



Article

# Isolation of Cysteine-Rich Peptides from *Citrullus colocynthis*

Behzad Shahin-Kaleybar <sup>1,2</sup>, Ali Niazi <sup>2</sup>, Alireza Afsharifar <sup>3</sup>, Ghorbanali Nematzadeh <sup>4</sup>, Reza Yousefi <sup>5</sup>, Bernhard Retzl <sup>1</sup>, Roland Hellinger <sup>1</sup>, Edin Muratspahić <sup>1</sup> and Christian W. Gruber <sup>1,\*</sup>

<sup>1</sup> Center for Physiology and Pharmacology, Medical University of Vienna, 1090 Vienna, Austria; shahin.bio65@gmail.com (B.S.-K.); bernhard.retzl@meduniwien.ac.at (B.R.); roland.hellinger@meduniwien.ac.at (R.H.); edin.muratspahic@meduniwien.ac.at (E.M.)

<sup>2</sup> Department of Plant Biotechnology, Shiraz University, Shiraz 7144165186, Iran; niazi@shirazu.ac.ir

<sup>3</sup> Department of Plant Protection, Shiraz University, Shiraz 7144165186, Iran; afshari@shirazu.ac.ir

<sup>4</sup> Department of Plant Breeding and Biotechnology, SANRU, Sari P.O. Box 578, Iran; gh.nematzadeh@sanru.ac.ir

<sup>5</sup> Department of Biology, Shiraz University, Shiraz 7194684795, Iran; ryousefi@shirazu.ac.ir

\* Correspondence: christian.w.gruber@meduniwien.ac.at; Tel.: +43-(0)-140160-31390

Received: 15 August 2020; Accepted: 11 September 2020; Published: 16 September 2020



**Abstract:** The plant *Citrullus colocynthis*, a member of the squash (Cucurbitaceae) family, has a long history in traditional medicine. Based on the ancient knowledge about the healing properties of herbal preparations, plant-derived small molecules, e.g., salicylic acid, or quinine, have been integral to modern drug discovery. Additionally, many plant families, such as Cucurbitaceae, are known as a rich source for cysteine-rich peptides, which are gaining importance as valuable pharmaceuticals. In this study, we characterized the *C. colocynthis* peptidome using chemical modification of cysteine residues, and mass shift analysis via matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. We identified the presence of at least 23 cysteine-rich peptides in this plant, and eight novel peptides, named citcol-1 to -8, with a molecular weight between ~3650 and 4160 Da, were purified using reversed-phase high performance liquid chromatography (HPLC), and their amino acid sequences were determined by de novo assignment of b- and y-ion series of proteolytic peptide fragments. In silico analysis of citcol peptides revealed a high sequence similarity to trypsin inhibitor peptides from *Cucumis sativus*, *Momordica cochinchinensis*, *Momordica macrophylla* and *Momordica sphaeroidea*. Using genome/transcriptome mining it was possible to identify precursor sequences of this peptide family in related Cucurbitaceae species that cluster into trypsin inhibitor and antimicrobial peptides. Based on our analysis, the presence or absence of a crucial Arg/Lys residue at the putative P1 position may be used to classify these common cysteine-rich peptides by functional properties. Despite sequence homology and the common classification into the inhibitor cysteine knot family, these peptides appear to have diverse and additional bioactivities yet to be revealed.

**Keywords:** Cucurbitaceae; *Citrullus colocynthis*; peptides; trypsin inhibitor; knottin

## 1. Introduction

Small molecules have been the dominant class of chemicals in drug discovery and development, in particular due to their favorable pharmacokinetic properties and low production costs [1]. However, they often lack target selectivity, which is generally associated with unwanted side-effects in clinical applications. Hence, more recently, biologics such as antibodies have emerged as promising molecular entities in drug development since they are highly selective for their target, but they have high

production costs. One drawback associated with using proteins as pharmaceutical drugs is their low metabolic and structural stability [2,3]. To overcome these limitations, small peptides (~5–50 amino acids in length) appear to be ideal molecules: peptide and medicinal chemists are able to design peptides with drug-like properties, such as metabolic stability, great efficacy and low off-target effects [4]. Interestingly, nature has provided the blueprints for many of these molecules. Nature-derived cyclic and/or cysteine-stabilized peptides are an emerging class of natural products as templates for drug development or as drug lead candidates [5,6]. In fact, there are numerous clinical candidates and approved peptide drugs available that were derived from natural sources [7–9]. To highlight two examples: the cone snail toxin-derived ziconotide [10] has been on the market since 2000 for management of neuropathic pain, and the plant-derived peptide T20K is in clinical development for multiple sclerosis [11].

Plants provide 'natural libraries' for the discovery of novel cysteine-stabilized peptides (also known as cysteine-rich peptides, CRPs). They are still an underexplored class of molecules, which exhibit great natural diversity [12]. They are gene-encoded and ribosomally produced natural products [13] and hence they are accessible not only by chemical extraction and mass spectrometry, but via genetic mining and bioinformatics analysis [14]. Their endogenous physiological function in planta appears to be defense against herbivores and microbes, and therefore CRPs have demonstrated several pharmacological activities including anthelmintic [15], antimalarial [16], antimicrobial [17], antiviral [18], antitumor [19], and protease inhibition [20], which makes them valuable molecules for drug discovery. Structurally, the disulfide bonds of CRPs are often arranged in a knotted configuration, which provides a stable structural fold [21]. Hence, CRPs have received much attention as templates or scaffolds for designing orally active drug lead molecules since they exhibit high resistance to the acidic pH conditions and proteolytic environment of the gastrointestinal tract [22].

In the present study, we explored the diversity of CRPs from *Citrullus colocynthis* (Cucurbitaceae). The Cucurbitaceae family is well known as a rich source of CRPs containing an inhibitor cystine knot (ICK) structural motif [23–26]. The species *C. colocynthis* (known under the common names bitter apple, desert gourd or vine of Sodom) is a desert vine plant native to the Mediterranean Basin and Asia. It is well known for its anti-inflammatory [27] and antidiabetic properties in evidence-based traditional medicine [28,29]. In addition, some isolated constituents of this plant like cucurbitacin have been explored as therapeutic option for cancer treatment: it showed growth inhibitory activity on human breast cancer cells [30] and cytotoxicity on multidrug-resistant cancer cells [31].

Following chemical extraction, we used mass spectrometry to characterize CRPs in *C. colocynthis*. We identified at least 23 different 'citcol' peptides. Several novel peptides were isolated for analysis of their amino acid sequence. Since they appeared to be homologous to other known squash trypsin inhibitors, we attempted to confirm their function using enzyme inhibition assays and used genome/transcriptome mining to identify structural features in Cucurbitaceae precursor sequences to classify this family of CRPs.

## 2. Materials and Methods

### 2.1. Plant Material and Extraction

The aerial parts of *C. colocynthis* (L.) Schrad. were collected in the summer of 2018 in the region around Toodej mountain of Estahban city in Fars Province, Iran at an altitude of 1300 m. A voucher specimen (no. 3842) was deposited and identified by the Agricultural Research Education and Extension Organization Institute, Iran. The plant material was dried and stored at 25 °C in a moisture-free environment. The dried seeds and fruits of *C. colocynthis* (400 g) were ground using a coffee grinder and, peptides were extracted with 2 L of dichloromethane/methanol (1:1; v/v) by continuous agitation overnight at 25 °C [32]. Next, the extract was filtered, and the peptide-rich fraction was separated by liquid–liquid fractionation by adding 1 L of ddH<sub>2</sub>O. The remaining aqueous extract was concentrated using a rotary evaporator and dried by lyophilization. The crude extract was dissolved in 5% of solvent

B (90% acetonitrile, 9.9% ddH<sub>2</sub>O and 0.1% trifluoroacetic acid, TFA, *v/v/v*) and peptides were enriched by solid-phase extraction with C<sub>18</sub> silica beads (ZEOPrep 60 Å, C<sub>18</sub> irregular 40–63 µm; ZEOCHEM, Uetikon, Switzerland), after activating with one bed volume of methanol and equilibration with two bed volumes of solvent A (99.9% ddH<sub>2</sub>O and 0.1% TFA, *v/v*). Plant compounds bound to the C<sub>18</sub> material were washed and eluted with increasing concentrations of solvent B (wash 1:10%, wash 2:25%, elution 1:35% and elution 2:70%). The resulting eluates were analyzed by matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry (MS) as outlined below. The peptide-rich fraction eluting between 35–70% solvent B was lyophilized and stored at –20 °C until further use.

### 2.2. Reversed-Phase High Performance Liquid Chromatography (RP-HPLC) Purification

A Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA) was used for further fractionation and purification. Aliquots of the peptide-rich fraction (100 mg) were applied and purified in multiple cycles on a preparative Kromasil C<sub>18</sub> column (dichrom GmbH, Marl, Germany; 250 × 21.2 mm, 10 µm, 110 Å), at a flow rate set to 8 mL/min with up to 75-min linear gradient (0.8%/min) from 10% (5 min) to 66% (75 min) of solvent C (90% acetonitrile, 9.92% ddH<sub>2</sub>O and 0.08% TFA, *v/v/v*). The peptide-containing fractions were identified by MS-analysis and further purified by semi-preparative RP-HPLC (Kromasil C<sub>18</sub> column, 250 × 10 mm, 5 µm) with a flow rate of 3 mL/min, and a linear gradient of 0.7%/min of solvent C. Elution profiles were monitored via a multiwavelength detector at 214, 254, and 280 nm and peaks were collected manually. The identity and purity of the peptides were determined by MALDI-TOF analysis and analytical HPLC using a Kromasil C<sub>18</sub> column (5 µm, 100 Å, 250 × 4.6 mm) with a flow rate of 1 mL/min.

