Sequence analysis and membrane partitioning energies of α-helical antimicrobial peptides

Xing Han* and Wenjun Kang†

DuPont Haskell Laboratory for Health and Environmental Sciences, P.O. Box 50, Newark, DE 19714, USA

Received on August 5, 2003; revised on November 2, 2003; accepted on November 10, 2003
Advance Access publication February 5, 2004

ABSTRACT
Sequences of 221 α-helical antimicrobial peptides (αAMPs) were compared and 63–166 of them were selected and analyzed using Perl programs. The results showed that aliphatic amino acids Gly, Leu, Ala, Ile and two positively charged amino acids Lys and Arg were composed of more than 63% of the first 20 residues of αAMPs. The weighed mean membrane partitioning energies at positions from 1 to 25 of αAMPs were calculated. Profile of the partitioning energies suggests oblique membrane insertion and an amphipathic α-helical structure of the N-terminus of αAMP (residues from 1 to 13), a bend structure at positions 13 and 14, and a less structured C-terminus that parallels the surface of the membrane. These structural features are in good agreement with the experimentally determined membrane structure of hemagglutinin fusion peptide from influenza virus. We hypothesize that this (N-terminal oblique α-helix)—central bend—(C-terminus) could be a common structural motif of membrane-disruptive peptides.

Contact: xing.han@usa.dupont.com

INTRODUCTION
Antimicrobial peptides (AMPs) are key components of the innate immune system and host defense mechanisms of all animals ranging from insects to mammals (Boman, 1991). All peptides of this class show activities against microorganisms and therefore provide good templates for design of novel anti-infectious agents for humans. AMPs kill microbes by permeabilizing their cytoplasmic membranes. This function is achieved by direct interaction of AMPs with the lipid bilayers of microbial membranes (Wade et al., 1990). α-Helical AMPs (αAMPs; typically 20–40 amino acids long) are those whose active form adopts a linear, amphipathic α-helical structure in a membrane or membrane-mimic environment. A couple of hundreds of αAMPs have been isolated to date from a variety of sources and they show a marked degree of variability in sequences and activities. Earlier efforts on sequence analysis revealed a number of interesting features of αAMPs in their sequences, such as the positional conservation of residue types, the net positive charge states, and 40–60% of hydrophobic residues (for detailed review of previous works, see Tossi et al., 2000). All these features to various degrees affect the activities of αAMPs. Based on the results of sequence analysis, a sequence template has been developed and potent peptide antibiotics have been designed (Tossi et al., 1997). Encouraged by these successful studies, we analyzed αAMP sequences by utilizing new algorithms in an effort to disclose the underlying principles of the structure–activity relationships of αAMPs.

ALGORITHM
Dataset L contained 221 randomly ordered αAMPs (s₁, s₂, s₃, ..., s₂₂₁). Two Perl programs were written for sequence selection and sequence analysis, respectively. Sequences that have the percentage sequence identity smaller than a threshold (values incremented from 0.3 to 0.9, Table 1) were saved in a subset S (gap was not allowed in sequence matching). The amino acid compositions and the weighed mean membrane partitioning energies of the sequences in S were analyzed.

The weighted mean membrane partitioning energy at position n (ranged from 1 to 25) of αAMPs was calculated according to the following formula:

\[
(\Delta G_{\text{mean}})n = \sum_{i=\text{AA}_1}^{\text{AA}_n} f_i \cdot \Delta G_i, \quad (1)
\]

where \( f \) is the frequency of an amino acid (AA) at the nth position of αAMPs and \( \Delta G \) is the free energy of the amino acid transferred from water to membrane interface (Wimley and White, 1996). Twenty amino acids are randomly numbered from AA₁ to AA₂₀.

*To whom correspondence should be addressed.
†Present address: University of Minnesota School of Dentistry, 515 Delaware ST. SE, RM 6-320, Minneapolis, MN 55455, USA.

