BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 11(1)

Genome-wide identification of antimicrobial peptides in the liver fluke, *Clonorchis sinensis*

Won Gi Yoo[1#], Sanghyun Lee[1#], Myoung-Ro Lee[2], Mi-Ran Yun[1], Taesoo Kwon[1] & Dae-Won Kim[1*]

[1]Division of Biosafety Evaluation and Control, Korea National Institute of Health, Chungbuk 363-951, Republic of Korea; [2]Division of Malaria and Parasitic Diseases, Centre for Immunology and Pathology, Korea National Institute of Health, Chungbuk 363-951, Republic of Korea; Dae-Won Kim – Email: todaewon@gmail.com; Phone: +82-43-719-8529; Fax: +82-43-719-8059; *Corresponding author
#Authors equaliy contributed

Received December 19, 2014; Accepted January 01, 2015; Published January 30, 2015

Abstract:
The increase in prevalence of antimicrobial resistance makes the search for new antibiotic agents imperative. Antimicrobial peptides (AMPs) from natural resources have been recognized as suitable tools to combat antibiotic-resistant bacteria. The liver fluke *Clonorchis sinensis* living in germ-filled environments could be a good source of antimicrobials. Here, we report the use of a rational protocol that combines AMP predictions based on their physicochemical properties and their *in vivo* stability to discover AMP candidates from the entire genome of *C. sinensis*. To screen AMP candidates, *in silico* analyses based on the physicochemical properties of known AMPs, such as length, charge, isoelectric point, and *in vitro* and *in vivo* aggregation values were performed. To enhance their *in vivo* stability, proteins having proteolytic cleavage sites were excluded. As a consequence, four high-activity, high-stability peptides were identified. These peptides could be potential starting materials for the development of new AMPs via structural modification and optimization. Thus, this study proposes a refined computational method to develop new AMPs and identifies four AMP candidates, which could serve as templates for further development of peptide antibiotics.

Keywords: Antibiotics, Infection, Antibacterial agent, Bioinformatics, Rational drug design

Background:
Infectious diseases remain a significant threat to public health, even with the development of antimicrobial chemotherapy. The extensive use of antibiotics has contributed to the emergence of antimicrobial resistance and the reduced efficacy of the available antimicrobial agents **[1].** Therefore, identification of novel antibiotics is urgently required to overcome the challenges presented by the emergence of multiple drug-resistant microorganisms **[2].**

Natural molecules provide a much wider and larger chemical space than that covered by synthetic molecules. Moreover, Natural molecules often have surprisingly beneficial properties in terms of penetration, absorption, and solubility **[3].** Antimicrobial peptides (AMPs) originating from natural resources have been recognized as a novel class of antibiotics to treat antibiotic-resistant bacterial infections **[4].** AMPs, which are typically less than 50 amino acids in length, are naturally occurring molecules that play a central role in the innate immune system and have specific physicochemical properties, such as charge, isoelectric point (pI) and aggregation **[5].**

Organisms living in germ-filled environments, such as liver flukes, are considered as an abundant source of antimicrobials **[3].** Such organisms make these natural molecules for their own needs; thus, these molecules may be deficient in certain medicinal properties required in humans. After ingestion by mammals, the metacercariae of the liver fluke excyst in the duodenum and larval flukes then migrate up into the small intrahepatic bile ducts **[6].** Based on this life cycle, flukes are vulnerable to microbes from both the external and internal environment. Indeed, a previous study showed that there is a

diverse microbial community associated with liver flukes and that some species are specific to the biliary system **[7],** suggesting that liver flukes may have the ability to produce various AMPs.

Currently, bioinformatic techniques are used to predict peptide sequences by a theoretical survey of the target transcriptome **[5].** The search for peptides using *in silico* analysis is performed by correlating the characteristics of AMPs previously reported in the literature with amino acid sequences. However, further studies are needed to optimize methods for *in silico* analysis.

Therefore, the aim of this study was to identify novel AMP candidates through refined computational methods.

