

# Synthesis, structural studies, molecular modeling and characterization of helminth antimicrobial peptides

---

Čopac, Roko

Master's thesis / Diplomski rad

2022

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Split, University of Split, Faculty of science / Sveučilište u Splitu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:166:228687>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-11-15**

*Repository / Repozitorij:*

[Repository of Faculty of Science](#)



University of Split  
Faculty of Science

**Synthesis, structural studies, molecular  
modeling, and characterization of helminth  
antimicrobial peptides**

Master Thesis

Roko Čopac

Split, March 18<sup>th</sup>, 2022.

## Temeljna dokumentacijska kartica

Sveučilište u Splitu  
Prirodoslovno – matematički fakultet  
Odjel za fiziku  
Ruđera Boškovića 33, 21000 Split, Hrvatska

Diplomski rad

### Sinteza, strukturne studije, molekularno modeliranje i karakterizacija antimikrobnih peptida u plošnjaka

Roko Čopac

Sveučilišni diplomski studij Fizika, smjer Biofizika

#### Sažetak:

Rezistencija bakterija na antibiotike, za koju je djelomično zaslužna neodgovorna i pretjerana uporaba, gorući je problem diljem svijeta u zadnjim desetljećima. Kao potencijalna alternativa klasičnim antibioticima nameću se antimikrobni peptidi (AMP), strukturno različite makromolekule koje imaju direktnu antimikrobnu i/ili imunomodulatornu aktivnost. Plošnjaci, skupina nametničkih organizama, uspješno ko-egzistira u okolini (najčešće) bogatoj bakterijama, a jedan od mogućih mehanizama obrane ovih organizama su AMP-i. Analizom genoma većeg broja plošnjaka, identificirali smo 31 AMP, te nakon detaljne bioinformatičke analize odabrali tri peptida za daljnje istraživanje. Istovremeno, u cilju ispitivanja utjecaja pojedinih segmenata peptidne sekvence na aktivnost i mehanizam djelovanja, sintetizirani su N- i C-fragmenti peptida identificiranog u *Taenia solium*. Svi odabrani peptidi, kao i fragmenti, karakterizirani su putem strukturalnih, antimikrobnih i citotoksičnih ispitivanja, te istraženi na molekularnoj razini korištenjem simulacija molekularne dinamike. Većina peptida pokazuje antimikrobnu aktivnost širokog spektra, dok je toksičnost često u prihvatljivim okvirima, što otvara vrata daljnjim ispitivanjima mehanizama djelovanja, kao i potencijalnoj biomedicinskoj uporabi.

**Ključne riječi:** antimikrobni peptidi, antimikrobna rezistencija, mehanizam djelovanja, plošnjaci

**Rad sadrži:** 63 stranice, 46 slika, 12 tablica, 75 literaturnih navoda, Izvornik je na engleskom

**Mentori:** izv. prof. dr. sc. Larisa Zoranić

**Neposredni voditelji:** dr. sc. Tomislav Rončević  
dr.sc. Andrea Caporale  
dr.sc. Sabrina Pacor

**Ocjenjivači:** izv. prof. dr. sc. Larisa Zoranić  
dr. sc. Tomislav Rončević  
doc. dr. sc. Željka Sanader Maršić  
dr.sc. Lucija Krce  
doc. dr. sc. Ivica Šamanić

**Rad prihvaćen:** 18.03.2022

Rad je pohranjen u knjižnici Prirodoslovno – matematičkog fakulteta, Sveučilišta u Splitu.

## Basic documentation card

University of Split  
Faculty of Science  
Department of Physics  
Ruđera Boškovića 33, 21000 Split, Croatia

Master thesis

### **Synthesis, structural studies, molecular modeling, and characterization of helminth antimicrobial peptides**

Roko Čopac

University graduate study programme Physics, orientation Biophysics

#### **Abstract:**

Antimicrobial resistance, which can be partially attributed to irresponsible and improper use of antibiotics, remains a major concern worldwide in recent decades. One of the potential alternatives to classical antibiotic treatment are antimicrobial peptides (AMPs) - structurally different (macro)molecules with direct antimicrobial and/or immunomodulatory properties. Helminths, a group of parasitic organisms, successfully co-exist in bacteria-rich surroundings during their life cycle, and among the possible mechanisms of defense of these organisms are likely be AMPs. By exploring the genomes of several helminth species, 31 AMPs were obtained, after which a bioinformatics analysis was performed and 3 were selected for further characterization. Moreover, several N- and C-terminal fragments of a peptide identified in *Taenia solium* were synthesized to conduct structure-activity studies. All selected peptides, as well as the fragments, were characterized by structural, antimicrobial, and cytotoxic tests, and modeled by molecular dynamics simulations. Most of the peptides showed broad-spectrum antimicrobial activity, with (mostly) acceptable levels of toxicity, which warrants further mode-of-action studies, and development for possible biomedical applications.

**Keywords:** antimicrobial peptides, antimicrobial resistance, mode of action, helminths

**Thesis consists of:** 63 pages, figures, 12 tables, 75 reference, Original language: English

**Supervisors:** Assoc. prof. Larisa Zoranić

**Assistant Supervisor:** dr. Tomislav Rončević  
dr. Andrea Caporale  
Assist. Prof. Sabrina Pacor

**Reviewers:** Assoc. prof. dr. sc. Larisa Zoranić  
dr. Tomislav Rončević  
Assist. Prof. Željka Sanader Maršić  
dr. Lucija Krce  
Assist. Prof. Ivica Šamanić

**Thesis accepted:** March 18<sup>th</sup>, 2022

Thesis is deposited in the library of the Faculty of Science, University of Split.

## STATEMENT

by which I declare with full material and moral responsibility, that I have made a thesis entitled “Synthesis, structural studies, molecular modelling and characterization of helminth antimicrobial peptides” independently under the guidance of Prof. Alessandro Tossi, Assoc. Prof. Larisa Zoranić and Dr. Tomislav Rončević. I have applied the methodology of scientific research and used the literature that is listed at the end of the thesis. Other people's knowledge, attitudes, conclusions, and theories that I stated directly or paraphrasing in my thesis in the usual, standard way, I quoted and connected with footnotes with the used bibliographic units. The paper is written in the spirit of the English language.

Student  
Roko Čopac

## ACKNOWLEDGMENT

*I would like to thank my mentors: Assoc. Prof. Larisa Zoranić and Dr. Tomislav Rončević, for their accessibility, devotion, numerous discussions, huge support, and patience as I wrote this thesis. Thank you for introducing me to the fascinating world of antimicrobial peptides and inspiring me with your knowledge and passion.*

*I am very much grateful to Prof. Alessandro Tossi, who welcomed me in his laboratory and, together with dr. Andrea Caporale, helped me in my professional and personal development. Thank you both for all the lessons and for making me feel like at home.*

*I would also like to express my gratitude to Assoc. Prof. Sabrina Pacor, who introduced me to the world of cytotoxicity and to all my friends at the Department of Life Sciences of the University of Trieste, for making my stay unforgettable.*

*In the end, I would like to thank my family and friends for all the support and patience. Thank you for believing in me.*

## Table of content

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1.1	Classification of AMPs.....	2
1.1.2	Biophysical properties of antimicrobial peptides .....	2
1.1.3	Mechanism of action of AMPs.....	3
1.2	Helminth peptides .....	5
<b>2</b>	<b>Aims and scope of the thesis.....</b>	<b>7</b>
<b>3</b>	<b>Materials and methods .....</b>	<b>8</b>
3.1	Bioinformatical analysis .....	8
3.1.1	Sequence data .....	8
3.1.2	Peptide analysis .....	9
3.2	Solid phase peptide synthesis .....	10
3.2.1	Fmoc based SPPS .....	11
3.2.2	Synthesis of TSO8 fragments .....	12
3.3	Structural analysis.....	13
3.3.1	Circular dichroism spectroscopy .....	13
3.3.2	Experimental set up .....	15
3.4	Antimicrobial activity assays.....	16
3.4.1	Experiment set up .....	16
3.4.2	MIC assay .....	16
3.4.3	MBC assay.....	18
3.5	Cytotoxicity assay – membrane permeabilization .....	18
3.6	Molecular modelling.....	18
3.6.1	Molecular dynamics simulation .....	18
3.6.2	Simulation setup .....	20
<b>4</b>	<b>Results and discussion .....</b>	<b>22</b>
4.1	Peptide identification and bioinformatical analysis.....	22

4.2	Structural studies .....	28
4.3	Antimicrobial activity .....	35
4.4	Cytotoxicity assay .....	36
4.4	Molecular modeling .....	39
<b>5</b>	<b>Conclusion .....</b>	<b>49</b>
<b>6</b>	<b>Literature .....</b>	<b>51</b>
<b>7</b>	<b>Appendix A .....</b>	<b>57</b>
<b>8</b>	<b>Appendix B .....</b>	<b>59</b>
<b>9</b>	<b>Appendix C .....</b>	<b>60</b>
<b>10</b>	<b>Appendix D .....</b>	<b>63</b>



## 1 Introduction

In recent decades, an alarmingly high rate of antibiotic resistance development in bacteria has resulted in millions of people infected and many thousands dying worldwide every year (1–3). During the golden age of antibiotics in the 1960s and ‘70s, new antibiotics were identified and put into clinical use with relative ease. Nowadays, however, few novel antibiotics enter the developmental pipeline, urging the need for alternative therapeutics. For this reason, World Health Organization (WHO) has appointed antimicrobial resistance as among the top ten threats for humanity (1). Novel antimicrobial drugs are needed to address this issue and some of those include plant-derived compounds, bacteriophages, RNA-based therapeutics, nanomaterials and antimicrobial peptides (2). In fact, some antibiotics in clinical use today, such as colistin and vancomycin, have a partially peptidic structure (2).

Antimicrobial peptides (AMP), or host defense peptides (HDP) as they are sometimes called, can be defined as multifunctional, and generally gene encoded effector molecules produced by almost all organisms, having a direct antimicrobial activity and/or immunomodulatory properties (4). In that sense they can be considered as an ancient innate immunity weapon produced by almost all organisms to fight pathogens such as bacteria, viruses, fungi, parasites, etc. (4–16).

To this date, an abundance of AMPs has been identified in all eukaryotic organisms and these data can be easily accessed from several devoted online databases such as: Antimicrobial Peptide Database (17), Collection of Anti-Microbial Peptides (CAMP<sup>R3</sup>) (18), Database of Anuran Defense Peptides (DADP) (19), Defensins (20), etc.

In 2014, about 2500 AMPs were reported in Antimicrobial Peptide Database (7), whilst in September 2021 the same database contained 3273 AMP peptide sequences, which emphasizes the rate of AMP research (17). On the other hand, the CAMP<sup>R3</sup> database in September 2021 contained 8164 entries for peptides (19) most of which were of animal origin (73.6%) (see Figure 1) (4).

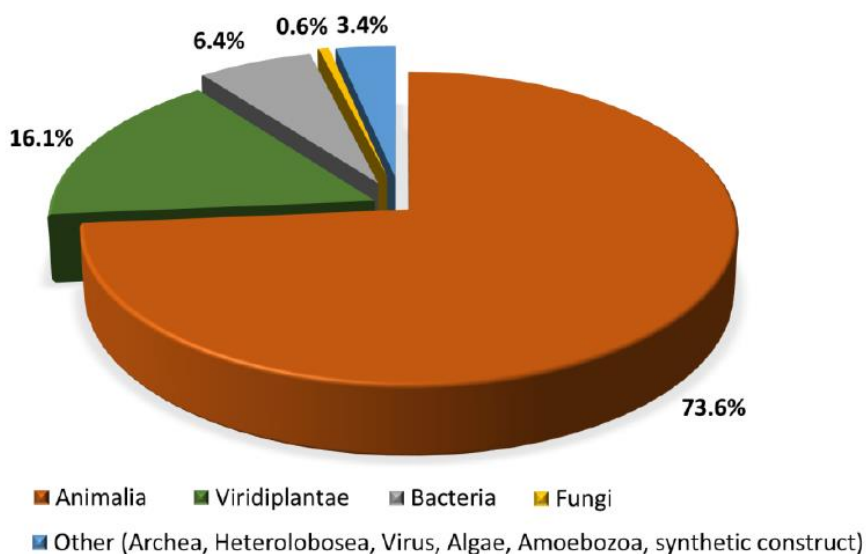


Figure 1 Distribution of antimicrobial peptides across kingdoms in CAMP<sup>R3</sup> database. Figure taken from Roncevic et al. (4)

### 1.1.1 Classification of AMPs

Antimicrobial peptides can be classified in several different ways (21). For example, classification according to source organism is based on origin (animal, fungi, insects, etc.). Classification according to activity can be based on whether AMPs are active against Gram positive or negative or both types of bacteria. Peptides can also be classified by the presence of some dominant amino acids in sequence. Another classification is on secondary structure diversity, where AMPs can be found in a form predominantly  $\alpha$ -helices,  $\beta$ -sheets,  $\beta$ -hairpins, random coils or different combinations of these (4,9). Although differing in secondary structure, most AMPs have an amphipathic design, which means peptides form topologically distinct hydrophobic and charged domains (8).

