioLabs (NEB, MA). Site-directed substitutions were done using the pPICZA-optCopsin construct as template. In brief, the initial PCR reaction was executed in a total volume of 25 µl with the Phusion high-fidelity DNA polymerase (Thermo), 3 ng of plasmid template, and the corresponding forward and reverse primer at 1 µM (**Table S1**). Circularization of the PCR product and removal of the template plasmid was performed in a total volume of 22 µl T4 ligase buffer (NEB) with 2 µl of the PCR reaction and the following enzymes: 0.3 µl FD-DpnI (Thermo), 0.3 U/µl T4 PNK (Thermo), 9 U/µl T4 ligase (NEB). The reaction mixture was incubated 45 min at room temperature, 30 min at 37 °C and chilled on ice. The entire solution was used for transformation of chemocompetent *E. coli* cells. The resulting mutated plasmid was isolated, verified by DNA sequencing (Microsynth, Switzerland), and transformed into *P. pastoris* as described previously[22,24].

Site-directed mutagenesis - Golden Gate cloning

The Golden Gate cloning strategy was performed according to Engler *et al.*[25] and used for direct transformation of *P. pastoris* cells. Three recipient expression plasmids were designed and synthesized based on the pPICZA-optCopsin construct and the BsmBI restriction enzyme (GenScript; **Figure S2**). DNA entry fragments were synthesized with the corresponding overhangs (Microsynth) and are listed in **Table S1**. First, the complementary forward and reverse DNA entry fragments were mixed in Tris-HCl buffer (10 mM pH 7.5, 50mM NaCl), each to a concentration of 10 µM. The annealing was performed in a temperature gradient of -1 °C/min starting from 80 °C to 40 °C. The Golden Gate assembly was executed in NEB 3.1 buffer in a total volume of 10 µl with 2 µM of annealed DNA fragment, 20 nM recipient plasmid, 1 mM ATP, 2 U/µl BsmBI, and 40 U/µl T4 ligase. The reaction conditions were as follows: 25 °C (2h), 37 °C (2h), 55 °C (30 min), 80 °C (20 min), and 4 °C. For linearizing the resulting plasmid, SacI-HF (NEB) was added to 2 U/µl and the reaction mixture was further incubated at 37 °C (3h) and finally at 65 °C (20 min). After DNA clean-up, 150 ng of the linearized plasmid was used for transformation of *P. pastoris* as previously described[22,24]. Transformants were confirmed by colony PCR using the following primer pair: 5'-AOX1, 5'-CCTCCAATACGGTCCCTACCCTC-3', and 3'-AOX1, 5'- CAGGCAAATGGCATTCTGACATCC-3' [26].

P. pastoris cultivation in 96-well microtiter plates

Positive *P. pastoris* transformants were inoculated in 200 µl BMDY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.3% (w/v) YNB without amino acids (BD Diagnostics), 100 mM potassium phosphate buffer pH 6, 1% (w/v) D-glucose) in a 96-half-deepwell plate (Enzyscreen, The Netherlands) and cultured for 24 h (28 °C, 225 rpm). After centrifugation (3000 x g, 20 °C, 10 min), the cell pellets were resuspended in 190 µl BMMY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.3% (w/v) YNB without amino acids (BD Diagnostics), 100 mM potassium phosphate buffer pH 6, 1% (v/v) methanol) and cultured for 72 h (28 °C, 225 rpm). Methanol was added to 1% (v/v) in a time interval of 12 h. Supernatant was harvested by centrifugation (2000 x g, 20 °C, 10 min) and stored at 4 °C.

Kinetics assays

Supernatants of the microtiter plate cultivations were diluted (1:5) in LB-Lennox medium and 5 µl of this dilution was added to 195 µl of a *B. subtilis* (strain 168) suspension in MHB (BD Diagnostics) at an optical density of 0.005 (OD_{600}) in a 96F-well polystyrene plate (TPP). The bacterial growth was monitored every 15 min at 37°C for 14-20 h in a Spectramax Plus microplate spectrophotometer (Molecular devices LLC, CA) at a wave-

length of 600 nm, with 5 s shaking before each measurement. Data were analysed with the SoftMax pro software (Molecular devices LLC).

P. pastoris cultivation in shake flasks

Selected *P. pastoris* transformants were used to inoculate 50 ml of BMDY medium and cultured for 24 h (28 °C, 180 rpm) in baffled shake flasks. Cells were harvested by centrifugation (2000 x g, 20 °C, 5 min) and resuspended in 100 ml FM22 medium supplemented with PTM_4 trace salts and 1% (v/v) methanol[26]. The cells were cultivated for 72 h (28 °C, 180 rpm) and methanol was added to 1% (v/v) in a time interval of 12 h. Cells were removed by centrifugation (3000 x g, 20 °C, 10 min) and supernatants were further processed as described for the fermentation product but omitting the 6-8 kDa dialysis step. The protein concentration was determined by a BCA protein assay, according to the manufacturer's instructions (Thermofisher).

Antimicrobial activity - Copsin mutants

The MIC was determined by the microdilution broth method[23] for *B. subtilis* and for *Enterococcus faecium* (ATCC 19434). Two-fold dilution series (1-8 µg/ml) of copsin and its mutants were prepared in 96-well polypropylene microtiter plates in CAMHB (BD Diagnostics). Bacteria were grown in LB-Lennox medium to an OD_{600} of 0.2-0.6 and added to the dilution series to a final OD_{600} of 0.005 for *B. subtilis* and of 0.0001 for *E. faecium*. The microtiter plates were incubated at 37 °C for 20 h and the MIC was determined as the lowest concentration of copsin without visible bacterial growth. All MIC measurements were done in duplicates. A kinetics assay was performed for *Staphylococcus aureus* (ATCC 35556) as described above. LB-Lennox medium (copsin mutants) or CAMHB (copsin mutants combined) was used with a starting OD_{600} of 0.0005 and a polypeptide concentration of 80 µg/ml for copsin and its mutants.

Membrane models

The starting coordinates of pre-equilibrated palmitoyloleoylphosphatidylglycerol (POPG)/paolmitoyloleoylphosphatidylethanolamine (POPE) bilayer and POPG/POPE bilayer with one molecule of lipid II were taken from the website of the Laboratory of Biomolecular Modelling of the Russian Academy of Science such as lipid II force field parameters (http://model.nmr.ru/files/lipid-II_bilayers.zip)[27]. The POPG/POPE molecule ratio was 3:1 to simulate the membrane of gram Positive bacteria such as *B. subtilis*. Lipids force fields were taken from the Biocomputing laboratory of the University of Calgary (http://moose.bio.ucalgary.ca/index.php?page=Structures_and_Topologies)[28] according to Berger *et al.*[29].

Protein models

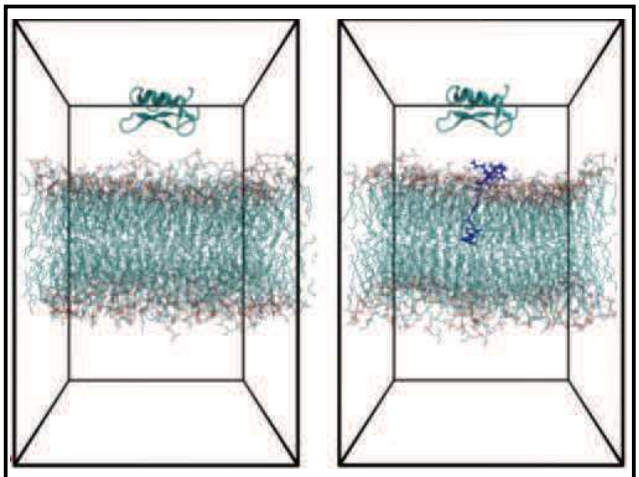


Figure 1: Examples of starting configurations without (left) and with (right) lipid II. Copsin is represented as cartoon in cyan, lipid II in dark blue and other lipids as thin lines.

The copsin starting structures were retrieved from the PDB entry 2MN5. Mutant version named k-copsin was created from 2MN5 using the Pymol mutate plug-in[30] and performing 2000 steps of energy minimization with GROMACS[31]. From NMR data, His26 was positively charged (i.e. doubly protonated) and His21 had a single proton on δ nitrogen[22]. New histidines of K-copsin were all protonated accordingly to PROPKA prediction at pH 6[32]. The principal axis of the peptides were aligned to cell x axis. Four independent simulations were performed rotating the starting cospin structure of 0°, 90°, 180° and 270° along the x axis and centring the peptide on the membrane xy plane, above the outer leaflet, setting a minimum distance with the membrane of 20 Å. System were solvated with single point charge (SPC) water molecules and a proper amount of Na and Cl atoms were added to counter ion the negative membrane and the positive peptide, respectively. Examples of starting configuration can be found in **Figure 1**.

Molecular dynamics simulations

All the simulations were performed in GROMACS v. 5.0.1[31] applying Gromos96 43a2x[33] force-field and 2 fs steps at 310 K with three separate thermostats for Protein, Membrane and water with ions groups. Membranes were centred in a triclinic box maintaining x and y axis length and with z axes length of 13 nm. The protein was added to the system centred on x and y axis and z coordinate set to 11.6 nm. The box was filled with spc water model and Na^+ and Cl^- ions were added to neutralize membrane and protein charges respectively. All the systems were subjected to a 20000 step energy minimization using a steepest descended integrator. Than a position restrained NVT equilibration of 50000 steps were performed using V-rescale algorithm for temperature coupling[34] and Berendsen pressure coupling setting[35]. PME algorithm was used for long-range electrostatic[36]. As the last preparation step, we performed a 50000 position restrained NPT simulation, switching to semiisotropic Parrinello-Rahman pressure coupling algorithm[37]. Finally, the protein position restraints were released and 300 ns production runs were performed. **Table 1** provides a summary of the performed simulations. The analysis of the runs was performed with GROMACS tool g_order, g_distance, g_mindist, g_mdmat and in house scripts. Secondary structure was computed using dssp[38]. Membrane thickness was computed using GridMAT-MD 2.0 and plotting results with Gnuplot (www.gnuplot.info)[39]. Membrane curvature was calculated using g_lomepro[40].

Protein	Membrane composition			Ions		Length
	lipid II	POPG	POPE	Na^+	Cl^-	
Copsin wt	1	186	66	186	6	4 x 300 ns
	0	216	72	216	6	4 x 300 ns
Copsin mutant	1	186	66	186	6	4 x 300 ns
	0	216	72	216	6	4 x 300 ns
Membrane-Lipid II control	1	186	66	186	8	300 ns
Membrane control	0	216	72	216	8	300 ns
Copsin	0	0	0	0	6	100 ns

Table 1: summary of performed simulations.

Results and discussion

Antimicrobial activity of copsin against Bacillus spp.

Bacteria of the genus *Bacillus* are rod-shaped, have aerobic metabolism, and produce protective endospores[41]. The minimal inhibitory concentration (MIC) of copsin was determined for vegetative cells of the soil bacterium *B. subtilis* and of the human pathogenic species *B. cereus* and *B. anthracis*, the causative agent of anthrax[42]. The latter are both members of the *B. cereus* group of *Bacilli* that share a high degree of chromosomal sequence conservation[43,44]. Using the microdilution broth method, copsin exhibited a distinct antimicrobial profile with MIC values of 128 µg/ml for *B. cereus*, 16-64 µg/ml for *B. anthracis* and 4 µg/ml for *B. subtilis* (**Table 2**). These findings further highlight the diversity and selective targeting of this CSαβ defensin against closely related species[11].

Bacillus strain	Source	MIC[1,2] [µg/ml]
B. subtilis 168		4
B. cereus	ATCC 10876	128
B. cereus	ATCC 11778	128
B. cereus BWB		128
B. anthracis Sterne	NCTC 8234	16-32
B. anthracis Vollum	NCTC 10340	16-32
B. anthracis Pasteur		16-32
B. anthracis[3]		64

Table 2: MIC of Copsin against strains of the genus *Bacillus*. [1]The MIC was determined according to the CLSI standard M07-A9. [2]Copsin used for these assays was produced in a fermentation run. [3]Clinical isolate, Switzerland, 2014.

Structure-function studies on copsin

To study the structure-function relationship of copsin, site-directed mutations were introduced into the polypeptide chain and evaluated according to their impact on the *in vitro* activity of copsin. Two experimental strategies were applied for cloning. The first approach was based on whole-plasmid PCR using primers generating mutations at defined positions of the genetic sequence of copsin (**Table S1**). The mutated plasmids were purified from *Escherichia coli* and linearized for transformation into the expression host *P. pastoris*. Although this method allows the creation of great variety of mutations, the initial PCR and cloning steps in *E. coli* are time consuming and error-prone. Therefore, as second approach, we adapted the so-called "Golden Gate" cloning for direct transformation into *P. pastoris*. Golden Gate cloning depends on type IIs restriction enzymes,

which cleave outside of their recognition site[25]. This allows to perform restriction digestion of an expression plasmid and ligation of a desired DNA fragment in one step (**Table S1** and **Figure S1**). We combined this cloning method with a linearization step for homologous recombination into *P. pastoris*. Total 152 copsin mutants were generated, containing single amino acid changes spread along the entire peptide sequence with the exception of cysteine residues (**Table S1**).

Expression of copsin mutants in *P. pastoris* was performed in a microtiter plate including empty vector and native copsin controls. As the recombinant pPICZA plasmid allows copsin to be secreted, the microculture supernatants were harvested and their antibacterial activity was evaluated in a kinetics assay. In detail, a defined volume of supernatant was mixed with a suspension of *B. subtilis* cells (strain 168) and the bacterial growth was monitored in a microtiter plate reader at an optical density of 600 nm (OD_{600}) until an OD_{600} of 0.10 was reached. According to this threshold, 17 copsin mutants showing delayed start of bacterial growth relative to native copsin (i.e. improved production level or/and activity of the recombinant copsin mutant) were selected for further experiments.

