BBA - General Subjects 1865 (2021) 129989

Contents lists available at ScienceDirect

BBA - General Subjects

journal homepage: www.elsevier.com/locate/bbagen

Schistocins: Novel antimicrobial peptides encrypted in the *Schistosoma mansoni* Kunitz Inhibitor SmKI-1

B.P.O. Santos [a,b], E.S.F. Alves [c], C.S. Ferreira [d], A. Ferreira-Silva [e], A. Góes-Neto [b,e], R.M. Verly [d], L.M. Lião [c], S.C. Oliveira [b,f,g,1], M.T.Q. de Magalhães [a,b,*,1]

[a] *Laboratório de Biofísica de Macromoléculas, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil*

[b] *Programa Interunidades de Pós-graduação em Bioinformática, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil*

[c] *Laboratório de Ressonância Magnética Nuclear, Instituto de Química, Universidade Federal de Goiás, Goiânia, Goiás 74690-900, Brazil*

[d] *Laboratório de Síntese e Estrutura de Biomoléculas, Departamento de Química, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, Minas Gerais 39100-000, Brazil*

[e] *Laboratório de Biologia Molecular e Computacional de Fungos, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil*

[f] *Laboratório de Imunologia de Doenças Infecciosas, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil*

[g] *Instituto Nacional de Ciência e Tecnologia em Doenças Tropicais (INCT-DT), Conselho Nacional de Desenvolvimento Científico e Tecnológico, Ministério da Ciência Tecnologia e Inovação Salvador, Bahia, Brazil*

ARTICLE INFO

Keywords:
Cryptides
Peptide conformation
Peptide design
Nuclear magnetic resonance (NMR)

ABSTRACT

Background: Here we describe a new class of cryptides (peptides encrypted within a larger protein) with antimicrobial properties, named schistocins, derived from SmKI-1, a key protein in *Shistosoma mansoni* survival. This is a multi-functional protein with biotechnological potential usage as a therapeutic molecule in inflammatory diseases and to control schistosomiasis.

Methods: We used our algorithm enCrypted, to perform an *in silico* proteolysis of SmKI-1 and a screening for potential antimicrobial activity. The selected peptides were chemically synthesized, tested *in vitro* and evaluated by both structural (CD, NMR) and biophysical (ITC) studies to access their structure-function relationship.

Results: EnCrypted was capable of predicting AMPs in SmKI-1. Our biophysical analyses described a membrane-induced conformational change from random coil-to-α-helix and a peptide-membrane equilibrium for all schistocins. Our structural data allowed us to suggest a well-known mode of peptide-membrane interaction in which electrostatic attraction between the cationic peptides and anionic membranes results in the bilayer disordering. Moreover, the NMR H/D exchange data with the higher entropic contribution observed for the peptide-membrane interaction showed that schistocins have different orientations upon the membrane.

Conclusions: This work demonstrate the robustness for using the physicochemical features of predicted peptides in the identification of new bioactive cryptides. Besides, it demonstrates the relevance of combining these analyses with biophysical methods to understand the peptide-membrane affinity and improve further algorithms.

Abbreviations: 2D, two-dimensional; AMPs, antimicrobial peptides; APD, antimicrobial peptide database; ATCC, American type culture collection; CD, circular dichroism; CFU, colony forming units; D_2O, deuterium; DPC, dodecylphosphocholine; FID, free induction decay; FN, False Negative; FP, False Positive; HSQC, heteronuclear single quantum coherence; ITC, isothermal titration calorimetry; KD, equilibrium dissociation constant; LUV, large unilamellar vesicles; MALDI-TOF/MS, matrix-assisted laser desorption/ionization-time of flight/mass spectrometry; MCC, Matthews correlation coefficient; MIC, minimum inhibitory concentration; NCCLS, national committee for clinical and laboratory standards; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; POPG, 2-oleoyl-1-palmitoyl-sn-glycero-3-phospho-rac-(1-glycerol); RF, random forest; RP-HPLC, reverse phase-high performance liquid chromatography; SDS, sodium dodecyl sulphate; SVM, support vector machine; TFE, 2,2,2-trifluoroethanol; TN, True Negative; TOCSY, total correlation spectroscopy; TP, True Positive; ΔG, Change in Gibbs free energy; ΔH, enthalpy change; ΔS, entropic change.

* Corresponding author at: Avenida Antônio Carlos, 6627, Instituto de Ciências Biológicas, Dep. Bioquímica e Imunologia, Laboratório de Biofísica de Macromoléculas, Bloco O4, Sala 125, Belo Horizonte, Minas Gerais, Brazil.

E-mail address: mquezado@icb.ufmg.br (M.T.Q. de Magalhães).

[1] Both authors share senior authorship.

https://doi.org/10.1016/j.bbagen.2021.129989
Received 31 May 2021; Received in revised form 30 July 2021; Accepted 6 August 2021
Available online 10 August 2021

General significance: Bioprospecting cryptides can be conducted through data mining of protein databases demonstrating the success of our strategy. The peptides-based agents derived from SmKI-1 might have high impact for system-biology and biotechnology.

1. Introduction

Antimicrobial peptides (AMPs) are widely distributed in nature and have broad spectrum of activity (antibiotic, anti-inflammatory, immunomodulatory) [1]. Most known AMPs are cationic molecules, a property that culminates in great affinity with negatively charged membranes. Their interaction with membranes can be through: barrel-stave, in which peptides interaction creates pores in the membrane (eg. pardaxins) [2,3]; detergent-like, forming micelles with the membrane lipids (eg. magainins and MSI-78) [4,5]; carpet model, in which the membrane is disrupted but peptides do not enter the cell (eg. LL-37) [6]; or toroidal, with pores formation without peptides interaction [7–10]. The knowledge of those processes culminated in the study of thousands of antimicrobial peptides in the past few years and allowed researchers to classify them according to their physicochemical properties, such as secondary structural content, hydrophobicity, net charge, and structure upon interaction with a membrane [11,12].

Most of synthetic antimicrobial peptides are designed as cationic molecules with hydrophilic and hydrophobic interfaces to allow the interaction with the phospholipids heads and tails, assuring a good peptide-membrane interaction. However, this is not state of art, and the balance between different physicochemical properties need to be evaluated such as: toxicity to eukaryotic cells of extremely charged peptides [13,14]; difficulties in administration of hydrophobic peptides and mixing these features can also culminate in inefficient AMPs [15]. This reinforces the necessity of studying AMPs features and the relevance of peptide design. Another factor to take into account is the peptide shortening strategy, which aims to improve the peptide activity by only keeping in the sequence and removing residues that could hinder the peptide's best performance. Shortening can be accomplished by removing N-terminal or C-terminal residues and the approaches take in consideration the peptide amphipathicity, charge and other physicochemical features [16–18]. Residue substitution can also be the key in peptide design since identifying strategic regions for modification are fundamental in peptide improvement.

Cryptides describe peptide fragments encrypted within a larger protein, displaying dissimilar biological functions from the parental, originary primary structures [20]. According to this idea, proteins may have multiple functions and activities, depending on enzymatic processing, metabolic target or relative concentration. A few relevant examples of cryptides are the hemoglobin-derived peptides that display several biological activities distinct from those well-described to its precursor [20–23], the antimicrobial peptides (AMPs) from the apolipoprotein-B, r(*P*)ApoBL/r(*P*)ApoBs [24], the lactoferrin-derived ones [25], the cryptides found in the C-terminal domain hemocyanins, a multi-subunit oxygen carrier proteins found in the hemolymph of arthropods and molluscs [26–28], and those from the human thrombin, such as GKY25 [29]. These peptides are generated either after protein maturation and metabolic processes [30,31] or during *in vitro* assays [32,33]. Most of these peptides have been overlooked and maybe this occurs because they are only active after proteolytic action of enzymes, environmental pH change or can be converted inside large sets of proteomic data [34,35]. The diversity of encrypted biologically active peptides in known proteins can be exploited by several theoretical and computational models [36–38]. This might lead researchers to get the most out of the already known protein sequences, looking into their multifunctionality, and reshaping their role in nature, especially those with unknown domains. In addition, bioactive peptides with potential therapeutic applications (e.g inflammation, cancer) can also be discovered [39].

Several approaches applied to the peptide design field have already been proposed to retrieve cryptides from natural molecules, and, usually, they consider the primary structure properties and the three-dimensional features [35,40–42]. The physicochemical properties of peptides influence their behavior and affinity to their targets [43]. The structure factor includes not only the peptide folding (secondary structure) but also the residues disposition in the structure (amphipathicity), and both will be influenced by the ionic and hydrophobic interactions with the target (*e.g.*: biological membrane). Atomically, membrane disruption may occur through the activity of AMPs that reside in the polar-apolar interface, forcing the lipid groups to distance from each other [44]; through peptide-peptide oligomerization and changes in the electrostatic surface potential in the lipid bilayer [45]. Also, peptide orientation is a crucial factor in its ability to cause membrane disruption, as seen for the amyloidogenic peptides IAPP, where the rat analogue was more capable of burying deeper in the membrane, while the human analogue was oriented in the surface of the micelle [46]. In the peptide rational design, it is possible to input as many features as desired regarding what may be crucial for a specific biological activity in a peptide [47]. Nonetheless, consolidating and analyzing all the information is not straightforward since it is more than inputting too many features in peptide rational design, as selecting the properties is more relevant.

Some predictors for antimicrobial peptides including iAMPpred [48], AmPEP [49], ClassAMP [50], amPEPpy 1.0 [51], AntiMPmod [52], MAMPs-Pred [53], AntiBP [54] have been frequently reported. Nevertheless, these tools do not take into account the cryptide concept in their algorithms, which impairs the discovery of peptides hidden within a protein. Since the cryptide concept is innovative in exploring proteins and their corresponding domains without known function, we focused on the proteins expressed by the water-borne human parasite, *Schistosoma mansoni*. For the past few years, our research group has been studying the properties of the *S. mansoni* Kunitz Inhibitor SmKI-1 (uniprot CCD77156) [55,56], a protein with a single Kunitz-type domain, which is also a C-terminus domain with no similarity outside the genera, with a vaccine applicability against *S. cercariae* infection in murine model [57]. The presence of the SmKI-1 in the schistosomula tegument [57,58] lead us to hypothesize that its C-terminal domain has membrane affinity features.