### 1.1.2 Biophysical properties of antimicrobial peptides

As mentioned, there are several important biophysical properties that contribute to antimicrobial peptide's activity against microbes. Generally, AMPs are normally, less than fifty amino acids long peptides, with a charge between -6 to +16 (4), the majority having a net positive charge of around +6. Positive charge is important for the peptide's initial contact with the negatively charged external leaflet of the outer membrane, which is a crucial step in microbe-killing mechanism (4,12,16,22,23). It is reported that a change of the net charge of AMPs can alter selectivity (12,13,23). Another important property of AMPs, which can impact antimicrobial activity and selectivity is hydrophobicity. Hydrophobicity can be defined

as numerical value quantifying a dislike for water for each amino acid sidechain (9). About 40–50% of residues in the primary sequence of natural AMPs are hydrophobic, which facilitates peptide to interact with lipid bilayer.

However, there is evidence that amphipathicity is even more important for peptides incorporation into microbial membranes than hydrophobicity (12,15,23). This property refers to the topographic distribution of hydrophobic and polar residues within the peptide sequence, which results in a formation of distinct hydrophobic and charged domains of peptide structure (4). Therefore, amphipathicity should be a priority in designing novel antimicrobial peptide sequences (23).

Other important properties that contribute to AMP's activity are length, solubility, and degree of conformational stability (e.g., helicity). For instance, at least seven to eight amino acid residues are needed to form amphipathic structure, whilst aggregation of certain AMPs results in losing the ability to interact with the cell membrane. Lastly, helicity or helical structuring is another important peptide property which seems to correlate more with toxicity than with antimicrobial activity (23).

### **1.1.3 Mechanism of action of AMPs**

Since their initial discovery, AMPs have been extensively studied to try and elucidate their mechanism of action and behavior upon interaction with host and bacteria cells. Cationic nature of most AMPs allows them to interact with negatively charged bacterial membrane which may cause membrane permeabilization, as the most common mode of action (24). Also, the fact that their mode of action is aimed at the membrane is critical for AMP's low tendency to elicit resistance, as bacteria can't easily change membrane lipid composition (25).

Several modes of action have been suggested for different peptides, although none with complete certainty. However, based on the peptide's behavior, AMPs are nowadays classified into *i*) membrane-active and *ii*) non-lytic (4). There are some examples in which AMPs can act using both of these two major mechanisms or even switching from one to the other, depending on the peptide concentration, membrane characteristics of a particular bacterial species or its growth phase (4,9).

Concerning mechanism of action of membrane-active peptides, the initial steps are often the same, regardless of the specific mechanism of action. These include *i*) initial attraction and interaction with the membrane, *ii*) adapting of an active structure and *iii*) concentration-

dependent accumulation on/in the bilayer (see Figure 2) (24). Several membrane-active mechanisms of action have been proposed and those include:

- a) Barrel-stave model predicts AMPs' insertion into the bilayer by orienting their hydrophobic regions in the lipid core of the bilayer, forming a barrel-like transmembrane pore. Relatively few peptides have been found to utilize this mode of action (i.e., pardaxin and alamethicin) (4,9,23,26).
- b) Toroidal or "wormhole" model predicts AMPs to insert within the membrane and cause the lipids within the bilayer to bend, forming a peptide-lined pore in which the AMPs associate with the polar head groups of the phospholipids, such is the case of magainin 2 (4,9,23,26).
- c) Carpet model or "detergent like model" predicts that accumulation of peptides on the surface induces local weaknesses and breaks down the bilayers into small areas lined by the peptide units, emulating a carpet-like appearance (4,9,23,26).
- d) Membrane thinning model predicts AMP's insert themselves into one leaflet of the lipid bilayer. This can result in a gap between lipid molecules at the chain region, which creates a force and pulls the neighboring lipid molecules to fill it (23).
- e) Aggregation model predicts that AMPs stick to the membrane parallel to the surface. AMPs reorient and insert into the membrane vertically to form sphere-like structures and micellization of the membrane (4,9,23,26).
- f) Other examples of possible AMPs mechanisms like depolarization or fusion, electroporation, and targeting of specific phospholipids are less disruptive (4).

Non-lytic peptides also act by different mechanisms, most of which include incubation of different intracellular (rarely extracellular) targets and processes (4,9). Examples of such activities include inhibition of: *i*) enzymatic activity, *ii*) protein biosynthesis (bovine cathelicidin Bac7), *iii*) protein folding (pyrrhocoricin), *iv*) cell division (CRAMP), *v*) nucleic acids replication or transcription (buforin 2), *vi*) cell wall biosynthesis (defensins), etc. (4,9,24).

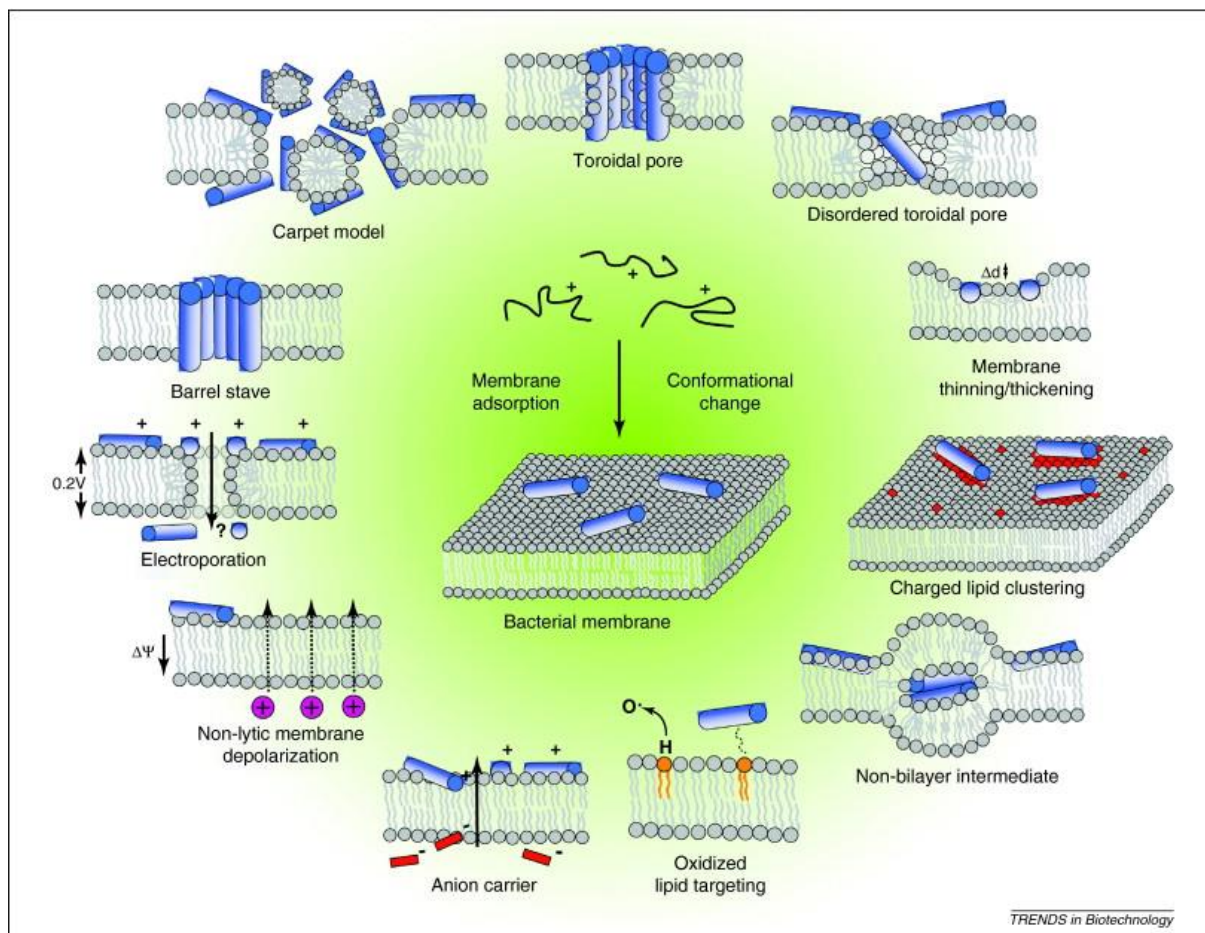


Figure 2 Possible interaction events of AMPs with bacterial membrane bilayer. Proposed models are not necessarily exclusive of each other. Figure 2 taken from Nguyen et al. (24).

## 1.2 Helminth peptides

During their complex life-cycle, parasitic worms or helminths coexist with the host, which sometimes can lead to development of severe disease or even death (27). It is estimated that about one-sixth of the world population is affected by these organisms (28) and although helminths are present worldwide, infection is most common in developing countries due the lack of hygiene, climate and diet (contaminated food or water). Helminth infection can be transmitted in several ways, such as ingestion of eggs or larvae, through vectors or ingestion through food.

Helminths can be divided into two major groups: flatworms (flukes, tapeworms) and roundworms. Flattened, leaf-shaped flukes can range in length from a few millimeters to 8 cm. Tapeworms can vary in length from 2 mm to 10 m, while roundworms can exceed more than 15 cm (29,30). In infected humans, majority of helminth organisms can usually be found in bowels (28), which is a complex environment containing commensal, symbiotic and

pathogenic bacteria, as well as fungi, viruses and archaea (31,32). Although it is generally accepted that helminths have immunomodulatory properties (31), which means that they can alter intestinal physiology, permeability, mucous secretion and the production of antimicrobial peptides (33), helminth/bacteria interaction is largely unexplored (31,33). It is rather obvious that helminths share their microenvironment with pathogens without being greatly affected by them, although the exact mechanisms underlying this are still not elucidated. A possible answer could lay in their ability to produce AMPs' as one way to fight control pathogens and could, therefore, be considered as candidates for identification and development of potential leads for antimicrobial therapeutics.

## 2 Aims and scope of the thesis

To study the structural and functional characteristics of three antimicrobial peptides identified in three helminth species: *Taenia solium*, *Taenia multiceps* and *Mesocestoides corti*.

This includes identification of AMPs in several helminths, selection for synthesis of selected peptides, structural studies using circular dichroism (CD) spectroscopy in different environments, determination of antimicrobial activity [minimal inhibitory concentration (MIC) and minimal bactericidal concentration (MBC)] towards indicative Gram-negative (*E. coli*) and Gram-positive (*S. aureus*) strains, toxicity towards human cell lines and molecular modeling by means of molecular dynamics (MD) simulations.

Moreover, one peptide, identified in *T. solium*, was selected for more extensive structure-activity studies. Alongside the full mature peptide sequence, specific fragments of this peptide were also synthesized and characterized with the aim of understanding the role of certain amino acids (i.e., tryptophan) and peptide regions (N- or C-terminal fragments) in determining the biological activity.

## 3 Materials and methods

### 3.1 Bioinformatical analysis

#### 3.1.1 Sequence data

Protein translations computationally generated from annotated gene models in the *Taenia solium* genome version Tsolium\_Mexico\_v1 (34) were downloaded from WormBase (35). These were screened with an in-house script developed at the Department of Life Sciences of the University of Trieste, aimed at identifying candidate potential AMPs, using a *de novo* approach, as previously done in the case of myticalins in *Mytilus galloprovincialis* (36). Briefly, protein sequences were considered for further analysis only if they had *i*) a length compatible with the usual size of an AMP precursor (i.e. < 200 amino acids); *ii*) included a highly supported signal peptide for secretion detected with SignalP (37); *iii*) had a net cationic charge. The presence of a previously characterized amphipathic alpha helix signature typically associated with peptides with antibacterial action (G-h-h-p-p-h-x-p-x-h-p-p-h-x-p-p-h-x-p-x, where h indicates hydrophobic; p indicates polar; and x indicates undefined) (38) was verified in resulting peptides. This approach allowed the initial identification of three candidate peptides, which showed a significant pairwise primary sequence similarity, indicating a shared evolutionary origin.