After expression in shake flasks and purification, MIC values of the 17 copsin mutants were determined for *B. subtilis* and *Enterococcus faecium* (**Table 3**). Moreover, a kinetics assay was performed for *S. aureus*, as described above, with 80 µg/ml of purified copsin mutants (**Table 3**). Overall, four mutants inhibited *B. subtilis* at higher extent than the native peptide, all of them carrying one amino acid substitution in loop regions of copsin. In detail, the S13H and T16H substitutions located in the n-loop connecting N-terminus and α-helix also showed enhanced activity against *S. aureus*, whereas the T41G and T46H substitutions located in the c-loop connecting the two β-strands did not affect the growth of *S. aureus* relative to native copsin. No improvement of MICs was observed in the case of *E. faecium*, with even higher values for T41S, T46S and P55R.

Copsin mutants	Production level [2] [µg]	MIC [ug/ml]		Time shift [min]
		B. subtilis 168	*E. faecium*	*S. aureus*
WT copsin [1]	38	8	8	Baseline (0 min)
Q1S	127	8	8	– 15
P4G	259	8	8	0
R7H	97	8	8	0
S13H	92	4	8	+ 240
T16H	312	4	8	+ 45
T16G	457	8	8	– 30
D29H	131	8	8	+ 30
V30H	309	8	8	– 15
V30R	264	8	8	– 30
S36H	56	4–8	8	Not determined
T41G	435	4–8	8	– 15
T41S	460	8	>8	– 45
F43W	384	8	8	0
L44W	188	8	8	– 15
T46H	182	4	8	0
T46S	128	8	>8	– 15
P55R	165	8	>8	0

Table 3: Antimicrobial properties for the WT and mutated copsins. [1] For the *S. aureus* kinetics assay, copsin produced in a fermenter run was used. [2] Total amount of peptide purified of a 100 ml *P. pastoris* culture.

The four mutations associated to enhanced activity against *B. subtilis* were combined in one construct (**Table S2**). The antimicrobial activity of the peptide (S13H, T16H, T41G, T46H), named k-copsin, led to a 4-fold lower MIC value for *B. subtilis* and to the most potent activity against *S. aureus*, as determined in a kinetics assay with 80 µg/ml of purified copsin mutants (**Table 4**). In addition, the yield of k-copsin after production and purification was about 15-fold higher in comparison to native copsin.

Compared to the addition of a positive charged amino acid such as Lys or Arg, mutations of native amino acids to histidine appear to be more effective in improving the antimicrobial peptide features. As possible explanation, if resistance mechanisms of bacteria target positively charged amino acids then, increasing the positive peptide charge may or may not lead to a better performance[45]. Thus, histidine could be an optimal choice, being neutral at physiological pH and switching to positively charged in a slightly acidic environment, such as bacterial surface[46].

Mutations [1]	Production level [2] [µg]	*B. subtilis* 168 MIC [3] [µg/ml]	*S. aureus* Time shift [3] [min]
Native copsin	17	8	Baseline (0 min)
T41G, S13H	213	8	+ 90
T41G, T16H	83	8	- 75
T41G, T46H, S13H	256	4	+ 195
T41G, T46H, T16H	158	4	- 45
T41G, T46H, S13H,T16H	259	2	+ 285

Table 4: Antimicrobial properties for combined copsin mutants. [1] Corresponding templates and primers are listed in suppl. Table 3. [2] Total amount of peptide purified of a 100 ml *P. pastoris* culture. [3] All experiments were performed in cation-adjusted MHB (CAMHB).

Copsin Lipid Preference

Copsin affinity to bacterial membranes

In order to investigate the copsin interaction with bacterial membranes, we performed a series of atomistic simulations reproducing a Gram positive membrane environment. Likewise, we studied the impact of mutated residues (S13H, T16H, T41G, T46H) of k-copsin on the binding to the bacterial cell surface. Copsin or k-copsin were posed 2 nm above a pre-equilibrated 3:1 POPG/POPE bilayer and the system was left free to progress for 300 ns at 310 K. In order to avoid data misinterpretation due to peptide initial placement, four independent simulations were executed for each experimental setup, changing peptide spatial orientation. In all simulation systems, copsin started to interact with the bacterial membrane within 2 - 10 ns and it stays bound during the 300 ns time frame (**Fig. S4**). As further evidence, the MD studies highlighted how copsin, both in native and k version, is able to displace lipids from the membrane external leaflet. Thus, copsin increased the interaction surface of the membrane and, at the same time, it causes a negative curvature without significantly decreasing the membrane thickness (**Figure 2**).

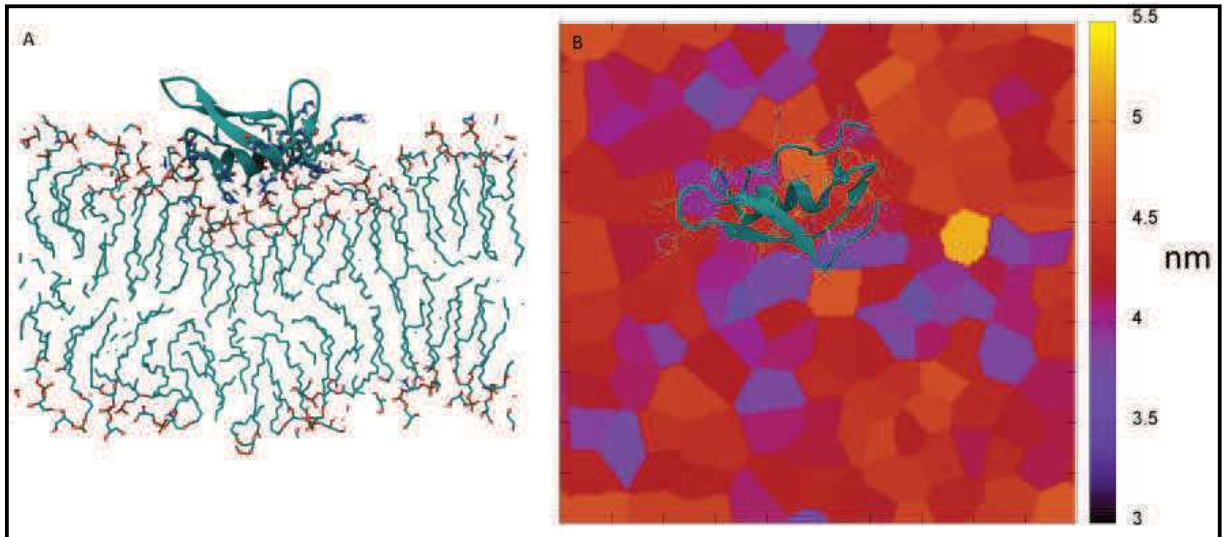


Figure 2: A: membrane bending induced by copsin (last frame of 300 ns simulation). B: membrane thickness at same time point computed using Gridmat-MD, with copsin positioned on the membrane surface.

Moreover, native copsin showed 3%-8% POPE contacts (defined as distance between two atoms $\leq$ 0.35 nm), compared to a value of 25% expected for POPG/POPE molar ratio. Thus, native copsin prefers negatively charged POPG lipids compared to zwitterionic POPE. This preference was partially lost in the mutated version (8%-16% POPE contacts) and the total number of contacts significantly increased by 30% ($p < 0.05$) (**Fig. S3**). To better study the role of mutated residues during membrane interactions, we calculated the number of contacts of the amino acids at positions 13, 16, 41, and 46. In comparison with those of native copsin, the four mutated residues showed a significantly increased number of POPG contacts (+390%, $p < 0.005$). Overall, these results suggest

that the introduced mutations enhance the copsin of affinity to membranes by preference for POPG lipids. As demonstrated for vancomycin[47], it is possible that copsin binds first to bacterial membrane and then diffuses on the membrane surface or waits until the target molecule comes into proximity to exert its function.

Copsin-lipid II interactions

As for other defensins, copsin action has been shown to involve binding to lipid II[22]. Such interaction can prevent cell wall biosynthesis and has also direct effect on bacterial cell division[48]. To gain insights into the copsin-lipid II interaction mechanism, we performed simulations in the presence of lipid II on the outer leaflet of a Gram positive membrane (**Figure S5 and S6**). Copsin interacted within 1 ns both with the membrane lipids and lipid II (**Figure S4**). Simulations exhibited a membrane bending around the copsin-lipid II complex, as also seen for copsin solely binding to membrane (**Figure 2**). The total contacts of copsin with POPE and POPG did not reveal a significant difference (<10%) between the two setups, with and without lipid II (**Figure S4**). This is unexpected especially when considering that steric effect of lipid II prevents copsin from binding part of the membrane surface. Thus, copsin could embed in a bacterial membrane more effectively when binding to lipid II.

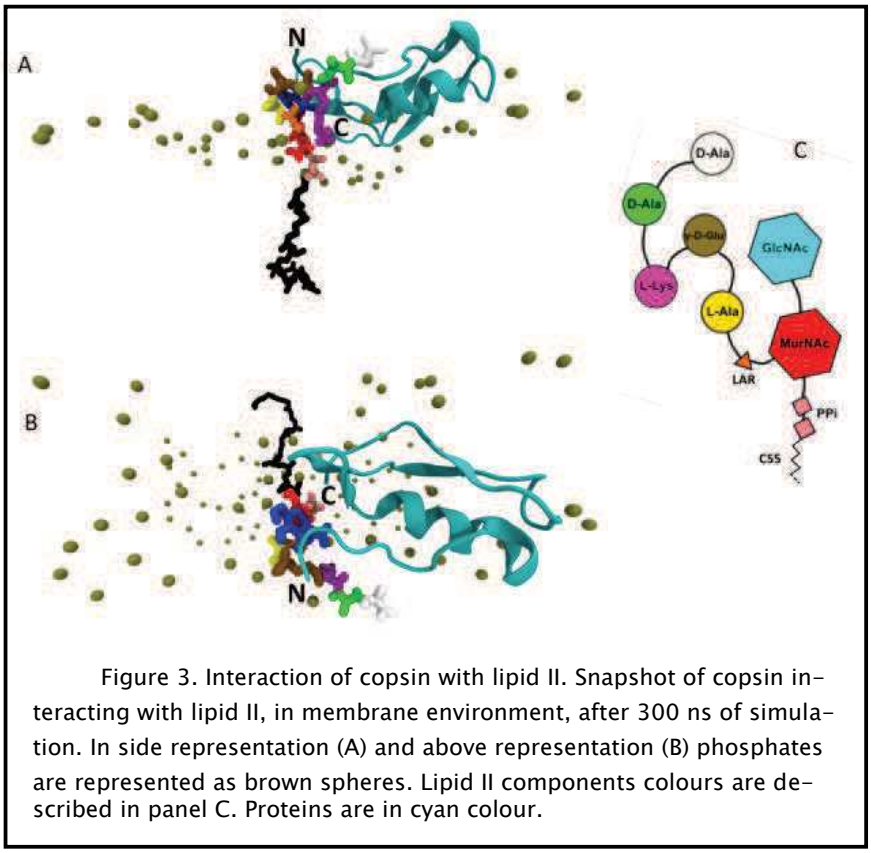


Figure 3. Interaction of copsin with lipid II. Snapshot of copsin interacting with lipid II, in membrane environment, after 300 ns of simulation. In side representation (A) and above representation (B) phosphates are represented as brown spheres. Lipid II components colours are described in panel C. Proteins are in cyan colour.

In all simulations, copsin and k-copsin binding to lipid II was irreversible. Taking into account the last 50 ns of the simulations, interactions were predominately restricted to the N-terminal tail (amino acids 1-8) and partially to the loop region between the α-helix and the β-strands (25 - 34) as well as to the C-terminus of copsin (**Figure 2 and Figure 3**). Introducing the four mutations of k-copsin, the binding pattern to lipid II changed, with contacts spanning the amino acids from position 1 to 34 and from position 41 to 49.

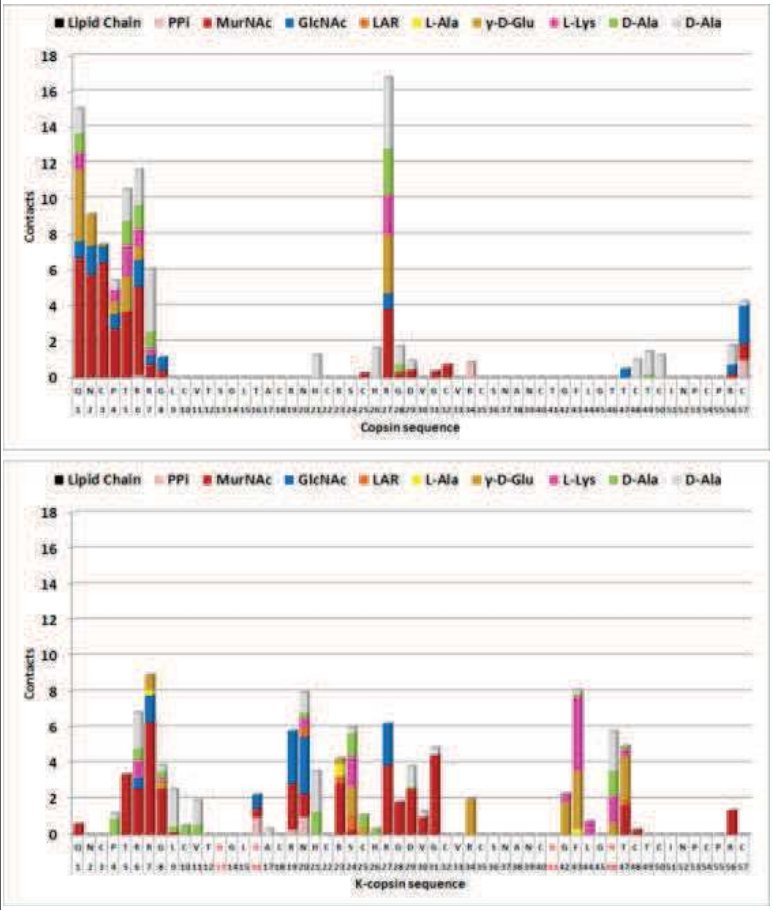


Figure 4. Number of contacts between copsin (top) or k-copsin (bottom) and lipid II components. Data are expressed as mean of contacts during the last 50 ns of the simulation, averaged for the four independent simulations, each with a different initial spatial orientation of copsin.