### 2.3. Reduction, Alkylation and Enzymatic Digestion of Peptides

The purified and lyophilized peptides (50 µg each) were dissolved in 0.2 mL of 100 mM ammonium bicarbonate/NaOH buffer, pH 8.6. The mixture was incubated for 30 min after addition of 12 µL freshly prepared dithiothreitol (DTT, Sigma-Aldrich, St. Louis, MO; final concentration: 10 mM), at 60 °C in the absence of light under gentle agitation. Iodoacetamide (Sigma-Aldrich, St. Louis, MO) was added to the reaction mixture at a final concentration of 50 mM and incubated for 1 min at 65 °C. Quenching of unreacted iodoacetamide was performed by adding 5.5 µL DTT solution (final concentration of 5 mM), then the mixture was diluted two-fold by adding ammonium bicarbonate buffer. The reduced and alkylated peptides (50 µg) were enzymatically digested by 0.5 µg of endoproteinase Glu-C, 0.2 µg of trypsin, or 0.5 µg of chymotrypsin (all proteases were MS-grade from Sigma-Aldrich, St. Louis, MO, USA) at 37 °C for 3 to 16 h, and the digestion was monitored over time by recording MALDI-TOF spectra in two hour intervals. The digestion was terminated by adding TFA at a final concentration of 0.3%, the samples were desalted and concentrated by using C<sub>18</sub> Zip-Tips™ according to manufacturer's instructions (Millipore, Billerica, MA) and stored at –20 °C prior to analysis.

### 2.4. MALDI-TOF/TOF Analysis and De Novo Peptide Sequencing

Molecular weight analysis of peptides was performed by MALDI-TOF mass spectrometry on a 4800 Analyzer (ABSciex, Framingham MA, USA) operated in reflector positive ion mode acquiring between 2000 and 10,000 total shots per spectrum with a laser intensity set between 4500 and 4800. The matrix  $\alpha$ -cyano-4-hydroxycinnamic acid (Sigma–Aldrich, St. Louis, MO, USA) was prepared by dissolving 5 mg of the matrix in 1 mL of 50% acetonitrile containing 0.1% TFA (*v/v*), and 0.5 µL of the samples were directly spotted onto the target plate after mixing in a 1:6 ratio (sample: matrix). Mass analysis of native plant extract and peptide-enriched fractions were acquired and processed using 4800 Analyzer Software in the range of 500–5000 *m/z* and analyzed by Data Explorer Software. Corresponding precursor masses were used for tandem mass fragmentation in positive ionization mode, with/without collision-induced dissociation with laser energy of 1 kV. Finally, the sequences were determined by manual fragment assembly of identified *N*-terminal *b*- and *C*-terminal *y*-ions series. The isobaric amino

acids leucine and isoleucine as well as glutamine and lysine were distinguished by identification of site-specific cleavage fragments of trypsin and chymotrypsin digestions. The determined amino acid sequences were confirmed using the ion fragmentation calculator in Data Explorer software, sequence homology and amino acid composition analysis (described below).

### 2.5. Amino Acid Composition Analysis

For confirmation of peptide sequences High Sensitivity Amino Acid Analysis (Macquarie University, Sydney, Australia) was performed. In brief, the lyophilized peptides were dissolved in 0.2 mL of 20% acetonitrile/0.1% TFA (*v/v*) and aliquots were dried for analysis. Samples underwent 24 h gas phase hydrolysis in 6 M hydrochloric acid at 110 °C. After hydrolysis, all amino acids were labelled using the Waters AccQTag Ultra chemistry (following supplier's recommendations) and analyzed on a Waters Acquity UPLC. Aliquots were analyzed in duplicate and results expressed as an average (units of ng/sample).

### 2.6. Trypsin, $\alpha$ -Chymotrypsin and $\alpha$ -Amylase Inhibitory Assays

The inhibitory assays were performed in 96-well microtiter plates using a Synergy H4 Hybrid reader (BioTek). For the trypsin inhibitor assay, the reaction mixture contained 140  $\mu$ L of 20 mM Tris-HCl buffer (pH 8.0), 20  $\mu$ L bovine pancreatic trypsin (Sigma Aldrich) at a final concentration of 2.5  $\mu$ g/well, 20  $\mu$ L citcol peptides at different concentrations and 20  $\mu$ L of the substrate benzoyl-arginine-p-nitroanilide dissolved in 20 mM Tris-HCl buffer (pH 8.0) containing 20% dimethylsulfoxide. The mixture was equilibrated for 10 min at 5 °C and then the substrate was added automatically (900  $\mu$ M final concentration) at the start of the assay (time point 0). The absorbance was recorded at a wavelength of 410 nm in 5 min intervals for 1 h at 37 °C. The  $\alpha$ -chymotrypsin inhibitory assay was performed with 1  $\mu$ g/well of pancreatic bovine  $\alpha$ -chymotrypsin (Sigma Aldrich) and Ala-Ala-Phe-7-amino-4-methylcoumarin (150  $\mu$ M final concentration) as substrate similar to the trypsin inhibitor assay with the difference that the fluorescence signal was recorded at 480 $\pm$ 20 nm after excitation at 360  $\pm$  20 nm. For the  $\alpha$ -amylase inhibition assay 50  $\mu$ L of citcol peptides at different concentrations were added to 150  $\mu$ L of 20 mM phosphate buffer (pH 6.9) containing 6.7 mM NaCl and 2 U of  $\alpha$ -amylase from *Aspergillus oryzae*, *Bacillus subtilis* or *Sus scrofa*; (all from Sigma-Aldrich). The mixture was pre-incubated for 10 min at 30 °C under mild agitation. Afterwards 50  $\mu$ L of potato starch solution (1% stock solution in water, *w/v*) was added and incubated for 10 min. Finally, the reaction was terminated by adding 250  $\mu$ L of color reagent mixture (96 mM 3,5-dinitrosalicylic acid solution prepared in 5.3 M sodium potassium tartrate solution and 2 M NaOH,) and heating for 10 min at 99 °C. After cooling to room temperature, the mixture was diluted with 0.5 ml of ddH<sub>2</sub>O, and the absorbance was measured at 540 nm. In all assays, enzyme activity without inhibitor (negative control) and complete inhibition of enzymes (positive control) were measured by replacing peptides with water or acarbose (for  $\alpha$ -amylase), phenylmethylsulfonylfluorid (for  $\alpha$ -chymotrypsin) and sunflower trypsin inhibitor I (for trypsin), respectively. For each sample, the background signal was measured by replacing the enzyme with water. All the assays were carried out in three biological replicates and are presented as percentage of inhibition (mean  $\pm$  standard deviation) relative to the total enzyme activity. Molar concentration of citcol peptides was determined by measuring absorption at 280 nm using Lambert-Beer's law and the molar extinction coefficients for tryptophan (5690 M<sup>-1</sup> cm<sup>-1</sup>), cystines (120 M<sup>-1</sup> cm<sup>-1</sup>) and tyrosine (1280 M<sup>-1</sup> cm<sup>-1</sup>).

### 2.7. Sequence Homology Analysis

For initial online homology blastp and tblastn searches against UniProt/NCBI (National Center for Biotechnology Information) databases, the amino acid sequences of citcol peptides were used as queries and results with E-values of less than 0.1 were further considered. In addition, representative annotated peptide sequences of different functional CRP families, comprising at least six Cys residues in the mature processed peptide domain and belonging to different CXC patterns were used for comparison

