

Methodology article

Open Access

A theoretical approach to spot active regions in antimicrobial proteins

Marc Torrent*, Victòria M Nogués and Ester Boix

Address: Dpt. Bioquímica i Biologia Molecular, Fac. Biociències, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain

Email: Marc Torrent* - marc.torrent@uab.cat; Victòria M Nogués - victoria.nogues@uab.cat; Ester Boix - ester.boix@uab.cat

* Corresponding author

Published: 11 November 2009

Received: 10 March 2009

BMC Bioinformatics 2009, 10:373 doi:10.1186/1471-2105-10-373

Accepted: 11 November 2009

This article is available from: <http://www.biomedcentral.com/1471-2105/10/373>

© 2009 Torrent et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Much effort goes into identifying new antimicrobial compounds able to evade the increasing resistance of microorganisms to antibiotics. One strategy relies on antimicrobial peptides, either derived from fragments released by proteolytic cleavage of proteins or designed from known antimicrobial protein regions.

Results: To identify these antimicrobial determinants, we developed a theoretical approach that predicts antimicrobial proteins from their amino acid sequence in addition to determining their antimicrobial regions. A bactericidal propensity index has been calculated for each amino acid, using the experimental data reported from a high-throughput screening assay as reference. Scanning profiles were performed for protein sequences and potentially active stretches were identified by the best selected threshold parameters. The method was corroborated against positive and negative datasets. This successful approach means that we can spot active sequences previously reported in the literature from experimental data for most of the antimicrobial proteins examined.

Conclusion: The method presented can correctly identify antimicrobial proteins with an accuracy of 85% and a sensitivity of 90%. The method can also predict their key active regions, making this a tool for the design of new antimicrobial drugs.

Background

Host defence anti-microbial proteins and peptides are important participants of the innate immune response in most multicellular organisms [1]. The innate immune system comprises the cells and mechanisms that defend the host from infection by other organisms in a non-specific manner. Unlike the adaptive immune system, the innate immune system does not confer a long-lasting or protective immunity to the host, but is thought to constitute an evolutionarily older defence strategy. It remains the dominant immune system in plants, fungi and insects, and plays a crucial role during the first steps of infection in multicellular organisms.

One of the major achievements of medicine was the development of antibiotics, which can kill a broad spectrum of microorganisms. Unfortunately, the emergence of antibiotic resistance has become a clinical threat [2,3]. Cationic proteins and peptides that are involved in innate immunity represent an alternative strategy to conventional antibiotics [4]. A considerable variety of peptides with different size and structure are associated with antimicrobial activity in eukaryotic hosts. Antimicrobial cationic peptides have some common structural characteristics. They are composed of 12-50 amino acids, with 2-9 cationic residues and up to a 50% hydrophobic amino acids [5]. Many peptides do not adopt a well-

defined secondary structure in aqueous solution, but adopt a specific secondary structure upon binding to acidic phospholipids or within lipid bilayers [6].

A major limitation in the design of new antimicrobial peptides lies in the difficulty in finding new structures with low toxicity for the host and a broad spectrum of action against pathogens. An important group of antimicrobial peptides are derived from fragments released by proteolytic cleavage of larger proteins or derived from known antimicrobial regions of proteins involved in the innate immune system, such as the C-terminal domain of cathelicidins [7] or the N-terminus of bactericidal/permeability increasing protein (BPI) [8,9]. Additionally, antimicrobial peptides often display immunomodulation properties which can be applied in the design of new drugs for the treatment of immune system dysfunctions, such as autoimmune diseases [10]. Knowledge of the key regions in antimicrobial proteins is of great importance in deriving antimicrobial peptides.

Following this approach, we present a method of predicting potentially active regions of bactericidal proteins that can serve as templates to develop agents against antimicrobial resistance.

Results and discussion

Method description

Experimental data based on high throughput screening and database mining techniques show that certain residues are preferred in antimicrobial peptides. Hilpert et al. screened all amino acid substitutions for each position of a 12-mer peptide and tested all substitutions for activity, which was measured as the bactericidal half-maximal inhibitory concentration (IC₅₀) [11].

Bactenecin (also called bovine dodecapeptide), the antimicrobial peptide used by Hilpert et al., was discovered in neutrophils and is active against both Gram-negative and Gram-positive bacteria [12]. These authors used a variant of bactenecin called Bac2A (RLRIVVIRVAR-NH₂), where internal disulfide bridges had been eliminated. Bac2A had a high antimicrobial activity against both Gram-positive and Gram-negative bacteria, and was used as a template to test the effect of each amino acid substitution on the peptide's activity. The IC₅₀ value against the *Pseudomonas aeruginosa* was recorded for each synthetic peptide. Based on the screening results, we calculated a propensity value for each amino acid.

Although the bacterial strain for the high-throughput assay is Gram-negative, the peptide has also been tested against other strains [12]. The results have demonstrated that Bac2A is a peptide with broad antimicrobial action, which is also active against Gram-positive strains and

fungi. No differences were found between proteins with distinct antimicrobial properties in both the training and testing datasets. Our selected dataset covers a wide range of antimicrobial proteins that have activity on distinct strains from Gram-negative and Gram-positive bacteria, and the prediction data discard any potential bias. Nevertheless, some discrepancies may arise between the predicted and experimental regions if a specific determinant is found in a protein conferring some specificity for a particular microorganism or strain.