In the simulated membrane environment, copsin and k-copsin primarily interacted with the two sugar moieties and the four terminal amino acids (D-Ala-D-Ala-L-Lys-D-Glu) of the lipid II pentapeptide (**Figure 4 and Table 5**). Even if the most solvent-accessible sugar is known to be GlcNAc[27], the most stable interaction of copsin in both versions was with MurNAc. This is consistent with the previous results showing the ability of copsin to bind lipid I, comprising a MurNAc monosaccharide, but not to lipid III, having a GlcNAc monosaccharide[22]. This evidence is also consistent with the capability of copsin to inhibit nisin-induced pore formation, since MurNAc is spatially close to the pyrophosphates, known to be the main target for nisin[22,49].

Component	Copsin Wt	Mutated
GlcNAc[1]	8.81%	16.11%
MurNAc[2]	46.80%	34.97%
ALR[3]	0.01%	3.07%
L-Ala[4]	0.50%	0.70%
γ-D-Glu[4]	8.42%	9.57%
L-Lys[4]	3.61%	11.84%
D-Ala[4]	10.08%	5.47%
D-Ala[4]	20.28%	17.10%
PPI[5]	1.50%	1.17%

Table 5: Contacts expressed as percentage of total between the two copsin versions and the lipid II components. 1) N-acetyl glucosamine. 2) N-acetylmuramic acid 3) Acid Lactic residue, 4) Amino acids, 5) Pyrophosphates.

In the final part of the simulations, a heavy displacement of lipids was evident around residues of β-loop (residues 41-46) and of the N-terminal-helix part, where the four mutated residues are located (**Figure 4**). Interestingly, the mutated residue S13H, having the most positive effect on copsin activity, deeply inserts itself in the membrane in the presence of lipid II. Moreover, the same copsin region has been shown to be important for the binding of other defensins to lipid II, such as plectasin[17]. In particular,

residues of the loop, connecting the two β-strands, and the N-terminal-helix part of plectasin interact with DPC micelles but are not involved in lipid II binding[17]. Mutations occurring in such peptide region significantly increased the plectasin and micasin activity against *S. aureus* (**Figure 5**)[21,50]. This suggests that the improved activity of k-copsin is firstly driven by membrane interactions and secondly by a change in the lipid II binding pattern.

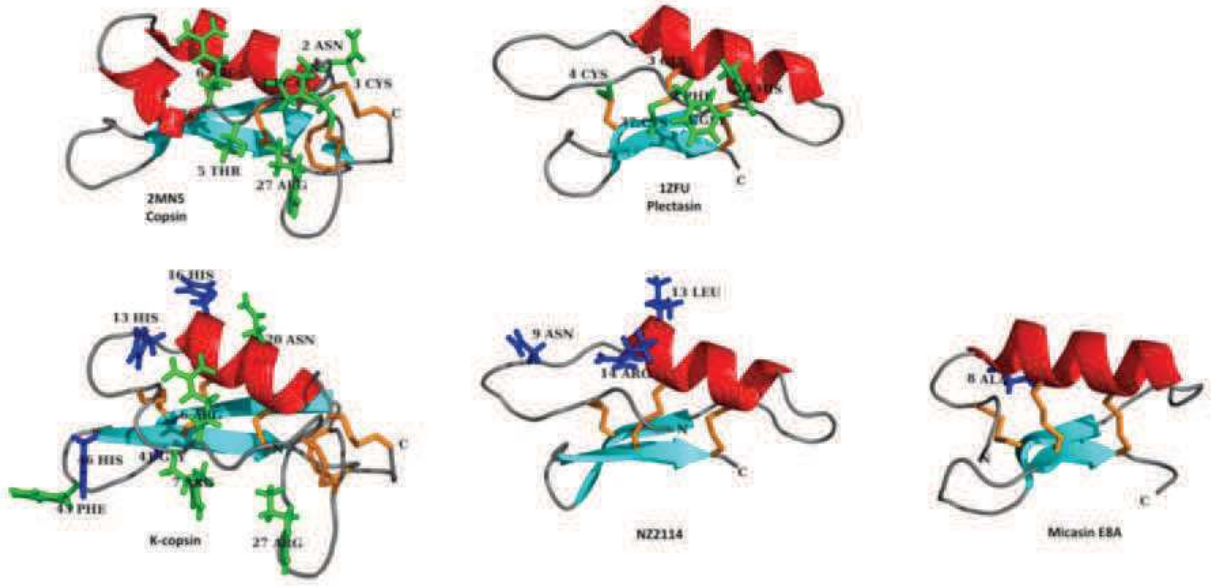


Figure 5. Comparison between copsin and other lipid II-binding peptides. Plectasin, Eurocin, and Micasin are fungal defensins consisting of a CSαβ fold interconnected by three disulphide bonds. Secondary structures are represented as cartoons. Green: amino acids proposed to interact with lipid II based on this work (copsin and k-copsin) or NMR data (plectasin)[17]. Yellow: putative amino acids involved in lipid II interactions, based on structural conservation with plectasin, Blue: amino acid substitution sites known to improve activity of represented peptides.

Chapter 5

Concluding remarks

Naturally derived fungal defensins commonly reveal MIC in the mid to low µg/ml range, values which are often not sufficient for an effective treatment of microbial infections (copsin, plectasin, micasin)[17,20,22]. Therefore, efficient screening strategies are needed to produce defensins with an improved *in-vitro* activity. Here, we present a new workflow, which combines the Golden Gate cloning with microscale expression in *P. pastoris*, for high-throughput production and selection of peptide mutants. This method is particularly suitable for cysteine-rich and secreted proteins and it allowed us to generate a copsin variant, named k-copsin, with enhanced activity against *B. subtilis* and *S. aureus*. However, the lack of improvement against other Gram positive bacteria such as Enterococci, demonstrates the difficulty of making fungal defensins effective towards a broader range of bacterial species.

MD simulations in a membrane lipid II environment allowed us to demonstrate how the activity improvement does not primarily derive from increased lipid II affinity, but rather from a different and stronger interaction with membranes. Such evidence supports the hypothesis that copsin inhibits the cell wall biosynthesis anchoring the lipid II to the membrane, with consequent prevention of processing and recycling. The protonation state of histidine residues may play a fundamental role in such interaction. After interaction with the lipid hydrophobic tails, protonated histidines could deprotonate and insert themselves deeply into the membrane.

In silico simulations also revealed a rather dispersed interaction pattern of copsin and k-copsin with lipid II. Copsin interacts with the two sugar moieties and the pentapeptide side chain whereas the glycopeptide vancomycin primarily binds to the two terminal alanine residues. Such a distributed binding pattern may be a way to counteract lipid II modifications and thus to circumvent the emergence of resistance. In agreement, copsin has potent activity against vancomycin-resistant Enterocooci with a Lac-Ala terminus instead of Ala-Ala[22]. The primary interaction side of copsin resulted to be the MurNAc residue of lipid II.

Even if MurNAc modifications have been identified in pathogenic strains[51], to our knowledge no direct evidence of AMP-resistance involving MurNAc has been reported so far. Together with enormous stability, these features make fungal defensins promising candidates for the development of new antibiotics less prone to bacterial resistance.

References

1. Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* **24,** 1551–1557 (2006).
2. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415,** 389–395 (2002).
3. Lehrer, R. I. Primate defensins. *Nat. Rev. Microbiol.* **2,** 727–738 (2004).
4. Yeaman, M. R. & Yount, N. Y. Unifying themes in host defence effector polypeptides. *Nat. Rev. Microbiol.* **5,** 727–740 (2007).
5. Mansour, S. C., Pena, O. M. & Hancock, R. E. W. Host defense peptides: front-line immunomodulators. *Trends Immunol.* **35,** 443–450 (2014).
6. Tu, J. *et al.* Molecular Evolutionary Analysis of β-Defensin Peptides in Vertebrates. *Evol. Bioinforma. Online* **11,** 105–114 (2015).
7. Dias, R. de O. & Franco, O. L. Cysteine-stabilized αβ defensins: From a common fold to antibacterial activity. *Peptides* **72,** 64–72 (2015).
8. Thomma, B. P. H. J., Cammue, B. P. A. & Thevissen, K. Plant defensins. *Planta* **216,** 193–202 (2002).
9. Terras, F. R. *et al.* Analysis of two novel classes of plant antifungal proteins from radish (Raphanus sativus L.) seeds. *J. Biol. Chem.* **267,** 15301–15309 (1992).
10. Zhu, S. Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSalphabeta defensins. *Mol. Immunol.* **45,** 828–838 (2008).
11. Schmitt, P., Rosa, R. D. & Destoumieux-Garzón, D. An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes. *Biochim. Biophys. Acta BBA - Biomembr.* **1858,** 958–970 (2016).
12. Landon, C., Sodano, P., Hetru, C., Hoffmann, J. & Ptak, M. Solution structure of drosomycin, the first inducible antifungal protein from insects. *Protein Sci. Publ. Protein Soc.* **6,** 1878–1884 (1997).
13. Wilmes, M. & Sahl, H.-G. Defensin-based anti-infective strategies. *Int. J. Med. Microbiol. IJMM* **304,** 93–99 (2014).
14. Brogden, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* **3,** 238–250 (2005).
15. Lee, J., Jung, S. W. & Cho, A. E. Molecular Insights into the Adsorption Mechanism of Human β-Defensin-3 on Bacterial Membranes. *Langmuir ACS J. Surf. Colloids* **32,** 1782–1790 (2016).
16. Breukink, E. & de Kruijff, B. Lipid II as a target for antibiotics. *Nat. Rev. Drug Discov.* **5,** 321–332 (2006).
17. Schneider, T. *et al.* Plectasin, a fungal defensin, targets the bacterial cell wall precursor Lipid II. *Science* **328,** 1168–1172 (2010).
18. Oeemig, J. S. *et al.* Eurocin, a new fungal defensin: structure, lipid binding, and its mode of action. *J. Biol. Chem.* **287,** 42361–42372 (2012).

19. Sass, V. *et al.* Human β-Defensin 3 Inhibits Cell Wall Biosynthesis in Staphylococci. *Infect. Immun.* **78,** 2793–2800 (2010).

20. Zhu, S., Gao, B., Harvey, P. J. & Craik, D. J. Dermatophytic defensin with antiinfective potential. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8495–8500 (2012).

21. Wu, J., Gao, B. & Zhu, S. Single-point mutation-mediated local amphipathic adjustment dramatically enhances antibacterial activity of a fungal defensin. *FASEB J.* **30,** 2602–2614 (2016).

22. Essig, A. *et al.* Copsin, a Novel Peptide-based Fungal Antibiotic Interfering with the Peptidoglycan Synthesis. *J. Biol. Chem.* **289,** 34953–34964 (2014).

23. *Clinical and Laboratory Standards Institute CLSI. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically.* **Ninth Edition M07-A9,** (2012).

24. Wu, S. & Letchworth, G. J. High efficiency transformation by electroporation of Pichia pastoris pretreated with lithium acetate and dithiothreitol. *BioTechniques* **36,** 152–154 (2004).

25. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* **3,** e3647 (2008).

26. Kombrink, A. Heterologous production of fungal effectors in Pichia pastoris. *Methods Mol. Biol. Clifton NJ* **835,** 209–217 (2012).

27. Chugunov, A. *et al.* Lipid-II forms potential 'landing terrain' for lantibiotics in simulated bacterial membrane. *Sci. Rep.* **3,** 1678 (2013).

28. Sapay, N. & Tieleman, D. P. Combination of the CHARMM27 force field with united-atom lipid force fields. *J. Comput. Chem.* **32,** 1400–1410 (2011).

29. Berger, O., Edholm, O. & Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **72,** 2002–2013 (1997).

30. Schrödinger, L. *The PyMOL Molecular Graphics System*. (2015).

31. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinforma. Oxf. Engl.* **29,** 845–854 (2013).

32. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).

33. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **38,** 236–240 (1999).

34. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 14101 (2007).

35. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. van, DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

36. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

37. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).

38. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).

39. Allen, W. J., Lemkul, J. A. & Bevan, D. R. GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. *J. Comput. Chem.* **30,** 1952–1958 (2009).

40. Gapsys, V., de Groot, B. L. & Briones, R. Computational analysis of local membrane properties. *J. Comput. Aided Mol. Des.* **27,** 845–858 (2013).

41. Slepecky, R. A. & Hemphill, H. E. in *The Prokaryotes* (eds. Dr, M. D. P., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 530–562 (Springer US, 2006).

42. Moayeri, M., Leppla, S. H., Vrentas, C., Pomerantsev, A. P. & Liu, S. Anthrax Pathogenesis. *Annu. Rev. Microbiol.* **69,** 185–208 (2015).

43. Patiño-Navarrete, R. & Sanchis, V. Evolutionary processes and environmental factors underlying the genetic diversity and lifestyles of Bacillus cereus group bacteria. *Res. Microbiol.* (2016). doi:10.1016/j.resmic.2016.07.002

44. Pilo, P. & Frey, J. Bacillus anthracis: molecular taxonomy, population genetics, phylogeny and patho-evolution. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **11,** 1218–1224 (2011).

45. Nawrocki, K. L., Crispell, E. K. & McBride, S. M. Antimicrobial Peptide Resistance Mechanisms of Gram-Positive Bacteria. *Antibiotics* **3,** 461–492 (2014).

46. Malik, E., Dennison, S. R., Harris, F. & Phoenix, D. A. pH Dependent Antimicrobial Peptides and Proteins, Their Mechanisms of Action and Potential as Therapeutic Agents. *Pharmaceuticals* **9,** 67 (2016).

47. Jia, Z., O'Mara, M. L., Zuegg, J., Cooper, M. A. & Mark, A. E. The effect of environment on the recognition and binding of vancomycin to native and resistant forms of lipid II. *Biophys. J.* **101,** 2684–2692 (2011).

48. Hasper, H. E. *et al.* An Alternative Bactericidal Mechanism of Action for Lantibiotic Peptides That Target Lipid II. *Science* **313,** 1636–1637 (2006).