In this work, we developed and validated an algorithm to predict cryptides in protein sequences – enCrypted. Our main goal was to use our tool to uncover the cryptides hidden in the SmKI-1 C-terminal domain and evaluate their potential biotechnological applications as antimicrobial peptides. Our approach retrieved 18 attributes and the use of machine-learning algorithms allowed us to describe a new class of antimicrobial cryptides: the schistocins. In order to validate our hypothesis, we tested their antimicrobial activity and performed biophysical characterization by using circular dichroism, calorimetry and nuclear magnetic resonance to access their structure-function relationship. Our data suggests that our computational and experimental framework was able to uncover new cryptides from an uncharacterized domain from SmKI-1 and it is effective to be applied to many other proteins domains of unknown function.

2. Material and methods

2.1. enCrypted: in silico prediction, selection, and downstream analyses of bioactive cryptides

enCrypted is a software that reads and identifies peptides from

FASTA protein files. A predictor model was built using curated sequences (3245 FASTA sequences) extracted from the APD database [59] and PepBank [60] (Date of Access: 20/06/2020). For the algorithm selection, we performed a grid-search to obtain a hyperparameter space for the best cross validation score. The machine learning classification was performed using scikit-learn with python codes [61]. The selected algorithms were linear Support Vector Machines (SVM), decision tree, and random forest (RF). The linear SVM creates an optimal hyperplane to separate data into classes [62]. The parameters tested were: regularization parameter ('C') [0.001, 0.01, 0.1, 1, 10] and loss function ('loss') [squared hinge, hinge]. The decision tree consists of a binary recursive partitioning classification method composed of nodes, edges and leaf nodes that split data into partitions and possibilities to test the different values [63]. The parameters tested were the criterion function ('criterion') [gini, entropy], the splitter strategy (splitter) [best, random], the maximum depth of the tree ('max_depth') [3,4,5,7,8,9], the number of features per split ('max_features') [3,5,8,9,10]. The RF has the same principle of decision trees, but with a large number, which one testing a class prediction [64]. The parameters tested were the number of trees ('n_estimators') [50, 100, 200], the criterion function ('criterion') ['gini','entropy'] and the maximum depth of the tree ('max_depth') [4,5,6,7,8]. The grid-search was performed 10 fold (Fig. S1). The metrics were evaluated using the Eqs. (1) and (2):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Matthews correlation coefficient} = \frac{TP \times TN - FP \times FN}{\sqrt{((TP+FP) \times (TN+FN) \times (TP+FN) \times (TN+FP))}} \quad (2)$$

The TP as true positive, the TN as true negative, the FP as false positive and the FN as the false negative. After selection, the method had its parameters optimized and exported to be used as the predictor model. The C-terminal SmKI-1 sequences were filtered by molecular range between 1400 and 2300 Da and a potential biological activity was predicted for each one of them. At the end, the algorithm generates a peptide list with two distinct groups: (a) antimicrobial activity and (b) non-antimicrobial activity.

In silico analyses for the selected sequences were also performed. The positive results obtained were compared with the original C-terminal domain of SmKI-1 sequence, and the visualization was provided by Jalview [65]. A multiple sequence alignment was performed with MUSCLE [66], using sequences from the APD database (similarity), previously selected using NCBI-Blast+ [67,68], and the visualization and pid (percent sequence identity) were provided by MView [69]. The amino acid residues disposition in potential α-helices were studied through helical wheels generated in the webserver Heliquest [70].

2.2. Peptide synthesis, purification and characterization

Four *S. mansoni* – derived antimicrobial peptides (cryptides) were designed and synthesized by solid-phase synthesis, purified by RP-HPLC, and the monoisotopic molecular mass and primary sequences were verified by mass spectrometry (Fig. S3 and Fig. S4).

The peptides schistocin-1 (GILDIKNKVSNLFKKIKGEK-NH_2), schistocin-2 (GILDIKNKVSNLFKKIK-NH_2), schistocin-3 (GILDIKNKVSNLF-NH_2), and schistocin-3.1 (GILDILNKVSNLF-NH_2) were obtained by solid-phase peptide synthesis *via* Fmoc (9-fluorenylmethoxycarbonyl) strategy [71]. Fmoc Rink-Amide® resin (0.63 mmol·g^{-1}) was used for the synthesis of all the peptides resulting in C-terminal amidated peptides. The amino acid derivatives were provided by Iris Biotech GmbH. Cleavage and final deprotection were conducted with TFA:Triisopropylsilane:Water (95.0:2.5:2.5, v:v:v) solution for 1.5 h at room temperature. Peptide purification was performed through RP-HPLC with a C18 column (Discovery® C18 HPLC Column 5 μm particle size, L × I.D. 25 cm × 10 mm) with a linear gradient of water containing 0.1% (v:v) TFA and acetonitrile containing 0.1% TFA (v:v). The peptide concentrations were determined according to the Murphy & Kies method, by measuring the absorbance of stock solutions at 205, 215, and 225 nm [56]. The identity of the peptides (0.001 mg·mL^{-1} in 0.1% TFA) were assessed by MALDI-TOF/MS (Bruker Daltonics Autoflex III) in the positive reflected mode, using α-cyano-4-hydroxycinnamic acid (5 mg·mL^{-1}, in 50% AcN, 0.1% TFA) as an ionization matrix in a proportion of 1:1 (v/v). Ions of interest were fragmented in the LIFT mode (MS/MS). The purity of the synthetic peptides was accessed by RP-C_{18}-HPLC (*not shown*).

2.3. Antibacterial assays

Certified American Type Culture Collection (ATCC; USA) strains of *E. coli* (ATCC 25922), *P. aeruginosa* (ATCC 27853) and *Staphylococcus aureus* (ATCC 29213) were provided by the National Institute of Quality Control in Health (INCQS-FIOCRUZ, Brazil), and were cultured at 37 °C on agar plate containing BHI medium (Brain Heart Infusion) from HiMedia (Mumbai, India) for 12 h. Subsequently, the isolated colonies were selected and transferred to a Mueller-Hinton culture medium (Kasvi; Brazil) according to the National Committee for Clinical and Laboratory Standards (NCCLS; USA) in order to perform the bioassays (NCCLS Institute protocols) [72]. The peptide was diluted 10-fold in Mueller-Hinton liquid broth. The highest peptide concentration used was 264 μM in a suspension composed of approximately 5 × 10^5 CFU. mL^{-1} (colony forming units·mL^{-1}), and the final volume was 100 μL per well 1:1 (peptide:inoculum). The experiments were performed for 20 h in stationary culture at 37 °C, and the optical density was measured at the beginning and at the end of the assay (A_{595} nm). Ampicillin and chloramphenicol (50 μM) were used as the standard antibiotics for the control of cell death, and the positive control was achieved by the addition of 50 μL of the inoculum and 50 μL of the sterile culture medium. All assays were performed in triplicate.

2.4. Antifungal assays

Antifungal activity were screened using *C. albicans* (ATCC 10231; resistant to anidulafungin, voriconazole, itraconazole, and fluconazole), *C. krusei* (ATCC 6258), *C. tropicalis* (ATCC 750), *C. parapsilosis* (ATCC 90018) and *Candida glabrata* (ATCC 2001) as provided by the National Institute of Quality Control in Health (INCQS-FIOCRUZ, Brazil). The minimum inhibitory concentration (MIC) was determined by a susceptibility test for microdilution in broth for the determination of MIC to yeasts (CSLI 2012) [73], with modifications. Tests were also performed in Sabouraud dextrose broth (Kasvi, Brazil). Serial dilutions of 20–500 μM were prepared from compounds on sterile 96-well microtitration plates. The density for yeasts per well corresponded to 1 × 10^3 to 5 × 10^3 CFU·mL^{-1}. The plates were incubated at 35 ± 2 °C for 44 h, and subsequently evaluated by visual reading. The positive control was nystatin (50 to 0.0225 μg·mL^{-1}). Controls were also prepared for viability assessment of the test microorganisms and sterility assessment of the culture medium. All assays were performed in triplicate.

2.5. Vesicle preparation

Large unilamellar vesicles (LUVs) were prepared using 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 2-oleoyl-1-palmitoyl-sn-glycero-3-phospho-rac-(1-glycerol) (POPG) lipids. The appropriate amounts of phospholipids required for the preparation of POPC or POPC:POPG (3:1, mol:mol) LUVs were solubilized in 1.0 mL of chloroform, which was evaporated using nitrogen gas inside the fume hood and the resulting lipid film was resuspended in an aqueous 50 mM Tris-HCl buffer (pH 7.5), containing 100 mM NaCl buffer. In order to form unilamellar vesicles, 8 cycles of freezing and warming the solution was performed: the tube was heated in a water bath (40 °C) and flash-frozen using liquid nitrogen. The vesicles size was regulated by extrusion procedure, where the lipid solution was pushed 11 times between a

mini-extruder and two syringes. The mini-extruder filter contained filters with 0.1 μm. The LUVs diameter was confirmed using the Dynamic Light Scattering analyzer/Malvern® Zetasizer nano ZS particles model BI-900 (Worcestershire, United Kingdom) from the Multi-user Laboratory of Research in Pharmacy (Department of Pharmacy – UFVJM). The hydrodynamic diameter analysis was accessed through the Malvern® model DTS1060 cuvette, using monochromatic light (Ne 4 mW laser, λ 633 nm) and was carried out in triplicates. LUVs were maintained at 4 °C and used within 24 h from preparation [74].

