UNIVERSITA' DEGLI STUDI DI NAPOLI “FEDERICO II”

Dottorato in Biologia Computazionale e Bioinformatica

26° ciclo

Coordinatore: prof. Sergio Cocozza

Identification, production and structural modelling of cationic antimicrobial peptides (CAMPs)

Tutor: dott. Eugenio Notomista

Studente: Lorenzo Durante

Co-tutor: dott. Cristina De Castro

Anno accademico 2013/2014

INDEX

RIASSUNTO

I peptidi antimicrobici cationici (*CAMPs*, *cationic antimicrobial peptides*) sono piccoli peptidi (15-50 residui) che esercitano una azione battericida diretta e costituiscono l'arma più antica del sistema immunitario innato degli eucarioti multicellulari. Queste molecole possiedono una carica netta positiva ed acquisiscono una struttura anfipatica, che rende possibile la loro interazione con la membrana plasmatica, destabilizzando la sua architettura e/o creando pori. I *CAMPs* sono promettenti agenti terapeutici ed il nostro gruppo di ricerca è focalizzato sullo sviluppo di nuovi *CAMPs* diretti contro i patogeni più comuni nelle infezioni polmonari dei malati di fibrosi cistica. Negli scorsi anni, sono state scoperte diverse proteine che mostrano un'attività antibatterica non correlata con la loro funzione primaria; queste proteine sembrano agire da trasportatrici, nella loro sequenza, di *CAMPs* criptici, che potrebbero essere rilasciati dall'azione di proteasi umane o batteriche.
I principali scopi di questo lavoro sono (1) lo sviluppo di un nuovo sistema di punteggio per l'identificazione di peptidi antimicrobici criptici nelle sequenze proteiche, (2) lo sviluppo e l'ottimizzazione di un costrutto di fusione per l'espressione dei nuovi *CAMPs* ed infine (3) l'elaborazione di strategie di *modelling* di *CAMPs* attraverso simulazioni Monte Carlo e funzioni di solvatazione implicita.
Per quanto riguarda il primo scopo, abbiamo sviluppato funzioni di punteggio basate sulla carica e sulla idrofobicità, due caratteristiche universalmente riconosciute come essenziali per l'attività antimicrobica. La sostanziale novità del nostro sistema di punteggio è la presenza di variabili ceppo-specifiche che possono essere calcolate utilizzando i dati di attività antimicrobica di un set di peptidi saggiato sui ceppi batterici di interesse. Le nostre funzioni di punteggio, quindi, possono essere “regolate” al fine di identificare *CAMPs* particolarmente attivi contro il ceppo di interesse.
Una validazione preliminare del sistema di punteggio è stata condotta *in silico* mediante l'analisi di un pannello di proteine contenenti peptidi antimicrobici criptici noti; il nostro sistema ha identificato quasi tutti i *CAMPs* criptici noti. Alcuni nuovi putativi *CAMPs* sono già stati prodotti con un nuovo sistema di espressione ricombinante in *Escherichia coli*. Il sistema è costituito da un costrutto di fusione dove la sequenza codificante il peptide è localizzata a valle di un *carrier*, l'onconasi, una ribonucleasi capace di formare corpi di inclusione con elevata efficienza e resa. I corpi di inclusione sequestrano il peptide antimicrobico, mascherando quindi la sua tossicità verso al cellula batterica. L'onconasi ed il peptide sono uniti da un *linker* che contiene il dipeptide Asp-Pro, capace di idrolizzarsi

spontaneamente in condizione relativamente blande di temperatura e pH.
Il costrutto di fusione è stato testato clonando il peptide criptico già noto della trombina umana (*ThrAP*, *thrombin antimicrobial peptide*) ed ottimizzato con diverse mutazioni puntiformi al fine di abolire siti di taglio interni. La versione finale contiene anche una sequenze di istidine che permette di purificare in un singolo passaggio il costrutto di fusione. Il peptide è stato invece purificato sfruttando la sua diversa solubilità, rispetto al costrutto di fusione, a pH neutro. L'espressione è stata condotta in un nuovo mezzo di coltura dalla composizione semi-definita e si è avuta una resa pari a circa 7-10 mg di peptide puro da un litro di coltura. Un nuovo peptide antimicrobico identificato nella apolipoproteina E (*ApoE-AP*, *apolipoprotein E antimicrobial peptide*) è stato prodotto attraverso la stessa strategia. I due peptidi ricombinanti sono stati caratterizzati mediante dicroismo circolare; in tampone sono apparsi privi di una struttura definita, mentre in presenza di agenti che mimano un ambiente di membrana hanno acquisito una struttura elicoidale. Inoltre, è stata osservata una possibile interazione tra *ThrAP* e due molecole di grande importanza da un punto di vista biologico, il lipopolisaccaride e l'alginato.
Infine, sono stati effettuati studi computazionali allo scopo di identificare funzioni di solvatazione implicita che potessero simulare strutture sperimentali di *CAMPs*; la conformazione di peptidi le cui strutture sono state risolte in presenza di micelle (SDS o DPC) e TFE è stata modellata con la strategia Monte Carlo nel vuoto, in acqua implicita, in ottanolo implicito ed in ottanolo implicito con attenuazioni dell'energia di solvatazione. Quest'ultima condizione è stata considerata per creare una sorta di ambiente che mimasse una membrana. Le simulazioni hanno mostrato che le strutture risolte in presenza di micelle vengono più correttamente simulate con la solvatazione da ottanolo “attenuata”, mentre le strutture risolte in presenza di TFE sono spesso simulate meglio nel vuoto.

SUMMARY

Cationic AntiMicrobial Peptides (CAMPs) are small peptides (15-50 residues) which exert a direct microbicidal activity and constitute the most ancient arm of the innate immune system of multicellular eukaryotes. They possess a positive net charge and acquire an amphipathic structure, which permits their interaction with the cell membrane, destabilizing its architecture and/or creating pores. CAMPs are promising therapeutic agents and our research group is focused on the development of new CAMPs against the most common pathogens in the lung infections of cystic fibrosis patients. In the last years, several proteins which show antibacterial activity not correlated with their primary function have been discovered; these proteins seem to act as carriers in their primary structure of "cryptic" CAMPs, that could be released by the action of human or bacterial proteases.

The main aims of this research work are (1) the development of a novel scoring system for the identification of "cryptic" antimicrobial peptides in protein sequences, (2) the development and optimization of a fusion construct for the preparation of the novel CAMPs and (3) the development of modelling strategies of CAMPs through Monte Carlo simulations and implicit solvation energy functions.

As for the first aim, we have developed scoring functions based on charge and hydrophobicity, two characteristics universally recognized as essential for the antimicrobial activity. The main novelty of our scoring system is the presence of strain-specific variables which can be estimated using antimicrobial activity data of a set of peptides assayed on bacterial strains of interest. Thus our scoring functions can be tuned to identify CAMPs particularly active against the strain of interest.

A preliminary i*n silico* validation of the scoring system was conducted through the analysis of a panel of protein sequences containing known cryptic antimicrobial fragments. Our system was able to identify almost all the antibacterial fragments. Several newly identified putative CAMPs were already produced with a novel recombinant expression system in *Escherichia coli*. This system is constituted by a fusion construct where the sequence coding the peptide is located downstream a carrier, onconase, a ribonuclease capable of forming inclusion bodies with high efficiency and yield. Inclusion bodies sequester the CAMP thus abolishing its toxicity towards the bacterial cell. Onconase and the peptide are joined by a linker which contains the dipeptide Asp-Pro which undergo spontaneous hydrolysis in relative mild conditions of temperature and pH.

The fusion construct was tested cloning the already known cryptic CAMP of human

thrombin (ThrAP, thrombin antimicrobial peptide) and optimized with several point mutations in order to abolish internal sites of cleavage which were discovered. The final optimized version contains also a tag of histidines which permits a one-step purification of the fusion construct. The peptide was instead purified by exploiting its different solubility at neutral pH with respect to the fusion construct. The expression was conducted in a novel semi-defined rich medium and a final yield of about 7-10 mg of pure peptide from one liter of culture was obtained. A novel CAMP identified in human apolipoprotein E, ApoE-AP (apolipoprotein E-antimicrobial peptide) was produced with the same strategy. The two recombinant peptides displayed antibacterial activity towards both Gram-negative and Gram-positive bacteria. The peptides underwent a preliminary structural characterization by means of circular dichroism; they were unstructured in buffer, but acquired a helical structure in membrane-mimicking environments. Moreover, a putative interaction between ThrAP and two molecules of great biological relevance, LPS and alginate, was observed.

Finally, computational studies were performed with the aim of identifying solvation energy functions which simulate experimental structures of CAMPs; the conformation of peptides whose structures were solved in the presence of micelles (SDS or DPC) and TFE was modeled using a Monte Carlo strategy in vacuum, implicit water, implicit octanol and implicit octanol with attenuations of the solvation energy. The last condition was employed in order to re-create a sort of membrane-like environment. The simulations showed that structures solved in the presence of micelles are better simulated with the “attenuated” octanol solvation, whereas, structures solved in the presence of TFE are often better simulated in vacuum.

ABBREVIATIONS

aa.	amino acids
AMAC	ammonium acetate
ApoE-AP	Apolipoprotein E-antimicrobial peptide
ASA	accessible surface area
BPA	3-bromopropylamine
BSA	bovine serum albumin
CAMP	cationic antimicrobial peptide
CD	circular dichroism
DPC	dodecylphosphocholine
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EEF-1	effective energy function-1
GuHCl	guanidinium chloride
HPLC	high-pressure liquid cromatography
hRNase 4-CAM	carboxamido-methyl-human ribonuclease 4
hRNase 4-PA	propylamine-human ribonuclease 4
hRNase 4-PE	pyridine-ethyl-human ribonuclease 4
IAA	iodoacetamide
IC_{50}	half maximal inhibitory concentration
IMAC	immobilized metal-affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LPS	lipopolysaccharide/s
Lyz-CAM	carboxamido-methyl-lysozyme
Lyz-PA	propylamine-lysozyme
Lyz-PE	pyridine-ethyl-lysozyme
MC	Monte Carlo
MD	molecular dynamics
MES	2-(*N*-morpholino)ethanesulfonic acid
MIC	minimum inhibitory concentration
NaP	sodium-phosphate
NMR	nuclear magnetic resonance
OD	optical density
ONC	onconase
PDB	Protein Data Bank
QSAR	quantity structure-activity relationship
RMSD	root-mean-square deviation
RNase/hRNase	ribonuclease/human ribonuclease
rpm	revolutions per minute
SD	standard deviation
SDRM	semi-defined rich medium
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	solvation energy coefficien
SUMO	small ubiquitin-related modifier
TEV	Tobacco Etch Virus
TFE	2,2,2-trifluoroethanol
ThrAP	Thrombin-antimicrobial peptide
TSA	trypticase soy agar
VP	4-vinylpyridine

1. INTRODUCTION

1.1 General properties of cationic antimicrobial peptides (CAMPs)

Cationic antimicrobial peptides (CAMPs) constitute the most ancient arm of the host-defense system and are thus widely diffused in nature, from bacteria to mammals. These molecules are able to exert a direct antimicrobial, antiviral and antifungin activity and their length usually spans from 10 to 50 residues. Regarding their structure, CAMPs are divided in three different classes: α-helical peptides, like magainins and mellitin, globular peptides with β-sheets linked by disulphide bridges and finally peptides without a well defined structure, which are often enriched in amino acids like proline, tryptophan, histidine or glycine (Figure 1).

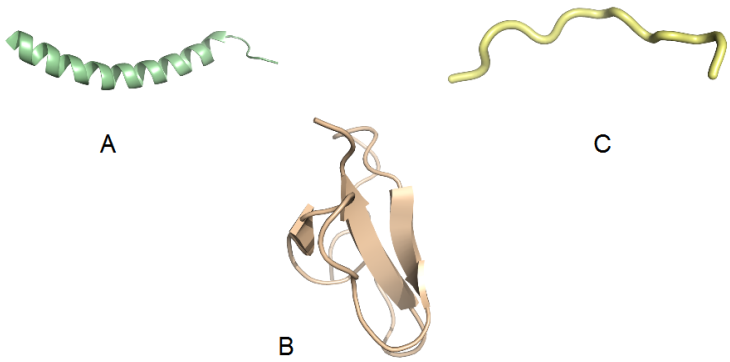


Figure 1: examples of the three different structural classes of CAMPs. **A**: human LL-37 bound to SDS micelles (PDB code: 2K6O); **B**: human β-defensin 2 (PDB code: 1E4Q); **C**: bovine indolicidin bound to SDS micelles (PDB code: 1G8C).

CAMPs, as their name suggests, have a positive net charge due to the abundance in lysines and arginines and are rich in hydrophobic residues; they are consequently able to acquire an amphipatic structure which allow them to perturb the bacterial membranes, their main target, leading to cell death. Bacterial membranes possess anionic phospholipids and thus the electrostatic interaction between CAMPs and these lipids can be easily imagined; in eukaryotic membranes, instead, anionic phospholipids are sequestered in the inner leaflet and thus no electrostatic interactions can be formed (Wiesner J. and Vilcinskas A., 2010). This important physicochemical characteristic is the

basis of the selectivity of cationic antimicrobial peptides towards bacteria (Figure 2).

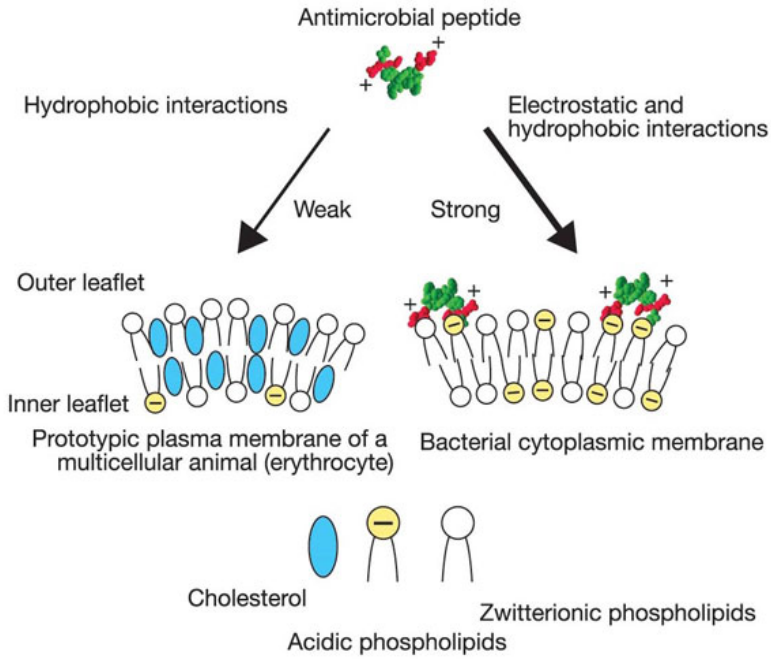


Figure 2: representation of the basis of specificity of CAMPs (from Zasloff M., 2002).

If we consider the emerging of multi-drug resistant pathogens in the last years, cationic antimicrobial peptides are very promising therapeutic agents, because the induction of a form of resistance is very unlikely, as bacteria should drastically change the membrane architecture. However, some “transient” forms of resistance have been described, like the modification of cell wall components in order to reduce the negative net charge, the increase of membrane rigidity, the production of a polysaccharide capsule which may act as a shield or even the alteration of the expression of CAMPs in the host (Guilhelmelli F. *et al.*, 2013). Moreover, the biofilm produced by some bacteria like *Pseudomonas aeruginosa* can reduce the action of antimicrobial peptides. For example, Chan C. *et al.*, (2004 and 2005) demonstrated that the exopolysaccharide alginate, the major component of the biofilm, can act as an auxiliary membrane, binding CAMPs and inducing their aggregation.
CAMPs are active also on resting bacteria, whereas commonly used antibiotics are instead only active on dividing cells. Finally, there are many reports of additional biological activities of these peptides, like immunomodulation, wound healing and anticancer activity (Pushpanathan M. *et al.*, 2013). An example of immunomodulation is the ability to inhibit lipopolysaccharide (LPS)-induced pro-inflammatory cytokine production. LPS is an endotoxin localized on the external membrane of Gram-negative bacteria and

characterized by an amphiphilic structure, with an hydrophobic portion, lipid A, and an hydrophilic moiety, divided into a polysaccharide part and antigen “O”, with several repeats of variable oligosaccharide units. Different CAMPs are able to bind LPS aggregates, making them “invisible” to transduction pathways that once activated lead to inflammation and even septic shock. The binding is driven by the cationic residues of the peptide, which interact with the phosphate groups of lipid A, while hydrophobic residues are located in the lipophilic core region; in addition, aromatic residues play a fundamental role because they stabilize a packed structure, that can also facilitate the translocation across the outer membrane in order to reach the plasma membrane (Pulido D. *et al.*, 2011).
Among the main drawbacks of CAMPs, we find the haemolytic activity, usually observed in particularly hydrophobic and amphiphilic peptides, and the susceptibility to proteolytic degradation (Aoki W. and Ueda M., 2013).

1.2 Models of action

CAMPs are able to perturb and even destroy bacterial membranes, leading to cell death for the disruption of the electrochemical gradient, the loss of metabolites and the final lysis. The exact mechanism of membrane perturbation and/or disruption is still not perfectly clear even if many experimental and computational works have tried to shed more light. Three models have been proposed: the barrel-stave model, the toroidal model and the carpet model. According to the barrel-stave model, the peptides insert into the membrane creating a pore which they line; in the toroidal model, the peptides create a toroidal pore, which is lined also by the phospholipids' heads; finally, in the carpet model, the peptides interact with the membrane surface and, when a critical concentration is reached, the membrane is destroyed with the formation of micelles and lipid-peptides aggregates (Figure 3).

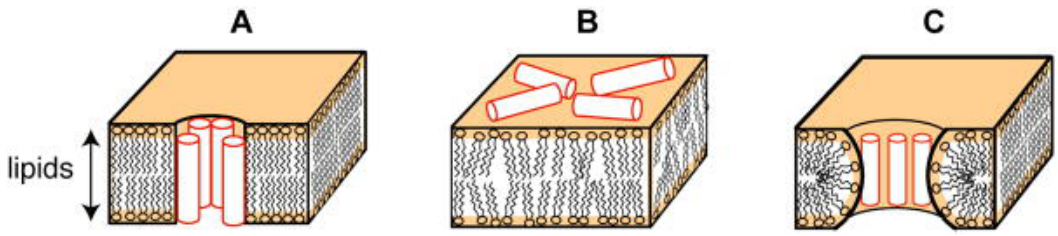


Figure 3: representation of the three proposed mechanisms of membrane perturbation. **A**: barrel-stave model; **B**: carpet model; **C**: toroidal model (from Tang M. and Hong M., 2009).

More recently, new models arose from the experimental observations of peptide-induced lipid segregation of anionic components from zwitterionic lipids (Figure 4); in detail, anionic lipids are clustered and consequently biophysical and biological alterations occur, like the modification of membrane curvature, which can alter cell division or sporulation, or the loss of functionality of protein-lipid complexes with important physiological properties (Teixeira V. *et al.*, 2012).

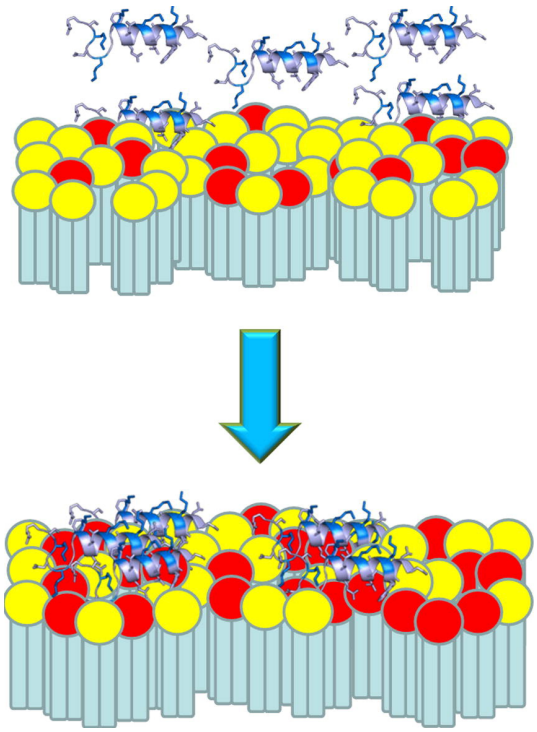


Figure 4: representation of the segregation of anionic lipids (in **red**) from zwitterionic ones (in **yellow**) induced by CAMPs (in **blue**) (from Teixeira V. *et al.*, 2012).

1.3 Structural studies of CAMPs

The structure of CAMPs has been studied using a variegated panel of both experimental and computational techniques. Structural studies are aimed to elucidate the structure of peptides in presence of different kinds of micelles (usually anionic, like SDS, to mimic a bacterial membrane or zwitterionic, like DPC, to mimic instead the eukaryotic membrane), lipid bilayers and also the structure of pores and the dynamic of their formation.

Experimental studies performed with circular dichroism focus on the study of the secondary structure of the peptides in aqueous buffers alone and in presence of micelles

and possible ligands like lipopolysaccharide and alginate (Chan C. *et al.*, 2004; Gopal R. *et al.*, 2012). Globular peptides with β-sheets are structured in water (Munyuki G. *et al.*, 2013), while instead α-elical peptides are usually unordered, possessing a typical random-coil spectrum, but micelles and ligands are able to induce an helicoidal structure. In particular, the structuring observed in presence of LPS is very important, as the binding to this molecule is a prerequisite for a possible detoxifying activity. Experiments with circular dichroism are often conducted in presence of trifluoroethanol, an agent able to induce α-helix (Roccatano D. *et al.*, 2002), in order to create a membrane-mimicking environment and to evaluate the propensity to acquire an ordered structure.

Solution NMR permits to solve the tertiary structure of CAMPs and is generally performed in presence of SDS or DPC micelles, trifluoroethanol and also lipopolysaccharide. Solid-state NMR gives instead a more realistic image of the membrane interactions of antimicrobial peptides, because it allows the study of CAMPs in the presence of liquid disordered phospholipid bilayers and gives information also on the dynamics and phase properties of lipids (Bechinger B. and Salnikov E. S., 2012).

Finally, force atomic microscopy is an experimental technique recently used to directly visualize the formation and the progressive lateral expansion of membrane pores (Rakowska P. D. *et al.*, 2013).

Computational techniques are employed to study at an atomic-level the perturbation of lipid bilayers by antimicrobial peptides or their interaction with micelles. Micelles mimic the main physicochemical property of a membrane: the presence of an hydrophobic core surrounded by an hydrophilic shell. They posses lower relaxation times than lipid bilayers and provide a direct link to NMR spectroscopy; however, they have an higher degree of curvature, a different chain structure than biological relevant lipids and could more easily deform. On the other hand, the more realistic modelling in lipid bilayers is biased by the choice of the initial position and orientation of the peptide and the accurate sampling of the phase space can be a challenge (Mátyus E. *et al.*, 2007; Langham A. and Kaznessis Y. N., 2010). Different molecular dynamics (MD) simulations pointed out that CAMPs bind more strongly to membrane pores, stabilizing them (Mihajlovic M. and Lazaridis T., 2010; Lam K. L. H. *et al.* 2012; He Y. *et al.*, 2013). Moreover, MD permitted to observe the spontaneous formation of toroidal pores which appeared disordered, without a regular packing and orientation of the peptides, in contrast with the classical cylindrical model (Sengupta D. *et al.*, 2008) (Figure 5).

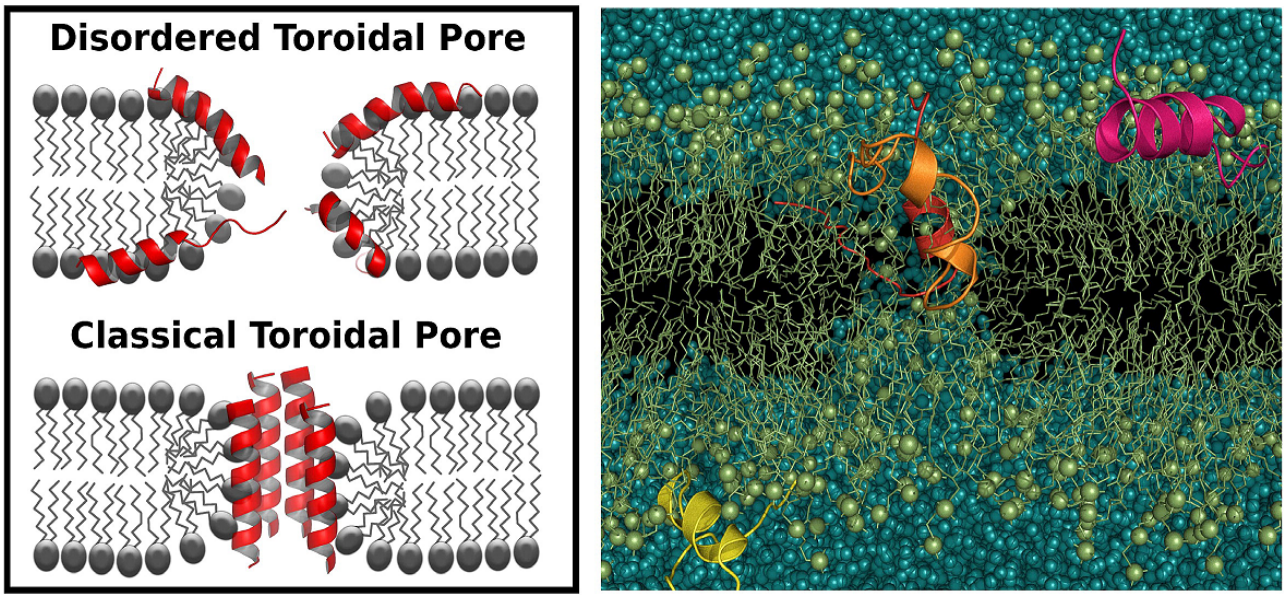


Figure 5: on the **left**, cartoon image which shows the difference between the disordered toroidal pore and the "classical" ordered model; on the **right**, snapshot of the disordered toroidal pore from the MD simulation (from Sengupta D. *et al.*, 2008).

Finally, MD studies have also shown a correlation between peptides' haemolytic activity and the strength of the interaction with zwitterionic membranes or micelles and the mechanism of insertion, thus permitting to predict models to design non-toxic peptides which still retain antimicrobial activity (Khandelia H. *et al.*, 2006; Sayyed-Ahmad A. *et al.*, 2009; Mihajlovic M. and Lazaridis T., 2010).

The computational studies described above are all conducted in the presence of a micelle or a lipid bilayer with explicit solvation, except for the works by Mihajlovic M. and Lazaridis T., (2010) and He Y. *et al.*, (2013), where instead simulations were run in implicit models of membrane and pores of different geometry. The use of implicit solvation simplify a fully atomistic simulation, reducing the number of atoms of the system, and the modelling of CAMPs in implicit solvations of different nature (polar *vs.* non polar) could highlight their possible influence on the structure of the peptides.

1.4 Proteins as carrier of cryptic CAMPs

In the last years, several fragments possessing antimicrobial activity were identified in many proteins, whose activity is often not correlated with immunity; some examples are given by human proteins like lysozyme (Ibrahim H. R. *et al.*, 2005), thrombin (Kasetty G. *et*

al., 2011a), cathepsin G (Shafer W. M. *et al.*, 1993), apolipoprotein E (Dobson C. B. *et al.*, 2006), RNase 3 (Boix E. *et al.*, 2012). These proteins seem to act as carrier of cryptic CAMPs, that could be cut and released by the action of endogenous and/or bacterial proteases; it can be hypothesized that evolution created proteins with hidden host-defense potential in their primary structure, besides “canonical” antimicrobial peptides encoded by genes (D'Alessio G., 2011). The cryptic antimicrobial peptides are often located at the N- or C-terminal of the protein, and thus the excision by proteases can be easily triggered (Kasetty G. *et al*., 2011b; Torrent M. *et al.*, 2013) (Figure 6).

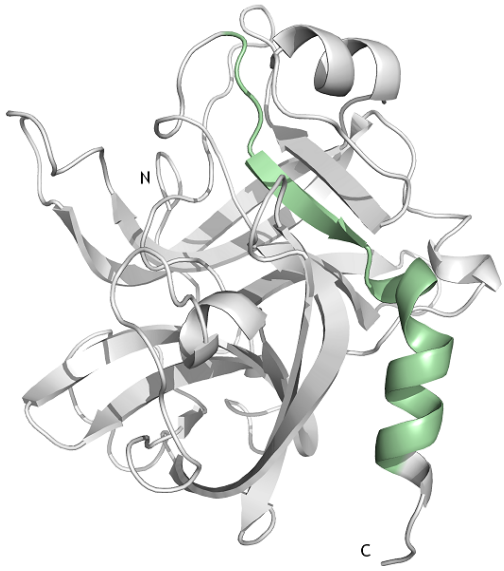


Figure 6: 3D-structure of human thrombin (PDB code: 1PPB) with the C-terminal antimicrobial region coloured in green (Kasetty G. *et al.*, 2011a).