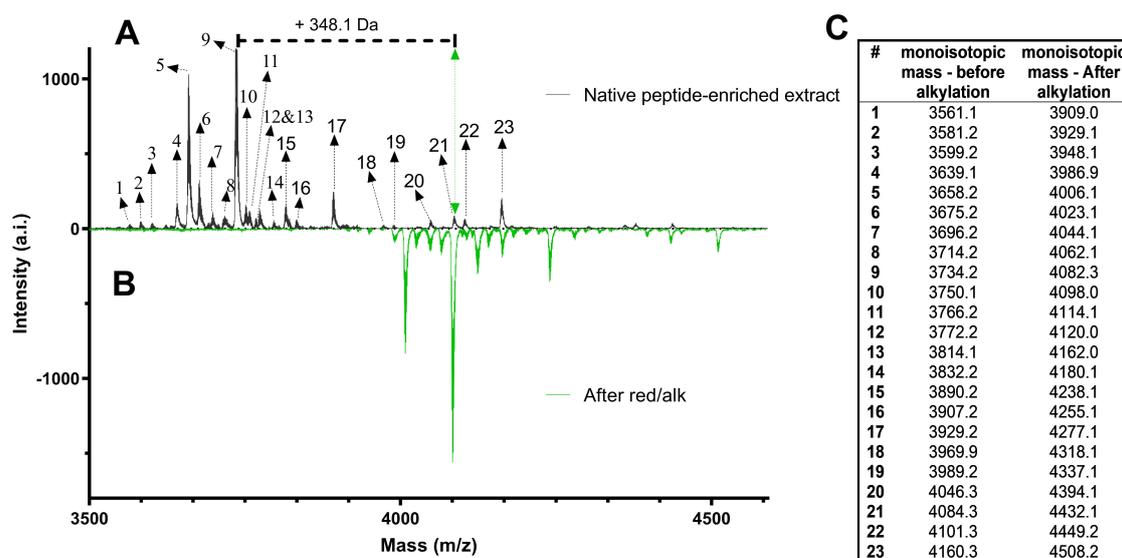
to novel citcol peptides: *Amaranthus hypochondriacus* (UniProt entry: P80403), *Spinacia oleracea* (P84781), *Phytolacca americana* (P81418), *Solanum tuberosum* (P01075), *Astragalus sinicus* (Q07A30), *Amaranthus retroflexus* (Q5I2B2), *Momordica cochinchinensis* (P82410). For similarity analysis of the initial hits a pairwise sequence comparison was carried out using the EMBOSS Needle online tool with citcol-8 as reference. For in-depth blastp searches, plant genomes and transcriptomes were downloaded from NCBI's FTP server, and additional 1341 transcriptomes (Data S1) were retrieved from the 1-kp project [33]. Prior to blastp analysis all data was six-frame translated using Python and Biopython [34]. As queries we used the following protein sequences: citcol 1-8, P01074, P10291, P10292, P10293, P10295, P11969, Q9S8D2 and Q9S8W3. Hits with E-values of less than 0.001 were considered as significant. Multiple sequence alignments were performed by Multiple Sequence Comparison by Log-Expectation (MUSCLE) [35], and sequence logos were generated using WebLogo application [36]. The phylogenetic tree construction and pairwise similarity analysis were performed by maximum-likelihood method with sequences found by blastp searches against six-frame translated genomes/transcriptomes with the software MEGA X [37], version 10.1.7, respectively and illustrated by iTOL [37], version 5.5.

### 3. Results and Discussion

Nature-derived peptides often exhibit high stability based on a network of disulfide bonds while maintaining great tolerance for sequence modifications of the intercysteine loops, a feature termed 'structural plasticity' [38]. Consequently, CRPs are under investigation as molecular scaffolds for engineering pharmaceutically useful peptides. CRPs have been studied from numerous plant families [39,40], and the squash family (Cucurbitaceae) is known as a rich source of these stabilized peptides [5,41]. In this study, we isolated and characterized novel CRPs from *C. colocynthis* a member of the Cucurbitaceae family. We confirmed the presence of at least 23 CRPs in this plant based on peptide mass fingerprinting, which were de novo sequenced by enzymatic digests, mass spectrometry and amino acid analysis. Based on sequence homology analysis we predicted the putative cysteine-connectivity of the novel citcol peptides. Genome/transcriptome mining and a subsequent bioinformatics analysis enabled us to elucidate the phylogenetic distribution and to propose a classification of Cucurbitaceae CRPs into trypsin inhibitors and antimicrobial peptides providing a better understanding of the distribution of these two subgroups of the manifold 'knottins' peptide family.

#### 3.1. Preparation of Peptide-Enriched Citcol Extracts

Since Cucurbitaceae plants are known as a rich source for CRPs we explored the proteome of *C. colocynthis* for discovering novel CRPs. To remove secondary metabolites and plant pigments such as chlorophyll, 400 g of dried plant material were extracted overnight by dichloromethane/MeOH. After filtration, addition of water and thorough liquid/liquid phase partitioning, the aqueous phase containing peptides, which yielded 430 mg crude dried material was used for subsequent experiments. Removal of polar plant compounds was achieved by C<sub>18</sub> solid-phase extraction in batch, while washing and eluting with increasing concentrations of acetonitrile (solvent B). Molecular weight analysis of the collected fractions obtained one sample with an elution condition of 35%–70% solvent B as major source of the desired peptides within a mass range of 3.5–4.5 kDa. Chemical modification of cysteine residues via alkylation confirmed the presence of at least 23 unique peptide mass signals with a mass shift of 348 Da, each corresponding to six carbamidomethylated cysteine residue (Figure 1), suggesting the citcol peptides may belong to a CRP family comprising three disulfide bonds.



**Figure 1.** Chemical analysis of the citcol peptidome using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) analysis. **(A)** The mass spectrum of a representative *Citrullus colocynthis* C<sub>18</sub> peptide-enriched extract is illustrated in the range of  $m/z$  3500–4600 (black trace). Each black dotted arrow indicated by numbers (#) represents a cysteine-rich peptide (refer to Table in C). **(B)** Chemical modification of cysteine residues was performed by reduction and alkylation. A representative spectrum after S-carbamidomethylation is illustrated in green. The mass shift of  $\Delta 348$  Da indicates the presence of six cysteine residues. For instance, the peptide signal at  $m/z$  3734.2 shifts by 348.1 Da to  $m/z$  4082.3. **(C)** List of monoisotopic mass signals  $[M + H]^+$  of citcol peptides before and after S-carbamidomethylation.

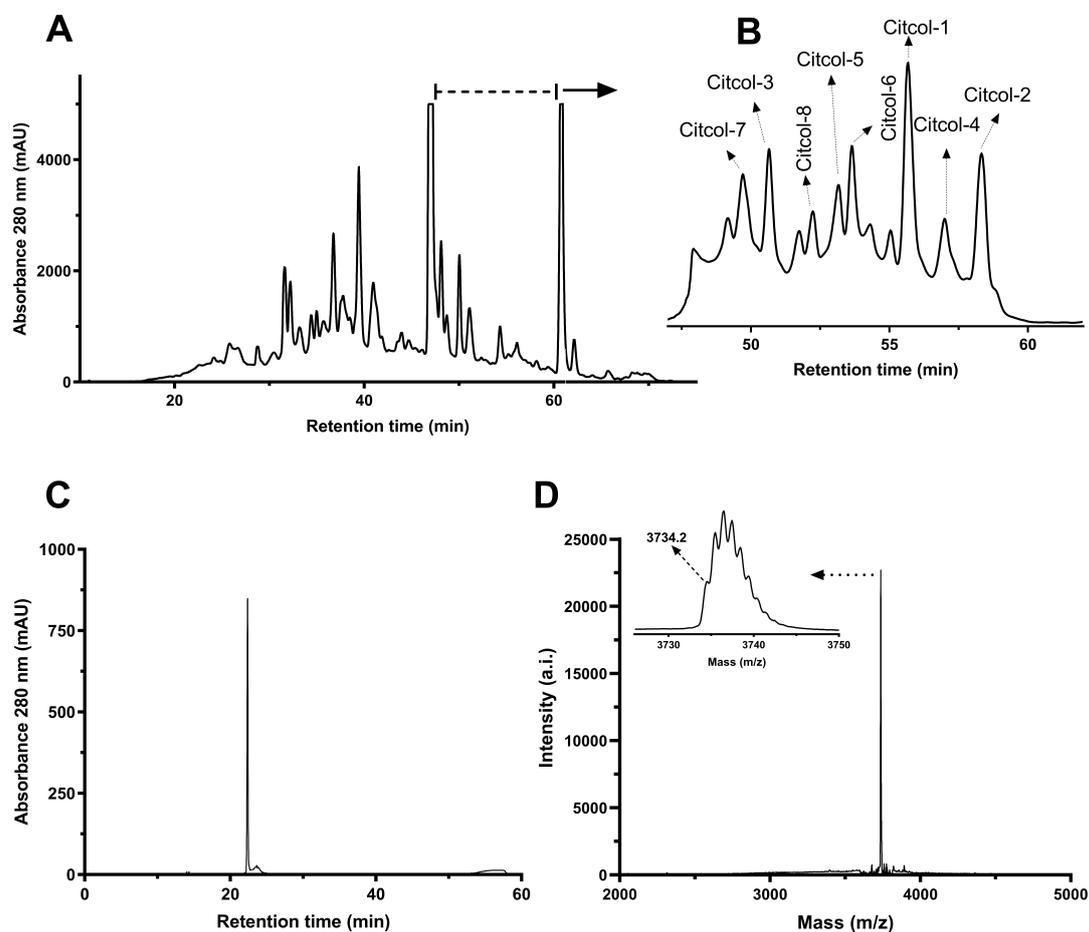
### 3.2. Reversed-Phase HPLC Fractionation and Purification

RP-HPLC was performed for purification of detected CRPs from the peptide-enriched extract of *C. colocynthis*. MALDI-TOF analysis of fractions collected from RP-HPLC enabled the identification of peptide peaks in the chromatogram, which typically eluted between 42 and 55% solvent C (Figure 2A). Preparative and semi-preparative RP-HPLC allowed isolation of eight different CRPs with mass signals between  $m/z$  3658.1 to 4160.3, which were named citcol-1 to -8 (*Citrullus colocynthis*) (Figure 2B). Analytical HPLC and MALDI-TOF analysis were used to confirm purity of the isolated citcol peptides  $\geq 90\%$  for individual peptides (exemplarily shown for citcol-2 in Figure 2C,D).