Methodology:

Identification of suitable candidates based on physicochemical properties and a machine learning algorithm

AMP candidates were picked according to the previously observed physicochemical properties of reported AMPs. All of the protein-coding genes were calculated using EMBOSS PEPSTATS **[8]** in terms of molecular weight (MW), number of residues (≤ 80-mer), pI (8 ≤ pI ≤ 12) **[9],** and a net positive charge [10]. *In vitro* and *in vivo* aggregation levels were calculated with default parameters using TANGO (AGG ≤ 500, 0 ≤ HELIX ≤ 25, 25 ≤ BETA ≤ 100) **[11]** and AGGRESCAN (-40 ≤ Na4vSS ≤ 60) **[12],** respectively. The continuous sequence of residues in each AMP candidate was determined using AMPA **[13].** These properties reflect the binding of AMPs to negatively charged microbial cell membranes, which results in bacterial cell death via a pore or carpet mechanism **[14].** In order to exclude non-AMP candidates, all *Clonorchis sinensis* protein sequences were examined using the artificial neural network (ANN) algorithm of CAMP at the default value **[15].**

In vivo stability

To enhance *in vivo* stability, *C. sinensis* protein sequences that had proteolytic cleavage sites were excluded. Proteolytic cleavage sites were identified using PeptideCutter (http://web.expasy.org/peptide_cutter/) for chymotrypsin and EMBOSS EPESTFIND **[8]** for PEST motifs, mainly enriched in proline (P), glutamic acid (E), serine (S), threonine (T).

Results & Discussion:

Parasites as a potential source of antimicrobials

The development of microbial resistance to AMPs is unlikely because it is difficult for microbes to alter cell-structural elements. Differences in membrane composition between microbes and higher eukaryotes are important for selectively targeting microbial membranes, reducing toxicity, and improving therapeutic use **[15]**. Recent research on multidrug resistant bacteria has shown that organisms living in germ-filled environments have developed means of protecting themselves against microbial pathogens **[3].** Such organisms could be very good sources of antimicrobials. Therefore, based on the life cycle of the liver fluke *C. sinensis*, this worm could be an untapped source of new antimicrobials.

Stepwise identification of novel antimicrobials

Because of the low sequence similarity of AMPs, novel AMP candidates from *C. sinensis* (CsAMPs) were identified by a computational approach following typical features of known AMPs **[16].** From all of the protein-coding genes collected, CsAMP candidates were screened for antimicrobial effects based on the following criteria **(Figure 1).** First, CsAMP candidates were examined for physicochemical properties of known AMPs, i.e., length (≤ 80-mer) **[17],** pI (8 ≤ pI ≤ 12), no aggregation *in vitro* but aggregation *in vivo* **[9]**, and a positive charge **[10].** Second, CsAMP candidates were examined based on *in vivo* stability of peptides, wherein CsAMP candidates having either chymotrypsin or PEST motifs as proteolytic cleavage sites were excluded.