Moreover, a new link between the complement system and the antimicrobial peptides, the two main branches of the innate immune system, was established with the discovery of antimicrobial fragments of complement factors C3a (Nordahl E. A. *et al.*, 2004) and C8a (Zhang Z. *et al.*, 2014), while the identification of an antimicrobial peptide released by fibrinogen suggests a new connection between immunity and blood coagulation (Påhlman L. I. e*t al.*, 2013). Other interesting examples of cryptic CAMPs stored inside proteins not correlated with immunity are instead given by different heparin-binding domains (Andersson E. *et al.*, 2004) and by growth-factors (Malmstem M. *et al.*, 2007). Following tissue damage and infection, the complement system and the coagulation cascade are activated, growth factors and antimicrobial peptides expression is enhanced and thus a

synergistic action of “canonical” and hidden CAMPs against pathogens can be imagined. These particular antimicrobial peptides of human origin possess an high potential, since they could be easily used as therapeutic molecules due to the “virtual” absence of immunogenicity.

1.5 State of the art in the field of cryptic CAMPs' discovery

The search of hidden antimicrobial peptides is often conducted in a random way, by synthesizing a set of overlapping peptides which cover the entire amino acidic sequence of the protein of interest and testing their activity or by focusing instead on the analysis of the fragments generated by proteases. These experimental procedures can be expensive and time-consuming, and thus faster bioinformatic approaches, capable of highlighting the presence of a putative antimicrobial region, could prove extremely useful. Several tools have already been developed and a brief review of them will be given in this paragraph. Torrent M. *et al.,* (2009 and 2012) used an antimicrobial propensity scale of the different amino acids to screen protein sequences with a sliding window system; the scale was based on the study of the antimicrobial activity of all amino acid substitutions for each position of a 12-mer peptide, the bovine bactenicin 2A. This system was able to correctly identify the 80-90% of known antimicrobial domains and identified new domains previously uncharacterised in antimicrobial proteins. Brand G. D. *et al.,* (2012) developed Kamal, a software that uncover putative antimicrobial sequences from proteins based on physicochemical similarity to a sample of known antimicrobial peptides. Some of the physicochemical properties taken into account are net charge, isoelectric point, hydrophobicity, hydropathy and the propensity to aggregation; several new antimicrobial fragments were identified in different organisms. PeptideLocator (Mooney C. *et al.*, 2013) focus instead on the identification of a broad panel of bioactive peptides, which include also CAMPs, while Niarchou A. *et al.*, (2013) scanned plant proteins for putative antimicrobial regions using a machine learning approach based on physicochemical descriptors of the amino acids and trained on a set of known CAMPs; sequences whose antimicrobial probability exceeded 90% were stored in a database.

It has to be noticed that an arbitrary choice of physicochemical parameters taken into account can lead to a biased selection of putative antimicrobial peptides, that does not consider their extreme variation in nature; the method by Torrent M. *et al.*, (2009 and 2012) could instead be biased by the choice of a single peptide for the study of the amino acids

substitutions.

1.6 Aims

The main aims of this research work are:

I. development of a novel bioinformatic tool allowing the identification of putative antimicrobial peptides inside human proteins;
II. development and optimization of a novel system for the recombinant expression of newly identified CAMPs, followed by the biological and structural characterization of peptides;
III. modelling of CAMPs by a Monte Carlo strategy with implicit solvents, in order to define the parameters which better reproduce the experimentally derived structures, thus paving the way to the development of strategies for *ab initio* modelling of CAMPs.

2. MATERIALS AND METHODS

BIOINFORMATIC SECTION

2.1 Development of the scoring system for the identification of hidden CAMPs

The “antimicrobial scores” of a peptide were calculated using the following formulas:

$$\textbf{Relative score (RS)} = (C^m H^n) / \textbf{MaxScore}$$

$$\textbf{Absolute Score (AS)} = \textbf{RS} \times L^s$$

Where

- **C** is the net charge of the peptide calculated by the algebraic sum
# of Arg + # of Lys - # of Glu - # of Asp + 1 (if the N-terminus is a free amino group) - 1 (if the C-terminus is a carboxylic group).

- **H** is the arithmetical sum of the hydrophobicity scores of all the residues of the peptide (taken from the scales derived from HPLC retention times as described in the Results and Discussion section).

- **L** is the number of residues in the peptide.

- **MaxScore** (maximum score) is the highest ($C^m H^n$) value obtainable for a peptide at given values of the coefficients *m* and *n*. Maximum scores where obtained by calculating the scores of all the possible peptides composed exclusively by Arg residues and the residue with the highest hydrophobicity score in the chosen hydrophobicity scale (Ile, in the case of Cowan’s scales; Phe, in the case of Monera’s scales; Trp, for all the other hydrophobicity scales).

More in detail, indicating with **#R** the number of arginine residues in the peptide

$0 \leq \#R \leq L$;

the number of hydrophobic residues, **#H**, will be exactly **#H = L - #R**;

the number of possible peptides will be **L+1** (only the composition of the peptides and not their primary sequence is considered);

The absolute maximum of the product ($C^m H^n$) is obtained when C = $\boldsymbol{m/(m+n)}$ and H = $\boldsymbol{n/}$

(*m* + *n*), however, as the ratios ***m*/(*m* + *n*)** and ***n*/(*m* + *n*)** can assume non integer values and the charge C can, obviously, only be an integer (C = #R + 1, for a peptide with a free amino terminal group and an amidated C-terminus) the highest (C^mH^n) value for a peptide will be obtained when **C** is as close as possible to ***m*/(*m* + *n*)**.
It should be noted that using hydrophobicity scales which do not assign a hydrophobicity score to arginine (see Results and Discussion section) the highest scoring peptide can indifferently contain arginine or lysine residues, whereas using hydrophobicity scales which assign a hydrophobicity score to arginine, arginines-containing peptides will have higher scores than lysine containing peptides.

- Coefficients ***m*** and ***n*** are strain dependent variables that were calculated correlating RS or AS values and experimental potency data of a selected peptide set (described in the Results and Discussion section) on a defined strain through the use of the linear regression option of Microsoft Excel. Experimental potency values were calculated as Log(1000/Effective Concentration) where the "effective concentration" can be the *half maximal inhibitory concentration* (IC_{50}) or the *minimum inhibitory concentration* (MIC) i.e. the lowest concentration that inhibits the visible growth of the microorganism. RS or AS values were calculated setting to 1 the initial values of *m* and *n* and calculating the R^2 value, hence the *m* and *n* values were manually changed and R^2 value re-calculated. By using this iterative procedure we defined the combination of *m* and *n* values providing the highest R^2 value. Coefficients ***m*** and ***n*** were calculated using two peptide sets described by Fjell C. D. *et al.*, (2009), RANDOM200 and RANDOM19 peptide sets, (described in details in the Results and Discussion section). The sequences of the RANDOM19 peptide set are reported in table 1.

- Coefficient ***s*** is a strain dependent variable that describes the dependence of the antimicrobial potency from the length of antimicrobial peptides. It was calculated by correlating AS values and the experimental potency data of a set composed by ten peptides of similar composition but different length (Wiradharma N. *et al.*, 2011) measured on *Bacillus subtilis* through the use of the linear regression option of Microsoft Excel. We used the same iterative procedure described for the determination of *m* and *n* values.

As defined, RS can assume all the values from 0 to 1, whereas AS can assume all the values from 0 to L^s.

Table 1: primary structures of the RANDOM19 set from Fjell C. D. *et al.*, (2009).

Primary structure
RLARIVVIRVAR
KIWWWWRKR
RWRRWKWWL
WRWWKIWKR
WKRWWKKWR
WKKWWKRRW
FRRWWKWFK
LRWWWIKRI
RKRLKWWIY
KKRWVWIRY
KWKIFRRWW
RKWIWRWFL
IWWKWRRWV
RRFKFIRWW
AVWKFVKRV
AWRFKNIRK
KRIMKLKMR
AIRRWRIRK
VVLKIVRRF

2.2 Validation of the scoring function

The validation of the scoring function was performed through the window analysis of a set of proteins with known antibacterial domains using the sets of exponents determined for *Staphilococcus aureus* ATCC 25923 and *Pseudomonas aurginosa* H103, the “Parker” scale zeroed at glycine (see Results and Discussion section) and a window size from 12 to 40 residues. The analysis was conducted using Microsoft Excel.

EXPERIMENTAL SECTION

2.3 Materials

Ampicillin, bovine serum albumin (purity > 97%), IPTG, urea, betaine, DTT, β-mercaptoethanol, guanidine chloride, agar were purchased from Sigma-Aldrich. Trypton was purcahsed from Applichem, yeast extract from Becton Dickinson. Sodium chloride and acrylamide (40% stock solution) were from Applichem.

2.4 General procedures

Cell transformation and Luria-Bertani medium preparation were performed according to Sambrook J. *et al.*, (1989). SDS-PAGE was carried out according to Laemmli U. K. (1970). Protein concentrations were determined by the method of Bradford, using BSA as the standard (Bradford M. M., 1976) and by UV spectroscopy using the theoretical, sequence-based extinction coefficients in table 2 (Gill S. C. and von Hippel P. H.,1989).

Table 2: sequence-based extinction coefficients of the optimized fusion construct and ThrAP.

	Extinction coefficient (M^{-1} cm^{-1})
ONC-DC*less*-HIS-ThrAP	24410
ThrAP	8480

2.5 Preparation of the semi-defined rich medium (SDRM)

The novel liquid growth medium was prepared by dissolving in one litre of deionized water 34 g of trypton, 12 mL of glycerol, 3 g of citric acid($1H_2O$), 2.31 g of KH_2PO_4, 12.54 g of K_2HPO_4, 4 g of glucose, 3 mL of NH_3 25%, 1 mL of betaine 1 M and 5 mL of a solution of micro-nutrients with the following composition expressed in g/L: 5.4 g of MgO, 1 g of $CaCO_3$, 0.72 g of $ZnSO_4(7H_2O)$, 0.56 g of $MnSO_4(H_2O)$, 0.125 g of $CuSO_4(5H_2O)$, 0.14 g of $CoSO_4(7H_20)$, 0.03 g of H_3BO_3, 25.6 mL of HCl, 30.1 of $MgSO_4$ (0.25M), 2.25 g of $FeSO_4(7H_2O)$ (Fe^{2+}16 mM), 2.502 g of $FeSO_4(7H_2O)$, 0.004 g of $NiCl_2(6H_2O)$, 0.006 g of $Na_2MoO_4(2H_2O)$. The pH of the medium was adjusted to 7.5 with NH_3 25%.

2.6 Heterologous expression and preliminary purification of the fusion construct

Escherichia coli BL21(DE3) cells (AMS Biotechnology) were used for recombinant protein expression; cells transformed with pET 22b(+)-fusion construct were grown in 1 liter of SDRM containing ampicillin (0.1 mg/mL). When the culture reached an $A_{600\ nm}$ of 3 OD unit, protein expression was induced by the addition of 0.4 mM IPTG and the bacterial culture was grown over-night. Cells were harvested by centrifugation (6000 rpm, 4°C, 10', JA-14 rotor, Beckman) and pellets were lysed by sonication in lysis buffer (0.1 M Tris-HCl, pH

7.4, containing 10 mM EDTA) at a final concentration of 100 OD/mL in an ultrasonic liquid processor (Misonix Ultrasonic Processor XL) with 30" impulses, each followed by a 30" rest, for a 30' total time, at 20 kHz. The suspension was then centrifuged at 12000 rpm for 30' at 4°C (JA-25.50 rotor, Beckman). The inclusion bodies were freed from membrane proteins by three washes in 0.1 M Tris-HCl, pH 7.4, containing 10 mM EDTA, 2 % Triton X-100 and 2 M urea, followed by repeated washes in 0.1 M Tris-HCl pH 7.4, containing 10 mM EDTA, to eliminate traces of Triton and urea. This procedure eliminated several contaminant proteins and cellular debris entrapped in inclusion body pellets. Inclusion bodies of the fusion constructs without His●tag® where dissolved in 0.1 M Tris-HCl pH 8, 10 mM EDTA, 6 M GuHCl and 25 mM DTT at a final concentration of 10 mg/mL, purged with N_2, and incubated at 37°C for 3 h. The protein solution was acidified to pH 5 with glacial acetic acid and extensively dialyzed against 0.1 M acetic acid (pH 3) at 4°C. Any insoluble material was removed by centrifugation (12000 rpm, 30', 4°C, rotor JA-25-50, Beckman). Inclusion bodies of the fusion constructs containing the His●tag® where purified with IMAC.

2.7 Immobilized metal-affinity chromatography (IMAC)

Inclusion bodies, following the preliminary washes, were dissolved in 50 mM Tris-HCl pH 8, 6 M GuHCl and 20 mM β-mercaptoethanol (binding buffer) at a final concentration of 8-10 mg/mL, purged with N_2, and incubated at 37°C for 3 h. Denatured and reduced inclusion bodies were over-night incubated in batch at 4°C with the chromatographic resin Ni-NTA Agarose (Quiagen), previously equilibrated in the binding buffer, under continuous stirring. Following binding of the tagged protein, the resin was extensively washed in batch with 50 mM Tris-HCl pH 8, 6 M GuHCl and 10 mM β-mercaptoethanol (wash buffer), at 4°C and under continuous stirring. The resin was finally packed into a column and the elution was performed by lowering the value of pH from 8 to 5 using sodium-acetate 0.1 M pH 5, 6 M GuHCl and 10 mM β-mercaptoethanol (elution buffer). The fractions of interest were pooled and extensively dialyzed against 0.1 M acetic acid (pH 3) at 4°C. Any insoluble material was removed by centrifugation (12000 rpm, 30', 4°C, rotor JA-25-50, Beckman) and through filtration with 0.2 μm filters (Corning).

2.8 Self-cleavage of the fusion construct and isolation of the peptide

The solution containing the fusion construct was acidified to pH 2 with HCl, purged with N_2, and incubated at 60°C for 24 h in a water bath. Following the cleavage, the solution was alkalized to pH 7-7.2 with the addition of NH_3 1 M, purged with N_2 after the addition of 2 mM β-mercaptoethanol and over-night incubated at 28°C in a water bath. The peptide was isolated from the insoluble components through repeated cycles of centrifugation (12000 rpm, 30', 4°C, rotor JA-25-50, Beckman) and was finally lyophilized. The purity of the peptide was checked through SDS-PAGE and mass spectrometry.

2.9 Bactericidal assays

The bactericidal assays were conducted by D. Anna Zanfardino (Department of Biology, University Federico II). A single colony of the different bacterial strains was re-suspended in 5 mL of TSA medium (Becton Dickinson) and over-night incubated at 37°C and 150 rpm. When the culture reached an $A_{600\ nm}$ of 1 OD unit, it was diluted to 1:1000 in NaP 20 mM, pH 7.0 buffer. Samples with a finale volume of 1 mL were then prepared; the bacterial cells constituted the 4% of the volume, and the different proteins at various concentrations were added, with 20 mM NaP, pH 7.4 buffer used to reach the final volume. The positive controls were represented by cells incubated without protein and with BSA at the same concentrations of the proteins tested, while instead the negative control was obtained by incubating the cells with ampicillin (0.05 mg/mL). Samples were incubated at 37°C and 150 rpm for 4 hours; serial dilutions (1:100, 1:1000) of all the samples were plated on solid TSA and the Petri dishes were over-night incubated at 37°C. The next day the amount of survived cells was estimated, by counting the number of colonies on each Petri dish and comparing it with the controls.

2.10 Circular dichroism measurements

Circular dichroism (CD) spectra were recorded with a Jasco J-715 spectropolarimeter at room temperature. The molar ellipticity per mean residue, [θ] in degrees square centimeters per decimole, was calculated from the equation $[\theta] = ([\theta]_{obs}\ mrw)/(10\ l\ C)$, where $[\theta]_{obs}$ is the ellipticity measured in degrees, mrw is the mean residue molecular weight (124.2 for ThrAP and 125.68 for ApoE-AP), C is the peptide concentration in grams

per milliliter, and *l* is the optical path length of the cell in centimeters. Cells with path lengths of 0.1 cm were used and CD spectra were recorded with a time constant of 4 s, a 2 nm bandwidth, and a scan rate of 20 nm/min; the signal was averaged over at least three scans and baseline corrected by subtraction of a buffer spectrum. Spectra were analyzed for secondary structure content using the PEPFIT tool (Reed J. and Reed T. A., 1997). Peptide concentrations were typically 35-40 μM (or 10 μM where specified) in 10 mM NaP pH 7.4 in the presence or absence of LPS and alginate, or in water with different concentrations of TFE or SDS 20 mM. *Escherichia coli* LPS 0111:B4 and seaweed alginate (Sigma Aldrich) were used at a final concentration of 0.2 mg/mL.

COMPUTATIONAL SECTION

2.11 Monte Carlo simulations

CAMPs NMR structures were downloaded from PDB and the first structure of each ensemble was chosen as starting conformation. The conformational space of peptides was explored by Monte Carlo (MC) simulations using the ZMM software (http://www.zmmsoft.com/) and the AMBER force field (Weiner S. J. *et al.*, 1984). Initial structures were subjected to 10000 MC steps, each followed by 2000 iterations of energy minimization, at constant temperature (T = 300 K). Simulations were performed in vacuum, implicit water, with the Effective Energy Function-1, EEF-1, (Lazaridis T. and Karplus M., 1999), implicit octanol (Hopfinger A. J. and Battershell R. D., 1976) and implicit octanol with an attenuation of solvation energy, obtained through the modification of the SEC (Solvation Energy Coefficient) parameter of ZMM. SEC values used for the analysis were 1.0, 0.5 and 0.25 corresponding to full octanol solvation energy, one half and one fourth of the octanol solvation energy, respectively. In all the calculations, a distance-dependent dielectric permeability ε = 4 x r and a cut off at the distance of 8 Å for non-bonded interactions were used. Secondary structure, structural alignments and ASAs were analyzed using Swiss-PDBViewer; images were produced with the PyMol and Swiss-PDB Viewer.

3. RESULTS AND DISCUSSION

BIOINFORMATIC SECTION

3.1 A novel scoring function for the identification of cryptic CAMPs

Several researchers have attempted to develop methods to predict the antimicrobial efficacy of CAMP or the presence of CAMP-like sequences inside the primary structure of (large) proteins that we call for simplicity “cryptic CAMPs”. All these methods include very large sets of properties (up to 23 molecular descriptors in a QSAR study Fjell C. D. *et al.*, (2009)) and sometime arbitrary chosen properties, like tendency to form amyloid structures: even if few well-known amyloid peptides have antimicrobial activity (Kagan B. L. *et al.*, 2012), at the moment there is no indication that the ability to form amyloid fibrils is a general requirement. Moreover, all the cited studies ignore the fact that if a correlation does exist between sequence and/or structure of CAMPs and their antimicrobial activity, this correlation could be strain specific. Even if membranes of bacteria share some molecular features as a net negative charge on both sides of the membrane, each strain has its peculiar composition: the abundance of the negatively charged phospholipids (prevalently cardiolipin and phosphatidylglycerol) can vary from about 20% (e.g. in *E. coli*) to almost 100% (e.g. in *Staphylococcus* and *Streptococcus*). Furthermore, even the proportion of cardiolipin and phosphatidylglycerol varies from strain to strain (Epand R. M. and Epand R. F., 2009).

In order to derive a set of functions able to predict the antimicrobial activity of peptides on specific strains, we started from the molecular model of a crucial step common to all the models of antimicrobial activity (Figure 7): the CAMP adopts an amphipathic structure and inserts into the membrane, parallel to the membrane surface. Two major forces stabilize this complex: the electrostatic interaction between anionic phospholipids and positively charged residues on the peptide and the hydrophobic interaction between hydrophobic residues of the peptide and the fatty acid chain of phospholipids. It is well-known that anionic phospholipids encircle the CAMP, a phenomenon known as “phase separation”, therefore, the interaction CAMP/membrane, from the electrostatic point of view, resembles the interaction between a poly-anion and a poly-cation. Assuming that the electrostatic and the hydrophobic components act synergistically to increase the stability of the complex and that antimicrobial activity increases proportionally to the stability of the complex, then, the

antimicrobial activity of a CAMP should be proportional to the product:

$$C^m H^n$$

where, C is a measure of the electrostatic attraction (e.g net charge of the CAMP), H is a measure of the hydrophobic interaction contribution and exponents *m* and *n* determine the relative contribution of the two forces to the stability of the complex CAMP/membrane. We want to underline that the exponents *m* and *n*, likely, are not "universal" and could be unique for each strain depending on the relative abundance of negatively charged lipids but also on other peculiarities of bacterial strains like, for example, the properties of the fatty acids (length, double bonds, ramifications, cyclopropane rings, etc.).

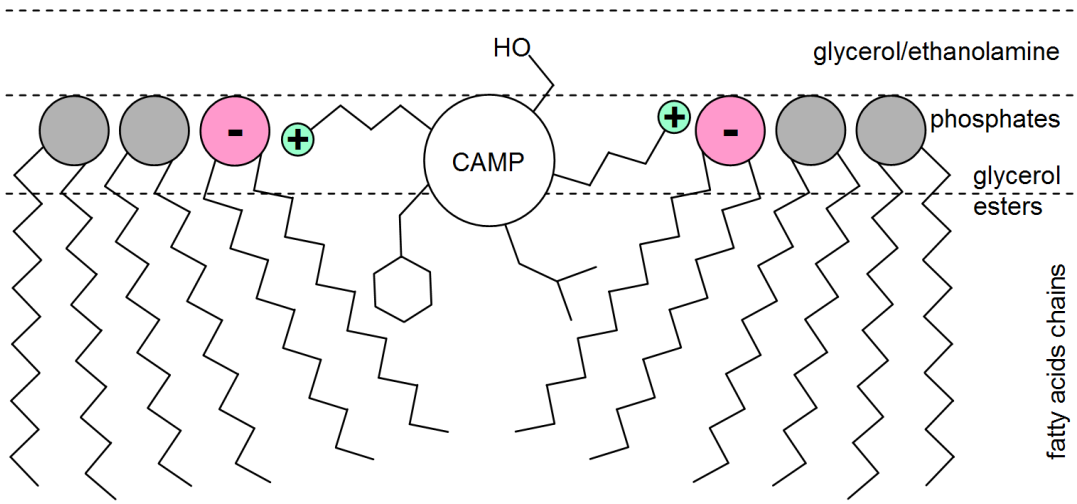


Figure 7: representation of the universal model of interaction between a CAMP and the bacterial membrane.

The product $C^m H^n$ is particularly well suited to describe any possible relative contribution of charge and hydrophobicity. Figure 8 (A and B) shows the dependence of the $C^m H^n$ product from exponents *m* and *n* for a set of model 10 aa long peptides composed only by arginine and tryptophan. For each peptide, C was calculated as the sum of the number of arginine residues (therefore, C is the net charge of the peptides), whereas H was calculated assigning arbitrarily a "hydrophobicity score = 1" to each tryptophan residues and considering additive the contributions of tryptophan residues (therefore, H the sum of the number of tryptophan residues). The arbitrariness of the choice is compensated by the exponents, in fact, if the exponents are identical then the highest scoring peptides are the peptides with five arginines and five tryptophan residues, whereas, if $n>m$, then the highest scoring peptides are the peptides with more tryptophan residues than arginine

residues and vice versa. For example if *m*=0.6 and *n*=1.4, then the highest scoring peptides have seven tryptophan and five arginine residues respectively. It should be noted that the percentage of tryptophan residues in the highest scoring peptides is simply given by the ratio n/(m+n). Moreover, at a fixed ratio n/m, the increase in the sum m+n causes an increase in the steepness of the score curve, i.e. the relative score of the non-optimal peptides decreases (compare the black, red and blue curves in figure 8).

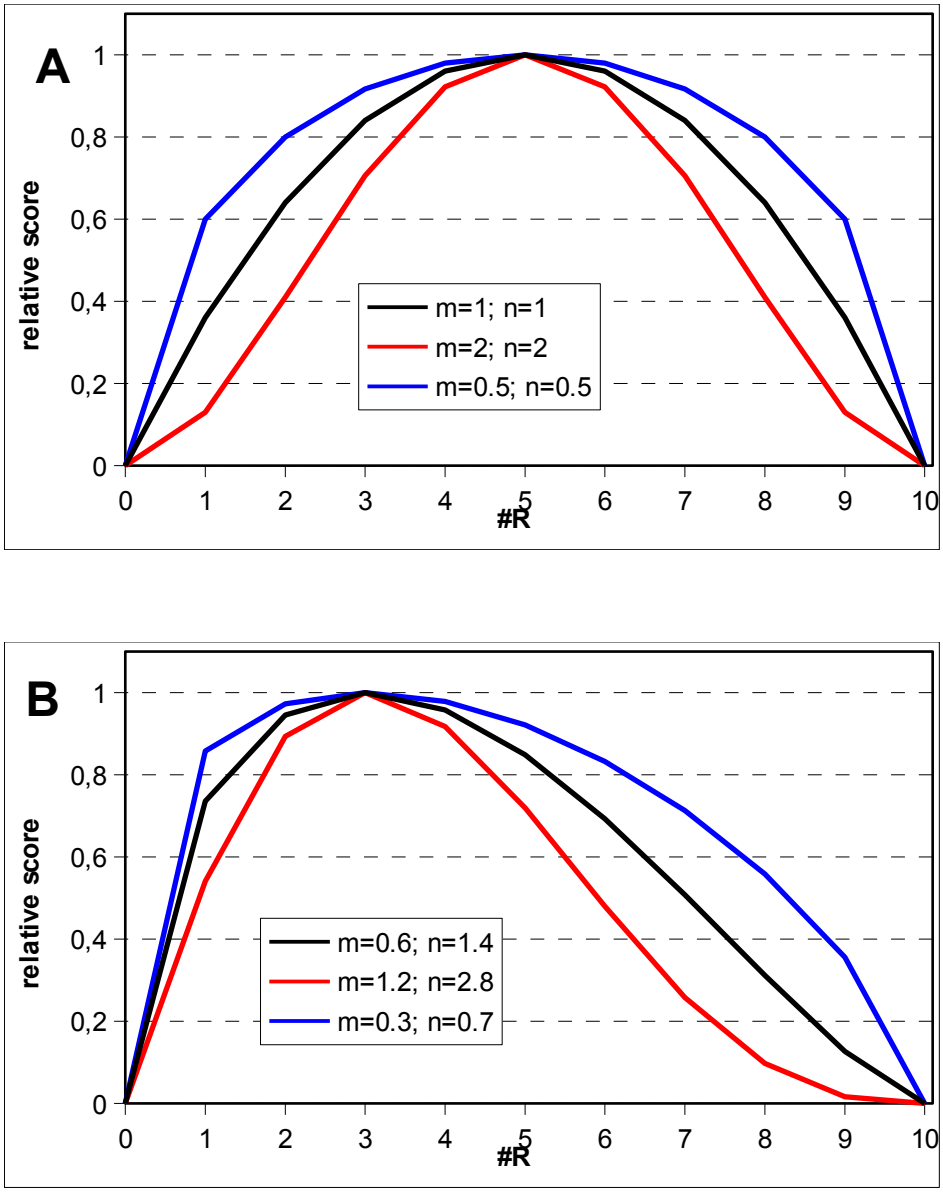


Figure 8: dependence of the C^mH^n product (relative score) from exponents *m* and *n* for a set of model 10 aa long peptides composed only by arginine and tryptophan.

Our calculation contains another arbitrary assumption: independently from the sequence,

all the arginine and tryptophan residues provides the same contribution. This is equivalent to assuming that each peptide, independently from the sequence, adopts a perfectly amphipathic conformation with all the tryptophan and arginine residues oriented in a productive way to contribute to the binding. This assumption is not unlikely for short peptides which are notoriously flexible but will not be true for peptides long enough to have a specific folding or for disulphide-rich peptides, whose folding is constrained by the disulphides.

In our simple example we have assigned a hydrophobicity score = 1 to tryptophan. More generally, it is necessary to derive a relative score of hydrophobicity for all the residue which are supposed to be placed on the hydrophobic side of the CAMP bound to the membrane. An impressive number of hydrophobicity scale have been published, however, as we need to score the ability of a side-chain inserted into a peptide framework to interact with the fatty acid chains of phospholipids, we have selected few scales (Table 3) derived by measuring retention times on C18 HPLC columns at pH 7 (in phosphate buffer) of free amino-acids or specific peptide libraries (Table 4). These scales, usually derived to predict the retention times of peptides in HPLC, are particularly well suited to estimate the relative contribution to membrane binding of hydrophobic amino-acids.

Table 3: hydrophobicity scales normalized between 0 and 1.