The calculated propensity index (Table 1), deduced from the corresponding IC₅₀ value for each amino acid substitution, makes a good reference for the assessment of the protein sequence determinants. Since low IC₅₀ values correspond to high antimicrobial activity, amino acids with a lower PV value represent the most favoured to take the part of an antimicrobial peptide. Positively charged residues (R, K) and some hydrophobic residues (W, Y, I, V) were favoured and present a low propensity index, whereas negatively charged residues are unfavoured and show a high propensity index. Positively charged residues would be necessary to direct antimicrobial proteins to the negatively charged bacterial cell wall and cytoplasmic membrane of bacteria, where they exert their antimicrobial effect [13]. Hydrophobic residues would be needed to interact with lipophilic regions of lipid bilayers in order to create pores or other destabilizing structures that lead to membrane depolarization or local disruption and eventually bacterial cell death [14]. Interestingly, W has the low-

Table 1: Bactericidal propensity values (PV) for each amino acid.

Amino acid	PV value
R	0.106
K	0.111
C	0.165
W	0.172
Y	0.185
I	0.198
V	0.200
H	0.202
N	0.240
T	0.242
F	0.246
L	0.246
Q	0.248
G	0.265
M	0.265
S	0.281
A	0.307
P	0.327
E	0.449
D	0.479

Calculations are based on the average half maximal inhibitory concentration, as detailed in the *Methods* section.

est PV value among the hydrophobic residues, whereas L has the highest value, while I and V rather than L, are preferred. In fact, W residues are known to be important for antimicrobial peptide action [15].

Because bactericidal action is generally located in a specific region, a sliding window system of 7 residues was chosen for the screening approach. To improve the predictive accuracy, 3 predictive lengths of 10, 12 and 14 residues were evaluated. For each length, the optimal number of permissible gaps was tested (2, 3 or 4 gaps). For each combination of parameters, an ROC curve was constructed, and the accuracy, sensitivity and selectivity of the method calculated in order to select the best parameters [see Additional file 1: Supplemental Figures S1, S2 and S3]. Optimal results were obtained using a predictive length of 12 amino acids allowing 2 gaps (Table 2). For these conditions, the AvPV threshold value was 0.225; thus residues with an AvPV <0.225 were considered favourable whereas residues with an AvPV >0.225 were considered unfavourable. Allowing a higher gap number drastically lowered both the specificity and sensitivity, so only 2 gaps were allowed. Using the selected cut-off value (a predictive length of 12 residues with 2 allowed gaps) we achieved an accuracy value of 85%, a sensitivity of 90%, and a specificity of 80% (Table 2).

Although this method detects antimicrobial regions with high accuracy, it may not be adequate in predicting antimicrobial regions with a high content of some specific amino acids, e.g. proline. Likewise, F residues are considered to favor antimicrobial activity [16], but this residue has a relatively high calculated index. Our method has been constructed with data provided from a peptide with particular amino acid content, and this may slightly alter the effect of some of the substitutions and the final output of their assigned propensity values. This may increase the number of target proteins missed, but do not alter the significance of the positive hits. In any case, the propensity indexes can be recalculated as other experimental data

Table 2: Evaluation of the tested parameter values for the selection of optimal predicting length and gap inclusion.

Predictive length (Gaps)	Sensitivity	Specificity
10 (2)	92%	78%
10 (3)	86%	82%
10 (4)	92%	78%
12 (2)	90%	80%
12 (3)	88%	82%
12 (4)	86%	84%
14(2)	86%	84%
14 (3)	86%	84%
14 (4)	84%	86%

become available, to update and improve the method prediction power.

Method testing and implementation

The proposed method was applied to a set of 100 proteins (50 bactericidal proteins and 50 non-bactericidal). Representative members of the main antimicrobial protein families described in the literature were included, and the results were compared and discussed with the available experimental data (Table 3).

Most of the known antimicrobial proteins were correctly identified and their active region was accurately predicted, when adequate information was available (Table 3).

To probe its reliability, the method was also applied to a positive testing dataset containing 20 antimicrobial proteins. It predicted 90% of the proteins in it. A negative testing dataset was also analyzed and 81% of the proteins were correctly identified as non-antimicrobial proteins. The results obtained are in good agreement with those presented for the training dataset [see Additional file 1: Supplemental Tables S1 and S2].

We describe below the main representative families of antimicrobial proteins that were examined, together with an exhaustive comparison between the predicted identified sequences and the experimentally active reported regions (Table 3).

Cathelicidins comprise a family of mammalian proteins expressed in epithelial and myeloid cells, involved in the innate immune response [10]. Cathelicidins contain a C-terminal cationic antimicrobial domain that becomes active on release from the N-terminal region of the holoprotein [17]. The most studied form is the human cathelicidin antimicrobial protein of 18 kDa (hCAP18). Its C-terminus (LL-37) has a wide spectrum of antimicrobial activity and other biological activities [18]. Analysis of the human cathelicidin sequence leads to the prediction that the LL-37 peptide region is responsible for antimicrobial activity (*hCAP 18*, Table 3).

This family comprises other known antimicrobial peptides active against a great variety of bacteria, including CAP11 or CRAMP, with a very low sequence identity and distinct assigned active regions. In all cases, antimicrobial regions predicted span the sequence corresponding to experimentally reported active peptides (*CAP11 and CRAMP*, Table 3) [19,20].