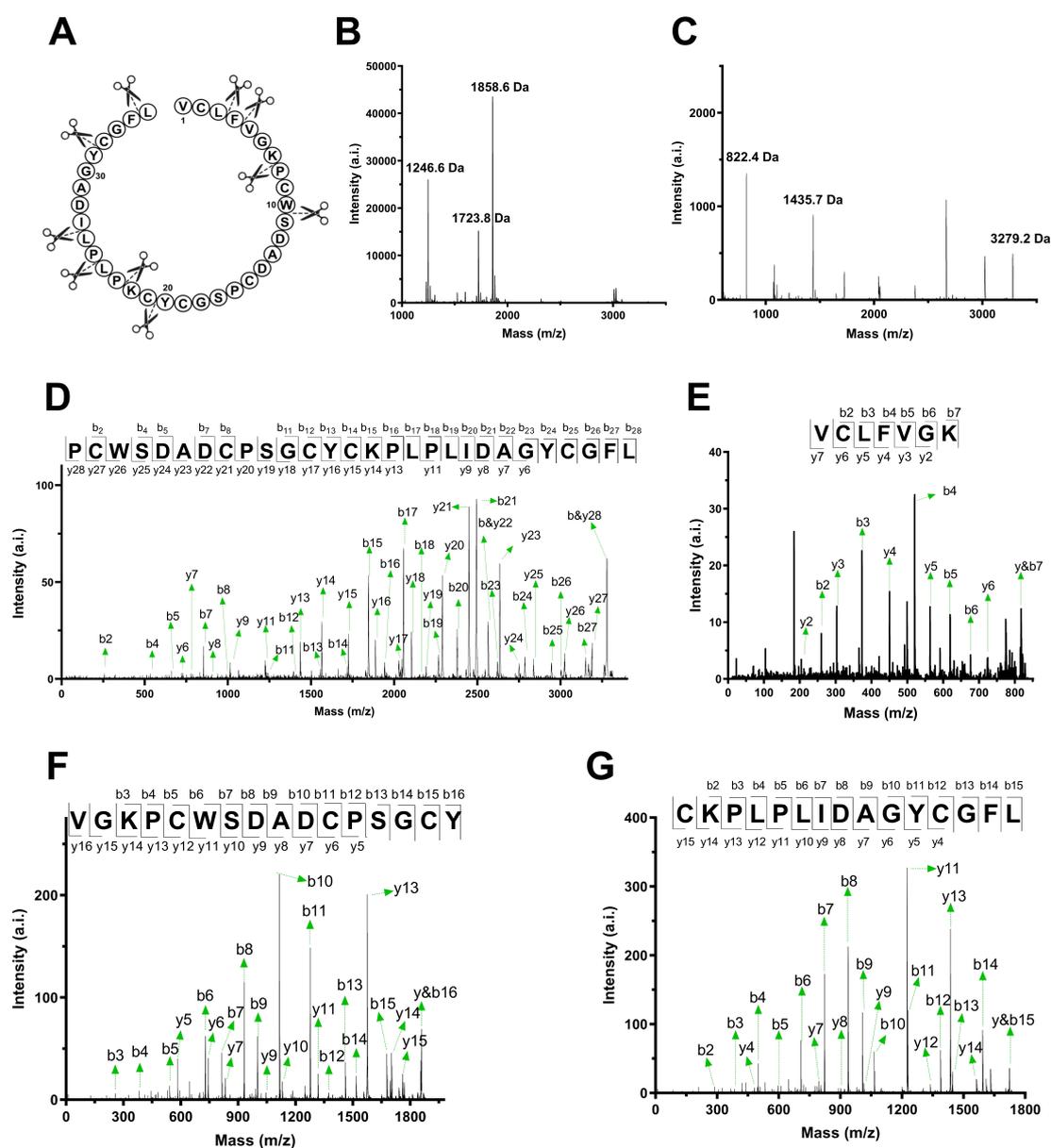
### 3.3. De Novo Sequencing of Citcol Peptides by MALDI-TOF/TOF

To determine the amino acid sequences of citcol peptides, well-established peptidomics protocols were applied [14]. Based on MALDI-TOF analysis, S-carbamidomethylation, resulted in a mass shift of  $+347.9 \pm 0.2$  Da corresponding to the presence of six cysteine residues in citcol peptides (Figure 1). To determine complete sequence of each peptide the Cys-modified peptides were partially digested by trypsin and chymotrypsin to produce short peptide fragments which are generally well amenable to collision-induced fragmentation. Exemplarily, the predicted cleavage sites and mass spectra obtained after digestion of citcol-2 are shown in Figure 3A–C. There were three major mass signals with an absolute intensity  $>500$  after site specific chymotryptic proteolysis with corresponding monoisotopic masses of 1246.6 Da, 1723.8 Da, 1858.6 Da, and three major fragments with 822.4 Da, 1435.7 Da, and 3279.2 Da resulted by tryptic digestion. The identified peaks of citcol-2 were used as precursor ions for collision-induced fragmentation and sequence assembly by assignment of N-terminal b- and C-terminal y-ion series (Figure 3D–G, Figure S1). Using sequence fragment assembly, we identified the correct order of overlapping fragments and full-length sequences of citcol peptides. For instance, citcol-2 has the sequence NH<sub>2</sub>-VCLFVGKPCWSDADCPSCGYCKPLPLIDAGYCGFL-COOH. Complete

tryptic/chymotryptic digestion allowed to distinguish between isobaric residues Gln/Lys and Leu/Ile, respectively and sequence similarity analysis, high-sensitivity amino acid analysis and automated ion fragmentation calculation confirmed the results obtained by de novo peptide sequencing. The sequences of other citcol peptides were elucidated in the same way and the mass spectrometry data (Figure S2) and sequences of these peptides are summarized in Table 1.



**Figure 2.** Fractionation and purification of cysteine-rich citcol peptides by reversed-phase HPLC. (A) Purification chromatogram by preparative RP-HPLC; the dashed line indicates the elution region of citcol peptides. (B) Semi-preparative RP-HPLC of the elution region of citcol peptides (retention time 47–60 min) indicating peaks for citcol-1 to -8. (C) Analytical RP-HPLC chromatogram of citcol-2 (purity ~95%). (D) Molecular weight analysis of citcol-2 corresponding to a monoisotopic mass  $[M + H]^+$  of 3734.1 Da. For HPLC analysis linear gradients of solvent C (90% acetonitrile, 9.92% H<sub>2</sub>O and 0.08% TFA) were used at flow rates of 8, 3 and 0.3 mL/min for preparative, semi-preparative and analytical RP-HPLC, respectively.



**Figure 3.** De novo sequencing of citcol-2 by MALDI-TOF/TOF. (A) Schematic structure and full sequence of citcol-2; scissors on the inside of the circle indicates potential trypsin cleavage sites and scissors on the outside indicate potential chymotrypsin cleavage sites. MS spectrum of citcol-2 after partial digestion with chymotrypsin (B) and trypsin (C). MS/MS fragmentation spectra of precursor peptides with a molecular weight of 3279.2 Da (D) and 822.4 Da (E) after tryptic digestion, and 1858.6 Da (F) and 1723.8 Da (G) after chymotrypsin digestion, which covers the entire sequence of citcol-2. The amino acid sequence was determined by assembly of digested fragments and assignment of N-terminal b- and C-terminal y-ions series. Complete tryptic digestion was used to distinguish between isobaric amino acids Gln/Lys and complete chymotrypsin for Leu/Ile. Comparison of the sequences of the fragments produced by these two enzymes showed the correct order of the fragments. The  $m/z$  value list of each of the precursor masses used for de novo sequencing are shown in Figure S1, where A, B, C and D correspond to precursors with a molecular weight of 3279.2 Da, 822.4 Da, 1723.8 Da and 1858.6 Da, respectively.