2.6. Circular dichroism (CD) spectroscopy

The schistocins 1, 2, 3 and 3.1 were evaluated for secondary structure contents by circular dichroism (CD) measurements using a JASCO® J-810 spectropolarimeter (Jasco Inc., Japan), equipped with Peltier temperature control system - PFD-425S (Jasco Inc., Japan). CD spectra were recorded at 20 °C using a 100 μm path length quartz cuvette (Uvonic Instruments, NY), with peptides (30 μM) in i) 50 mM Tris-HCl pH 7.0 and 100 mM NaCl, ii) TFE:water solutions, iii) presence of 100-500 μM sodium dodecyl sulphate (SDS) micelles, and iv) LUVs. All the spectra were recorded with three scans from 260 to 190 nm using a 1.0 nm spectral bandwidth, 200 nm.min^{-1} scan speed, and 1 s response time, and properly corrected for their respective blank measurements performed without peptides. Blank subtraction was performed with Spectra Analysis software (Jasco Inc., Japan). The resulting curves were smoothed without data distortion by using the Savitzky-Golay filter [74] with a polynomial order of 3 and a smoothing window of 51 points. A customized algorithm for deconvolution was used using the following Eq. (3):

$$\text{Ellipticity, in degrees}\cdot\text{cm}^2\cdot\text{dmol}^{-1} = 100 \times \theta\lambda/m \times d \quad (3)$$

The ellipticity calculation was proceeded as previously informed [76]. The K2D2 method was used for spectra deconvolution and average helicity value was extracted [77].

2.7. Isothermal titration calorimetry (ITC)

ITC experiments were performed at 35 °C with a high-precision VP®-ITC microcalorimeter (Malvern Instrument Ltd., UK) for 50 μM peptide solutions in 50 mM Tris-HCl, pH 8.5, containing 100 mM NaCl. The ITC equipment was electrically and chemically calibrated before performing the experiments [62]. All the solutions and buffers employed in the experiments were filtered and degassed under reduced pressure (140 mbar) for 5 min. The peptide solutions in the calorimeter cell were titrated with fifty-one successive injections of 20 mM POPC or POPC: POPG (3:1, mol:mol) LUVs (the first 1 μL injection was discarded in order to eliminate diffusion effects of the material from syringe to calorimeter cell, and this injection was followed by fifty injections of 5 μL). Injection times of 2 s with intervals of 240 s have been used in the experiment. The background experiments were performed titrating the buffer solution with the respective LUVs suspensions to determine the corresponding heats of dilution. In order to measure the heat produced by the peptide-to-lipid interactions, each peak of the calorimetric curve was integrated following standard procedures as described elsewhere [78]. The total lipid concentration was used to estimate the degree of membrane association as well as in the determination of the thermodynamic binding parameters. The raw data were analyzed with Microcal Origin 5.0 (OriginLab Corporation, MA).

2.8. Nuclear magnetic resonance spectroscopy (NMR)

The NMR measurements were performed at 1 mM peptide in a micellar solution containing 100 mM of deuterated dodecylphosphocholine (DPC-d_{38}), 10% D_2O, 1.0 mM 2,2-dimethyl-2-silapentane-5-sulfonate (DSS; internal reference) and 50 mM sodium phosphate pH 3.5. Two-dimensional (2D) NMR spectra of schistocin-1, 2, 3 and 3.1 were recorded on a Bruker AVANCE III 500 NMR spectrometer operating at 500.013 MHz equipped with a 5 mm broadband inverse (BBI) probe head (Chemistry Institute, Federal University of Goiás - UFG, Brazil) at 300 K, and pre-saturation for proper water suppression.

Total correlation spectroscopy (TOCSY) spectra were acquired using the DIPSI-2 pulse sequence [75,79]. The water supression was performed using the pulse sequence dipsi2esgpph (BRUKER nomenclature), using a spin-lock of 80 ms. The spectral width was 9615 Hz, and 512 t_1 increments were collected with 16 transients of 4096 points. Nuclear Overhauser Spectroscopy (NOESY) spectra were acquired using a pulse sequence with watergate suppression (noesygpphw5) and mixing times of 150 and 200 ms [80]. The spectral width was 9615 Hz, 512 t_1 increments were collected with 16 transients of 4096 points for each FID. ^{1}H-^{13}C Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired in edited mode with phase sensitive multiplicity-edited HSQC using echo-antiecho sequence (hsqcedetgp) with F1 and F2 spectral widths of 72,010 Hz and 9615 Hz, respectively [81]. A total of 128 t1 increments were collected with 128 transients of 1024 points. The HSQC experiments allowed the user to observe methyl groups in different ways: CH and CH_3 correlations show positive phase and CH_2 correlations show negative phase.

2.9. NMR data analysis and structure calculations

All the NMR data were processed using TopSpin® software 4.1 (academic license, Bruker, Germany). Two-dimensional ^{1}H resonance assignments were performed by simultaneously analyzing 2D ^{1}H-^{1}H TOCSY and NOESY spectra [82], using the CCPNMR analysis software [83]. NOE intensities obtained at mixing times of 200 ms were converted into semi-quantitative distance restraints [84], resulting in upper limits of distances of 2.8, 3.4 and 8.0 Å (for strong, medium, and weak NOEs, respectively). The dihedral angle (φ (phi) and ψ (psi)) restraints have been obtained from analysis of C_α, H_α, C_β and H_β chemical shifts with the DANGLE software, part of CCPNMR platform [85]. Structure calculations were performed using the Aria2.3 software, coupled with CNS [86,87]. Starting with an extended conformation, 200 structures were generated using a simulated annealing protocol, followed by 20,000 steps of simulated annealing at 1000 K and a subsequent decrease in temperature in 15,000 steps in the first slow-cool annealing stage. The twenty lowest energy structures were refined by using a sa_new.inp protocol and the stereochemical quality was analyzed by MolProbity [88], Procheck [89] and Prosa [90]. The display, analysis, potential hydrogen bonds counting and manipulation of the three-dimensional structures were performed with Chimera (UCSF, USA) [91]. The solvation potential energy calculation was measured by converting the PDB files into PQR, that served as input for APBD, a Poisson-Boltzmann Solver calculator [92,93].

2.10. Hydrogen-deuterium exchange

Hydrogen exchange data were recorded on a Bruker AVANCE III 500 NMR spectrometer operating at 500.013 MHz (Chemistry Institute, Federal University of Goiás – IQ/UFG, GO, Brazil). Hydrogen-deuterium exchange experiments were initiated by gently reconstituting dry lyophilized peptide (originally 1 mM) in dry DPC-d_{38} lipids (originally 100 mM) with 95% of D_2O sufficient to achieve their original concentration. A series of TOCSY spectra were acquired in intervals of 1.33 h to monitor the decay of the HN-H_α due to exchange for deuterium. All spectra were acquired with water suppression by using dipsi2esgpph pulse sequence with 4096 time domain data size, 512 scans, 16 dummy scans and 32 number of scans, with a spin-lock time of 80 ms. All data were processed using TopSpin 4.1 (Bruker) and analyzed using CCPNMR to assign and to follow the intensity changes. The data exchange kinetic was adjusted for a monoexponential decay using I = A.exp.(-B.t), where A and B are the fit parameters and t is time after complete reconstitution

with D_2O.

3. Results

3.1. Identification of new cryptides from SmKI-1 C-terminal domain, their solid-phase synthesis, and characterization

SmKI-1 is located in the schistosomula tegument bound to a plasma membrane and overlain by a secreted membranocalyx (a second membrane). Considering its affinity to membranes as desired feature for novel AMP, we attempted to identify potential bioactive cryptides from the SmKI-1 protein by using enCrypted (details in the *Material and Methods* section). A set of theoretical, potential AMP (183 peptides) and non-AMP (265 peptides) were obtained (Fig. 1A), with the AMP typically starting at the position Ser-21 and ending at Leu-69 (numbers corresponds to full SmKI-1 C-terminal domain amino acid sequence).

In order to select potential new antimicrobial cryptides from SmKI-1, and study their biophysical properties, we selected one out of the three peptides presenting similarity with the greatest number of sequences of another known antimicrobial peptides (Supporting CSV File). Firstly, we chose the theoretical AMP 20-mer peptide found by the C-terminal of SmKI-1, hereby named schistocin-1. Secondly, we used the minimization strategy to shorten the peptide and evaluate its effects caused by changing important features such as charge, molecular weight and hydrophobicity, resulting in the theoretical peptides schistocin-2 (17 mer), and schistocin-3 (13 mer). The net charge in the peptides schistocin-1 and schistocin-2 is the same (+5), but schistocin-3 presents a lower value (+2) (Fig. 1B). Despite the lowest charge, schistocin-3 is the most hydrophobic of the three peptides (0.463) and has similar hydrophobic moment with schistocin-2 (0.551 and 0.579, respectively).

The schistocins exhibited primary structure similarity to previously deposited antimicrobial peptides in the APD database, which were originally identified in amphibians (percentage of identity: 26.3–31.6% uperin, 7.9–10.5% brevinin, 10.8–16.7% ranateurin), and in scorpions (9.1% hadrurin) (Fig. 1C). A helical wheel plot of schistocin-1, schistocin-2, schistocin-3 were obtained to confirm and compare the cryptides amphiphilic nature when in an α-helical configuration (Fig. S2). After this analysis, we increased the α-helix amphipathicity (0.701) by replacing Lys6 > Leu in schistocin-3, leading to schistocin-3.1. The quality of the predictor for this analysis can be evaluated through its accuracy value of 87% and its efficacy compared with other predictors reported elsewhere (Table S1). The physical-chemical features used in the model construction and full peptides selection are listed in Table S2.