These sequences were used as queries for additional BLASTp searches to identify other similar peptides belonging to the same family that did not comply with the aforementioned selection criteria, using a p-value threshold of 0.05. Moreover, tBLASTn searches were carried out both against the genome assembly and the *T. solium* transcriptome (*de novo* assembled and based on the reads included in the SRA submission SRX1899230 using the CLC Genomics Workbench v.12), to ensure the detection of other sequences that may have been missing in the current version of the genome annotation. The *T. solium* peptides identified as mentioned above were then used as queries in BLASTp searches carried out against the protein sets derived from the annotations of the genomes available for other *Cestoda* species. Namely, the selected species were *Taenia multiceps* (version ASM192302v3), *Hidratigera taeniformis* (version H\_taeniaeformis\_Canary\_Islands\_0011\_upd), *Mesocestoides corti* (version M\_corti\_Specht\_Voge\_0011\_upd), *Echinococcus granulosus* (version EGRAN001), *Echinococcus multilocularis* (version EMULTI002) and *Hymenolepis diminuta* (version H.diminuta\_WMSil1), all retrieved from Wormbase (35). Data from *Moniezia expansa* (39)



and *Rodentolepsis nana* (40) were retrieved from the respective reference studies. The sequence alignment between all hits characterized by an e-value lower than 0.05 and query sequences were manually screened to remove false positives and only sequences denoting full-length proteins (i.e., those encoded by a complete ORF, including both an ATG start codon and a stop codon) were further considered.

### 3.1.2 Peptide analysis

Bioinformatical analysis of mature peptide sequences (see Appendix A) we performed in order to select suitable candidates for further characterization. This included structural prediction of each peptide, its biophysical properties and finally, biological origin and diversity.

Predictions of secondary and tertiary structures were obtained with PSIPRED and Contact-assisted QUARK (cQUARK)/ PEP-FOLD 3.5/ Phyre<sup>2</sup>, respectively (41–44), while net charges were calculated with BACHEM peptide calculator (45). PSIPRED software is based on neural networks which performs an analysis on output obtained from PSI-BLAST (Position Specific Iterated - BLAST) (41). The cQUARK is a method for *ab initio* protein structure prediction based on fragment assembly simulations (42), PEP-FOLD 3.5 is a *de novo* method for peptide structural prediction based on structural alphabet SA letters (43), while Protein Homology/AnalogY Recognition Engine (Phyre<sup>2</sup>) is a program which determines structure based on the protein threading method (44).

Peptide molecular diversity was screened with Basic Local Alignment Search Tool (BLAST (46)). In general, BLAST is used to find regions of similarity between nucleotide or protein query sequences, comparing them with sequence databases, calculating statistical significance and assigning a score (46). Each sequence was “blasted” using the “BLASTp” method (compares a protein query to a protein database) to search the non-redundant protein database, the non-redundant protein database filtered for only insect sequences and Protein Data Bank protein structure database. Moreover, we also searched translated nucleotide sequence using a protein query (tBLASTn) against the whole genome shotgun contigs (wgs) database (46). Algorithm parameters used for BLASTp and tBLASTn search are given in Figure 3.

The image shows a screenshot of the BLAST algorithm parameters interface. It is organized into three main sections: General Parameters, Scoring Parameters, and Filters and Masking. The General Parameters section includes: Max target sequences (100), Short queries (checked: Automatically adjust parameters for short input sequences), Expect threshold (0.05), Word size (6), and Max matches in a query range (0). The Scoring Parameters section includes: Matrix (BLOSUM62), Gap Costs (Existence: 11 Extension: 1), and Compositional adjustments (Conditional compositional score matrix adjustment). The Filters and Masking section includes: Filter (Low complexity regions) and Mask (Mask for lookup table only, Mask lower case letters), all of which are currently unchecked.

Figure 3 Algorithm parameters used for BLAST search.

### 3.2 Solid phase peptide synthesis

Unlike peptide extraction from natural sources, which generally requires a lot of raw biological sample, solid phase peptide synthesis (SPPS) provides peptides of at high yield and purity. Furthermore, SPPS allows relatively easy modification of original sequence, which is very important when studying a peptides' structure-activity relationships. In the AMPs research field, SPPS quickly became the preferred method to obtain peptides with high quality and to introduce several modifications to optimize antimicrobial features (26). In SPPS, the sequence is synthesized starting from C-terminal amino acids to N-terminal amino acids; the direction of the synthesis is inverted with respect to biological synthesis. The first amino acid is linked to a resin, with different reactions that result in a different functionalization in C-terminal; amide, carboxyl or alcohol. The resin is complex composed of matrix polymer (bead) and linker, which provides solid support for chain elongation. A main feature of the resin is the loading capacity (the number of active sites available per unit of weight in grams).

The reaction is carried out in a homogeneous phase microscopically and in heterogeneous phase macroscopically. For these reasons, it is possible to use an excess of reagents to increase yield of the single reaction steps up 99%. Following the first reaction, the amine group is deprotected from the protecting group on the N-terminal  $\alpha$ -amino group of the inserted residue, under acid or basic conditions, depending on the SPPS chemistry used. The subsequent incoming amino acid in the sequence is added with an appropriate coupling

reaction. Coupling is the process of forming the peptide bond using an active ester on the incoming residue, generated in situ. After the whole sequence has been synthesized, the peptide is released from the solid support (resin) generally by applying very acid conditions, in a process known as called cleavage (9,26,47). SPPS avoids the intermediate purification steps typically required by liquid phase (solution) synthesis, increasing the speed of synthesis. On the other hand, SPPS is not cheap, so that sometimes liquid phase synthesis is preferred in industry for large scale production of peptides especially if they are small (9), although new approaches and methodologies to purify the crude peptide have been developed. In our case, we have prepared TSO8 fragments: TSO8 (1-23), TSO8 (3-23) and TSO8 (23-39), while TSO8, TSO8 (3-39), MESCO 2 and TAEMU 1 were commercially synthesized by SPPS by GenicBio (Shanghai, China).

Prior to synthesis the TSO8 fragments were analyzed using the Peptide Companion Software to identify regions of potential synthetic difficulty (see Figure 4). On y-axes, values between 0.6-0.8 represent relatively easy coupling (colored green), whereas values  $> 0.8$  are associated with more difficult process (color yellow) (48). The value expressed on y-axis is a statistical average of the propensity of single residue to aggregate during the synthesis of growing-up peptide.

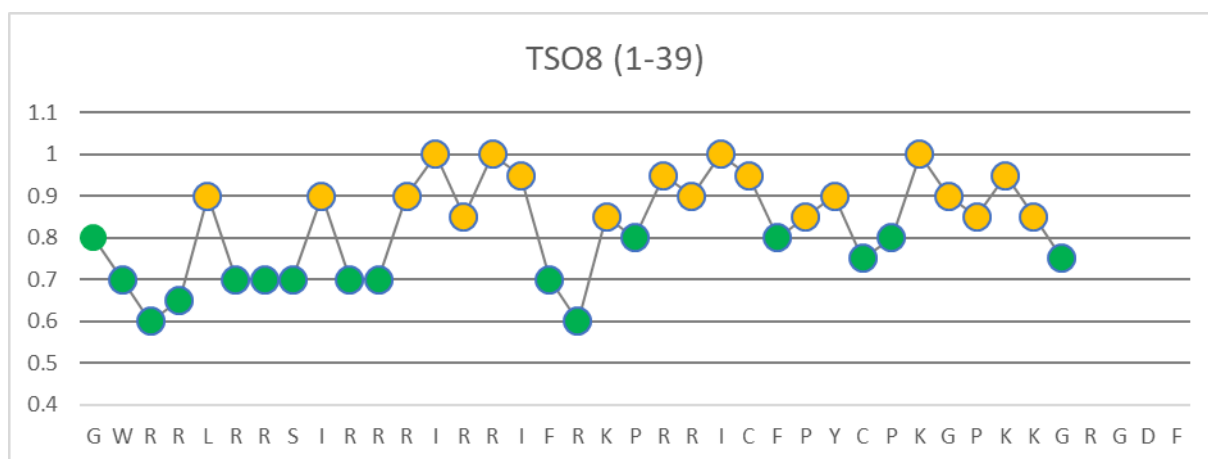


Figure 4 Prediction of synthesis difficulty by Peptide Companion Software (48) for TSO8 (1-39)

### 3.2.1 Fmoc based SPPS

In our synthetic approach, the amino acids used in SPPS usually contain a protection on the  $\alpha$ -amino group that can be removed in basic condition. Generally, the protocols used in SPPS are based on the particular chemistry used for  $\alpha$ -amino group protection, either Boc (tert-butyloxycarbonyl) or Fmoc (fluorenylmethoxycarbonyl) molecules. The Boc method was that originally proposed by Merrifield in 1963, but the Boc chemistry requires use of highly toxic

hydrofluoric acid for the final cleavage from the solid support, meanwhile trifluoroacetic acid is used to remove the protective group at each step. Fmoc-based SPPS is now the method of choice for routine synthesis of peptides because cleavage from resin is carried out using trifluoroacetic acid, while the amino terminal protective group is removed using a piperidine solution that is relatively easy to handle (47). This chemistry uses “orthogonal” protection, as the amino protecting group is removed by base while deprotection of sidechains and release from the resin is simultaneously effected by acid treatment.

### 3.2.2 Synthesis of TSO8 fragments

For synthesis of the N-terminal fragments, Fmoc Rink Amide MBHA resin (loading 0.56 mmol/g, 40 mg) (Novabiochem, Darmstad, Germany) was used, while for synthesis of C-terminal fragments Fmoc-Rink AmideAM-PEG (loading 0.19 mmol/g, 70 mg) (IRIS BioTech, Marktredwitz, Germany) was used. Synthesis was carried out with microwave synthesizer (Biotage, Sweden), the power applied is measured using temperature in Celsius. Resin swelling was ensured by Dimethylformamide (DMF) and N-methyl pyrrolidone (NMP) (IRIS BioTech, Marktredwitz, Germany). The use of NMP is recommended to avoid aggregation phenomena during the synthesis of peptide. Each deprotection was done using 20% piperidine in DMF. Activation chemical for coupling reaction was 1-hydroxybenzotriazole (HBTU) (IRIS BioTech, Marktredwitz, Germany), acetylation chemical was acetic anhydride (Ac<sub>2</sub>O) (Sigma-Aldrich, St. Louis, Missouri, USA). Lastly, the cleavage mixture was prepared using trifluoroacetic acid (TFA, Sigma-Aldrich, St. Louis, Missouri, USA), Triisopropylsilane (TIS, Sigma-Aldrich, St. Louis, Missouri, USA) and water.

For synthesis of specific fragments, specific protocols were used. The one used for the synthesis of fragments TSO8 (1-23) and TSO8 (3-23) was

*Table 1 The protocol used for the synthesis of fragments TSO8 (1-23) and TSO8 (3-23)*

<b>Reaction</b>	<b>Reagent</b>	<b>Temperature</b>	<b>Time</b>
Fmoc Cleavage	Pip 20% in NMP	70°C	First cycle 3 min Second cycle 5 min
Coupling	HOBt/HBTU/Dipea	70°C	First coupling 7 min Second coupling 7 min
Acetilation (end-capping)	Ac <sub>2</sub> O/ Dipea	R.T.	10 min

The protocol used for the synthesis of fragment TSO8 (23-39) was:

Table 2 The protocol used for the synthesis of fragment TSO8 (23-39)

Reaction	Reagent	Temperature	Time
Fmoc Cleavage	Pip 20% in NMP	50°C	First cycle 3 min Second cycle 5 min
Coupling	HOBt/HBTU/Di pea	50°C	First coupling 7 min Second coupling 7 min
Acetylation (end-capping)	Ac <sub>2</sub> O/ Dipea	R.T.	10 min

The different protocol is related to the resin used. Rink amide resin used in the synthesis of N-terminal fragment of TSO8 can be more stable to the power used in the coupling than Trityl Resin, used in the synthesis of C-terminal fragment.

### 3.3 Structural analysis

Circular dichroism spectroscopy is spectroscopic technique, which measure the difference of absorption of left and right circular polarized light as function of wavelengths. CD is a consequence of the interaction of circularly polarized light with chiral molecules (i.e. amino acids) (49), and in particular how the chiral centers in protein restudies are arranged in asymmetric conformations, such as the right handed helix ore extended  $\beta$ -strand.

#### 3.3.1 Circular dichroism spectroscopy

CD spectroscopy is powerful method in structural analysis of chiral molecules, and it can be used to examine secondary structure of peptides. Circular dichroism can be defined as difference in the absorption of left and right-handed circularly polarized light (respectively  $A_L$  and  $A_R$ ):

$$\Delta A = A_L - A_R \quad 3.1$$

where  $\Delta A$  stands for difference of absorption. The above equation is an extension of Beer-Lambert's law (49) and therefore can be rewritten as:

$$A = (\varepsilon_L - \varepsilon_R) \times c \times l \quad 3.2$$

where  $\varepsilon_R$  and  $\varepsilon_L$  are the molar extinction coefficients for left and right circularly polarized light component respectively,  $c$  is molar concentration, and  $l$  is the path length in centimeters.

$\Delta\epsilon$  is known as the molar circular dichroism and depends on concentration, temperature, and the chemical environment (49). For historical reasons, CD spectra are normally reported in degrees of ellipticity ( $\theta$ ). After light passes through a solution of chiral molecules, the polarized electrical fields from the two circularly polarized light components that have been differentially absorbed sum up to form an elliptically polarized electrical field (see Figure 5).