Bactericidal/permeability-increasing protein (BPI) is a 456-residue cationic protein stored in the polymorphonuclear leukocytes primary granules [21]. During phagocytosis and degranulation, proteases cleave BPI in the 236-241

Table 3: Detail of the predicted sequences from the studied antimicrobial proteins.

Proteins	Identified sequences	Experimental data reported
hCAP 18	RKSKEKIGKEFKRIVQRIKD	LLGDFFRKSKEKIGKEFKRIVORIKDFLRNLVPRTES [17]
CAP 11	LGRRFRRMVGLRKKFRKTRKRIQKLRKIGKTGRKV WKAW	<u>GLRKKFRKTRKRIQKLRKIGKTGRKVWKAW</u> REYGG IPYPCRI [19]
CRAMP	RFKKKISLRAGLLRKGK... EKLKIGQKIKNFFQ SVRFRVKETV	<u>ISLRAGLLRKGGEKIGEKLLKIGQKIKNFFQ</u> KLVPQPE [21]
BPI	SVHVHISKSKVWLIQLFHK NCIKISGKWKAQKRFLK	<u>KSKVGWLIQLFHK</u> [23] ISNANIKISGKWKAQKRFLKMSGNFDLSI*
Histone H2A	GRGKQGGKVRAKAKTRS//GRVHRLLRKG KKTRIIIPRHLQL	<u>SGRGKQGGKVRAKAKSSRSR</u> AGLQFPVGRVHRLLRK GNY [25]
Lysozyme	AKSRWYNQTPNRAKRVITTFRT	WDEAAVN <u>LAKSRWYN</u> [27]
Azurocidin	SGGRLSRFPFRVNV PNNVCTGVLTRRGGI	50% sequence identity with protegrin antimicrobial peptides
α-defensin	YGTCIYQGRWLWAFCC	<u>RYGTCIYQGRWLWAFS</u> [36]
β-defensin	TLQKYYCRVRGGRCVAVLSC KCSTRGRKCCRRKK	GIINT <u>LQKYYSRVGGR</u> <u>KSSTRGRKSSRRK</u> [37]
θ-defensin	GLRCICTRGFCRLL	<u>RCICTRGFCRLL</u> [39]
Magainin precursor	IGKFLHSAKKFGK	<u>GIGKFLHSAKKFGK</u> AFVGEIMNS [40]
Moronecidin	HIFRGIVHVGKTIHRLVT	FFH <u>HIFRGIVHVGKTIHKLVTG</u> [44]
Pleurocidin	FFKAAAHVKGKLVG	GWGS <u>FFKAAAHVKGKLVG</u> KAAALTHYL [45]
Bacteriocin enterocin	SCNKKGSCPGVKYGGKLLGG	KYYGNGV <u>SCNKKGCSV</u> [49]
Helveticin J	VVQKGNVGSYYVGLQLRKA	No active peptide reported
Sarcotoxin IA	WLKKIGKKIERVGG	W35 is described as a key residue in endotoxin neutralizing activity [52]
RNase 3	INNYRWRCKNQNTFLR VNVCGNQSIRCPHNRTLNNCHRFRVPL	TIAMRA <u>INNYRWRCKNQNTFLR</u> [61]
RNase 7	NINKHTKRCK SHGRVSLTMCKLTSGKYPNCRYKEKRQNKSYVVA	No active peptide reported
SPAG 11K	QLLRHPVKRAPIIRIP	N -terminus suggested [68]
Hepcidin	LCRFCCCKCRNKGGCYCKCF	No active peptide reported
Ace AMPI	ICPRVNRIVT VNTRNLRRACRCLVGVVNRNPGLRRNPR RNTFVRPFVWRPQICGRI	No active peptide reported

* United States Patent n° 5830860

The first column includes the identified sequences by the prediction approach. The corresponding active peptides previously described in the literature and tested experimentally are included in the adjacent column for comparison. Matching residues between the identified and experimentally active sequences are underlined and highlighted in bold.

region, releasing the fragment corresponding to the N-terminus end [8]. This fragment is responsible for its antibacterial activity [22]. Our method predicts 2 potential antimicrobial regions in the N-terminus of, in good agreement with the reported experimental data (BPI, Table 3). One of these regions (rBPI21) is reported to be clinical useful [23]. The other region is a potent endotoxin neutralizing peptide, thus providing a potential therapeutic value for peptides tested (USA Patent 5830860).

Histone H2A is one of the 5 main histone proteins involved in the structure of chromatin. Buforin I is a 39 amino acids peptide encoded by the same gene as histone H2A. A specific protease responsible for the generation of buforin I from histone H2A is in the crude extracts of the toad stomach, suggesting the presence of a specific functional regulation mechanism which converts toad histone H2A to buforin I. Moreover, a more potent antimicrobial peptide of 21 amino acids, buforin II, derived from buforin I [24,25] shows high antimicrobial activity

against a broad spectrum of microorganisms. The histone H2A sequence screening analysis shows 2 potential antimicrobial regions. The first predicted sequence is part of buforin II sequence (Histone H2A, Table 3). There seems to be no report of any experimental evidence for the second region predicted. However, the ratio between charged and hydrophobic residues makes this peptide a good target to test.

Lysozyme is an antibacterial protein with activity against Gram-positive and Gram-negative bacteria; muramidase activity is considered responsible for its bactericidal activity [26]. However, the denatured protein is also active, showing that lysozyme retains its antimicrobial activity when muraminidase activity is absent. In addition, some derived peptides from the lysozyme C-terminus, e.g. peptide A4 (residues 143-155), also retain antimicrobial activity [27,28]. These results agree with our prediction about this region in bacteriophage T4 lysozyme (Lysozyme, Table 3), further supporting the hypothesis that bacteri-

cidal activity of lysozyme is not uniquely attributed to its muramidase activity.