3.2. Schistocins exhibit different antimicrobial activities against human pathogenic bacteria and yeasts

The schistocins peptides were evaluated for antimicrobial activity against both gram-positive and gram-negative bacterial (*S. aureus* ATCC 29213, *E. coli* ATCC 25922, *P. aeruginosa* ATCC 7853) and yeast strains (*C. albicans* ATCC 10231, *C. krusei* ATCC 6258, *C. tropicalis* ATCC 750, *C. parapsilosis* ATCC 90018, *C. glabrata* ATCC 2001). Schistocin-3 and schistocin-3.1 displayed a similar antibacterial activity when compared to the controls (ampicillin and chloramphenicol) whereas higher MIC values were detected in cryptides schistocin-1 (>130 μM) and schistocin-2 (>130 μM) against the tested bacterial strains (Table 1). A similar behavior was observed in antifungal assays, except for *C. krusei*, to which all the peptides were ineffective. Overall, schistocin-3 and schistocin-3.1 were more effective either against bacteria or fungi than schistocin-1 and schistocin-2.

3.3. Schistocins conformation depends on environmental hydrophobicity

In order to evaluate whether the resulting enCrypted output provided membrane-active peptides capable of folding into an ordered α-helix, we performed circular dichroism measurements in the presence of membrane-mimetic solvents and phospholipid vesicles. The four schistocins adopted a random coil conformation when free in solution, and transitioned into α-helix-rich structures when in the presence of *zwitterionic* (a.k.a. amphoteric) or anionic POPC:POPG (3:1) vesicles (Table 2 and Fig. 2). In POPC:POPG vesicles the helix content of schistocin-1, schistocin-2 and schistocin-3 are high (84%), even at mid concentration, while schistocin-3.1 exhibited lower helix content (50% in P/L = 1/8, 76% in P/L = 1/16) (Fig. 2). In addition, in TFE:water mixtures and in the presence of SDS micelles, all the schistocins have a characteristic α-helix pattern, with two minimums, in 208 nm and 222 nm (Fig. S5). Altogether, these findings demonstrate the self-assembly of schistocins into helix-rich structure upon recognition and binding into hydrophobic environment.

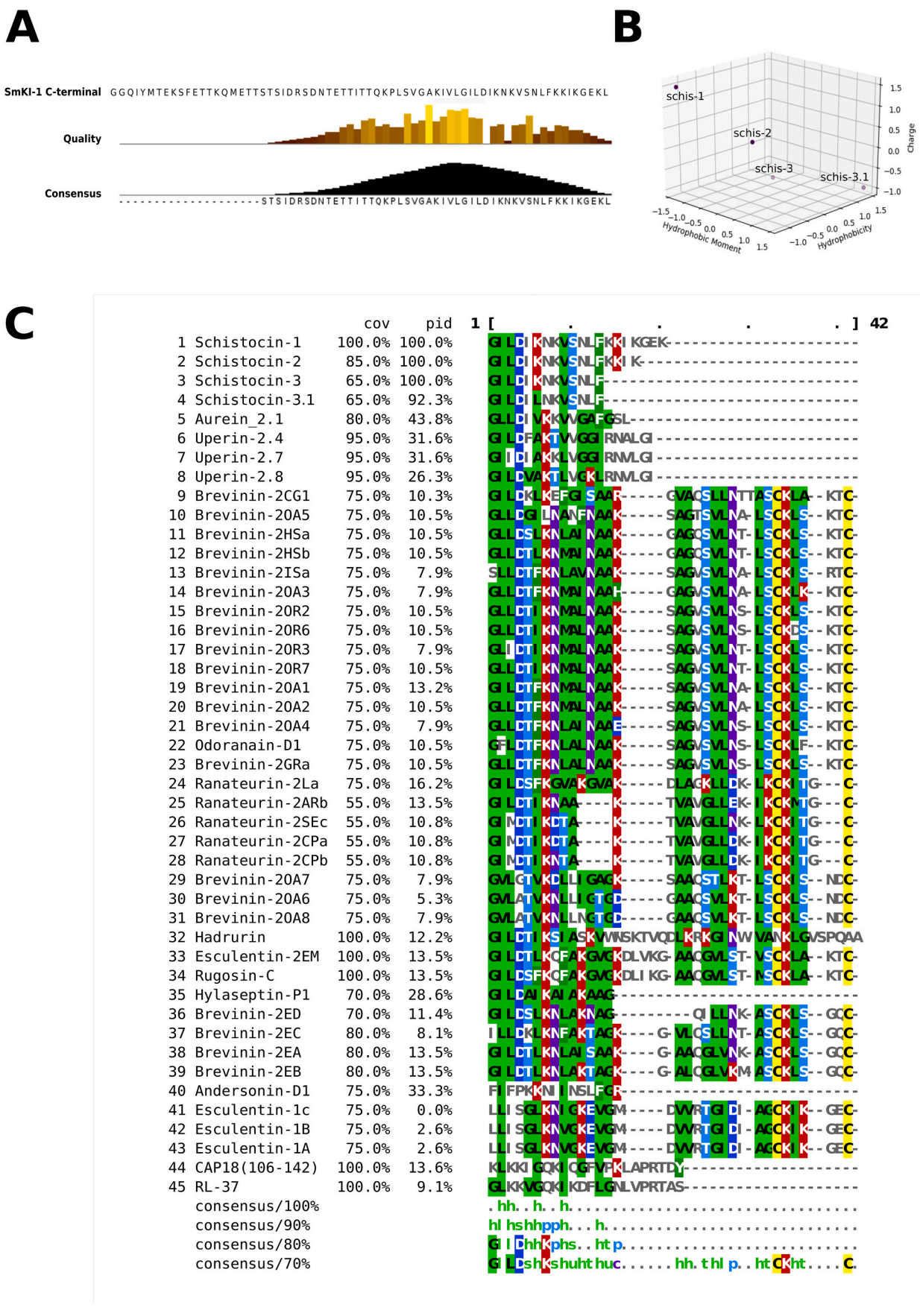
3.4. Interaction and binding affinity of schistocins are favored by negative enthalpy changes

In order to access the coupled energetics of schistocins binding to membrane and peptide folding into helix we conducted isothermal titration calorimetry (ITC). The overall interactions of schistocins with anionic vesicles were observed to be exothermic in nature as suggested by the integrated heat values (Fig. 3). Decreases in the heat flow until values near to the dilution are due to the smaller amount of free peptide in solution at the end of the experiment, which indicates the affinity of schistocins to the POPC:POPG LUVs (Table 3). Exergonic processes take place for the interaction of the peptide with the vesicles, and the highest change in Gibbs free energy was observed for schistocin-3.1 (-6.2 kcal. mol^{-1}). Consequently, greater binding constant (k_{app}) is observed for schistocin-3-1 when compared with others schistocins. Although, the negative enthalpy values indicate exothermic binding of the peptides to the anionic vesicles and the predominance of attractive coulomb forces, the binding process for all peptides are mainly driven by entropic factors, since $|\Delta H| << |T\Delta S|$. Therefore, these data suggest a well-known mode of peptide-membrane interaction in which electrostatic attraction between the cationic peptides and anionic membranes results in the bilayer disordering.

3.5. NMR solution structures of schisctocins revealed peptide interaction and orientation upon the mimetic-membrane model

CD spectroscopy and ITC analyses jointly described a membrane-induced conformational change from random coil-to-α-helix and a peptide-membrane equilibrium for schistocins. In order to gain more detailed insights on the structural basis of schistocin interaction with membranes, solution NMR spectra were recorded in the presence of dodecylphosphocholine (DPC-d_{38}) micelles. As expected, several strong sequential connectivity $d_{NN}(i,i+1)$ and medium range $d_{\alpha N}(i,i+3)$ correlations characteristic of α-helix were observed in all the schistocins NOESY spectra. All inter-residual NOE correlations are graphically summarized in Fig. S6.

The NOEs distance constraints in both schistocin-1 and schistocin-2 showed short-range constraints and medium-range constraints comprising residues Ile-2 to Lys-14 for schistocin-1 and Ile-2 to Ile-16 for schistocin-2 (Fig. S6). The calculated structures show an α-helix in these segments with a random-coil tail in the C-terminal. The ensemble of the lowest-energy structures shows relatively small all residues RMSD value for both schistocin-1 (0.32 ± 0.11) and schistocin-2 (0.10 ± 0.02) (Table 4). The residues Lys-15/Glu-19 and Glu-19/Lys-20 had hydrogen bond interactions responsible to stabilize the C-terminus of schistocin-1, while the Phe-13/Ile-16, Phe-13/Lys-17, Lys-14/Lys-17 hydrogen bonds stabilized the C-terminus of schistocin-2. Schistocin-1 and schistocin-2 have an apolar cluster with aliphatic residues Gly-1, Ile-2, Leu-3, Ile-5, Val-9, Leu-12, Phe-13, Ile-16 (Fig. 4A and Fig. 4B). Furthermore, the electrostatic potential on the surface of the peptides generated a solvation potential energy of 1.16×10^4 $kJ{\cdot}mL^{-1}$ for schistocin-1 and $1.33.10^4$ $kJ{\cdot}mL^{-1}$ for schistocin-2, demonstrating that they are highly



(caption on next page)

Fig. 1. Peptides Identification. Antibacterial prediction was performed using *in house* enCrypted software. Cryptides were generated through simulation of PAM enzyme activity and limiting them between 1400 Da and 2300 Da. A) Alignment of predicted peptides with the C-terminal sequence of SmKI-1. The analysis shows a preference for a fixed region of C-terminal SmKI-1, between residues 22 and 60. B) Physicochemical distribution of schistocins, taking into account hydrophobic moment, hydrophobicity and charge. C) Sequence alignment with AMPs sequences. A total of forty one sequences, in a rank of similarity, from the APD database were compared with schistocins peptides. The MUSCLE tool was used for multiple sequence alignment. The coloured residues are those with a consensus threshold of 70% and identical to the reference sequence, the schistocin-1. Colors: A - bright-green, C - yellow, D - bright-blue, E - bright-blue, F - dark-green, G - bright-green, H - dark-green, I - bright-green, K - bright-red, L - bright-green, M - bright-green, N - purple, P - bright-green, Q - Purple, R - bright-red, S - dull-blue, T - dull-blue, V - bright-green, W - dark-green, Y - dark-green. Color schemes from MView.