**Table 1.** Properties and sequence comparison of isolated citcol peptides.

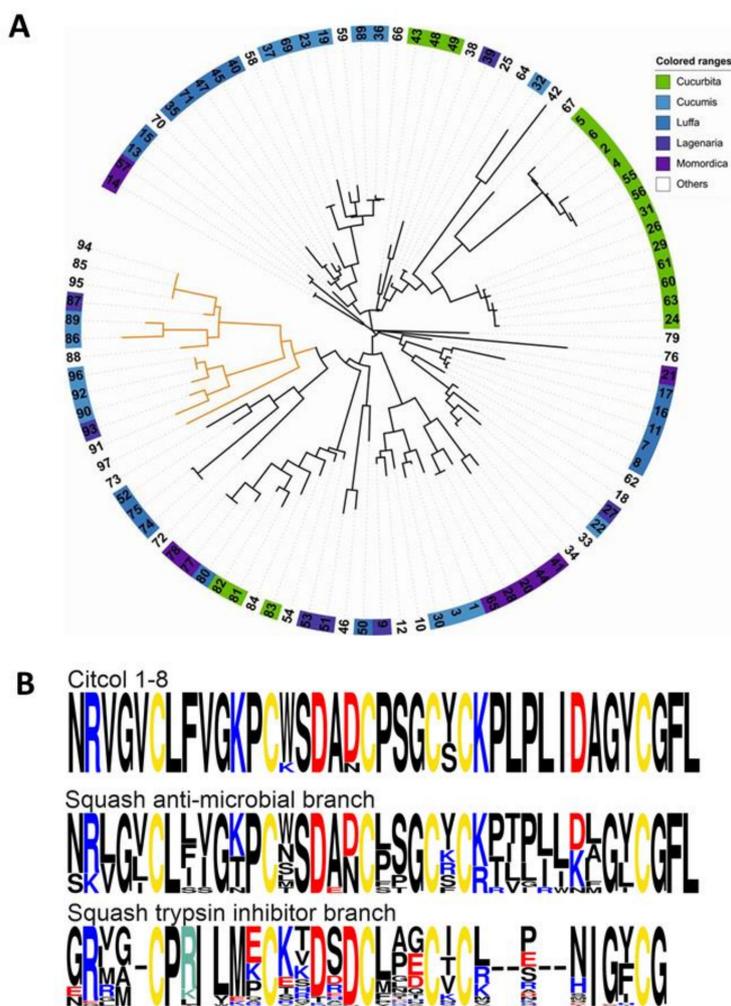
Name	[M + H] <sup>+</sup> <sup>1</sup>	Sequence <sup>2</sup>					Length
		loop 1	2	3	4	5	
citcol-1	3658.1	—VCLFVGKPCWSDADCP	SGC	SCK	PLPLIDAGYCGFL		35 aa
citcol-2	3734.1	—VCLFVGKPCWSDADCP	SGCYCK	PLPLIDAGYCGFL			35 aa
citcol-3	3814.4	–VGVCLFVGKPCWSDADCP	SGC	SCK	PLPLIDAGYCGFL		37 aa
citcol-4	3890.2	–VGVCLFVGKPCWSDADCP	SGCYCK	PLPLIDAGYCGFL			37 aa
citcol-5	4046.9	–RVGVCLFVGKPCWSDADCP	SGCYCK	PLPLIDAGYCGFL			38 aa
citcol-6	4084.1	NRVGVCLFVGKPCWSDADCP	SGC	SCK	PLPLIDAGYCGFL		39 aa
citcol-7	4101.2	NRVGVCLFVGKPCSDANCP	SGCYCK	PLPLIDAGYCGFL			39 aa
citcol-8	4160.3	NRVGVCLFVGKPCWSDADCP	SGCYCK	PLPLIDAGYCGFL			39 aa
		*****	***.****	*****	*****	*****	

<sup>1</sup> monoisotopic mass signal; <sup>2</sup> Conserved and similar residues in all sequences have been marked by asterisks and colons below the bottom line, respectively. All sequences are in N- to C-terminal direction. asterisk '\*' and colon ':' indicate identical and similar residues in the alignment.

Our peptidomics approach does not allow a definite conclusion as to whether all citcol peptides are translated from individual genes, or whether they are produced by splicing events or via posttranslational processing, but there are examples from other plants showing that indeed truncated analogs of CRPs can be made by alternative splicing [42], which has also been reported for other proteins [43–47]. Alignment of citcol-1 to -8 showed that they have a conserved cysteine spacing with the structure X<sub>1-5</sub>CX<sub>6</sub>CX<sub>5</sub>CX<sub>3</sub>CX<sub>1</sub>CX<sub>10</sub>CX<sub>3</sub> in which X represents any amino acid residue except cysteine (Figure 4A). By sequence similarity to other characterized *Momordica* peptides, we predict the citcol peptides to have the following putative cysteine connectivity C<sub>I</sub>-C<sub>IV</sub>, C<sub>II</sub>-C<sub>V</sub>, C<sub>III</sub>-C<sub>VI</sub> (Table S1), which is yet to be confirmed in a future study.

#### 3.4. Genome Mining and Sequence Homology Analysis

To obtain information on the putative biological function of citcol peptides, we developed custom bioinformatics approaches. Standard database searching and sequence homology analysis against UniProt (via blastp) using citcol peptide sequences as queries did not yield any hits with significant homology (E-value ≤ 0.1) to functional characterized peptides. For instance, in our search against UniProt, five hits, which are classified as uncharacterized proteins, were found (A0A0A0KY29, A0A5A7TRX6, A0A5D3E1P7, A0A5A7TM06, A0A5D3E1J3). Additional mining of deposited genome/transcriptome data using the tblastn algorithm yielded a hit with identity to citcol-8 in the translated chromosome 10 of *Citrullus lanatus* (GenBank: VOOL01000010.1) and one hit differing in six amino acids in the translated chromosome 7 of the genome of *Cucumis melo* (GenBank: LN713261.1) (Table 2).



**Figure 4.** Homology and phylogenetic distribution of citcol peptides. (A) Maximum-likelihood phylogenetic tree calculated with sequences found by blastp searches against six-frame translated genomes/transcriptomes (E-value  $\leq 0.1$ ). The antimicrobial branch is colored in orange and the trypsin inhibitory branch is colored in black. The sequences are encoded by numbers (Data 2), and the Cucurbitaceae subfamilies summarized with one color (Cucurbita—green, Cucumis—light blue, Luffa—dark blue, Lagenaria—dark purple, Momordica—violet, Others—white). (B) Alignment of frequency logos calculated with citcol-1 to -8, sequences from the squash antimicrobial peptides branch and the squash trypsin inhibitor branch (see Materials and Methods for details). Cysteines are colored in yellow, amino acids with positively charged side chains (at pH 7) are colored in blue (H, K, R) and with negative charged sidechains (at pH 7) are colored in red (D, E). Gaps ‘-’ were introduced to maximize the alignment.

**Table 2.** Homologous sequences of citcol peptides found by a tblastn search against Genbank <sup>1</sup>.

Species	Sequence <sup>2</sup>	E-Value <sup>3</sup>	Accession No
<i>C. colocynthis</i> (citcol-8)	NRVGVCLEFVGKPCWSDADCPGSGCYCKPLPLIDAGYCGFL	n.a.	this work
<i>Citrullus lanatus</i> (chr10 <sup>4</sup> )	NRVGVCLEFVGKPCWSDADCPGSGCYCKPLPLIDAGYCGFL	$3 \times 10^{-23}$	VOOL01000010.1
<i>Cucumis melo</i> (chr7 <sup>4</sup> )	-RLGVCLLVGKPCMSDADCPGSGCYCKPVLLDIGYCGFL	$5 \times 10^{-16}$	LN713261.1
	*.****.***** *****.***.*****		

<sup>1</sup> only sequences containing six cysteine residues are reported; <sup>2</sup> conserved and similar residues in all sequences have been marked by asterisks and colons below the bottom line, respectively. <sup>3</sup> E-value cut-off of 0.001 was applied; <sup>4</sup> chr: chromosome. n.a.—not applicable, asterisk ‘\*’ and colon ‘:’ indicate identical and similar residues in the alignment.

UniProt/NCBI database analysis by blastp search was used to understand the similarity of citcol peptides to other functional peptide families and annotated CRPs. Initially this search provided one functionally annotated hit, i.e., a trypsin inhibitor peptide isolated from *Momordica cochinchinensis* (E-value >0.1). To evaluate the similarities and Cys loop pattern of citcol peptides with other functional CRPs (or precursors thereof), we selected representative cysteine-rich peptides and sequences sporadically from different families based on functionality and comprising at least six Cys residues (Table 3). A pairwise distance analysis was performed against the following sequences: an  $\alpha$ -amylase inhibitor (from *Amaranthus hypochondriacus*), two trypsin inhibitors (from *Momordica cochinchinensis* and *Spinacia oleracea*), an anti-fungal/antimicrobial peptide (from *Amaranthus retroflexus*), a metallo-carboxypeptidase inhibitor (from *Solanum tuberosum*), a nodule-specific protein (from *Astragalus sinicus*), and an antimicrobial peptide (from *Phytolacca americana*) (Table 3). Homology analysis revealed that the trypsin inhibitor peptide MCoTI-III isolated from *Momordica cochinchinensis* (Cucurbitaceae) shares the highest, i.e., 43.6% sequence homology with citcol-8. In addition, sequence comparison of citcol peptides with several other trypsin inhibitors from *Momordica* spp. (namely MCoTI-I, -II, -III and EETI-II, Table S1) [48–50] revealed that loops 1 to 4 of these peptides share a similar length. Therefore, to determine the functional properties of novel citcol peptides, we tested their protease inhibitory activity. We assayed different concentrations of pure citcol peptides for inhibition of bovine trypsin and  $\alpha$ -chymotrypsin. Furthermore, we also tested the inhibitory activity against  $\alpha$ -amylase enzymes of *Aspergillus oryzae*, *Bacillus subtilis* and *Sus sucrofa*. Surprisingly, the citcol peptides did not exhibit any inhibitory activity against these enzymes (Figure S3), despite their apparent similarity to functional enzyme inhibitor peptides (Table 3). To understand this functional discrepancy, i.e., why citcol peptides do not inhibit enzyme activity (based on the tested assays), a detailed genome/transcriptome mining approach was performed using sequence homology analysis to other squash family (Cucurbitaceae) CRPs. This detailed analysis yielded 97 sequences (Data 2) that were homologous to citcol peptides (E-value  $\leq$  0.1) from several sub-families (genus) of Cucurbitaceae, including Cucurbita, Cucumis, Luffa, Lagenaria and Momordica. The phylogenetic distribution of these publicly available sequences revealed that they cluster into two distinct branches: an antimicrobial and a trypsin inhibitory branch (Figure 4A). Furthermore, the discovered citcol peptides are more similar to antimicrobial peptides than trypsin inhibitory peptides (Figure 4B), as indicated for instance by the length of loop 5.