Serprocidins are 25-37 kDa serine proteases localized in neutrophil granules with cytotoxic activity against both Gram-negative and Gram-positive bacteria [29].

There is no evidence for a defined antimicrobial region in azurocidin, the human serprocidin. Although a peptide comprising the region 20-44 has been suggested [30], subsequent studies found no direct correlation between this segment and antimicrobial activity [31].

Our results also suggest that 2 other regions are involved in the bactericidal activity of azurocidin. Of special interest is the predicted antimicrobial region SGGRLSRF-PRFVNV that shows a 53% of identity with protegrins (*Azurocidin*, Table 3), a group of antimicrobial peptides in porcine leukocytes that exhibit *in vitro* broad-spectrum antimicrobial activity [32,33].

Defensins are a family of antimicrobial peptides [34] showing antimicrobial activity against Gram-negative and Gram-positive bacterial strains, fungi, and some parasites and enveloped viruses. Defensins can be classified by origin and structure. Vertebrate α and β defensins share a common fold and are mainly distinguished according to their disulfide bridge pattern, while θ defensins are α -derived cyclic peptides.

Our results predict a main antimicrobial region for α -defensins in agreement with published data (α -defensin, Table 3). The prediction screening points to the C-terminus of α -defensins as the key region that retain antimicrobial activity. Recent studies on the antimicrobial activity of α -defensins also support this hypothesis [35].

In analysing human β -defensin, we identified a potential active sequence at both the N- and C-termini. The C-terminal sequence of β -defensins retains the antimicrobial capacity of the whole protein against Gram-negative bacteria. Moreover, the N-terminal sequence seems to be necessary for activity against Gram-positive bacteria and fungi. Thus, the contribution of both N- and C-termini is not completely understood, but seems necessary to have broad spectrum antimicrobial activity [36,37]. These findings coincide with the prediction results, suggesting that the whole protein is involved in its antimicrobial action (β -defensin, Table 3).

Finally, we have predicted the active domain described in *Rhesus macaque* θ -defensin (θ -defensin, Table 3) [38,39].

Magainins are a class of antimicrobial peptides discovered in the skin of *Xenopus laevis* [40-42]. In analyzing the

magainin precursor that contains 5 copies of the active peptide, and our method can identify the active segments (*Magainin precursor*, Table 3).

Piscidins were the first antimicrobial cationic peptides to be isolated from the mast cells of striped bass fish [43]. We have analyzed the potential antimicrobial regions of 2 proteins: moronecidin and pleurocidin (*Moronecidin*, *pleurocidin*, Table 3). The predictions are in good agreement with the experimentally tested peptides, which, in both cases, correspond to the mature peptide released after the cleavage of the propeptide region from the protein [44,45].

Bacteriocins, antimicrobial proteins or peptides produced by bacteria, are expressed in lactic acid bacteria (LAB) [46,47].

We have analyzed bacteriocins, which have been the most studied peptides. For bacteriocin enterocin CRL35, the region predicted at its N-terminus overlaps with some sequences tested in literature (*Bacteriocin enterocin*, Table 3) [48]. We have also analyzed bacteriocin helveticin-J, and identified a potential antimicrobial region, but there is no data on the region responsible for its activity to corroborate this prediction (*Helveticin-J*, Table 3).

Sarcotoxin IA is a cecropin-like polycationic peptide that is active against a wide range of both Gram-positive and Gram-negative bacteria [49]. The natural 62 amino acid precursor of sarcotoxin is processed, resulting in a 39 amino acid long mature peptide [50]. No region has been described for sarcotoxin IA as primarily responsible for its antimicrobial action. However, key residues have been identified as important in the bactericidal activity of sarcotoxin IA, specifically W35, implicated in endotoxin-neutralizing activity [51]. W35 residue lies in the predicted potential antimicrobial region (*Sarcotoxin IA*, Table 3). Moreover, our predicted sequence shows 45% amino acid identity with salmocidins, an antimicrobial peptide group isolated from *Salmo gairdneri*, which has still to be characterized (deposited in SwissProt [P81369](#)).

Antimicrobial ribonucleases The RNase A family includes RNases with cytotoxic and antipathogen properties [52-54].

The eosinophil cationic protein (ribonuclease 3 or ECP) is a human host defence ribonuclease involved in inflammatory processes mediated by eosinophils [55,56]. ECP is a potent cytotoxic molecule, with bactericidal and helminthotoxic properties [57]. ECP antimicrobial activity is dependent on its action at the cytoplasmic membrane and bacteria wall [58,60]. Our recent experimental data confirm that the first predicted sequence (*RNase 3*,

Table 3) is involved in the protein bactericidal activity [61]. Screening of the membrane lysis and bactericidal activity of RNase 3 derived peptides corroborated that the protein N-terminus region retains most of its antimicrobial activity. The role of the second predicted region remains to be determined.