Table 1
Antibacterial and antifungal activity of schistocins using determination of the Minimum Inhibitory Concentration (MIC). Peptides activity was tested against three bacterial strains and 5 fungi strains. Nystatin, DS01, Ampicillin and Chloramphenicol were used as control drugs.

Microorganisms	MIC(μM)[a]						
	schistocin-1	schistocin-2	schistocin-3	schistocin-3.1	Nystatin	Amp.[c]	Chloram.[c]
S. aureus ATCC 29213	>130	>130	31.2	8	NT[b]	<11	>30
E. coli ATCC 25922	>130	>130	62.5	62.5	NT[b]	46	25
P. aeruginosa ATCC 27853	>130	>130	62.5	31.2	NT[b]	20	30
C. albicans ATCC 10231	250	250	62.5	31.2	0.39	NT[b]	NT[b]
C. krusei ATCC 6258	125	62.5	62.5	62.5	1.56	NT[b]	NT[b]
C. tropicalis ATCC 750	500	500	31.2	31.2	0.78	NT[b]	NT[b]
C. parapsilosis ATCC 90018	>500	>500	125	125	0.78	NT[b]	NT[b]
C. glabrata ATCC 2001	>500	>500	250	125	0.09	NT[b]	NT[b]

[a] MIC - Minimal agent concentrations required for total inhibition of cell growth.
[b] NT- Antimicrobial activity not tested.
[c] Amp and Chloram - ampicilin and chloramphenicol.

Table 2
Helix percentage of schistocins structures from circular dichroism spectra. Data were fitted using K2D2 online software.

Helix content (%)							
SDS (mM)							
	0	100	200	300	400	500	
schistocin-1	11.33	69.46	69.46	75.57	69.46	69.46	
schistocin-2	8.6	84.27	84.27	84.27	84.27	84.27	
schistocin-3	6.27	47.81	47.93	47.81	47.93	36.77	
schistocin-3.1	56.82	75.57	80.68	69.46	80.68	69.46	
TFE (%)							
	0	10	20	30	40	50	60
schistocin-1	11.33	20.49	69.46	80.68	80.68	69.46	75.57
schistocin-2	8.6	50.16	84.27	84.27	84.27	84.27	84.27
schistocin-3	6.27	10.33	12.98	47.81	47.93	47.81	47.81
schistocin-3.1	56.82	27.25	84.27	84.27	84.27	84.27	84.27
POPC (mM)							
	0			250		500	
schistocin-1	7.96			8.6		75.57	
schistocin-2	8.6			56.82		75.57	
schistocin-3	6.27			7.84		11.27	
schistocin-3.1	6.71			7.96		7.96	
POPC:POPG(3:1) (mM)							
	0			250		500	
schistocin-1	6.27			84.27		84.27	
schistocin-2	8.6			84.27		84.27	
schistocin-3	7.96			84.27		84.27	
schistocin-3.1	6.71			50.16		75.57	

cationic.

Schistocin-1 has a partially amphipathic conformation, impaired by Lys-6 and Gly-18. In addition, the hydrophilic side chains residues were less rigid than the hydrophobic residues (Fig. 4A). On the other hand, for schistocin-3 and schistocin-3.1 (13-mer peptides), the NOEs distance restraints evidenced sequential and medium-range constraints in virtually all backbone from Gly-1 up to Phe-13 (Fig. S6). A high stereochemical quality of the most stable structures obtained for all the peptides was obtained as depicted by the validation analysis (Table 4). The calculated structures show a full α-helix structure, compatible with the previous solution results by CD. The ensemble of the 10 lowest energy structure show that the N-terminal portion (2-12 amino acids residues) has relatively low conformational flexibility, including the side chains for both peptides (Fig. 4C and Fig. 4D) with RMSD values of 0.16 ±0.05 schistocin-3 and 0.11 ±0.05 for schistocin-3.1. As for schistocin-3.1, the substitution of Lys6 > Leu improved the helix amphipathicity. Taken together, these results led us to suggest that the schistocins display a well-defined, stable α-helix conformation.

Accordingly, NMR H/D exchange experiments allowed us to evaluate the intensities of HN-Hα correlation and fit exponential decays to determine the exchange rates of amidic hydrogen from the residues with large decay (Fig. S7). In all schistocins investigated here, most residues exchange occurred at a very fast rate (1.33 h). Nonetheless, in schistocin-1 and schistocin-2, the HN-Hα correlations involving Val-9, Ile-12, Phe-13, and Ile-16 were in slow exchange, suggesting that they are located in the micelle interface with low water accessibility. The Lys-6 presented in schistocin-1, schistocin-2, and schistocin-3 rapidly exchanged with deuterium, while in schistocin-3.1 the Leu-6 was in slow exchange indicating a deeper insertion of the N-terminus of this peptide into the micelles. This data together with the higher entropic contribution observed for the peptide-membrane interaction led us to propose a model of peptide insertion into the membrane interface (Fig. 5), suggesting that schistocin-1 and schistocin-3.1 appeared to attach almost in parallel to the membrane surface, while schistocin-2 and schistocin-3 present a slight tilted angle in relation to the membrane surface.

4. Discussion

In this paper, we developed and validated a machine learning algorithm (a random forest model) for the prospection of novel peptides encrypted within protein sequences. From the results obtained with enCrypted, we further designed, produced and evaluated the antimicrobial and biophysical properties of this new family of cryptides in the *Schistosoma mansoni* Kunitz Inhibitor protein, SmKI-1, hereby named schistocins. Our present data demonstrate the robustness for using the physical-chemical features of predicted peptides in the identification of new bioactive peptides encrypted within larger proteins.

Experimental and computational approaches in discovering cryptides have been previously described, and some studies have shown that cryptides can be found by top-down proteomics in tissues followed by bioactivity assays [35], such as the discovery of hemorphin-like peptides

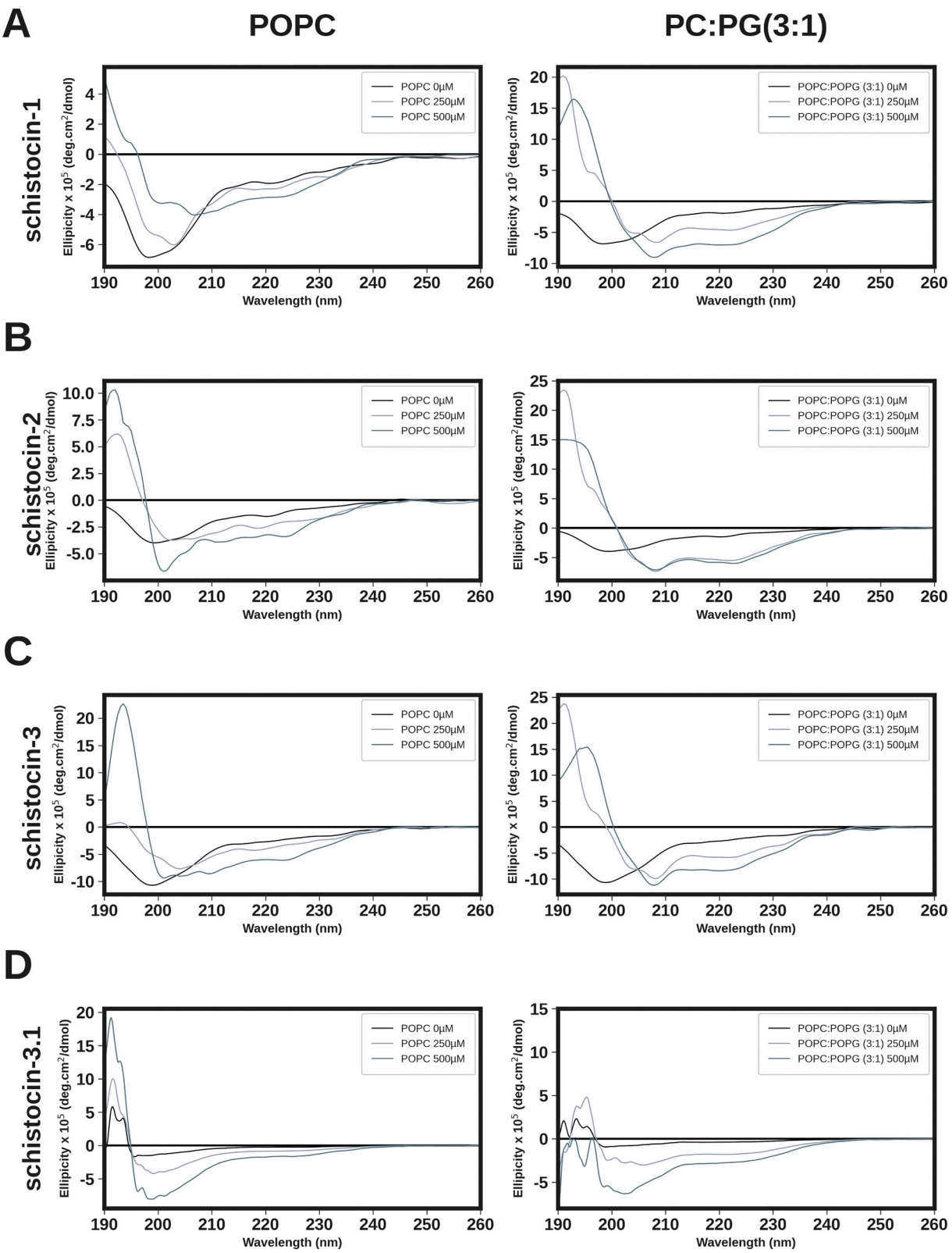


Fig. 2. CD spectra of 30 μM of schistocins in POPC and POPC:POPG (3:1) vesicles (A-H). Left column: POPC vesicles (P/L = 1/8, P/L = 1/16); right column: POPC:POPG (3:1) vesicles (P/L = 1/8, P/L = 1/16). A) schistocin-1. B) schistocin-2. C) schistocin-3. D) schistocin-3.1. All spectra were recorded in 50 mM Tris-HCl pH 7.0 and 100 mM NaCl. Spectra recorded from 260 nm until 190 nm. The black line represents peptides without vesicles, the gray line represents peptides in P/L = 1/8 peptide, and the blue line represents peptides in P/L = 1/16. Negative bands at 222 nm and 208 nm are characteristic of alpha-helix peptides, as seen in peptide-POPC:POPG vesicles; negative bands lower at 200 nm and lower are characteristic of random coil peptides, as seen in peptides-POPC vesicles.

