OPEN ACCESS Freely available online

PLOS ONE

Cyclotide Structure–Activity Relationships: Qualitative and Quantitative Approaches Linking Cytotoxic and Anthelmintic Activity to the Clustering of Physicochemical Forces

Sungkyu Park, Adam A. Strömstedt, Ulf Göransson*

Division of Pharmacognosy, Department of Medicinal Chemistry, Uppsala University, Uppsala, Sweden

Abstract

Cyclotides are a family of plant-derived proteins that are characterized by a cyclic backbone and a knotted disulfide topology. Their cyclic cystine knot (CCK) motif makes them exceptionally resistant to thermal, chemical, and enzymatic degradation. Cyclotides exert much of their biological activity via interactions with cell membranes. In this work, we qualitatively and quantitatively analyze the cytotoxic and anthelmintic membrane activities of cyclotides. The qualitative and quantitative models describe the potency of cyclotides using four simple physicochemical terms relevant to membrane contact. Specifically, surface areas of the cyclotides representing lipophilic and hydrogen bond donating properties were quantified and their distribution across the molecular surface was determined. The resulting quantitative structure-activity relation (QSAR) models suggest that the activity of the cyclotides is proportional to their lipophilic and positively charged surface areas, provided that the distribution of these surfaces is asymmetric. In addition, we qualitatively analyzed the physicochemical differences between the various cyclotide subfamilies and their effects on the cyclotides' orientation on the membrane and membrane activity.

Citation: Park S, Strömstedt AA, Göransson U (2014) Cyclotide Structure–Activity Relationships: Qualitative and Quantitative Approaches Linking Cytotoxic and Anthelmintic Activity to the Clustering of Physicochemical Forces. PLoS ONE 9(3): e91430. doi:10.1371/journal.pone.0091430

Editor: Paul Taylor, University of Edinburgh, United Kingdom

Received November 6, 2013; **Accepted** February 11, 2014; **Published** March 28, 2014

Funding: U.G. is supported by the Swedish Research Council (#621-2007-5167) and the Swedish Foundation for Strategic Research (#F06-0058). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ulf.goransson@fkog.uu.se

Introduction

The cyclotides are a family of proteins characterized by their cyclic backbone and knotted disulfide topology (Figure 1) [1,2]. The cyclic cystine knot (CCK) motif makes the cyclotides exceptionally resistant to thermal, chemical, and enzymatic degradation [3]. However, it also means that their structure is ''inside out'' compared to that of most other globular peptides and proteins: their hydrophobic residues are forced onto their external surface because the inner core is occupied by the cystines. Currently, cyclotides have been found in limited set of plant families, namely the Rubiaceae (coffee) [4], the Violaceae (violet) [5], the Fabaceae (legume) [6] and the Solanaceae (potato) families [7]. In addition, peptides with similar structures and sequences have been found in the Poaceae (grass) [8] and Cucurbitaceae (gourd) [9] families. Sequence variation is immense: it has been estimated that the number of different cyclotides in Rubiaceae alone exceeds 50,000 [4]. Cyclotides form a combinational peptide library in those plant species that express them, and their function appears to be related to plant defense as reflected in their potent insecticidal [10] and antimicrobial activity [11,12]. Cyclotides also have pharmaceutically relevant properties including anti-cancer [13] and anti-HIV activity [14], and have proven to be good scaffolds for protein engineering [15]. These features make cyclotides potentially valuable peptides for pharmaceutical and agrochemical applications.

The known cyclotides have been divided into two main subfamilies based on the presence or absence of a conceptual 180° twist in the cyclic backbone caused by a conserved *cis*-Pro residue in loop 5, which lies between two cysteine residues [1]. Cyclotides that contain this twist are referred to as Möbius cyclotides; those without it are referred to as bracelet cyclotides. Some loops (loops 1 and 4) have high sequence similarity between the subfamilies while others (loops 2 and 3) are conserved only within individual subfamilies. Recent studies [16] have described hybrid cyclotides that exhibit sequence characteristics of both the Möbius and bracelet subfamilies. A third minor subfamily, known as the trypsin inhibitors, has been discovered in gourd plants. These peptides contain the CCK motif but do not otherwise exhibit any sequence homology with the other subfamilies [9].

Although the cyclotides exhibit diverse biological effects, there is a growing body of evidence suggesting that most of these activities are due to their ability to interact with and disrupt cell membranes [17,18,19,20]. This ability arises from a combination of hydrophobic and electrostatic interactions between the membrane and the cyclotide. In particular, exposed hydrophobic and electrostatic patches on the surface of the peptide contribute to its binding and dictate the orientation of its binding with respect to the CCK-motif [21]. The integrity of the hydrophobic patch is important for

the cyclotides' biological activity, as demonstrated by experiments using chemically modified analogues [18,22] and an alanine scan [19].

Although the extremely stable and rigid cyclotide structure appears ideal for structure-activity relationship (SAR) analyses, no attempt has yet been made to establish quantitative SARs (QSARs) for these peptides. Several authors have presented QSAR models for other membrane-active peptides, especially antimicrobial peptides [23]. Like cyclotides, these peptides are thought to exert their bioactivities via membrane disruption and/or lysis. Some of these studies [24,25] included variables that represent the conformational energy or the shape of the folded peptide in their model equations, probably because the studied proteins are quite flexible. However, this approach does not produce models that allow for easy geometrical interpretation or provide guidelines for the design of more active compounds. Another potential problem in QSAR modeling is the overfitting of data due to the inclusion of multiple variables that describe similar physicochemical properties. While this can yield models with very high regression coefficients, they can also be easily misinterpreted [23,25]. Recently, Frecer [26] developed a method that classifies the positions of amino acid residues within a peptide based on their orientations and physiochemical properties, but this method appears to be limited to the analysis of minor changes in peptide sequences.

In this work, we present a method for qualitatively and quantitatively characterizing the lipophilic and electrostatic properties of cyclotides using only four variables. We use this method to classify cyclotides, and to build a QSAR model linking those physicochemical properties to their cytotoxic and anthelmintic activities. In contrast to previous SAR/QSAR studies on peptides of similar size and activity, our approach accounts for both the total extent of the molecular surface area that exhibits a given physicochemical property and also the orientation and distribution of those surfaces.

Materials and Methods

Template selection

MUSCLE [27] was used to perform a multiple sequence alignment of each target sequence against 13 cyclotides whose structures have been solved by NMR: kalata B1 (pdb code: 1nb1) [28], kalataB2 (1pt4) [29], kalata B8 (2b38) [16], [W19K,P20N,V21K]-kalata B1 (2f2j) [30], [P20D,V21K]-kalata B1 (2f2i) [30], circulin A (1bh4) [31], circulin B (2eri) [32], cycloviolacin O1 (1nbj) [28], cycloviolacin O2 (2knm) [33], cycloviolacin O14 (2gj0) [34], varv F (2k7g) [35], vhr1 (1vb8) [36] and tricyclon A (1yp8) [37]. Optimized alignments were created by adjusting the initial alignments generated using MUSCLE with respect to conserved residues. A Neighbor Joining (NJ) tree [27] was constructed based on the multiple sequence alignment, using nonparametric bootstrapping with 1000 replicates. The cyclotide closest to the target sequence in the resulting tree was then used as a template for molecular modeling. All of these phylogenetic analyses, including the MUSCLE sequence alignment and the construction of the NJ tree, were performed using version 4.03 of the SeaView software package [27]. The template used for the linear cyclotide psyle C was violacin A (2fqa) [38]. The first conformation of the NMR ensemble reported in the PDB was used as template and, when applicable, for calculation of molecular descriptors.

Structure construction

Structures were constructed based on the pairwise sequence alignments of each target sequence with the selected templates using Modeller 9v8 [39]. Twenty structures were generated for each sequence, and the modeled structures were evaluated using the DOPE potential and the GA341 score as implemented in Modeller 9v8. Of the structures whose GA341 scores were closest to 1 (indicating a native-like fold), the three with the lowest DOPE potentials were selected for structure refinement. In cases where two templates were chosen for one target sequence, six structures were refined.

Refinement of models

The selected structures were assigned appropriate ionization states using the Protonate3D module of the MOE package and then refined by performing conformational searches with LowMode MD [40] using the AMBER94 force field and Generalized Born solvation (GB). The root mean square deviation (RMSD) gradient was set at 0.5 Å and the RMSD limit at 0.75 Å, with a maximum of 10000 iterations of energy minimization and a rejection limit of 100. In cases where the refinement of the overall structure failed for all of the conformations imported from Modeller 9v8, a few loops were selected and refined individually with LowMode MD. Predicted structures established through conformational searches covering the entire peptide structure were accepted if the RMSD value for the Cα atoms was less than 2.0 Å (relative to their positions in the initial conformation) and the Φ and ψ angles of the predicted structure were within the allowable regions of the Ramachandran plot. In cases where conformational searches covering the entire peptide structure did not produce acceptable results, constrained searches were performed in which only one or two loops were allowed to vary their conformations. Loops were selected for constrained optimization if they contained residues whose side chains had different physicochemical properties relative to their counterparts in the template loop, or if they contained multiple gap regions compared to the template loop. The backbones of the unselected loops and the cystine residues were constrained by fixing their atomic positions, and the side-chains of these fixed loops were subjected to low frequency mode

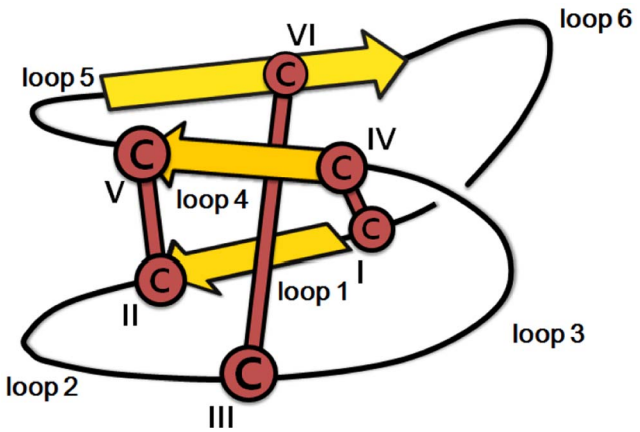

Figure 1. Schematic illustration of cyclotide structure. A 3D illustration of the cyclotide structure showing the CCK motif (with cysteines labeled I–IV), and the inter-cysteine loops (labeled 1–6). The CCK motif is characterized by a cyclic peptide backbone and the knotted disulfide bonds (I–IV, II–V, III–VI), which are indicated by thick brown bars. The structure usually contains antiparallel beta strands (indicated by large yellow arrows) and several tight turns.
doi:10.1371/journal.pone.0091430.g001