**Table 3.** Comparison of citcol-8 with representative plant cysteine-rich sequences and peptides.

Species	Sequence	Uniprot ID	Function	Cys Loops Pattern	Similarity (%) <sup>1</sup>	Reference
<i>C. colocythis</i> (citcol-8)	NRVGVCLFVGKPCWSDADCPGSCYCKPLPLIDAGYCGFL	n.a.	n.a.	CX <sub>6</sub> CX <sub>5</sub> CX <sub>3</sub> CX <sub>1</sub> CX <sub>10</sub> C <sup>2</sup>	n.a.	this work
<i>Momordica cochinchinensis</i>	QRACPRLKCKRRDSDCPGECICKENGYCG	P82410	trypsin inhibitor	CX <sub>6</sub> CX <sub>5</sub> CX <sub>3</sub> CX <sub>1</sub> CX <sub>5</sub> C	43.6	[51]
<i>Solanum tuberosum</i>	HADPICNKPKCTHDDCSGAWFCQACWNSARTCGPY	P01075	metallocarboxy peptidase inhibitor	CX <sub>3</sub> CX <sub>5</sub> CX <sub>5</sub> CX <sub>2</sub> CX <sub>6</sub> C	38.5	[52]
<i>Phytolacca americana</i>	AGCIKNGGRCNASAGPPYCCSSYCFQIAGQSYGVCKNR	P81418	antimicrobial peptide	CX <sub>6</sub> CX <sub>8</sub> CCX <sub>3</sub> CX <sub>10</sub> C	31.7	[53]
<i>Spinacia oleracea</i>	EDKCSPSGAICSGFGPPEQCSCGACVPHPLRIFVCQ	P84781	trypsin inhibitor	CX <sub>6</sub> CX <sub>8</sub> CCX <sub>3</sub> CX <sub>10</sub> C	29.5	[54]
<i>Amaranthus hypochondriacus</i>	CIPKWNRCGPKMDGVPCCPEPYTCTSDYYGNCS	P80403	α-amylase inhibitor	CX <sub>6</sub> CX <sub>8</sub> CCX <sub>4</sub> CX <sub>7</sub> C	27.3	[55]
<i>Amaranthus retroflexus</i>	AGECVQGRCPGSMCCSQFGYCGRGPKYCGR	Q5I2B2	antimicrobial peptide	CX <sub>4</sub> CX <sub>4</sub> CCX <sub>5</sub> CX <sub>6</sub> C	26.7	[17]
<i>Astragalus sinicus</i>	TYSCGGHIDCKDFCKSEGYRGFKCTPKKTCTCFH	Q07A30	nodule specific protein	CX <sub>5</sub> CX <sub>3</sub> CX <sub>9</sub> CX <sub>5</sub> CX <sub>1</sub> C	20.0	[56]

<sup>1</sup> the percentage of similarity was calculated based on pairwise alignment of citcol-8 as reference using the EMBOSS Needle online tool; <sup>2</sup> X represents any amino acid residue except cysteine. n.a.—not applicable.

At the molecular detail, we noticed a minor but apparently important modification in the sequence of citcol peptides: comparison of squash family trypsin inhibitor peptides isolated from the seeds of Cucurbitaceae showed that the two basic amino acids Arg and Lys, respectively, are important residues responsible for trypsin inhibitory activity in the active site ( $P_1$  position) in loop 1 (Figure S4), whereas in citcol peptides the  $P_1$  position has been occupied by a Phe residue. This could be one reason for the observed lack of trypsin inhibitory activity of citcol peptides. Nevertheless, such a mutation may not explain the lack of activity against chymotrypsin and  $\alpha$ -amylase. For instance in Bowman-Birk inhibitors a Lys to Phe mutation reportedly was sufficient for a functional change from trypsin to chymotryptic inhibitory capacity [57]. Therefore, it is likely that other modifications, such as the observed differences in loop 5 of citcol peptides as compared to CRPs of *Momordica spp.* (Figure 4A, Table S1) are contributing factors to the observed lack of enzyme inhibition.

#### 4. Conclusions

There are several reports attributing diverse biological activities to plant-derived CRPs [58–60], and in silico analysis such as genome mining has had an important role in CRP discovery [61]. In fact, CRPs have also been found in other organisms, including humans [62] and animals [63,64]. Isolation and characterization of peptides with novel sequences and/or cysteine motifs or new functions is an important process in the discovery of novel potential pharmaceuticals. Here, we purified and characterized eight novel CRPs (citcol-1 to citcol-8) from *C. colocynthis*. Several plant peptides that share sequence similarity to the here-described citcol peptides are well-known protease inhibitors [5,17,48,49]. However, detailed homology and phylogenetic analysis of these citcol peptides in comparison to other Cucurbitaceae CRPs suggested that they were a novel subtype of CRPs with minor variations in loops 1 and 5, leading to a functional specialization, since these novel citcol CRPs did not inhibit representative protein or sugar processing enzymes. For instance, it has been shown that a single amino acid substitution in loop 1 of inhibitory peptides can alter their activity and specificity or inactivate their function [65,66]. It is likely that such minor molecular changes led to the loss of enzyme inhibitory activity, as confirmed in several inhibition assays. At a more general level, the apparent sequence similarities between different subfamilies of plant “trypsin inhibitor” peptides may be misleading for a correct functional classification. Our results underline that detailed bioinformatics analysis is an expedient strategy applicable for discovery of cysteine-rich plant peptides. Diversity in the intercysteine loop sequences despite a common cysteine pattern suggest additional bioactivities for these CRPs that are yet to be identified. Overall, our study adds to the biodiversity of CRPs, a family of plant biomolecules that experiences increasing attention for applications in drug design and agriculture [6,67].

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2218-273X/10/9/1326/s1>, Figure S1: The  $m/z$  value list of the b- and y-ions series for the precursor masses used for de novo sequencing of citcol-2, Figure S2: De novo sequencing of citcol peptides (citcol-1 to 7) by MALDI-TOF/TOF, Figure S3: Enzyme inhibition assays of citcol peptides, Figure S4: Comparison of active site amino acid in squash family trypsin inhibitory CRPs from Cucurbitaceae, Table S1: Comparison of citcol-8 with CRPs from *Momordica spp.*, Data S1: Overview and accession codes of genome and transcriptome data, Data S2: Cucurbitaceae sequences most similar to citcol peptides that were used for phylogenetic analysis.

**Author Contributions:** B.S.-K. performed experiments with the help from B.R., R.H. and E.M.; B.S.-K., A.N., A.A., G.N., R.Y., B.R. and C.W.G. analyzed data; R.H., and E.M. contributed reagents/tools; B.S.-K., B.R., A.A., G.N. and C.W.G. prepared figures and wrote manuscript with the help of R.H. E.M., A.N. and R.Y. All authors approved the final version of the manuscript.