49. Bonev, B. B., Breukink, E., Swiezewska, E., Kruijff, B. D. & Watts, A. Targeting extracellular pyrophosphates underpins the high selectivity of nisin. *FASEB J.* **18,** 1862–1869 (2004).

50. Xiong, Y. Q. *et al.* Efficacy of NZ2114, a novel plectasin-derived cationic antimicrobial peptide antibiotic, in experimental endocarditis due to methicillin-resistant Staphylococcus aureus. *Antimicrob. Agents Chemother.* **55,** 5325–5330 (2011).

51. Vollmer, W. Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol. Rev.* **32,** 287–306 (2008).

Supplementary material

```
1   ATG AAA CTT AGT ACA TCA CTT TTG GCT ATT GTC GCA GTC GCC TCT ACC TTT ATT GGA AAC
    M   K   L   S   T   S   L   L   A   I   V   A   V   A   S   T   F   I   G   N
    >------------------------------- Signal peptide ----------------------------

61  GCA TTG TCA GCA ACC ACC GTC CCT GGA TGT TTT GCT GAA TGC ATT GAT AAG GCT GCC GTT
    A   L   S   A   T   T   V   P   G   C   F   A   E   C   I   D   K   A   A   V
    ----------< >-------------------- Pro peptide --------------------------------

121 GCT GTC AAC TGT GCA GCT GGA GAT ATT GAC TGC TTG CAA GCC TCT TCC CAG TTC GCA ACT
    A   V   N   C   A   A   G   D   I   D   C   L   Q   A   S   S   Q   F   A   T
    ---------------------------------------------------------------------------

181 ATC GTT TCT GAG TGT GTC GCT ACT TCC GAC TGC ACA GCC CTT TCA CCA GGA TCA GCA AGT
    I   V   S   E   C   V   A   T   S   D   C   T   A   L   S   P   G   S   A   S
    ---------------------------------------------------------------------------

241 GAT GCT GAC AGT ATT AAC AAG ACT TTT AAC ATC TTG TCT GGT CTT GGT TTT ATT GAT GAA
    D   A   D   S   I   N   K   T   F   N   I   L   S   G   L   G   F   I   D   E
    ---------------------------------------------------------------------------

301 GCT GAC GCC TTC AGT GCC GCA GAT GTT CCT GAA GAG AGA GAC TTG ACA GGT CTT GGA AGA
    A   D   A   F   S   A   A   D   V   P   E   E   R   D   L   T   G   L   G   R
    ---------------------------------------------------------------------------

361 GTT TTG CCA GTC GAG AAA AGA CAA AAT TGT CCT ACC AGA AGA GGT TTG TGT GTT ACC TCC
    V   L   P   V   E   K   R   Q   N   C   P   T   R   R   G   L   C   V   T   S
    --------------------------< >-----------------------------------------------

421 GGA CTT ACT GCT TGT AGA AAC CAT TGT AGA TCA TGC CAC AGA GGA GAT GTT GGA TGT GTC
    G   L   T   A   C   R   N   H   C   R   S   C   H   R   G   D   V   G   C   V
    --------------------------------- Mature copsin ------------------------------

481 AGA TGC TCT AAT GCT CAA TGT ACT GGT TTC TTG GGA ACT ACT TGC ACC TGT ATC AAT CCT
    R   C   S   N   A   Q   C   T   G   F   L   G   T   T   C   T   C   I   N   P
    ---------------------------------------------------------------------------

541 TGT CCT AGA TGC TAA
    C   P   R   C   -
    -------------<
```

Supplementary Figure 1: Codon-optimized prepro-copsin

Supplementary Table 1: Primers used for initial mutant generation

Position	Amino acid	Change	Cloning method	Fw Primer/DNA fragment (5'-3')	Rw Primer/DNA fragment (5'-3')
1	Q	G	Synthesized (GenScript)		
		R	GoldenGate	AAGACGAAATTGTCCTAC-CAGAAGAGGTTTG	CACACAAACCTCTTCTGGTAG-GACAATTTCG
		S	GoldenGate	AAGATCGAATTGTCCTAC-CAGAAGAGGTTTG	CACACAAACCTCTTCTGGTAG-GACAATTCGA
2	N	F	GoldenGate	AAGACAATTCTGTCCTAC-CAGAAGAGGTTTG	CACACAAACCTCTTCTGGTAG-GACAGAATTG
		H	GoldenGate	AAGACAACACTGTCCTAC-CAGAAGAGGTTTG	CACACAAACCTCTTCTGGTAG-GACAGTGTTG
		R	GoldenGate	AAGACAACGATGTCCTAC-CAGAAGAGGTTTG	CACACAAACCTCTTCTGGTAG-GACATCGTTG
3	C				
4	P	G	GoldenGate	AAGACAAAATTGTGGTAC-CAGAAGAGGTTTG	CACACAAACCTCTTCTGGTAC-CACAATTTTG
		S	GoldenGate	AAGACAAAATTGTTCAAC-CAGAAGAGGTTTG	CACACAAACCTCTTCTGGTT-GAACAATTTTG

		R	GoldenGate	AAGACAAAATTGTCGAAC-CAGAAGAGGTTTG	CACACAAAC-CTCTTCTGGTTCGACAATTTTG
		W	GoldenGate	AAGACAAAATTGTTGGAC-CAGAAGAGGTTTG	CACACAAAC-CTCTTCTGGTCCAACAATTTTG
		H	GoldenGate	AAGACAAAATTGTCACAC-CAGAAGAGGTTTG	CACACAAAC-CTCTTCTGGTGTGACAATTTTG
5	T	G	GoldenGate	AAGACAAAATTGTCCTGG-TAGAAGAGGTTTG	CACACAAACCTCTTCTACCAG-GACAATTTTG
		S	GoldenGate	AAGACAAAATTGTCCTTCAA-GAAGAGGTTTG	CACACAAACCTCTTCTT-GAAGGACAATTTTG
		R	GoldenGate	AAGACAAAATTGTCCTCGAA-GAAGAGGTTTG	CACACAAACCTCTTCTTCGAG-GACAATTTTG
		W	GoldenGate	AAGACAAAATTGTCCTTGGA-GAAGAGGTTTG	CACACAAAC-CTCTTCTCCAAGGACAATTTTG
6	R	S	GoldenGate	AAGACAAAATTGTCCTAC-CTCAAGAGGTTTG	CACACAAACCTCTTGAGGTAG-GACAATTTTG
		H	GoldenGate	AAGACAAAATTGTCC-TACCCACAGAGGTTTG	CACACAAACCTCTGTGGGTAG-GACAATTTTG
7	R	S	GoldenGate	AAGACAAAATTGTCCTAC-CAGATCAGGTTTG	CACACAAACCTGATCTGGTAG-GACAATTTTG
		G	GoldenGate	AAGACAAAATTGTCCTAC-CAGAGGTGGTTTG	CACACAAACCACCTCTGGTAG-GACAATTTTG
		H	GoldenGate	AAGACAAAATTGTCCTACCAGA-CACGGTTTG	CACACAAACCGTGTCTGGTAG-GACAATTTTG
		W	GoldenGate	AAGACAAAATTGTCCTAC-CAGATGGGGTTTG	CACACAAACCCCATCTGGTAG-GACAATTTTG
		K	GoldenGate	AAGACAAAATTGTCCTACCAGA-GACGGTTTG	CACACAAACCGTCTCTGGTAG-GACAATTTTG
8	G	F	GoldenGate	AAGACAAAATTGTCCTAC-CAGAAGATTCTTG	CACACAAGAATCTTCTGGTAG-GACAATTTTG
		S	GoldenGate	AAGACAAAATTGTCCTAC-CAGAAGATCATTG	CACACAATGATCTTCTGGTAG-GACAATTTTG
		R	GoldenGate	AAGACAAAATTGTCCTAC-CAGAAGACGATTG	CACACAATCGTCTTCTGGTAG-GACAATTTTG
		H	GoldenGate	AAGACAAAATTGTCCTAC-CAGAAGACACTTG	CACACAAGTGTCTTCTGGTAG-GACAATTTTG
9	L	G	GoldenGate	AAGACAAAATTGTCCTAC-CAGAAGAGGTGGT	CACAACCACCTCTTCTGGTAG-GACAATTTTG
		W	GoldenGate	AAGACAAAATTGTCCTAC-CAGAAGAGGTTGG	CACACCAACCTCTTCTGGTAG-GACAATTTTG
		D	GoldenGate	AAGACAAAATTGTCCTAC-CAGAAGAGGTGAC	CACAGTCACCTCTTCTGGTAG-GACAATTTTG
10	C				
11	V	W	GoldenGate	GTGTTGGACCTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCG-GAGGTCCA
		R	GoldenGate	GTGTCGGACCTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCG-GAGGTCCG
		G	GoldenGate	GTGTGGCACCTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCG-GAGGTGCC
		H	GoldenGate	GTGTCACACCTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCG-GAGGTGTG
		D	GoldenGate	GTGTGACACCTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCG-GAGGTGTC
12	T	N	GoldenGate	GTGTGTTAACTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCGGAG-TTAAC

		E	GoldenGate	GTGTGTTGAGTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCG-GACTCAAC
		H	GoldenGate	GTGTGTTCACTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCGGAG-TGAAC
		W	GoldenGate	GTGTGTTTGGTCCG-GACTTACTGCT	TACAAGCAGTAAGTCCGGAC-CAAAC
13	S	E	GoldenGate	GTGTGTTAC-CGAGGGACTTACTGCT	TACAAGCAGTAAGTCCCTCGG-TAAC
		G	GoldenGate	GTGTGTTACCGGTG-GACTTACTGCT	TACAAGCAGTAAGTCCACCGG-TAAC
		H	GoldenGate	GTGTGTTACCCAC-GGACTTACTGCT	TACAAGCAGTAAGTCCGTGGG-TAAC
14	G	R	GoldenGate	GTGTGTTAC-CTCCCGGCTTACTGCT	TACAAGCAGTAAGCCGG-GAGGTAAC
		S	GoldenGate	GTGTGTTAC-CTCCAGCCTTACTGCT	TACAAGCAGTAAGGCTG-GAGGTAAC
15	L	D	GoldenGate	GTGTGTTACCTCCGGAGA-TACTGCT	TACAAGCAGTATCTCCG-GAGGTAAC
		H	GoldenGate	GTGTGTTACCTCCGGACACAC-TGCT	TACAAGCAGTGTGTCCG-GAGGTAAC
		G	GoldenGate	GTGTGTTACCTCCG-GAGGGACTGCT	TACAAGCAGTCCCTCCG-GAGGTAAC
		F	GoldenGate	GTGTGTTACCTCCGGATTCAC-TGCT	TACAAGCAGTGAATCCG-GAGGTAAC
16	T	L	GoldenGate	GTGTGTTACCTCCG-GACTTCTGGCT	TACAAGCCAGAAGTCCG-GAGGTAAC
		H	GoldenGate	GTGTGTTACCTCCGGACTTCAC-GCT	TACAAGCGTGAAGTCCG-GAGGTAAC
		F	GoldenGate	GTGTGTTACCTCCG-GACTTTTCGCT	TACAAGCGAAAAGTCCG-GAGGTAAC
		G	GoldenGate	GTGTGTTACCTCCGGACTT-GGTGCT	TACAAGCACCAAGTCCG-GAGGTAAC
17	A	R	GoldenGate	GTGTGTTACCTCCG-GACTTACTCGT	TACAACGAGTAAGTCCG-GAGGTAAC
		H	GoldenGate	GTGTGTTACCTCCG-GACTTACTCAC	TACAGTGAGTAAGTCCG-GAGGTAAC
		F	GoldenGate	GTGTGTTACCTCCG-GACTTACTTTC	TACAGAAAGTAAGTCCG-GAGGTAAC
18	C				
19	R	H	Whole Plasmid	CACAACCATT-GTAGATCATGCCACAGAG	ACAAGCAGTAAGTCCG-GAGGTAAC
		V	Whole Plasmid	TACTGCTTGTAACAACCATT-GTAGATCATG	AGTCCGGAGGTAACACAC
		F	Whole Plasmid	TACTGCTTGTTTCAACCATT-GTAGATCATG	AGTCCGGAGGTAACACAC
20	N	R	Whole Plasmid	TGCTTGTAGACGTCATT-GTAGATCATGCCAC	GTAAGTCCGGAGGTAACAC
		H	Whole Plasmid	TGCTTGTAGACACCATT-GTAGATC	GTAAGTCCGGAGGTAACAC
		V	Whole Plasmid	TGCTTGTAGAGTCCATT-GTAGATCATGC	GTAAGTCCGGAGGTAACAC
21	H	R	Whole Plasmid	AGAAACCGTT-GTAGATCATGCCACAGAG	ACAAGCAGTAAGTCCG-GAGGTAAC
22	C				