	Cowan	Kovacs(a)[a]	Kovacs(b)[a]	Parker	Monera	AVE2[b]	AVE3[c]
Trp	0.879	1.000	1.000	1.000	0.983	1.000	1.000
Phe	0.965	0.916	0.931	0.959	1.000	0.962	0.969
Leu	0.992	0.76	0.792	0.959	0.983	0.976	0.916
Ile	1.000	0.707	0.74	0.902	0.990	0.950	0.882
Met	0.817	0.551	0.59	0.711	0.833	0.763	0.715
Val	0.872	0.486	0.538	0.686	0.843	0.778	0.693
Tyr	0.460	0.514	0.549	0.597	0.760	0.616	0.639
Cys	0.731	0.318	0.382	0.432	0.670	0.580	0.497
Pro	0.751	0.355	0.422	0.397	0.173	0.573	0.333
Ala	0.628	0.174	0.266	0.397	0.620	0.512	0.430
His	0.377	0.190	0.266	0.397	0.403	0.386	0.357
Arg	0.163	0.174	0.338	0.289	0.263	0.227	0.298
Thr	0.472	0.174	0.243	0.241	0.437	0.356	0.309
Gln	0.307	0.103	0.182	0.216	0.290	0.253	0.231
Lys	0.153	0.000	0.266	0.200	0.207	0.184	0.225
Gly	0.540	0.056	0.182	0.200	0.357	0.378	0.248
Ser	0.382	0.090	0.171	0.175	0.323	0.278	0.224
Asn	0.291	0.084	0.165	0.149	0.173	0.221	0.163
Glu	0.050	0.044	0.012	0.108	0.157	0.080	0.093
Asp	0.000	0.034	0.000	0.000	0.000	0.000	0.000

[a] the two scales were derived using the same set of peptides and the same experimental conditions (phosphate buffer pH 7.0) except that in the case of the Kovacs(b) scale 0.1 M $NaClO_4$ was added to the buffer (Kovacs J. M. *et al.*, 2006).

[b] this scale is an average of the scales of Cowan (Cowan R. and Whittaker R. G., 1990) and Parker (Parker J. M. R. *et al.*, 1986). Trp value was arbitrarily set to 1.

[c] this scale is an average of the scales of Kovacs(b), Parker and Monera (Monera O. D. *et al.*, 2005). The scales of Parker and Monera, like the scale Kovacs(b), were obtained with buffers containing $NaClO_4$.

Table 4: peptide sets used to derive the hydrophobicity scales.

Scale	Peptide library[a]	notes
Cowan	NH_2-**X**-COOH	amino-acids
Kovacs	Ac–**X**GAKGAGVGL–amide	random coil peptides
Parker	Ac–G**XX**LLLKK–amide	random coil peptides
Monera	Ac–EAEKAAKE**X**EKAAKEAEK–amide	helical peptides

[a] X denotes any of the twenty canonical amino-acids.

The scale described by Monera *et al*. was derived using an helical peptide and is strongly influenced by structural effects, for example proline, an helix-breaking residue, in this scale is a very hydrophilic residues. The scales of Kovacs, on the other hand, are influenced by the very high solvent exposure of the variable residue. The scale of Parker is intermediate, presenting two adjacent copies of the variable residue, one more exposed at the N-terminus of the peptide and one packed between the preceding residue and a leucine

residue. Not surprisingly, a scale obtained by averaging the previous three scales is very similar to the scale of Parker and co-workers. Finally, the scale of Cowan, being based on the retention times of free amino-acids, shows some peculiarities as the fact that aliphatic residues isoleucine and leucine are the most hydrophobic and proline and glycine are much more hydrophobic than in the other scales. We have also prepared an average scale between the scales of Cowan and Parker.

It should be remembered that, according to our model, the H value in the product C^mH^n is the sum of the hydrophobic contribution of the residues located on the hydrophobic face of the CAMP. Therefore, the scales in table 3 are just the starting point to derive the hydrophobicity scores of the residues that will be frequently located on the hydrophobic face of the CAMP, thus contributing to the binding. As a score = 0 can be arbitrarily assigned to different residues, each scale can give rise to more than one score list as shown in tables 4 A-G.

Table 4-A: hydrophobicity scales derived from the Cowan's scale

	Cowan	**Cowan-Ser0**	**Cowan-Gly0**
Ile	1.000	1.000	1.000
Leu	0.992	0.988	0.984
Phe	0.965	0.943	0.923
Trp	0.879	0.805	0.738
Val	0.872	0.793	0.721
Met	0.817	0.703	0.601
Pro	0.751	0.598	0.459
Cys	0.731	0.565	0.415
Tyr	0.638	0.415	0.213
Ala	0.628	0.398	0.191
Gly	0.540	0.256	0
Thr	0.472	0.146	0
Ser	0.382	0	0
His	0.377	0	0
Gln	0.307	0	0
Asn	0.291	0	0
Arg	0.163	0	0
Lys	0.153	0	0
Glu	0.050	0	0
Asp	0.000	0	0

Table 4 B: hydrophobicity scales derived from the Parker's scale.

	Parker	Parker-Ser0	Parker-Gly0	Parker-Arg0
Trp	1.000	1.000	1.000	1.000
Leu	0.959	0.952	0.949	0.944
Phe	0.959	0.952	0.949	0.944
Ile	0.902	0.879	0.873	0.859
Met	0.711	0.648	0.631	0.592
Val	0.686	0.618	0.599	0.556
Tyr	0.597	0.509	0.484	0.430
Cys	0.432	0.309	0.274	0.197
Ala	0.397	0.267	0.229	0.148
Pro	0.397	0.267	0.229	0.148
His	0.397	0.267	0.229	0.148
Arg	0.289	0.139	0.096	0
Thr	0.241	0.079	0.032	0
Gly	0.216	0.048	0	0
Lys	0.200	0.048	0	0
Gln	0.200	0.030	0	0
Ser	0.175	0	0	0
Asn	0.149	0	0	0
Glu	0.108	0	0	0
Asp	0	0	0	0

Table 4-C: hydrophobicity scales derived from the Kovacs's scale (a).

	Kovacs(a)	Kovacs(a)-Glu0	Kovacs(a)-Gln0	Kovacs(a)-Arg0
Trp	1.000	1.000	1.000	1.000
Phe	0.916	0.912	0.906	0.898
Leu	0.760	0.749	0.733	0.709
Ile	0.707	0.694	0.674	0.645
Met	0.551	0.531	0.500	0.457
Tyr	0.514	0.492	0.458	0.411
Val	0.486	0.463	0.427	0.377
Pro	0.355	0.326	0.281	0.219
Cys	0.318	0.287	0.240	0.174
His	0.190	0.153	0.097	0.019
Ala	0.174	0.137	0.080	0
Thr	0.174	0.137	0.080	0
Arg	0.174	0.137	0.080	0
Gln	0.103	0.062	0	0
Ser	0.090	0.049	0	0
Asn	0.084	0.042	0	0
Gly	0.056	0.013	0	0
Glu	0.044	0	0	0
Asp	0.034	0	0	0
Lys	0	0	0	0

Table 4-D: hydrophobicity scales derived from the Kovacs's scale (b).

	Kovacs(b)	Kovacs(b)-Ser0	Kovacs(b)-Gln0
Trp	1.000	1.000	1.000
Phe	0.931	0.916	0.915
Leu	0.792	0.749	0.746
Ile	0.740	0.686	0.682
Met	0.590	0.505	0.498
Tyr	0.549	0.456	0.449
Val	0.538	0.443	0.435
Pro	0.422	0.303	0.293
Cys	0.382	0.254	0.244
Arg	0.338	0.202	0.191
Lys	0.266	0.115	0.102
Ala	0.266	0.115	0.102
His	0.266	0.115	0.102
Thr	0.243	0.087	0.074
Gln	0.182	0.014	0
Gly	0.182	0.014	0
Ser	0.171	0	0
Asn	0.165	0	0
Glu	0.012	0	0
Asp	0	0	0

Table 4-E: hydrophobicity scales derived from the Monera's scale.

	Monera	Monera-Ser0	Monera-Gln0
Phe	1.000	1.000	1.000
Ile	0.990	0.985	0.984
Trp	0.983	0.975	0.974
Leu	0.983	0.975	0.974
Val	0.843	0.769	0.757
Met	0.833	0.754	0.741
Tyr	0.760	0.645	0.627
Cys	0.670	0.512	0.487
Ala	0.620	0.439	0.409
Thr	0.437	0.168	0.125
His	0.403	0.118	0.073
Gly	0.357	0.050	0
Ser	0.323	0	0
Gln	0.290	0	0
Arg	0.263	0	0
Lys	0.207	0	0
Pro	0.173	0	0
Asn	0.173	0	0
Glu	0.157	0	0
Asp	0	0	0

Table 4-F: hydrophobicity scales derived from the scale “average 2”.

	AVE2	AVE2-Ser0	AVE2-Gly0
Trp	1.000	1.000	1.000
Leu	0.976	0.967	0.962
Phe	0.962	0.948	0.940
Ile	0.950	0.931	0.920
Val	0.778	0.693	0.644
Met	0.763	0.672	0.620
Pro	0.573	0.408	0.314
Cys	0.581	0.419	0.326
Tyr	0.617	0.469	0.384
Ala	0.512	0.323	0.215
Gly	0.378	0.137	0
His	0.386	0.149	0
Thr	0.356	0.108	0
Ser	0.278	0	0
Gln	0.253	0	0
Asn	0.221	0	0
Arg	0.227	0	0
Lys	0.184	0	0
Glu	0.080	0	0
Asp	0	0	0

Table 4-G: hydrophobicity scales derived from the scale “average 3”.

	AVE3	AVE3-Gln0	AVE3-Gly0	AVE3-Arg0
Trp	1.000	1.000	1.000	1.000
Phe	0.969	0.960	0.959	0.956
Leu	0.916	0.891	0.888	0.880
Ile	0.882	0.847	0.843	0.832
Met	0.715	0.629	0.621	0.594
Val	0.693	0.601	0.592	563
Tyr	0.639	0.531	0.520	0.486
Cys	0.497	0.346	0.331	0.283
Ala	0.430	0.259	0.242	0.188
His	0.357	0.164	0.145	0.084
Pro	0.333	0.133	0.113	0.050
Thr	0.309	0.101	0.081	0.016
Arg	0.298	0.087	0.066	0
Gly	0.248	0.022	0	0
Gln	0.231	0	0	0
Lys	0.225	0	0	0
Ser	0.224	0	0	0
Asn	0.163	0	0	0
Glu	0.093	0	0	0
Asp	0	0	0	0

As for the determination of the exponents *m* and *n*, these values are not only strain dependent, as already discussed above, but also condition dependent. In fact, the ionic strength at which the antimicrobial activity is assayed could influence the electrostatic component of the CAMP/membrane interaction and hence the relative contribution of the ionic and hydrophobic components. Therefore, the sole way to determine the values of the two exponents is to analyze the antimicrobial activity values of a set of peptides with very different composition (i.e. different ratios between hydrophobic and basic residues) determined using not only the same strain, but exactly the same assay. Fjell C. D. *et al.*, (2009) published the antimicrobial activity on *Pseudomonas aeruginosa* H103 (measured as IC_{50} by an assay based on luciferase) of a set of 200 peptides 9 aa long of (almost) random composition. Even if some amino-acids are not present (e.g. Asp, Glu, Pro, Cys) and other are underrepresented (e.g. Thr, His, Tyr) this set, that we will call the RANDOM200 set, provides the opportunity to verify if antimicrobial activity is correlated to the product C^mH^n.

The graph in figure 9 shows the **relative scores** (**RS**) of the RANDOM200 set as function of the **antimicrobial potency**. Antimicrobial potency was expressed as $Log(1000/IC_{50})$, whereas the relative scores were calculated by the equation:

$$\mathbf{RS = (C^mH^n) / MaxScore} \qquad \mathbf{(1)}$$

where **MaxScore** is the highest score that a peptide can obtain at given *m* and *n* values and, hence, corresponds to the score of the “optimal” CAMP.

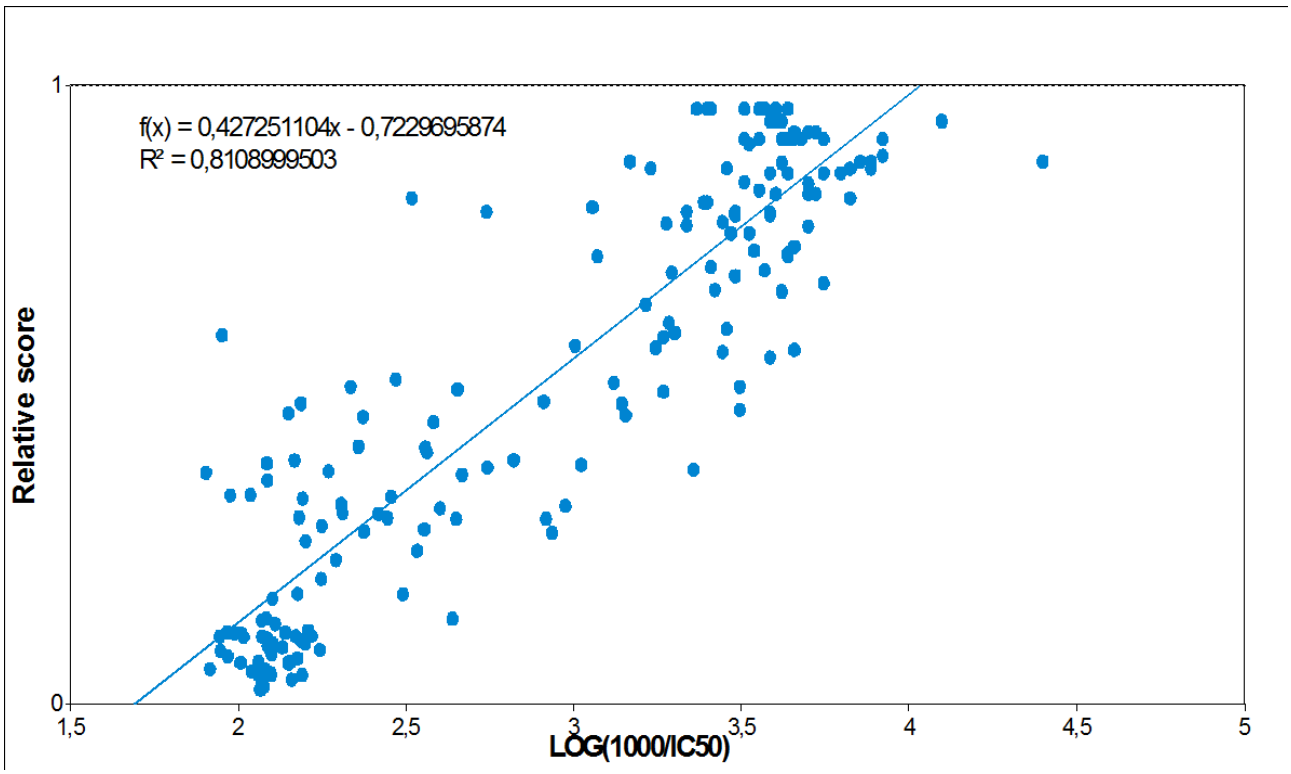


Figure 9: linear correlation between the relative scores and the antibacterial potency of the 200 peptides of the RANDOM200 set (Parker-Gly0 scale).

The *m* and *n* values shown in the graph were obtained setting the initial values of both exponents to 1 and then progressively changing them to find the values that maximise the R^2 value of the least squares line. As no systematic exploration was performed, we cannot exclude that we found *m* and *n* values corresponding to a local maximum, nonetheless the correlation between our RS and the experimental antimicrobial potency values is very good. All the hydrophobicity scales of tables 4 A-G performed almost equally well, with the not surprising exception of the Cowan's scale. The R^2, *m* and *n* values are summarized in table 5.

Table 5: values of the exponents *m* and *n* obtained through the linear fit with the RANDOM200 set.

Scale	R^2	*m*	*n*	*f*H[a]
Kovacs(a)-Arg0	0.833	1.30	1.30	0.50
Kovacs(a)-Gln0	0.825	1.33	1.50	0.53
Kovacs(a)-Glu0	0.816	1.32	1.60	0.55
Kovacs(b)-Gln0	0.813	1.30	2.00	0.61
Kovacs(b)-Ser0	0.814	0.92	1.52	0.62
Parker-Arg0	0.824	1.85	1.90	0.51
Parker-Gly0	0.812	1.88	2.14	0.53
Parker-Ser0	0.809	1.86	2.34	0.56
Monera-Gln0	0.806	2.88	2.88	0.50
Monera-Ser0	0.806	2.88	3.00	0.51
AVE2-Gly0	0.810	2.25	2.30	0.51
AVE2-Ser0	0.814	2.35	2.36	0.50
AVE3-Arg0	0.829	2.00	2.00	0.50
AVE3-Gly0	0.821	1.98	2.15	0.52
AVE3-Gln0	0.818	1.98	2.21	0.53
Cowan-Gly0	0.634	1.60	1.60	0.50
Cowan-Ser0	0.656	2.35	2.30	0.49

[a] *f*H, the ratio *n*/(*n*+*m*), is the fraction of hydrophobic residues in the highest scoring peptides.

All the peptides of the RANDOM200 set have the same length, allowing an easy comparison of their score and potency. Wiradharma N. *et al.*, (2011) have characterized another smaller set of peptides with molecular features complementary to those of the RANDOM200 set. Their set is composed by peptides of very similar composition but of different lengths of general sequence $(XXYY)_n$ where n is 2, 3 or 4, X is Phe, Leu, Ala and Y is Arg or Lys.

The graph in figure 10 shows the **absolute scores** (**AS**) of the peptides as function of their **antimicrobial potency**. Antimicrobial potency was expressed as Log(1000/MIC), whereas the absolute scores were calculated by the equation:

$$\mathbf{AS = RS \times L^{s}} \qquad \mathbf{(2)}$$

With all the hydrophobicity scales the best correlation was obtained with $s = 1$. Using the Parker-Arg0 or the AVE2-Ser0 scales we obtained $R^2 = 0.85$ with the complete set of ten peptides and $R^2 = 0.95$ by omitting the peptide $(LLKK)_2$ whose activity is slightly lower than that predicted by equation 2.

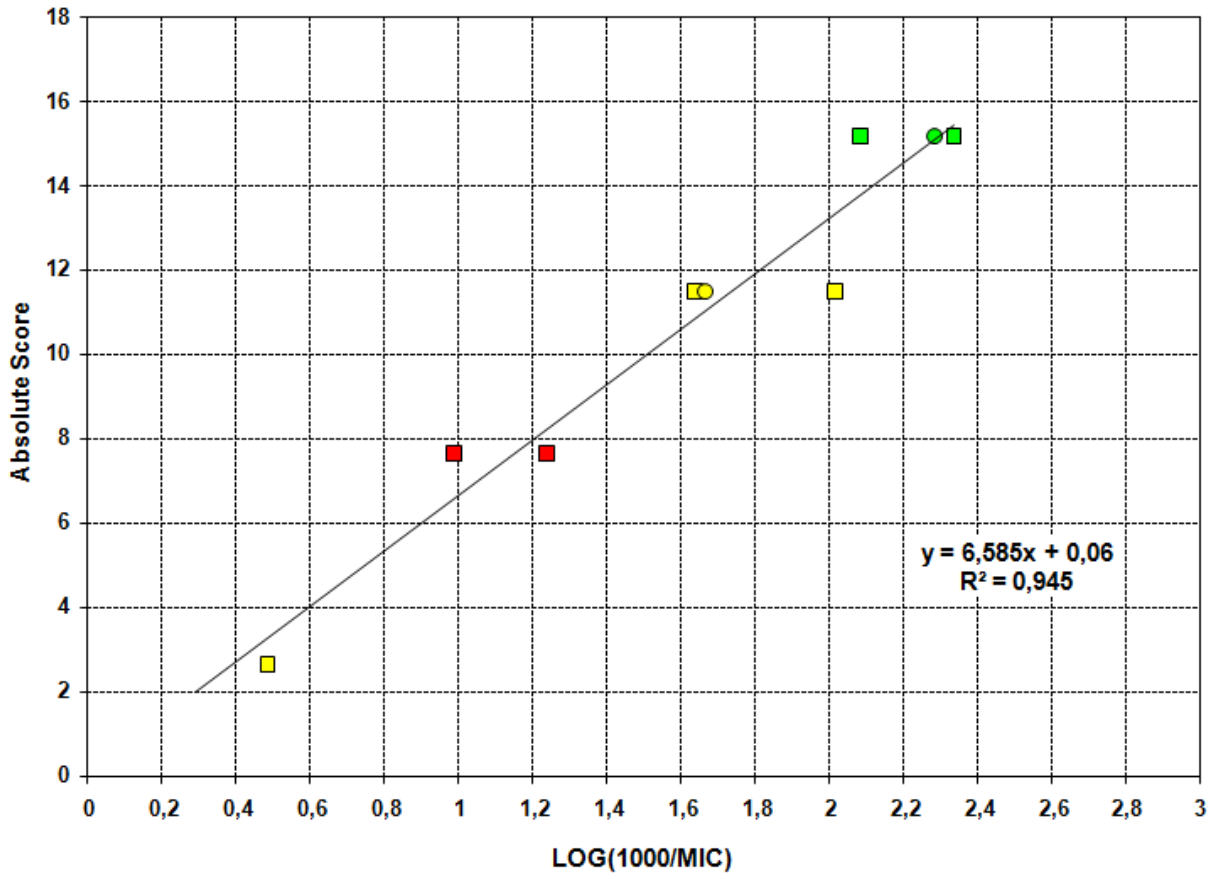


Figure 10: linear correlation between the absolute scores and the antimicrobial potency of the $(XXYY)_n$ peptides. In **yellow**: $(XXYY)_3$ peptides; in **red**: $(XXYY)_2$ peptides, in **green**: $(XXYY)_4$ peptides (Parker-Arg0 scale).

These results clearly indicate that, in a pool of peptides of similar composition (hence similar RS), the antimicrobial potency increases linearly with the length of the peptides (at least up to about 16 residues).

Using equation 2 in the simplified form:

$$\mathbf{AS = RS \times L} \qquad \mathbf{(3)}$$

we have analyzed a second data set reported by Fjell and co-workers. This set, hereafter called RANDOM19, is composed by 18 representative peptides from the RANDOM200 set and bactenicin 2A (Bac2A), a natural CAMP 12 residues long. Fjell and co-workers measured the MIC values of the 19 peptides on twenty strains, including seven strains of *P. aeruginosa,* two of *Staphilococcus aureus*, two of *Escherichia coli* etc; interestingly, some of the strain are clinical isolates. For almost all the strains examined we found a very good linear correlation between the absolute scores calculated by equation 3 and the experimental potency, Log(1000/MIC). Figure 11 (A and B) and tables 6-ABC show some

representative results.

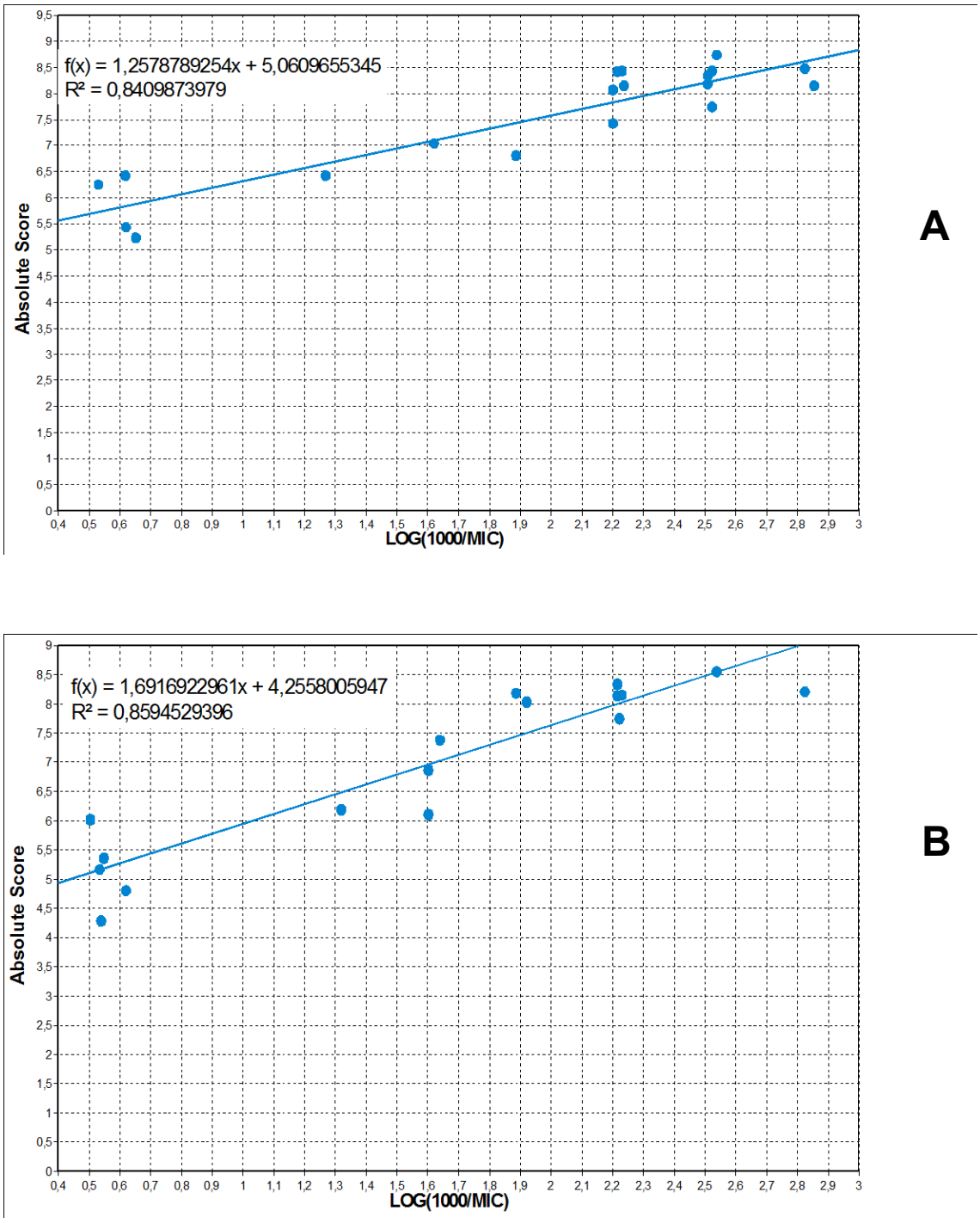


Figure 11: linear correlation between the absolute scores and the antimicrobial potency of the RANDOM19 peptides against *Staphilococcus aureus* ATCC 25923 (**A**) and *Pseudomonas aeruginosa* H103 (**B**) (Parker-Gly0 scale).

Table 6-A: values of the exponents obtained for three strains of *Pseudomonas aeuruginosa.*

	***P. aer.* H103**				***P. aer.* Brazil 9**				***P. aer.* LES400**			
	m	***n***	***f*H[a]**	**R^2**	***m***	***n***	***f*H[a]**	**R^2**	***m***	***n***	***f*H[a]**	**R^2**
Parker-Ser0	0.97	1.62	0.63	0.856	1.76	2.62	0.60	0.849	1.10	1.94	0.64	0.803
Parker-Gly0	0.96	1.45	0.60	0.860	1.64	2.22	0.58	0.851	1.00	1.62	0.62	0.804
Parker-Arg0	0.73	1.04	0.59	0.860	1.53	2.02	0.57	0.848	0.87	1.32	0.60	0.802
Kovacs(a)-Glu0	0.47	0.99	0.68	0.906	0.80	1.44	0.64	0.878	0.52	1.23	0.70	0.851
Kovacs(a)-Gln0	0.38	0.81	0.68	0.904	0.67	1.20	0.64	0.875	0.42	0.99	0.70	0.847
Kovacs(a)-Arg0	0.29	0.60	0.67	0.895	0.50	0.86	0.63	0.866	0.34	0.88	0.72	0.835
AVE2-Ser0	1.36	1.71	0.56	0.834	3.20	3.76	0.54	0.822	1.73	2.30	0.57	0.774
AVE3-Gln0	0.86	1.30	0.60	0.881	2.04	2.86	0.58	0.866	1.20	1.94	0.62	0.825
AVE3-Gly0	0.86	1.30	0.60	0.883	2.03	2.81	0.58	0.866	1.00	1.62	0.62	0.825
AVE3-Arg0	0.73	1.05	0.59	0.880	1.57	2.13	0.58	0.860	0.87	1.35	0.61	0.820

[a] *f*H, the ratio *n*/(*n*+*m*), is the fraction of hydrophobic residues in the highest scoring peptides.

Table 6-B: values of the exponents obtained for two strains of *S. aureus* and *Escherichia coli*.

	***S. aureus* ATCC 25923**				***S. aureus* C623**				***E. coli* 63103**			
	m	***n***	***f*H[a]**	**R^2**	***m***	***n***	***f*H[a]**	**R^2**	***m***	***n***	***f*H[a]**	**R^2**
Parker-Ser0	0.92	1.18	0.56	0.838	0.90	1.20	0.57	0.862	0.65	1.03	0.61	0.722
Parker-Gly0	0.92	1.07	0.54	0.841	0.88	1.07	0.55	0.864	0.65	0.93	0.59	0.723
Parker-Arg0	0.73	0.79	0.52	0.846	0.72	0.83	0.53	0.862	0.55	0.75	0.58	0.713
Kovacs(a)-Glu0	0.55	0.75	0.58	0.906	0.51	0.75	0.60	0.862	0.36	0.69	0.66	0.745
Kovacs(a)-Gln0	0.48	0.63	0.57	0.907	0.44	0.62	0.58	0.914	0.31	0.59	0.66	0.738
Kovacs(a)-Arg0	0.41	0.49	0.54	0.901	0.37	0.50	0.57	0.902	0.25	0.46	0.65	0.724
AVE2-Ser0	1.16	1.20	0.51	0.812	1.12	1.20	0.52	0.834	0.85	1.02	0.55	0.680
AVE3-Gln0	0.91	1.06	0.54	0.863	0.86	1.06	0.55	0.882	0.64	0.93	0.59	0.729
AVE3-Gly0	0.85	0.98	0.54	0.866	0.80	0.97	0.55	0.882	0.62	0.88	0.59	0.727
AVE3-Arg0	0.75	0.82	0.52	0.867	0.72	0.83	0.54	0.878	0.55	0.75	0.58	0.718

[a] *f*H, the ratio *n*/(*n*+*m*), is the fraction of hydrophobic residues in the highest scoring peptides.