**Funding:** This research was funded by the Austrian Science Fund (FWF) grant numbers I3243 and P32109.

**Acknowledgments:** Behzad Shahin-Kaleybar wishes to thank the Ministry of Science, Research and Technology of the Islamic Republic of Iran for financial support during a research stay in Austria. Open Access Funding by the Austrian Science Fund (FWF).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Craik, D.J.; Fairlie, D.P.; Liras, S.; Price, D. The future of peptide-based drugs. *Chem. Biol. Drug Des.* **2013**, *81*, 136–147. [[CrossRef](#)] [[PubMed](#)]
2. Xu, C.; Lei, C.; Yu, C. Mesoporous silica nanoparticles for protein protection and delivery. *Front. Chem.* **2019**, *7*, 290. [[CrossRef](#)] [[PubMed](#)]
3. Bruno, B.J.; Miller, G.D.; Lim, C.S. Basics and recent advances in peptide and protein drug delivery. *Ther. Deliv.* **2013**, *4*, 1443–1467. [[CrossRef](#)] [[PubMed](#)]
4. Gilad, Y.; Firer, M.; Gellerman, G. Recent innovations in peptide based targeted drug delivery to cancer cells. *Biomedicines* **2016**, *4*, 11. [[CrossRef](#)] [[PubMed](#)]
5. Hellinger, R.; Gruber, C.W. Peptide-based protease inhibitors from plants. *Drug Discov. Today* **2019**, *24*, 1877–1889. [[CrossRef](#)]
6. Wang, C.K.; Craik, D.J. Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nat. Chem. Biol.* **2018**, *14*, 417–427. [[CrossRef](#)] [[PubMed](#)]
7. Uhlig, T.; Kyprianou, T.; Martinelli, F.G.; Oppici, C.A.; Heiligers, D.; Hills, D.; Calvo, X.R.; Verhaert, P. The emergence of peptides in the pharmaceutical business: From exploration to exploitation. *EuPA Open Proteom.* **2014**, *4*, 58–69. [[CrossRef](#)]
8. Lien, S.; Lowman, H.B. Therapeutic peptides. *Trends Biotechnol.* **2003**, *21*, 556–562. [[CrossRef](#)]
9. Lau, J.L.; Dunn, M.K. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg. Med. Chem.* **2018**, *26*, 2700–2707. [[CrossRef](#)]
10. McIntosh, M.; Cruz, L.; Hunkapiller, M.; Gray, W.; Olivera, B. Isolation and structure of a peptide toxin from the marine snail *Conus magus*. *Arch. Biochem. Biophys.* **1982**, *218*, 329–334. [[CrossRef](#)]
11. Thell, K.; Hellinger, R.; Sahin, E.; Michenthaler, P.; Gold-Binder, M.; Haider, T.; Kuttke, M.; Liutkeviciute, Z.; Goransson, U.; Grundemann, C.; et al. Oral activity of a nature-derived cyclic peptide for the treatment of multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3960–3965. [[CrossRef](#)]
12. Das, D.; Jaiswal, M.; Khan, F.N.; Ahamad, S.; Kumar, S. PlantPepDB: A manually curated plant peptide database. *Sci. Rep.* **2020**, *10*, 1–8. [[CrossRef](#)] [[PubMed](#)]
13. Arnison, P.G.; Bibb, M.J.; Bierbaum, G.; Bowers, A.A.; Bugni, T.S.; Bulaj, G.; Camarero, J.A.; Campopiano, D.J.; Challis, G.L.; Clardy, J.; et al. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30*, 108–160. [[CrossRef](#)]
14. Hellinger, R.; Koehbach, J.; Soltis, D.E.; Carpenter, E.J.; Wong, G.K.-S.; Gruber, C.W. Peptidomics of circular cysteine-rich plant peptides: Analysis of the diversity of cyclotides from *Viola tricolor* by transcriptome and proteome mining. *J. Proteome Res.* **2015**, *14*, 4851–4862. [[CrossRef](#)] [[PubMed](#)]
15. Colgrave, M.L.; Kotze, A.C.; Ireland, D.C.; Wang, C.K.; Craik, D.J. The anthelmintic activity of the cyclotides: Natural variants with enhanced activity. *ChemBioChem* **2008**, *9*, 1939–1945. [[CrossRef](#)]
16. Baraguey, C.; Blond, A.; Cavelier, F.; Pousset, J.-L.; Bodo, B.; Auvin-Guette, C. Isolation, structure and synthesis of mahafacyclin B, a cyclic heptapeptide from the latex of *Jatropha mahafalensis*. *J. Chem. Soc. Perkin Trans.* **2001**, *1*, 2098–2103. [[CrossRef](#)]
17. Lipkin, A.; Anisimova, V.; Nikonorova, A.; Babakov, A.; Krause, E.; Bienert, M.; Grishin, E.; Egorov, T. An antimicrobial peptide Ar-AMP from amaranth (*Amaranthus retroflexus* L.) seeds. *Phytochemistry* **2005**, *66*, 2426–2431. [[CrossRef](#)] [[PubMed](#)]
18. Hallock, Y.F.; Sowder, R.C.; Pannell, L.K.; Hughes, C.B.; Johnson, D.G.; Gulakowski, R.; Cardellina, J.H.; Boyd, M.R. Cycloviolins A–D, Anti-HIV Macrocyclic Peptides from *Leonia cymosa*. *J. Org. Chem.* **2000**, *65*, 124–128. [[CrossRef](#)]
19. Blanco-Aparicio, C.; Molina, M.A.; Fernández-Salas, E.; Frazier, M.L.; Mas, J.M.; Querol, E.; Avilés, F.X.; de Llorens, R. Potato carboxypeptidase inhibitor, a T-knot protein, is an epidermal growth factor antagonist that inhibits tumor cell growth. *J. Biol. Chem.* **1998**, *273*, 12370–12377. [[CrossRef](#)]
20. Hellinger, R.; Koehbach, J.; Puigpinós, A.; Clark, R.J.; Tarragó, T.; Giralt, E.; Gruber, C.W. Inhibition of human prolyl oligopeptidase activity by the cyclotide psysol 2 isolated from *Psychotria solitudinum*. *J. Nat. Prod.* **2015**, *78*, 1073–1082. [[CrossRef](#)]

21. Lobo-Ruiz, A.; Tulla-Puche, J. Synthetic approaches of naturally and rationally designed peptides and peptidomimetics. In *Peptide Applications in Biomedicine, Biotechnology and Bioengineering*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 23–49.
22. Wang, C.K.; Gruber, C.W.; Cemazar, M.A.; Siatskas, C.; Tagore, P.; Payne, N.; Sun, G.; Wang, S.; Bernard, C.C.; Craik, D.J. Molecular grafting onto a stable framework yields novel cyclic peptides for the treatment of multiple sclerosis. *ACS Chem. Biol.* **2014**, *9*, 156–163. [[CrossRef](#)]
23. Felizmenio-Quimio, M.E.; Daly, N.L.; Craik, D.J. Circular Proteins in Plants solution structure of a novel macrocyclic trypsin inhibitor from *Momordica cochinchinensis*. *J. Biol. Chem.* **2001**, *276*, 22875–22882. [[CrossRef](#)] [[PubMed](#)]
24. He, W.-J.; Chan, L.Y.; Clark, R.J.; Tang, J.; Zeng, G.-Z.; Franco, O.L.; Cantacessi, C.; Craik, D.J.; Daly, N.L.; Tan, N.-H. Novel inhibitor cystine knot peptides from *Momordica charantia*. *PLoS ONE* **2013**, *8*, e75334. [[CrossRef](#)]
25. Heitz, A.; Hernandez, J.F.; Gagnon, J.; Hong, T.T.; Pham, T.T.; Nguyen, T.M.; Le-Nguyen, D.; Chiche, L. Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins. *Biochemistry* **2001**, *40*, 7973–7983. [[CrossRef](#)]
26. Zhu, S.; Darbon, H.; Dyason, K.; Verdonck, F.; Tytgat, J. Evolutionary origin of inhibitor cystine knot peptides. *FASEB J.* **2003**, *17*, 1765–1767. [[CrossRef](#)]
27. Marzouk, B.; Marzouk, Z.; Haloui, E.; Fenina, N.; Bouraoui, A.; Aouni, M. Screening of analgesic and anti-inflammatory activities of *Citrullus colocynthis* from southern Tunisia. *J. Ethnopharmacol.* **2010**, *128*, 15–19. [[CrossRef](#)] [[PubMed](#)]
28. Abdel-Hassan, I.A.; Abdel-Barry, J.A.; Mohammeda, S.T. The hypoglycaemic and antihyperglycaemic effect of *Citrullus colocynthis* fruit aqueous extract in normal and alloxan diabetic rabbits. *J. Ethnopharmacol.* **2000**, *71*, 325–330. [[CrossRef](#)]
29. Ostovar, M.; Akbari, A.; Anbardar, M.H.; Iraj, A.; Salmanpour, M.; Ghoran, S.H.; Heydari, M.; Shams, M. Effects of *Citrullus colocynthis* L. in a rat model of diabetic neuropathy. *J. Integr. Med.* **2020**, *18*, 59–67. [[CrossRef](#)]
30. Tannin-Spitz, T.; Grossman, S.; Dovrat, S.; Gottlieb, H.E.; Bergman, M. Growth inhibitory activity of cucurbitacin glucosides isolated from *Citrullus colocynthis* on human breast cancer cells. *Biochem. Pharmacol.* **2007**, *73*, 56–67. [[CrossRef](#)] [[PubMed](#)]
31. Saeed, M.E.; Boulos, J.C.; Elhaboub, G.; Rigano, D.; Saab, A.; Loizzo, M.R.; Hassan, L.E.; Sugimoto, Y.; Piacente, S.; Tundis, R. Cytotoxicity of cucurbitacin E from *Citrullus colocynthis* against multidrug-resistant cancer cells. *Phytomedicine* **2019**, *62*, 152945. [[CrossRef](#)]
32. Gruber, C.W.; Elliott, A.G.; Ireland, D.C.; Delprete, P.G.; Dessein, S.; Göransson, U.; Trabi, M.; Wang, C.K.; Kinghorn, A.B.; Robbrecht, E. Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* **2008**, *20*, 2471–2483. [[CrossRef](#)]
33. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **2019**, *574*, 679–685. [[CrossRef](#)] [[PubMed](#)]
34. Cock, P.J.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [[CrossRef](#)]
35. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
36. Crooks, G.; Hon, G.; Chandonia, J.; Brenner, S. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)]
37. Kumar, S.; Stecher, G.; Li, M.; Niyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
38. Clark, R.J.; Daly, N.L.; Craik, D.J. Structural plasticity of the cyclic-cystine-knot framework: Implications for biological activity and drug design. *Biochem. J.* **2006**, *394*, 85–93. [[CrossRef](#)] [[PubMed](#)]
39. Lima, R.M.; Kylarova, S.; Mergaert, P.; Kondorosi, E. Unexplored Arsenal of Legume Peptides With Potential for Their Applications in Medicine and Agriculture. *Front. Microbiol.* **2020**, *11*, 1307. [[CrossRef](#)]
40. Parisi, K.; Shafee, T.M.A.; Quimbar, P.; van der Weerden, N.L.; Bleackley, M.R.; Anderson, M.A. The evolution, function and mechanisms of action for plant defensins. *Semin. Cell Dev. Biol.* **2019**, *88*, 107–118. [[CrossRef](#)]