1Sequences and Perl programs are available for download at http://www.huhaha.com/AMP.html
RESULTS AND DISCUSSION

221 αAMP sequences were retrieved from Antimicrobial Sequences Database (AMSDb; http://www.bbcm.univ.trieste.it/~tossi/antimic.html). The length and the first amino acid at the N-terminus of the sequences were determined based on the definition in the Swiss-Prot protein sequence database. Because significant numbers of sequences in the AMSDb database are point mutations, an algorithm of sequence comparison was used for the selection of the sequence representatives of αAMPs. Based on this algorithm, only a peptide that does not exceed a certain percentage (threshold, 0.3–0.9) of sequence match with other peptides will be accepted. The number of peptides selected at different threshold values were shown in Table 1. Because no attempts were made to categorize αAMPs from different subclasses, such as cecropins, magainins, BLPs, caerins, cathelicidins, etc., we believe this sequence selection step provides an objective algorithmic approach to eliminate unwanted repetition of very similar sequences, and by varying the threshold values the extent of the effect of the sequence similarity on sequence analysis can be evaluated.

Figure 1 shows the result of statistical analysis of amino acid compositions of pre-selected αAMPs. A random distribution of all 20 amino acids in αAMP sequences would result in 5% frequency for each amino acid. Figure 1A shows the mean values and SDs of the frequencies of all 20 amino acids in the first 20 residues of αAMPs. Aliphatic amino acids Gly, Leu, Ala, Ile and two basic amino acids Lys and Arg are clearly preferred amino acids with average frequencies ranging from 17.3 to 8.2%. These six amino acids on an average made up more than 63% of the first 20 residues of αAMPs. On the other hand, His, Gln, Asp, Glu, Trp, Tyr, Asn, Pro, Met and Cys have minor contributions to αAMP sequences. The frequencies of these 10 amino acids are all below 2.3% and their total contribution to αAMP sequences is close to 16%.2 The SD for glycine is noticeably larger than the rest of the amino acids, which indicates glycine frequency is more dependent on the threshold values. The larger the threshold, i.e. the more the tolerance to sequence similarity, the higher the tendency to give higher glycine frequency. Figure 1B shows the most frequently occurred amino acids (combined average frequency >40%) at positions from 1 to 14 of αAMPs. Amino acids in this figure were grouped according to their preference of membrane partitioning (Wimley and White, 1996). Most notably, Gly (44%) and Leu/Ieu (43%) prevailed at positions 1 and 2, respectively. Acidic amino acid, Asp, despite its low combined frequency (2.0%) in the first 20 residues of αAMPs (Fig. 1A), has surprisingly high percentage (12%) at position 4. Other two positions that have high percentages of hydrophobic residues (Leu/Ieu) were found at position 6 (36%) and 13 (32%), respectively. It is worth to note that even though the positional conservation of residue types is also observed in Figure 1B, there are significant discrepancies between our data and the data reported by Tossi et al. (2000) in the absolute frequencies of certain amino acids at some positions. For example, Gly at position 1 and Lys at position 8 were reported as high as 70 and 50%, respectively in Tossi et al. (1997, 2000). Whereas in our study, these two amino acids are also predominant at these two positions, but with a lot lower percentages: 43 and 27%, respectively. Because peptides that were analyzed in both studies were from the same database, we conclude that the discrepancies are a direct result of the different sequence selection criteria used in these two studies. In the work of Tossi et al., αAMPs were pre-grouped into several subclasses based on their homologies (Tossi et al., 1997). Whereas in our study, the sequence selection was solely dependent on the extent of sequence similarity, which significantly reduced the possibility of over-weighing certain class of αAMPs, and therefore enabled broader representation of αAMPs.