In Figure 5 c) we can see that:

$$\tan\theta = \frac{E_L - E_R}{E_L + E_R} \quad 3.3$$

where  $E_L$  and  $E_R$  represents component of left and right electrical field respectively.  $\theta$  is measured in radians and can be expressed as  $\theta = \frac{2.303}{4}(A_L - A_R)$ . In this form,  $\theta$  is a function of wavelength. If we multiply the above equation by  $\frac{180}{\pi}$ ,  $\theta$  is now measured in degrees and is expressed as  $\theta = 32.98 \Delta A$ . Historically, CD has been expressed in terms of molar ellipticity  $[\theta] = 3298 \Delta\epsilon$  to avoid concentration and path length dependence (49).

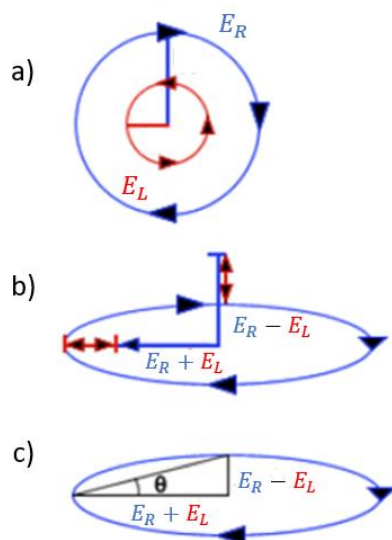


Figure 5 Left and right circularly polarized light with different magnitudes shown in a), which together form elliptically polarized light upon interacting with chiral molecule show in b) and characterized by angle  $\theta$  as presented in c).

### 3.3.2 Experimental set up

CD spectra were obtained on Jasco J-815 CD spectrometer (Tokyo, Japan). The spectra are accumulation of three scans measured in sodium phosphate buffer (SPB), in mixture of trifluoroethanol (TFE) and SPB and lastly, in the presence of sodium dodecyl sulfate micelles (10 mM SDS in SPB). SDS was used as simple model of biological membranes.

All scans were done at far ultraviolet wavelengths, ranging from 197-247 nm. To eliminate background noise, spectra of all blank media were obtained. Firstly, spectra of 400  $\mu$ l of a 10 mM SPB buffer solution was obtained, followed by gradual addition of TFE to obtain the blank spectra of mixtures with different TFE proportions, ranging from 5% TFE in 10mM SPB to 50% TFE in 10 mM SPB. Lastly, spectra of SDS media were obtained in the same ultraviolet range.

This was followed by measuring CD spectra of each individual peptide. Firstly, 6.25  $\mu$ l peptide (40  $\mu$ M final concentration) was mixed with 393.75  $\mu$ l of SPB (10 mM) in a quartz cuvette. Next, specific volumes of TFE were added to the same cuvette to acquire peptide spectra depending on different proportions of TFE (see Table 1). For instance, to measure the spectra in 5% TFE in 10 mM SPB, 21.05  $\mu$ l of TFE was added to obtain the required percentage of TFE in mixture. To obtain 10% TFE in the mixture, 44.44  $\mu$ l of was required, and therefore, an additional 23.39  $\mu$ l of TFE was added to the mixture. This procedure was repeated for all other TFE mixtures (see Table 1). The absorption was then adjusted for dilution by using an appropriate factor as shown in the table.

*Table 3 Calculation of required TFE volume to add in the cuvette.*

<b>%TFE</b>	<b>Volume TFE (needed)</b>	<b>Volume TFE (added)</b>	<b>Dilution factor</b>
5	21.05	21.05	1.05
10	44.44	23.39	1.11
15	100	55.56	1.25
25	171.42	71.43	1.43
40	266.67	95.24	1.67
50	400	133.3	2

CD spectra analysis was performed by Bestsel (50–52) and confirmed by Model Spectra designed by J. Reed and T.A.Reed (53).

### 3.4 Antimicrobial activity assays

To examine antibiotic potency of selected peptides, minimal inhibitory concentration (MIC) and minimal bactericidal concentration (MBC) was determined. Minimum inhibitory concentration is the lowest concentration of AMP needed to inhibit visible growth of bacteria after incubation, while minimal bactericidal concentration represents lowest concentration of the same peptide required to kill bacteria. Model microorganisms used in these experiments were *Escherichia coli* ATCC 25922 and *Staphylococcus aureus* ATCC 25923, as representatives of Gram-negative and Gram-positive bacteria, respectively.

#### 3.4.1 Experiment set up

Initial experimental step was to prepare two-four ml of overnight bacterial cultures incubated in Mueller-Hinton broth (MH) at 37°C for 18-20 hours. This suspension is then diluted by mixture 300 µl of overnight culture with 6 ml of MH medium and then incubated at 37°C for 1-1.5 h, until bacteria reach logarithmical growth phase. The final bacterial concentration required for the experiment is  $1 \times 10^6 \frac{CFU}{ml}$  in 6 ml. The concentration of bacteria is determined by measuring scattering at 600 nm with spectrophotometer, followed by determining the volume of bacterial suspension the needs to be diluted to obtain the required volume and concentration. This is done by applying the relationship:

$$c_i \times V_i = c_f \times V_f \quad 3.4$$

where  $c_i$  represents the concentration of bacteria in a suspension,  $c_f$  is the final concentration (in our case  $5 \times 10^5 \frac{CFU}{ml}$ ),  $V_f$  is the final volume (in our case 6 ml).  $V_i$  is the volume of bacterial suspension that needs to be diluted with MH medium to obtain  $V_f$ :

$$V_f = V_i + V_{MH} \quad 3.5$$

where  $V_{MH}$  is volume of MH medium that needs to be added to  $V_i$ .

#### 3.4.2 MIC assay

Stock solutions of peptides used for MIC assay ranged from 0.75 mM (TSO8\_full) to 2.24 mM (TSO8\_3-23). The MIC value of selected peptides was determined by a standard microdilution method using 96-well microtiter plates (see Figure 6). The plate contains eight rows with twelve columns and each row was used for a given peptide while each column contained different peptide concentrations. Initially, 50µl of MH medium was added to all wells from the second to the twelfth column. In the first well of each row (or first column of



the plate), 100  $\mu\text{l}$  of mixture containing peptide and MH medium was prepared, in a such way that concentration of each peptide was 128  $\mu\text{M}$ . This was followed by serial two-fold dilution by transferring 50  $\mu\text{l}$  of the suspension from the first well into the second, from the second into the third etc., finishing with the 10<sup>th</sup> column.

Finally, 50  $\mu\text{l}$  of bacteria was added in each well with peptide mixture (from first column to eleventh), thus lowering concentration of the peptide (64  $\mu\text{M}$  in the first column) and bacteria by half. Eleventh column was positive control to ensure validation of experiment by viability of bacteria. To the twelfth column was added with only MH medium to ensure there was no contamination of the medium. The microtiter plate was incubated for 18 hours at 37°C followed by visual validation of the results to determine the MIC. Each experiment was repeated six times to ensure accuracy of presented results.

well	1	2	3	4	5	6	7	8	9	10	11	12
A	64 $\mu\text{l}$	32 $\mu\text{l}$	16 $\mu\text{l}$	8 $\mu\text{l}$	4 $\mu\text{l}$	2 $\mu\text{l}$	1 $\mu\text{l}$	0.5 $\mu\text{l}$	0.25 $\mu\text{l}$	0.125 $\mu\text{l}$	0 $\mu\text{l}$	MH
B												
C												
D												
E												
F												
G												
	a)		b)					c)		d)		

Figure 6 Example of microtiter plate set-up for MIC assay. a) First column contains 100  $\mu\text{l}$  of peptide suspension in MH medium, with peptide concentration of 64  $\mu\text{M}$ . b) Peptide was diluted by pipetting 50  $\mu\text{l}$  of peptide/from the first column into the second and resuspended at least ten times. Same process was repeated inclusive with 10<sup>th</sup> column (serial two-fold dilution). c) Eleventh column contained bacteria in MH medium (bacterial positive control) d) twelfth column contained only MH medium to monitor any possible contamination.

### 3.4.3 MBC assay

MBC assay was performed after MIC assay was completed to determine whether the peptides have a bacteriostatic or bactericidal effect. For this purpose, 25  $\mu$ l of bacterial suspension from each well corresponding to MIC, 2 $\times$ MIC and 4 $\times$ MIC was inoculated on MH agar and incubated for 18 - 20 hours. The MBC value was determined as the peptide concentration causing at least a 99.9% reduction of the number of bacteria present at the beginning of MIC assay (9).

## 3.5 Cytotoxicity assay – membrane permeabilization

Cells of the human lymphocyte B precursor cell line MEC-1 (54) were cultivated in suspension at 37°C with 5% CO<sub>2</sub>. MEC1 cells were maintained in complete medium consisting of RPMI medium (Sigma-Aldrich, St. Louis, Missouri, USA) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich, St. Louis, Missouri, USA), 2 mM glutamine, 100 U/mL penicillin, and 100  $\mu$ g/mL streptomycin, and passaged twice a week. For experimental set-up, MEC1 cells were collected from culture flask, washed with PBS buffer, and diluted to 10<sup>6</sup> cell/ml, by counting viable cells with Trypan Blue. Cells in PBS were aliquoted into tubes and exposed to peptides at concentrations ranging from 0.1 - 100  $\mu$ M, for 30 min at 37°C in thermostatic bath. At the end of the incubation period, cell viability was assessed using a flow-cytometer Attune NxT (Life Technologies, Carlsbad, California, USA), collecting 10<sup>4</sup> cells for each measurement. Data were stored as list-mode files and subsequently analyzed by FCSv7 Express De Novo Software.

## 3.6 Molecular modelling

### 3.6.1 Molecular dynamics simulation

Molecular modeling was performed by using semiclassical models adapted for the simulations of molecular dynamics (MD). This is a powerful computational method that gives insight into microscopic world, by calculating time evolution of an atomic or molecular system using Newtonian mechanics:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\nabla_i V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad i = 1, \dots, N \quad 3.6$$

where  $m_i$  stands for mass of  $i$ -th particle,  $\mathbf{r}_i$  for position of  $i$ -th particle,  $\mathbf{F}_i$  force acting on  $i$ -th particle, and  $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$  for potential energy function. By defining the initial conditions for each atom in the system, the set of coupled differential equations (3.6) can be solved (55).

Modeling quantum mechanical motions and chemical reactions of large molecular systems is often computationally too intensive (56). Therefore, a semi-classical approach is used in MD modeling in which both classical and quantum effects are involved in describing and defining interaction parameters or force fields (56,57). Typical force fields are described as pair interactions including bound, non-bound and force field specific terms  $E_{other}$ :

$$E_{total} = E_{bonded} + E_{non-bonded} + E_{other} \quad 3.7$$

Bound terms include stretching of bonds, the bending of valence angles, and the rotation of dihedrals:

$$E_{bonded} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi[1 + \cos(n\chi - \sigma)] \quad 3.8$$

The bonded potential energies, in equation 3.7, have forms of harmonic potentials, where  $b$  is bond length,  $K_b$  stiffness constant,  $b_0$  equilibrium length,  $\theta$  is the angle formed by the two bond vectors,  $K_\theta$  stiffness constant,  $\theta_0$  equilibrium angle,  $\chi$  is the value of the dihedral,  $K_\chi$  is the energetic parameter that determines barrier heights,  $n$  is the periodicity or multiplicity, and  $\sigma$  is the phase (57).

Non-bound terms include electrostatics, dispersion, and Pauli exclusion:

$$E_{non-bonded} = \sum_{\substack{non-bonded \\ pairs\ ij}} (4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}}) \quad 3.9$$

where  $r_{ij}$  is distance between particles, non-bonded Lennard Jones potential (first two terms in Equation 3.9) has parameters  $\varepsilon_{ij}$  depth of potential well,  $\sigma_{ij}$  is effective distance between adjacent particles (first neighbors), and Coulomb potential (third term in Equation 3.9) is defined by  $q_i$  charge of the particle.

The  $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$  is equivalent to equation 3.7. However, the numerical algorithms may use different approaches when calculating bonded, non-bonded force contributions or field specific terms (58).

Several force fields are commonly used in molecular dynamics simulations, including AMBER (59), CHARMM (60), and GROMOS (61). With developments of computational power and optimized algorithms, MD simulations have become a proven valuable method in molecular biology and drug discovery (56). They can be used for simulating behavior of

proteins in various mediums, ligand-receptor binding, understanding mechanism of action of biomolecules and etc. (56).