23	R	L	Whole Plasmid	CTGTCATGCCACAGAGGA-GATGTTG	ACAATGGTTTCTACAAGCAG-TAAGTC
		H	Whole Plasmid	AAACCATTGTCAC-TCATGCCACAGAGG	CTACAAGCAGTAAGTCCG
24	S	A	Whole Plasmid	AGAGCATGCCACAGAGGA-GATGTTGG	ACAATGGTTTCTACAAGCAG-TAAGTC
		H	Whole Plasmid	CCATTGTAGACAC-TGCCACAGAGGAGATG	TTTCTACAAGCAGTAAGTCC
		F	Whole Plasmid	CCATTGTAGAT-TCTGCCACAGAGG	TTTCTACAAGCAGTAAGTCC
25	C				
26	H	K	Whole Plasmid	AGATCATGCAAGAGAGGA-GATGTTGG	ACAATGGTTTCTACAAGCAG-TAAGTC
27	R	G	Whole Plasmid	AGATCATGCCACGGTGGA-GATGTTGG	ACAATGGTTTCTACAAGCAG-TAAGTC
		K	Whole Plasmid	AGATCATGCCACAAGGGA-GATGTTGG	ACAATGGTTTCTACAAGCAG-TAAGTC
		F	Whole Plasmid	ATCATGCCACTTCGGAGATGTT-GGATG	CTACAATGGTTTCTACAAGC
		H	Whole Plasmid	ATCATGCCACCACGGAGATGTT-GGATG	CTACAATGGTTTCTACAAGC
28	G	Y	Whole Plasmid	AGATCATGCCACAGATAC-GATGTTGGATGTGTC	ACAATGGTTTCTACAAGCAG-TAAGTC
		K	Whole Plasmid	AGATCATGCCACAGAAAGGATGTTGGATGTGTC	ACAATGGTTTCTACAAGCAG-TAAGTC
		H	Whole Plasmid	ATGCCACAGACACGATGTT-GGATGTGTCAG	GATCTACAATGGTTTCTACAAG
		D	Whole Plasmid	ATGCCACAGAGACGATGTT-GGATG	GATCTACAATGGTTTCTACAAG
29	D	K	Whole Plasmid	AAGGTT-GGATGTGTCAGATGCTCTAATG	TCCTCTGTGG-CATGATCTACAATGG
		F	Whole Plasmid	TTCGTT-GGATGTGTCAGATGCTCTAATG	TCCTCTGTGG-CATGATCTACAATGG
		V	Whole Plasmid	CCACAGAGGAGTGGTTGGATGTG	CATGATCTACAATGGTTTCTAC
		H	Whole Plasmid	CCACAGAGGACACGTT-GGATGTGTC	CATGATCTACAATGGTTTCTAC
30	V	G	Whole Plasmid	GATGGTG-GATGTGTCAGATGCTCTAATG	TCCTCTGTGG-CATGATCTACAATGG
		F	Whole Plasmid	GATTTCG-GATGTGTCAGATGCTCTAATG	TCCTCTGTGG-CATGATCTACAATGG
		H	Whole Plasmid	CAGAGGAGATCAC-GGATGTGTCAGATG	TGGCATGATCTACAATGG
		R	Whole Plasmid	CAGAGGAGATCGTG-GATGTGTCAG	TGGCATGATCTACAATGG
31	G	R	Whole Plasmid	GATGTTAGGTGTGTCAGATGCTCTAATG	TCCTCTGTGG-CATGATCTACAATGG
		H	Whole Plasmid	AGGAGATGTTCAC-TGTGTCAGATGCTC	CTGTGGCATGATCTACAATG
		D	Whole Plasmid	AGGAGATGTT-GACTGTGTCAGATGC	CTGTGGCATGATCTACAATG
32	C				
33	V	R	Whole Plasmid	GATGTTGGATGTAGGA-GATGCTCTAATGCTC	TCCTCTGTGG-CATGATCTACAATGG
		H	Whole Plasmid	TGTT-	TCTCCTCTGTGGCATGATC

				GGATGTCACAGATGCTCTAATG	
		F	Whole Plasmid	TGTT-GGATGTTTCAGATGCTCTAATG	TCTCCTCTGTGGCATGATC
34	R	G	Whole Plasmid	GATGTTGGATGTGTCGGTT-GCTCTAATGCTC	TCCTCTGTGG-CATGATCTACAATGG
		H	Whole Plasmid	TGGATGTGTCCAC-TGCTCTAATGCTCAATG	ACATCTCCTCTGTGGCATG
		F	Whole Plasmid	TGGATGTGTCTTCTGCTCTAATGCTCAATG	ACATCTCCTCTGTGGCATG
35	C				
36	S	R	Whole Plasmid	AG-GAATGCTCAATGTACTGGTTTCTTGGG	GCATCTGACACATCCAACATCTCC
		H	Whole Plasmid	TGTCAGATGCCACAATGCTCAATGTAC	CATCCAACATCTCCTCTG
		F	Whole Plasmid	TGTCAGATGCTTCAATGCTCAATG	CATCCAACATCTCCTCTG
37	N	L	Whole Plasmid	TCTCTGGCTCAATGTACTGGTTTCTTGGG	GCATCTGACACATCCAACATCTCC
		H	Whole Plasmid	CAGATGCTCTCAC-GCTCAATGTACTG	ACACATCCAACATCTCCTC
38	A	G	Whole Plasmid	TCTAATGGTCAATGTACTGGTTTCTTGGG	GCATCTGACACATCCAACATCTCC
		H	Whole Plasmid	ATGCTCTAATCAC-CAATGTACTGGTTTC	CTGACACATCCAACATCTC
		F	Whole Plasmid	ATGCTCTAATTTCCAATGTACTGGTTTC	CTGACACATCCAACATCTC
39	Q	F	Whole Plasmid	TCTAATGCTTTCTG-TACTGGTTTCTTGGG	GCATCTGACACATCCAACATCTCC
		H	Whole Plasmid	CTCTAATGCTCACTG-TACTGGTTTC	CATCTGACACATCCAACATC
40	C				
41	T	G	GoldenGate	ATGTGGTGGTTTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAAGAAAC-CACC
		A	GoldenGate	ATGTGCTGGTTTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAAGAAAC-CAGC
		R	GoldenGate	ATGTAGGGGTTTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAA-GAAACCCCT
		S	GoldenGate	ATGTTCGGGTTTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAA-GAAACCCGA
		H	GoldenGate	ATGTCACGGTTTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAAGAAAC-CGTG
42	G	R	GoldenGate	ATGTACTAGGTTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAAGAAC-CTAGT
		W	GoldenGate	ATGTACTTGGTTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAAGAAC-CAAGT
43	F	W	GoldenGate	ATGTACTGGTTGGTTGGGAAC-TACT	TGCAAGTAGTTCCCAACCAAC-CAGT
		R	GoldenGate	ATGTACTGGTAGGTTGGGAAC-TACT	TGCAAGTAGTTCCCAACCTAC-CAGT
		H	GoldenGate	ATGTACTGGTCACTTGGGAAC-TACT	TGCAAGTAGTTCCCAAGTGAC-CAGT
		V	GoldenGate	ATGTACTGGTGTCTTGGGAAC-TACT	TGCAAGTAGTTCCCAAGACAC-CAGT

44	L	W	GoldenGate	ATGTACTGGTTTCTGGGGAAC-TACT	TGCAAGTAGTTCCCCAGAAAC-CAGT
45	G	W	GoldenGate	ATGTACTGGTTTCTTGTGGAC-TACT	TGCAAGTAGTCCACAAGAAAC-CAGT
		R	GoldenGate	ATGTACTGGTTTCTTGCGAAC-TACT	TGCAAGTAGTTCGCAAGAAAC-CAGT
		H	GoldenGate	ATGTACTGGTTTCTTGCACAC-TACT	TGCAAGTAGTGTGCAAGAAAC-CAGT
		T	GoldenGate	ATGTACTGGTTTCTTGAC-TACTACT	TGCAAGTAGTAGTCAAGAAAC-CAGT
46	T	M	GoldenGate	ATGTACTGGTTTCTTGG-GAATGACT	TGCAAGTCATTCCCAAGAAAC-CAGT
		S	GoldenGate	ATGTACTGGTTTCTTGG-GATCGACT	TGCAAGTCGATCCCAAGAAAC-CAGT
		R	GoldenGate	ATGTACTGGTTTCTTGG-GAAGGACT	TGCAAGTCCTTCCCAAGAAAC-CAGT
		G	GoldenGate	ATGTACTGGTTTCTTGGGAG-GAACT	TGCAAGTTCCTCCCAAGAAAC-CAGT
		W	GoldenGate	ATGTACTGGTTTCTTGG-GATGGACT	TGCAAGTCCATCCCAAGAAAC-CAGT
		H	GoldenGate	ATGTACTGGTTTCTTGG-GACACACT	TGCAAGTGTGTCCCAAGAAAC-CAGT
47	T	G	GoldenGate	ATGTACTGGTTTCTTGG-GAACTGGA	TGCATCCAGTTCCCAAGAAAC-CAGT
		W	GoldenGate	ATGTACTGGTTTCTTGG-GAACTTGG	TGCACCAAGTTCCCAAGAAAC-CAGT
		R	GoldenGate	ATGTACTGGTTTCTTGG-GAACTAGG	TGCACCTAGTTCCCAAGAAAC-CAGT
		H	GoldenGate	ATGTACTGGTTTCTTGG-GAACTCAC	TGCAGTGAGTTCCCAAGAAAC-CAGT
48	C				
49	T	K	Whole Plasmid	TGCAAGTGTATCAATCCTTGTCC	AGTAGTTCCCAAGAAACCAG-TACATTG
		F	Whole Plasmid	TGCTTCTGTATCAATCCTTGTCC	AGTAGTTCCCAAGAAACCAG-TACATTG
		H	Whole Plasmid	AACTACTTGCCACTG-TATCAATCCTTG	CCCAAGAAACCAGTACATTG
		V	Whole Plasmid	AACTACTTGCGTCTG-TATCAATCCTTG	CCCAAGAAACCAGTACATTG
		D	Whole Plasmid	AACTACTTGCGACTG-TATCAATCCTTG	CCCAAGAAACCAGTACATTG
50	C				
51	I	Y	Whole Plasmid	TGCACCTGTTACAATCCTT-GTCCTAG	AGTAGTTCCCAAGAAACCAG-TACATTG
		R	Whole Plasmid	TGCACCTGTAGGAATCCTT-GTCCTAG	AGTAGTTCCCAAGAAACCAG-TACATTG
		H	Whole Plasmid	TTGCACCTGTCACAATCCTTGTC	GTAGTTCCCAAGAAACCAG
		D	Whole Plasmid	TTGCACCTGTGACAATCCTTGTC	GTAGTTCCCAAGAAACCAG
		T	Whole Plasmid	TTGCACCTGTACGAATCCTT-GTCCTAG	GTAGTTCCCAAGAAACCAG
52	N	K	Whole Plasmid	TGCACCTGTATCAAGCCTT-GTCCTAGATG	AGTAGTTCCCAAGAAACCAG-TACATTG
		H	Whole Plasmid	CACCTGTATCCACCCTT-GTCCTAG	CAAGTAGTTCCCAAGAAAC

		D	Whole Plasmid	CACCTGTATCGACCCTT-GTCCTAG	CAAGTAGTTCCCAAGAAAC
		W	Whole Plasmid	CACCTGTATCTGGCCTT-GTCCTAGATG	CAAGTAGTTCCCAAGAAAC
		T	Whole Plasmid	CACCTGTATCACGCCTT-GTCCTAGATG	CAAGTAGTTCCCAAGAAAC
53	P	V	Whole Plasmid	CTGTATCAATGTGTGTCCTA-GATGCTAAGTC	GTGCAAGTAGTTCCCAAG
		H	Whole Plasmid	CTGTATCAATCACTGTCCTA-GATGCTAAG	GTGCAAGTAGTTCCCAAG
54	C				
55	P	R	Whole Plasmid	CCTTGTAGGAGATGCTAAGTCG	ATTGATACAGGTGCAAGTAG-TTCC
		V	Whole Plasmid	CAATCCTTGTGTGA-GATGCTAAGTCG	ATACAGGTGCAAGTAGTTC
		F	Whole Plasmid	CAATCCTT-GTTTCAGATGCTAAGTCG	ATACAGGTGCAAGTAGTTC
56	R	S	Whole Plasmid	CCTTGTCCTTCGTGCTAAGTCG	ATTGATACAGGTGCAAGTAG-TTCC
		D	Whole Plasmid	TCCTT-GTCCTGACTGCTAAGTCGAC	TTGATACAGGTGCAAGTAG
		V	Whole Plasmid	TCCTT-GTCCTGTGTGCTAAGTCGACC	TTGATACAGGTGCAAGTAG
		H	Whole Plasmid	TCCTTGTCCTCAC-TGCTAAGTCGAC	TTGATACAGGTGCAAGTAG
57	C				

A

QNCPTRRGLCVTSGLTACRNHCRSCHRGDVGCVRCSNAQCTGFLGTTCTCINPCPRC

n-terminus n-loop c-loop

B

CGAGACGAAAATTGTCCTACCAGAAGAGGTTTCGTCTCGTGTGTTACCTCCGGACTTACTGCTTGTAGAAACCAT
TGTAGATCATGCCACAGAGGAGATGTTGGATGTGTCAGATGCTCTAATGCTCAATGTACTGGTTTCTTGGGAACT
ACTTGCACCTGTATCAATCCTTGTCCTAGATGCTAA

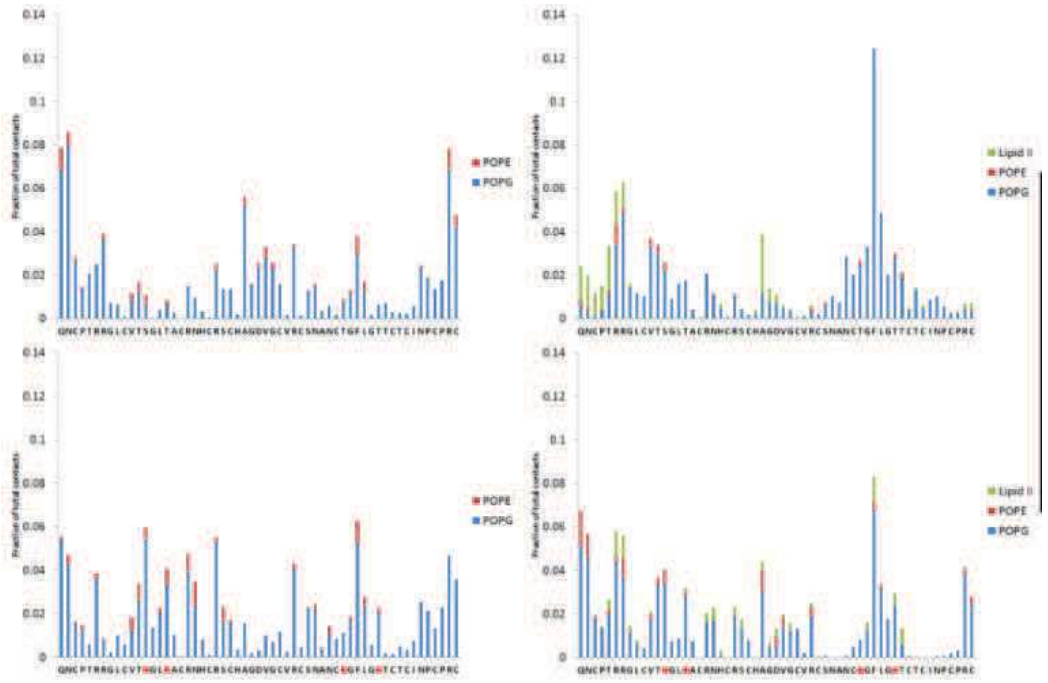
C

CAAAATTGTCCTACCAGAAGAGGTTTGTGTGGAGACGTTACCTCCGGACTTACTGCCGTCTCTTGTAGAAACCAT
TGTAGATCATGCCACAGAGGAGATGTTGGATGTGTCAGATGCTCTAATGCTCAATGTACTGGTTTCTTGGGAACT
ACTTGCACCTGTATCAATCCTTGTCCTAGATGCTAA