Table 6-C: values of the exponents obtained for *Klebsiella pneumoniae*, *Enterococcus faecium* and *Enterococcus faecalis*.

	***K. pneumonie* 63575**				***E. faecium* t62764**				***E. faecalis* f43559**			
	m	***n***	***fH*[a]**	**R^2**	***m***	***n***	***fH*[a]**	**R^2**	***m***	***n***	***fH*[a]**	**R^2**
Parker-Ser0	1.50	3.40	0.69	0.715	0.51	0.81	0.61	0.790	0.04	1.00	0.96	0.756
Parker-Gly0	1.50	3.10	0.67	0.714	0.43	0.60	0.58	0.775	0.05	0.95	0.95	0.754
Parker-Arg0	1.50	3.10	0.67	0.716	0.43	0.60	0.58	0.759	0.00	0.68	1.00	0.730
Kovacs(a)-Glu0	0.80	3.60	0.82	0.770	0.28	0.55	0.66	0.803	0.01	0.57	0.98	0.711
Kovacs(a)-Gln0	0.74	3.80	0.84	0.771	0.24	0.46	0.65	0.788	0.00	0.50	1.00	0.689
Kovacs(a)-Arg0	0.74	4.00	0.84	0.767	0.20	0.37	0.65	0.758	0.00	0.35	1.00	0.658
AVE2-Ser0	1.81	2.25	0.55	0.697	0.73	0.88	0.55	0.755	0.16	1.31	0.89	0.719
AVE3-Gln0	1.55	3.40	0.69	0.722	0.50	0.73	0.59	0.793	0.05	0.92	0.95	0.746
AVE3-Gly0	1.55	3.40	0.69	0.723	0.49	0.71	0.59	0.787	0.02	0.87	0.98	0.740
AVE3-Arg0	1.55	3.50	0.69	0.722	0.43	0.60	0.58	762	0.00	0.68	1.00	0.721

[a] *f*H, the ratio $n/(n+m)$, is the fraction of hydrophobic residues in the highest scoring peptides.

Among the different hydrophobicity scales, the worst R^2 values were obtained with the AVE2-Ser0 scale. Using the scales derived from Kovacs' scale we obtained the highest R^2 values, but *m* and *n* values significantly lower than those obtained with the other scales. However, we noticed that, increasing proportionally both *n* and *m,* the R^2 values obtained with these scales decreased very slightly so that for *n* and *m* values similar to those obtained with the other scales also the R^2 values were similar. In other words, using the scales derived from Kovacs' scale only the ratio between *n* and *m* is well defined, whereas their absolute values cannot be determined accurately. It should be noted that all the scales provide very similar *f*H values for the same strain, but these values are different from strain to strain (Table 7)

Table 7: average *f*H calculated from the different values obtained with the hydrophobicity scales.

Strain	average *f*H	SD[a]
***P. aer.* H103**	0.62	0.043
***P. aer.* Braz9**	0.59	0.034
***P. aer.* LES400**	0.64	0.051
***S. aureus* ATCC 25923**	0.54	0.022
***S. aureus* C623**	0.56	0.025
***E. coli* 63103**	0.60	0.038
***K. pneumonie* 63575**	0.72	0.091
***E. faecium* t62764**	0.61	0.038
***E. faecalis* f43559**	0.97	0.035

[a] standard deviation.

3.2 *In silico* validation of the scoring function

Equation 3 is not only an effective tool to analyse the sensitivity to CAMPs of strains of interest, but also a very simple tool for the identification of new "cryptic" CAMPs. In fact, a protein sequence can be divided in all the possible peptides of a desired length and the absolute scores of these peptides can be reported in a graph as function of the position. As the absolute score increases linearly with the antimicrobial potency, the analysis not only allows to determine the position of the CAMP inside the precursor, but also to estimate the MIC of the peptides, at least for the strains described above.

We verified the reliability of this approach by analysing two pools of sequences containing known CAMPs:

I. proteins containing known "cryptic" CAMPs;
II. cathelicidins, protegrins, α-defensins, and some other antimicrobial peptides which are secreted as large precursors successively cleaved to release the mature peptide (from this point of view several traditional CAMPs could be considered "cryptic" CAMPs, the only difference being the fact that the propeptide has no function besides the secretion of the CAMP).

The first group is very heterogeneous and contains CAMPs identified by a variegated panel of experimental and theoretical approaches. For example, some CAMPs were isolated by synthesizing overlapping peptides which covered the entire sequence or a specific region of the protein of interest. It should be noted that, in these cases, the location of the CAMPs within the sequence of their precursors is known with good precision. In other cases the CAMPs were identified by fragmentation of the precursor with

different chemical or enzymatic strategies and, hence, the “optimal” peptide is not known. Several CAMPs were identified by homology or by the recognition of “consensus sequences”. For example, as the N-terminus of human RNase 3 (also known as eosinophil cationic protein) is a known antimicrobial peptide, some research groups systematically analysed the corresponding region in all the homologous RNases. Other groups analysed the antimicrobial activity of heparin-binding sequences or the consensus sequence X-[PFY]-X-[AFILV]-[AFY]-[AITV]-X-[ILV]-X(5)-W-[IL]-X found in serine proteases.

For the analysis, we used the parameters reported in tables 6-ABC. For the sake of brevity, we will discuss only the results obtained using the sets of exponents determined for *Staphilococcus aureus* ATCC 25923 and *Pseudomonas aeruginosa* H103, using the Parker-Gly0 scale. Table 8 shows the correspondence between absolute score values and MIC values for these two strains. These values are reported in graphs of figure 12 as thresholds for the identification of potential CAMPs. In each case the peak above the thresholds corresponds exactly to the known CAMP.

Table 8: correspondence between absolute score values and MIC values for the two selected strains.

Strain	**Absolute Scores**			
	MIC: 100 µM	**MIC: 50 µM**	**MIC: 10 µM**	**MIC: 1 µM**
***Staphilococcus aureus* ATCC 25923**	6.33	6.73	7.64	8.95
***Pseudomonas aurginosa* H103**	5.95	6.46	7.64	9.33

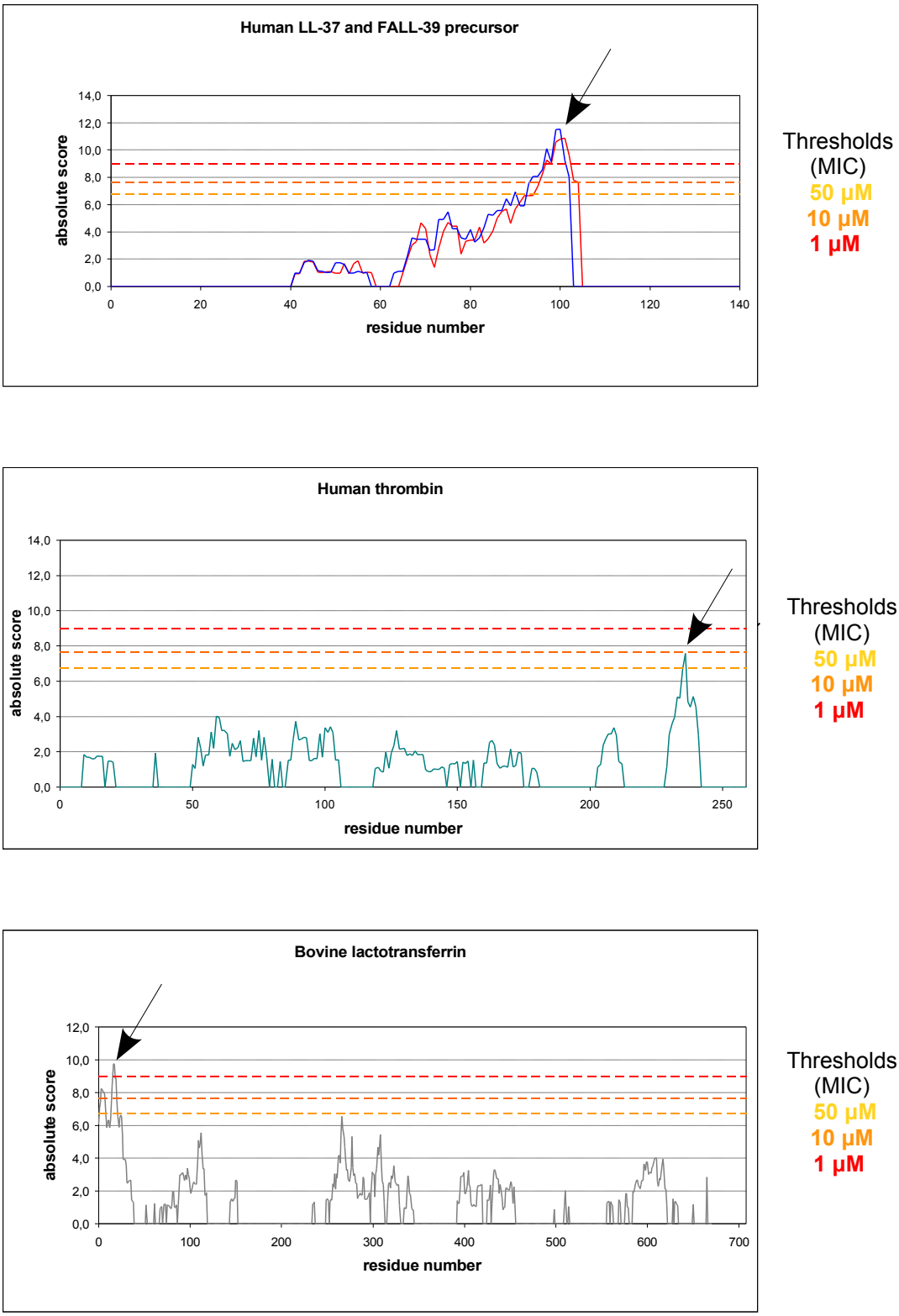


Figure 12: window analysis of human LL-37 and FALL-39 precursor, human thrombin and bovine lactotransferrin, with the following window lengths, chosen according to the length of the known antimicrobial fragments: 39 aa. (**blue**), 37 aa (**red**), 25 aa. (**grey**) and 19 aa. (**green**). The arrows indicate known cryptic CAMPs described in literature, localized by the scoring function.

Tables 9 A-F report for each analysed protein the position and the score of the highest scoring peptide and of selected peptides corresponding to local maxima of the absolute score profile.

Table 9-A: cryptic CAMPs identified by overlapping peptides.

Protein (UniProt ID)	Organism	Known cryptic CAMP/s (ref.)	Absolute maximum *S. aureus* ATCC 25923 (score)	Local maxima *S. aureus* ATCC 25923 (score)	Score of known CAMPs *S. aureus* ATCC 25923	Absolute maximum *P. aer.* H103 (score)	Relative maxima *P. aer.* H103 (score)	Score of known CAMPs *P. aer.* H103
Prothrombin (P00734)	*Homo sapiens*	**598-617** (a)	422-456* (8)	422-453* (6.7) **599-617** (7.6)	7.3	**599-617** (6.3)	422-453* (6.1)	5.9
Cathepsin G (P08311)	*Homo sapiens*	**81-100**; **127-156**; **218-243** (b)	**92-131** (9.7)	**89-122** (8.5) **92-120** (7.6) **214-248** (6)	2.7; 4.3; 3.7	**92-131** (7.0)	**89-122** (6.1) **92-120** (5.5) **214-248** (4.5)	1.7; 2.9; 2.6
Heparin-cofactor 2 (P05546)	*Homo sapiens*	**119-144** (c)	202-241* (11.8)	**113-142** (10.1) **116-142** (9.4) 201-234* (10.1)	6.7	202-241* (9.4)	**113-142** (8.4) **116-142** (7.8) 201-234* (8.3)	5.1
Complement C3 (P01024)	*Homo sapiens*	**675-695; 690-715; 716-742; 728-748** (d)	471-508* (9.8)	**678-716** (6.5) **704-740** (6.9)	2.4; 4.3; 3.8; 2.4; 3.2	471-508* (6.2)	**678-716** (4.3) **704-740** (4.9)	1.5; 2.5; 1.6; 2.2
Thimic stromal lymphopoietin (Q969D9)	*Homo sapiens*	**124-158** (e)	**124-157**, **123-157**, **118-157** (10.9)	/	10.7	**124-157**, **123-157** (8)	/	7.7
Kininogen-1 (P01042)	*Homo sapiens*	**294-319; 497-516** (f)	362-393* (7.4)	**297-324** (7) **295-324** (6.8)	6.3; 1.5	362-393* (5.5)	**297-324** (5.4) **295-324** (5.2)	4.7; 0.6
Tissue factor pathway inhibitor (P10646)	*Homo sapiens*	**278-304** (g)	**264-302, 263-302** (13)	**264-296** (12.3) **274-302** (9.1)	7.6	**264-302, 263-302** (9.6)	**263-296** (9.1) **274-304** (6.7)	5.5

References: (a) Kasetty G. *et al.*, 2011a, (b) Shafer W. M. *et al.*, 1993, (c) Kalle M. *et al.*, 2013, (d) Nordahl E. A. *et al.*, 2004, (e) Sonesson A. *et al.*, 2011, (f) Sonesson A. *et al.*, 2011; Frick I. M. *et al.*, 2006, (g) Papareddy P. *et al.*, 2010.

* putative CAMP.

Table 9-B: cryptic CAMPs identified by analysis of peptides in body fluids.

Protein (UniProt ID)	Organism	Known cryptic CAMP/s (ref.)	Absolute maximum *S. aureus* ATCC 25923 (score)	Local maxima *S. aureus* ATCC 25923 (score)	Score of known CAMPs *S. aureus* ATCC 25923	Absolute maximum *P. aer.* H103 (score)	Relative maxima *P. aer.* H103 (score)	Score of known CAMPs *P. aer.* H103
Fibrinogen beta chain (P02675)	*Homo sapiens*	**45-72** (a)	195-212* (5.9)	**40-77** (5.2); 191-212* (5.8)	2.4	195-212* (4.8)	**40-77** (3,3); 191-212* (4.5)	1.4
Tissue factor pathway inhibitor 2 (P48307)	*Homo sapiens*	**202-235** (b)	**204-235** (13.7)	**204-233** (12.8); **208-233** (11.8)	11.5	**204-235** (10.1)	**204-233** (9.2); **208-233** (8.7)	8.2

References: (a) Påhlman L. I. *et al.*, 2013, (b) Papareddy P. *et al.*, 2012.

* putative CAMP.

Table 9-C: cryptic CAMPs identified for their physicochemical properties.

Protein (UniProt ID)	Organism	Known cryptic CAMP/s (ref.)	Absolute maximum *S. aureus* ATCC 25923 (score)	Local maxima *S. aureus* ATCC 25923 (score)	Score of known CAMPs *S. aureus* ATCC 25923	Absolute maximum *P. aer.* H103 (score)	Relative maxima *P. aer.* H103 (score)	Score of known CAMPs *P. aer.* H103
Platelet factor 4 (P02776)	*Homo sapiens*	**89-101** (a)	**72-99** (9.4)	**76-99** (8.7); **80-99** (7.6)	4.2	**72-99** (7.7)	**76-99** (7.1); **80-99** (6.5)	3.4
Mucin-7 (Q8TAX7)	*Homo sapiens*	**52-71** (b)	**45-84** (10.3)	**51-74** (9.3); 355-371* (7.4)	7.0	**45-84**, **45-74** (7.6)	**51-74** (7.3); 355-371* (6.5)	5.2

References: (a) Darveau R. P. *et al.*, 1992, (b) Bobek L. A. and Situ H., 2003.

* putative CAMP.

Table 9-D: cryptic CAMPs found in ribonucleases, lysozymes and lactotransferrins.

Protein (UniProt ID)	Organism	Known cryptic CAMP/s (ref.)	Absolute maximum *S. aureus* ATCC 25923 (score)	Local maxima *S. aureus* ATCC 25923 (score)	Score of known CAMPs *S. aureus* ATCC 25923	Absolute maximum *P. aer.* H103 (score)	Relative maxima *P. aer.* H103 (score)	Score of known CAMPs *P. aer.* H103
Ribonuclease pancreatic (P07998)	*Homo sapiens*	**29-76** (a)	**57-75** (4.5)	**32-71** (4.3)	**29-68** (3.2)	**57-75** (3.1)	**32-71** (2.5)	**29-68** (1.7)
Non-secretory ribonuclease (P10153)	*Homo sapiens*	28-72 **inactive** (a)	57-96 (5.5)		28-67 (3.1)			
Eosinophil cationic protein (P12724)	*Homo sapiens*	**28-71** (a)	**34-72** (9.1)	**37-72** (8.3); **34-65** (7.9)	**28-67** (8.5)	**34-72** (6.8)	**37-72** (6.3); **34-65** (6)	28-67 (6.2)
Ribonuclease 4 (P34096)	*Homo sapiens*	**29-75** (a)	**51-74** (7.9)	**55-74** (7.1); **51-84** (7.3)	**29-68** (3.2)	**51-74** (6.2)	**55-74** (5.7); **51-84** (5.6)	29-68 (2.1)
Angiogenin (P03950)	*Homo sapiens*	25-71 **inactive** (a)	53-80* (6.5)	48-81 (6.3)	25-64 (2.5)	53-80* (4.7)	48-81 (4.3)	25-64 (1.5)
Ribonuclease K6 (Q93091)	*Homo sapiens*	**24-68** (a)	88-127 (5.7)	80-116 (5.5)	24-63 (4.7)	88-127 (3.7)	80-116 (3.6)	24-63 (3.1)
Ribonuclease 7 (Q9H1E1)	*Homo sapiens*	**29-73** (a)	103-140* (6.6)	**29-66** (6); 106-140 (6.3)	**29-68** (5.8)	103-140 (4.2)	**29-66** (3.8); 106-140 (3.9)	**29-68** (3.7)
Ribonuclease 8 (Q8TDE3)	*Homo sapiens*	28-72 **inactive** (a)	87-125 (3.5)	82-121 (3.2)	28-67 (2.8)	87-125 (2.1)	82-121 (1.9)	28-67 (1.7)
Leukocyte ribonuclease A-2 (Q27J90)	*Gallus gallus*	**92-100**; **112-128** (b)	**90-127** (9.3)	**94-130** (8.5); **90-124** (8.7); **102-134** (7.3)	- ; 3.3	**90-127** (6.5)	**94-130** (5.9); **90-124** (6.2); **102-134** (5.1)	- ; 2.1
Lysozyme C (P61626)	*Homo sapiens*	**105-133**; **105-119**; **125-133** (c)	**111-143**, **110-143** (6.4)	/	1.1; 3.5; (125-136) 3.0	26-52 (4.8)	**111-143**, **110-143** (4.4)	0.7; 2.4; (125-136) 2.2
Lysozyme C (P00698)	*Gallus gallus*	**105-132**; **105-118**; **125-132** (c)	**110-147**, **112-147**, **114-147** (5.6)	/	1.0; 2.3; (125-136) 3.0	**110-147**, **114-147** (3.9)	/	0.6; 1.4; (125-136) 2.1
Lactotransferrin (P02788)	*Homo sapiens*	**20-29**; **39-49** (d)	**21-58** (8.2)	**37-58** (6.5); **35-58** (6.3)	(20-31) 2.5; (39-50) 3.7	**21-58** (5.6)	/	(20-31) 1.6; (39-50) 2.7
Lactotransferrin (P24627)	*Bos taurus*	**36-60**; **287-303** (e)	**21-60**, **22-61**, **22-60** (12.1)	**35-68** (11.6); **35-62** (11.2); **285-320** (7.9); **285-308** (6.8)	9.7; 4.1	**22-60**, **35-68** (9.3)	**35-62** (9.1); **285-320** (5.9)	7.7; 2.8

References: (a) Torrent M. *et al.*, 2013, (b) Nitto T. *et al.*, 2006, (c) Ibrahim H. R. *et al.*, 2001b, (d) Nibbering P. H. *et al.*, 2001, (e) Hoek K. S. *et al.*, (1997), van der Kraan M. I. A. *et al.*, (2004) * putative CAMP.

Table 9-E: cryptic CAMPs located in conserved regions.

Protein (UniProt ID)	Organism	Known cryptic CAMP/s (ref.)	Absolute maximum *S. aureus* ATCC 25923 (score)	Local maxima *S. aureus* ATCC 25923 (score)	Score of known CAMPs *S. aureus* ATCC 25923	Absolute maximum *P. aer.* H103 (score)	Relative maxima *P. aer.* H103 (score)	Score of known CAMPs *P. aer.* H103
Coagulation factor X (P00742)	*Homo sapiens*	**448-467** (a)	**443-475**, **444-475**, **445-475** (9.1)	**448-475** (7.9); **436-475** (8.9)	5.3	**443-475**, **444-475**, **445-475** (6.6)	**436-475** (6.4); **445-473** (6.2)	4.1
Granzyme H (P20718)	*Homo sapiens*	**225-244** (a)	**215-246** (9.3)	**221-246** (8.7); **228-246** (8.1)	5.9	**215-246** (7.2)	**221-246** (6.8); **228-246** (6.9)	4.7
Laminin subunit beta-1 (P07942)	*Homo sapiens*	**223-239** (b)	**223-243** (8.1)	**220-243** (7.8); **227-243** (7.1)	7.4	**223-243**, **223-240** (6.9)	**220-243** (6.4)	6.2
Plasma serine protease inhibitor (P05154)	*Homo sapiens*	**283-302** (b)	**285-308** (10.5)	**285-299** (8.4); 369-405* (9)	6.3	**285-308** (8.8)	**287-306** (8.4); 369-405* (7.3)	4.9

Coagulation factor X and granzyme H are serine proteases containing an antimicrobial region at the C-terminus, with the conserved pattern X-[PFY]-X-[AFILV]-[AFY]-[AITV]-X-[ILV]-X(5)-W-[IL]-X, while laminin subunit beta-1 and plasma protease inhibitor are proteins containing an heparin-binding domain.

References: (a) Kasetty G. *et al.*, 2011b, (b) Andersson E. *et al.*, 2004.

* putative CAMP.

Our strategy, in almost all cases, identifies the antimicrobial regions described in literature. When the highest scoring peptide does not match exactly the known peptide, the known CAMP is always contained inside the highest scoring peptide. We want to underline that the new absolute maxima identified in kininogen-1, complement C3 and heparin cofactor 2 (Table 9-A) are located in regions not studied by the authors of the works cited, and thus they are new putative antimicrobial agents.

In the case of fibrinogen beta chain (Table 9-B), the peptide receives a very low score and the function identifies a “weak” absolute maximum in another region not studied; this peculiar peptide is more active towards Group A and Group B streptococci, in particular if entrapped in the fibrin cloth, and thus, apparently, its mechanism of action could deviate from the behaviour of conventional CAMPs. The antimicrobial peptide of tissue factor pathway inhibitor 2 (Table 9-B), cleaved by neutrophil elastase, is instead perfectly recognized.

The highly cationic and hydrophobic N-terminus of eosinophil cationic protein (Table 9-D) is highlighted by the prediction system, while new putative CAMPs are located towards the C-terminus of angiogenin and ribonuclease 7. Finally, the antimicrobial regions of chicken RNase A-2, lysozymes and lactotransferrins are well recognized. The analysis of human

ribonucleases leads to two false negatives: pancreatic ribonuclease and ribonuclease K6. As for the case of fibrinogen beta chain, these fragments are weakly hydrophobic and with a low content of positive charges, thus they could act by a mechanism different from that of classic CAMPs.

The great majority of the hidden CAMPs located in heparin-binding domains and at the C-terminus of serine proteases are very weak CAMPs. These peptides are particular rich in cationic and hydrophilic residues and lack hydrophobic residues; their action is very often abolished when antibacterial assays are conducted in presence of salt and sometimes have little activity even at physiological conditions (Pasupuleti M. *et al.*, 2009; Malmsten M. *et al.*, 2006; Kasetty G. *et al.*, 2011b). Therefore we have analyzed only few of these peptides endowed with good antimicrobial activity (Table 9-E).

Table 9-F reports the analysis of antimicrobial peptides secreted as proproteins.

Table 9-F: “conventional” antimicrobial peptides secreted as proproteins.

Protein (UniProt ID)	Organism	Known cryptic CAMP/s*	Absolute maximum *S. aureus* ATCC 25923 (score)	Local maxima *S. aureus* ATCC 25923 (score)	Score of known CAMPs *S. aureus* ATCC 25923	Absolute maximum *P. aer.* H103 (score)	Relative maxima *P. aer.* H103 (score)	Score of known CAMPs *P. aer.* H103
Protegrin-1 (P32194)	*Sus scrofa*	**131-148**	**130-148** (6.3)	/	5.9	**130-148** (4.7)	/	4.3
Protegrin-2 (P32195)	*Sus scrofa*	**131-146**	**130-146** (6.1)	/	5.8	**130-146** (4.8)	/	4.4
Protegrin-3 (P32196)	*Sus scrofa*	**131-148**	**130-148** (5.2)	/	4.9	**130-148** (3.9)	/	3.5
Protegrin-4 (P49933)	*Sus scrofa*	**131-148**	**130-148** (5.3)	/	5.0	**130-148** (4.1)	/	3.8
Protegrin-5 (P49934)	*Sus scrofa*	**131-148**	**130-148** (5.5)	/	5.1	**130-148** (4.1)	/	3.8
Cathelicidin antimicrobial peptide (P49913)	*Homo sapiens*	**132-170**; **134-170**	**130-167** (11.8)	**130-162** (10.3); **138-167** (9.8)	8.0; 7.6	**130-167** (9.2)	**130-162** (8.0); **138-167** (7.6)	6.0; 5.6
Cathelicidin-1 (Q6QLQ5)	*Gallus gallus*	**123-148**	**122-148** (11.2)	**125-147** (8.7)	10.9	**122-148** (9.2)	**125-147** (7.2)	9.2
Cathelicidin-4 (P33046)	*Bos taurus*	**131-143**	**130-143** (6.4)	/	6.3	**130-143** (5.9)	/	5.8
Prophenin and tritrpticin precursor (P51524)	*Sus scrofa*	**112-124**; **131-209**	**113-152** (12.9)	**112-129** (11.6); **117-129** (8.6); **121-152** (8.6)	7.6; -	**112-129** (10.5)	**112-132** (10.3); **117-152** (8.2)	7.1; -
Cathelicidin-2 (Q2IAL7)	*Gallus gallus*	**123-154**	**122-153** (12.7)	**126-147** (11); **128-143** (9.7)	12.0	**122-153**, **122-147** ()	**126-147** (9.0); **128-143** (8.2)	9.4
Cathelicidin-3 (Q2IAL6)	*Gallus gallus*	**123-151**	**122-151** (10.1)	/	9.9	**122-151** (8.2)	/	7.9
Cathelicidin-1 (P22226)	*Bos taurus*	**144-155**	**130-153** (7.2)	/	(143-155) 4.8	**130-153** (5.7)	/	(**143-155**) 4.0
Cathelicidin-2 (P19660)	*Bos taurus*	**131-173**	**130-169** (13.7)	**141-169** (8.3); **149-176** (8.5)	(131-170) 13.2	**130-169** (11.2)	**136-162** (7.4); **149-176** (6.8)	(**131-170**) 10.6
Cathelicidin-3 (P19661)	*Bos taurus*	**131-189**	**127-166** (13.8)	**146-180** (10.3); **156-184** (8.3)	(131-170) 13.6	**127-166** (10.6); **130-166** (10.6)	**139-174** (8.6); **156-188** (7.3)	(**131-170**) 10.2
Cathelicidin-5 (P54229)	*Bos taurus*	**132-159**	**134-158** (11.4)	**137-158** (10.0); **139-158** (9.8)	10.3	**134-158** (9.8)	**137-158** (8.7); **139-158** (8.5)	8.4
Cathelicidin-6 (P54228)	*Bos taurus*	**132-158**	**133-157** (15.2)	**133-152** (12.6); **128-157** (14.9)	14.2	**133-157** (13.3)	**133-155**, **128-157** (12.6)	12.0
Cathelicidin-7 (P56425)	*Bos taurus*	**131-164**	**133-164** (9.3)	**128-164** (9.1)	9.0	**133-164** (7.0)	**128-164** (6.6)	6.6
Histon H2A (O13260)	*Bufo gargarigans*	**16-36**	**5-43** (9.4)	**9-36** (8.9)	5.4	**5-43** (6.4)	**9-36** (6.3)	4.0
Alpha-defensin 1 (P11477)	*Mus musculus*	**59-93**	**59-93** (7.1)	/	7.1	**59-93** (4.9)	/	4.9

* the location of the antimicrobial peptides was taken from the UniProt sequence annotations.