Generally, MD simulations for peptides are performed in several steps starting with the definition of a system which includes components, size, thermodynamic conditions or other. The starting peptide structure if known can be obtained from structure databases (for example PDB) or if not by using some of the programs for 3D modeling and prediction. The 3D structure is usually in pdb format, which is adapted for use in molecular dynamics programs. The peptide is placed in a simulation box and solvated by adding molecules of water, and electro-neutralized by adding ions. A further step is to perform energy minimization, a process by which a system is relaxed to a conformation of minimum energy to avoid clashes or inappropriate geometry. Then a system is brought into equilibrium state for defined conditions of temperature, pressure, or volume. Usually, in the process of equilibration the solvent is relaxed to adapt equilibrium position and orientation with respect to the protein at defined thermodynamic conditions. Generally, equilibration is done in two steps. Firstly, a system is equilibrated in canonical NVT ensemble to reach the defined temperature, and secondly, in isothermal-isobaric NpT ensemble where the system is placed under constant pressure until an equilibrium density is established. After these procedures, a system is ready for observing and calculation of simulation data.

### 3.6.2 Simulation setup

MD simulations were carried out for TSO8 (1-39) and TSO8\_lin in water. The cQUARK structure predictor was used to obtain models for the initial peptide structures (42). The simulations were done by Gromacs version 2021.2 (62) with the CHARMM36m force field (63) and TIP3 model for water molecules (64). The initial conformations were prepared by CHARMM-GUI solution builder (65). The peptide's charge was defined for pH 7 considering a charged N-terminus and C-terminus. The simulation box is cubic with the size of 7.6 nm filled with 13540 water molecules. The system was neutralized with addition of ions by using Gromacs “-neutral” command, and 42  $Cl^-$  and 25  $K^+$  were added. Energy minimization was done in 513 steps using the steepest descent algorithm followed by 125 ps equilibration runs in the isothermal isochoric (NVT) ensemble. The temperature was fixed over the course of equilibration and production run at 310K. During equilibration positional harmonic restraints were applied to heavy atoms of the peptides (force constants of protein backbone and sidechain were 1 and 0.1 [kcal/(molÅ<sup>2</sup>), respectively] (66). The isothermal-isobaric (NpT) ensemble conditions were imposed by the Nose-Hoover thermostat and Parrinello–Rahman

barostat, with 1.0 ps time constant for temperature and 5.0 ps for pressure (compressibility equal to  $4.5 \times 10^{-5}$  bar) (67,68). The leapfrog integrator time step was fixed at 2 fs, and the bonds were handled by the LINCS option (69). The particle-mesh-Ewald method (70) was used for calculation of the electrostatic interaction with Coulomb cut-off on 1.2 nm and the van der Waals cut-off was set to 1.2 nm with force-switch on 1.0 nm. GROMACS modules were used for the analysis, Gnuplot program for graphs and VMD program as visualization tool (71). DSSP program (database of secondary structure assignments for proteins) was used for calculating time dependence of secondary structural content (72).

## 4 Results and discussion

### 4.1 Peptide identification and bioinformatical analysis

Genome mining predicted thirty-one potential AMPs (Appendix A), all of which were analyzed in order to predict structure, biophysical properties and biological diversity. Obtained results are presented in (Appendix B). Biological diversity was evaluated with BLASTp and tBLASTp, tertiary structural predictions were carried out by cQuark, PEP-FOLD 3.5 and Phyre<sup>2</sup> software, while secondary structure predictions were obtained by PSIPRED software. Three peptides were chosen based on the specificity in predicted structuring, favorable biophysical properties (such as positive charges, structure, etc.) and especially as they are unique from the point of molecular diversity. The peptides' names, sequence and lengths are listed in Table 1, along with the fragments of TSO8 which are also included in the further investigation. The results of the bioinformatical analysis are presented in the following paragraphs.

Table 1 Peptide names, sequences and length

Peptide	Sequence	Length
	<div style="display: flex; justify-content: space-around; font-size: small;"> <span>10</span> <span>20</span> <span>30</span> <span>40</span> </div> <div style="display: flex; justify-content: space-around; font-size: x-small;"> <span>-----</span> <span>-----</span> <span>-----</span> <span>-----</span> </div>	
TAEMU 1	H-KPRLTPLVYRCFYDFRTGKKRCYPQTKLILLR-OH	32 AA
MESCO 2	H-GFFRRIGRAFSRVGRGIGRGFRQLGRLMPRGNYKICLGRCP-OH	41 AA
TSO8 (1-39)	H-GWRRLRRSIRRRIRRIFRKPRRICFPYCPKGPKKGRGDF-OH	39 AA
TSO8_lin	H-GWRRLRRSIRRRIRRIFRKPRRIAFPYAPKGPKKGRGDF-OH	39 AA
TSO8 (3-39)	H-RRLRRSIRRRIRRIFRKPRRICFPYCPKGPKKGRGDF-OH	37 AA
TSO8 (1-23)	H-GWRRLRRSIRRRIRRIFRKPRRI-NH <sub>2</sub>	22 AA
TSO8 (3-23)	H-RRLRRSIRRRIRRIFRKPRRI-NH <sub>2</sub>	20 AA
TSO8(21-39)	Ac-PRRICFPYCPKGPKKGRGDF-OH	20 AA

(a) H- = free N-terminal amine; Ac- = acetylated N-terminus; -OH = free C-terminal acid, -NH<sub>2</sub> = amidated C-terminus (due to C-terminal Gly residue amidation signal)

- **TAEMU 1**

TAEMU 1 is a peptide identified in *Taenia multiceps*. BLASTp search against non-redundant database results in no significant similarity with other sequences indicating that TAEMU 1 is a unique peptide. Moreover, tBLASTn search against whole genome shotgun contigs (wgs) database also results in no hits. From biophysical point of view, TAEMU 1 is a 32 amino acids long peptide, with +8 net charge at pH 7.0. Secondary structure prediction done by PSIPRED suggests that about ~20% is helical, ~25% beta sheet and ~55% random coil (see Figure 7 a)).

As to the tertiary structure, both PEP-FOLD 3.5 and cQUARK predict a N-terminal section and disordered C-terminal section of the peptide. Homology modeling prediction obtained by Phyre<sup>2</sup> resulted in peptides only eight residues long (25% of sequence), modelled with 13.2% confidence by the single highest scoring template (see Figure 7) to be in the form of random coils. The relatively low coverage and confidence might be due to the lack of TAEMU 1 homologues in the database that Phyre<sup>2</sup> is using for the construction of the template.

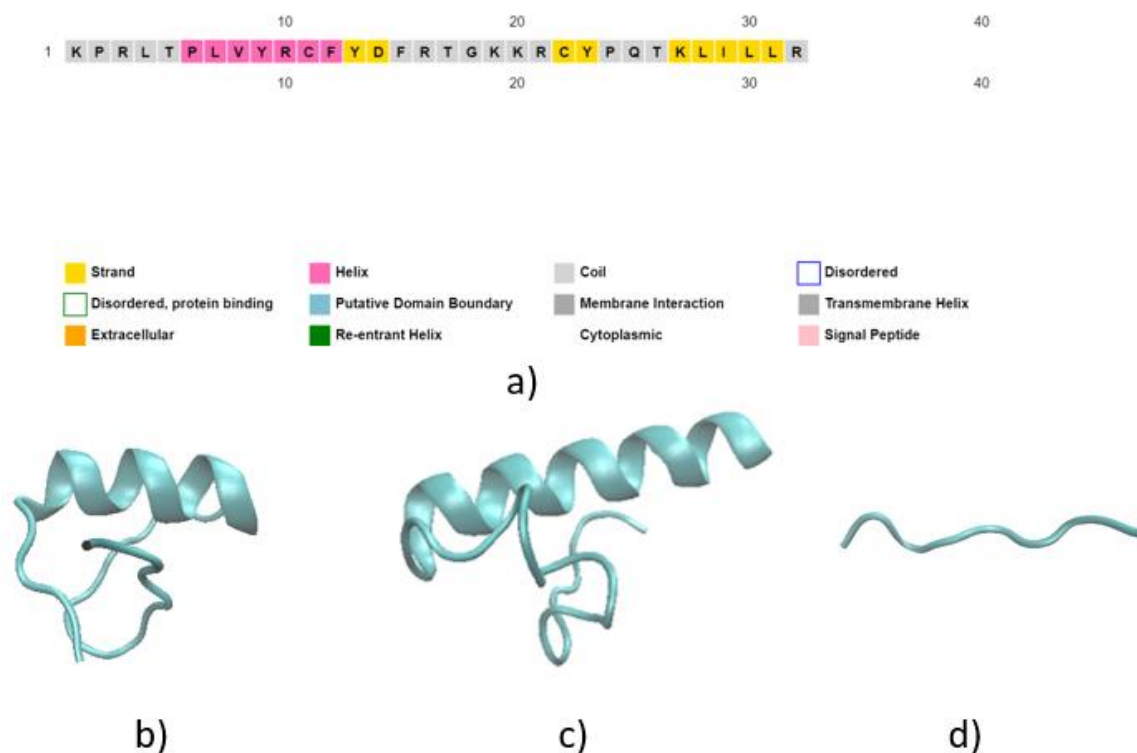


Figure 7 a) Secondary structure prediction of TAEMU 1 obtained by PSIPRED. Tertiary structure prediction obtained by b) PEP-FOLD, 3.5 c) cQUARK and d) Phyre<sup>2</sup>.

- MESCO 2

MESCO 2 is the peptide identified in *Mesocestoides corti*. A BLASTp search against a non-redundant database result in only three peptides with significant similarity (E-value <0.05) when compared to other known sequences (Appendix B). A tBLASTn search against the whole genome shotgun contigs (wgs) database results in no similarities. Therefore, MESCO 2 can be labeled as a quite unique peptide. From the biophysical point of view, MESCO2 is 41 amino acids long peptide, with +11 net charge at pH 7.0. According to PSIPRED the secondary structure prediction is ~ 50% of helical content, ~10% of beta sheet and 40% of random coil (see Figure 8 a)).

Tertiary structure prediction of PEP-FOLD 3.5 and cQUARK (see Figure 8 b) and c)) differ, in the placement of the helical segment, with three small helices are predicted by the first method, two larger helices and random coil at the C-terminal by the second method. The homology modeling prediction obtained by Phyre<sup>2</sup> yielded predictions for only 9 residues in the shape of helix (22% of the sequence) modeled with 18.8% confidence using a single template with the highest score. As in the case of TAEMU 1, this result might be due to the lack of MESCO 2 homologues in Phyre<sup>2</sup> database (see Figure 8 d)).

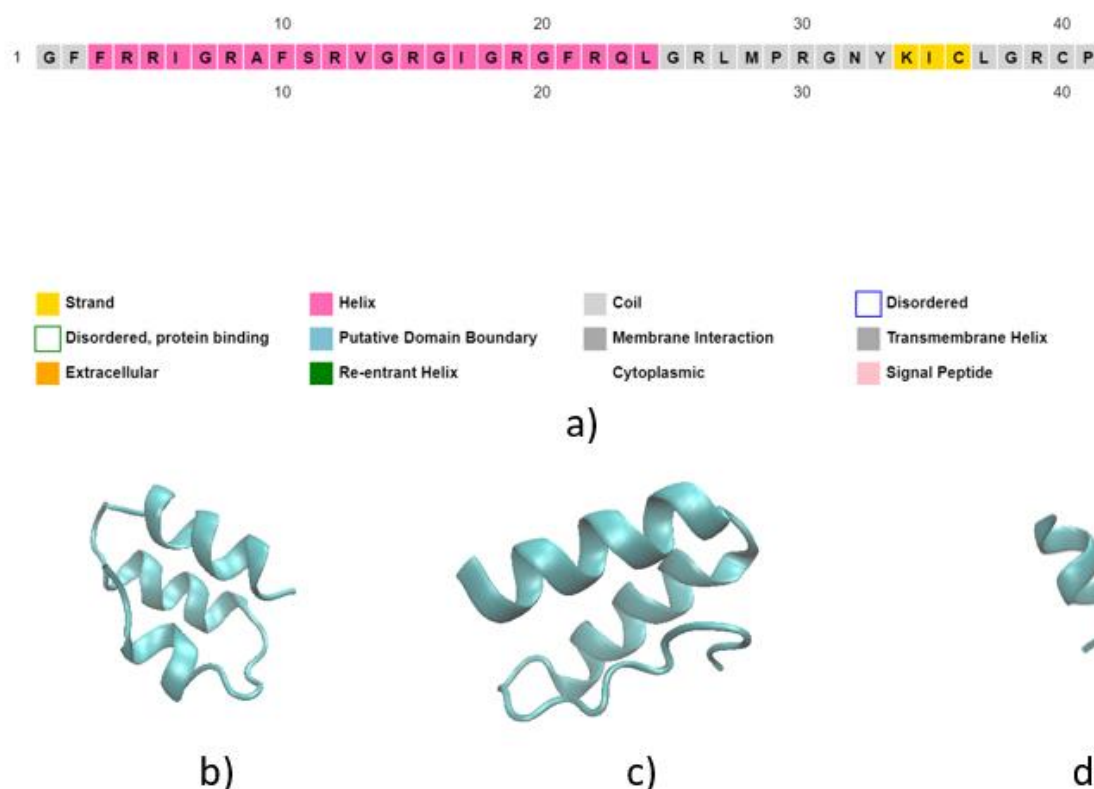


Figure 8 a) Secondary structure prediction of MESCO 2 obtained by PSIPRED. Tertiary structure prediction obtained by b) PEP-FOLD 3.5, c) cQUARK and d) Phyre<sup>2</sup>.