Ribonuclease 7 is an antimicrobial protein expressed in skin, liver, kidney, skeletal muscle and heart. RNase 7 has a high antimicrobial activity against *P. aeruginosa* and *P. Pastoris*, and a lower effect in *S. aureus* and *E. coli* cells [62-64]. RNase 7 site-directed mutagenesis studies indicate that some lysine clusters are necessary for the protein antimicrobial action, although not every cluster is of equal importance [64]. In this context, the regions predicted by our theoretical approach include a great number of lysine residues, giving them a high cationic content (RNase 7, Table 3). Nevertheless, the selected sequences present a low hydrophobic residue content in contrast to most known antimicrobial regions. As the RNase 7 antibacterial mechanism has not been identified, we cannot yet explain these results.

Conclusion

Recently, much attention has been paid to develop computational methods to screen and synthesize antimicrobial peptides. High throughput screening tests provide a powerful tool to design predictive methods. A method using high throughput screening that predicts antimicrobial action of peptides in *P. aeruginosa* has been published [65]. Other screening and bioinformatics approaches in the quest for new antimicrobial agents have recently been described. Lata et al. [66] designed a program to predict antimicrobial peptides, based in the observation that certain types of residues are preferred over others, particularly at the N- and C- termini. Using a support vector machine, this approach is a powerful tool with which to predict or identify antimicrobial peptides.

Our method differs from previously reported predictive algorithms because it applies experimentally derived high throughput screening values obtained in synthetic peptides to analyse protein sequences. Few attempts have been made to analyse bactericidal proteins to identify the structural determinants for their mechanism of action.

The proposed predictive screening approach has been applied to the main characterized antimicrobial protein families, allowing a direct comparison between the identified sequences and previous experimental data. We corroborate that the predicted sequences mostly match the regions experimentally reported by others. Moreover, our own experimental data testing a RNase 3 peptide collection was also proven successful [61]. Other antimicrobial proteins (SPAG 11K, Hecpudin and Ace AMP1) have also

been analysed and successfully predicted as antimicrobial proteins (Table 3). In these cases, no other experimental results have been reported, which precludes direct confirmation of the capacity of the prediction method. However, the results presented here provide an opportunity to study predicted peptides. To further optimize the method prediction power we are planning to update the provided indexes as new high throughput screening results become available.

Thus, this method can give a first approach in spotting the key regions of bactericidal proteins that give them their activity. The selected regions may provide a useful starting point in the development of new antimicrobial peptide derived drugs.

Methods

Sequence analysis

Protein sequences were obtained from the Swiss-Prot/TrEMBL Database <http://www.ebi.ac.uk/swissprot/>. Pairwise and multiple alignments were made with *ClustalW* <http://www.ebi.ac.uk/Tools/clustalw2>. Multiple alignments with known antimicrobial peptides were performed using the Antimicrobial Peptide Database <http://aps.unmc.edu/AP/main.php/>[67].

Data processing and model design

To predict potential antimicrobial regions, a bactericidal propensity index value (PV) was calculated for each amino acid. To ascribe a PV value to each, we took the experimental data of Hilpert et al. [11] where a complete library of 12 amino acid peptides was generated, starting from a template based on a linearized variant of bactericidin, a bovine antimicrobial peptide (Bac2A). From a high-throughput screening assay, activity against *Pseudomonas aeruginosa* for each amino acid substitution was tested and the bactericidal inhibitory concentration (IC₅₀) value for each peptide was estimated [11]. Using these experimental data, the average minimal inhibitory concentration (called the propensity value, PV) was calculated for each residue (Table 1). Each propensity value corresponds to the arithmetical average between the 12 positions tested in the peptide Bac2A. The standard error associated with each score has a constant value of approximately 0.05.

The screening for antimicrobial regions is based on the calculated bactericidal propensity values per amino acid (PV), using the IC₅₀ experimental values, as already discussed. The method calculates the average propensity value (AvPV) over a sliding window of the length of 7 residues and assigns this value to the central residue inside the window. The window size was selected to ensure the best signal/noise ratio for the tested sequences. The use of smaller window size may overestimate some regions,

whereas using larger window sizes may lose some information.

A protein is classified as antimicrobial if it has at least one antimicrobial region. To define a region in a sliding window system, 2 parameters were considered: the minimal length of the region (called the predictive length) and the allowed gap inclusion. In this case, 3 different predicting lengths of 10, 12 and 14 residues were tested, making allowance for 2, 3 or 4 gaps in each case.

Protein datasets

A training dataset of 50 antimicrobial proteins, comprising some of the best characterized antimicrobial protein families in the literature, was selected for evaluation. Some of the proteins sequence determinants have been reported experimentally, allowing a further checking of predicted results and experimental data. To complete the training dataset, 50 non-antimicrobial proteins recorded in the Swiss-Prot database were included. These proteins were randomly selected among those reported as soluble and having between 50 and 200 amino acid residues.

To assess the predictability of the method, a positive testing dataset has also been constructed, containing 20 antimicrobial proteins verified and annotated as antimicrobial according to Swiss-Prot. Another dataset, containing 20 soluble proteins has been constructed to provide a negative testing dataset. No proteins belonging to the training dataset were included in both cases [see Additional file 1: Supplemental Tables S1 and S2].

Evaluation of the method

To assess the performance of the method and determine the best cut-off AvPV value for each case, we used the Receiver Operating Characteristic (ROC) curves, considering the parameters described below:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100$$

$$\text{Specificity} = \frac{TN}{TN+FP} \times 100$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \times 100$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$$

True positive (TP) and true negative (TN) are correctly predicted as antimicrobial proteins and non-antimicrobial proteins, respectively. False positive (FP) and false negative (FN) are wrongly predicted antimicrobial proteins and predicted non-antimicrobial proteins, respectively. MCC is the Matthew's correlation coefficient.