41. Mahatmanto, T.; Mylne, J.S.; Poth, A.G.; Swedberg, J.E.; Kaas, Q.; Schaefer, H.; Craik, D.J. The evolution of Momordica cyclic peptides. *Molecular Biol. Evol.* **2015**, *32*, 392–405. [[CrossRef](#)]
42. Giacomelli, L.; Nanni, V.; Lenzi, L.; Zhuang, J.; Serra, M.D.; Banfield, M.J.; Town, C.D.; Silverstein, K.A.; Baraldi, E.; Moser, C. Identification and characterization of the defensin-like gene family of grapevine. *Mol. Plant Microbe Interact.* **2012**, *25*, 1118–1131. [[CrossRef](#)] [[PubMed](#)]
43. Gassmann, W.; Hinsch, M.E.; Staskawicz, B.J. The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant J.* **1999**, *20*, 265–277. [[CrossRef](#)]
44. Jordan, T.; Schornack, S.; Lahaye, T. Alternative splicing of transcripts encoding Toll-like plant resistance proteins—what’s the functional relevance to innate immunity? *Trends Plant Sci.* **2002**, *7*, 392–398. [[CrossRef](#)]
45. Lawrence, G.J.; Finnegan, E.J.; Ayliffe, M.A.; Ellis, J.G. The L6 gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene RPS2 and the tobacco viral resistance gene N. *Plant Cell* **1995**, *7*, 1195–1206.
46. Moran, Y.; Weinberger, H.; Reitzel, A.M.; Sullivan, J.C.; Kahn, R.; Gordon, D.; Finnerty, J.R.; Gurevitz, M. Intron retention as a posttranscriptional regulatory mechanism of neurotoxin expression at early life stages of the starlet anemone *Nematostella vectensis*. *J. Mol. Biol.* **2008**, *380*, 437–443. [[CrossRef](#)]
47. Schornack, S.; Ballvora, A.; Gürlebeck, D.; Peart, J.; Ganai, M.; Baker, B.; Bonas, U.; Lahaye, T. The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant J.* **2004**, *37*, 46–60. [[CrossRef](#)] [[PubMed](#)]
48. Favel, A.; Matras, H.; Coletti-Previero, M.; Zwilling, R.; Robinson, E.; Castro, B. Protease inhibitors from *Ecballium elaterium* seeds. *Int. J. Pept. Protein Res.* **1989**, *33*, 202–208. [[CrossRef](#)]
49. Hara, S.; Makino, J.; Ikenaka, T. Amino acid sequences and disulfide bridges of serine proteinase inhibitors from bitter melon (*Momordica charantia* LINN.) seeds. *J. Biochem.* **1989**, *105*, 88–92. [[CrossRef](#)]
50. Hayashi, K.; Takehisa, T.; Hamato, N.; Takano, R.; Hara, S.; Miyata, T.; Kato, H. Inhibition of serine proteases of the blood coagulation system by squash family protease inhibitors. *J. Biochem.* **1994**, *116*, 1013–1018. [[CrossRef](#)]
51. Hernandez, J.F.; Gagnon, J.; Chiche, L.; Nguyen, T.M.; Andrieu, J.P.; Heitz, A.; Trinh Hong, T.; Pham, T.T.; Le Nguyen, D. Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure. *Biochemistry* **2000**, *39*, 5722–5730. [[CrossRef](#)] [[PubMed](#)]
52. Hass, G.M.; Nau, H.; Biemann, K.; Grahn, D.T.; Ericsson, L.H.; Neurath, H. Amino acid sequence of a carboxypeptidase inhibitor from potatoes. *Biochemistry* **1975**, *14*, 1334–1342. [[CrossRef](#)] [[PubMed](#)]
53. Liu, Y.; Luo, J.; Xu, C.; Ren, F.; Peng, C.; Wu, G.; Zhao, J. Purification, characterization, and molecular cloning of the gene of a seed-specific antimicrobial protein from pokeweed. *Plant Physiol.* **2000**, *122*, 1015–1024. [[CrossRef](#)] [[PubMed](#)]
54. Kowalska, J.; Pszczola, K.; Wilimowska-Pelc, A.; Lorenc-Kubis, I.; Zuziak, E.; Lugowski, M.; Legowska, A.; Kwiatkowska, A.; Sleszynska, M.; Lesner, A.; et al. Trypsin inhibitors from the garden four o’clock (*Mirabilis jalapa*) and spinach (*Spinacia oleracea*) seeds: Isolation, characterization and chemical synthesis. *Phytochemistry* **2007**, *68*, 1487–1496. [[CrossRef](#)]
55. Chagolla-Lopez, A.; Blanco-Labra, A.; Patthy, A.; Sánchez, R.; Pongor, S. A novel alpha-amylase inhibitor from amaranth (*Amaranthus hypocondriacus*) seeds. *J. Biol. Chem.* **1994**, *269*, 23675–23680. [[PubMed](#)]
56. Chou, M.-X.; Wei, X.-Y.; Chen, D.-S.; Zhou, J.-C. Thirteen nodule-specific or nodule-enhanced genes encoding products homologous to cysteine cluster proteins or plant lipid transfer proteins are identified in *Astragalus sinicus* L. by suppressive subtractive hybridization. *J. Exp. Bot.* **2006**, *57*, 2673–2685. [[CrossRef](#)]
57. McBride, J.D.; Watson, E.M.; Brauer, A.B.; Jaulent, A.M.; Leatherbarrow, R.J. Peptide mimics of the Bowman–Birk inhibitor reactive site loop. *Pept. Sci. Orig. Res. Biomol.* **2002**, *66*, 79–92. [[CrossRef](#)]
58. Huang, J.; Wong, K.H.; Tay, S.V.; Serra, A.; Sze, S.K.; Tam, J.P. Astratides: Insulin-Modulating, Insecticidal, and Antifungal Cysteine-Rich Peptides from *Astragalus membranaceus*. *J. Nat. Prod.* **2019**, *82*, 194–204. [[CrossRef](#)]
59. Kam, A.; Loo, S.; Fan, J.S.; Sze, S.K.; Yang, D.; Tam, J.P. Roseltide rT7 is a disulfide-rich, anionic, and cell-penetrating peptide that inhibits proteasomal degradation. *J. Biol. Chem.* **2019**, *294*, 19604–19615. [[CrossRef](#)]
60. Molesini, B.; Treggiari, D.; Dalbeni, A.; Minuz, P.; Pandolfini, T. Plant cystine-knot peptides: Pharmacological perspectives. *Br. J. Clin. Pharmacol.* **2017**, *83*, 63–70. [[CrossRef](#)]

61. Shelenkov, A.; Slavokhotova, A.; Odintsova, T. Predicting Antimicrobial and Other Cysteine-Rich Peptides in 1267 Plant Transcriptomes. *Antibiotics (Basel)* **2020**, *9*, 60. [[CrossRef](#)]
62. Lavergne, V.; Taft, R.J.; Alewood, P.F. Cysteine-rich mini-proteins in human biology. *Curr. Top. Med. Chem.* **2012**, *12*, 1514–1533. [[CrossRef](#)] [[PubMed](#)]
63. Gruber, C.W.; Muttenthaler, M. Discovery of defense- and neuropeptides in social ants by genome-mining. *PLoS ONE* **2012**, *7*, e32559. [[CrossRef](#)] [[PubMed](#)]
64. Koehbach, J.; Craik, D.J. The Vast Structural Diversity of Antimicrobial Peptides. *Trends Pharmacol. Sci.* **2019**, *40*, 517–528. [[CrossRef](#)]
65. de Veer, S.J.; Wang, C.K.; Harris, J.M.; Craik, D.J.; Swedberg, J.E. Improving the Selectivity of Engineered Protease Inhibitors: Optimizing the P2 Prime Residue Using a Versatile Cyclic Peptide Library. *J. Med. Chem.* **2015**, *58*, 8257–8268. [[CrossRef](#)] [[PubMed](#)]
66. Legowska, A.; Debowski, D.; Lesner, A.; Wysocka, M.; Rolka, K. Introduction of non-natural amino acid residues into the substrate-specific P1 position of trypsin inhibitor SFTI-1 yields potent chymotrypsin and cathepsin G inhibitors. *Bioorg. Med. Chem.* **2009**, *17*, 3302–3307. [[CrossRef](#)]
67. Gruber, C.W.; Cemazar, M.; Anderson, M.A.; Craik, D.J. Insecticidal plant cyclotides and related cystine knot toxins. *Toxicon* **2007**, *49*, 561–575. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).