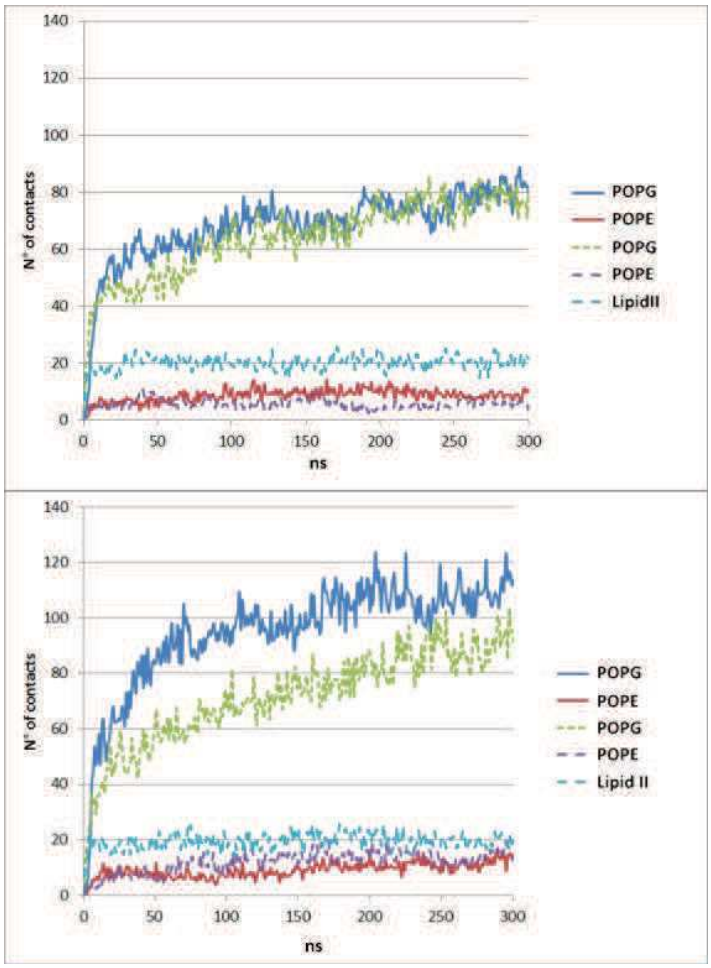
D

CAAAATTGTCCTACCAGAAGAGGTTTGTGTGTTACCTCCGGACTTACTGCTTGTAGAAACCATTGTAGATCATGC
CACAGAGGAGATGTTGGATGTGTCAGATGCTCTAATGCTCAATGTAGAGACGCTGGTTTCTTGGGAACTACCGTC
TCTTGCACCTGTATCAATCCTTGTCCTAGATGCTAA

Supplementary Figure 2: Golden Gate recipient expression constructs. (**A**) Amino acid sequence of mature copsin. Regions, which were replaced during the Golden Gate cloning, are indicated in orange. The α-helix is indicated in blue and the β-strands in green. (**B–D**) Nucleotide sequence of codon optimized mature copsin. Introduced BsmBI recognition sites are indicated in red and bold. Regions replaced during Golden Gate cloning are shown in orange corresponding to the amino acid sequence (**B**: N-terminus; **C**: n-loop; **D**: c-loop).



Supplementary Figure 3: Molecular dynamics simulations. Number of contacts expressed as fraction of total between membrane components and copsin in wt (top) and k versions (bottom). On the left only membrane, on the right membrane with lipid II. Mutated residues of k-copsin are marked in red. Initial pyroglutamate was named Q.



Supplementary Figure 4: Molecular dynamics simulations. Number of contacts in function of time between peptides and membrane components, averaged for the four independent simulations. Continuous line: POPG/POPE membrane. Dashed lines: POPG/POPE membrane with one lipid II molecule.

Supplementary Table 2: Primers used for improved peptide versions.

Mutations	Cloning method	Template	Fw Primer/DNA fragment (5'-3')	Rw Primer/DNA fragment (5'-3')
T41G	Golden-Gate	pPICZA-optCopsin_c-loop	ATGTGGTGGTTTCTTGGGAACTACT	TGCAAGTAGTTCCCAA-GAAACCACC
T41G, T46H	Golden-Gate	pPICZA-optCopsin_c-loop	ATGTGGTGGTTTCTTGGGACACACT	TGCAAGTGTGTCCCAA-GAAACCACC
T41G, S13H	Whole plasmid	pPICZA-optCopsin_T41G	GTGTGTTACCCACGGACTTACTG	AAACCTCTTCTGGTAGGAC
T41G, T16H	Whole plasmid	pPICZA-optCopsin_T41G	CTCCGGACTTCACGCTT-GTAGAAACCATTG	GTAACACACAAACCTCTTC
T41G, T46H, S13H	Whole plasmid	pPICZA-optCopsin_T41G,T46H	GTGTGTTACCCACGGACTTACTG	AAACCTCTTCTGGTAGGAC
T41G, T46H, T16H	Whole plasmid	pPICZA-optCopsin_T41G,T46H	CTCCGGACTTCACGCTT-GTAGAAACCATTG	GTAACACACAAACCTCTTC
T41G, T46H, S13H, T16H	Whole plasmid	pPICZA-optCopsin_T41G,T46H	GGACTTCACGCTTGTAGAAAC-CATTG	GTGGGTAACACACAAAC-CTCTTC

Acknowledgments

The authors would like to thank Tim Keys and Markus Künzler for their support and helpful discussions.

Conclusions

Now-a-days, the prediction of antimicrobial molecules represents a great challenge. Actually bioinformatics tools designed for this scope have an insufficient accuracy[1]. Difficulties arises from the AMP variety, their physical-chemical properties, mechanisms of action and, consequently, spectrum of activity. Moreover, single or few amino acids substitutions can dramatically influence the peptide activity, in some case completely changing their bacterial targets. Since structural data could be more informative than the simple sequence data, we aimed to improve the prediction of defensin-like peptide folding[2-4].

In any case, the possibility to miss the goal in such a jungle of molecules is really high, because of the production of peptides active against bacterial strains with no medical, veterinary or agronomical interest or having a different role *in vivo*. Functional studies on organisms naturally expressing specie-specific AMPs by real-time PCR, western blot or other assays, can give useful hints, but could not be sufficient.

Difficulties can also arise from the production and purification of AMPs. Chemical peptide synthesis becomes the more expensive and complex the longer are the peptides. Moreover, finding a good protocol for post-translational modifications, for example disulfide oxidation, can be tricky and time consuming[5]. This hampers also the heterologous peptide expression, especially when the natural AMP folding is guided by specific and controlled processes[6,7]. The use of recombinant systems other than bacteria has been sometimes successful, as demonstrated for CgDef[8] and Snakin-1[9], but decades can pass between the peptide discovery and its recombinant production. Therefore, most often it is necessary to screen many sequences before retrieving a bioactive candidate molecule[10].

According to WHO, microbial resistance is one of the most challenging and threatening problem for the future, both for human health and economic reasons[11]. Despite the difficulties in the development of new peptide-based antibiotics, several research groups in the world are working on this topic and, starting from 2002, over 10000 papers per year have been recorded in PubMed[12]. Actually, there is a great difference in the number of AMPs claimed to be potent antibacterial drug candidates and the number of AMPs entering the clinical trials. Most of the times, degradation or toxicity *in vivo* limit the the use of AMPs to topical treatments[13].

For all these reasons, novel and more effective investigation paths should be used to explore the AMPs jungle, to finally get its hidden treasures.

References

1. Torrent, M., Nogués, M. V. & Boix, E. Discovering new in silico tools for antimicrobial peptide prediction. *Curr. Drug Targets* **13,** 1148–1157 (2012).

2. Taboureau, O. in *Antimicrobial Peptides* (eds. Giuliani, A. & Rinaldi, A. C.) 77–86 (Humana Press, 2010).

3. Cherkasov, A. *et al.* QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **57,** 4977–5010 (2014).

4. Franzoi, M., Sturlese, M., Bellanda, M. & Mammi, S. A molecular dynamics strategy for CSαβ peptides disulfide-assisted model refinement. *J. Biomol. Struct. Dyn.* **0,** 1–25 (2016).

5. Bulaj, G. & Olivera, B. M. Folding of Conotoxins: Formation of the Native Disulfide Bridges During Chemical Synthesis and Biosynthesis of Conus Peptides. *Antioxid. Redox Signal.* **10,** 141–156 (2007).

6. Wu, Z. *et al.* Impact of pro segments on the folding and function of human neutrophil alpha-defensins. *J. Mol. Biol.* **368,** 537–549 (2007).

7. Pazgier, M. & Lubkowski, J. Expression and purification of recombinant human α-defensins in Escherichia coli. *Protein Expr. Purif.* **49,** 1–8 (2006).

8. Gueguen, Y. *et al.* Characterization of a defensin from the oyster Crassostrea gigas. Recombinant production, folding, solution structure, antimicrobial activities, and gene expression. *J. Biol. Chem.* **281,** 313–323 (2006).

9. Kuddus, M. R. *et al.* Expression, purification and characterization of the recombinant cysteine-rich antimicrobial peptide snakin-1 in Pichia pastoris. *Protein Expr. Purif.* **122,** 15–22 (2016).

10. Elhag, O. *et al.* Screening, Expression, Purification and Functional Characterization of Novel Antimicrobial Peptide Genes from Hermetia illucens (L.). *PLOS ONE* **12,** e0169582 (2017).

11. WHO | Antimicrobial resistance. *WHO* Available at: http://www.who.int/mediacentre/factsheets/fs194/en/. (Accessed: 11th January 2017)

12. pubmeddev. Home - PubMed - NCBI. Available at: https://www.ncbi.nlm.nih.gov/pubmed. (Accessed: 11th January 2017)

13. Kang, H.-K., Kim, C., Seo, C. H. & Park, Y. The therapeutic applications of antimicrobial peptides (AMPs): a patent review. *J. Microbiol.* **55,** 1–12 (2017).

Appendix

Other elements of the mussel innate immunity

Franzoi M.[1], Rosani U.[1], Sharma N.[2], Domeneghetti S.[1], Pretto T.[3], Venier P.[1]

[1]Department of Biology, University of Padova , Via Ugo Bassi 58/B, 35131 Padova, Italy.
[2]Sgt University Gurgaon, Gurgaon-Badli Road, 122505, Budhera, India.
[3]Istituto Zooprofilattico Sperimentale delle Venezie, Viale dell'Università 10, Padua, 35020, Italy.

Developmental and Comparative Immunology 2016 Jun;59:136-44.

Abstract

Mussels innate immunity is extensively studied in order to highlight its wide-ranging resilience mechanisms, considering the low incidence of infective diseases in comparison to other bivalve molluscs and the ability to live in sites significantly contaminated by toxic metals and persistent organic chemicals. For this reason, our group analysed transcriptomic data to find out new immunity effectors and analyse their role in various stress conditions.

Dealing with structure modelling, I participated in the study of the *M. galloprovincialis* and *C. gigas* serum amyloid A (SAA), already known as potent acute phase protein in vertebrates and found to be induced after bacterial challenge in both the mentioned organisms.

Moreover, I've also participated in the characterization of the *C. gigas* spat for its competence to express chaperone proteins like HSP70, providing immunohistochemistry (IHC) data in oysters subjected to different temperature regimes.

Finally, starting from the evidence that most of transcriptomic data analyses are not able to identify short sequences, I locally developed a tool able to extract small secreted protein sequences and permitting the identification of new putative antimicrobial peptides (AMP).

Introduction

SAA

Serum amyloid A (SAA) is one of the most studied acute phase protein (APP) in vertebrates, also for its role in human amyloidosis diseases[1]. Human SAAs (1-4) are produced in liver and then secreted in circulating fluids where they associate to high density lipoproteins. SAAs show immune-related functions and can induce extracellular matrix degradation, cytokine expression connected to attraction of immuno-related cells, cholesterol transport to the liver and retinol binding and transport[2]. Some evidences suggested also bacterial opsonisation and bacterial growth inhibition[3,4]. Based on such highly different and primary roles, these proteins have been proposed to be a link between innate and adaptive immunity[5].

The SAAs genes are conserved across all vertebrate genomes, where they can be found in different copy number, and, at the same time, they have been identified in several invertebrates, including echinoderms such as sea cucumber *Holothuria glaberrima*[6], *Daphnia pulex*[7],chelicerates[8] and bivalves, specifically *Crassostrea hongkongensis*[9] and *Meretrix meretrix*[10].

Many aspects concerning the evolution and functional features of invertebrate SAAs have been so far overlooked. For this reason we tracked the presence of SAA-like sequences in the genomes of basal metazoans, protostomes and deuterostomes, providing a phylogenetic overview of their relationships. Based on RNA-seq data and quantitative PCR assays, we also investigated the expression and inducibility of SAAs in two aquaculture bivalves the Mediterranean mussel *Mytilus galloprovincialis* and the Pacific oyster *C. gigas*. Here, the structural model of mussel SAAs, based on homology and molecular dynamics, is presented.