- **TSO8**

TSO8 (1-39) is a peptide identified in *Taenia solium*. A BLASTp search against a non-redundant database result in only six peptides with significant similarity (E-value <0.05) with other sequences, whereas a tBLASTn search against the whole genome shotgun contigs (wgs) database resulted in no similar peptides. From the biophysical point of view, TSO8 is a 39 amino acids long peptide and has relatively high net charge (+16) at pH 7.0. According to PSIPRED, the secondary structure consists of about 50% helix and 50% random coil (see Figure 9 a)). This is in line with PEP-FOLD 3.5 prediction (see Figure 9 b)), cQUARK (see Figure 9 c)) and Phyre<sup>2</sup> (see Figure 9 d)) predictions. However, Phyre<sup>2</sup> yielded predictions for only 22 residues (56% of the sequence) modeled with 36.2% confidence. Like in the case of TAEMU1 and MESCO 2, TSO8 (1-39) is quite a unique peptide, and this result might be due to a few good templates for homology modeling method.

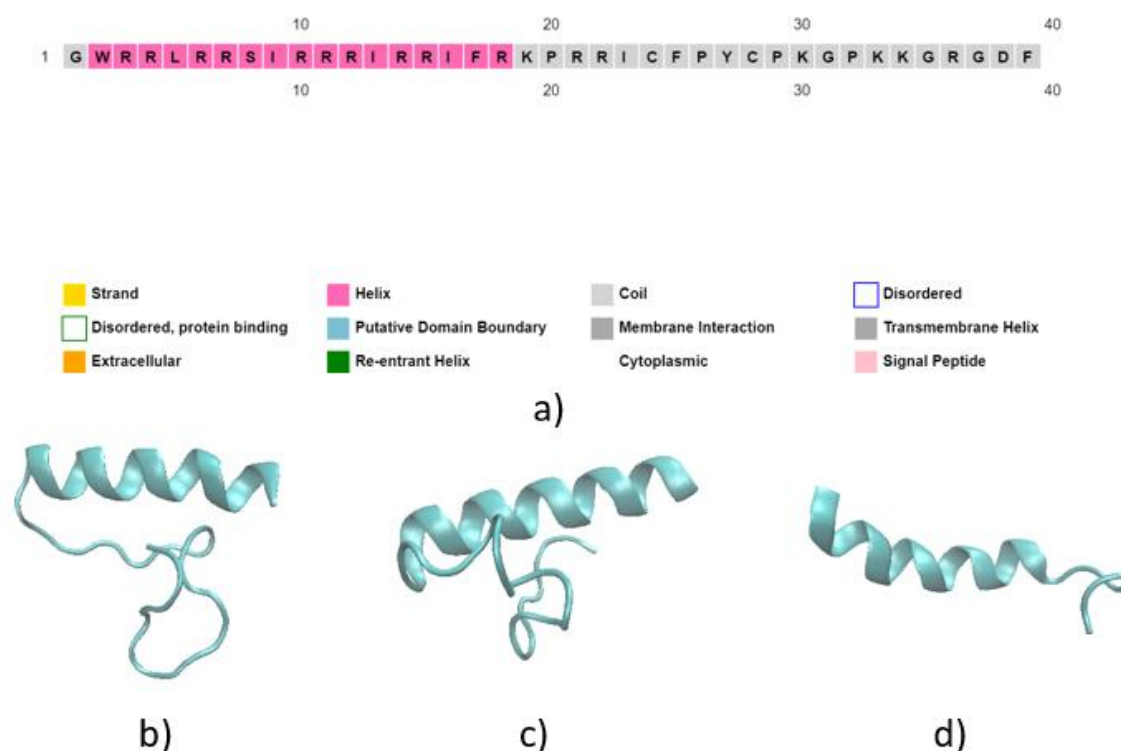


Figure 9 a) secondary structure prediction of TSO8 (1-39) obtained by PSIPRED. Tertiary structure prediction obtained by b) PEP-FOLD 3.5, c) cQUARK and d) Phyre<sup>2</sup>.

Analysis of TSO8 was expanded by synthesizing linear full-length peptide (TSO8\_lin) and specific fragments of the original sequence [TSO8 (3-39), TSO8 (1-23), TSO8 (3-23), TSO8

(21-39)] (see Table 3). Bioinformatical analysis was performed for all additional peptides and results are presented in the following paragraphs.

For TSO8\_lin, cysteine residues were replaced with alanine preventing disulfide bonds formation. A BLASTp search against non-redundant database results in only three peptides with significant similarity (E-value <0.05) with other sequences. A tBLASTn against whole genome shotgun contigs (wgs) database, got no hits. From the biophysical point of view, TSO8\_lin is a 39 amino acids long peptide, with +17 net charge at pH 7.0.

Similarly to TSO8, PSIPRED predicted about 50% of helical content and 54% of random coil (see Figure 10 a)), which was confirmed by PEP-FOLD 3.5 (see Figure 10 b)), cQUARK (see Figure 10 c)) and Phyre<sup>2</sup> (see Figure 10 d)). To conclude, although primary structure TSO8\_lin differs from TSO8 (1-39) in two amino acid residues, results of secondary and tertiary structure predictions are very similar.

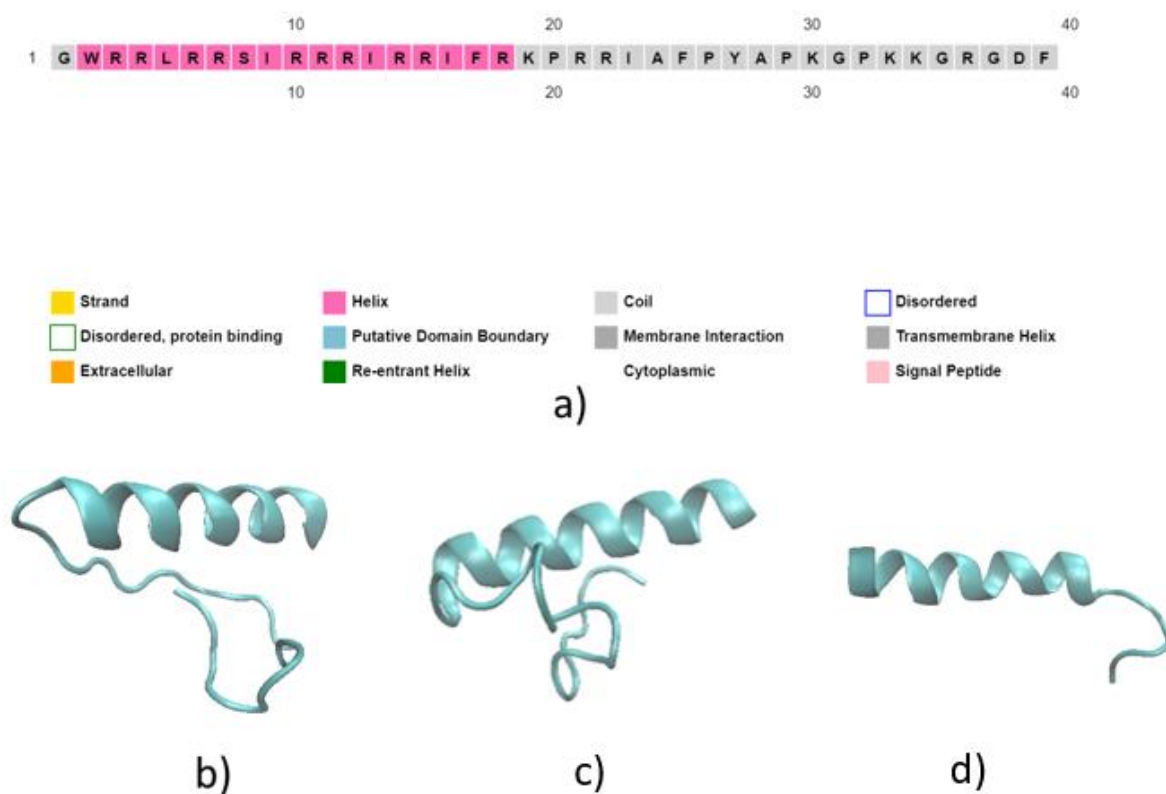


Figure 10 a) secondary structure prediction of TSO8\_lin obtained by PSIPRED. Tertiary structure prediction obtained by b) PEP-FOLD 3.5, c) cQUARK and d) Phyre<sup>2</sup>.

TSO8 fragments were synthesized to explore the role of GW residues at N-terminus, as well as the roles of N-terminal and C-terminal domains (see Figure 11 for alignments). The

net charges at pH of 7.0 differ significantly and are +14 for TSO8 (1-23) and TSO8 (3-23), +3.9 for TSO8 (21-39), and +15.9 for TSO8 (3-39).

TSO8_(21-39)	-----PRRICFPYCPKGPKKGRGDF	20
TSO8_(3-39)	--RRLRRSIRRRIRRIIFRKPRRICFPYCPKGPKKGRGDF	37
TSO8_(1-39)	GWRRLRRSIRRRIRRIIFRKPRRICFPYCPKGPKKGRGDF	39
TSO8_lin	GWRRLRRSIRRRIRRIIFRKPRRIAFPYAPKGPKKGRGDF	39
TSO8_(1-23)	GWRRLRRSIRRRIRRIIFRKPRRI-----	23
TSO8_(3-23)	--RRLRRSIRRRIRRIIFRKPRRI-----	21

\*\*\*\*

Figure 11 Clustal w alignment of TSO8 peptides

The prediction of the secondary structure with PSIPRED could not be obtained as TSO8 fragments are too short to be determined this method. Tertiary structures of TSO8 fragments obtained by PEP-FOLD 3.5 are shown in Figure 12. TSO8 (3-39) prediction is similar to the previous results with the N-terminal helix and C-terminal coil. TSO8 (1-23) and TSO8 (3-23) also have helical structure, and lastly, predicted structure of TSO8 (21-39) is random coil, which is also expected considering C-terminal of TSO8 (1-39) is random coil.

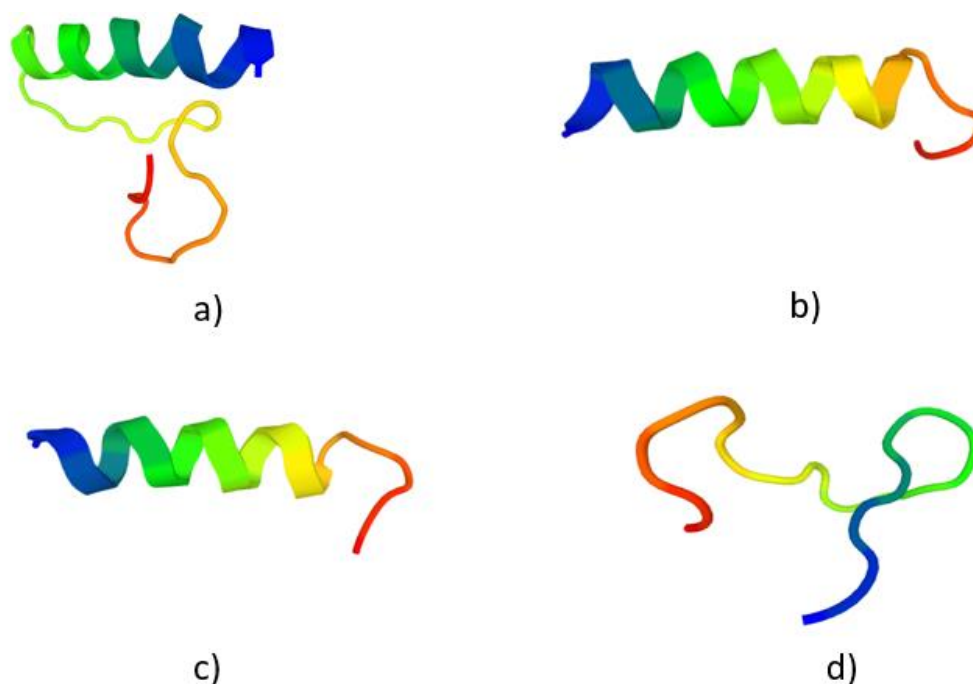


Figure 12 Tertiary structure predictions obtained by PEP-FOLD 3.5 for TSO8 fragments a) TSO8 (3-39) b) TSO8 (1-23), c) TSO8 (3-23) and d) TSO8 (21-39).