The antimicrobial peptides in table 9-F are in all the cases located in the absolute maximum of the scoring function, and often the minimal discrepancies at the extremities are due to the inclusion in the highest scoring peptide of the cleavage signals that, often, are pair of basic residues.

In conclusion, our scoring function properly identifies almost all the hidden antimicrobial domains here reported, failing only in the case of non-canonical CAMPs with a low content of positive residues and/or hydrophobic residues, that likely are not well described by the model in figure 7.

A more complete validations will be performed by automating the calculation of the absolute score, in order to analyse larger pools of known CAMP precursors. For example, at the moment, more than 190 cathelicidins and 140 α-defensins have been described.

Finally, an experimental validation has been performed by preparing and characterizing a new CAMP contained inside human apolipoprotein E, as described in the next section.

3.3 A novel fusion system for the recombinant expression of CAMPs

The production of large quantities of pure peptides is crucial for biological, biophysical and structural studies. The direct isolation from the organism is difficult and time-consuming, whereas chemical synthesis has high costs, especially when high purity, high quantities, long peptides or $^{15}N/^{13}C$ labeled peptides are needed. An economical alternative is given by the recombinant expression of peptides in *Escherichia coli*, one of the most diffused host. However, as direct expression of CAMPs in a bacterial host is made complex by their toxicity, usually CAMPs are produced as fusion proteins. The carrier in the fusion system protects the peptide from proteolytic cleavage and at the same time masks it, abolishing or reducing its toxicity. Two kinds of carriers are usually employed: solubility-enhancing carriers, like thioredoxin, glutathione transferase and small ubiquitin-related modifier (SUMO), and aggregation-promoting carriers, like ketosteroid isomerase. The formation of aggregates, called inclusion bodies, permits to protect more efficiently the peptide from degradation and the bacterial cell from the toxicity of the antimicrobial agent. Moreover, inclusion bodies allow a more rapid purification of the fusion proteins. The peptide can be isolated from the fusion construct using chemical reagents like cyanogen bromide, formic acid and hydroxylamine, which, however, can modify some side chains of the peptide (residues like cysteine, methionine, tryptophan but also lysine, serine, threonine and histidine often undergo undesired reactions). Alternatively, the peptide can be cleaved using proteases like enterokinase, thrombin, factor Xa etc.. Specific proteases are more selective than chemical reagents and do not damage sensitive residues of the peptide, but they are expensive, the yields are unpredictable and sometimes very low – e.g. when the cleavage site is partially hidden and/or unfavorable surrounding residues are present. TEV and SUMO proteases are still active in mild denaturing conditions that, inducing a less compact structure of the fusion protein, allow higher cleavage yields. Generally, proteases cannot be used when the fusion proteins are insoluble and/or extracted from inclusion bodies using strong denaturing conditions. Finally, there are also two examples of self-cleavable carriers: inteins, which excise themselves as introns, but can be prone to an uncontrolled auto-cleavage, and the N-terminal protease N^{pro} of classical swine fever virus, which instead needs extensive dilution, long incubation times and the cleavage is not complete (Li Y., 2011).

In collaboration with the group of Dr. Valeria Cafaro (Department of Biology, Federico II University, Naples) I have developed a new fusion system which allows to obtain very high yields of recombinant peptides without using expensive proteases or harsh cleavage conditions. We chose to express CAMPs as fusion proteins by attaching the desired peptide to the C-terminus of (M23L)-onconase (ONC), a frog ribonuclease (Figure 13). ONC is a very well suited partner for several reasons (Notomista E. *et al.*, 1999): (i) it can be expressed at very high levels as inclusion bodies (about 150 mg/L in Terrific Broth); (ii) no soluble onconase can be detected in the cultures, thus minimizing the risk of toxic effects of the CAMPs; (iii) it is a very small protein (104 aa), thus allowing higher yields of the peptides after the cleavage. Moreover, denatured ONC extracted from inclusion bodies is soluble at acidic pH (<4), but completely insoluble at pH 7. Therefore, if the cleavage of the fusion protein is performed at acidic pH – or at neutral pH in the presence of denaturants (e.g. guanidinium chloride) – and successively the pH is increased to 7 – or the denaturants are removed – ONC forms precipitate, whereas the majority of the CAMPs will remain in solution, thus allowing a very simple purification of the peptide. Moreover, ONC does not contain methionine residues, Asp-Pro or Asn-Gly dipeptides, thus allowing to cleave the desired peptide using the three most common chemical cleavage reagents: cyanogen bromide (CNBr), which cleaves at the C-side of methionine, formic acid, which cleaves the bond between aspartate and proline, and hydroxylamine, which cleaves the bond between asparagine and glycine (Li Y., 2011).

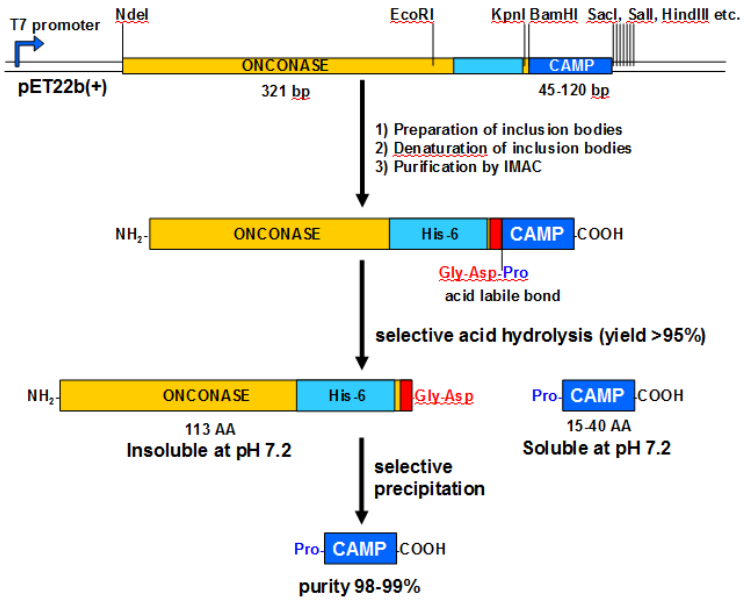


Figure 13: schematic representation of the novel fusion construct ONC-ThrAP.

In order to optimize this method we have prepared a fusion protein between ONC and the human **Thr**ombin-derived **A**ntimicrobial **P**eptide, ThrAP, a CAMP well described in the literature (Kasetty G. *et al.*, 2011a), that we have chosen as positive control. The acid-cleavable sequence GTGDP was inserted between ONC and ThrAP.

The initial fusion protein ONC-ThrAP has been expressed effectively in *E. coli* with a yield of about 150 mg per liter of culture in Terrific Broth. Moreover, the fusion protein was present exclusively in the insoluble fraction of cell lysates, thus confirming that ONC is able to efficiently deliver ThrAP to inclusion bodies. After denaturation of purified inclusion bodies and dialysis in an acidic buffer (pH 3), the fusion protein was heated at 60°C to perform the selective hydrolysis of the Asp-Pro bond.

Different acids were tested:

I. 70% formic acid (the reaction mixture usually described in literature (Landon M., 1977));
II. 0.1 M acetic acid, pH 3;
III. 10 mM HCl, pH 2;
IV. 0.1 M acetic acid/HCl, pH 2.

These analysis revealed that the mixture of acetic and hydrochloric acid at pH 2 allows to obtain an high efficient cleavage (>90%), avoiding the use of formic acid which at high concentrations can give undesired reactions (e.g. formylation of serine and threonine Li Y., (2011)). Interestingly, acetic acid (pH 3.0) or HCl (pH 2.0) alone gave very low cleavage yields suggesting that both pH 2.0 and the presence of acetic acid are necessary to obtain the cleavage. An intriguing hypothesis is that acetic acid, which at pH 2.0 is completely undissociated, acts as a catalyst. However, this aspect has not been further investigated.

Unfortunately, these first attempts also revealed two unexpected problems:

I. The mild acidic hydrolysis (independently from the acid used) cleaves ONC in three fragments, in spite of the fact that it does not contain Asp-Pro dipeptides.
II. Some *E. coli* proteins, present in small amount as contaminants in the inclusion bodies, during the mild acidic hydrolysis release small fragments of length comparable to that of ThrAP.

When the pH of samples was increased from 2 to 7, uncleaved ONC and *E. coli* proteins precipitated completely, whereas the shortest fragments released from these proteins remained in solution as contaminants, thus making more complex the purification of ThrAP. By N-terminal sequencing and mass spectrometry analysis of the fragments, we have assessed that two Asp-Cys dipeptides contained in the ONC sequence (at positions 18

and 67) undergo acid-catalyzed hydrolysis at almost the same efficiency of the Asp-Pro dipeptide, a result not described in literature that suggests the intriguing possibility to prepare recombinant peptides with a single additional cysteine at the N-terminus.
Moreover, at least two other Asp-X sequences of ONC were hydrolyzed with a very low efficiency (1-2%).
In order to solve these problems and to optimize the purification of ThrAP, I have characterized a series of mutated fusion proteins described in the table 10.

Table 10: mutated fusion proteins for the optimization of the carrier.

Protein	**Mutated ONC residues / insertion of His_6**
ONC-YY-ThrAP	C19Y,C68Y
ONC-EYEY-ThrAP	D18E,C19Y,D67E,C68Y
ONC-EYEY-His-ThrAP	D18E,C19Y,D67E,C68Y / His_6
ONC-DC*less*-His-ThrAP	D2E,D16E,D18E,D20E,D32E,D67E,C19Y,C30Y,C48L,C68Y,C75Y,C87I, C90I / His_6

The His_6 tag has been added to purify the fusion proteins by Nickel-chelate affinity chromatography. This chromatographic technique can be performed also in the presence of denaturants. Therefore, it allows to purify the fusion proteins immediately after denaturation of inclusion bodies, thus reducing time and the number of steps necessary to obtain a purified fusion protein suitable for the mild acidic hydrolysis. The mutations in the ONC sequence were chosen to determine the minimal changes necessary to avoid fragmentation of the carrier without changing the expression level of the fusion protein and the pH dependent solubility of ONC. We decided to prepare also a mutant with no cysteine residue (ONC-DC*less*-His-ThrAP), to reduce the possibility of unwanted oxidations and the formation of intra- and inter-chain disulphides. Moreover, this ONC mutant could be used to cleave peptides by reagents specific for cysteines (e.g. 2-nitro-5-thiocyanatobenzoic acid (Ryan R. O. *et al.*, 2003)) as an alternative to acid-catalysed hydrolysis.
Since the solubility and tendency to form aggregates - like inclusion bodies - of a protein is generally related to its net charge and hydrophobicity, but also to its secondary structure, the amino acidic substitutions were chosen in order to preserve these properties. Therefore, aspartic acid residues were replaced with glutamic acid residues in order to preserve the net charge. In the case of cysteine residues, we considered hydrophobicity and propensity to form specific secondary structures. Cysteine residues adjacent to aspartic acid (DC sequences) were replaced with tyrosine residues. The replacement of a

dipeptide DC with EY allows to keep unchanged the secondary structure propensity, in fact, the replacement aspartate / glutamate decreases the preference for loop structures, whereas the cysteine / tyrosine substitution increases the preference for loop structures. The cysteine residues not adjacent to residues of aspartic acid were replaced on the basis of secondary structure of native ONC (PDB code: 1ONC). Accordingly, the single cysteine residue present within α-helix 3 was replaced with a leucine residue (an “helix-preferring” residue). Two cysteine residues located in β-strands were replaced with isoleucine (a “β-preferring” residue). Finally, two cysteine residues located in loops were replaced with tyrosine residues.

All the mutated fusion proteins were expressed with a yield similar to, or even higher, that of ONC-ThrAP and entirely in the insoluble fraction, thus demonstrating that the chosen mutations do not interfere with the formation of inclusion bodies. Moreover, the mutations progressively reduced the number of unwanted cleavage sites. In particular, by changing the DC sequences to DY (ONC-YY-ThrAP) we obtained a drastic reduction but not a complete abolition of unwanted hydrolysis. Only replacing the residue of aspartate with glutamate (ONC-EYEY-ThrAP) it was possible to completely abolish the hydrolysis at these two sites. The analysis of ONC-EYEY-ThrAP has however shown that this protein still undergoes hydrolysis at other aspartic residues, as suggested by the mass spectrometry analysis on the first fusion protein. Finally, the hydrolysis of the protein ONC-DC*less*-His-ThrAP led to two protein bands of molecular weight corresponding to the carrier and ThrAP. The SDS-PAGEs in figure 14 display the different patterns of hydrolysis of the variants of the fusion construct, while the SDS-PAGE in figure 15 recapitulates the purification steps of recombinant ThrAP, starting from the final optimized variant, ONC-DC*less*-His-ThrAP.

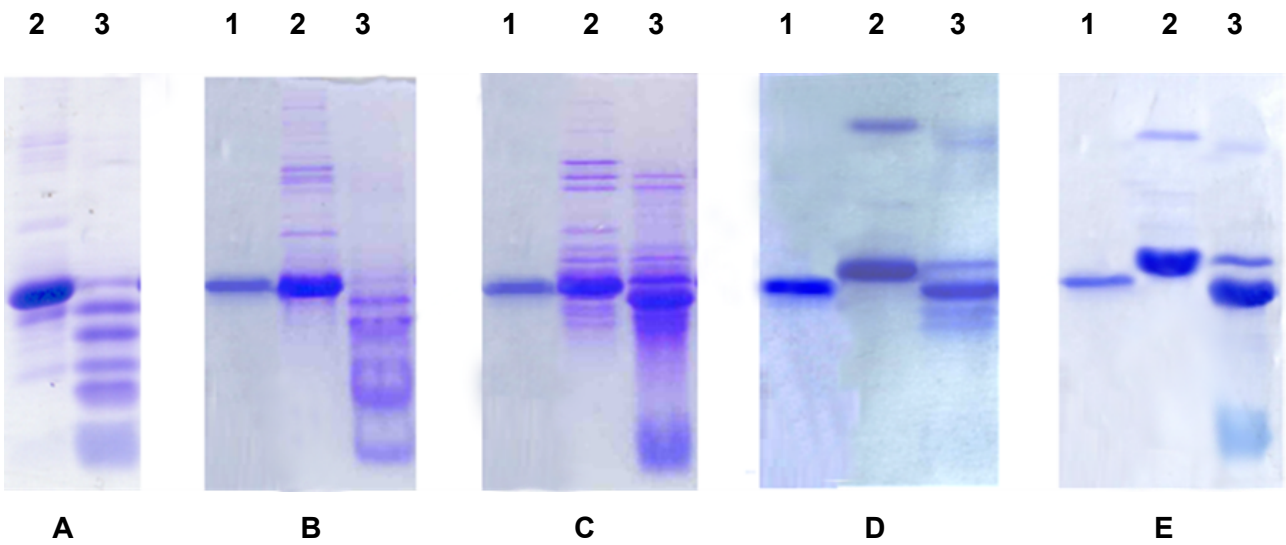


Figure 14: comparison of the acidic hydrolysis patterns of ONC-ThrAP (**A**), ONC-YY-ThrAP (**B**), ONC-EYEY-His-ThrAP (**C**) and ONC-DC*less*-His-ThrAP (**D**). The samples were analyzed on 20% SDS-PAGE. **Lanes 1**: *Gallus gallus* lysozyme (14 kDa, 2 µg); **lanes 2**: fusion construct (5 µg); **lanes 3**: cleaved fusion construct (5 µg).

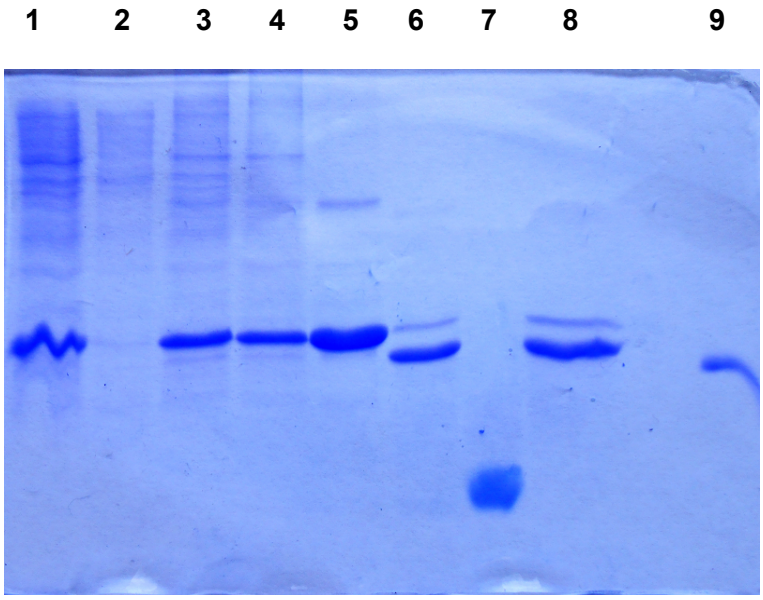


Figure 15: 20% SDS-PAGE of the purification steps of recombinant ThrAP. **Lane 1:** induced culture (0.063 OD); **lane 2**: soluble fraction after cell lysis (0.063 OD); **lane 3**: insoluble fraction after cell lysis (0.063 OD); **lane 4**: insoluble fraction after the preliminary purification (0.063 OD); **lane 5**: purified fusion construct after IMAC (5 µg); **lane 6**: cleaved fusion construct (5 µg); **lane 7**: soluble fraction after alkalinization of the cleavage reaction (5 µg); **lane 8**: insoluble fraction after alkalinization of the cleavage reaction (5 µg); **lane 9**: *Gallus gallus* lysozyme (14 kDa, 2 µg).

After the precipitation step at pH 7, the peptide, analyzed by mass spectrometry by Dr. Andrea Carpentieri (Department of Chemistry, University Federico II, Naples), showed a molecular weight of 2609.47 Da (Figure 16), almost identical to the expected one (2609.1 Da), and a purity >98%.

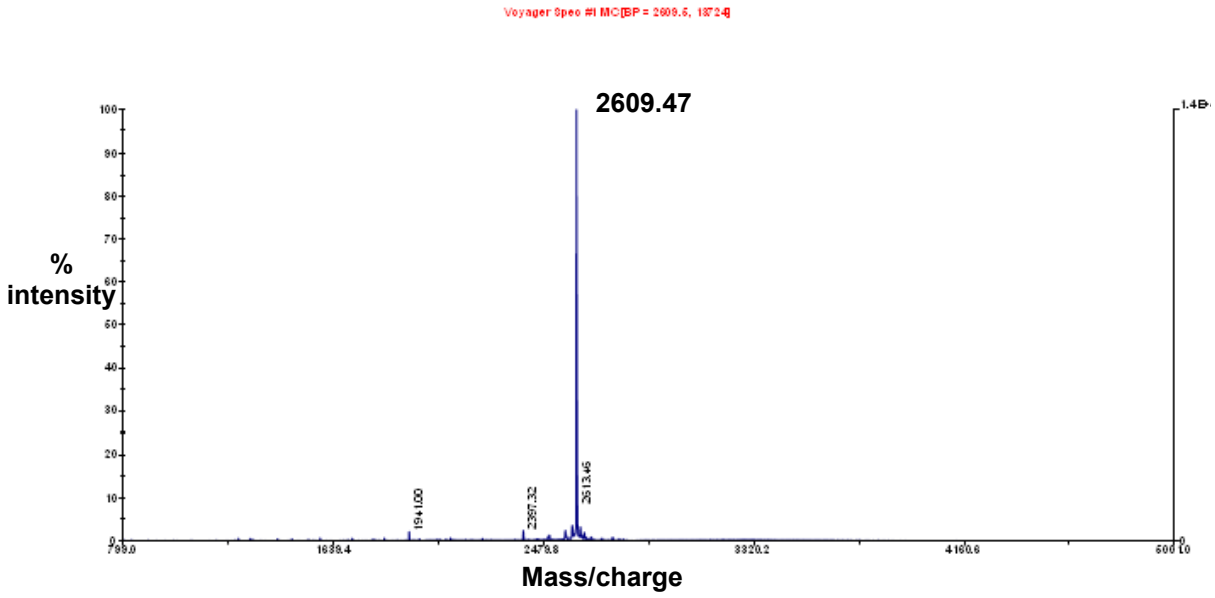


Figure 16: mass spectrum of purified ThrAP.

On the basis of several preparations, I have estimated an average yield of about 7-10 mg of ThrAP per liter of broth. These results confirm that ONC-DC*less* is an optimal carrier for the preparation of recombinant CAMPs and that our method is competitive with respect to conventional chemical synthesis of peptides.

Moreover, Dr. E. Pedone and co-workers (Istituto di Biostrutture e Bioimmagini, CNR, Naples; personal communication), using ONC-EYEY-His, have successively prepared the 21 aa peptide PHGLASTLTRWAHYNALIRAF (a membrane translocation peptide derived from herpes simplex virus type 1 glycoprotein H (Falanga A. *et al.*, 2011)) with yields of 7 mg per liter of culture in the case of the non-labeled peptide, and 1.9 mg per liter of culture in the case of the ^{15}N-labeled peptide, thus demonstrating that the carriers we have developed allow the efficient preparation of labeled peptides for NMR studies.

3.4 Development and optimization of a new rich broth

Recombinant proteins are generally prepared using very rich and complex broths like Luria-Bertani or Terrific Broth. Usually these broths give high expression levels. However,

unfortunately, the reproducibility is low. The variability has been attributed prevalently to yeast extract which shows differences in the composition not only depending on the manufacturer, but also from batch to batch of the same manufacturer (Huang C.-Jr *et al.*, 2012)

For this reason, I have contributed to the development and optimization of a new Semi-Defined Rich Medium (SDRM) in which the yeast extract was completely replaced with nutrient with defined – or less undefined – composition (Table 11).

Table 11: comparison between TB and SDRM composition.

ORGANIC COMPONENTS	TB	SDRM
Yeast extract	24 g/L	-
Trypton	12 g/L	34 g/L
Glycerol	4 ml/L	12 ml/L
Glucose	-	4 g/L
Betaine	-	1 mM
BUFFER COMPONENTS		
KH_2PO_4	2.31 g/L	2.31 g/L
K_2HPO_4	12.54 /L	12.54 /L
$(NH_4)_3$Citrate	-	15 mM
Micronutrients	-	+

Yeast extract was prevalently substituted with trypton, which, being an hydrolyzate of bovine casein, has a defined amino acidic composition. As different batches of trypton can contain different amounts of lactose which could determine undesired induction of the recombinant proteins, we have added glucose to the medium. Glucose acts as a repressor of lac promoter and prevents the induction of the T7-RNA-polymerase of the BL21(DE3) *E. coli* strain until the addition of IPTG. Our SDRM contains also a mixture of salts providing all the necessary metals at optimal concentration (e.g. Mg, Ca, Zn, Fe, Mn, Cu etc.). Ammonium citrate was added both as a source of inorganic nitrogen and as an additional buffer. Moreover, citrate, acting as a chelating agent, avoids the precipitation of transition metal cations. Finally, betaine was added to the medium as this compound is one of the best osmolytes for *E. coli*. This bacterium can synthesize betaine from choline, a component likely present in yeast extract but not in trypton.

Using our SDRM we have reproducibly obtained about 180 mg of protein/L of medium, a result only occasionally obtained with “the best preparations” of Terrific Broth.

3.5 Selection and preparation of a panel of promising new hypothetical CAMP

On the basis of the results described in the bioinformatic section we selected six small to medium sized promising hypothetical human CAMPs (from 18 to 47 aa) (Table 12):

I. a 18 aa peptide derived from apolipoprotein E (ApoE-AP, **Apo**lipoprotein **E**-derived **A**ntimicrobial **P**eptide);
II. three peptides derived from the C-termini of fibrinogen alpha, beta and gamma subunits (α-, β-, γ-FAP, **α**-, **β**-,**γ**-**F**ibrinogen-derived **A**ntimicrobial **P**eptide);
III. a 47 aa peptide derived from pepsin activation peptide (PA3-AP, the activation peptide of pepsinogen A3);
IV. a 29 aa peptide derived from isoform 2 of hydroxysteroid 11-beta-dehydrogenase 1-like protein (H11bD1-AP).

Table 12: primary structure of the six selected novel hypothetical cryptic CAMPs.

Peptide	UniProt ID	position	Primary structure[a]
ApoE-AP	P02649	151-168	LRVRLASHLRKLRKRLLR
α-FAP	P02671	841-866	GVVWVSFRGADYSLRAVRMKIRPLVTQ
β-FAP	P02675	464-491	GVVWMNWKGSWYSMRKMSMKIRPFFPQQ
γ-FAP	P02679	392-421	GIIWATWKTRWYSMKKTTMKIIPFNRLTIG
PA3-AP	P0DJD8	16-62	IMYKVPLIRKKSLRRTLSERGLLKDFLKKHNLNPARKYFPQWKAPTL
H11bD1-AP	Q7Z5J1-2	250-276	GVFYPWRFRLLCLLRRWLPRPRAWFIR

[a] basic residues are colored in blue, hydrophobic residues in green, borderline residues in gray, hydrophilic residues in yellow, acidic residues in red.

All these peptides were expressed as fusion proteins with ONC-DC*less* by the groups of Dr. V. Cafaro and Dr. E. Pizzo (Department of Biology, Federico II University, Naples), with yields similar or higher than that of ThrAP, thus demonstrating that our strategy is of general utility and can be used also to express peptides longer than ThrAP (at least up to 47 aa).

For the sake of brevity, I will describe only the considerations that led us to choose ApoE-AP. The presence of an antimicrobial peptide in apolipoprotein E was described by Dobson C. B. *et al.*, (2006). They selected a 9 aa peptide of ApoE coming from the receptor binding region of the protein. This sequence, however, is a very weak antimicrobial peptide, and so Dobson C. B. and coworkers prepared an artificial head to tail “duplicated” peptide with a good antimicrobial activity. The analysis of the ApoE sequence with our

strategy, on the other hand, suggests that a 18 aa peptide, including the previously identified 9 aa peptide, could be a very effective CAMP (with a relative score close to 0.5, this peptide is one of the highest scoring peptides with length <20 aa). The comparison between the artificial "duplicated" peptide and the peptide we have identified reveals intriguing similarities (Table 13). The colours in table 13 also highlight some interesting differences in the composition and distribution of residues in ThrAP and ApoE-AP: ApoE-AP is rich in aliphatic residues and contains homogeneously distributed basic and aliphatic residues, whereas ThrAP is rich in aromatic residues and shows a C-terminal amphipathic domain (helical in thrombin, see figure 19, paragraph 3.7) and a N-terminal aromatic and hydrophobic domain (extended in thrombin, see figure 19, paragraph 3.7).

Table 13: primary structure of the "duplicated" ApoE derived peptide, ApoE-AP and ThrAP.

Peptide	Primary structure[a]	Net charge
"duplicated" peptide	LRKLRKRLLLRKLRKRLL repeat1 repeat2	+10
ApoE-AP[b]	LRVRLASHLRKLRKRLLR	+8
ThrAP[b]	GKYGFYTHVFRLKKWIQKVI	+5

[a] basic residues are colored in blue, aliphatic and aromatic residues in green and cyan respectively, borderline residues in gray, hydrophilic residues in yellow.

[b] recombinant peptides produced as fusion proteins with ONC-DC*less* have an additional proline at the N-terminus derived from the acid-labile sequence Gly-Asp-Pro.

3.6 Antibacterial activity of recombinant ThrAP and ApoE-AP

The novel antimicrobial peptide ApoE-AP was produced by Dr. E. Pizzo's group using the strategy described above and characterized along with the control peptide ThrAP.

The antibacterial activity of the recombinant peptides was assessed on a lab strain of *Staphylococcus aureus* (ATCC 6538P, Gram-positive bacterium) and on a clinical isolate of *Pseudomonas aeruginosa* (KK27, Gram-negative bacterium), kindly provided by D. Alessandra Bragonzi (San Raffaele Hospital, Milan). The results in figure 17 show that ApoE-AP is more active than the control peptide ThrAP on both strains.