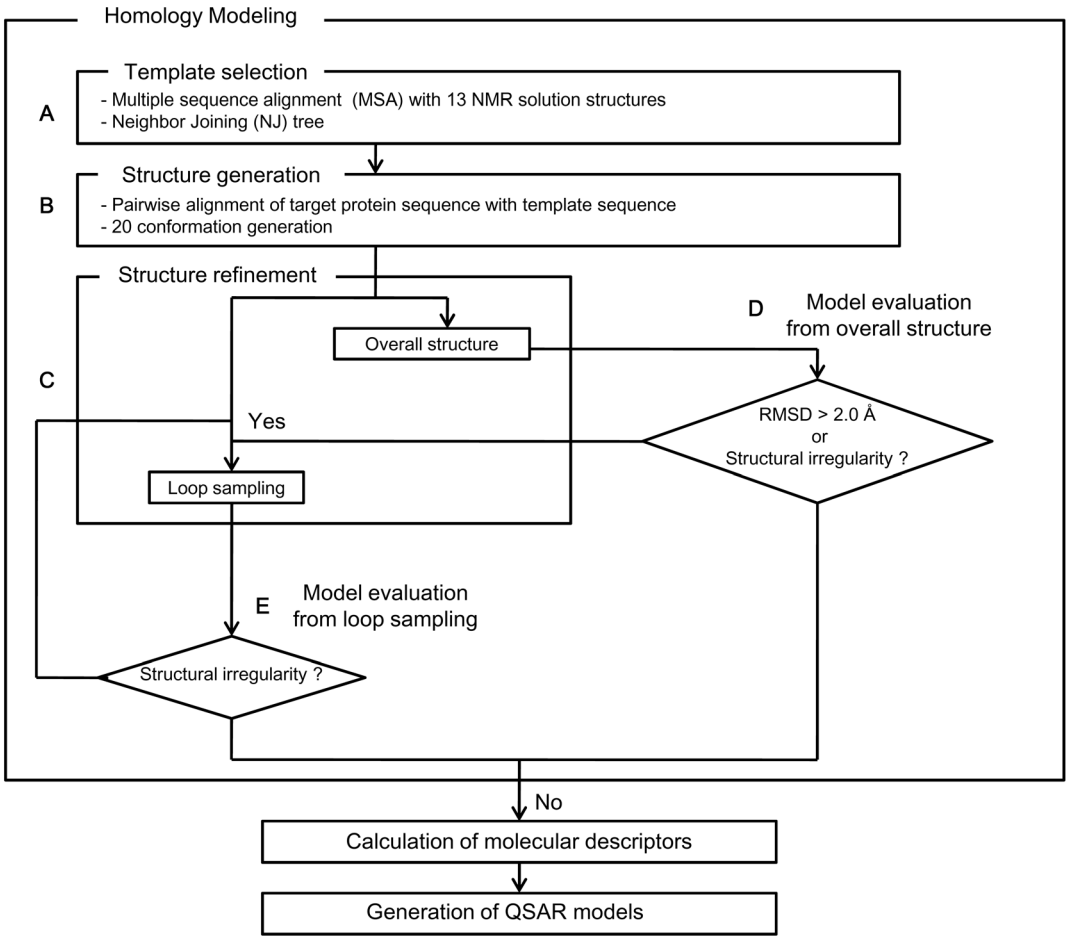

Figure 2. Flow chart depicting the methodology used for QSAR modeling. Predicted structures for the cyclotides were obtained through homology modeling using the sequences of the 13 cyclotides whose solution-phase structures had previously been determined by NMR. A) To identify appropriate templates, each target sequence was aligned with those of the 13 peptides with known structures. A neighbor joining (NJ) tree was then constructed based on the multiple sequence alignment. The resulting cladogram was used to identify the closest relatives of the target sequence with a known structure, which were then used as templates for modeling the unknown structure of the target peptide. B) For each template sequence, 20 PDB structures were generated based on the pairwise sequence alignment of the target sequence with the selected templates. These 20 conformations were evaluated using the DOPE potential and GA341 score, and the best three conformations were selected. If two templates were chosen for a target sequence, a total of 6 PDB structures were selected. C), D) and E) The structures generated during step B were subjected to conformation searches. If structural refinement of the protein as a whole yielded a result with an unacceptable geometry or one for which the RMSD of the Cα-atoms was greater than 2.0 Å relative to the starting conformation, two loops were selected for loop sampling. Loops were selected based on their sequences, focusing on those containing residues with different physicochemical properties relative to those in the corresponding positions of the template sequence or those that aligned with gaps in the template sequence. After structure refinement, one conformation of each peptide sequence was selected for use in calculating the molecular descriptors. The molecular descriptors of the cyclotides, together with their activity data were used as variables in QSAR modeling.
doi:10.1371/journal.pone.0091430.g002

vibration. Loop conformations were then sampled and ranked according to their potential energy.

Calculation of molecular descriptors

Descriptors were calculated using scientific vector language (SVL) as implemented in the Molecular Operating Environment (MOE) [41]. These descriptors include lipophilic moment (L_M), exclusive lipophilicity (L_S^+), total lipophilicity (L_S), Hydrogen Bond Donor (HBD) surface area (E_S) and HBD amphipathic moment (E_M).

Generation of a lipophilicity scale and the calculation of solvent-accessible areas

The Molinspiration Property Calculation Service [42] was used to predict the lipophilicity of amino acid side chains. The amino acids' logP values were used as the starting point for these calculations. The logP values were parameterized, scaled, and normalized with respect to glycine. The maximum solvent accessible surface area (SASA) of the side chain of each amino acid (X) was then calculated based on the structure of the tripeptide Gly-X-Gly in an extended conformation ($\Phi = \psi = 180^\circ$) [43]. For natural amino acids (X), the tripeptide was constructed using data from rotamer libraries. For tripeptides that contained a non-natural amino acid, the appropriate structural modification was introduced into the relevant precursor amino acid using the builder module implemented in MOE, and the conformation of the modified side chain was relaxed using the MFF94x force field with GB solvation.

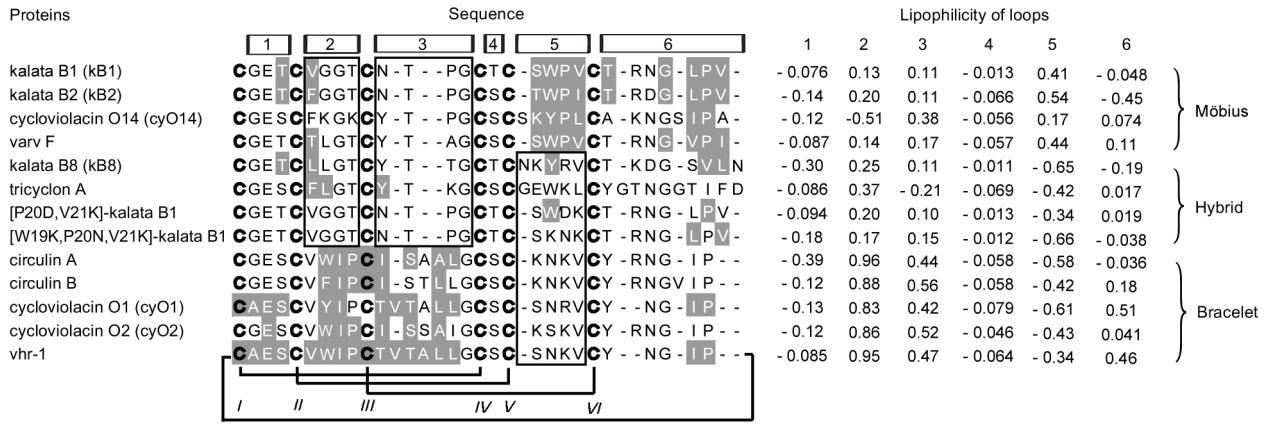

Figure 3. The sequences of the membrane-buried regions of selected cyclotides and their lipophilicity profiles. The multiple sequence alignment shows all of the cyclotides for which 3D structures (based on solution-phase NMR studies) are available. These sequences were used as template candidates for homology modeling of the cyclotides of unknown structure. The cysteine of loop 1 was arbitrarily selected as the N-terminus for the multiple sequence alignments. The residues required for cyclization, i.e. the conserved C-terminal residue (Asn or Asp) and N-terminal residue (Gly), are positioned within loop 6. The membrane-buried residues are highlighted with a gray background. OPM was used to predict the orientation of the selected cyclotides on the membrane. Among the members of the Möbius subfamily, many of the loop 5 and 6 residues are membrane-buried. Among the bracelet cyclotides, large proportions of loops 2 and 3 are buried. Only a couple of residues from the cyclotides of the hybrid subfamily penetrate the membrane interface. The lipophilicity of the loops was calculated as the total lipophilicity (L_S) of all of the residues within the loop. It should be noted that the membrane-buried regions identified by OPM are consistent with those exhibiting high lipophilicity.
doi:10.1371/journal.pone.0091430.g003

Statistical analysis

Materials studio 6.0 [44] was used to generate all QSAR models. Minitab 16 [45] was used to transform the molecular descriptor data.

Hierarchical clustering analysis

Ward's reciprocal nearest neighbor method was used to perform a hierarchical cluster analysis of 75 cyclotides. Distances were measured in Euclidean terms, using the four molecular descriptors E_M, E_S, L_M and L_S^+ as variables.

Results and Discussion

The lipophilic and electrostatic properties of cyclotides determine how they adsorb to membranes and the mechanisms by which they disrupt membrane integrity [46]. Here we describe a QSAR model that uses four molecular descriptors representing the lipophilic and electrostatic properties of cyclotides in terms of their surface area - a scalar quantity - and their orientation relative to the protein surface (the moment). We show that these descriptors are useful for characterizing and classifying cyclotides, and for understanding their structure-activity relationships. To this end, we analyze the physicochemical properties of the 68 known cyclotides with cytotoxic activity [13,18,22,47,48,49,50] against the lymphoma cell line U937-GTB or anthelmintic activity [20,51] against the caterpillar larvae of *H. contortus*.

Cyclotide homology modeling

We used cyclotide structures reported in the Protein Data Bank as solved by NMR, when possible. However, the majority of structures had to be constructed by homology modeling. Molecular properties of cyclotides were calculated based on the lowest energy conformation that was identified for each one. These low-energy conformers were identified through a homology modeling process involving template selection followed by structure generation and refinement (Figure 2).

Traditionally, template selection is done using either protein threading [52,53] or database searches using BLAST [54] or FASTA [55]. Unfortunately, none of these methods are well suited to cyclotides because of their high sequence similarity. To find the most appropriate template for each cyclotide from the 13 cyclotide NMR structures available in the Protein Data Bank (PDB), we used a neighbor joining (NJ) tree to describe the homology of the cyclotides based on sequence similarity. As shown in Figure 3, the first cysteine residue on the N-terminal side of loop 1 was arbitrarily defined as the N-terminus of each peptide. Multiple sequence alignments were performed using MUSCLE and then corrected manually. Manual corrections were performed to ensure that, among other things, all of the cysteine residues were kept in fixed positions, as well as the N-and C-terminal residues of loop 6, which must be conserved to enable circularization. Based on their bootstrap values, one template (or two in some cases) was selected for each sequence of unknown structure (Figure 4).

Modeller was then used to generate candidate structures, which were refined using LowMode MD [29] to identify the lowest energy conformation of each cyclotide. To avoid the pitfall of local energy traps, which can stop the refinement process once the first stable conformation is encountered, any refined structure for which the RMSD value of the Cα atoms was greater than 2.0 Å relative to the template structure was regarded as a decoy. The refinement process was conducted until a global minimum structure was located that had an RMSD value of less than 2.0 Å and which satisfied the requirements of protein geometry [56]. It is adequate to use one conformation only to represent the structure of cyclotides in solution as judged from experimentally determined NMR ensembles, at least for the calculation of molecular descriptors. For typical cyclotides, the average RMSD values of peptide backbone and side chains of the NMR ensemble do not exceed 2.0 Å, and the exposure of side chains at the surface is more or less constant (Figure S1).

Design and calculation of molecular descriptors

The cyclotides were characterized using descriptors of the lipophilicity and electrostatic properties that are important in

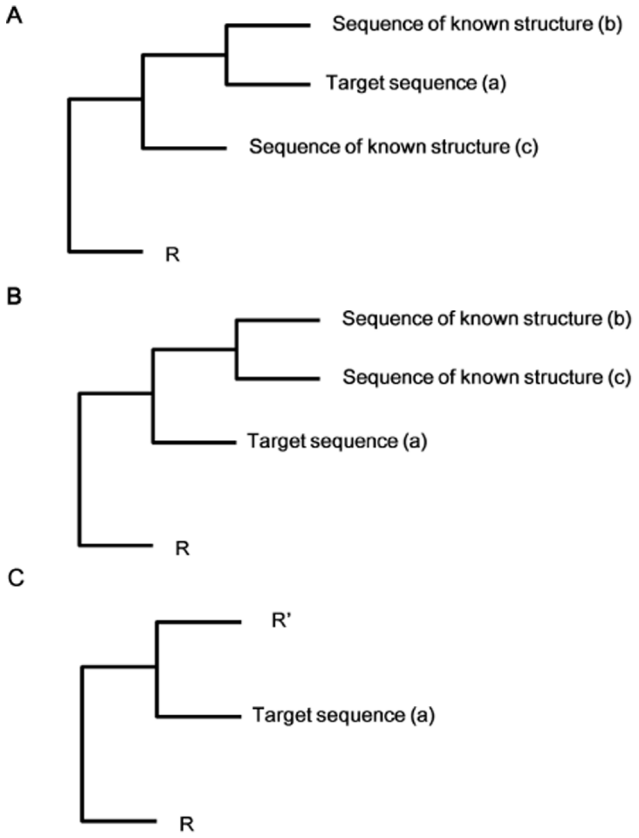

Figure 4. Template selection strategy based on the NJ tree. This figure illustrates the strategy of a template selection using a cladogram in which the target sequence (a) relates differently to various sequences of known structure. The symbols R and R′ represent the other branches of the cladograms. These branches contain sequences of known structure. Our aim was to identify sequences of known structure that could reasonably be used as templates for the target sequence based on the bootstrapped consensus tree. The procedure used to select the template sequences most closely related to the target sequence from the bootstrap NJ tree depended on the structure of the tree and the position of the target sequence relative to the nearest neighbors of known structure. In cases such as that shown in subfigure A, a single peptide of known structure (b) was selected as the template for target sequence (a) if the bootstrap value for the two taxa, a and b, was greater than 50%. Two taxa, b and c were selected if the bootstrap value for taxa a and b was below 50%. In cases such as that shown in subfigure B, two taxa, b and c were selected as templates, regardless of the bootstrap value of the common ancestor of the three taxa, a, b and c. In case C, we randomly chose two templates from the sister taxa (R′) of the target sequence (a) if the sister taxa (R′) included two or more sequences of known structure.
doi:10.1371/journal.pone.0091430.g004

peptide-membrane interactions. We associate each such physico-chemical property with a scalar value and a moment. The scalar value quantifies the molecular property without providing any information concerning its distribution on the molecular surface. For example, the surface area scalar quantifies the extent of the molecular surface that exhibits a certain property. Being scalar quantities, surface areas can be summed without regard for their orientation or position on the molecular surface. However, their distribution can also be described using a vector whose direction reflects the relative orientation of the property on the molecular surface. The moment, i.e. the length of the vector, is proportional to the degree of asymmetry in the distribution of the physico-chemical property of interest across the molecular surface.