HSP-70

The HSP-70 family includes ubiquitously expressed stress proteins, primary involved in protein folding. They are characterized by three functional domains, essentially an ATP-binding domain, a substrate binding domain and an alpha helical domain. It is also involved in preventing apoptosis[11]. The HSP-70 expression is enhanced by stresses such as heat shock, heavy metals and pathogens[12].

The study of of Crassostrea spp. spat can be informative of developmental stages and could help the identification of biomarkers for spat quality for breeders interested in improving the *C. gigas* aquacolture. Even if a link between stresses and HSP-70 expression in *C. virginica* spats has already been investigated[13], we decided to extend analysis to the economic relevant *C. gigas* spat subjected to different temperature re-

gimes. Here, immunostochemistry data aimed to highlight differential expression of HSP-70 in the tested conditions are presented.

Small bivalve proteins

As previously mentioned, mussel immunity can reveal information on innate mechanism of protection against pathogens. At the same time the great amount of transcriptomic and genomic data continuously give new insights on the mussel immunity[14]. Usually, sequence data are only partially annotated and new interesting sequences can be identified just changing the ORF predictor settings[15]. In particular, small proteins (<100 amino acids) are often overlooked owing to their difficult annotation and detection[16].

For this reasons, we decided to put our attention on the *M. galloprovincialis* and *C. gigas* secretome starting from raw transcriptomic data, looking to small proteins by means of in-house scripts and common predictors, in order to identify new putative AMPs.

Material and methods

SAA

Modelling

The protein structure model of the putative MgSAAs and CgSAA were inferred using the Modeller algorithm[17]. We applied psi-blast to identify the best templates in the PDB database: namely, mouse serum amyloid A3 (PDB entry: 4Q5G); human serum amyloid A1 (4IP8) and native human serum amyloid A1 (4IP9). The latter protein structure was chosen as template to perform the modelling of the hexameric SAA complex. We computed 100 structures of the conserved regions (corresponding to amino acids 35-138 for CgSAA, 76-196 for MgSAAa and 106-235 for MgSAAb), loops were refined and the final structures with the best DOPE score were selected[17]. Alignment used for structure calculation can be found in Supplementary Fig. 1. Images and superimpositions were rendered using PyMOL[18]. The interaction surface between chains was computed with Molsurfer Map online[19]. The protein surfaces were computed using YASARA 14[20] simulating 50 mM salt concentration.

Stability

Selected models for the three SAA hexamers were subjected to molecular dynamics simulation to check stability using GROMACS[21]. All the simulations were performed using 2 fs time steps, in a triclinic box with edge distance from the molecule set to 1.2 nm and filled with water molecules described by the SPC/E water model. Initially, 20000 steps of energy minimization at 300 K were performed using a steepest descent algorithm. For the simulation, the long-range Coulomb interactions were computed with the Particle Mesh Ewald (PME) method[22]. The V-rescale algorithm[23] for temperature coupling (300 K) and the Parrinello-Rahman algorithm for pressure coupling at 1.0 atm[24] were imposed. For neighbour searching, a Verlet scheme with a cut-off of 1.2 was set and the grip update method was imposed. Then, 5000 steps of solvent equilibration were performed using an Isothermal-isobaric (NPT) ensemble and applying a positional constraint on the peptide. The leap-frog integrator was used to integrate Newton's equations of motion. A non-bonded cut-off distance of 9 Å was imposed. Finally, a free dynamics of 100 ns at 300 K, setting the Coulomb type to Reaction-field and the non-bonded cut-off to 1.2 nm. Simulation analysis were performed using GROMACS tools g_rms and g_gyrate.

HSP-70

C. gigas spat treatment

C. gigas spats were collected from Goro lagoon and divided in four groups subjected to different temperature treatments in reconstituted sea water (Table 1): a control group maintained at 24 °C (24), group treated with increasing temperature from 24 to 32 °C (Up), treatment with temperature decreasing from 32 to 24 °C (Down), spat maintained at 32 °C (32). Treatments was performed in ten days and then 100 spats for each tank were collected. Each treatment was performed in three biological replicate. Three spat for each treatment (one per biological replicate) were dissected and fixed in Davison's solution for 12 hours and dehydrated in dehyol and finally treated with K-clear (Kaltek, Padua) and Paraplast (Sigma). Sections were used for both in situ hybridization (ISH) and immunohistochemistry (IHC).

Immunohistochemistry

Formalin-fixed tissues were de-paraffinized and rehydrated in solutions with xylene and increasing water/ethanol ratio solutions. Antigens were retrieved boiling the slides in sodium citrate buffer (10 mM sodium citrate, 0.05% Tween 20, pH 6) for two minutes.

Slides were subsequently washed with TBST (50 mM Tris, 150 mM NaCl, 0.05% Tween 20, pH 7.6) and blocked with 10% anti-goat serum and 1 % BSA in TBS (50 mM Tris, 150 mM NaCl, pH 7.6) for 3 hours at room temperature. Slides were incubated with primary antibody Anti-HSP70 (1:1000; Sigma) ON at 4 °C. After washing, endogenous peroxidases were inactivated with 0.3 % H_2O_2 in TBS for 15 minutes. The secondary antibody were incubated for 2 hours at room temperature. Signal was detected using DAB leading to dark brown positive signal. Slides were counterstained with hematoxylin before dehydration and mounting. Slides images were taken with bright-field microscope (Leica DM2000).

Small bivalve proteins

Protein identification

De novo C. gigas and *M. galloprovincialis* transcriptome assembly was performed as previously reported[25]. DNA sequences were translated in all six frames and all the sequences following a methionine and with a stop codon between 25 and 100 aminoano acids after the methionine were selected as initial database. Sequence were clustered imposing 95% identity using CD-HIT[26]. The sequences with predicted signal peptide[27] were further processed in two ways: the set was analysed with CAMPsign to identify known AMPs features[28]; we extracted the sequences with more than four cysteines to check for new cysteine rich peptides.

For *C. gigas* the presence of the sequence on the genome was performed using the blast tool of Ensembl Metazoa (http://metazoa.ensembl.org/index.html). *M. galloprovincialis* sequences presence on the genome was inferred from draft genome reads reassembled using hits as template.

Results and discussion

SAA

From transcriptome analysis we found two *M. galloprovincialis* and one *C. gigas* SAA sequences. Based on the structural modelling of the SAA conserved domain, the interaction between monomers of both the mussel and oyster proteins is compatible with the formation of a hexameric complexes, a fact suggesting not only sequence and structure conservation but also a functional role similar to that of murine and human SAAs. Hexameric complexes result to be stable through the whole 100 ns molecular dynamics basing on RMSD and radius of gyration (Supplementary Figure 1). When the protein monomers arrange as a hexamer, the first helix of the conserved domain of bivalve SAAs could still potentially form a conserved positively charged pore, despite its low sequence similarity with human SAA1. At the N-terminal tail and before the conserved domain, MgSAAa and MgSAAb present a long disordered serine-rich tail which might be cleaved before activation, reasonably in agreement to a regulatory role. CgSAA possesses a loop similar in length to that of human SAA, whereas the corresponding regions of MgSAAa and MgSAAb are longer and display alternated negatively and positively charged residues. It remains unclear whether the loop region can be rearranged to extend the helices, as reported by secondary structure prediction (Figure 1 and Supplementary Figure 1).

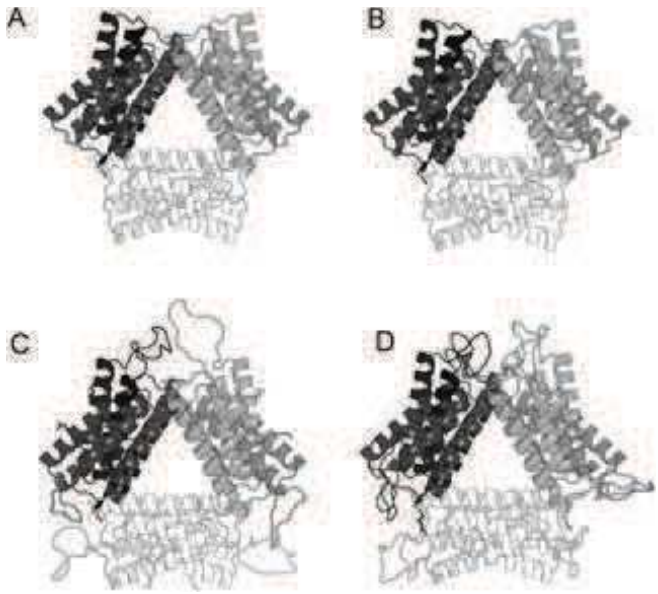


Figure 1: Hexameric model structure of A. Human native SAA (4IP9); B. CgSAA; C. MgSAAa; D. MgSAAb. The chains of the six monomers are represented by an increasing scale of grey colour.

HSP-70

Immunohistochemical localization

Immunostaining for HSP-70 was performed using the same monoclonal antibody as for western blot. Panels in Figure 2 indicate the distribution pattern of HSP 70 (positive brown signals) in tissues of oyster spat under different experimental conditions (24, Up, 32 and Down indicate the applied temperature regimes). Ducts of digestive gland and circulating haemocytes in the vicinity of digestive gland and gills resulted to be the major sites where the HSP 70 expression was recorded.

The upper panels of Figure 2 show the tissues of oyster spat in the control tank maintained at 24°C at two different magnifications (20 and 40X respectively): the positive signals indicate the presence of HSP 70 within the ducts of digestive gland. At higher magnification, the positive signals identify haemocytes circulating in the intraepithelial cells of the ducts. The presence of positive signals in the control group reflects the constitutive expression of HSP 70. Few positive cells were identified in the bottom panels, with some signals maintained among ducts and intraepithelial cells. An highly diffuse expression could be detected in the Up regime (temperature increase from 24 to 32 °C), suggesting the ubiquitous HSP 70 expression in the spat cells likely experiencing heat stress.

In the 32 group, the signals were found to be confined to the epithelia of digestive gland, as already shown after treatment of individuals with heavy metals[29]. Having haemocytes the ability to migrate into the connective tissues and also to the surface epithelia of various organs, HSP70 expression could be maintained in haemocytes only after acclimation to the high temperature, reproducing the expression pattern observed in the 24 °C tank, except for the localization of haemocytes.

Appendix

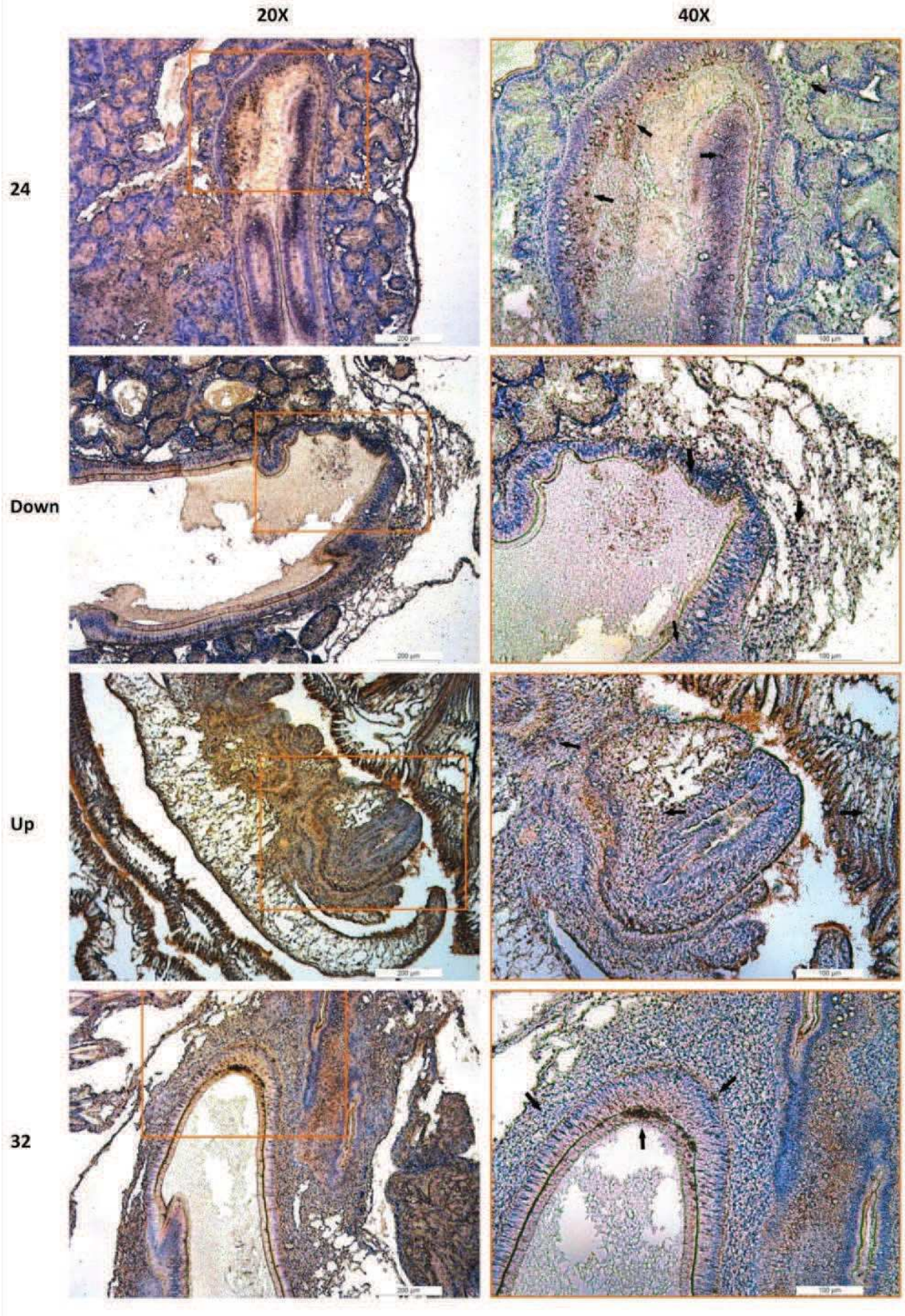


Figure 2: Oyster spat slides, positive signals for HSP70 in brown, hematoxylin counterstaining in blue. From top to bottom: 24, Down, Up and 32 tanks. 20x magnification on the left, 40x on the right.