Table 1. The number of sequences that were selected for sequence analysis

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Number of selected sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>63</td>
</tr>
<tr>
<td>0.4</td>
<td>80</td>
</tr>
<tr>
<td>0.5</td>
<td>108</td>
</tr>
<tr>
<td>0.6</td>
<td>114</td>
</tr>
<tr>
<td>0.7</td>
<td>127</td>
</tr>
<tr>
<td>0.8</td>
<td>145</td>
</tr>
<tr>
<td>0.9</td>
<td>166</td>
</tr>
</tbody>
</table>

2 The reported frequency values are statistical analysis of the compositions of αAMP residues and not to imply the importance of certain amino acid to the activities of αAMPs.

Fig. 1. Statistical analysis of αAMP sequences obtained from AMSDb database: (A) amino acid frequencies in the first 20 residues of αAMPs. (B) Distributions of most frequent amino acids (combined frequency >40%) at positions 1 to 14 of αAMPs. SDs are shown in the figure.

Sequence analysis of antimicrobial peptides
Binding of αAMPs to microbial membranes is always accompanied by a random coil ⇒ α-helix transition of the structure of αAMPs. This transition can be easily understood if one carefully examined the positional conservation of the residue types of αAMPs (Tossi et al., 2000). Therefore, the information about the secondary structure of αAMPs in lipid bilayers, which is critical for the activity, is well coded in αAMP sequences. In our present study, a new algorithm was developed for decoding of αAMP sequences.

The preference of an amino acid transferring from water to membrane interface is dictated by its free energy of membrane partitioning, which has been determined for all 20 amino acids by Wimley and White (1996). The weighed mean membrane partitioning energies at positions from 1 to 25 of αAMPs were calculated in our study and their profile is shown in Figure 2A. By convention, negative values of free energies favor membrane partitioning. The most interesting finding of Figure 2A is the periodic pattern of the partitioning energies between residue 1 and 14. This pattern was successfully fitted to an equation that describes a regular α-helix with an oblique angle (∼35° to the membrane plane). Outline of residues in Figure 2A (dashed lines) clearly shows a central bend at residue 13 and 14. The periodicity of residues after 14 is less regular, and the pattern can no longer be defined by a regular structure. Therefore, according to their free energies of membrane partitioning, αAMPs on an average seem to feature an obliquely membrane inserted, amphipathic, α-helical N-terminus, a central bend, and a less structured C-terminus that parallels the surface of the membrane. Indeed, it has been shown that the N-terminal helix and the central bend structure are adopted by many αAMPs in a membrane-mimetic environment, such as maculatin 1.1 (Chia et al., 2000), caerin 1.1 (Wong et al., 1997), magainin 2 (Gregory et al., 1997), melittin (Basso et al., 1988), alamethincin (Esposito et al., 1987), buforin II (Yi et al., 1996), PMAP-22 (Park et al., 2002) and gaegurin 4 (Park et al., 2000). An oblique membrane insertion (35° to the membrane normal) of the helical segment of maculatin 1.1 has also been observed (Chia et al., 2002). All these findings agreed well with the membrane structure that was proposed based on Figure 2A.

Very interestingly, the membrane structure of hemagglutinin fusion peptide from influenza virus, which was determined by electron paramagnetic resonance spectroscopy and nuclear magnetic resonance spectroscopy (Han et al., 2001), shares similar structural features, as shown in Figure 2B. Figure 2B clearly demonstrated that hemagglutinin fusion peptide in lipid bilayers adopted an obliquely (37° to the membrane plane) inserted N-terminal α-helix and a bend structure in the center of the peptide.

αAMPs function by disrupting the integrity of lipid bilayers that they bind to. Viral fusion peptides (Tamm and Han, 2000; Tamm et al., 2002), in addition to disturbing the structure of lipid bilayers, enable the merge of two adjacent lipid bilayers. It is rational for αAMPs and viral fusion peptides in some degree to share similar features in their membrane structures because it makes sense for nature to select one common structural motif for one particular task, which is the peptide-regulated membrane disruption.

REFERENCES