Table 1. Normalized lipophilicity and solvent accessible surface area (SASA) values for natural and modified amino acids, assuming maximal side-chain exposure.

A.A.	LogP[a]	SASA[b]
Ala	0.015	65.85
Arg^+	−0.62	200.62
Asn	−0.18	128.88
Asp^-	−0.51	112.53
Cys[c]	0.068	99.87
Gln	−0.14	148.60
Glu^-	−0.47	146.78
Gly[d]	0.00	0.00
His[e]	−0.060	157.99
Ile	0.38	157.72
Leu	0.19	166.64
Lys^+	−0.366	195.16
Met	0.096	166.80
Phe	0.26	188.38
Pro[f]	0.25	105.72
Ser	−0.083	82.22
Thr	−0.015	114.60
Trp	0.28	229.66
Tyr	0.19	214.65
Val	0.20	125.98
Ack[g]	0.012	263.09
Cdr^{+h}	−0.49	336.44
Kyw[i]	0.10	236.08
Mee[j]	−0.012	182.28

[a]Scaled parameter = (raw parameters+3.800)/6.457, where the value of 3.800 is the raw parameter value for logP(Arg^+) and the value of 6.457 is the sum of the absolute values of two raw parameters, logP(Arg^+) and logP(Ile).
The normalized parameter values are equal to the scaled parameter values minus the scaled parameter value for logP(Gly).
[b]The SASA of amino acid (X) is the exposed solvent-accessible surface area of its side chain in the tripeptide Gly-X-Gly, assuming that $\psi = \varphi = 180^\circ$.
C_α was not considered to be part of the side chain of X.
[c]Oxidized cysteine (Cys) has 50% of the lipophilicity of cystine (Css), i.e. the oxidized dimer of cysteine.
logP(Cys) = [logP(Css)]/2, where logP(Css) = logP(CH_2-S-S-CH_2).
[d]log(Gly) = log(H).
[e]At pH = 7.4, histidine exists in protonated (His^+) and unprotonated forms (His^0) with mole fractions of 11.2% and 88.8%, respectively.
logP(His) = logP(His^+)·mole fraction (His^+)+logP(His^0)·mole fraction (His^0).
[f]log(Pro) = log(CH_2-CH_2-CH_2-N).
[g]Ack: acetylated lysine.
[h]Cdr^+: modified arginine with 1,2-cyclohexanedione.
[i]Kyw: kynurenine, a metabolite of tryptophan.
[j]Mee: methyl-δ-glutamate.
doi:10.1371/journal.pone.0091430.t001

A lipophilicity scale for natural and non-standard amino acids. Because some of the cyclotides examined in this work contain chemically modified amino acids, it was necessary to create a lipophilicity scale that can describe such residues. The lipophilicity of the side chains was therefore predicted using their logP values (Table 1). To validate the lipophilicity values for the chemically modified amino acids, the calculated lipophilicity values for the standard amino acids were compared to those used in the hydrophobicity scales of Kyte-Doolittle [57] and Black-Mould [58]. With the exception of amino acids having pi-conjugated aromatic side chains, our predicted lipophilicity values

1. For a protein consisting of n resides, the reference center, $\overrightarrow{r_o r_c}$, is defined as the center of the alpha carbons (C_α) of all residues: $\overrightarrow{r_o r_c} = \frac{1}{n} \cdot \sum_{i=1}^{n} \overrightarrow{r_o r_i}$, where the vector $\overrightarrow{r_o r_i}$ is the C_α coordinate of the i^{th} residue.

2. The unit vector $(\vec{u_i})$ is calculated by normalizing the vector, $\overrightarrow{r_c r_i}$. Normalization is achieved by dividing the vector by its length: $\vec{u_i} = \overrightarrow{r_c r_i} / \left\| \overrightarrow{r_c r_i} \right\|$.

3. The solvent exposure ratio (p_i) of the i^{th} residue is defined as the ratio of the exposure of its side chain, i.e. its solvent accessible surface area (SASA) in the cyclotide structure, to that of a tripeptide (G-X-G). We assumed that the exposed SASA of the side chain is equal to that of the same amino acid (X) in G-X-G, when the side chain is maximally exposed on the cyclotide structure. Table 1 shows the SASA values for all amino acids: p_i = (SASA of i^{th} residue in cyclotide) / (SASA of X in tripeptide).

4. The lipophilic intensity of the i^{th} residue (w_i) is the product of the lipophilic scale value (l_i) for the corresponding amino acid and its exposed rate (p_i): $w_i = p_i \cdot l_i$.

5. The lipophilic vector of the i^{th} residue $(\vec{L_i})$ is defined as the unit vector $(\vec{u_i})$ weighted by the lipophilic intensity (w_i): $\vec{L_i} = w_i \cdot \vec{u_i}$.

6. The lipophilic vector $(\vec{L})$ of the protein is the normalized sum of the lipophilic vectors of all residues: $\vec{L} = \frac{1}{n} \cdot \sum_{i=1}^{n} \vec{L_i}$, where $\vec{L_i} = p_i \cdot l_i \cdot \vec{u_i}$.

7. The lipophilic moment (L_M) is the length of the lipophilic vector: $L_M = \frac{1}{n} \left\| \sum_{i=1}^{n} p_i \cdot l_i \cdot \vec{u_i} \right\| = \sqrt{{x_H}^2 + {y_H}^2 + {z_H}^2}$, where ($x_H$, y_H, z_H) is the coordinate of $\vec{L}$.

Figure 5. The calculation of lipophilic moment (L_M).
doi:10.1371/journal.pone.0091430.g005

1. The electrostatic potential is determined using the Possion-Boltzman Equation (PBE) with a 0.5 Å grid spacing. At a given point r = (x, y, z), the electrostatic potential, u(r), represents the work per unit charge that we must do to move a unit charge from infinity to a position r within the electric field. Each set of coordinates is assigned an electrostatic potential value in the form [x, y, z, u(r)].

2. The location of the peptide's hydrogen bond donor (HBD) center is defined as the center of the volume enclosed by the isosurface that interacts with an aqueous oxygen probe derived from the SPC water model [62] at -1.6 kcal/mol. This point was taken to represent the center of the molecular surface that displays a positive charge. At a given position r, the hydrogen bond potential, HBP(r), is defined as the interaction energy with the aqueous oxygen probe: $HBP(r) = q_o \cdot u(r) + v_o$. Here, q_o is -0.82 and represents the partial charge of the aqueous oxygen probe, while v_o is the van der Waals potential of the oxygen atom. The HBP(r) term also incorporates a pairwise summation of the Coulombic $(q_o \cdot u(r))$ and van der Waals interactions (v_o) in the form of a Lennard-Jones potential: $v_o = \sum_i [A_i \,|\, r - r_i \,|^{-12} - B_i \,|\, r - r_i \,|^{-6}]$, where A_i and B_i are OPLS van der Waals parameters and r_i is the coordinate of the i^{th} atom of the protein. After the assignment of hydrogen bonding potentials on the grids [x, y, z, HBP(r)], the coordinates are selected to define the isocontour surface on which HBP(r) < -1.6 kcal/mol in Cartesian coordinates (x, y, z).

3. The hydrophobic center is defined as a center of the grid surface reactive to a DRY probe at -0.2 kcal/mol. The DRY probe calculates the hydrophobic energy at each grid point: $E_{Dry} = \sum_i E_{vdw} + S - \sum_i E_{hb}$, where S is the entropy, which is taken to be a constant (-0.848), E_{vdw} is the van der Waals interaction energy, and E_{hb} is the energetic cost of disrupting the hydrogen bonding network in the protein hydration shell.

4. The HBD amphipathic moment measures the distance from the center of the hydrophobic surface to the center of the hydrogen bond donor surface.

Figure 6. The calculation of the HBD amphipathic moment (E_M).
doi:10.1371/journal.pone.0091430.g006

for natural amino acid side chains were consistent with those used in these existing scales. It should be noted that conventional measures of lipophilicity primarily reflect the strength of hydrophobic interactions. However, they are also affected by polar interactions [59] because the model lipid used to determine partition coefficients (i.e. logP values) is octanol, which has two

Table 2. Labels used to refer to each cyclotide considered in this work and their activities against *H. contortus* and cells from the U-937GTB line.

Label	Proteins	*H. contortus* IC_{50} (μM)	U-937GTB IC_{50} (μM)	Ref
	Möbius			
M1	kB1	2.48	6.9	[18,20]
M2	[G1K]-kB1[a]	0.5	-	[20]
M3	[L2K]-kB1[a]	6.2	-	[20]
M4	[P3K]-kB1[a]	>11.5	-	[20]
M5	[V4K]-kB1[a]	11.1	-	[20]
M6	[G6K]-kB1[a]	>11.5	-	[20]
M7	[E7K]-kB1[a]	>11.5	-	[20]
M8	[T8K]-kB1[a]	>11.5	-	[20]
M9	[V10K]-kB1[a]	>11.5	-	[20]
M10	[G11K]-kB1[a]	5.3	-	[20]
M11	[G12K]-kB1[a]	>11.5	-	[20]
M12	[T13K]-kB1[a]	2.8	-	[20]
M13	[N15K]-kB1[a]	>11.5	-	[20]
M14	[T16K]-kB1[a]	>11.5	-	[20]
M15	[P17K]-kB1[a]	3.6	-	[20]
M16	[G18K]-kB1[a]	1.1	-	[20]
M17	[T20K]-kB1[a]	0.9	-	[20]
M18	[T20K, G1K]-kB1[a]	0.4	-	[20]
M19	[T20K, S22K]-kB1[a]	0.7	-	[20]
M20	[T20K, N29K]-kB1[a]	0.4	-	[20]
M21	[T20K, N29K, G1K]-kB1[a]	0.2	-	[20]
M22	[S22K]-kB1[a]	2.3	-	[20]
M23	[W23K]-kB1[a]	>11.5	-	[20]
M24	[V25K]-kB1[a]	>11.5	-	[20]
M25	[T27K]-kB1[a]	1.6	-	[20]
M26	[R28K]-kB1[a]	3.9	-	[20]
M27	[N29K]-kB1[a]	0.4	-	[20]
M28	kB2	1.59	2.6	[18,51]
M29	kB6	0.87	-	[51]
M30	kB7	6.29	29	[18,51]
M31	kB13	-	3.8	[18]
M32	cyH3	0.85	-	[51]
M33	cyO14	0.41	-	[51]
M34	cyO15	0.38	-	[51]
M35	cyO16	0.27	-	[51]
M36	cyO24	1.74	-	[51]
M37	vaby A	-	7.6	[47]
M38	vaby D	-	2.8	[47]
M39	varv A	1.13	8.2	[18,51]
M40	[Cdr]-varv A	-	9.1	[18]
M41	[Cdr, Mee]-varv A	-	46	[18]
M42	[Mee]-varv A	-	34	[18]
M43	varv E	0.9	4	[49,51]
M44	varv F	-	7.1	[13]
M45	vibi D	-	>30	[49]
	Bracelet			
B1	circulin A	-	-	-

Table 2. Cont.