Mussel small secreted proteins

Starting from virtually translated transcriptomes, we were able to find more than 1 million potential ORFs ranging between 25 and 100 amino acids length for both *M. galloprovincialis* and *C. gigas*, then lowered to about 600000 by imposing 95% identity in the clustering process. Such big amount of possible new proteins is biased by a great number of false positive. In order to reduce the set complexity we focused on sequences with a signal peptide, thus filtering out many false positives (the probability of a random sequence reproducing a signal peptide is about 1%[27]) and, at the same time, concentrating on secreted proteins which are expected relevant in organism-pathogen interactions[30]. In this way, the sequence number was reduced to about 10000 for both organisms. All these sequences were filtered using CampSIGN[28] and the searches for known AMPs signatures, lead to the identification of 26 new sequences for *M. galloprovincialis*, 23 confirmed in the genome, and 13 for *C. gigas*, 10 found also on the deposited genome. Indeed, many of these sequences, not yet annotated, can be found in multiple copies on the genome, with slight sequence differences.

Due to the potential applications and economic relevance of such small proteins including many peptides, the sequencedata are not presented in this work, but the set up of the algorithm has been proposed to highlight the importance of a critical and careful analysis of genomic and transcriptomic data in order to rescue sequences of interest overlooked by the automatic annotations of already deposited data.

References

1. Urieli-Shoval, S., Linke, R. P. & Matzner, Y. Expression and function of serum amyloid A, a major acute-phase protein, in normal and disease states. *Curr. Opin. Hematol.* **7,** 64–69 (2000).

2. Mullan, R. H. *et al.* Acute-phase serum amyloid A stimulation of angiogenesis, leukocyte recruitment, and matrix degradation in rheumatoid arthritis through an NF-kappaB-dependent signal transduction pathway. *Arthritis Rheum.* **54,** 105–114 (2006).

3. Kovacevic, N. & Belosevic, M. Molecular and functional characterization of goldfish (Carassius auratus L.) Serum Amyloid A. *Fish Shellfish Immunol.* **47,** 942–953 (2015).

4. Shah, C., Hari-Dass, R. & Raynes, J. G. Serum amyloid A is an innate immune opsonin for Gram-negative bacteria. *Blood* **108,** 1751–1757 (2006).

5. Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* **272,** 50–53 (1996).

6. Santiago, P., Roig-López, J. L., Santiago, C. & García-Arrarás, J. E. Serum amyloid A protein in an echinoderm: Its primary structure and expression during intestinal regeneration in the sea cucumber Holothuria glaberrima. *J. Exp. Zool.* **288,** 335–344 (2000).

7. Colbourne, J. K. *et al.* The Ecoresponsive Genome of Daphnia pulex. *Science* **331,** 555–561 (2011).

8. Ribeiro, J. M. C. *et al.* The sialotranscriptome of Antricola delacruzi female ticks is compatible with non-hematophagous behavior and an alternative source of food. *Insect Biochem. Mol. Biol.* **42,** 332–342 (2012).

9. Qu, F., Xiang, Z. & Yu, Z. The first molluscan acute phase serum amyloid A (A-SAA) identified from oyster Crassostrea hongkongensis: Molecular cloning and functional characterization. *Fish Shellfish Immunol.* **39,** 145–151 (2014).

10. Zou, L. & Liu, B. Identification of a Serum amyloid A gene and the association of SNPs with Vibrio-resistance and growth traits in the clam Meretrix meretrix. *Fish Shellfish Immunol.* **43,** 301–309 (2015).

11. Beere, H. M. *et al.* Heat-shock protein 70 inhibits apoptosis by preventing recruitment of procaspase-9 to the Apaf-1 apoptosome. *Nat. Cell Biol.* **2,** 469–475 (2000).

12. Mayer, M. P. & Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **62,** 670

13. Ueda, N. & Boettcher, A. Differences in heat shock protein 70 expression during larval and early spat development in the Eastern oyster, Crassostrea virginica (Gmelin, 1791). *Cell Stress Chaperones* **14,** 439–443 (2009).

14. Gerdol, M. & Venier, P. An updated molecular basis for mussel immunity. *Fish Shellfish Immunol.* **46,** 17–38 (2015).

15. Thierry-Mieg, D. & Thierry-Mieg, J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* **7 Suppl 1,** S12.1–14 (2006).

16. Storz, G., Wolf, Y. I. & Ramamurthi, K. S. Small Proteins Can No Longer Be Ignored. *Annu. Rev. Biochem.* **83,** 753–777 (2014).

17. Chen, Y., Shang, Y. & Xu, D. Multi-dimensional scaling and MODELLER-based evolutionary algorithms for protein model refinement. in *2014 IEEE Congress on Evolutionary Computation (CEC)* 1038–1045 (2014). doi:10.1109/CEC.2014.6900443

18. *The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.*

19. Gabdoulline, R. R., Wade, R. C. & Walther, D. MolSurfer: a macromolecular interface navigator. *Nucleic Acids Res.* **31,** 3349–3351 (2003).

20. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins Struct. Funct. Bioinforma.* **47,** 393–402 (2002).

21. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinforma. Oxf. Engl.* **29,** 845–854 (2013).

22. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

23. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 14101 (2007).

24. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).

25. Rosani, U., Domeneghetti, S., Pallavicini, A. & Venier, P. Target Capture and Massive Sequencing of Genes Transcribed in Mytilus galloprovincialis. *BioMed Res. Int.* **2014,** (2014).

26. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinforma. Oxf. Engl.* **28,** 3150–3152 (2012).

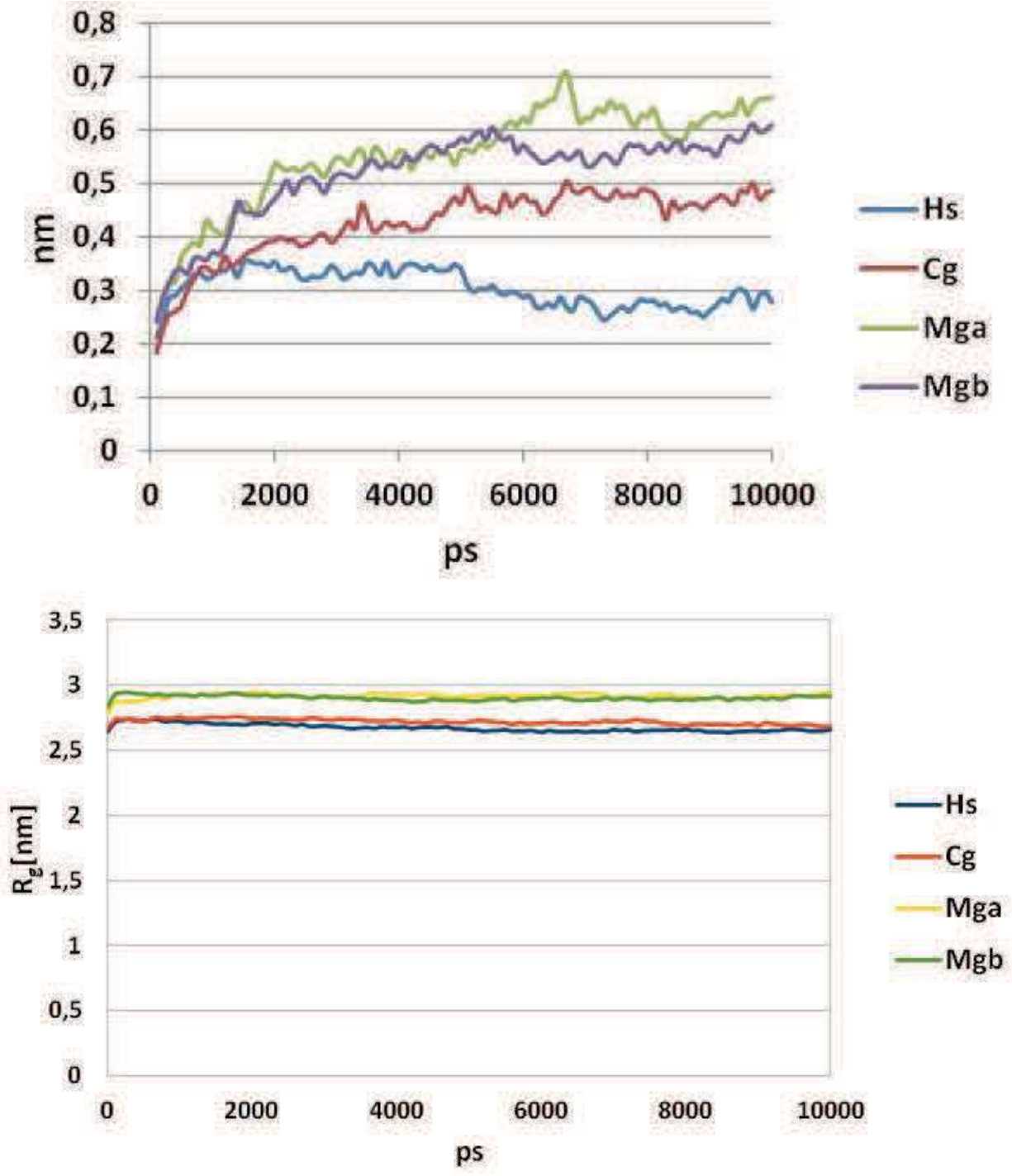
27. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8,** 785–786 (2011).

28. Waghu, F. H., Barai, R. S. & Idicula-Thomas, S. Leveraging family-specific signatures for AMP discovery and high-throughput annotation. *Sci. Rep.* **6,** 24684 (2016).

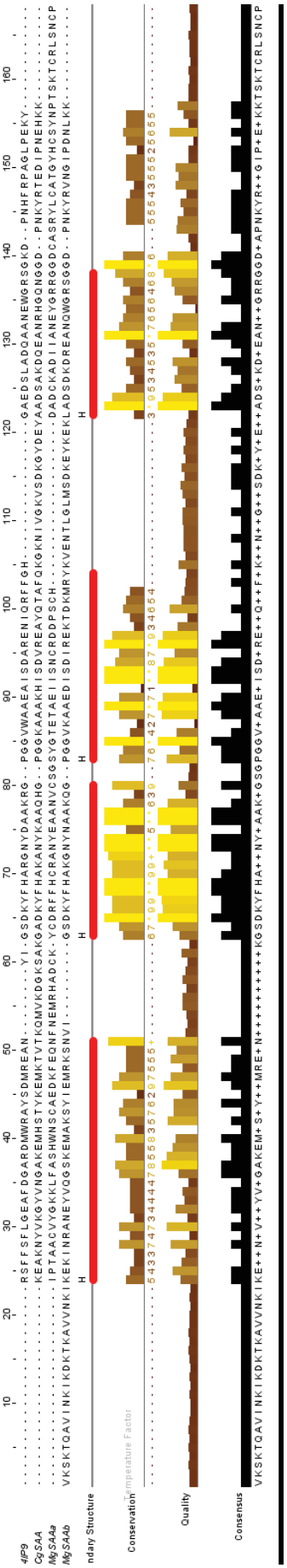
29. Moraga, D. *et al.* Stress response in Cu2+ and Cd2+ exposed oysters (Crassostrea gigas): An immunohistochemical approach. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* **141,** 151–156 (2005).

30. Meinken, J., Walker, G., Cooper, C. R. & Min, X. J. MetazSecKB: the human and animal secretome and subcellular proteome knowledgebase. *Database* **2015,** bav077 (2015).

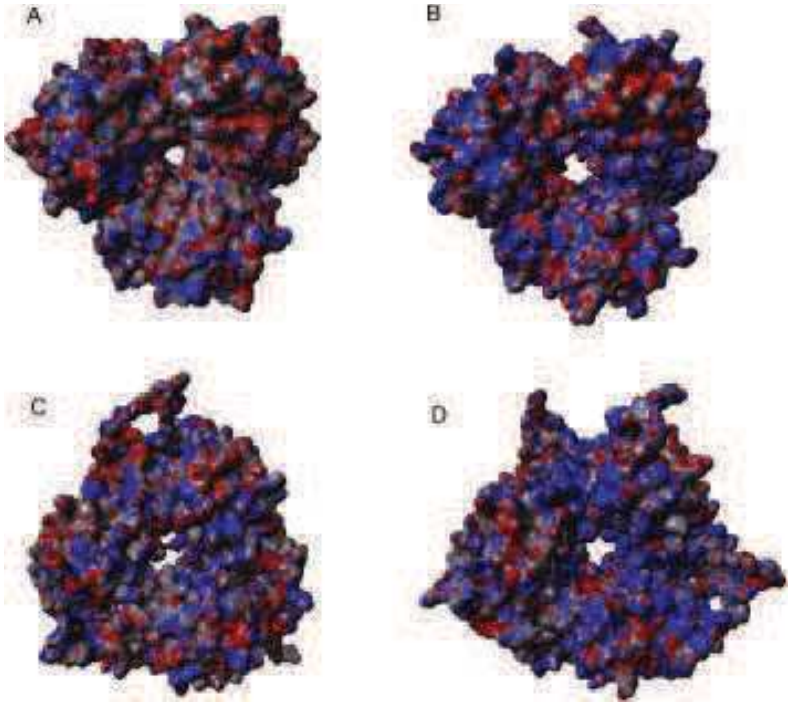
Supplementary material



Supplementary figure 1:on top RMSD during simulation for SAAs. On bottom radius of gyration during simulation for SAAs. Human SAA is reported as control. Hs: *Homo sapiens* SAA, Cg: *Crassostrea gigas* SAA, Mga: *Mytilus galloprovicialis* SAA-a, Mgb: *Mytilus galloprovincialis* SAA-b



Supplementary Figure 2: Sequence alignments of mussels SAAs with human template (pdb 4IP9) used for structure calculation. Secondary structure of template is also reported.



Supplementary Fig. 2.

SAA structure models. Surface electrostatic potential was computed using the Poisson-Boltzmann approach. Positive (blue) and negative (red) surfaces are shown. A. Human native SAA (4IP9); B. CgSAA; C. MgSAAa; D. MgSAAb.