For each case, we constructed an ROC curve and determined its parameters as described in Table 2. Once the best ROC curve had been gained, the cut-off AvPV value was evaluated and the protein sequences were assayed using these values.

Authors' contributions

MT designed the method, analyzed the data, and drafted the manuscript. EB participated in the data analysis, results discussion, and manuscript writing. VMN participated in the results discussion, critical reading and correction of the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Supplementary tables and figures. The data provided includes the statistical ROC analysis and a detailed description of the testing datasets used to validate the method.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-10-373-S1.doc>]

Acknowledgements

This project was funded by a grant from *Ministerio de Educación y Cultura*, Spain, (BFU2006-15543-C02-01) and by the "Fundació La Marató de TV3" (TV3-031110). MT is a recipient of a predoctoral fellowship from the *Generalitat de Catalunya*.

References

1. Scott MG, Hancock RE: **Cationic antimicrobial peptides and their multifunctional role in the immune system.** *Crit Rev Immunol* 2000, **20**:407-431.
2. Martinez JL, Baquero F, Andersson DI: **Predicting antibiotic resistance.** *Nat Rev Microbiol* 2007, **5**:958-965.
3. Wright GD: **The antibiotic resistome: the nexus of chemical and genetic diversity.** *Nat Rev Microbiol* 2007, **5**:175-186.
4. Mookherjee N, Hancock RE: **Cationic host defence peptides: innate immune regulatory peptides as a novel approach for treating infections.** *Cell Mol Life Sci* 2007, **64**:922-933.
5. Jenssen H, Hamill P, Hancock RE: **Peptide antimicrobial agents.** *Clin Microbiol Rev* 2006, **19**:491-511.
6. Lohner K, Blondelle SE: **Molecular mechanisms of membrane perturbation by antimicrobial peptides and the use of biophysical studies in the design of novel peptide antibiotics.** *Comb Chem High Throughput Screen* 2005, **8**:241-256.
7. Ramanathan B, Davis EG, Ross CR, Blecha F: **Cathelicidins: microbicidal activity, mechanisms of action, and roles in innate immunity.** *Microbes Infect* 2002, **4**:361-372.
8. Gray PW, Flagg G, Leong SR, Gumina RJ, Weiss J, Ooi CE, Elsbach P: **Cloning of the cDNA of a human neutrophil bactericidal protein.** *J Biol Chem* 1989, **264**:9505-9509.
9. Ooi CE, Weiss J, Doerfler ME, Elsbach P: **Endotoxin-neutralizing properties of the 25 kD N-terminal fragment and a newly isolated 30 kD C-terminal fragment of the 55-60 kD bactericidal/permeability-increasing protein of human neutrophils.** *J Exp Med* 1991, **174**:649-655.
10. Mookherjee N, Rehaume LM, Hancock RE: **Cathelicidins and functional analogues as antiseptics molecules.** *Expert Opin Ther Targets* 2007, **11**:993-1004.