Label	Proteins	*H. contortus* IC_{50} (μM)	U-937GTB IC_{50} (μM)	Ref
B2	circulin B	-	-	-
B3	cyO1	2.82	-	[51]
B4	cyO2	0.12	1.03	[18,51]
B5	[Cdr]-cyO2	-	0.95	[22]
B6	[Cdr, Ack]-cyO2	-	5.1	[22]
B7	[Mee]-cyO2	0.76	36	[22,51]
B8	[Ack]-cyO2	2.3	2.3	[22,51]
B9	[Kyw]-cyO2	-	5.2	[18]
B10	cyO3	0.21	-	[51]
B11	cyO8	0.24	-	[51]
B12	cyO13	0.21	-	[51]
B13	cyO19	-	0.52	[18]
B14	cyY4	2.01	-	[51]
B15	cyY5	2.28	-	[51]
B16	psyle E	-	0.76	[48]
B17	vhl-1	2.06	-	[51]
B18	vibi E	-	3.2	[49]
B19	vibi G	-	0.96	[49]
B20	vibi H	-	1.6	[49]
B21	vitri A	-	0.6	[50]
B22	vodo O	-	3.2	[18]
	Hybrid			
H1	[W23K, P24N, V25K]-kB1[a]	-	-	-
H2	[P24K]-kB1[a]	-	-	[20]
H3	[P24D, V25K]-kB1[a]	-	-	-
H4	kB8	-	18	[18]
H5	psyle A	-	2	[48]
H6	tricyclon A	-	-	-
	Linear			
L1	psyle C	-	3.5	[48]
L2	violacin A	-	-	-

[a]The first residue in the sequence of kalata B1 is defined relative to the C-terminal amino acid involved in the biosynthetic ligation in loop 6. (See Figure 3).
doi:10.1371/journal.pone.0091430.t002

distinct functional groups: a hydrophobic (non-polar) alkyl chain and a hydrophilic (polar) hydroxyl group. We suggest that the inconsistency between our scale and those presented previously with respect to the lipophilicity values for aromatic amino acids (and tryptophan in particular) is because these molecules interact preferentially with the interfacial region of the lipid, in a way that is not primarily driven by classical hydrophobicity [60,61].

Lipophilicity parameters (Ls and L_S^+) and the lipophilic moment (L_M). The value of the lipophilicity scalar for a given peptide depends on the exposed surface area of its amino acid side chains and their lipophilicity scale values. Two lipophilicity-related scalar quantities were calculated: the total lipophilicity (L_S) and the exclusive lipophilicity (L_S^+). The total lipophilicity (L_S) is based on the lipophilicity values of every residue in the peptide, regardless of their propensity for participating in lipophilic interactions. Conversely, the exclusive lipophilicity (L_S^+) is based only on the lipophilicity values of residues that interact favorably with lipids. The total lipophilicity (L_S) is thus proportional to the difference between the peptide's lipophilic and lipophobic surface areas, whereas the exclusive lipophilicity (L_S^+) is proportional to the lipophilic surface area alone.

The lipophilic moment (L_M) is calculated from the sum of the lipophilic vectors of each residue in the protein, as shown in Figure 5. In brief, the lipophilic vector of a given residue is calculated based on its value on the lipophilicity scale and its position relative to the protein center. If the majority of the peptide's lipophilic residues are positioned on one side of molecular surface, and its lipophobic residues are positioned on the opposite side, its overall distribution of lipophilicity will be asymmetric. The greater the asymmetry in the distribution of lipophilicity across the peptide's surface, the greater its lipophilic moment. Introducing lipophobic residues into a lipophilic cluster will decrease the lipophilic moment by increasing the symmetry of the lipophilicity distribution. The procedure used to calculate L_M is described in the Figure S2.

The affinity of the protein for the membrane surface is proportional to the contact area, provided that the contact surface does not contain lipophobic moieties. In conjunction with the

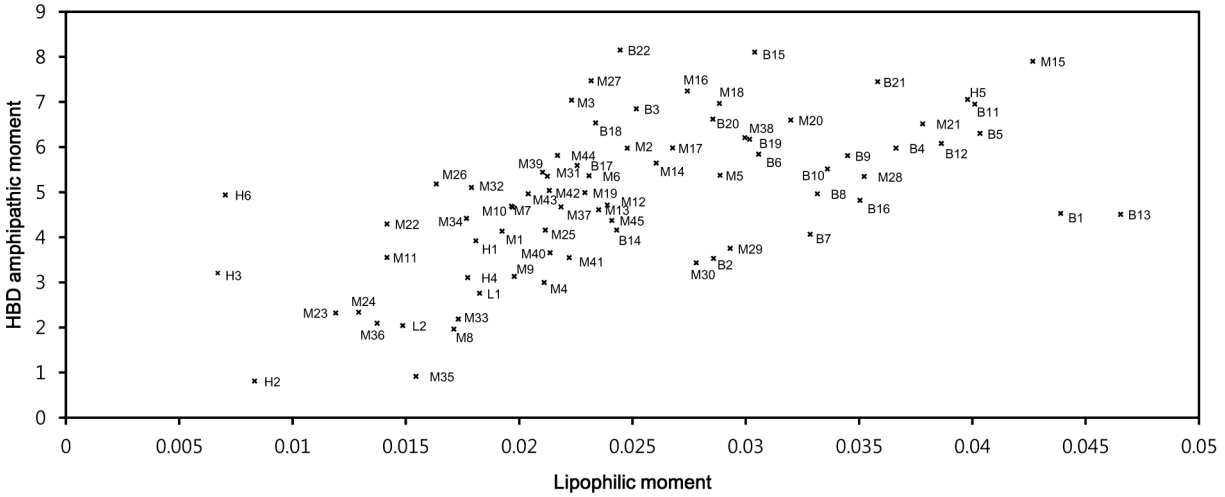

Figure 7. Moment plot (L_M, E_M) for the studied cyclotides. The cyclotides, labeled according to their subfamily (M for Möbius, B for bracelet, H for hybrid) and with their unique numbers, were plotted on a map based on their lipophilic moments (L_M) and HBD amphipathic moments (E_M). They cluster into two groups based on their moments: one group has low moments and exhibits low activity while members of the other group have high moments and exhibit moderate or high activity. The synthetic hybrid cyclotides, [W23K, P24N, V25K]-kB1 (H1), [P24K]-kB1 (H2) and [P24D, V25K]-kB1 (H3) cluster in low moment group. The activity of the cyclotides H1 and H3 against the U937GTB cell line and *H. contortus* is not known, but they are reported to be inactive against human type A erythrocytes [30].
doi:10.1371/journal.pone.0091430.g007

lipophilic moment (L_M), the exclusive lipophilicity (L_S^+) can be used to characterize the lipophilic strength of a membrane-active protein more accurately than is possible when using the total lipophilicity (L_S) alone. This is because the lipophilic moment (L_M) can be used to determine whether there is any significant lipophobic interruption of the contact surface, and the exclusive lipophilicity(L_S^+) provides information on the proportion of the protein's total lipophilic surface area that lies within the contact region. Most biologically active cyclotides have large lipophilic moments, and their potency correlates positively with their exclusive lipophilicity (L_S^+). Lipophobic residues on the opposite side of the peptide surface to the contact region reduce its total lipophilicity (L_S), but do not adversely affect its lipophilic interactions with membranes.

The positively charged surface area (E_S) and the HBD amphipathic moment (E_M). The positively charged surface area of the peptide is defined as the hydrogen bond donor (HBD) surface (Figure 6). The HBD surface is determined from the electrostatic potential (*E*) of the van der Waals surface of the protein. It reflects the exposed surface area of the positively charged residues and their surrounding electrostatic environment (i.e. their proximity to other charged entities and their location on the interior or exterior of the protein).

The HBD amphipathic moment (E_M) measures the imbalance in the electrostatic distribution between positively charged and hydrophobic (i.e. neutrally charged) surfaces. The magnitude of the moment is proportional to the distance between the hydrophobic center and the center of the positively charged surface. The moment is maximized in cases where the positively charged surfaces are localized on one side of the protein's surface and the hydrophobic surfaces are localized on the opposite side.

Classification of cyclotides and their activities

The molecular descriptors discussed above were determined for every modeled structure, and compared qualitatively with the reported activities of the corresponding cyclotides. Cytotoxic [13,18,22,47,48,49,50] and anthelmintic [20,51] activity data were obtained from the literature. Table 2 summarizes the potency of cyclotides and shows their IC_{50} values.

Physicochemical distribution of cyclotides. The cyclotides clustered closely with similar sequences and levels of biological activity when their lipophilic (L_M) and HBD amphipathic (E_M) moments were plotted against one-another (Figure 7). Most bracelet cyclotides have larger moments than those from the Möbius and hybrid subfamilies. Moreover, Möbius cyclotides generally have higher moments than hybrid cyclotides; the exceptions are the Möbius cyclotides that have mutations attenuating lipophilicity in loop 5. When combined with activity data, the low moment group coincides with cyclotides that we classify as having low activity, and cyclotides having moderate or high moments have moderate to high activity. The relative activity was calculated by comparing IC_{50} values in the cytotoxicity and anthelmintic activity assays to the IC_{50} of kB1. Cyclotides with ratios (IC_{50} of cyclotide/IC_{50} of kB1) <0.2 were considered being highly active; ratios between 0.2 and 1 as moderately active; and ratios >1 (i.e. less active than kB1) as being low activity cyclotides.

The trend was reinforced when all four of the new molecular descriptors (E_M, E_S, L_M and (L_S^+)) were used to analyze the relationships between the cyclotides: three distinct groups of different levels of activity emerged as demonstrated by the dendrogram in Figure 8. The cyclotide subfamilies clustered into different subgroups, with most of the Möbius cyclotides being in groups 1 and 2, and all of the bracelets being in group 3.

Aside from psyle A (H5), all hybrid cyclotides clustered in group 2. Two of them, kB8 (H4) and tricyclone A (H6), clustered with synthetic cyclotides that are known to be biologically inactive in antihelmintic assays, [W23K]-kB1 (M23), [V25K]-kB1 (M24) [20], or hemolytic assays, [W23K, P24N, V25K]-kB1 (H1), and [P24D, V25K]-kB1 (H3) [30]. These mutant cyclotides differ from native Möbius cyclotides, and resemble native hybrids, in that they have relatively low degree of lipophilicity in loop 5. These hybrid cyclotides exhibit a high degree of sequence identity with their Möbius counterparts in all domains other than loops 5 and 6. As such, it seems that reducing the lipophilicity of loop 5 decreases membrane activity of the Möbius cyclotide subfamily. We propose