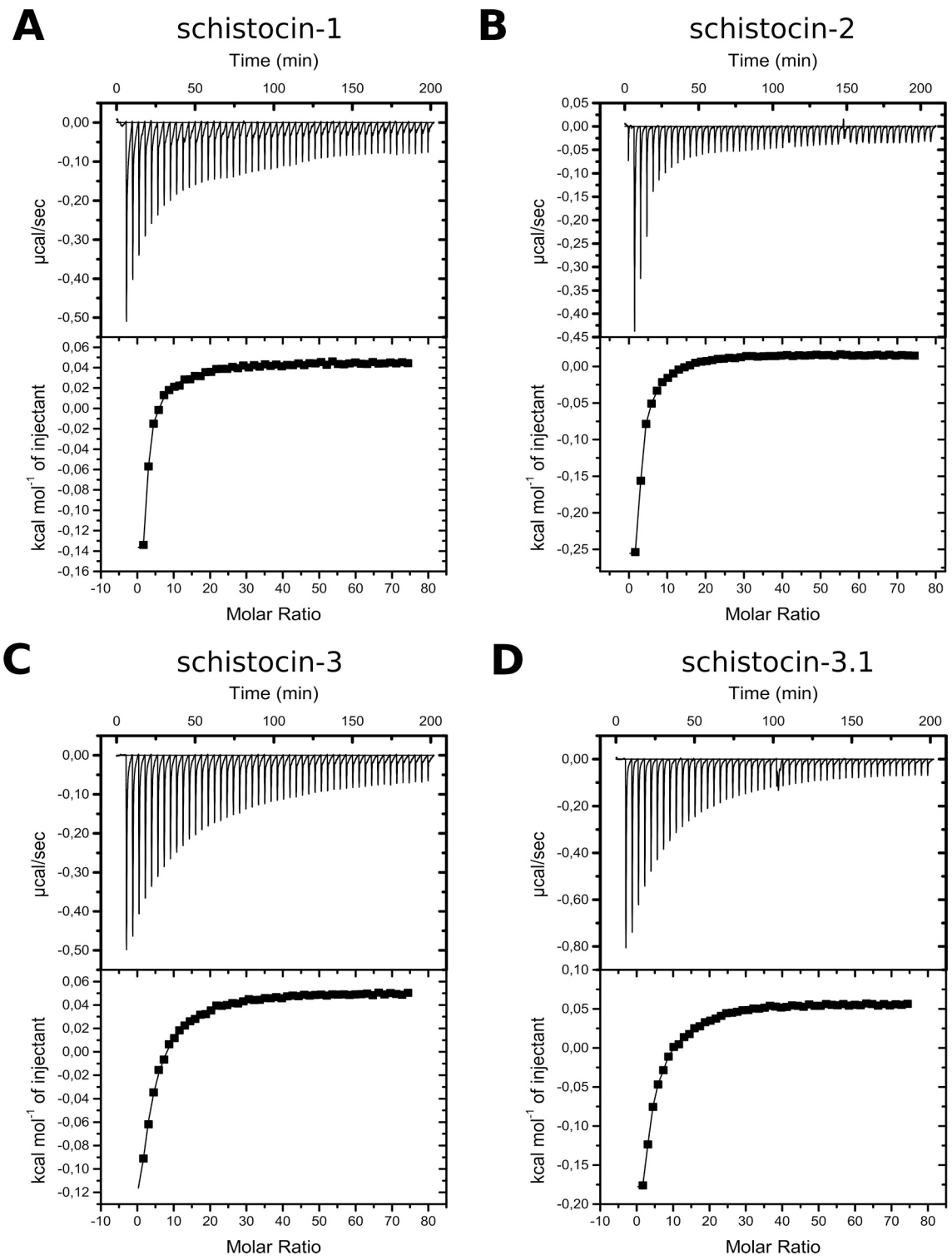


Fig. 3. POPC:POPG (3:1) vesicles - schistocins isothermal titration calorimetry. Titration calorimetry of schistocins solutions with unilamellar POPC:POPG (3:1) vesicles in buffer (35 °C; 50 mM Tris-HCl, 100 mM NaCl, pH 8.5). A total of 5 µL of the lipid dispersion (concentration = 20 mM) was injected into the calorimeter cell containing the schistocin-1 (A) or schistocin-2 (B) or schistocin-3 (C) or schistocin-3.1 (D) (concentration = 50 µM) in the same buffer. The upper image of each letter corresponds to the raw data. The lower image corresponds to the integrated heat change per injection with the best fitting curve.

[94], from *P. patens* cryptides [95], and human salivary peptides [96]. Furthermore, several research groups have been developing computational methodologies for the identification of cryptides. For instance, mitocryptides were identified in mitochondrial proteins obtained from the *Swiss-Prot* database, which followed by proteolytic cleavage *in silico*, and the selection of peptides with physical-chemical properties similar to *mastoparans* [97], such as identification of positive charges and the presence of specific residues [19,30,31]. Another example is the *Kamal* software proposed by Brand et al. (2012), which identifies physicochemical properties and resumes them as principal component analysis, which are used for bioactive peptides clusterization and allows identifying new antimicrobial peptides located in similar regions. The

Table 3
Thermodynamics parameters of ITC experiment of schistocin peptides in anionic LUVs.

Peptides	K	DH Kcal/mol	TDS Kcal.mol^{-1}	DG
schistocin-1	4.06E+03	-1.03	4.07	-5.09
schistocin-2	9.34E+03	-0.40	5.17	-5.57
schistocin-3	2.38E+03	-1.14	3.60	-4.74
schistocin-3.1	2.55E+04	-0.11	6.13	-6.23

aforementioned software included selection of whole-genome information of a large range of organisms, along with selected AMPs as a positive control [36–38]. The software *PeptideLocator* has the general goal of identifying bioactive peptides in protein sequences, with the bioactive peptide prediction being conducted by algorithms that identify residue positions and structural probability (helix, beta or turn) [98–100]. The *FeptideDB* has a similar framework but with different purpose: it identifies bioactive peptides of food proteins in input sequences, but limits for peptides already present in their database, having no prediction tool [101,102]. The present enCrypted algorithm has a pipeline for protein sequence screening, followed by sequence proteolysis and physical-chemical properties extraction, culminating in a list of peptides, and uses a supervised classification machine-learning algorithm (RF) for antibacterial prediction. The software concept proposes to join tasks that generally are separately performed: (i) *in silico* digestion and (ii) physical-chemical and bioactivity prediction; and opens up a field to expand the bioactivity prediction: antifungal, antiparasitic, immunostimulatory and others, provide that the current databases are curated and available. In addition, we propose a new approach to exploring unknown proteins by taking a closer look at encrypted information.

We hypothesised that SmKI-1 has to be cleaved from the tegument to exhibit an anti-inflammatory activity [57,58], and its other domain could be digested by degradation or proteolysis of the proteins resulting in the release of bioactive fragments in the host circulation. We selected three peptides from a set of 351 and, after a secondary pattern *in silico* analysis, we designed the fourth (schistocin-3.1) with a single substitution to improve amphipathicity (Fig. S2). These peptides were chemically synthesized and *in vitro* tested against bacterial and fungal strains that are pathogenic to humans, and the MICs were comparable to other antimicrobial peptides [103–105]. The peptides schistocin-3 and schistocin-3.1 displayed a wide antimicrobial spectrum, while schistocin-1 and schistocin-2 showed mild activity against *C. krusei* strains. Comparing our findings to other antimicrobial cryptides with known high activity, we can observe that schistocin-3 and schistocin-3.1 are nearly as efficient as apoB cryptides (from apolipoprotein), which have antimicrobial/antibiofilm activity with MIC values of 20 μM [24]. They are also similar to rpS30-2, from p62 eukaryotic intracellular protein, with MIC value of 100 μM against antimycobacterial activity [106]. In order to gain insights to explain the biological activity divergence among schistocins, the peptides-membrane interactions were investigated by a set of biophysical techniques [107–109]. Biomimetic membranes and micelles are useful to understand the peptides structuring in environments similar to bilayers found in both prokaryotes and eukaryotes, besides simulating the interfacial electrostatics responsible for the peptides' affinity and the analyses of different peptide-membrane stages [110,111].