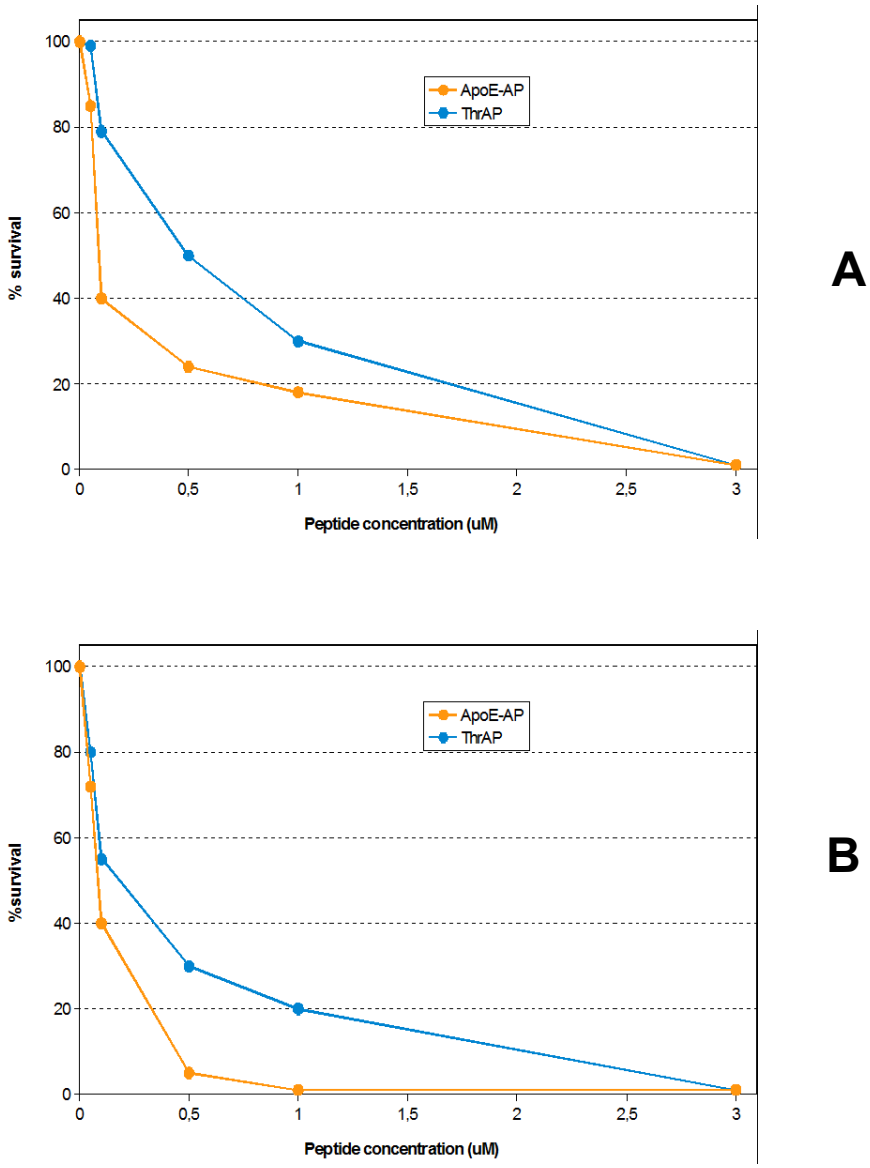
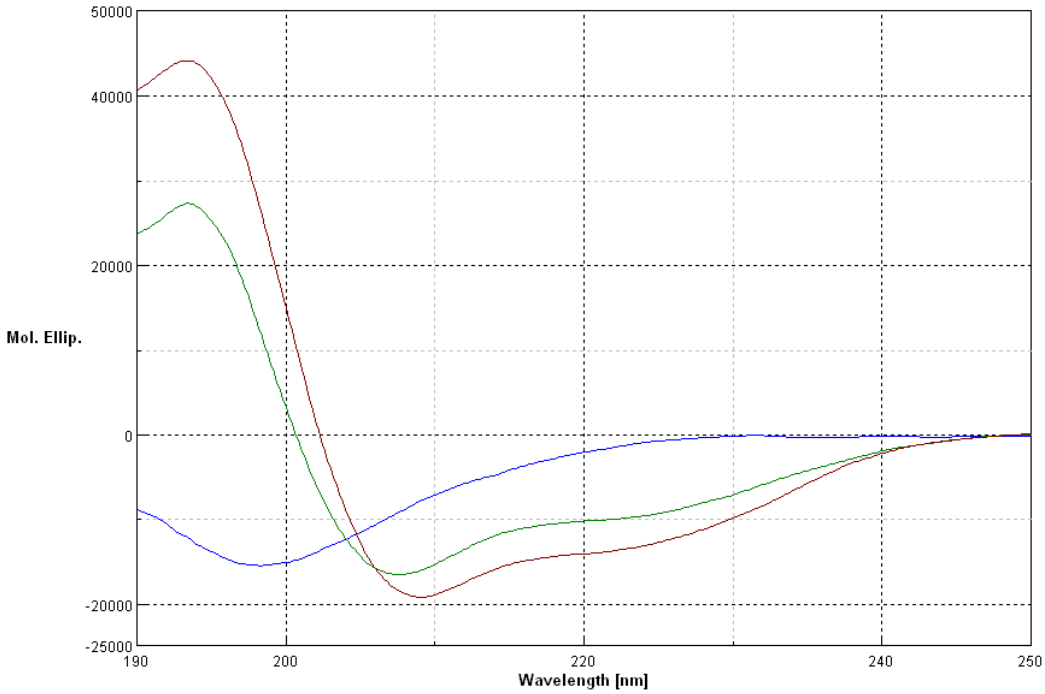


Figure 17: antibacterial activity of the recombinant peptides towards *Staphylococcus aureus* ATCC (**A**) and *Pseudomonas aeruginosa* KK27 (**B**). The relative error was always lower than the 10% of the measure.

3.7 Structural characterization of recombinant ThrAP and ApoE-AP

Circular dichroism studies were performed to characterize ThrAP and ApoE-AP in different environments. The two peptides were largely unordered in buffer, while TFE and SDS, two membrane-mimicking agents (see paragraph 1.3), induced a pronounced α-helix in both (Figure 18).



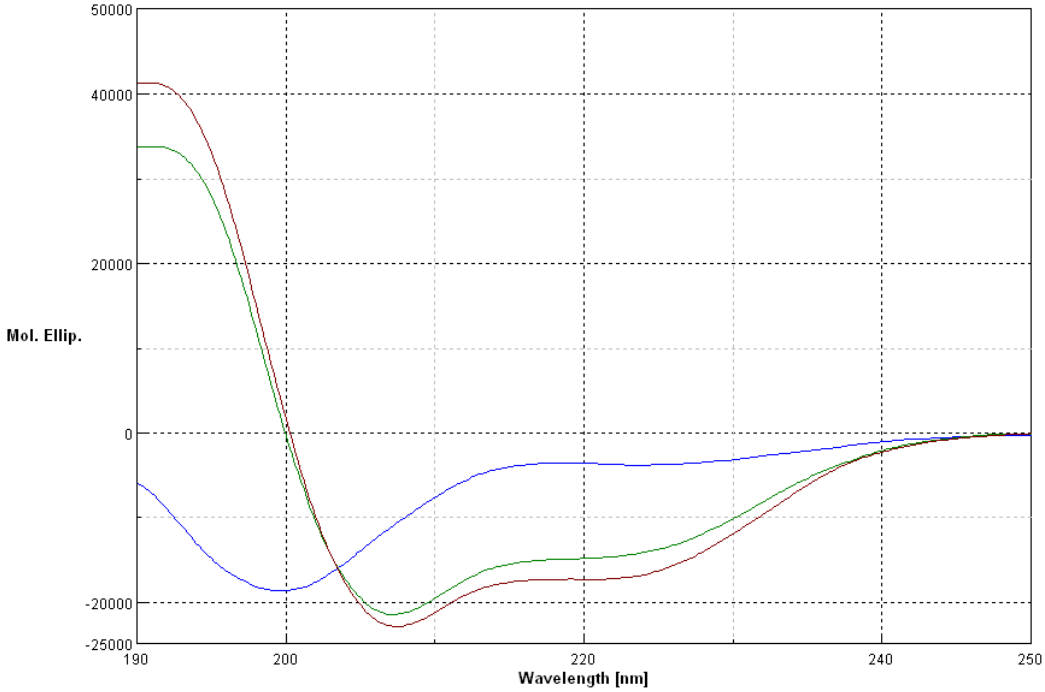


Figure 18: CD spectra of the recombinant peptides in buffer and in the presence of membrane-mimicking agents. In **blue**: spectra registered in sodium-phosphate 10 mM pH 7.4; in **green**: spectra registered in SDS 20 mM; in **brown**: spectra registered in TFE 30%.

Secondary structure content was estimated through the PEPFIT tool (Reed J. and Reed T. A., 1997), and the results are shown in table 14.

Table 14: secondary structure content estimated from circular dichroism spectra in phosphate buffer and in the presence of membrane-mimicking agents.

	Random coil	**α**	**β**	**turn**	**R^2**
ThrAP **buffer**	47%	/	/	53%	0.9979
ApoE-AP **buffer**	72%	10%	13%	5%	0.9927
ThrAP + **SDS 20 mM**	28%	32%	8%	32%	0.9793
ApoE-AP + **SDS 20 mM**	14%	50%	/	36%	0.9923
ThrAP + **TFE 10%**	42%	/	/	58%	0.9941
ApoE-AP + **TFE 10%**	69%	8%	15%	8%	0.9965
ThrAP + **TFE 30%**	2%	51%	/	47%	0.9868
ApoE-AP + **TFE 30%**	18%	69%	/	13%	0.9945
ThrAP + **TFE 50%**	/	57%	/	43%	0.9826
ApoE-AP + **TFE 50%**	14%	74%	/	12%	0.9938
ThrAP + **TFE 70%**	/	54%	/	46%	0.9859
ApoE-AP + **TFE 70%**	13%	78%	/	9%	0.9946

The value of R^2 is a measure of the discrepancy between the experimental spectrum and the calculated one obtained with the PEPFIT tool; a value of 1 indicates a perfect match between the spectra.

It is worth noting that the helix content did not exhibit a significant change at concentrations of TFE higher than 30%, denoting a high propensity to acquire an ordered structure; peptides with pronounced helical-propensity reach, in fact, the maximum helical content at concentrations of TFE between 30% and 50% (Sönnichsen F. D. *et al.*, 1992).
However, it should be noted that in all the conditions tested ApoE-AP has a helix content significantly higher than ThrAP. This could indicate that the isolated peptides, in the presence of structure-inducing compounds (TFE, SDS), tend to adopt structures similar to those seen in the intact proteins (Figure 19): in the ApoE structure the peptide corresponding to ApoE-AP is entirely helicoidal, whereas in the thrombin structure, only 10 out of 20 residues of the region corresponding to ThrAP adopt an helical conformation (the last ten residues of ThrAP).

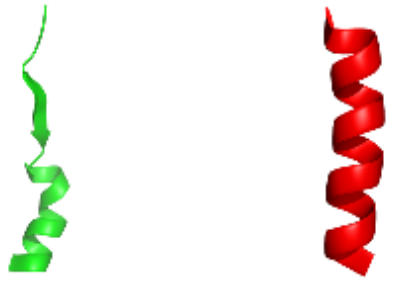
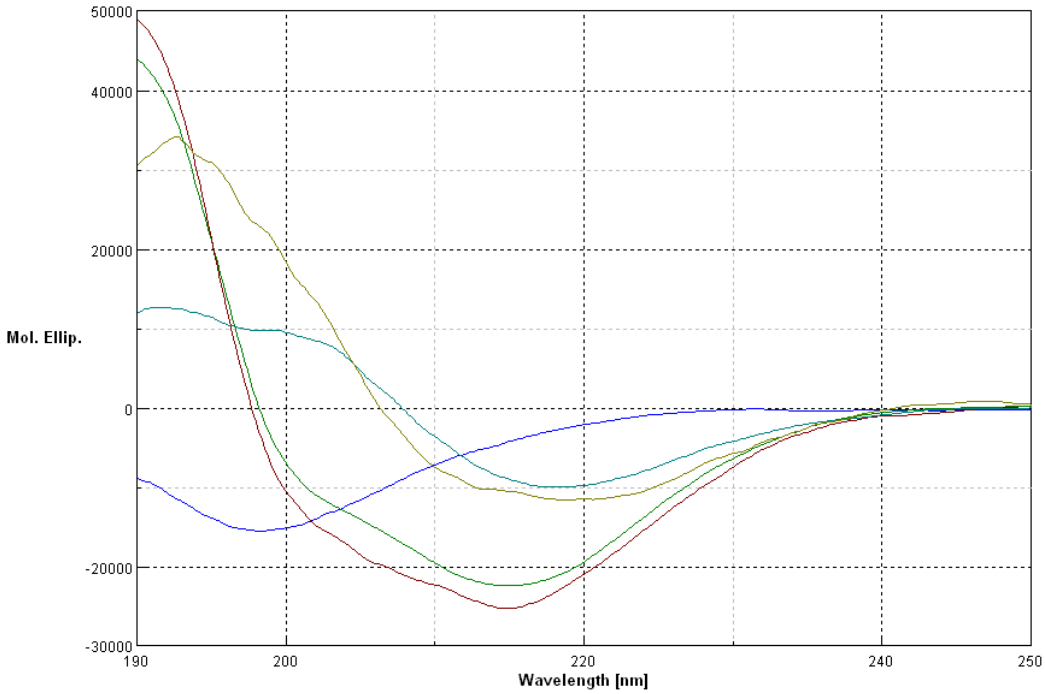


Figure 19: comparison between the structure of ThrAP (in **green**) and ApoE-AP (in **red**) in the respective entire proteins (PDB codes: 1PPB for thrombin and 2L7B for ApoE). The molecules are depicted with the N-terminus up and the C-terminus down.

To further characterize the structural and biological properties of the two peptides we studied their binding to alginate and LPS, two bacterial molecules of great biological relevance: alginate, an acidic polysaccharide, can sequester antimicrobial peptides inhibiting their action, whereas LPS, the main constituents of the outer membrane of Gram negative bacteria, are “endotoxins” that can induce septic shock. Some CAMPs, in addition to antimicrobial activity, can bind and neutralize LPS thus preventing septic shock (see paragraph 1.1). The binding of ThrAP to LPS has already been described by Kasetty G. *et al.*, (2011a): LPS induced an helical structure and the peptide displayed an immunomodulatory activity *in vivo* in macrophage cultures. Therefore we compared the binding of ThrAP and ApoE-AP to LPS and alginate.

At similar concentrations (about 3-fold higher than the work previously cited), ThrAP and ApoE-AP reacted differently to both alginate and LPS: ThrAP aggregated in the presence of LPS and, at a minor extent, also in presence of alginate; on the contrary, ApoE-AP remained soluble in both conditions. The addition of sodium chloride reduced the aggregation (data not shown), but it was abolished only at very low ThrAP concentrations. The spectra obtained with the ligands are in figure 20.



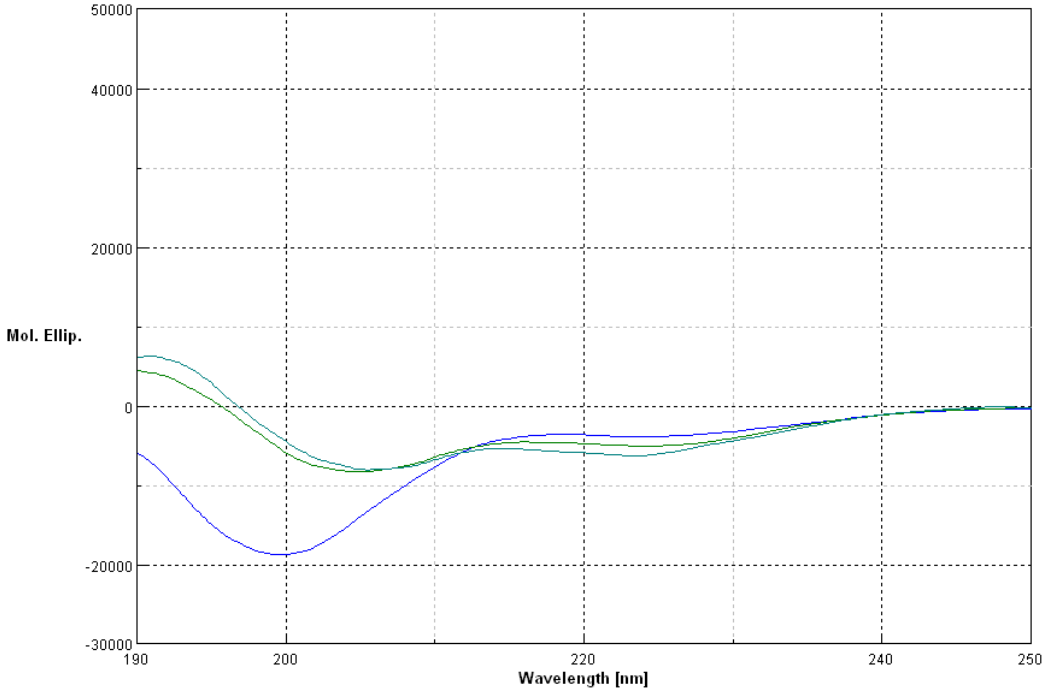


Figure 20: CD spectra of the recombinant peptides in buffer and in the presence of LPS and alginate. In **blue**: spectra registered in sodium-phosphate 10 mM pH 7.4; in **green**: spectra registered in the presence of alginate 0.2 mg/mL; in **light blue**: spectra registered in the presence of LPS 0.2 mg/mL. Only for ThrAP, at a concentration of 10 μM, two other spectra with LPS (**olive green**) and alginate (**brown**) were recorded.

Secondary structure content was estimated as described before, and the results are shown in table 15.

Table 15: secondary structure content estimated from circular dichroism spectra in the presence of LPS and alginate.

	Random coil	α	β	turn	R^2
ThrAP + **LPS 0.2 mg/mL**	9%	20%	/	71%	0.9636
ThrAP (10 uM)* + **LPS 0.2 mg/mL**	/	22%	20%	58%	0.9907
ApoE-AP + **LPS 0.2 mg/mL**	45%	14%	22%	19%	0.9949
ThrAP + **alginate 0.2 mg/mL**	/	45%	/	55%	0.9621
ThrAP (10 uM)* + **alginate 0.2 mg/mL**	3%	52%	/	45%	0.9581
ApoE-AP + **alginate 0.2 mg/mL**	48%	13%	17%	22%	0.9930

The value of R^2 is a measure of the discrepancy between the experimental spectrum and the calculated one obtained with the PEPFIT tool; a value of 1 indicates a perfect match between the spectra.

* diluted solutions which displayed no aggregation.

ApoE-AP's spectra are indicative of a relative small perturbation of the random-coil state, whereas ThrAP's spectra suggest a more relevant conformational change induced by the ligands. Bhunia A. *et al.*, (2009) found that the antimicrobial peptide fowlicidin-1, which has two LPS-binding regions, aggregated in presence of LPS. Interestingly the two regions, once separated, remained soluble in the presence of LPS. This observation lets to propose the hypothesis that only peptides with multiple LPS-interaction sites cause the aggregation phenomenon. As discussed above, ThrAP shows the presence of two distinct regions, an extended more hydrophobic region at the N-terminus and an amphipathic helical charged region at the C-terminus which could behave as distinct LPS-binding modules. Further studies will be necessary to confirm this hypothesis. Finally, it should be noted that ThrAP has several aromatic residues which, as underlined in Pulido D. *et al.*, (2011), play a fundamental role in LPS interaction. On the contrary no aromatic residue is present in ApoE-AP.

It is worth noting that ApoE-AP is active on Gram-negative strains even if it interacts weakly with LPS. Therefore, LPS-binding, likely, is not necessary for antibacterial activity on Gram-negative strains. Moreover, the presence of high concentrations of free LPS and/or capsular polysaccharides as alginates could scavenge and inhibit the antimicrobial activity of ThrAP (as described for other CAMPs that, like ThrAP, bind these bacterial secretion products), whereas, ApoE-AP will retain its antimicrobial efficacy. On the other

hand, ApoE-AP, likely, will not be able to prevent septic shock. The data presented here strongly suggest that both from the structural and functional point of view ThrAP and ApoE-AP are complementary. This conclusion has relevant consequences for a future pharmacological application of the two peptides.

COMPUTATIONAL SECTION

3.8 Modelling of CAMPs by implicit solvation

An accurate representation of solvent is crucial in biological simulations in order to obtain meaningful and realistic results. Simulations can be carried out in presence of explicit solvation, but the relative high number of degree of freedoms can have a strong computational cost, and thus implicit solvation, that approximate the effects of solvent through a potential of mean force, can reduce the computational complexity (Feig M. and Brooks C. L. III, 2004). A recent work from Huang A. and Stultz C. M. (2007) compared the local energy minima of a small peptide obtained with explicit solvation and with three different models of implicit solvations, and found that all the different approaches mapped similar regions of the conformational space. Moreover, different authors developed solvent models for the simulations of peptides in lipid bilayers (Efremov R. G. *et al.*, 1999a and b; Maddox M. W. and Longo M. L., (2002); Lazaridis T., 2003) and in pores (Mihajlovic M. and Lazaridis T., 2010; He Y. *et al.*, 2013).

In this thesis, the influence of different implicit solvations was studied on a panel of experimental structures of CAMPs, by means of the Monte Carlo strategy, in order to define the best possible conditions which preserve the initial structure. In detail, the simulation of each peptide was carried out in vacuum, water (with the effective energy function-1, EEF-1, developed by Lazaridis T. and Karplus M., (1999)), octanol (Hopfinger A. J. and Battershell R. D., 1976) and octanol with an attenuation of the solvation energy (0.5 x solvation energy and 0.25 x solvation energy). The last two conditions were studied with the aim of recreating a sort of “hybrid” ambient, “partially unpolar”, which could mimic a micellar environment and TFE solvation; NMR structures of CAMPs are in fact usually solved in presence of SDS or DPC micelles and TFE, as described in paragraph 1.3. The majority of the CAMPs selected for this study are helical peptides which are, usually, unordered in water and fold in the presence of micelles or TFE. This kind of peptides is very well studied, and several NMR structures are available. The only two exceptions in our set are indolicidin and tritrpticin, that adopt a prevalently extended structure (PDB codes: 1G89 and 2I1D).

Tables 16-ABC describe the results of the structural alignments between the initial experimental structure and the lowest energy structure of the Monte Carlo ensemble in the different simulated environments, divided in three groups: structures solved in presence of

DPC micelles, SDS micelles and TFE.

Table 16-A: RMSD of the structural alignments between the peptide structure solved in the presence of DPC micelles and the lowest energy model obtained with the different simulations.

	Structures solved in DPC micelles				
	Water solvation	**Vacuum**	**Octanol solvation**	**0.5x Octanol Solvation**	**0.25x Octanol Solvation**
RP-1 (2RLH) 18 aa.					
RMSD backbone (Å)	12.56	8.2	13.84	7.61	3.25
RMSD (portion (Å)	/	(4-16) 6.34 (5-11) 1.66	/	(4-16) 3.83 (6-14) 1.21	(4-16) 1.57
Piscidin (2JOS) 22 aa.					
RMSD backbone (Å)	12.06	4.35	17.78	7.27	9.55
RMSD (portion) (Å)	/	(8-21) 0.92	/	(8-21) 0.94	(8-21) 8.57 (8-14) 0.93
Indolicidin (1G89) 13 aa.					
RMSD backbone (Å)	9.16	8.18	9.24	5.04	4.21
RMSD (portion) (Å)	/	(5-10) 1.29	/	(5-10) 3.97 (6-8) 1.47	(5-10) 1.84 (5-9) 1.48
Tritrpticin-1 (2I1D) 14 aa.					
RMSD backbone (Å)	8.89	9.87	9.25	9.92	11.08
RMSD (portion) (Å)	/	/	/	/	/
CM15 (2JMY) 15 aa.					
RMSD backbone (Å)	13.22	8.14	14.07	7.62	10.93
RMSD (portion) (Å)	/	(2-9) 1.88 (1-8) 1.26	/	(1-8) 1.57 (2-9) 1.21	(1-8) 3.35 (2-9) 4.50 (2-7) 1.4
Magainin-2 (2MAG) 23 aa.					
RMSD backbone (Å)	10.9	12.98	13.89	6.55	5.68
RMSD (portion) (Å)	/	/	/	(5-18) 1.46 (6-18) 1.46	(5-18) 1.93 (6-18) 1.27
LL-23 (2LMF) 23 aa.					
RMSD backbone (Å)	11.98	10.13	14.43	3.20	3.31
RMSD (portion) (Å)	/	(5-20) 6.82 (11-18) 1.06	/	(5-20) 1.25	(5-20) 1.38

PDB IDs are shown in brackets near the peptide's name.

Table 16-B: RMSD of the structural alignments between the peptide structure solved in the presence of SDS micelles and the lowest energy model obtained with the different simulations.

	Structures solved in SDS micelles				
	Water solvation	**Vacuum**	**Octanol solvation**	**0.5x Octanol Solvation**	**0.25x Octanol Solvation**
RP-1 (2RLG) 18 aa.					
RMSD backbone (Å)	14.44	7.81	12.46	5.06	1.90
RMSD (portion) (Å)	/	(4-14) 4.51 (5-10) 1.58	/	(4-14) 1.24	(4-14) 1.03
Piscidin-1 analogue (2JON) 22 aa.					
RMSD backbone (Å)	12.4	10.59	14.41	14.58	9.64
RMSD (portion) () (Å)	/	(9-14) 2.38 (10-15) 2.38 (9-13) 2.29	/	(9-14) 2.56 (10-15) 2.12	(10-15) 1.82 (9-14) 2.22
Indolicidin (1G8C) 13 aa.					
RMSD backbone (Å)	8.44	8.22	7.82	4.88	5.45
RMSD (portion) (Å)	/	(5-9) 1.75	/	(5-9) 1.77	(5-9) 1.73
Tritrpticin-1 (1D6X) 14 aa.					
RMSD backbone (Å)	10.48	4.65	6.76	6.94	11.24
RMSD (portion) (Å)	/	(6-10) 2.37	(6-10) 4.19 (5-8) 1.87	(6-10) 1.47	/
Latarcin-2a (2G9P) 26 aa.					
RMSD backbone (Å)	8.69	4.66	19.14	7.06	6.55
RMSD (portion) (Å)	/	(13-22) 1.29	/	(13-22) 0.63	(13-22) 0.69
LL-37 (2KFO) 37 aa.					
RMSD backbone (Å)	8.03	6.11	13.06	7.10	8.55
RMSD (portion) (Å)	(13-30) 3.35 (17-26) 1.32	(13-30) 1.30	/	(13-30) 1.59 (13-29) 1.28	(13-30) 2.24 (14-27) 0.96
Piscidin-1 (2OJM) 22 aa.					
RMSD backbone (Å)	10.45	8.73	16.31	1.75	9.76
RMSD (portion) (Å)	(5-19) 5.82 (8-15) 1.44	(5-19) 5.96 (10-20) 1.39	/	(5-19) 0.87	(5-19) 6.04 (10-18) 1.51

PDB IDs are shown in brackets near the peptide's name.

Table 16-C: RMSD of the structural alignments between the peptide structure solved in the presence of TFE and the lowest energy model obtained with the different simulations.

	Structures solved in TFE				
	Water solvation	**Vacuum**	**Octanol solvation**	**0.5x Octanol Solvation**	**0.25x Octanol Solvation**
Meucin-24 (2KFE) 24 aa.					
RMSD backbone (Å)	10.1	4.39	11.8	7.7	6.22
RMSD (portion) (Å)	(5-15) 1.01	(5-15) 1.33	/	(5-15) 1.31	(5-15) 1.21
Fowlicidin-3 (2HFR) 27 aa.					
RMSD backbone (Å)	17.15	5.88	23.11	9.29	6.31
RMSD (portion) (Å)	/	(9-20) 1.33	/	(9-20) 1.49	(9-20) 1.3
Fowlicidin-2 (2GDL) 31 aa.					
RMSD backbone (Å)	11.56	14.83	21.68	17.02	14.82
RMSD (portion) (Å)	/	/	/	/	/
CAP18(106-137) (1LYP) 32 aa.					
RMSD backbone (Å)	8.96	5.2	15.91	9.52	5.18
RMSD (portion) (Å)	(7-21) 4.59 (13-23) 1.46	(7-21) 2.24 (11-21) 1.36	/	(7-21) 1.54 (11-21) 1.10	(7-21) 3.08 (14-25) 1.62
Fowlicidin-1 (2AMN) 26 aa.					
RMSD backbone (Å)	14.63	7.26	23.15	8.28	7.35
RMSD (portion) (Å)	(10-16) 1.22	(7-17) 1.18	/	(7-17) 1.67 (8-17) 1.15	(7-17) 1.25
Phylloseptin-2 (2JP1) 19 aa.					
RMSD backbone (Å)	12.13	6.82	8.51	4.95	6.45
RMSD (portion) (Å)	/	(5-13) 1.69 (6-10) 1.42	/	(5-13) 1.11	(5-13) 1.19
Ranatuerin-2CSa (2K10) 32 aa.					
RMSD backbone (Å)	11.67	10.49	18.87	12.72	11.57
RMSD (portion) (Å)	(14-25) 3.25 (13-23) 1.44	14-25 (1.43)	/	(14-25) 3.79 (15-22) 1.03	(14-25) 3.58 (12-21) 1.30

PDB IDs are shown in brackets near the peptide's name.

The results shown in tables 16-ABC point out that a relevant portion of the experimental structures solved in the presence of SDS or DPC micelles is preserved in simulations run using the “attenuated” octanol implicit solvation; to a lesser degree, also simulations in vacuum are able to preserve the experimental structure. As for the structures solved in presence of TFE, the results are less clear and both vacuum and the “attenuated” octanol

seem to provide similar results. Apparently, the “attenuated” octanol solvation is less able to preserve the conformation induced by TFE. TFE is a known strong helix inducer and the structures obtained in this solvent are likely more ordered and compact than those obtained in true micelles. The results obtained with magainin-2 are in good agreement with the simulations conducted by Efremov R. G. *et al.*, (1999b), where the same peptide was studied in implicit water, an implicit membrane-like environment and vacuum. The most significant differences are in the conformer modelled in water, which completely looses secondary structure in our simulation, while instead still retains helical stretches in the cited work; they both share, however, a similar compact structure (Figure 21).