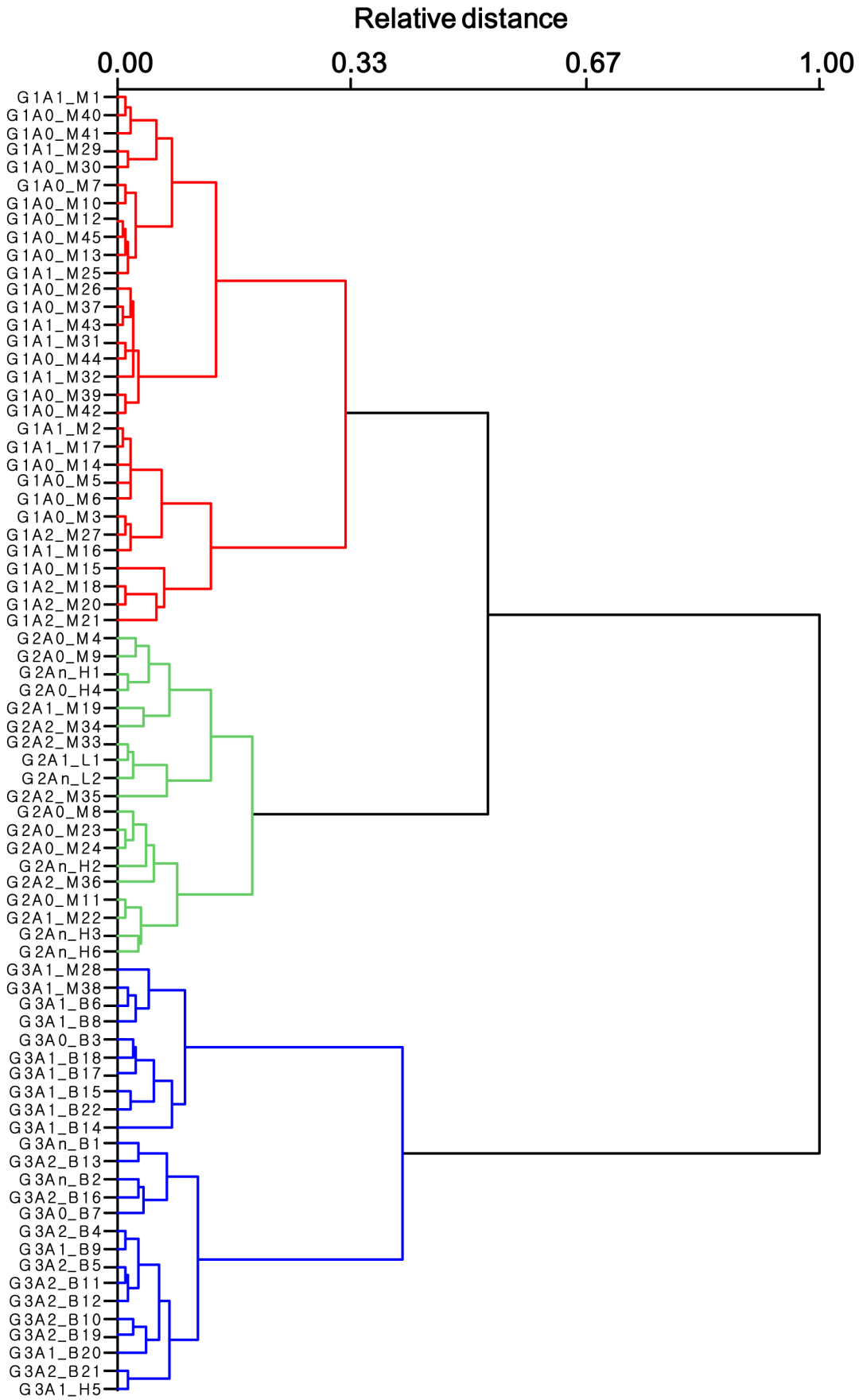
Relative distance
0.00
0.33
0.67
1.00
G1A1_M1
G1A0_M40
G1A0_M41
G1A1_M29
G1A0_M30
G1A0_M7
G1A0_M10
G1A0_M12
G1A0_M45
G1A0_M13
G1A1_M25
G1A0_M26
G1A0_M37
G1A1_M43
G1A1_M31
G1A0_M44
G1A1_M32
G1A0_M39
G1A0_M42
G1A1_M2
G1A1_M17
G1A0_M14
G1A0_M5
G1A0_M6
G1A0_M3
G1A2_M27
G1A1_M16
G1A0_M15
G1A2_M18
G1A2_M20
G1A2_M21
G2A0_M4
G2A0_M9
G2An_H1
G2A0_H4
G2A1_M19
G2A2_M34
G2A2_M33
G2A1_L1
G2An_L2
G2A2_M35
G2A0_M8
G2A0_M23
G2A0_M24
G2An_H2
G2A2_M36
G2A0_M11
G2A1_M22
G2An_H3
G2An_H6
G3A1_M28
G3A1_M38
G3A1_B6
G3A1_B8
G3A0_B3
G3A1_B18
G3A1_B17
G3A1_B15
G3A1_B22
G3A1_B14
G3An_B1
G3A2_B13
G3An_B2
G3A2_B16
G3A0_B7
G3A2_B4
G3A1_B9
G3A2_B5
G3A2_B11
G3A2_B12
G3A2_B10
G3A2_B19
G3A1_B20
G3A2_B21
G3A1_H5

Figure 8. A dendrogram showing the hierarchical clustering of 75 cyclotides based on their molecular descriptors. The descriptors considered include the lipophilic moment (L_M), exclusive lipophilicity (L_S^+), positively charged surface area (E_S) and HBD amphipathic moment (E_M). The left column shows the cyclotides' labels, which indicate the subgroup to which they belong (indicated by the letter 'G'), relative activity (indicated by the letter 'A') and a subfamily-specific letter (M for Möbius, B for bracelet, H for hybrid and L for linear). The relative distance measures the physicochemical dissimilarity between cyclotides in respect to the normalized scale. The relative activity was calculated with respect to the cytotoxicity of kB1, and three levels of activity were defined: A0 for peptides whose relative activity ranges of >1.0, A1 for those with relative activities between 0.2 and 1.0, and A2 for those with relative activities of <0.2. Those with unknown activity were labeled An. Aside from [Mee]-cyO2 and varv A, all of the cyclotides were assigned to the same activity groups for both cytotoxicity and anthelmintic activity. The cyclotide subfamilies clustered into different subgroups, with most of the Möbius cyclotides being in groups 1 (red) and 2 (green), and all of the bracelets being in group 3 (blue). Aside from psyle A (H5), all of the hybrid cyclotides clustered in group 2. Group 2 also contains some unusual cyclotides that exhibit high or intermediate activity, namely cyO14 (M33), cyO15 (M34), cyO16 (M35), [T20K, S22K]-kB1 (M19) and [S22K]-kB1 (M22). These cyclotides are all rich in lysines.
doi:10.1371/journal.pone.0091430.g008

that this loop is important because it contributes significantly to the lipophilic contact area through which the Möbius cyclotide adsorbs to the membrane, and to the asymmetric distribution of lipophilicity within those peptides.

Loop lipophilicity and its impact on membrane orientation. The cyclotide subfamilies interact with membranes in ways that reflect their differences in terms of the distribution of lipophilicity across the molecular surface. Due to the high levels of sequence similarity within cyclotide subfamilies and their very similar folds, cyclotides from the same subfamily generally have very similar spatial distributions of lipophilicity and therefore adopt very similar orientations when associated with membranes (Figure 3 and 9). Strongly lipophilic loops are typically buried in the interior of the membrane. This is consistent with the predictions of the "Orientations of Proteins in Membranes" (OPM) database for these peptides [63]. The Möbius and hybrid subfamilies adopt different orientations on the membrane because of their differences with respect to the lipophilicity of loop 5: in the Möbius cyclotides, this loop is rich in lipophilic residues but in the hybrids, it is relatively lipophobic. Consequently, whereas the Möbius cyclotides have large membrane-buried regions that include this loop, hybrid cyclotides cannot interact with membranes in this way. Loop 6 is the largest loop in all of the cyclotide subfamilies, and is the loop with the highest degree of sequence diversity within individual subfamilies. Therefore, the lipophilicity of this loop differs significantly between individual members of specific subfamilies. Even though it has negative lipophilicity values in many members of the Möbius and hybrid subfamilies, some residues from loop 6 may nevertheless become buried in the membrane.

The conformational constraints imposed by the cyclic cystine knot (CCK) motif increase the "lipophilic intensity" (see Figure 5) of the cyclotides by increasing the solvent-accessible surface area (SASA) of the lipophilic residues and reducing the SASA of lipophobic residues. The cystine knot is deeply buried in the protein core, and each pair of cystines fixes the end points of the loops in close proximity to one-another. Thus, rather than interacting directly with the membrane, the cystines force the backbones of the loops to curve outwards. Because of this knot structure, the lipophilic side-chains of cyclotides are generally oriented towards the exterior of the peptide in convex loops with a high rate of solvent accessible surface, despite this energetically unfavorable state. This combination of high lipophilicity and a convex backbone is illustrated in Figure 9, which shows loops 5 and 2 of the Möbius and bracelet subfamilies, respectively. Loop 5 and loop 2 are the most lipophilic loops in these subfamilies, and both contain aromatic residues in close proximity to a proline residue. In loop 5 of the Möbius subfamily, the aromatic residue (X^{Ar}) is directly adjacent to a *cis*-proline (X^{Ar}+*cis*-Pro). In loop 2 of the bracelet subfamily, the aromatic residue is separated from a *trans*-proline residue by a single intermediate residue (X) that has an alkyl side chain (X^{Ar}+X+*trans*-Pro). In contrast to loop 5 of the Möbius cyclotides, in which the aromatic residue stacks on top of its neighboring proline, the aromatic side chain in loop 2 of the bracelet subfamily protrudes outwards and does not interact with its neighboring residues. This extraordinary degree of protrusion increases the solvent-accessible surface area of the aromatic side chain and therefore significantly increases the peptide's lipophilicity. Furthermore, the convexity of the backbone increases when the loop is constrained by the CCK motif. Figure 9 shows the lipophilic intensity of each residue in three representative cyclotides in two conformations: the native fold and a hypothetically unfolded conformation in which $\Phi = \psi = 180^\circ$. The lipophilicity values for most residues are comparable in both conformations, but there are some important exceptions.

QSAR analysis

QSARs assume that the biological activity of a compound is dependent on its physicochemical properties. Such relationships can be illustrated quantitatively when a series of related compounds are considered. The complexity of QSAR models stems from a number of different factors. One of these is that the introduction of additional terms into the equation can easily increase the square correlation coefficient (r^2) via a phenomenon that is known as overfitting. This is especially problematic if the new term represents a variable that conveys physicochemical information that is also present in one of the other terms used in the model. In addition, the inclusion of polynomial terms, products of multiple terms, or descriptors whose meaning is obscure can produce models that are very accurate but hard to interpret. We therefore aimed to develop models that use a relatively small number of easily interpreted descriptors (Table 3).

The QSAR models were established using an equation that relates the activity of the studied cyclotides to the four physicochemical descriptors discussed in the preceding sections: the exclusive lipophilicity (L_S^+), the positively charged surface area (E_S), and the corresponding moments (L_M and E_M) (Table 4). A process termed "dummy variable regression" was used to explain the observation that the cyclotides could be divided into two groups that had different moment values, with the high moment value group having high or intermediate activity and the low moment group having low activity. We hypothesized that relative activity only correlates with L_S^+ and E_S when the corresponding quantities were asymmetrically distributed over the molecular surface as a whole. The two groups of cyclotides were analyzed using a single regression equation featuring a dummy variable regressor, τ, which assigns a value of one to the first group and zero to the second. The hypothesis is embodied in the following regression equation: $\mathrm{Activity} = k\cdot(\tau\cdot L_S^+) + l\cdot(\tau\cdot E_S) + m$, where k, l and m are constants. We then sought to define a critical point in terms of a pair of threshold moment values that would discriminate between cyclotides from the high and low activity