## 4.2 Structural studies

The CD spectra of all six peptides were measured under different conditions, including aqueous buffer (SPB), anionic SDS micelles, and different proportions of TFE (5-50%). The results are discussed, separately for each peptide, in the following paragraphs and the ones obtained by Bestsel are presented in Appendix C.

- **TAEMU 1**

CD spectra suggest that the peptide is quite unstructured in an aqueous environment (blue curve 0% TFE, Figure 13), which is not unexpected for an AMP, and the addition of the helix-stabilizing solvent TFE at 5% or 10% (dark red curve and green curve, Figure 13) did not alter this. Some helical structuring is observed at  $\geq 15\%$  TFE (see Figure 13 and Figure 14) but does not surpass  $\sim 15\%$  even at 50% TFE or in membrane mimicking SDS micelles (see also Appendix C), although the formation of even a short helical segment may be relevant for the peptide interaction with the membrane.

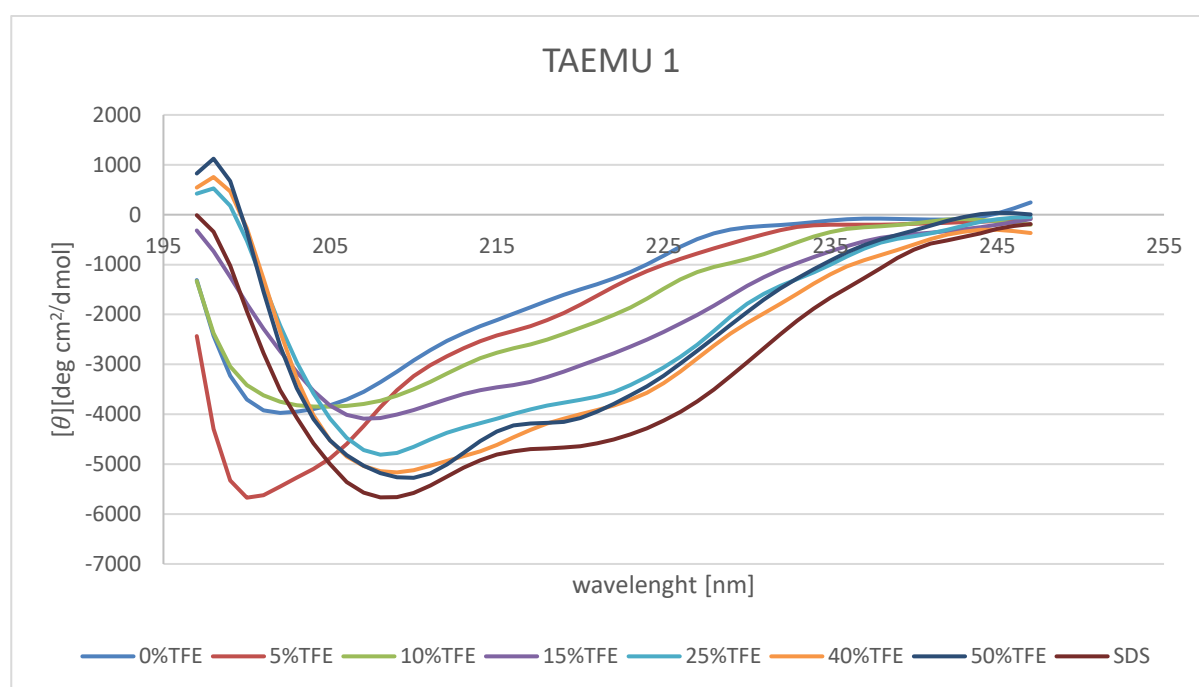


Figure 13 CD spectra of TAEMU 1 in SPB, TFE and SDS

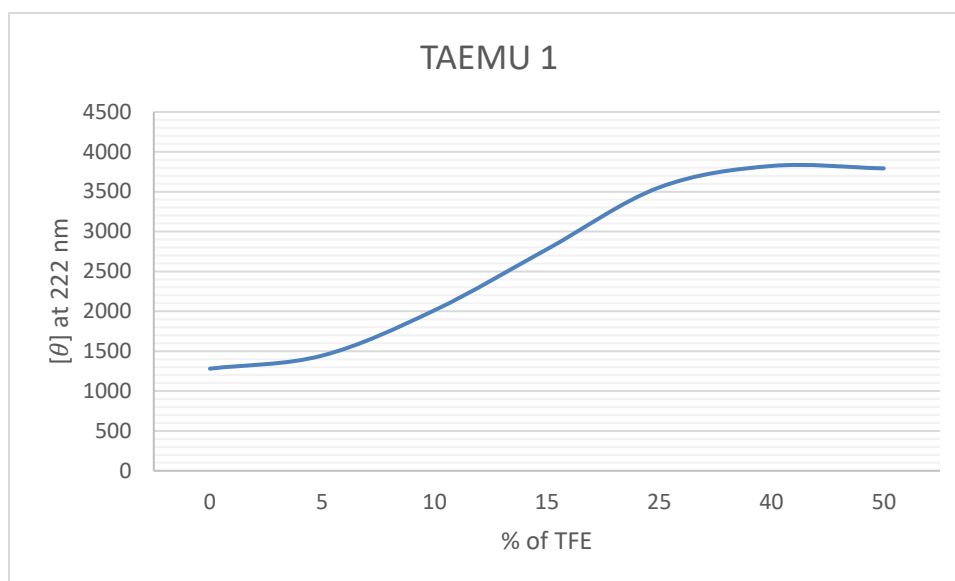


Figure 14 Molar ellipticity at 222 nm for different percentage of TFE

- **MESCO 2**

CD spectra indicate that MESCO 2 is also unstructured in an aqueous environment (blue curve, 0% TFE in Figure 15), and remains so up to 10% TFE (green curve, Figure 15). It however increases to somewhat higher values than TEAMU 1 at increased % TFE (see Figure 16). Again, based on Bestsel, we may suppose that similar helical structuring is present for peptide in 50 % TFE solution and in the presence of SDS micelles (see Appendix C).

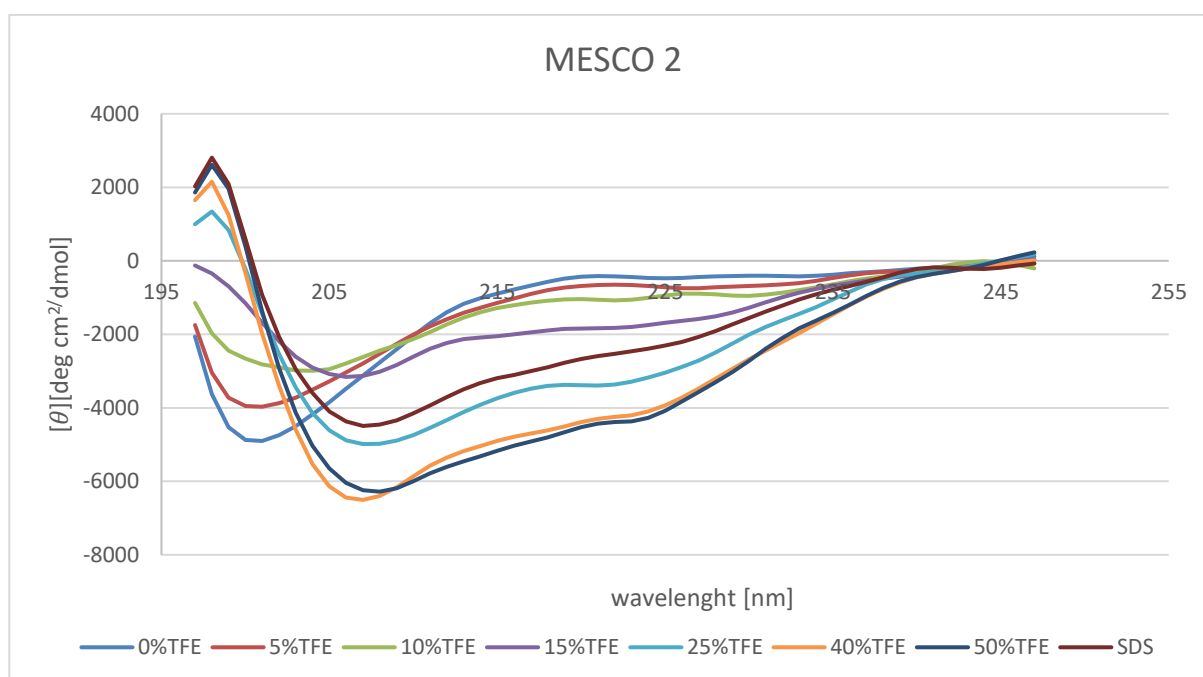


Figure 15 CD spectra of MESCO 2 in SPB, TFE and SDS

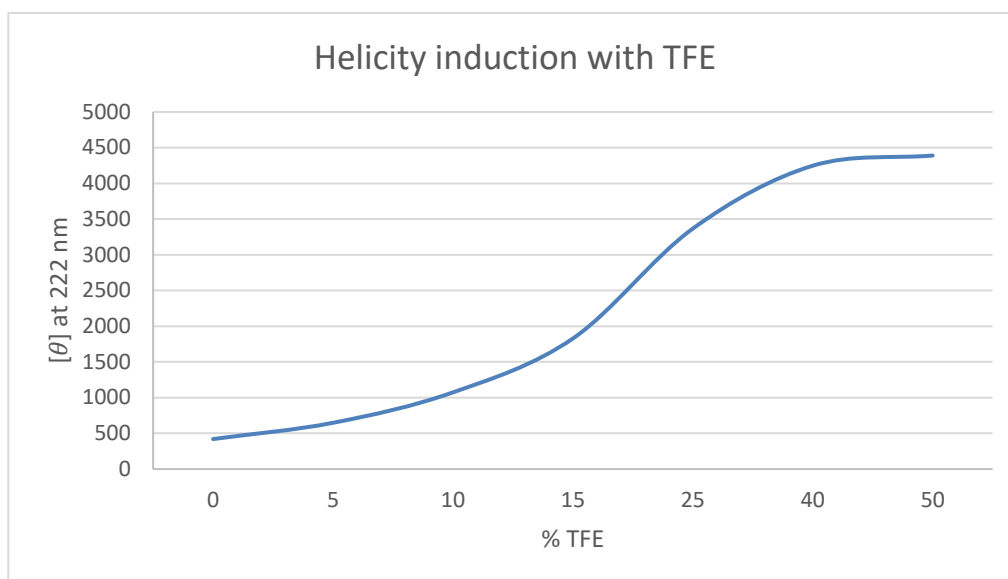


Figure 16 Molar ellipticity at 222 nm for different percentage of TFE

- **TSO8**

In aqueous environment TSO8 (1-39) is unstructured (dark blue curve, 0% TFE in Figure 17) with more than 50% of the peptide in random coil conformation, as calculated by Betsel. The curve of molar ellipticity at 222 nm shows a linear decrease to more negative values which is apparent also for at lower TFE proportions, saturating at  $\geq 25\%$  (see Figure 18). However, compared to MESCO 2 and TAEMU 1, % of helicity is lower, with 7.2% and 9.2% of helical content in 50% TFE and SDS, respectively.