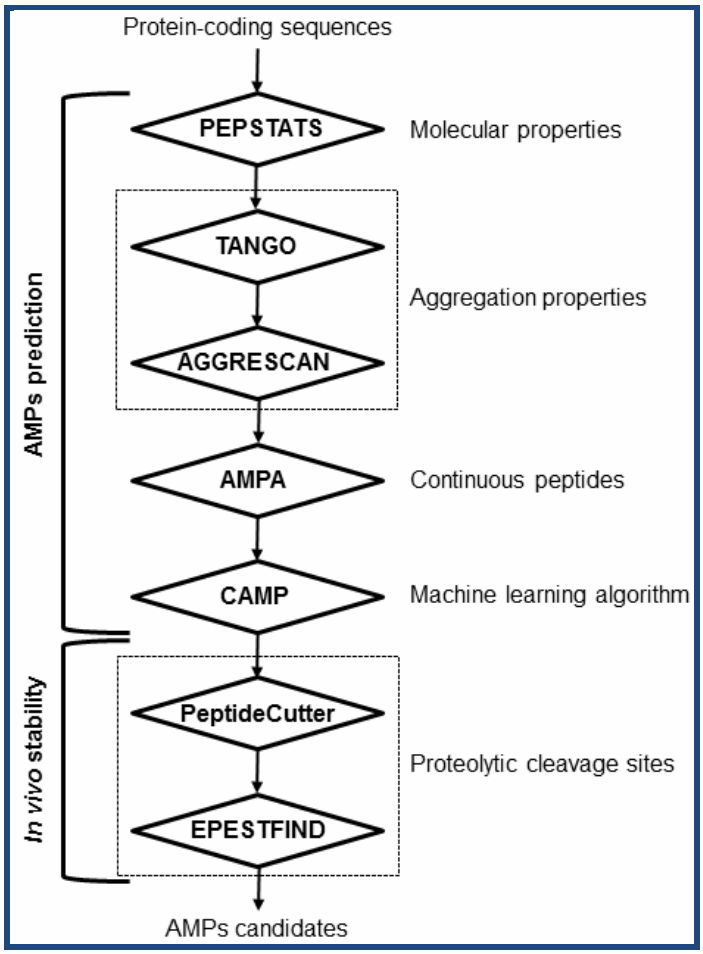

Figure 1: Schematic representation of AMP predictions based on physicochemical properties and predictions of their *in vivo* stability to discover AMP candidates from the entire genome of *C. sinensis*.

Predictions based on physicochemical properties

Based on the average length of known AMPs, we selected 317 sequences that were less than or equal to 80 amino acids in length. This was a stringent feature that dramatically reduced the number of possible CsAMPs. Subsequently, 230 sequences were selected based on their positive charge, another feature of most known AMPs. Further analysis of pI yielded 198 sequences having a pI between 8 and 12 **[9].**

In vitro and *in vivo* aggregation is important for the formation of aggregates on cell surfaces, an essential step required for AMP activity, similar to the carpet-like mechanism **[10].** Peptide aggregation on the bacterial outer membrane could help clear bacteria. Recent studies have shown that AMP aggregation

decreases in solution but increases at the bacterial membrane, conditions that are more hydrophobic. Thus, we examined aggregation in solution and bacteria using TANGO **[11]** and AGGRESCAN **[12]**, respectively. Through these two predictions, only six sequences remained.

Within the sequence of each CsAMP candidate, it is important to identify continuous fragments that exhibit antimicrobial activity. Using the AMPA server, we found four CsAMPs with at least one antimicrobial fragment **Table 1 (see supplementary material).** The AMPA server applies an algorithm to assign an antimicrobial index to each amino acid and builds a model for the entire protein sequence through a sliding-window approach **[18].** The antimicrobial index is calculated from the half-maximal inhibitory concentration values and is experimentally affected by the individual substitutions in bactenecin 2A **[19].** The predicted continuous fragments were confirmed to exhibit antimicrobial activity for the entire sequence **[20]** and could be a valuable starting point for the development of novel AMPs **[5].**

Finally, we tested CsAMP candidates using the ANN algorithm of CAMP **[15]**, which applies the linear association of independent variables to predict the group membership for each of the dependent variables. All four CsAMP candidates were classified as AMPs using this analysis.

In vivo stability

AMPs often exhibit limited bioavailability because they can be aggregated or degraded by proteases **[21].** It appears that most known AMPs have proteolytic cleavage sites for some proteases, such as SUMO protease, furin protease, or 3C protease. Trypsin, proteinase K, and pepsin do or do not affect the antimicrobial activity under different experimental conditions. However, some endopeptidases induce a dramatic reduction in antimicrobial activity. The α-chymotrypsin attacks internal basic residues, as an essential feature of antimicrobial peptides **[22].** PEST motifs are known to elevate the degradation of polypeptides containing them **[23].** In some cases, conserved binding motifs, such as lanthionine rings, may enhance the stability of AMPs against proteases **[22].** None of the three CsAMP candidates contained protease cleavage sites for chymotrypsin or PEST motifs, except for one CsAMP candidate (Acc. No. GAA53434.1). Nonetheless, after partial removal of the problematic region, the 'SRHCRVCSNRHGLIRK' peptide was confirmed to be antimicrobial based on the previous stepwise analysis (data not shown).

Conclusion:

In this study, we have suggested a refined computational approach to discover new AMPs from the liver fluke *C. sinensis*. The procedure can be divided into two distinct but complementary parts: rational data-driven predictions and *in vivo* stability predictions. We used a prediction strategy based on physicochemical properties to reduce the huge amount of *C. sinensis* protein fragments. Following this theoretical screening few CsAMP candidates were selected, and their *invivo* stability was analyzed. Eventually, four high-activity, high-stability peptides were identified. These could serve as the starting material for further development of new AMPs via structural modification and optimization. We believe that the strategic computational approach proposed in this study can also be applied to other bioactive peptide design.