As many other cationic peptides with antimicrobial activity [105,112–114] when in the presence of membrane mimetic models, schistocins display an α-helical fold (Fig. 2). These amphipathic peptides with net positive charge adopt α-helical structure upon contact with hydrophilic/hydrophobic surface, such as the membrane of microorganisms [115]. Examples are the GKY25, an AMP derived from the human thrombin protein [115], and the 23-mer PvHCt peptide from the shrimp *L. vannamei* hemocyanin [25]. The CD results describe schistocins random-to-helix transition from aqueous buffer to (3:1) POPC: POPG LUVs (Fig. 2). The peptides schistocin-1 and schistocin-2 presented weak helix-forming intensity with POPC and moderate helicity with POPC:POPG. Conversely, schistocin-3 and schistocin-3.1 upon membrane titration showed significant peptide-phospholipid helicity, highlighted in anionic vesicles (Fig. 2). In general, thermodynamic studies regarding the membrane interactions of schistocins indicated that these peptide presents from moderated to higher affinity to negatively charged vesicles only ranging from 35 °C and that result in exothermic process. The schistocins interaction at 25 °C did not show any significant binding and no thermodynamic parameters could be obtained (data not shown). The temperature-dependence of peptide-membrane interaction suggest the entropic component is predominant, as a consequence of the hydrophobic effect caused by water desolvation, which occurs during the peptide-membrane interaction. In fact, the major driving forces of interaction between schistocins and phospholipid membranes are the entropic contributions [117]. Interestingly, schistocin-3.1 interaction reveals the highest entropic factor and binding constant, which is in accordance with NMR data that shows better partitioning of the hydrophylic and hydrophobic residues withing the bilayer interface. On the other hand, a higher enthalpic contribution observed in the schistocin-1, schistocin-2 and schistocin-3 interaction suggests that the electrostatic attraction may still be an important phenomenon to the membrane binding process of these peptide [118]. Even though the schistocin-3.1 has net charge +1, the schistocin-1, 2 and 3 presents higher cationic hydrophilic surfaces to establish attractive coulomb forces with the negatively charged surface of the POPC:POPG membranes [119,120]. Thus, the higher antimicrobial activity of schis-3.1 could be related to the higher membrane affinity due an optimal insertion of the hydrophobic residues, pointing to membrane insertions of N-terminus anchored by Leu-6 residue. Recent studies showed that peptides derived from the C-terminal of SARS CoV E protein fold into an α-helix, which is correlated to the precursor protein host specificity and its inherent function to bind to the membrane [121]. In a similar way, the α-helical structure observed for schisctocins when upon the membrane may also be related to the potential C-terminal domain adhering function of the kunitz domain in the *S. mansoni* tegument.

Our NMR data confirmed that the peptides fold into well-defined α-helix three-dimensional structures in DPC-d_{38}, and corroborated the hypothesis that the Lys-6 was impairing the peptide amphipathicity, besides giving information on peptide flexibility in micelles (Fig. S5). The peptides schistocin-1 and schistocin-2 have a more flexible C-terminal, while schistocin-3 and schistocin-3.1 have a rigid helix, with a Phe-13 C-terminal side chain in slow exchange, showing its affinity to the membrane and increasing the specificity of these peptides. Moreover, hydrophobic residues are known to be important in peptide stabilization upon the membrane [122,123]. Future studies should consider taking a deeper look on the schistocins orientation in DPC-d_{38} micelles using paramagnetic probes in solution NMR. A previous study showed that gadolinium is inert with proteins and DPC-d_{38} micelles. Its addition to the NMR tube results in paramagnetic relaxation enhancements that make it possible to inform the orientation of α-helical structures [124,125].

Our dataset was constructed with peptides of up to 20 amino acid residues, provided that the peptides used as negative controls were obtained from a database that limited their peptides size. Restricting the AMPs sizes was essential to not construct an imbalanced dataset, leading to prediction errors [126]. An alternative approach could be the generation of peptides from Uniprot databases with a large molecular weight range and use them as the negative control. Still, an extensive curation would be necessary to distinguish the activities and to not impair the bioactivity prediction. Furthermore, our data suggests, as shown in Table 1, that two of the selected peptides did not exhibit strong antimicrobial activity against the studied bacterial and fungal strains. Nonetheless, this is not a drawback because the sequences of schistocin-1 and schistocin-2 are precursors of schistocin-3, which presented antimicrobial activity and helped us to understand which structural differences were responsible for the activity divergence. If taken as an

Table 4
Summary of structural statistics from NOE experiments.

Summary of Structural Statistics																
	schistocin-1				schistocin-2				schistocin-3				schistocin-3.1			
Distance and dihedral restraints																
NOEs																
Intraresidue	101				101				97				100			
Interesidue	71				70				48				76			
Sequential	39				44				25				44			
Short range	28				25				21				28			
Medium range	4				1				2				4			
Long range	0				0				0				0			
Total unambiguous	172				171				145				176			
Ambiguous	45				65				14				29			
Dihedral angles (psi + phi)	34				28				20				20			
Structural statistics																
Structural precision																
RMSD Backbone (all/2ndary structure)	0.32 (+/- 0.12) / 0.75 (+/- 0.30)				0.10 (+/- 0.02) / 0.24 (+/- 0.05)				0.16 (+/- 0.05) / 0.37 (+/- 0.05)				0.11 (+/- 0.05) / 0.45 (+/- 0.19)			
RMSD Heavyatom (all/2ndary structure)	0.90 (+/- 0.11) / 1.31 (+/- 0.38)				0.40 (+/- 0.03) / 0.47 (+/- 0.03)				0.60 (+/- 0.09) / 0.63 (+/- 0.11)				0.68 (+/- 0.09) / 0.93 (+/- 0.28)			
	value	error	min	max	value	error	min	max	value	error	min	max	value	error	min	max
PROCHECK:																
most favored regions	99.4	1.82	94.1	100	100	0	100	100	98.3	3.41	91.7	100	95	5.67	83.3	100
allowed regions	0.59	1.82	0	5.9	0	0	0	0	1.66	3.41	0	8.3	4.99	5.67	0	16.7
generously allowed regions	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
disallowed regions	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
labelled residues (all Ramachandrans)	0	0	0	0	0	0	0	0	0.1	0.308	0	1	0.25	0.55	0	2
labelled residues (Chi1-chi2 plots)	1.2	0.523	1	3	0.65	0.489	0	1	0.15	0.366	0	1	0	0	0	0
bad contacts	0.05	0.224	0	1	0	0	0	0	0	0	0	0	0	0	0	0
G-factor dihedrals	-0.145	0.578	-0.24	0	-0.124	0.055	-0.2	-0.01	0.131	0.0759	-0.01	0.27	0.154	0.123	-0.12	0.31
G-factor covalent	0.609	0.671	0.59	0.62	0.528	0.048	0.47	0.6	0.604	0.0131	0.58	0.63	0.641	0.008	0.62	0.65
G-factor overall	0.135	0.352	0.08	0.22	0.118	0.044	0.06	0.2	0.307	0.0456	0.22	0.39	0.339	0.073	0.17	0.43
WHATCHECK:																
Ramachandran plot appearance *Z*-score (RAMCHK)	-6.63	1.35	-9.73	-4.69	-4.27	1.28	-6.32	-2.39	-1.84	1.18	-3.66	0.751	-1.12	2.23	-4.62	1.76
Chi-1 chi-2 rotamer normality *Z*-score (C12CHK)	-7.28	1.21	-9.88	-4.92	-9.7	1.09	-11.4	-7.65	-8.08	0.978	-9.54	-5.84	-7.19	1.7	-10.1	-4.79
Backbone conformation *Z*-score (BBCCHK)	-1.84	0.597	-3.62	-1.38	0.382	0.0789	0.288	0.445	0.305	0	0.305	0.305	0.303	0	0.303	0.303
Bond lengths RMS Z-score (BNDCHK)	0.3	0.128	0.278	0.325	0.382	0.0272	0.323	0.419	0.312	0.0137	0.284	0.335	0.282	0.011	0.261	0.307
Bond angles RMS *Z*-score (ANGCHK)	0.395	0.168	0.367	0.43	0.492	0.0757	0.377	0.578	0.325	0.02	0.302	0.361	0.271	0.034	0.261	0.383
Omega angle restraints RMS Z-score (OMECHK)	0.734	0.707	0.606	0.877	0.648	0.0601	0.563	0.733	0.455	0.0548	0.329	0.549	0.447	0.057	0.323	0.549
Side chain planarity RMS *Z*-score (PLNCHK)	0.301	0.82	0.18	0.536	0.306	0.12	0.103	0.566	0.384	0.125	0.117	0.56	0.248	0.111	0.075	0.52
Improper dihedral distribution RMS Z-score (HNDCHK)	0.51	0.194	0.471	0.549	0.499	0.0492	0.418	0.579	0.468	0.0245	0.423	0.505	0.458	0.034	0.41	0.562
MOLPROBITY:																
Clashscore	17.7	4.6	11.6	29	43.1	9.23	26.4	59.2	18	3.47	13.7	22.8	8.8	5.6	0	18.5
Clashscore percentile	41.4	14.3	16	64	7.5	5.38	2	19	39.4	11.6	25	55	75.3	22.1	36	100
Clashscore Z-score	-0.195	0.365	-0.947	0.365	-1.48	0.336	-1.97	-0.809	-0.242	0.28	-0.605	0.125	nan	nan	-0.306	inf