11. Hilpert K, Volkmer-Engert R, Walter T, Hancock RE: **High-throughput generation of small antibacterial peptides with improved activity.** *Nat Biotechnol* 2005, **23**:1008-1012.
12. Wu M, Hancock RE: **Improved derivatives of bactenecin, a cyclic dodecameric antimicrobial cationic peptide.** *Antimicrob Agents Chemother* 1999, **43**:1274-1276.
13. Matsuzaki K, Nakamura A, Murase O, Sugishita K, Fujii N, Miyajima K: **Modulation of magainin 2-lipid bilayer interactions by peptide charge.** *Biochemistry* 1997, **36**:2104-2111.
14. Sitaram N, Nagaraj R: **Host-defense antimicrobial peptides: importance of structure for activity.** *Curr Pharm Des* 2002, **8**:727-742.
15. Chan DI, Prenner EJ, Vogel HJ: **Tryptophan- and arginine-rich antimicrobial peptides: structures and mechanisms of action.** *Biochim Biophys Acta* 2006, **1758**:1184-1202.
16. Wang G: **Structures of human host defense cathelicidin LL-37 and its smallest antimicrobial peptide KR-12 in lipid micelles.** *J Biol Chem* 2008, **283**:32637-32643.
17. Sorensen OE, Follin P, Johnsen AH, Calafat J, Tjabringa GS, Hiemstra PS, Borregaard N: **Human cathelicidin, hCAP-18, is processed to the antimicrobial peptide LL-37 by extracellular cleavage with proteinase 3.** *Blood* 2001, **97**:3951-3959.
18. Yu J, Mookherjee N, Wee K, Bowdish DM, Pistolic J, Li Y, Rehaume L, Hancock RE: **Host defense peptide LL-37, in synergy with inflammatory mediator IL-1beta, augments immune responses by multiple pathways.** *J Immunol* 2007, **179**:7684-7691.
19. Nagaoka I, Yomogida S, Tamura H, Hirata M: **Antibacterial cathelicidin peptide CAP11 inhibits the lipopolysaccharide (LPS)-induced suppression of neutrophil apoptosis by blocking the binding of LPS to target cells.** *Inflamm Res* 2004, **53**:609-622.
20. Gallo RL, Kim KJ, Bernfield M, Kozak CA, Zanetti M, Merluzzi L, Genaro R: **Identification of CRAMP, a cathelicidin-related antimicrobial peptide expressed in the embryonic and adult mouse.** *J Biol Chem* 1997, **272**:13088-13093.
21. Schultz H, Weiss JP: **The bactericidal/permeability-increasing protein (BPI) in infection and inflammatory disease.** *Clin Chim Acta* 2007, **384**:12-23.
22. Gazzano-Santoro H, Parent JB, Grinna L, Horwitz A, Parsons T, Theofan G, Elsbach P, Weiss J, Conlon PJ: **High-affinity binding of the bactericidal/permeability-increasing protein and a recombinant amino-terminal fragment to the lipid A region of lipopolysaccharide.** *Infect Immun* 1992, **60**:4754-4761.
23. Levin M, Quint PA, Goldstein B, Barton P, Bradley JS, Shemie SD, Yeh T, Kim SS, Cafaro DP, Scannon PJ, Giroir BP: **Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal sepsis: a randomised trial.** *rBPI21 Meningococcal Sepsis Study Group.* *Lancet* 2000, **356**:961-967.
24. Park CB, Kim MS, Kim SC: **A novel antimicrobial peptide from *Bufo bufo gargarizans*.** *Biochem Biophys Res Commun* 1996, **218**:408-413.
25. Kim HS, Park CB, Kim MS, Kim SC: **cDNA cloning and characterization of buforin I, an antimicrobial peptide: a cleavage product of histone H2A.** *Biochem. Biophys Res Commun* 1996, **229**:381-387.
26. Tsugita A: **Phage Lysozyme and Other Lytic Enzymes.** In *The Enzymes* 3rd edition. Edited by: Boyer PD. New York: Academic Press; 1971:343-411.
27. Doring K, Porsch P, Mahn A, Brinkmann O, Gieffers W: **The non-enzymatic microbicidal activity of lysozymes.** *FEBS Lett* 1999, **449**:93-100.
28. Ibrahim HR, Matsuzaki T, Aoki T: **Genetic evidence that antibacterial activity of lysozyme is independent of its catalytic function.** *FEBS Lett* 2001, **506**:27-32.
29. Levy O: **Antimicrobial proteins and peptides of blood: templates for novel antimicrobial agents.** *Blood* 2000, **96**:2664-2672.
30. Pereira HA, Erdem I, Pohl J, Spitznagel JK: **Synthetic bactericidal peptide based on CAP37: a 37-kDa human neutrophil granule-associated cationic antimicrobial protein chemotactic for monocytes.** *Proc Natl Acad Sci USA* 1993, **90**:4733-4737.
31. McCabe D, Cukierman T, Gabay JE: **Basic residues in azurocidin/HBP contribute to both heparin binding and antimicrobial activity.** *J Biol Chem* 2002, **277**:27477-27488.
32. Steinberg DA, Hurst MA, Fujii CA, Kung AH, Ho JF, Cheng FC, Loury DJ, Fiddes JC: **Protegrin-1: a broad-spectrum, rapidly microbicidal peptide with in vivo activity.** *Antimicrob Agents Chemother* 1997, **41**:1738-1742.
33. Aumelas A, Mangoni M, Roumestand C, Chiche L, Despaux E, Grassy G, Calas B, Chavanieu A: **Synthesis and solution structure of the antimicrobial peptide protegrin-1.** *Eur J Biochem* 1996, **237**:575-583.
34. Yang D, Liu ZH, Tewary P, Chen Q, de la Rosa G, Oppenheim JJ: **Defensin participation in innate and adaptive immunity.** *Curr Pharm Des* 2007, **13**:3131-3139.
35. Lundy FT, Nelson J, Lockhart D, Greer B, Harriott P, Marley JJ: **Antimicrobial activity of truncated alpha-defensin (human neutrophil peptide (HNP)-1) analogues without disulphide bridges.** *Mol Immunol* 2008, **45**:190-193.
36. Pazgier M, Prah A, Hoover DM, Lubkowsky J: **Studies of the biological properties of human beta-defensin 1.** *J Biol Chem* 2007, **282**:1819-1829.
37. Hoover DM, Wu Z, Tucker K, Lu W, Lubkowsky J: **Antimicrobial characterization of human beta-defensin 3 derivatives.** *Antimicrob Agents Chemother* 2003, **47**:2804-2809.
38. Selsted ME: **Theta-defensins: cyclic antimicrobial peptides produced by binary ligation of truncated alpha-defensins.** *Curr Protein Pept Sci* 2004, **5**:365-371.
39. Tran D, Tran PA, Tang YQ, Yuan J, Cole T, Selsted ME: **Homodimeric theta-defensins from rhesus macaque leukocytes: isolation, synthesis, antimicrobial activities, and bacterial binding properties of the cyclic peptides.** *J Biol Chem* 2002, **277**:3079-3084.
40. Matsuzaki K: **Magainins as paradigm for the mode of action of pore forming polypeptides.** *Biochim Biophys Acta* 1998, **1376**:391-400.
41. Maloy WL, Kari UP: **Structure-activity studies on magainins and other host defense peptides.** *Biopolymers* 1995, **37**:105-122.
42. Soravia E, Martini G, Zasloff M: **Antimicrobial properties of peptides from *Xenopus granular gland secretions*.** *FEBS Lett* 1988, **228**:337-340.
43. Noga EJ, Silphaduang U: **Piscidins: a novel family of peptide antibiotics from fish.** *Drug News Perspect* 2003, **16**:87-92.
44. Lauth X, Shike H, Burns JC, Westerman ME, Ostland VE, Carlberg JM, Van Olst JC, Nizet V, Taylor SW, Shimizu C, Bulet P: **Discovery and characterization of two isoforms of moronecidin, a novel antimicrobial peptide from hybrid striped bass.** *J Biol Chem* 2002, **277**:5030-5039.
45. Cole AM, Weis P, Diamond G: **Isolation and characterization of pleurocidin, an antimicrobial peptide in the skin secretions of winter flounder.** *J Biol Chem* 1997, **272**:12008-12013.
46. Drider D, Fimland G, Hechard Y, McMullen LM, Prevost H: **The continuing story of class IIa bacteriocins.** *Microbiol Mol Biol Rev* 2006, **70**:564-582.
47. Klaenhammer TR: **Genetics of bacteriocins produced by lactic acid bacteria.** *FEMS Microbiol Rev* 1993, **12**:39-85.
48. Salvucci E, Saavedra L, Sesma F: **Short peptides derived from the NH2-terminus of subclass IIa bacteriocin enterocin CRL35 show antimicrobial activity.** *J Antimicrob Chemother* 2007, **59**:1102-1108.
49. Okada M, Natori S: **Purification and characterization of an antibacterial protein from haemolymph of *Sarcophaga peregrina* (flesh-fly) larvae.** *Biochem J* 1983, **211**:727-734.
50. Skosyrev VS, Kuleskiy EA, Yakhnin AV, Temirov YV, Vinokurov LM: **Expression of the recombinant antibacterial peptide sarco-toxin IA in *Escherichia coli* cells.** *Protein Expr Purif* 2003, **28**:350-356.
51. Okemoto K, Nakajima Y, Fujioka T, Natori S: **Participation of two N-terminal residues in LPS-neutralizing activity of sarco-toxin IA.** *J Biochem* 2002, **131**:277-281.
52. Boix E, Nogues MV: **Mammalian antimicrobial proteins and peptides: overview on the RNase A superfamily members involved in innate host defence.** *Mol Biosyst* 2007, **3**:317-335.
53. Rosenberg HF: **RNase A ribonucleases and host defense: an evolving story.** *J Leukoc Biol* 2008, **83**:1079-1087.
54. Pizzo E, D'Alessio G: **The success of the RNase scaffold in the advance of biosciences and in evolution.** *Gene* 2007, **406**:8-12.
55. Boix E, Torrent M, Sanchez D, Nogues MV: **The antipathogen activities of eosinophil cationic protein.** *Curr Pharm Biotech* 2008, **9**:141-152.