Acknowledgement:

This research was supported by a fund (4800-4847-311) by Research of Korea Centers for Disease Control and Prevention.

References:

[1] Amirkia VD & Qiubao P, *Bioinformation* 2011 **5:** 365 [PMID: 21383924]
[2] Saha SB & Verma V, *Bioinformation* 2013 **9:** 518 [PMID: 23861568]
[3] Lee S *et al. Pathog Global Health.* 2012 **106**: 218 [PMID: 23265422]
[4] Singh N & Rai V, *Bioinformation* 2012 **8**: 1021 [PMID: 23275700]
[5] Yoo WG *et al. Funct & Integr Genomics*. 2014 **14:** 275 [PMID: 24652097]
[6] Yoo WG *et al. PLoS Negl Trop Dis*. 2011 **5**: e1208 [PMID: 21738807]
[7] Plieskatt JL *et al. FASEB J.* 2013 **27**: 4572 [PMID: 23925654]
[8] Rice P *et al. Trends Genet*. 2000 **16**: 276 [PMID: 10827456]
[9] Torrent M *et al. PloS one*. 2011 **6:** e16968 [PMID: 21347392]
[10] Shai Y, *Biopolymers* 2002 **66**: 236 [PMID: 12491537]
[11] Fernandez-Escamilla AM *et al. Nat Biotechnol*. 2004 **22:** 1302 [PMID: 15361882]
[12] Conchillo-Sole O *et al. BMC bioinformatics*. 2007 **8**: 65 [PMID: 17324296]
[13] Torrent M *et al. Bioinformatics* 2012 **28**: 130 [PMID: 22053077]
[14] Yeaman MR & Yount NY, *Pharmacol Rev*. 2003 **55:** 27 [PMID: 12615953]
[15] Thomas S *et al. Nucleic acids Res*. 2010 **38**: D774 [PMID: 19923233]
[16] Fjell CD *et al. Nat Rev Drug Discov*. 2011 **11**: 37 [PMID: 22173434]
[17] Brogden KA, *Nat Rev Microbiol*. 2005 **3:** 238 [PMID: 15703760]
[18] Torrent M *et al. BMC bioinformatics*. 2009 **10**: 373 [PMID: 19906288]
[19] Hilpert K *et al. Nat Biotechnol*. 2005 **23**: 1008 [PMID: 16041366]
[20] Torrent M *et al. J Med Chem*. 2011 **54:** 5237 [PMID: 21696142]
[21] Parachin NS *et al. Peptides*. 2012 **38**: 446 [PMID: 23022589]
[22] Hancock RE & Sahl HG, *Nat Biotechnol*. 2006 **24:** 1551 [PMID: 17160061]
[23] Xing H *et al. Cell Stress Chaperones*. 2010 **15**: 301 [PMID: 19768582

Edited by P Kangueane

Citation: **Yoo *et al.*** Bioinformation 11(1): 017-020 (2015)

BIOINFORMATION

open access

Supplementary material:

Table 1: A summary of the four *C. sinensis* AMP candidates by *in silico* analysis

Genbank ID	Molecular properties				Prediction tools						
	Length	MW (kDa)	Charge	pI	TANGO			AGGRESCAN	AMPA	PeptideCutter	EPESTFIND
					AGG	HELIX	BETA				
ABZ82029.1	20	2.08	1.0	8.05	44.10	3.82	44.97	48.6	2 to 16 (15-mer)	No sites	No sites
GAA53434.1	56	6.62	10.5	10.45	3.75	4.11	92.52	-13.7	18 to 53 (36-mer)	Four sites	No sites
GAA49532.1	57	6.68	7.0	10.27	11.32	20.59	95.26	-10.6	10 to 21 (12-mer)	No sites	No sites
GAA47811.1	60	7.22	8.0	10.94	138.94	23.33	80.75	-21.8	20 to 31 (12-mer)	No sites	No sites