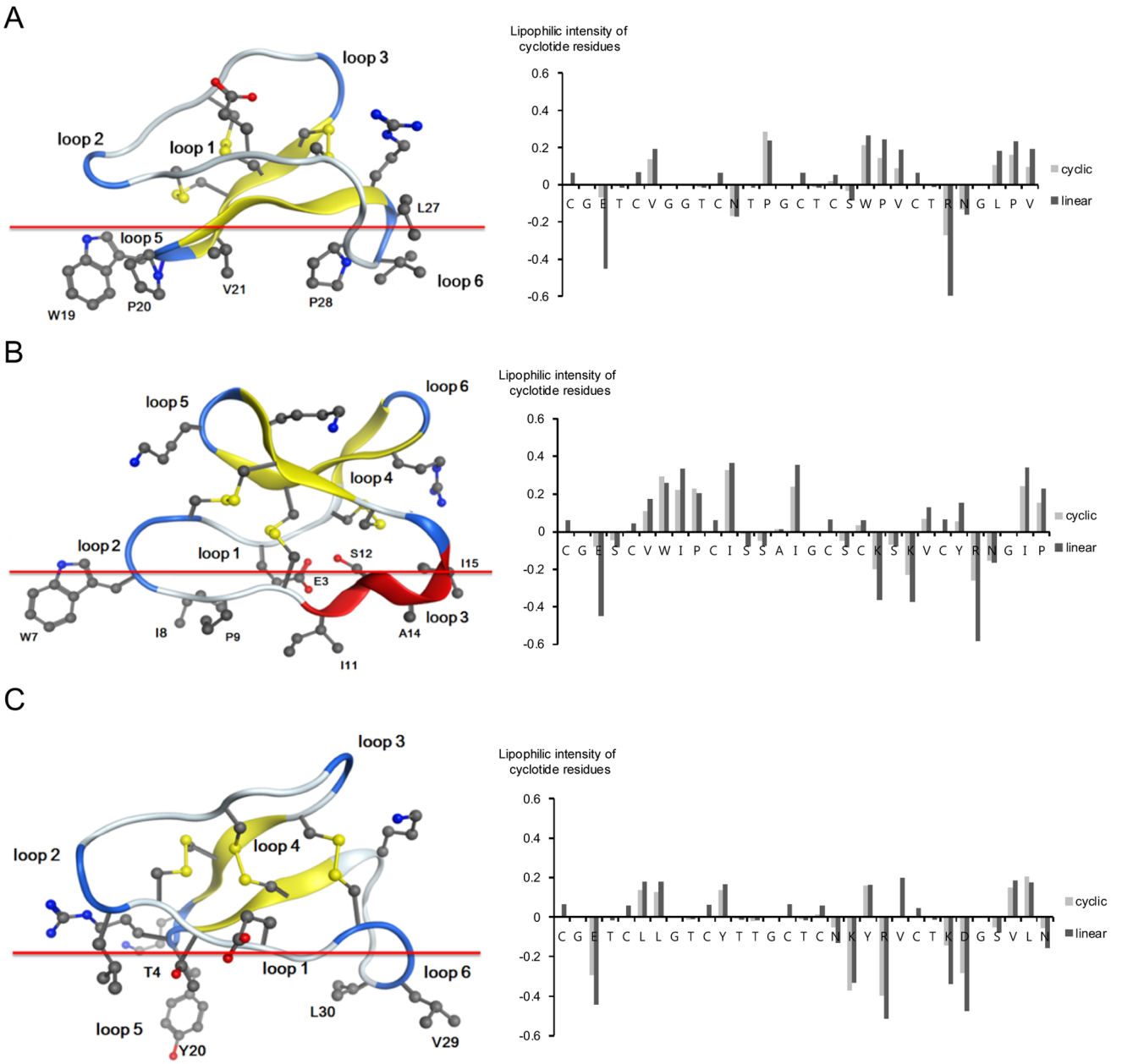

Figure 9. Predicted membrane binding modes (left) and lipophilicity profiles (right) for kalata B1 (A), cycloviolacin O2 (B) and kalata B8 (C) in their native cyclic conformations and in hypothetically unfolded conformations. The outer phosphate layer of a phospholipid membrane is represented by a red line in each case. Strained backbone regions with little structural mobility are indicated in blue. Backbone regions with a defined secondary structure are colored in yellow and red to indicate β-strands and α-helices, respectively. Only selected side chains are depicted, including those that form the hydrophobic patch, residues carrying positive charges, and the conserved glutamic acid in loop 1. The cyclotide structures were obtained from the PDB server; their PDB IDs are 1NB1 (kalata B1) [28], 2KNM (cycloviolacin O2) [33], and 2B38 (kalata B8) [16]. The lipophilicity profile of each residue was calculated for the natively folded cyclotides and for their unfolded conformations. The hypothetical unfolded cyclotide conformations were defined such that $\psi = \varphi = 180^{\circ}$. It should be noted that in the natively folded conformations of kalata B1 (kB1) and cycloviolacin O2 (cyO2), the degree of solvent exposure for lipophilic residues is high while that for lipophobic residues is low. For example, in loop 2 of cyO2, the lipophilicity of Trp and Pro is higher in the native fold than in the unfolded conformation. Moreover, the lipophobicity values for polar residues are relatively high in natively folded kB8 compared to natively folded kB1 and cyO2. In contrast to the situation for kB1 and cyO2, the polar residues of kB8 exhibit similar lipophobicities in the native and unfolded conformations. Notably, the glutamic acid residue in loop 1 of kB8 participates in an internal hydrogen-bonding network within the core of the protein that effectively minimizes its SASA value. Consequently, it exhibits a much higher level of lipophobicity in the native fold than do the equivalent residues of kB1 and cyO2.
doi:10.1371/journal.pone.0091430.g009

groups, such that cyclotides in one group would have moment values above the critical point, while those in the other group would have moments below the critical point. A range of potential critical points was evaluated by considering the equations arising in each case; the optimal critical point was identified as that for which the correlation coefficient (r^2) and cross-validated correlation coefficient (r^2_{cv}) for the regression equation were maximized.

Table 3. Molecular descriptors used in the QSAR models.

Descriptors	Description
τ	The variable τ is a diagonal matrix whose diagonal elements (τ_{ii}) consist of the spline terms of E_M and L_M. The value of a diagonal element (τ_{ii}) is zero if both moments of the i^{th} cyclotide are below those of the critical point. Otherwise, the value of τ_{ii} is one.
E_S	The positively charged surface area
L_S^+	The exclusive lipophilicity. This quantity represents the sum of the lipophilic intensities of residues that have positive values on the lipophilicity scale and are located within regions that are, overall, attractive to lipids. In contrast, the quantity $\tau \cdot L_S^+$ gives a measure of the lipophilicity of the lipophilic domain, i.e. the lipophilicity of all residues located on regions of the molecular surface that would be in contact with lipids when the protein is associated with a membrane.

doi:10.1371/journal.pone.0091430.t003

The generation of the model and the introduction of the dummy variable τ is described in Figure 10.

The quality of the QSAR models could be improved to a remarkable extent by explaining the activity of cyclotides in a non-linear fashion using the exclusive lipophilicity (L_S^+) and positively charged surface area (E_S) variables. This was done because we identified two groups of cyclotides with different moment values whose contact surfaces affect their activity in different ways, which we took to be indicative of a nonlinear correlation. Figure 11 shows the evaluation of model equations with these statistical parameters on the moment plot (L_M, E_M). The equation becomes a linear model when all of the diagonal components of τ are set to either 0 or 1. For all i, τ_{ii} is 1 when a critical point is assigned on the grid at (0,0), and τ_{ii} is 0, when a critical point is assigned on the grid at (0.050, 9.000). If a valid model (i.e. one with acceptable values of r^2 and r_{cv}^2) is obtained when the critical point is assumed to be located at (0, 0), the implication is that there is a positive correlation between activity and the surface area variables L_S^+ and E_S for all cyclotides. Conversely, if a valid model is obtained when the critical point is assumed to be located at (0.050, 9.000) on the grid, the implication is that there is no correlation between the activity and L_S^+ or E_S that is valid for all cyclotides. For cytotoxic activity, the critical point was determined to be (0.03750, 4.5000), and 7 of the 30 cyclotides for which the literature contains quantitative cytotoxicity data have moment values that place them below the critical point. For anthelmintic activity, the critical point was determined to be (0.02813, 5.6250); 19 of the 46 cyclotides for which the literature contains quantitative anthelmintic activity data have moments below this threshold.

When constructing the QSAR model for anthelmintic activity, the cyclotides were classified as non-active if their IC_{50} values exceeded the highest tested concentration, 11.5 μM [20]. The logarithm of the IC_{50} value is conventionally used to describe the potency of peptides. However, in this work we used a log(x+1) data transformation rather than a conventional ln(x) transformation in order to enable the inclusion of non-active cyclotides in the QSAR model. For cyclotides that were classified as active, the value of x was calculated based on their relative activity compared to kalata B1. Put simply, x is the ratio of the IC_{50} values of the cyclotide and kalata B1. Non-active cyclotides were assumed to have x values of 0, in which case the log(x+1) transformation also yields a value of 0 (= log1). This prevents the transformed values from approaching negative infinity as they would if a simple ln(x) transformation were used but does not affect the normality of the data and keeps the variance relatively constant.

After inspecting the output of the initial QSAR models, outliers were removed from the data sets and new QSAR models were created. For the cytotoxicity QSAR model, the atypical linear cyclotide psyle C was removed. Similarly, cyO14, cyO15 and cyO16 were removed as outliers before generating the refined anthelmintic activity model. These cyclotides exhibit high cytotoxic activity despite having exceptionally low E_M and L_M moments, are rich in lysines, and stands out in when compared to other Möbius cyclotides. This discrepancy may indicate that these cyclotides exert their cytotoxic activity by different mechanisms to the other cytotoxic cyclotides. These three peptides have large positively charged surface areas that may form favorable electrostatic interactions with the membrane. In this context, it should be noted that a deep lipophilic patch is not an absolute prerequisite for membrane interaction [64]; it may be that these proteins adhere to the membrane surface due to electrostatic interactions with its charged phospholipid head groups and the moderately polar glyceryl/carbonyl groups just below them.

Our model for predicting cyclotide potency does not require any specific assumptions based on prior knowledge such as the location of PE binding sites or bioactive faces. Instead, it assumes that the potency of membrane-active peptides is determined by their affinity for membranes, which is in turn dictated by the size and heterogeneity of the contact surface. The location of the contact surface can be estimated by considering the orientation of the lipophilic vector; if the predicted contact surface features a large proportion of lipophobic side chains, the peptide would not be predicted to interact strongly with membranes. This hypothesis explained the observed activity of most of the studied cyclotides quite well, as indicated by the dendrogram shown in Figure 8. We suggest that those peptides whose activity is not well explained by this hypothesis exert their effects by some alternative mechanism. Recently, the first such specific cyclotide receptor interaction was reported: the binding of kalata B7 to the oxytocin receptor [65]. Is it possible that the current model can be used to study such

Table 4. QSAR equations for activity against the U-937GTB line and *H. contortus*.

No	Equations	r^2	r_{cv}^2	F	F^c
1	$\ln(IC_{50}) = -0.4748\tau \cdot L_S^+ - 0.003000\tau \cdot E_S + 3.0962$	0.65	0.56	20.13	4.58
2	$\ln(Rel+1) = -1.7445\tau \cdot L_S^+ + 0.004703\tau \cdot E_S + 0.51755$	0.63	0.56	35.01	4.32

Equations 1 and 2 represent the QSAR models for activity against the U-937GTB line and *H. contortus*, respectively. The raw values of L_S^+ were transformed using the Box-Cox method, according to which $y' = y^{\lambda}$, where $\lambda = 1.0$ and -0.5 in equation 1 and 2, respectively. In equation 2, this transformation inverts the order of the L_S^+ data set. Therefore, the negative sign of the coefficient reflects a positive correlation between L_S^+ and the measured activity values. The F^c is the critical F-value at confidence level of 95%. "Rel" refers to the relative activity of the cyclotide in question compared to kalata B1.

doi:10.1371/journal.pone.0091430.t004

If the activity of the cyclotides is assumed to be quantitatively related to their physicochemical properties, a model equation can be introduced to correlate this relationship with the regression coefficients (β_0, β_1, β_2). For the i^{th} cyclotide, the activity y_i in a matrix Y, depends on the two molecular descriptors, E_S and L_S^+, as x_{i1} and x_{i2} in matrix X, respectively:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}, \; X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{i1} & x_{i2} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}_{n\times 2}, \; B = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{2\times 1}, \; B' = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix}_{n\times 1},$$

$$\tau = \begin{bmatrix} \tau_{11} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \tau_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & \tau_{ii} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \tau_{nn} \end{bmatrix}_{n\times n} \text{ and } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}.$$

All of the elements of the diagonal matrix, τ, are equal to zero (τ_{ij} for $i \neq j$) except for the diagonal elements. The diagonal elements (τ_{ii}) have binary values (0 or 1), and are determined by the E_M and L_M values of i^{th} cyclotide relative to the critical point (a, b):

$$\tau_{ii} = \left(\frac{\langle L_M - a \rangle}{(L_M - a)} \vee \frac{\langle E_M - b \rangle}{(E_M - b)} \right),$$

where "v" is the Boolean logical operator that relates two binary entities with "or", and (p v q) is equivalent to the algebraic equation: p + q – p·q. The spline term [69], <F(x) - t>, is equal to zero if the value of (F(x) - t) is negative; otherwise it has the same value as (F(x) - t). Here, F(x) is a function of x as a variable, and t is constant. The matrix ε represents the error of model, and for the i^{th} cyclotide, the error is the difference between the observed the activity (y_i) and the predicted activity, $\tau_{ii} \cdot x_{i1} \cdot \beta_1 + \tau_{ii} \cdot x_{i2} \cdot \beta_2 + \beta_0$.

All of the cyclotides are scattered on the plot made using their E_M and L_M values as shown in figure 5. The values range from 0 to 0.0500 for L_M, and 0 to 9.00 for E_M. The moment plot (L_M, E_M) is divided over 25 grid points by grid spacing (0.0125, 2.250). On each grid point, the equation, $Y = \tau \cdot X \cdot B + B' + \varepsilon$ was evaluated according to the r^2 (square correlation coefficient) and r_{cv}^2 (cross-validated r^2) values for the model. The four grid points with the highest r^2 and r_{cv}^2 values were selected and the rectangular plane with those four grid points as its end points, was subdivided using nine further grid points, at which point the process was repeated iteratively. The grid spacing on the k^{th} iteration was determined using the expression $\frac{1}{2^k} \cdot (0.0125, 2.250)$. The iterative grid search was halted when the fit of the equation stopped improving. Because the matrix τ has different values for different critical points, each equation is associated with different values of r^2 and r_{cv}^2. We identified the critical point that maximized the values of r^2 and r_{cv}^2, and selected the corresponding equation as the optimized QSAR model.