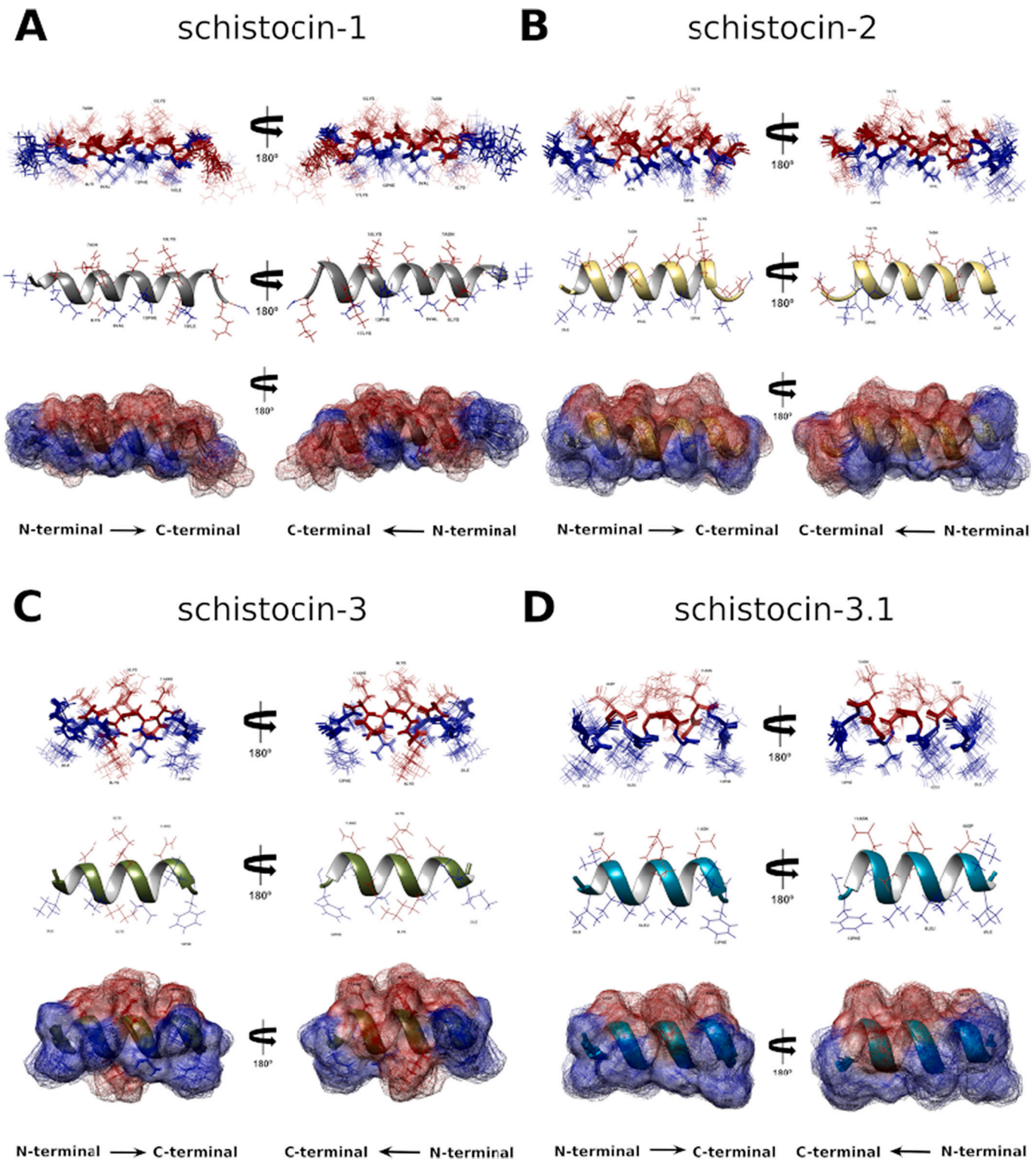


Fig. 4. Solution NMR structure of 400 μM in 20 mM DPC-d_{38} micelles. (A) schistocin-1, (B) schistocin-2, (C) schistocin-3, (D) schistocin-3.1. All peptide (400 μM) structures were obtained in Bruker Avance III 500 (11.75 T) in DPC-d_{38} micelles (20 mM). First, the ten lowest energy structures (atoms superposition representation), evidencing the hydrophobic residues, in blue, and the hydrophilic residues, in red. Second, the lowest energy structure (ribbon representation). Third, the lowest energy structure (surface representation), with the hydrophobic regions, in blue, and the hydrophilic regions, in red.

algorithm impediment, this could be solved by adding new layers to the bioactive peptide prediction, as proposed by Lin et al. (2019), as well as other aspects in peptide predictions, such as position-specific residue preference [49,53]. This study robustness is the approach of combining the computational effort and the experimental verification, which led us to a great understanding of cryptides identification.

To our knowledge, this is the first report on SmKI-1 cryptides with antimicrobial activity. This work strengthens several others that came across cryptides derived from pathways responsible for protein degradation or other ones with many functions, often completely different

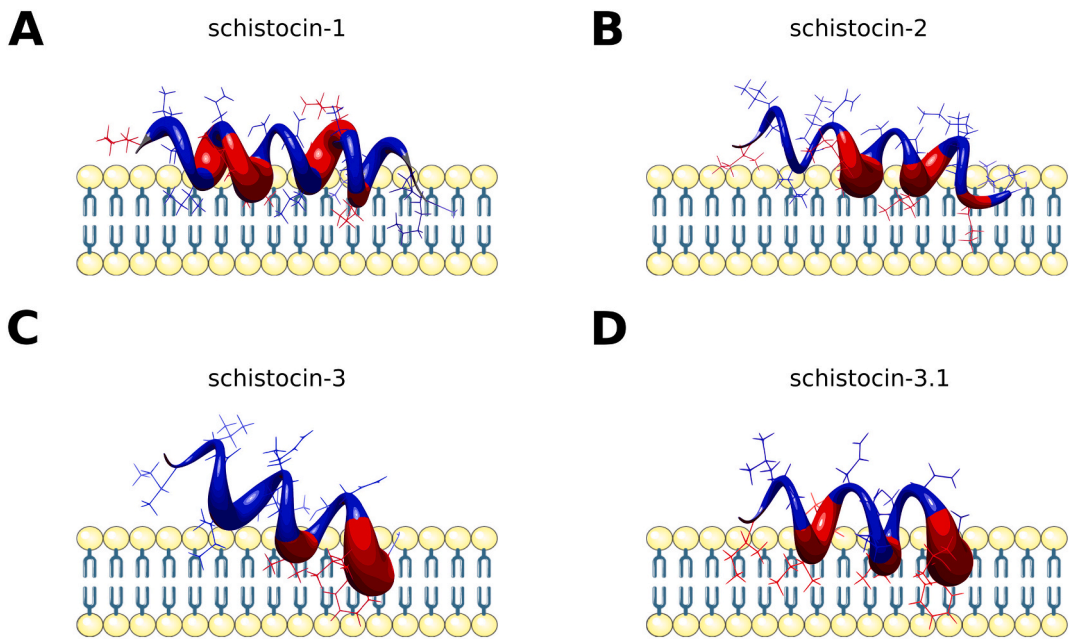


Fig. 5. Amide hydrogen exchange shown in volume rate for each residue of the schistocins. Residues with low exchange rate are shown in higher volume, while the residues with low volume are those with fast exchange during the hydrogen-deuterium TOCSY experiments. Cartoon representation in blue: residues with fast exchange. Cartoon representation in red: residues with slow exchange.

from those of the parent protein [24,31,127]. Additionally, it brings a new functionality for SmKI-1 C-terminal domain by using the concept of cryptides. We demonstrated that one could look into protein sequences differently and get more information about proteins' function. This multifaceted protein has more to come and to be learned, and most likely many other antimicrobial peptides described in the literature, further design is still necessary until converging into potential drug candidates. Therefore, we described a new tool to hunt for new peptides and uncover proteins' hidden functions. In addition, we interestingly observed the affinity of the peptides originated from the membrane domain to interact with vesicles only at a temperature similar to that in the human body. This effect could be related to: a) the energy required for peptide-peptide interactions in an aggregated form to give rise to peptide-phospholipid interactions during the binding process; b) a parasite specificity in the adaptation to the human host, making the schistosomula retain the SmKI-1 in the tegument and liberating only at specific temperature oscillation, pH variation or external enzymatic action.

The presented enCrypted algorithm is being improved to extend its application to other classes of bioactive peptides and the design of point-mutation and chemical modifications. Our perspective is that cryptides screening and peptide rational design will enable us to focus our efforts on biophysical characterization and using experimental data to create, in a near future, a more efficient encrypted bioactive peptide predictor.

5. Conclusions

While the identification of novel bioactive peptides can be performed by a top-down, untargeted approach by high throughput solution proteomic, we showed here that a bioprospection of encrypted, "hidden" peptides can satisfactorily be conducted by data mining using predictive analyses of protein database based on physicochemical properties of the sequence-only approach. The enCrypted algorithm presented here was capable of identifying hundreds of predicted AMPs, successfully electing three representatives for biological validation by antimicrobial assay and structural and thermodynamic analysis, and further optimization by point mutation. These observations demonstrated that SmKI-1 protein is a protein with multiple biotechnological applications and the bioactive peptide-based agents derived from it might have high impact for system biology and biotechnology can be a successful strategy. The extension of this study for other biological systems and further optimization of enCrypted will assist in the rapid identification and design of novel bioactive peptide-based agents. Future investigations will focus in learning the mechanism of action of release of shisctocins and how the protein gets accessible to the main proteases able to cut them off.

Author contributions

Conceptualization: BS, SO, MM, Methodology: BS, EA, CF, RV, LL, SO, MM, Formal analyses: BS, EA, RV, MM. Investigation: BS, EA, CF, RV, LL, SO, MM. Writing: BS, MM, Writing-revision and edition: BS, EA, AG-N, RV, LL, SO, MM. All the authors contributed to the article and approved the manuscript.

Funding

This research was supported by grants from Universidade Federal de Minas Gerais (UFMG/PRPQ/PQ; grant #23853-405 01/2017 to MM), Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq; grants #465229/2014-0 to SO, and #303044/2020-9 to SO, by Fundação de Amparo a Pesquisa de São Paulo (FAPESP; grant #2017/24832-6 to SO), Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG; grant# REDE-00140-16 to SO and MM). BS was supported by a DSc fellowship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; grant #88882.349264/2019-01; Finantial code #001). The funding agencies had no role in the study design, data collection and analysis, or decision to publish or prepare the manuscript.

Data availability

All coordinates presented in this paper have been deposited in the Protein Data Bank (PDB) with the following codes: schistocin-1: 7M67; schistocin-2: 7M73; schistocin-3: 7M77; schistocin-3.1: 7M79. All NMR data have been deposited in the Biological Magnetic Resonance Bank (BMRB) with the following accession numbers: schistocin-1: 30891,

schistocin-2: 30892, schistocin-3: 30893 and schistocin-3.1: 30894. All other datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

None.

Acknowledgements

We thank Dr. Talita Lopes Santos for her technical support on peptide synthesis at Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM). We thank Jamil de Oliveira for its technical support in some experiments. We thank Dr. Yulia Pustovalova and Dr. L. Maurício T. R. Lima for critical reading of the manuscript. We would like to thank Laboratório Multiusuário de Proteômica (LMProt/UFMG), Laboratório de Ressonância Magnética Nuclear (RMN/UFG/IQ), LASEB (UFVJM) and their staff for the access to their analytical facilities. This publication is part of BS DSc thesis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2021.129989.