56. Venge P, Bystrom J, Carlson M, Hakansson L, Karawacjzyk M, Peterson C, Seveus L, Trulsson A: **Eosinophil cationic protein (ECP): molecular and biological properties and the use of ECP as a marker of eosinophil activation in disease.** *Clin Exp Allergy* 1999, **29**:1172-1186.
57. Boix E: **Eosinophil cationic protein.** *Methods Enzymol* 2001, **341**:287-305.
58. Torrent M, Cuyas E, Carreras E, Navarro S, Lopez O, de la Maza A, Nogues MV, Reshetnyak YK, Boix E: **Topography studies on the membrane interaction mechanism of the eosinophil cationic protein.** *Biochemistry* 2007, **46**:720-733.
59. Torrent M, Navarro S, Moussaoui M, Nogues MV, Boix E: **Eosinophil cationic protein high affinity binding to bacteria wall lipopolysaccharides and peptidoglycans.** *Biochemistry* 2008, **47**:3544-3555.
60. Carreras E, Boix E, Rosenberg HF, Cuchillo CM, Nogues MV: **Both aromatic and cationic residues contribute to the membranalytic and bactericidal activity of eosinophil cationic protein.** *Biochemistry* 2003, **42**:6636-6644.
61. Torrent M, De la Torre BG, Nogués MV, Andreu D, Boix E: **Bactericidal and membrane disruption activities of the eosinophil cationic protein are largely retained in an N-terminal fragment.** *Biochem J* 2009, **421**:425-434.
62. Zhang J, Dyer KD, Rosenberg HF: **Human RNase 7: a new cationic ribonuclease of the RNase A superfamily.** *Nucleic Acids Res* 2003, **31**:602-607.
63. Harder J, Schroder JM: **RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin.** *J Biol Chem* 2002, **277**:46779-46784.
64. Huang YC, Lin YM, Chang TW, Wu SJ, Lee YS, Chang MD, Chen C, Wu SH, Liao YD: **The flexible and clustered lysine residues of human ribonuclease 7 are critical for membrane permeability and antimicrobial activity.** *J Biol Chem* 2007, **282**:4626-4633.
65. Jenssen H, Lejon T, Hilpert K, Fjell CD, Cherkasov A, Hancock RE: **Evaluating different descriptors for model design of antimicrobial peptides with enhanced activity toward P. aeruginosa.** *Chem Biol Drug Des* 2007, **70**:134-142.
66. Lata S, Sharma BK, Raghava GP: **Analysis and prediction of antibacterial peptides.** *BMC Bioinformatics* 2007, **8**:263.
67. Wang Z, Wang G: **APD: the Antimicrobial Peptide Database.** *Nucleic Acids Res* 2004, **32**:D590-2.
68. Yenugu S, Hamil KG, French FS, Hall SH: **Antimicrobial actions of human and macaque sperm associated antigen (SPAG) 11 isoforms: influence of the N-terminal peptide.** *Mol Cell Biochem* 2006, **284**:25-37.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