Figure 10. Generation of QSAR models.
doi:10.1371/journal.pone.0091430.g010

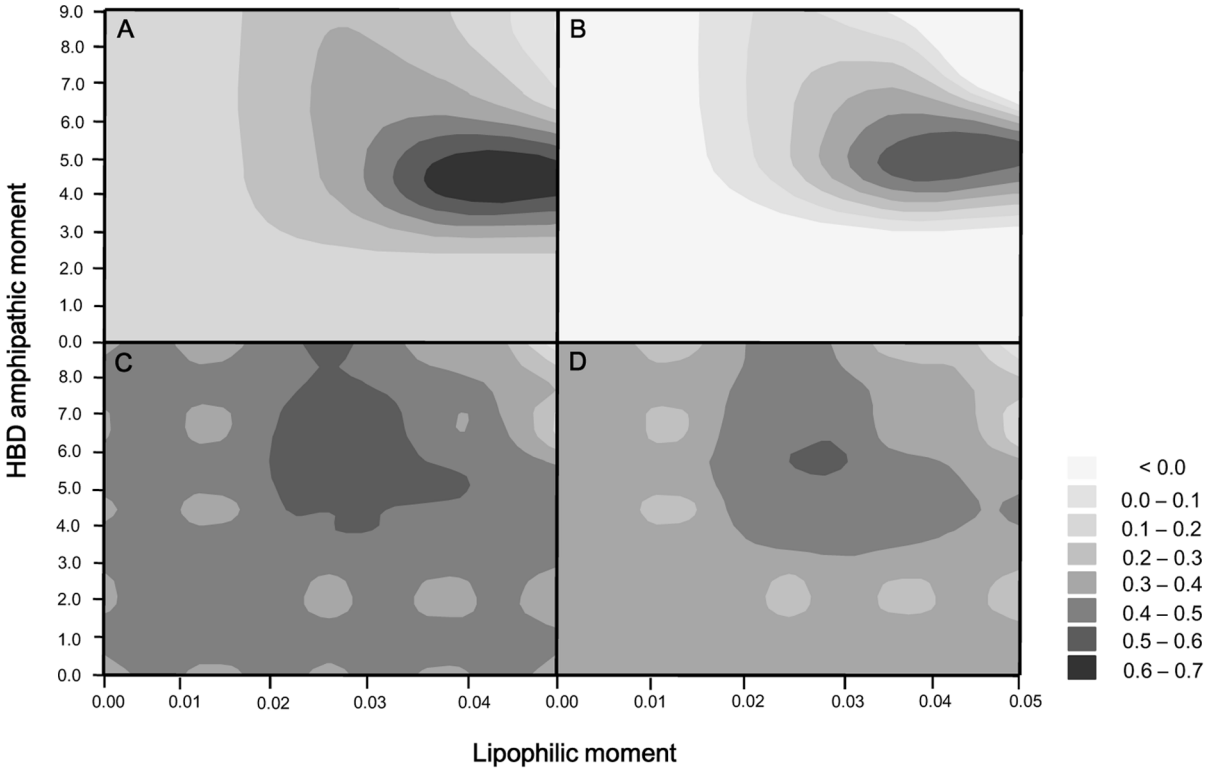

Figure 11. Statistical parameters for the moment plots (L_M, E_M). The grayscale values indicate the performance of the QSAR models based on the correlation coefficient (r^2) and the cross-validated correlation coefficient (r^2_{cv}) for the moment plot. Panels A and B show the values of r^2 and r^2_{cv} for the model describing cyclotide cytotoxicity against the lymphoma cell line, while panels C and D show the r^2 and r^2_{cv} values for anthelmintic activity. Cyclotides whose moment values were below the critical point were placed in the low moment group, while those whose moments exceeded the critical point were placed in the high moment group. The critical points were determined to be (0.03750, 4.5000) and (0.02813, 5.6250) for activity against the U937GTB line and *H. contortus*, respectively.
doi:10.1371/journal.pone.0091430.g011

specific, ''classical'', protein receptor – peptide ligand interactions too? The answer is yes, because the fundamental principle of affinity is the same between protein-peptide interactions and protein-membrane interactions: the affinity increases with contact surface area. However, in the current case, the suitability of the model will depend heavily on rigidity and size of the peptide and the type of force mediating molecular interaction to the target. For extremely rigid peptides, such as cyclotides, in which there is no room for structural rearrangements upon docking into a receptor, it is clear that the full surface area that interacts with the target must be taken into account. The model could also be used to discover possible specific interactions: a peptide that clusters together with peptides of similar physicochemical properties but stands out in terms of activity indicates the presence of a specific interaction.

To facilitate the interpretation of the QSAR models, we chose to disregard molecular descriptors that exhibited any co-linearity with molecular size such as molecular weight and flexibility [66]. Because cyclotide activity correlates positively with the size of the membrane contact area, the use of descriptors that contain information on molecular size could potentially result in misleading causal interpretations. In particular, if it is assumed that the surface distribution of lipophilicity and high electrostatic potential do not vary, the potency of the molecules would increase with their size (in proportion to $\sqrt{L_S^2+E_S^2}$). There is a positive linear correlation between flexibility and molecular size ($r^2 = 0.79$) for molecules containing the CCK motif (Figure 12). The cyclic backbone and the disulfide bond network constrain the peptide's degree of freedom (which is proportional to flexibility), and the degree of freedom increases with the number of atoms (which is proportional to molecular weight). However, it has been reported that enhanced flexibility also reduces cyclotide activity due to reduced membrane adsorption [67] and stability [68].

The cytotoxicity and anthelmintic activity models both have correlation coefficients between 0.6 and 0.7. We believe that the main sources of error in the models relate to the fact that membrane-active proteins necessarily bind to membranes via non-specific interactions. For example, cyclotides drift over the fluid membrane surface, and can adopt a wide range of tilt angles as they do so. In addition, cyclotides can form oligomers on the membrane surface. Our model explains the potency of the cyclotides purely in terms of their physical contact with membranes. However, this approach implicitly assumes that the peptides interact as monomers and adopt a single static conformation with respect to the membrane surface. While the physical parameters (the lipophilic and HBD amphipathic surface areas and their distribution) used in this work were derived directly from the protein structure alone, their relevance for membrane binding is supported by the OPM database. Strong correlations have often been reported for QSAR models based on specific interactions such as ligand-protein interactions, but lower coefficients are reasonable for models based on non-specific interactions such as those that govern the binding of membrane-active proteins. Indeed, with one exception, no previous QSAR study on membrane-active peptides has yielded a regression coefficient

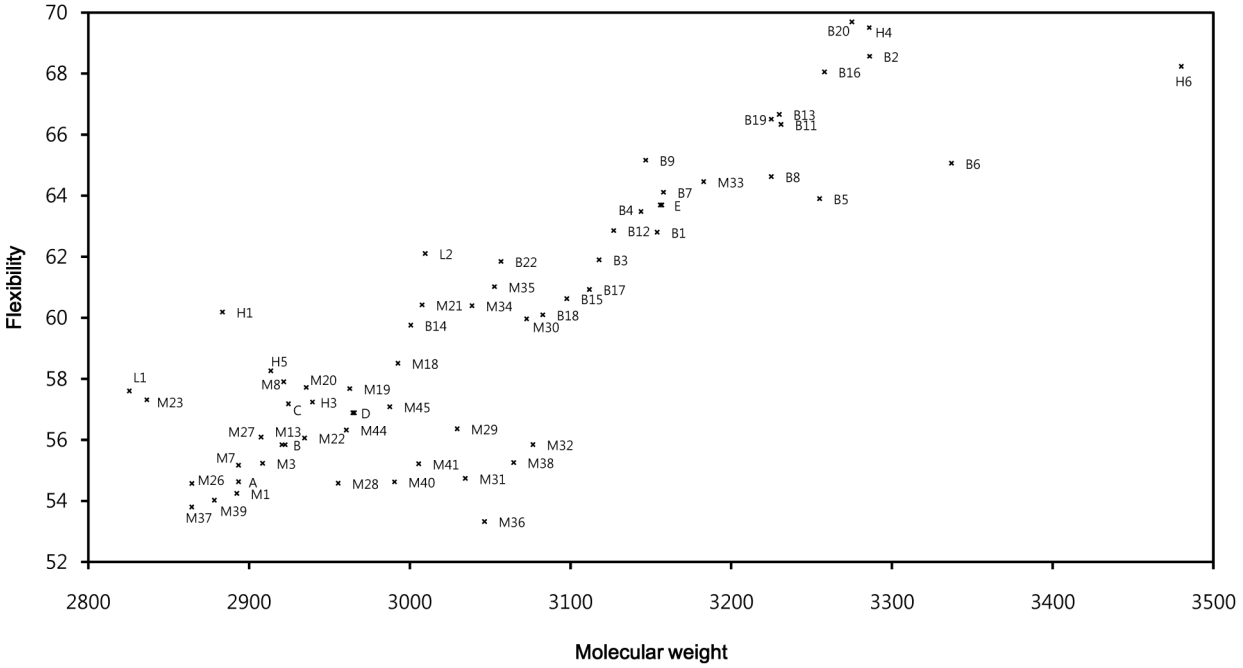

Figure 12. Chart of molecular weight and flexibility in cyclotides. This chart illustrates the correlation between flexibility and molecular weight. Points where multiple cyclotides overlap are indicated by the letters A-E; the cyclotides at these points are: A (M42, M43), B (M5, M9, M12, M14, M17, M24, M25), C (M4, H2), D (M2, M6, M10, M11, M16) and E (B10, B21). The linear cyclotide (L1 and L2) is relatively flexible since it lacks a cyclic backbone. The members of the hybrid cyclotide subfamily (H4, H6) have high molecular weights and flexibility values.
doi:10.1371/journal.pone.0091430.g012

in excess of 0.7 [23,24,26]. The sole exception is a study that used more than twenty molecular descriptors to obtain an r^2 value of more than 0.9 [25].

Conclusions

Cyclotides exert their biological effects via membrane disruption. We have presented a quantitative analysis of their structure-activity relationships based on two molecular properties that are important in membrane interactions: lipophilicity and an electrostatic property. Both of these properties were characterized using two dimensional quantities (a scalar and a moment), and their correlations with the peptides' biological activities were evaluated using one model equation. Our QSAR model suggests that there is a non-linear positive correlation between cyclotide potency and the (scalar) size of the molecular surface that interacts with the membrane. The nonlinearity is explained by the moments; a linear positive correlation is only observed for peptides whose lipophilic and electrostatic properties are unevenly distributed on the molecular surface. Furthermore, we qualitatively demonstrated that these molecular descriptors are useful in explaining how the physicochemical properties of cyclotides differ between subfamilies, and how these differences affect the orientation of cyclotides on the membrane and their membrane activity. Our results also illustrate how the cyclic cystine knot (CCK) motif forces cyclotides to adopt a conformation that enhances their lipophilicity.

The approach presented herein could be extended to cover other membrane-active proteins that contain non-standard amino acids. This is because lipophilicity and electronic charge data are readily determined for all amino acids, including non-standard ones. In addition, it can provide clear guidelines for drug design by predicting how specific changes will affect the surface properties of the peptide (i.e. the number of surface-accessible lipophilic and positively charged residues) and their relative distribution (i.e. their location on the molecular surface relative to other residues with similar properties).

Supporting Information

Figure S1 RMSD distribution and solvent surface area rate of cyclotides' NMR ensembles.
(PDF)

Figure S2 Example of calculation procedure of lipophilic parameters (L_M and L_S).
(PDF)

Acknowledgments

We wish to thank Dr. Pravech Ajawatanawong for providing a great deal of help with molecular modeling. We also acknowledge Drs. Kiwoong Nam and Inseok Doo for invaluable discussions regarding statistical analysis. We thank Dr. Andrei Lomize for critically reading the manuscript.

Author Contributions

Conceived and designed the experiments: SP AAS UG. Performed the experiments: SP. Analyzed the data: SP AAS UG. Contributed reagents/materials/analysis tools: UG. Wrote the paper: SP AAS UG.