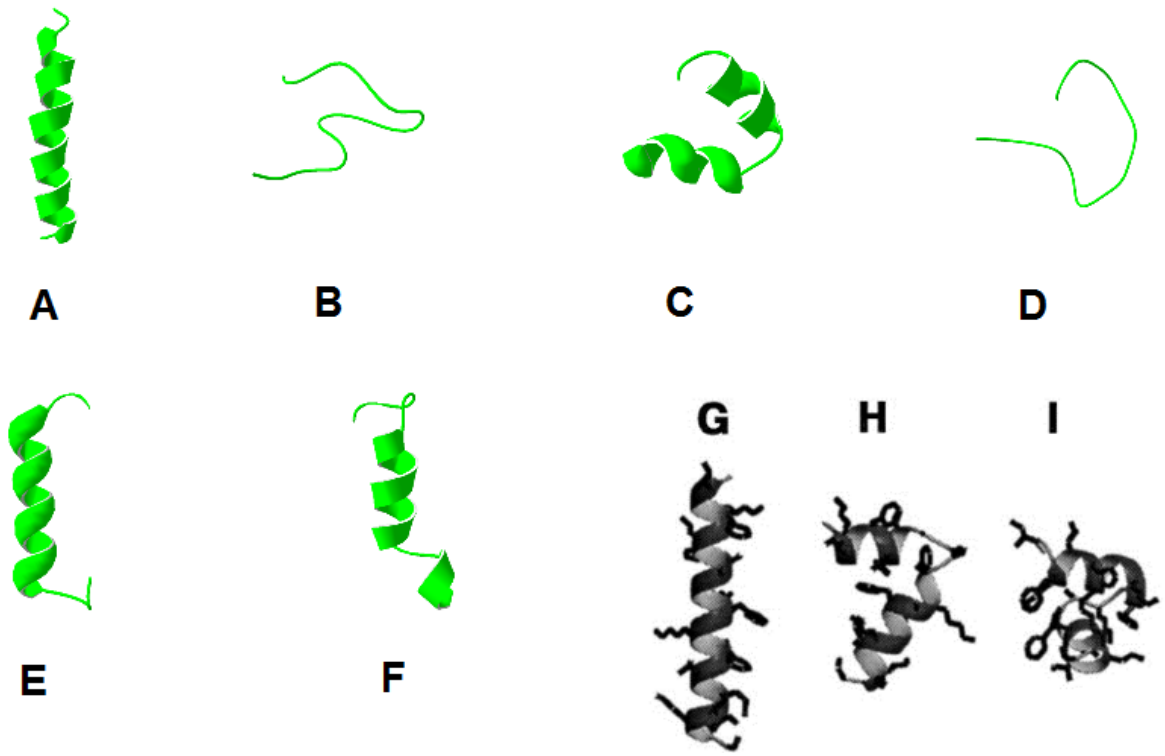


Figure 21: comparison between the simulations outputs for magainin-2. In **green**: initial experimental structure (A) and models obtained in water (B), vacuum (C), octanol (D), 0.5x octanol (E) and 0.25x octanol (F) with our simulations; in **grey**: models obtained in the membrane-mimicking ambient (G), water (H) and vacuum (I) in the simulations run by Efremov R. G. *et al.*, (1999b). (The initial structure, according to the article, perfectly matches the model G). The molecules are depicted with the N-terminus up and the C-terminus down.

In order to further analyse the results of the modelling procedures we have determined

the number of residues in α-helix, the accessible surface area (ASA) and the volume of models and reference structures. The results are shown in table 17.

Table 17: helicity (N_α), accessible solvent area (ASA) and volume of the experimental structures and the lowest-energy conformers obtained in different environments.

Peptide	Experimental structure	Water solvation	Vacuum	Octanol solvation	0.5x Octanol Solvation	0.25x Octanol Solvation
Indolicidin (1G89) 13 aa.						
N_α	0	0	0	0	0	0
ASA (Å^2)	1400	1261	1164	1500	1221	1206
Volume (Å^3)	1837	1907	1936	1810	1934	1925
Tritrpticin-1 (2I1D) 14 aa.						
N_α	0	0	0	0	0	0
ASA (Å^2)	1161	1336	1097	1557	1341	1223
Volume (Å^3)	1852	1905	1956	1736	1834	1916
CM15 (2JMY) 15 aa.						
N_α	11	0	9	0	6	7
ASA (Å^2)	1187	1270	1080	1495	1329	1109
Volume (Å^3)	1791	1880	1863	1780	1822	1895
RP-1 (2RLH) 18 aa.						
N_α	13	0	8	0	8	11
ASA (Å^2)	1342	1744	1326	1703	1588	1539
Volume (Å^3)	2188	2122	2319	2122	2183	2224
Piscidin (2JOS) 22 aa.						
N_α	10	0	19	0	13	14
ASA (Å^2)	1739	1611	1594	2041	1745	1513
Volume (Å^3)	2580	2779	2609	2424	2570	2670
Magainin-2 (2MAG) 23 aa.						
N_α	18	0	16	0	14	12
ASA (Å^2)	1621	1637	1395	1992	1656	1613
Volume (Å^3)	2395	2632	2569	2391	2465	2514
LL-23 (2LMF) 23 aa.						
N_α	18	10	15	0	20	21
ASA (Å^2)	1799	1962	1622	2270	1848	1783
Volume (Å^3)	2839	3020	2963	2709	2855	2875

Indolicidin (1G8C) **13 aa.**						
N_α	0	0	0	0	0	0
ASA (Å^2)	1436	1224	1142	1045	1335	1222
Volume (Å^3)	1881	2039	1939	1993	1895	1947
Tritrpticin-1 (1D6X) **14 aa.**						
N_α	0	0	0	0	0	0
ASA (Å^2)	1306	1319	1179	1499	1335	1128
Volume (Å^3)	1868	1900	1939	1774	1848	1976
RP-1 (2RLG) **18 aa.**						
N_α	11	0	9	0	10	13
ASA (Å^2)	1358	1764	1351	1890	1563	1455
Volume (Å^3)	2143	2247	2322	2076	2258	2366
Piscidin-1 analogue (2JON) **22 aa.**						
N_α	5	4	13	0	8	13
ASA (Å^2)	1690	1786	1495	2125	1815	1589
Volume (Å^3)	2622	2764	2744	2523	2640	2713
Latarcin-2a (2G9P) **26 aa.**						
N_α	17	5	15	0	15	18
ASA (Å^2)	1941	2105	1831	2415	2042	1991
Volume (Å^3)	2995	3062	3045	2715	2882	2976
Piscidin-1 (2OJM) **22 aa.**						
N_α	19	13	15	0	18	17
ASA (Å^2)	1696	1533	1477	2038	1652	1523
Volume (Å^3)	2581	2797	2687	2398	2628	2643
LL-37 (2KFO) **37 aa.**						
N_α	29	25	28	0	30	25
ASA (Å^2)	2940	2998	2598	3519	2807	2629
Volume (Å^3)	4570	4661	4682	4311	4597	4777
Phylloseptin-2 (2JP1) **19aa.**						
N_α	13	7	10	0	13	12
ASA (Å^2)	1420	1256	1250	1666	1415	1362
Volume (Å^3)	2169	2290	2210	1981	2099	2140
Meucin-24 (2KFE) **24 aa.**						
N_α	19	12	18	6	17	16
ASA (Å^2)	1826	1882	1698	2058	1782	1661
Volume (Å^3)	2731	2775	2735	2655	2749	2767

Fowlicidin-1 (2AMN) **26 aa.**						
N_α	11	5	17	0	14	14
ASA (Å^2)	2077	2299	1935	2599	2079	2063
Volume (Å^3)	3218	3281	3187	2952	3208	3209
Fowlicidin-3 (2HFR) **27 aa.**						
N_α	12	9	17	0	12	16
ASA (Å^2)	2095	2156	1901	2487	2051	1964
Volume (Å^3)	3123	3127	3154	2908	3153	3146
Fowlicidin-2 (2GDL) **31 aa.**						
N_α	9	4	15	0	12	10
ASA (Å^2)	2466	2859	2128	3130	2564	2382
Volume (Å^3)	4035	3852	4164	3577	3823	4062
CAP18(106-137) (1LYP) **32 aa.**						
N_α	26	12	25	0	25	26
ASA (Å^2)	2561	2795	2425	3191	2473	2500
Volume (Å^3)	3997	3882	4030	3620	4081	3899
Ranatuerin-2CSa (2K10) **32 aa.**						
N_α	21	16	18	10	28	24
ASA (Å^2)	2057	2238	1991	2515	2127	2088
Volume (Å^3)	3383	3356	3383	3193	3288	3282

Peptides are coloured according to their experimental origin, as in tables 16-ABC.

The analysis of the data in table 17 shows that almost all the structures obtained in vacuum have the lowest ASA, maintain a good amount of helicity and often display more residues in α-helix than the experimental structures. In general, these models deviate from the experimental conformation to a greater extent than the structures obtained in attenuated octanol, except for the structure in TFE; in this case, as described before, the two strategies give similar results. The highest ASA is instead found for the conformers modelled in octanol, where the secondary structure is completely lost and all the residues are exposed, thus inducing a simulated “denaturation”. Helix structure is lost also in water, in particular for peptide structures determined in DPC; moreover, ASA in water is in some cases lower than the starting experimental value. These results are due to the fact that water solvation reproduces the hydrophobic effect, thus inducing the “collapse” of hydrophobic residues. Models obtained using attenuated octanol are characterized by values of ASA and levels of helicity intermediate between values of the initial structures and values of the vacuum models, thus giving a more realistic picture than in vacuum.

As a control, we also examined two β-sheets peptides, human β-defensin 1 and pig protegrin-1, whose NMR structure has been solved in water. These peptides possess disulphide bridges which make their structures more rigid with respect to helical peptides so that, whereas helical peptides are structured only in membranes (or membrane mimetics), defensins and protegrins are structured also in water. Therefore we modelled human β-defensin 1 and pig protegrin-1 in water, using the NMR structures solved in water as initial structures. The models obtained were very similar to the experimental structures (data not shown).

The main conclusions of our analysis are:

I. the attenuated implicit octanol solvation provides the best results when models are compared to structures obtained with SDS or DPC micelles;
II. modelling in vacuum and sometimes attenuated implicit octanol solvation provide the best results when models are compared to structures obtained in TFE;

These results will be useful for the development of non-demanding *ab initio* modelling procedures of CAMPs in membrane-like environments.

4. CONCLUSIONS

The work presented here is inserted in the wide field of research on cationic antimicrobial peptides (CAMPs), molecules thoroughly studied for their potential pharmacological use. Employing bioinformatic, experimental and computational approaches we have developed a panel of tools for the identification, production and characterization of new CAMPs.

Sequence studies permitted to define a novel scoring system capable of locating, with very good accuracy, putative CAMP-like fragments inside protein sequences. The main novelty of our method is the introduction in the scoring of the putative CAMPs of strain dependant variables which allow to search putative CAMPs particularly active against the strains of interest. A preliminary *in silico* validation shows that our scoring system accurately detect all the known antimicrobial fragments in proteins. We are currently automating the procedure in collaboration with Prof. O. Crescenzi (Department of Chemical Sciences, University of Naples, Federico II) in order to analyze large sets of proteins, like for example human secretome and hence to identify potential new human CAMPs. Preliminary results indicate that a very high number of potential new CAMPs are contained in our extracellular proteins, thus suggesting that the phenomenon of “cryptic” CAMPs is much more widespread than currently believed.

Parallely we have developed a novel fusion construct for the recombinant expression of CAMPs. This fusion construct allows to obtain pure peptides in high yield with just a single chromatographic step and mild conditions for the chemical cleavage of the peptide from the carrier. Moreover, it is very versatile allowing the production of relative long peptides (40-50 aa.) and $^{15}N/^{13}C$ labeled peptides, thus providing an interesting alternative to the expensive chemical synthesis. Our fusion system will make easier the characterization of the newly identified CAMPs.

Two human CAMPs, ThrAP and ApoE-AP, respectively located in thrombin and apolipoprotein E, were produced using our fusion construct. Their biological and structural characterization revealed that the two peptides possess complementary molecular and biological features, in fact, ApoE-AP displays a stronger antibacterial activity and no binding to CAMP-scavenging bacterial molecules like LPS and alginate, whereas ThrAP, characterized by a weaker antibacterial activity, is instead able to bind LPS and thus could prevent septic shock. Our data clearly suggest the two peptides are non equivalent from the pharmacological point of view and further widen the possible application of CAMPs in medicine.

Finally, the last part of this thesis focuses on preliminary studies of the structural modelling of CAMPs. Monte Carlo simulations performed using different implicit solvation functions allowed to define *in silico* conditions which reproduce with good accuracy the starting experimental structures. These data will be the stating point for the development of *ab initio* strategies for the prediction of CAMPs' structures.

Concluding, we believe that the methods we have developed will significantly stimulate the research in the field of antimicrobial peptide and the development of pharmacological strategies based on them.

APPENDIX

Antibacterial activity of basic and disulphide-rich proteins

5. BRIEF INTRODUCTION AND AIM

Lysozyme from *Gallus gallus* and human RNase 4 are two proteins which share an abundance of basic residues and cysteines and a relative small weight (around 14 kDa), as shown in table 18.

Protein	Primary structure	Isoelectric point
Gallus gallus lysozyme	KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNT QATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCS ALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDVQA WIRGCRL	9.32
Human RNase 4	QDGYYQRFMRQHLHVEETGGSDRYPNLMMQRRRMTLYHC CRFNTFIHEDIWNIRSICSTTNIQCKNGKMNCHEGVVKVTDC RDTGSSRAPNCRYRAIASTRRVVIACEGNPQVPVHFDG	9.05

Table 18: primary structures of *Gallus gallus* lysozyme (UniProt ID: P00698) and human RNase 4 (UniProt ID: P34096); basic residues, hydrophobic residues and cysteines are respectively colored in blue, green and yellow.

Lysozyme is an antibacterial protein which cleave the bacterial cell wall; it retains the antibacterial property also when the enzymatic activity is abolished by heat denaturation (During K. *et al.,* 1999) or by point mutations (Ibrahim H. R. *et al*., 2001a), and CAMP-like fragments are released upon digestion with pepsin (Ibrahim H. R. *et al.,* 2005; Ibrahim H. R. *et al*., 2001b). These studies point out that the bactericidal activity can also be independent from the catalytic one and the tertiary structure, and this peculiar phenomenon has been demonstrated for many different proteins as described in detail in the introduction of this thesis.

Several human ribonucleases are active protagonists of the immune system, exploiting their defense activity in various districts of the human body (Sorrentino S., 2010). Also for these proteins, the antibacterial activity can be separated from the enzymatic one, which consists of RNA degradation, and experimental evidences, which enforce this hypothesis, were already found for some human ribonucleases (Torrent M. *et al*., 2013) and homologous ribonucleases from *Gallus gallus* (Nitto T. *et al*., 2006), *Danio rerio* (Pizzo E. *et al*., 2011; Zanfardino A. *et al*., 2010) and *Salmo salar* (Pizzo E. *et al.,* 2008).

In conclusion, the presence of one or more potential CAMP-like regions in lysozyme and bactericidal RNases seems to be the major requisite for their activity against pathogens.

The main aim of this experimental work is to verify if *Gallus gallus* lysozyme and human ribonuclease 4 can be turned into more efficient carrier of antimicrobial regions if maintained in a stable denatured and reduced form through the alkylation of cysteines; two works (Pizzo E. *et al*., 2008; Pizzo E. *et al.,* 2011) show, in fact, that the antibacterial activity of RNases from *Danio rerio* and *Salmo salar* is enhanced when they are administered in a denaturated form, while Schroeder B. O. *et al.*, (2011) discovered that the antibacterial potential of human β-defensin 1 is augmented after the reduction of disulphide bridges. Whereas *Gallus gallus* lysozyme's antibacterial activity and cryptic CAMPs are already known, the biological activity of human RNase 4 is still unclear, and a possible bactericidal role is just an hypothesis (Sorrentino S., 2010).

6. MATERIALS AND METHODS

6.1 Materials

Ampicillin, bovine serum albumin (purity > 97%), IPTG, urea, DTT, Ellman's reagent, L-cysteine, guanidine chloride, agar and the alkylating agents (3-bromopropylamine hydrobromide, iodoacetamide and 4-vinylpyridine) were purchased from Sigma-Aldrich. *Gallus gallus* lysozyme with a grade of purity of 95% was purchased from Sigma-Aldrich and used without further purification. Trypton and yeast extract were purchased from Becton Dickinson. Sodium chloride and acrylamide (30% stock solution) were from Applichem. Trifluoroacetic acid and acetonitrile used for HPLC were purchased from Romil.

6.2 General procedures

Cell transformation and growth medium preparation were performed according to Sambrook J. *et al.*, (1989). SDS-PAGE was carried out according to Laemmli U. K. (1970). Protein concentrations were determined by the method of Bradford, using BSA as the standard (Bradford M. M., 1976) and by UV spectroscopy using the theoretical, sequence-based extinction coefficients in table 19 (Gill S. C. and von Hippel P. H.,1989).

Table 19: sequence-based extinction coefficients of the variants of *Gallus gallus* lysozyme and human RNase 4.

	Extinction coefficient of the native form (M^{-1} cm^{-1})	**Extinction coefficient of the alkylated form (M^{-1} cm^{-1})**
hRNase 4	11960	11460
***Gallus gallus* lysozyme**	37970	37470

The content of free cysteines after the alkylation reactions was assessed according to Ellman G. L. (1959).

6.3 Heterologous expression and preliminary purification of human RNase 4

Escherichia coli BL21(DE3) cells (AMS Biotechnology) were used for recombinant protein expression; cells transformed with pET 22b(+)-hRNase 4 were grown in 1 liter of Luria-Bertani medium containing ampicillin (0.1 mg/mL). When the culture reached an $A_{600\ nm}$ of 0.7 OD unit, protein expression was induced by the addition of 0.4 mM IPTG and the bacterial culture was grown over-night. Cells were harvested by centrifugation (7000 rpm, 4°C, 15', JA-14 rotor, Beckman) and pellets were lysed by sonication in 20 mL of lysis buffer [50 mM Tris-acetate, pH 8.4, containing 10 mM EDTA and protease inhibitor (Roche)] in an ultrasonic liquid processor (Misonix Ultrasonic Processor XL) at 20 kHz with 30" impulses, each followed by a 30" rest, for a 15' total time. The suspension was then centrifuged at 12000 rpm for 60' at 4°C (JA-25.50 rotor, Beckman). The inclusion bodies were freed from membrane proteins by two washes in 0.1 M Tris-acetate, containing 10 mM EDTA, 2 % Triton X-100 and 2 M urea, followed by repeated washes in 0.1 M Tris-acetate pH 8.4, containing 10 mM EDTA, to eliminate traces of Triton and urea. This procedure eliminated several contaminant proteins and cellular debris entrapped in inclusion body pellets.

Inclusion bodies were then dissolved in 0.1 M Tris-acetate pH 8.4, 10 mM EDTA, 6 M GuHCl and 25 mM DTT, purged with N_2, and incubated at 37°C for 3 h. The protein solution was acidified to pH 5 with glacial acetic acid and dialyzed over-night against 0.1 M acetic acid (pH 3) at 4°C. Any insoluble material was removed by centrifugation (12000 rpm, 30', 4°C, rotor JA-25-50, Beckman) and the supernatant, containing the RNase in the completely reduced form, was lyophilized.

6.4 Preparation of denatured and reduced *Gallus gallus* lysozyme

Gallus gallus lysozyme (lyophilized powder) was dissolved in 0.1 M Tris-acetate pH 8.4, 10 mM EDTA, 6 M GuHCl and 25 mM DTT at a final concentration of about 14 mg/mL, purged with N_2, and incubated at 37°C for 3 h. The protein solution was acidified to pH 5 with glacial acetic acid, dialyzed over-night against 0.1 M acetic acid (pH 3) at 4°C and finally lyophilized.

6.5 Alkylation of cystein residues

The optimal conditions for the alkylation reactions are shown in table 20. The lyophilized protein was dissolved in the adequate reaction buffer at a final concentration of 1 mg/mL, the appropriate amount of reactive was immediately added and the solutions were finally purged with N_2. At the end of incubation, the reactions were stopped by the addition of β-mercaptoethanol (at a final concentration double with respect to the alkylating agent), acidified to pH 5 with glacial acetic acid and exhaustively dialyzed against 20 mM AMAC, pH 4.5, at 4°C. Any insoluble material was removed by centrifugation (12000 rpm, 30', 4°C, rotor JA-25-50, Beckman).

Table 20: optimal conditions for the alkylation of cysteine residues.

Alkylating agent	Reaction buffer	Incubation	Ratio cysteine/alkylating agent
3-bromopropylamine hydrobromide (BPA)	Tris HCl 0.2 M, pH 9.5, EDTA 7 mM,GuHCl 6 M	24 hours 37°C	1:108
Iodoacetamide (IAA)	MES 0.2 M, pH 6.1, EDTA 7 mM, GuHCl 6 M	Over-night 25°C	1:10
4-vinylpyridine (VP)	Tris HCl 0.1 M, pH 8.5, EDTA 7 mM, GuHCl 6 M	Over-night 25°C	1:36

6.6 High pressure liquid cromatography (HPLC)

The chromatography was conducted on a Perkin-Elmer series 200 instrument and monitored at $\lambda_{278\ nm}$. The alkylated variants of hRNase 4 were loaded on a reverse-phase C-4 column (Phenomenex) equilibrated in 100% solution A (composed of 5% acetonitrile (v/v) and 0.1% trifluoroacetic acid (v/v)). The column was eluted with a gradient in which the concentration of solution B (composed of 95% acetonitrile (v/v) containing 0.1% trifluoroacetic acid (v/v)) was raised in 70'.

6.7 Acetic acid-urea Polyacrylamide Gel Electrophoresis

The electrophoretic migration of the alkylated variants of the two proteins was checked on a 12% polyacrilamide-gel containing 2 M urea and 50 mM acetic acid-NaOH, pH 4.5. The loading buffer contained 2 M urea, 10% glycerol, 50 mM acetic acid-NaOH, pH 4.5, and 0.01% bromophenol blu. The electrophoretic run was conducted for 3 hours at 150 V using 50 mM acetic acid-NaOH, pH 4.5 with 2 M urea as running buffer.

6.8 Bactericidal assays

See paragraph 2.9

7. RESULTS

7.1 Preparation of the alkylated variants of *Gallus gallus* lysozyme

The alkylation of cysteines permits to maintain the protein in a stable denatured and reduced form and to modulate protein net charge and/or hydrophobicity. The variant modified with 3-bromopropylamine (Lyz-PA), in fact, receives eight additional positive charges, whereas the variant modified with 4-vinylpyridine (Lyz-PE) contains eight aromatic groups which act as weak bases with a pKa ≈ 6. Finally, iodoacetamide adds a polar carboxamidomethyl group to each cysteine (Lyz-CAM) (Figure 22).

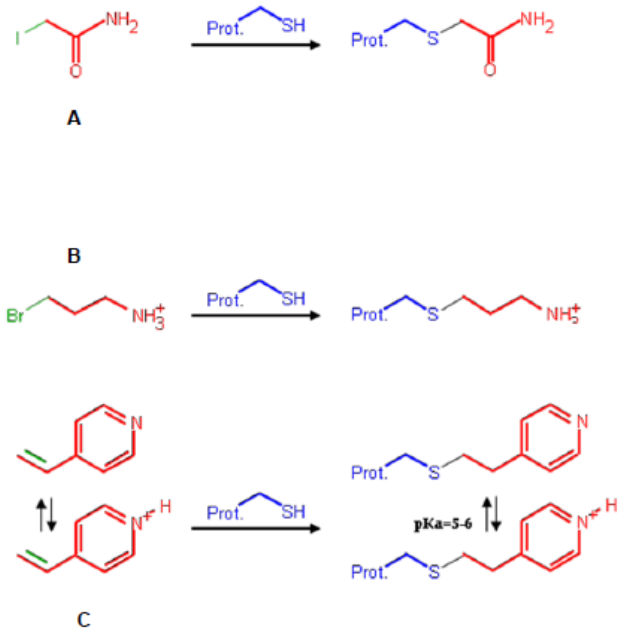


Figure 22: chemical structures of the alkylating agents: iodoacetamide (**A**), 3-bromopropylamine (**B**) and 4-vinylpyridine (**C**); the leaving group, the group that binds to cysteine and the cysteine residue are respectively coloured in green, red and blue.

The three alkylated variants were produced as described in material and methods and the free cysteines content, assessed with the Ellman test, was lower than 10%.

7.2 Acetic acid-urea Polyacrylamide Gel Electrophoresis of the alkylated variants of *Gallus gallus* lysozyme

Gallus gallus lysozyme's alkylated variants were analysed through an acetic acid-urea PAGE at pH 4.5. In these conditions, the velocity of migration depends on three factors: molecular weight, the compactness of the structure and net charge. A denatured protein maintains the same net charge and molecular weight of the native form, but is less compact and thus has a lower velocity of migration. As shown in figure 23, lysozyme variants display a reduced rate of migration compared to native lysozyme, as expected. The discrepancy between the variants can be explained considering that in Lyz-PE the pyridinic groups are not completely protonated at pH 4.5. In conclusion, the chemical modifications have a strong impact on lysozyme structure and the presence of an almost

single electrophoretic band for each variant suggests that they are homogeneously modified.

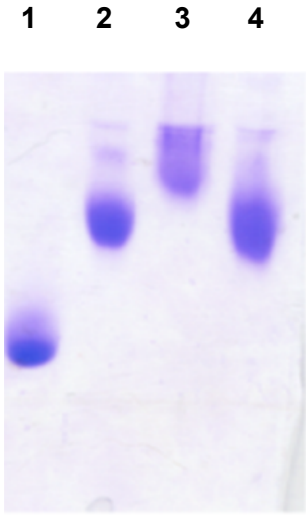


Figure 23: 12% acetic acid-urea PAGE at pH 4.5 of the alkylated variants of *Gallus gallus* lysozyme. **Lane 1**: native lysozyme (5 µg); **lane 2**: lyz-PA (5 µg); **lane 3**: lyz-PE (5 µg); **lane 4**: lyz-PA (5 µg).

7.3 Bactericidal activity of the alkylated variants of *Gallus gallus* lysozyme

As described in the introduction, the bactericidal activity of native lysozyme can be attributed both to its enzymatic activity and to the presence of CAMP-like regions in its primary structure; lysozyme's alkylated variants, instead, have completely lost their enzymatic activity (data not shown), and thus their bactericidal action can be explained only considering the presence of antimicrobial determinants in their sequence. The alkylated variants show levels of activity comparable and in some cases superior to the native form (Figure 24).

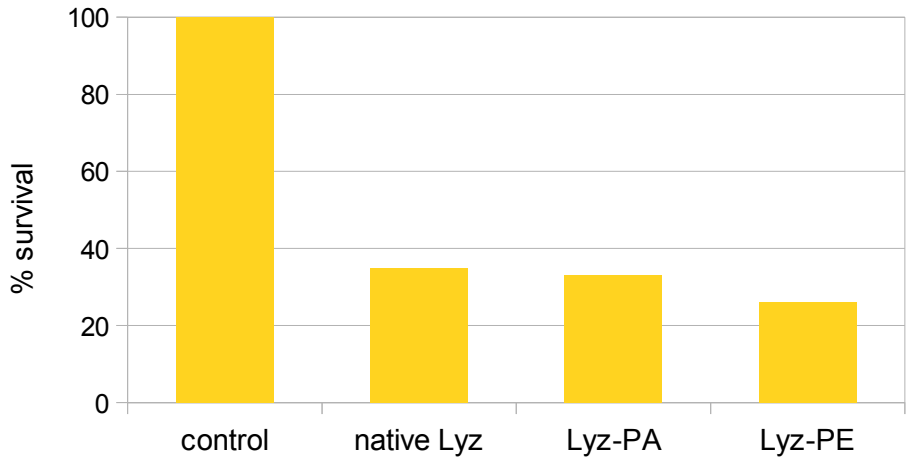


Figure 24: antibacterial assays of native *Gallus gallus* lysozyme and of two variants at a final concentration of 3 μM against *Staphilococcus aureus* ATCC 6538P. The relative error was always lower than the 10% of the measure.

7.4 Over-expression, alkylation and purification of human ribonuclease 4

The results observed with *Gallus gallus* lysozyme point out that the irreversible denaturation obtained through the chemical modifications of cysteines seems to be a valid strategy to enhance the antimicrobial potency of basic and disulphide-rich proteins and, consequently, the same study was performed with human ribonuclease 4. It should be noted that this protein, due to the human origin, could be more useful in the clinical field than the chicken protein.

The expression of human ribonuclease 4 was conducted as described in materials and methods and its level was analyzed by SDS-PAGE 15% (Figure 25).