References

1. Craik DJ, Daly NL, Bond T, Waine C (1999) Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J Mol Biol 294: 1327–1336.
2. Göransson U, Burman R, Gunasekera S, Strömstedt AA, Rosengren KJ (2012) Circular proteins from plants and fungi. J Biol Chem 287: 27001–27006.

3. Colgrave ML, Craik DJ (2004) Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. Biochemistry 43: 5965–5975.
4. Gruber CW, Elliott AG, Ireland DC, Delprete PG, Dessein S, et al. (2008) Distribution and evolution of circular miniproteins in flowering plants. Plant Cell 20: 2471–2483.
5. Hashempour H, Koehbach J, Daly NL, Ghassempour A, Gruber CW (2013) Characterizing circular peptides in mixtures: sequence fragment assembly of cyclotides from a violet plant by MALDI-TOF/TOF mass spectrometry. Amino Acids 44: 581–595.
6. Poth AG, Colgrave ML, Philip R, Kerenga B, Daly NL, et al. (2011) Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins. ACS Chem Biol 6: 345–355.
7. Poth AG, Mylne JS, Grassl J, Lyons RE, Millar AH, et al. (2012) Cyclotides associate with leaf vasculature and are the products of a novel precursor in *petunia* (Solanaceae). J Biol Chem 287: 27033–27046.
8. Nguyen GK, Lian Y, Pang EW, Nguyen PQ, Tran TD, et al. (2013) Discovery of linear cyclotides in monocot plant *Panicum laxum* of Poaceae family provides new insights into evolution and distribution of cyclotides in plants. J Biol Chem 288: 3370–3380.
9. Craik DJ, Daly NL, Mulvenna J, Plan MR, Trabi M (2004) Discovery, structure and biological activities of the cyclotides. Curr Protein Pept Sci 5: 297–315.
10. Jennings C, West J, Waine C, Craik D, Anderson M (2001) Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from *Oldenlandia affinis*. Proc Natl Acad Sci U S A 98: 10614–10619.
11. Tam JP, Lu YA, Yang JL, Chiu KW (1999) An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. Proc Natl Acad Sci U S A 96: 8913–8918.
12. Ovesen RG, Brandt KK, Göransson U, Nielsen J, Hansen HC, et al. (2011) Biomedicine in the environment: cyclotides constitute potent natural toxins in plants and soil bacteria. Environ Toxicol Chem 30: 1190–1196.
13. Lindholm P, Göransson U, Johansson S, Claeson P, Gullbo J, et al. (2002) Cyclotides: a novel type of cytotoxic agents. Mol Cancer Ther 1: 365–369.
14. Gustafson KR, Sowder RC, Henderson LE, Parsons IC, Kashman Y, et al. (1994) Circulins A and B. Novel Hiv-Inhibitory Macrocyclic Peptides from the Tropical Tree *Chassalia parvifolia*. J Am Chem Soc 116: 9337–9338.
15. Aboye TL, Ha H, Majumder S, Christ F, Debyser Z, et al. (2012) Design of a novel cyclotide-based CXCR4 antagonist with anti-human immunodeficiency virus (HIV)-1 activity. J Med Chem 55: 10729–10734.
16. Daly NL, Clark RJ, Plan MR, Craik DJ (2006) Kalata B8, a novel antiviral circular protein, exhibits conformational flexibility in the cystine knot motif. Biochem J 393: 619–626.
17. Henriques ST, Huang YH, Rosengren KJ, Franquelim HG, Carvalho FA, et al. (2011) Decoding the membrane activity of the cyclotide kalata B1: the importance of phosphatidylethanolamine phospholipids and lipid organization on hemolytic and anti-HIV activities. J Biol Chem 286: 24231–24241.
18. Burman R, Herrmann A, Tran R, Kivela JE, Lomize A, et al. (2011) Cytotoxic potency of small macrocyclic knot proteins: Structure-activity and mechanistic studies of native and chemically modified cyclotides. Org Biomol Chem 9: 4306–4314.
19. Simonsen SM, Sando L, Rosengren KJ, Wang CK, Colgrave ML, et al. (2008) Alanine scanning mutagenesis of the prototypic cyclotide reveals a cluster of residues essential for bioactivity. J Biol Chem 283: 9805–9813.
20. Huang YH, Colgrave ML, Clark RJ, Kotze AC, Craik DJ (2010) Lysine-scanning Mutagenesis Reveals an Amendable Face of the Cyclotide Kalata B1 for the Optimization of Nematocidal Activity. J Biol Chem 285: 10797–10805.
21. Wang CK, Colgrave ML, Ireland DC, Kaas Q, Craik DJ (2009) Despite a conserved cystine knot motif, different cyclotides have different membrane binding modes. Biophys J 97: 1471–1481.
22. Herrmann A, Svangard E, Claeson P, Gullbo J, Bohlin L, et al. (2006) Key role of glutamic acid for the cytotoxic activity of the cyclotide cycloviolacin O2. Cell Mol Life Sci 63: 235–245.
23. Ostberg N, Kaznessis Y (2005) Protegrin structure-activity relationships: using homology models of synthetic sequences to determine structural characteristics important for activity. Peptides 26: 197–206.
24. Langham AA, Khandelia H, Schuster B, Waring AJ, Lehrer RI, et al. (2008) Correlation between simulated physicochemical properties and hemolycity of protegrin-like antimicrobial peptides: predicting experimental toxicity. Peptides 29: 1085–1093.
25. Bhonsle JB, Venugopal D, Huddler DP, Magill AJ, Hicks RP (2007) Application of 3D-QSAR for identification of descriptors defining bioactivity of antimicrobial peptides. J Med Chem 50: 6545–6553.
26. Frecer V (2006) QSAR analysis of antimicrobial and haemolytic effects of cyclic cationic antimicrobial peptides derived from protegrin-1. Bioorg Med Chem 14: 6065–6074.
27. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27: 221–224.
28. Rosengren KJ, Daly NL, Plan MR, Waine C, Craik DJ (2003) Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. J Biol Chem 278: 8606–8616.
29. Jennings CV, Rosengren KJ, Daly NL, Plan M, Stevens J, et al. (2005) Isolation, solution structure, and insecticidal activity of kalata B2, a circular protein with a twist: do Mobius strips exist in nature? Biochemistry 44: 851–860.
30. Clark RJ, Daly NL, Craik DJ (2006) Structural plasticity of the cyclic-cystine-knot framework: implications for biological activity and drug design. Biochem J 394: 85–93.
31. Daly NL, Koltay A, Gustafson KR, Boyd MR, Casas-Finet JR, et al. (1999) Solution structure by NMR of circulin A: a macrocyclic knotted peptide having anti-HIV activity. J Mol Biol 285: 333–345.
32. Koltay A, Daly NL, Gustafson KR, Craik DJ (2005) Structure of Circulin B and Implications for Antimicrobial Activity of the Cyclotides. Int J Pept Res Ther 11: 99–106.
33. Göransson U, Herrmann A, Burman R, Haugaard-Jonsson LM, Rosengren KJ (2009) The conserved glu in the cyclotide cycloviolacin O2 has a key structural role. Chembiochem 10: 2354–2360.
34. Ireland DC, Colgrave ML, Craik DJ (2006) A novel suite of cyclotides from *Viola odorata*: sequence variation and the implications for structure, function and stability. Biochem J 400: 1–12.
35. Wang CK, Hu SH, Martin JL, Sjogren T, Hajdu J, et al. (2009) Combined X-ray and NMR analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold. J Biol Chem 284: 10672–10683.
36. Trabi M, Craik DJ (2004) Tissue-specific expression of head-to-tail cyclized miniproteins in Violaceae and structure determination of the root cyclotide *Viola hederacea* root cyclotide1. Plant Cell 16: 2204–2216.
37. Mulvenna JP, Sando L, Craik DJ (2005) Processing of a 22 kDa precursor protein to produce the circular protein tricyclon A. Structure 13: 691–701.
38. Ireland DC, Colgrave ML, Nguyencong P, Daly NL, Craik DJ (2006) Discovery and characterization of a linear cyclotide from *Viola odorata*: implications for the processing of circular proteins. J Mol Biol 357: 1522–1535.
39. Sali A, Blundell TL (1993) Comparative Protein Modeling by Satisfaction of Spatial Restraints. J Mol Biol 234: 779–815.
40. Labute P (2010) LowModeMD—implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. J Chem Inf Model 50: 792–800.
41. Molecular Operating Environment (2012) 10. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: Chemical Computing Group Inc., 2012.
42. Molinspiration Cheminformatics (2014) www.molinspiration.com/cgi-bin/properties
43. Lee B, Richards FM (1971) Interpretation of Protein Structures - Estimation of Static Accessibility. J Mol Biol 55: 379–400.
44. Discovery Studio Modeling Environment (2007) Release 6.0 San Diego: Accelrys Software Inc.
45. Minitab Inc. (2010) Minitab 16 statistical Software. State College, PA
46. Strömstedt AA, Ringstad L, Schmidtchen A, Malmsten M (2010) Interaction between amphiphilic peptides and phospholipid membranes. Current Opinion in Colloid & Interface Science 15: 467–478.
47. Yeshak MY, Burman R, Asres K, Göransson U (2011) Cyclotides from an Extreme Habitat: Characterization of Cyclic Peptides from *Viola abyssinica* of the Ethiopian Highlands. J Nat Prod 74: 727–731.
48. Gerlach SL, Burman R, Bohlin L, Mondal D, Göransson U (2010) Isolation, Characterization, and Bioactivity of Cyclotides from the Micronesian Plant *Psychotria leptothyrsa*. J Nat Prod 73: 1207–1213.
49. Herrmann A, Burman R, Mylne JS, Karlsson G, Gullbo J, et al. (2008) The alpine violet, *Viola biflora*, is a rich source of cyclotides with potent cytotoxicity. Phytochemistry 69: 939–952.
50. Svangard E, Göransson U, Hocaoglu Z, Gullbo J, Larsson R, et al. (2004) Cytotoxic cyclotides from *Viola tricolor*. J Nat Prod 67: 144–147.
51. Colgrave ML, Kotze AC, Ireland DC, Wang CK, Craik DJ (2008) The anthelmintic activity of the cyclotides: Natural variants with enhanced activity. Chembiochem 9: 1939–1945.
52. Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253: 164–170.
53. Jones DT, Taylor WR, Thornton JM (1992) A new approach to protein fold recognition. Nature 358: 86–89.
54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
55. Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183: 63–98.
56. Morris AL, Macarthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical Quality of Protein-Structure Coordinates. Proteins-Structure Function and Genetics 12: 345–364.
57. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132.
58. Black SD, Mould DR (1991) Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. Anal Biochem 193: 72–82.
59. van de Waterbeemd H, Karajiannis H, Eltayar N (1994) Lipophilicity of Amino-Acids. Amino Acids 7: 129–145.
60. de Planque MR, Bonev BB, Demmers JA, Greathouse DV, Koeppe RE 2nd, et al. (2003) Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions. Biochemistry 42: 5341–5348.

61. Yau WM, Wimley WC, Gawrisch K, White SH (1998) The preference of tryptophan for membrane interfaces. Biochemistry 37: 14713–14718.
62. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In *Intermolecular Forces*, edited by Pullman B: D. Reidel publishing company. pp. 331–342.
63. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22: 623–625.
64. Chen Y, Guarnieri MT, Vasil AI, Vasil ML, Mant CT, et al. (2007) Role of peptide hydrophobicity in the mechanism of action of alpha-helical antimicrobial peptides. Antimicrob Agents Chemother 51: 1398–1406.
65. Koehbach J, O'Brien M, Muttenthaler M, Miazzo M, Akcan M, et al. (2013) Oxytocic plant cyclotides as templates for peptide G protein-coupled receptor ligand design. Proc Natl Acad Sci U S A 110: 21183–21188.
66. Hall LH, Kier LB (1991) The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. Reviews in Computational Chemistry 2:367–422.
67. Burman R, Strömstedt AA, Malmsten M, Göransson U (2011) Cyclotide-membrane interactions: defining factors of membrane binding, depletion and disruption. Biochim Biophys Acta 1808: 2665–2673.
68. Wang CK, Clark RJ, Harvey PJ, Rosengren KJ, Cemazar M, et al. (2011) The role of conserved Glu residue on cyclotide stability and activity: a structural and functional study of kalata B12, a naturally occurring Glu to Asp mutant. Biochemistry 50: 4077–4086.
69. Rogers D (1994) Application of genetic function approximation to quantitative structure-activity relationship and quantitative structure-property relationships. J Chem Inf Comput Sci 34: 854–866.