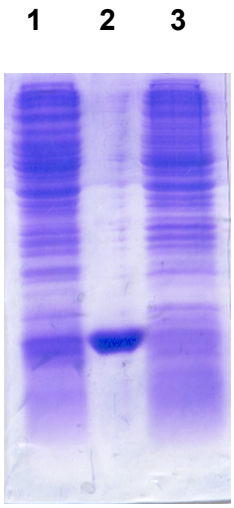


Figure 25: 15% SDS-PAGE with the analysis of the expression of hRNase 4. **Lane 1**: induced culture (0.126 OD); **lane 2**: RNase A (kDa, 5 µg); **lane 3**: non-induced culture (0.126 OD).

A densitometric analysis permitted to estimate a protein yield of about 20 mg per liter of culture. Cells were harvested by centrifugation and then lysed by sonication; the lisate was centrifuged in order to separate the soluble fraction from the inclusion bodies. The protein was expressed only in the insoluble fraction (data not shown) and was partially purified from membrane debris and other contaminants by several washes of the inclusion bodies with a buffer containing a detergent (Triton X-100) and a mild denaturing agent (Urea 2 M); aliquots of the supernatants of the washes were analyzed on SDS-PAGE 15% (Figure 26, lanes 3-7), in order to check any possible loss of protein during the preliminary purification. Inclusion bodies were finally dissolved in a denaturing and reducing buffer and were extensively dialyzed against a solution of acetic acid 0.1 M (pH 3), with the aim of maintaining the protein in a denatured and reduced state for the next step of chemical modification. An aliquot of the partially purified protein after dialysis was analyzed on SDS-PAGE 15% (Figure 26, lane 8).

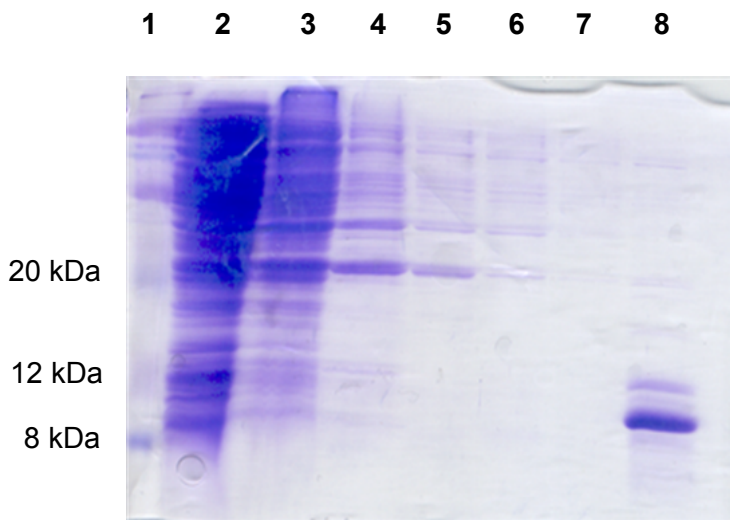


Figure 26: 15% SDS-PAGE of the preliminary purification of hRNase 4. **Lane 1**: molecular weight markers (Color-Burst, Sigma Aldrich); **lane 2**: soluble fraction after cell lysis (5 µL); **lanes 3-7**: supernatant of the five inclusion bodies washes (5 µL); **lane 8**: partially purified hRNase 4 (5 µg).

The protein was finally lyophilized and the three different variants (hRNase 4-PA, hRNase 4-CAM and hRNase 4-PE) were prepared as described in materials and methods and were finally purified to homogeneity by means of HPLC; chromatograms are shown in figure 27, whereas the SDS-PAGE analysis is shown in figure 28.

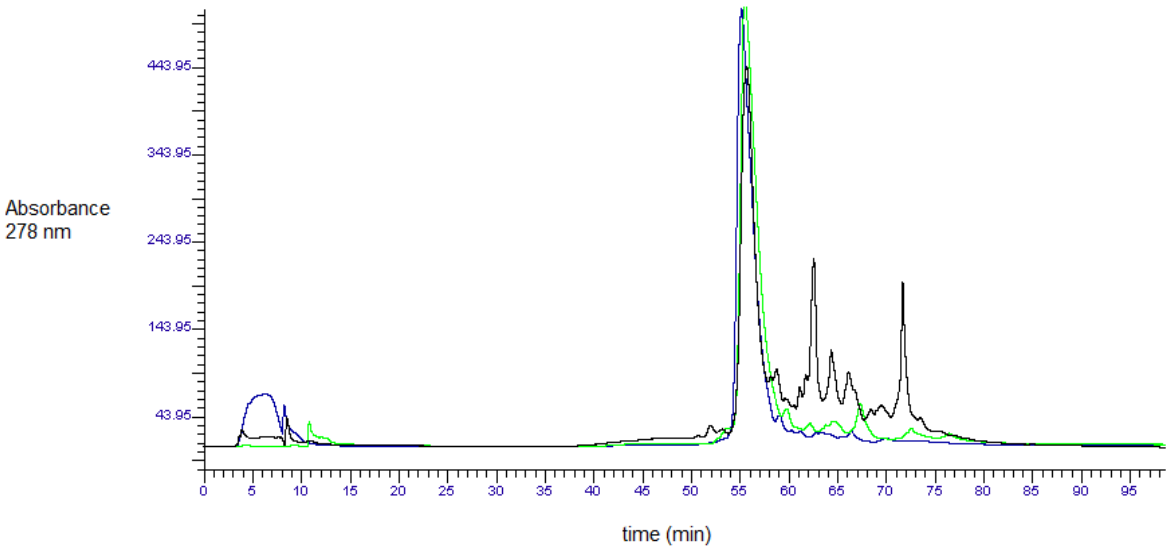


Figure 27: HPLC chromatograms of hRNase 4-CAM (black line), hRNase 4-PA (green line) and hRNase 4-PE (blue line).

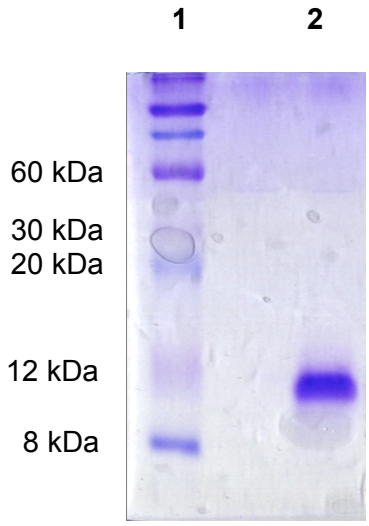


Figure 28: 15% SDS-PAGE of the purification of hRNase 4-CAM. **Lane 1**: molecular weight markers (Color-Burst, Sigma Aldrich); **lane 2**: protein after HPLC (5 µg).

The free cysteine content was lower than 10% for all the modified forms of hRNase 4.

7.5 Acetic acid-urea Polyacrylamide Gel Electrophoresis of the alkylated variants of human ribonuclease 4

The alkylated variants of hRNase 4 showed a pattern of electrophoretic migration analogous to the lyosozyme's modified forms, as shown in figure 29. Due to the difficulties in obtaining the native variant of hRNase 4, the migration of the denatured forms was compared to the migration of native RNase A.

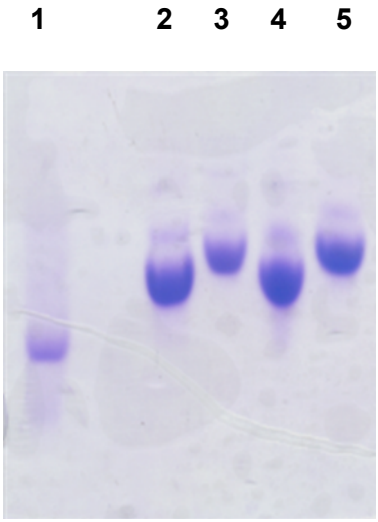


Figure 29: 12% acetic acid-urea PAGE at pH 4.5 of the alkylated variants of hRNase 4. **Lane 1**: native RNase A (5 µg); **lane 2**: hRNase 4-PA (5 µg); **lane 3**: hRNase 4-PE (5 µg); **lane 4**: hRNase 4-PA (5 µg); **lane 5**: hRNase 4-CAM (5 µg).

The modified form with the highest net charge, hRNase 4-PA, migrates faster than the other two modified forms, but more slowly than the native form, which is characterized by a more compact structure. In conclusion, as observed for the alkylated variants of lysozyme, the alklylation procedures did not generate a relevant heterogeneity of modified forms.

7.6 Bactericidal activity of the alkylated variants of human ribonuclease 4

The first antibacterial assays of the ribonuclease 4 variants hRNase 4-PA and hRNase 4-PE were conducted on different lab strains, including both Gram-negatives and Gram-positives; the results are displayed in figure 30.

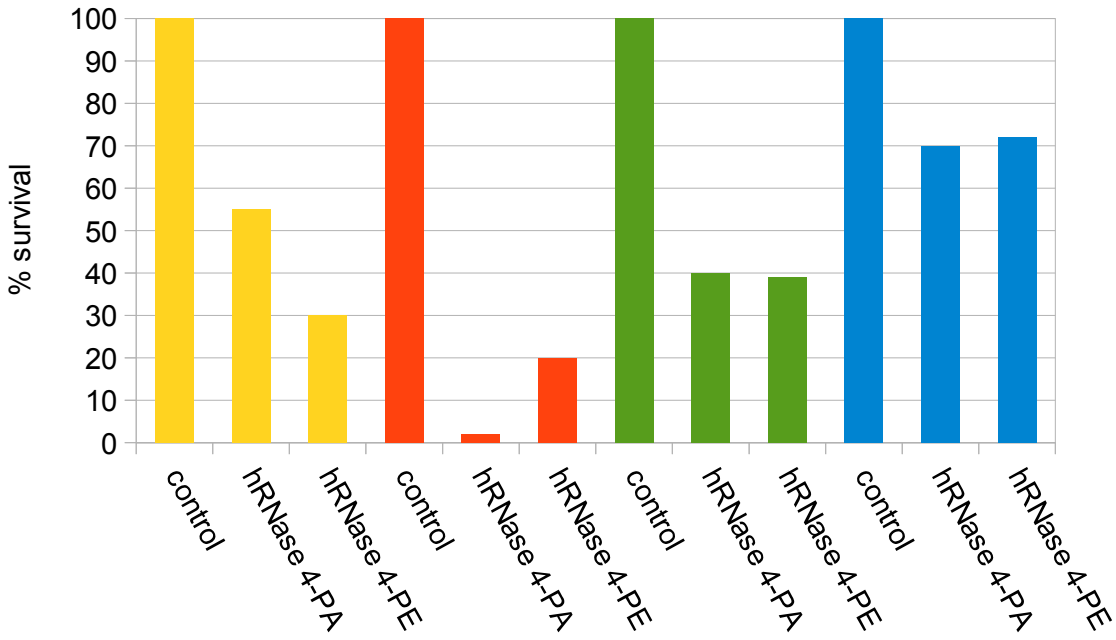


Figure 30: antibacterial assays of two variants of human ribonuclease 4 at a final concentration of 0.3 μM against *Staphilococcus aureus* ATCC 6538P (yellow), *Bacillus subtilis* PY79 (red), *Pseudomonas aeruginosa* PAO1 (green) and *Escherichia coli* DH5α (blue). The relative error was always lower than the 10% of the measure.

Both variants display an antibacterial activity, but it is not possible to define the best modification, as the level of activity seems to depend on the bacterial strain. This result is confirmed by the antibacterial essays conducted on different clinical isolates of *Pseudomonas aeruginosa*, kindly provided by D. Alessandra Bragonzi (San Raffaele Hospital, Milan), which are shown in figure 31.

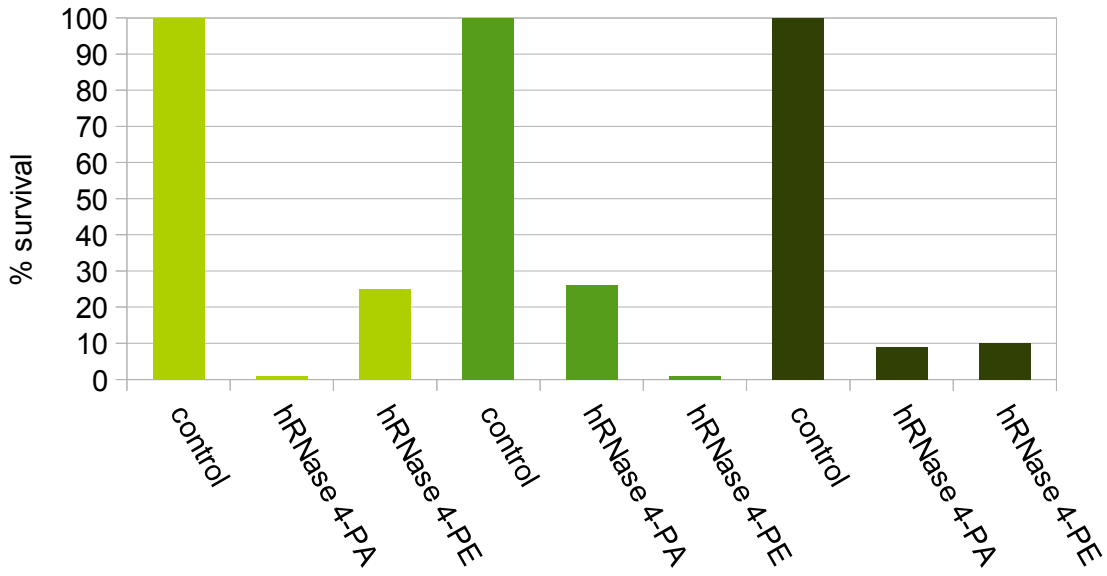


Figure 31: antibacterial assays of two variants of human ribonuclease 4 at a final concentration of 0.3 μM against three clinical isolates of *Pseudomonas aeruginosa*: AA2 (light green), PA14 (green) and BT72 (olive green). The relative error was always lower than the 10% of the measure.

8. CONCLUSIONS

The work reported here demonstrates that basic and disulphide-rich proteins with known or putative CAMP-like regions can be easily produced and maintained in a stable denatured and reduced form, with the possibility of obtaining variants with different biological actions by simply modifying the chemical nature of the group attached to cysteines. The denatured forms are active on both lab strains and clinical isolates from lungs of cystic fibrosis patients, and thus are promising therapeutic agents. The strategies described here can be applied to other basic and disulphide-rich human proteins and, moreover, other modifying agents could be tested, in order to obtain a broad panel of antimicrobial and pharmacologically relevant proteins.

BIBLIOGRAPHY

Andersson E. *et al.*, (2004). Antimicrobial activities of heparin-binding peptides. *European Journal of Biochemistry* **271:**1219-1226.

Aoki W. and Ueda M., (2013). Characterization of antimicrobial peptides toward the development of novel antibiotics. *Pharmaceuticals* **6:**1055-1081.

Bechinger B. and Salnikov E. S. (2012). The membrane interactions of antimicrobial peptides revealed by solid-state NMR spectroscopy. *Chemistry and Physics of Lipids* **165:**282-301.

Bhunia A. *et al.*, (2009). Lipopolysaccharide bound structures of the active fragments of fowlicidin-1, a cathelicidin family of antimicrobial and antiendotoxic peptide from chicken, determined by transferred nuclear Overhauser effect spectroscopy. *Biopolymers.* **92(1):**9-22.

Bobek L. A. and Situ H., (2003). MUC7 20-Mer: investigation of antimicrobial activity, secondary structure, and possible mechanism of antifungal action. *Antimicrob. Agents Chemother.* **47(2):**643-52.

Boix E. *et al.*, (2012). Structural determinants of the eosinophil cationic protein antimicrobial activity. *Biological Chemistry* **393:**801-815.

Bradford M. M., (1976). A rapid and sensitive method for the quantization of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem*. **72:**248-254.

Brand G. D. *et al.*, (2012). Probing protein sequences as sources for encrypted antimicrobial peptides. *PLoS ONE* **7(9):**e45848.

Chan C. *et al.*, (2004). Helix induction in antimicrobial peptides by alginate in biofilms. *The Journal of Biological Chemistry* **279(37):**38749-38754.

Chan C. *et al.*, (2005). Alginate as an auxiliary bacterial membrane: binding of membrane-

active peptides by polysaccharides. *J. Peptide Res.* **(65):**343–351.

Cowan R. and Whittaker R. G., (1990). Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography. *Pept. Res.* **3(2):**75-80.

D'Alessio G., (2011). Denatured bactericidal proteins: active per se, or reservoirs of active peptides? *FEBS Letters* **585:**2403-2404.

Darveau R. P. *et al.*, (1992). Peptides related to the carboxyl terminus of human platelet factor IV with antibacterial activity. *J. Clin. Invest.* **90(2):**447-55.

Dobson C. B. *et al.*, (2006). The receptor-binding region of human Apolipoprotein E has direct anti-infective activity. *The Journal of Infectious Diseases* **193:**442-450.

During K. *et al*., (1999). The non-enzymatic microbicidal activity of lysozymes. *FEBS Lett*. **23;449(2-3):**93-100.

Efremov R. G. *et al.*, (1999a). A solvent model for simulations of peptides in bilayers. I. Membrane-promoting alpha-helix formation. *Biophys. J.* **76(5):**2448-59.

Efremov R. G. *et al.*, (1999b). A solvent model for simulations of peptides in bilayers. II. Membrane-spanning alpha-helices. *Biophys. J.* **76(5):**2460-71.

Ellman G. L., (1959). Tissue sulfhydryl groups. *Arch. Biochem. Biophys*. **82(1):** 70–7.

Epand R. M. and Epand R. F., (2009). Domains in bacterial membranes and the action of antimicrobial agents. *Molecular ByoSystems*. **5:**580-587.

Falanga A. *et al.*, (2011). A peptide derived from herpes simplex virus type 1 glycoprotein H: membrane translocation and applications to the delivery of quantum dots. *Nanomedicine: Nanotechnology, Biology, and Medicine*. **7:**925-934.

Feig M. and Brooks C. L. 3rd, (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol*.

14(2):217-24.

Fjell C. D. *et al.*, (2009). Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem*. **52(7):**2006-15.

Frick I. M. *et al.*, (2006). The contact system - a novel branch of innate immunity generating antibacterial peptides. *EMBO J*. **25(23):**5569-78.

Gill S. C. and von Hippel P. H., (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem*.**182:**319–326.

Gopal R. *et al.*, (2012). Applications of circular dichroism for structural analysis of gelatin and antimicrobial peptides. *Int. J. Mol. Sci.* **13:**3229-3244.

Guilhelmelli F. *et al.*, (2013). Antibiotic development challenges: the various mechanisms of action of antimicrobial peptides and of bacterial resistance. *Frontiers in Microbiology* **4:**253.

He Y. *et al.*, (2013). Modeling peptide binding to anionic membrane pores. *Journal of Computational Chemistry* **34:**1463–1475.

Hoek K. S. *et al.*, (1997). Antibacterial activity of bovine lactoferrin-derived peptides. *Antimicrobial Agents and Chemotherapy*. **41(1):**54-59.

Hopfinger A. J. and Battershell R. D., (1976). Application of SCAP to drug design. 1. Prediction of octanol-water partition coefficients using solvent-dependent conformational analyses. *J. Med. Chem*. **19(5):**569-73.

Huang A. and Stultz C. M. (2007). Conformational sampling with implicit solvent models: application to the PHF6 peptide in tau protein. *Biophysical Journal*. **92(1):**34-45.

Huang C.-Jr *et al.,* (2012). Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancement. *J. Ind. Microbiol. Biotechnol.* **39:**383-399.

Ibrahim H. R. *et al*., (2001a). Genetic evidence that antibacterial activity of lysozyme is independent of its catalytic function. *FEBS Lett*. **28;506(1):**27-32.

Ibrahim H. R. *et al.*, (2001b). A helix-loop-helix peptide at the upper lip of the active site cleft of lysozyme confers potent antimicrobial activity with membrane permeabilization action. *Journal of Biological Chemistry*. **276(47):**43767-43774.

Ibrahim H. R. *et al*., (2005). Processing of lysozyme at distinct loops by pepsin: a novel action for generating multiple antimicrobial peptide motifs in the newborn stomach. *Biochimica et Biophysica Acta* **1726:**102-114.

Kagan B. L. *et al.*, (2012). Antimicrobial properties of amyloid peptides. *Mol. Pharm*. **9(4):**708-717.

Kalle M. I. *et al.*, (2013). Proteolytic activation transforms heparin cofactor II into a host defense molecule. *J. Immunol.* **190(12):**6303-10.

Kasetty G. *et al*., (2011a). Structure-activity studies and therapeutic potential of host defense peptides of human thrombin. *Antimicrobial Agents Chemotherapy* **55(6):**2880-2890.

Kasetty G. *et al*., (2011b). The C-terminal sequence of several human serine proteases encodes host defense functions. *Journal of Innate Immunity* **3:**471-482.

Khandelia H. *et al.*, (2006). Driving engineering of novel antimicrobial peptides from simulations of peptide-micelle interactions. *Biochimica et Biophysica Acta* **1758:**1224-1234.

Kovacs J. M. *et al.*, (2006).Determination of intrinsic hydrophilicity/hydrophobicity of amino acid side chains in peptides in the absence of nearest-neighbor or conformational effects. *Biopolymers.* **84:**283–297.

Laemmli U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227(5259)**:680-5.

Lam K. L. H. *et al.* (2012). Mechanism of structural transformations induced by antimicrobial peptides in lipid membranes. *Biochimica et Biophysica Acta* **1818:**194-204.

Landon M., (1977). Cleavage at aspartyl-proline bonds. *Methods Enzymol.* **47:**145-157.

Langham A. and Kaznessis Y. N., (2010). Molecular simulations of antimicrobial peptides. *Methods Mol. Biol.* **618:**267-285.

Lazaridis T. and Karplus M., (1999). Effective energy function for proteins in solution. *Proteins*. **35(2):**133-52.

Lazaridis T., (2003). Effective energy function for proteins in lipid membranes. *Proteins*. **52(2):**176-92.

Li Y., (2011). Recombinant production of antimicrobial peptides in *Escherichia coli*: a review. *Protein Expression and Purification.* **80(2):**260-267.

Maddox M. W. and Longo M. L., (2002). A Monte Carlo study of peptide insertion into lipid bilayers: equilibrium conformations and insertion mechanisms. *Biophys. J*. **82(1):**244-63.

Malmstem M. *et al.*, (2007). Antimicrobial peptides derived from growth factors. *Growth Factors* **25(1):**60-70.

Malmsten M. *et al.*, (2006). Bacterial killing by heparin-binding peptides from PRELP and thrombospondin. *Matrix Biol*. **25(5):**294-300.

Mátyus E. *et al.*, (2007). Computer simulation of antimicrobial peptides. *Current Medicinal Chemistry* **14:**2789-2798.

Mihajlovic M. and Lazaridis T., (2010). Antimicrobial peptides bind more strongly to membrane pores. *Biochimica et Biophysica Acta.* **1798:**1494-1502.

Monera O. D. *et al.*, (2005). Relationship of side-chain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J. Pept. Sci.*

1(5):310-329.

Mooney C. *et al.*, (2013). PeptideLocator: prediction of bioactive peptides in protein sequences. *Bioinformatics* **29(9):**1120-1126.

Niarchou A. *et al.*, (2013). C-PAmP: large scale analysis and database construction containing high scoring computationally predicted antimicrobial peptides for all the available plant species. *PLoS ONE* **8(11):**e79728.

Nibbering P. H. *et al.*, (2001). Human lactoferrin and peptides derived from its N terminus are highly effective against infections with antibiotic-resistant bacteria. *Infection and Immunity*. **69(3):**1469-1476.

Nitto T. *et al*., (2006). Evolution and function of leukocyte RNase A ribonucleases of the avian species Gallus gallus. *Journal of Biological Chemistry*. **281:**25622-25634.

Nordahl E. A. *et al.*, (2004). Activation of the complement system generates antibacterial peptides. *PNAS* **101(48):**16879-16884.

Notomista E. *et al.*, (1999). Effective expression and purification of recombinant onconase, an anti tumor protein. *FEBS Lett.* **463(3):**211-215.

Påhlman L. I. e*t al.*, (2013). Antimicrobial activity of fibrinogen and fibrinogen-derived peptides – a novel link between coagulation and innate immunity. *Thrombosis and Haemostasis* **109(5):**769-975.

Papareddy P. I. *et al.*, (2010). C-terminal peptides of tissue factor pathway inhibitor are novel host defense molecules. *J. Biol. Chem*. **285(36):**28387-98.

Papareddy P. I. *et al.*, (2012). Tissue factor pathway inhibitor 2 is found in skin and its C-terminal region encodes for antibacterial activity. *PLoS One*. **7(12):**e52772.

Parker J. M. R. *et al.*, (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*. **25*:***5425-5432.

Pasupuleti M. *et al.*, (2009). Antimicrobial activity of a C-terminal peptide from human extracellular superoxide dismutase. *BMC Res. Notes*. **15:2:**13.

Pizzo E. *et al*., (2008). Ribonucleases with angiogenic and bactericidal activities from the Atlantic salmon. *FEBS J.* **275(6):**1283-95.

Pizzo E. *et al*., (2011). A new RNase sheds light on the RNase/angiogenin subfamily from zebrafish. *Biochem J*. **433(2):**345-55.

Pulido D. *et al.*, (2011). Lipopolysaccharide neutralization by antimicrobial peptides: a gambit in the innate host defense strategy. *Journal of Innate Immunity* **4:**327-336.

Pushpanathan M. *et al.*, (2013). Antimicrobial peptides: versatile biological properties. *International Journal of Peptides* **2013:**675391.

Rakowska P.D. *et al.*, (2013). Nanoscale imaging reveals laterally expanding antimicrobial pores in lipid bilayers. *PNAS* **110(22):** 8918-8923.

Reed J. and Reed T. A., (1997). A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism. *Analytical Biochemistry.* **254:**36-40.

Roccatano D. *et al.*, (2002). Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study. *PNAS* **99(19):**12179-12184.

Ryan R. O. *et al.*, (2003). Optimized bacterial expression of human apolipoprotein A-I. *Protein Expr. Purif.* **27(1):**98-103.

Sambrook, J. *et al*., (1989). Molecular Cloning: A Laboratory Manual, *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY.

Sayyed-Ahmad A. *et al.*, (2009). Relative free energy of binding between antimicrobial peptides and SDS or DPC micelles. *Mol. Simul.* **35(10-11):**986-997.

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

Sengupta D. *et al.*, (2008). Toroidal pores formed by antimicrobial peptides show significant disorder. *Biochimica et Biophysica Acta*.**1778:**2308-2317.

Shafer W. M. *et al.*, (1993). Synthetic peptides of human lysosomal cathepsin G with potent antipseudomonal activity. *Infection and immunity* **61:**1900-1908.

Sonesson A. I. *et al.*, (2011). Antifungal activities of peptides derived from domain 5 of high-molecular-weight kininogen. *Int. J. Pept*. **2011:**761037.

Sonesson A. I. *et al.*, (2011). Thymic stromal lymphopoietin exerts antimicrobial activities. *Exp. Dermatol*. **20(12):**1004-10.

Sönnichsen F. D. *et al.*, (1992). Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry*. **31(37):**8790-8.

Sorrentino S., (2010). The eight human "canonical" ribonucleases: molecular diversity, catalytic properties, and special biological actions of the enzyme proteins. *FEBS Lett*. **584(11):**2194-200.

Tang M. and Hong M., (2009). Structure and mechanism of beta-hairpin antimicrobial peptides in lipid bilayers from solid-state NMR spectroscopy. *Molecular Biosystems* **5(4):**317-322.

Teixeira V. *et al.*, (2012). Role of lipids in the interaction of antimicrobial peptides with membranes. *Progress in Lipid Research* **51:**149–177.

Torrent M. *et al.*, (2009). A theoretical approach to spot active regions in antimicrobial proteins. *BMC Bioinformatics* **10**:373.

Torrent M. *et al.*, (2012). AMPA: an automated web server for prediction of protein antimicrobial regions. *Bioinformatics* **28(1)**:130-131.

Torrent M. *et al.*, (2013). Ribonucleases as a host-defence family: evidence of evolutionarily conserved antimicrobial activity at the N-terminus. *Biochemical Journal* **456(1)**:99-108.

van der Kraan M. I. A. *et al.*, (2004). Lactoferrampin: a novel antimicrobial peptide in the N1-domain of bovine lactoferrin. *Peptides*. **25:**177-183.

Weiner S. J. *et al.*, (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *JACS.* **106:**765-784.

Wiesner J. and Vilcinskas A., (2010). Antimicrobial peptides the ancient arm of the human immune system. *Virulence* **1(5):**440-464.

Wiradharma N. *et al.*, (2011). Synthetic cationic amphiphilic α-helical peptides as antimicrobial agents. *Biomaterials*. **(32):**2204-2212.

Zanfardino A. *et al*., (2010). The bactericidal action on Escherichia coli of ZF-RNase-3 is triggered by the suicidal action of the bacterium OmpT protease. *FEBS J*. **277(8):**1921-8.

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

Zhang Z. *et al.*, (2014). Four novel antimicrobial peptides derived from human C8a-MACPF. *Biotechnology Letters* **36(2):**319-325.