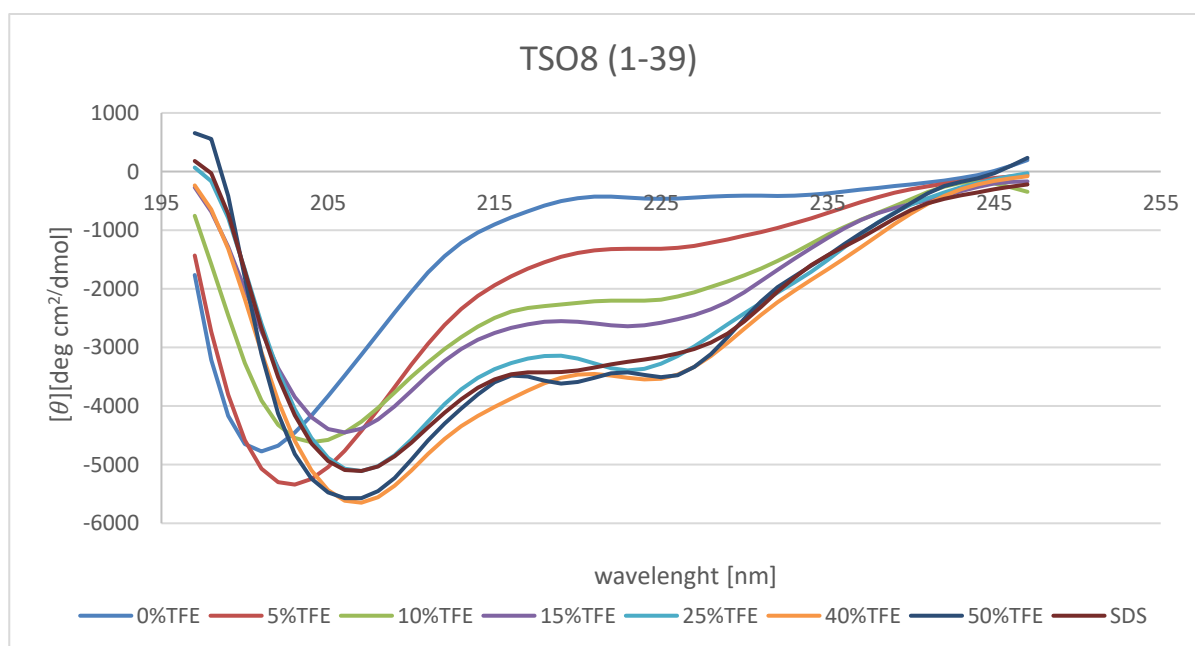


Figure 17 CD spectra of TSO8 (1-39) in SPB, TFE and SDS

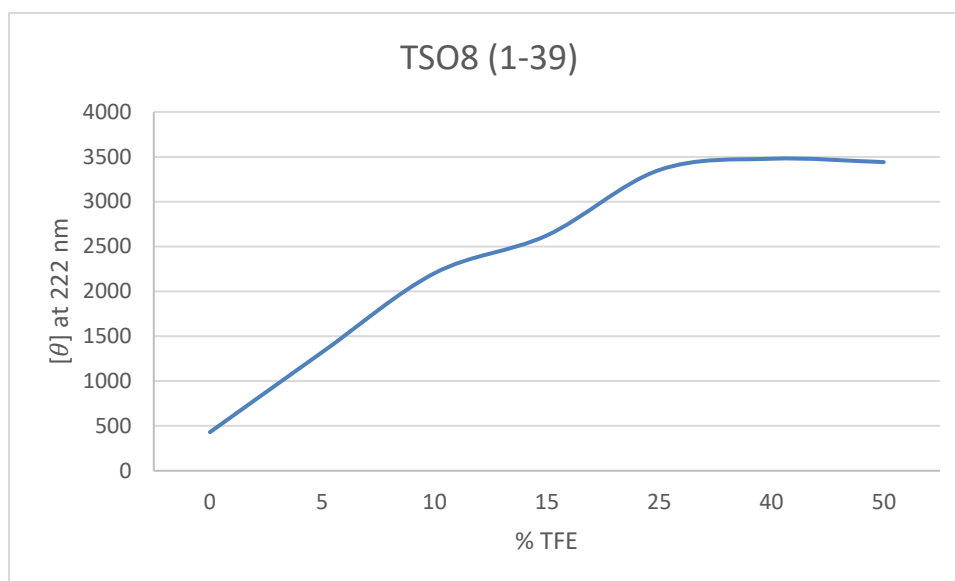


Figure 18 Molar ellipticity at 222 nm for different percentage of TFE

CD spectra of TSO8 (1-39) and its fragment TSO8 (3-39) are similar, as one might expect for two peptides that differ by only 2 amino acids (see Figure 17 and Figure 19). The variation of molar ellipticity at 222 nm in the presence of TFE shows similar behavior (Figure 18 and Figure 20), and Bestsel analysis shows that they adopt an ~ 15% helical content in 50% TFE or SDS, respectively.

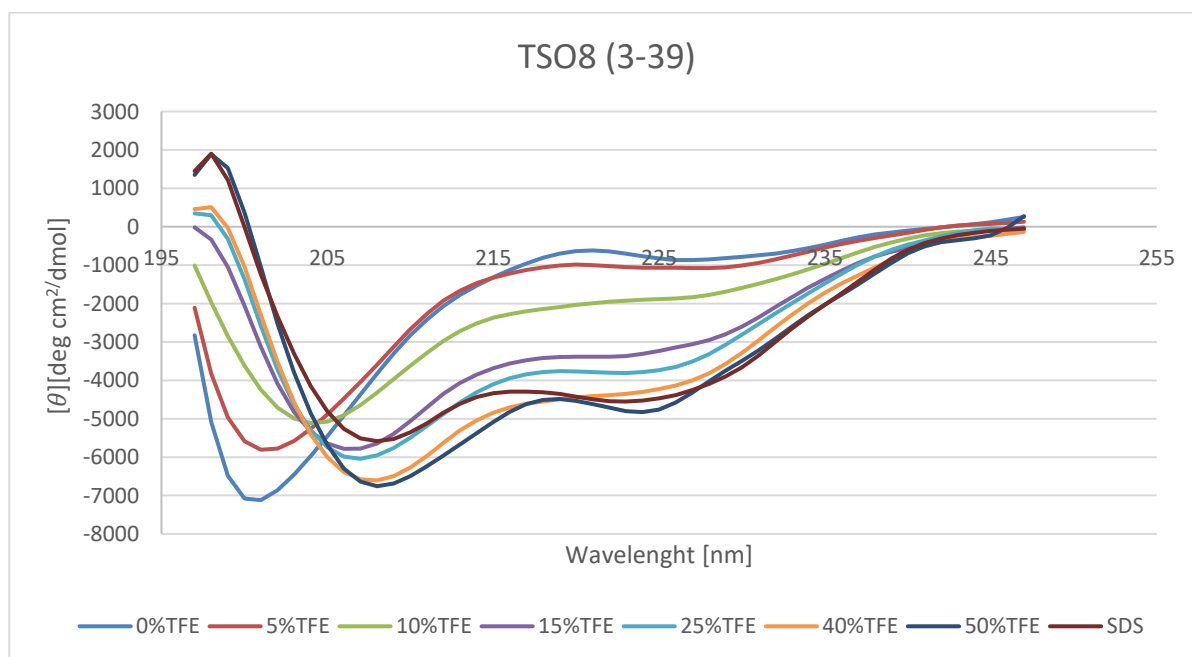


Figure 19 CD spectra of TSO8 (3-39) in SPB, TFE and SDS

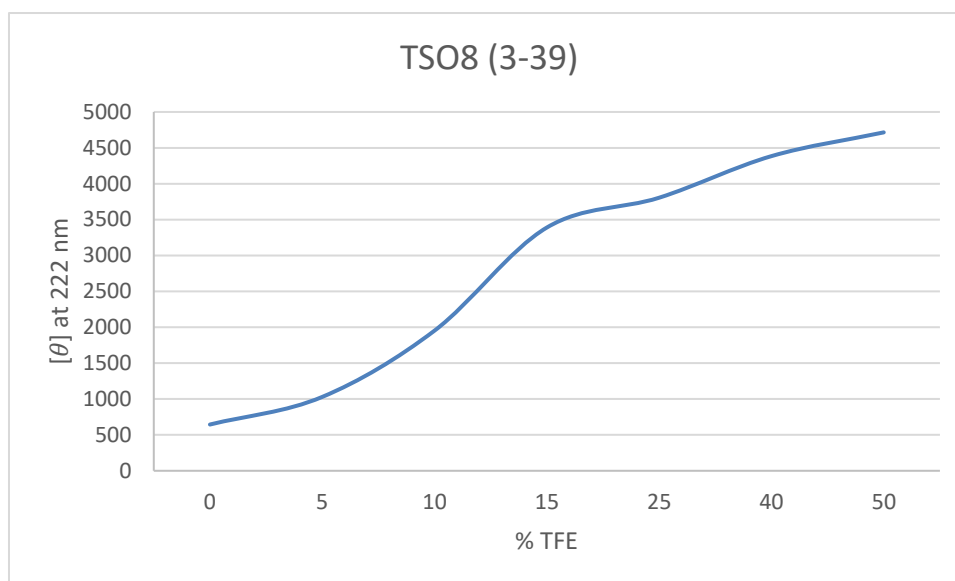


Figure 20 Molar ellipticity at 222 nm for different percentage of TFE

The N-terminal fragment, TSO8 (1-23), is also fairly unstructured in aqueous environment but adopts some helical structure, saturating at > 15% TFE solutions (Figure 21 and Figure 22). and reaching >30% at the higher TFE concentration. However, structuring in SDS (brown curve, Figure 21) differs from those in TFE suggesting different behavior and perhaps more affinity for a membrane-like environment. Bestsel analysis shows that the peptide has 32.5% of helical content in 50% TFE and 20.9% in SDS,

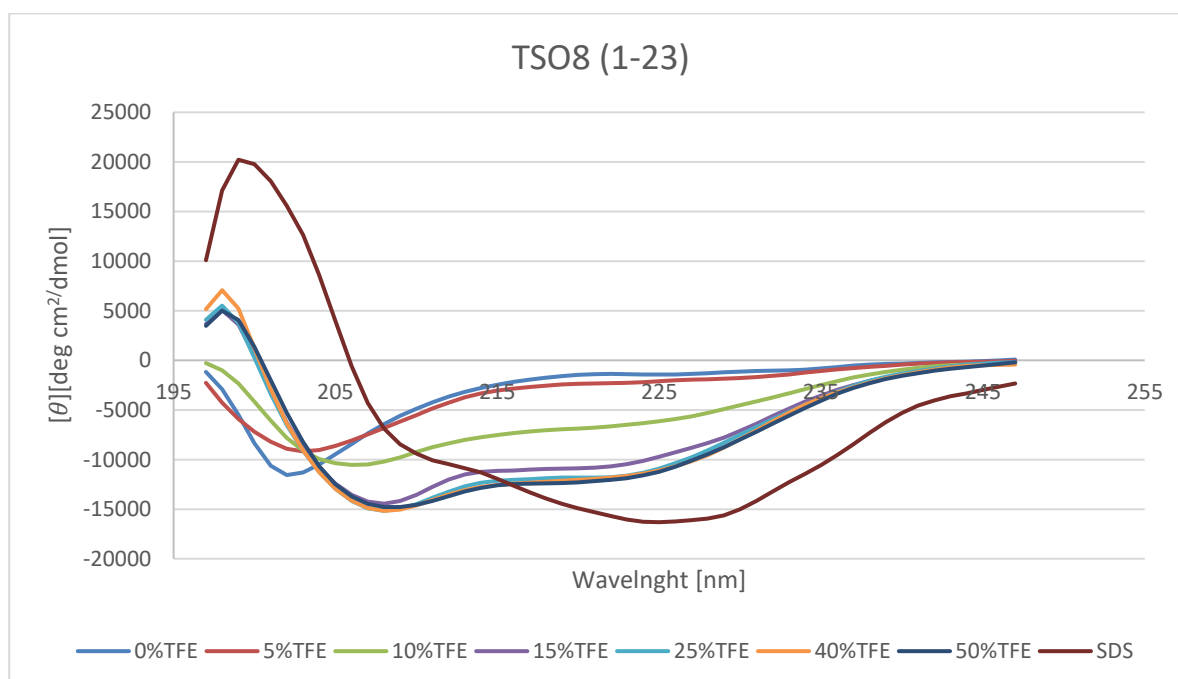
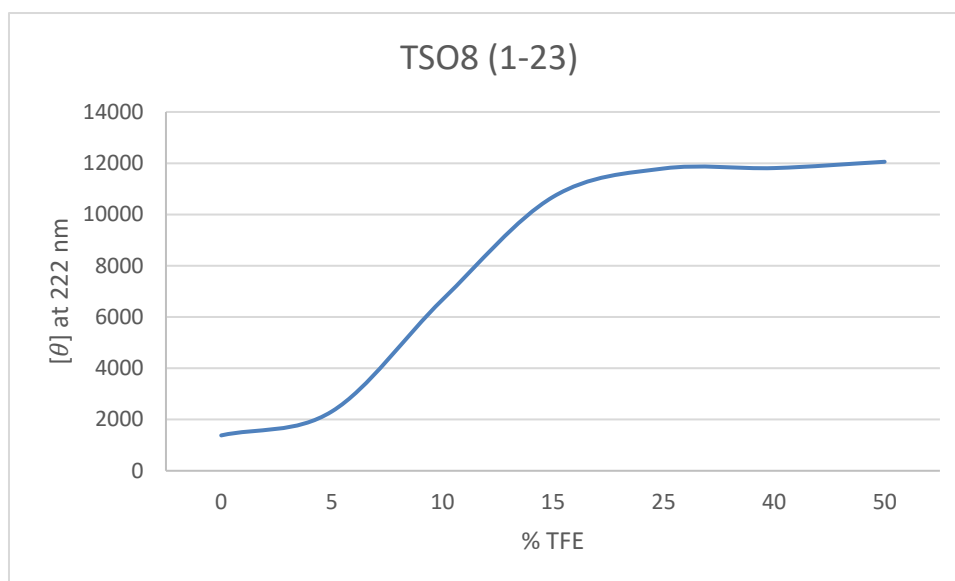


Figure 21 CD spectra of TSO8 (1-23) in SPB, TFE and SDS





*Figure 22 Molar ellipticity at 222 nm for different percentage of TFE*

The fragment TSO8 (3-23) shows similar CD spectra as TSO8 (1-23) in water and water-TFE mixture, but different behavior in SDS solutions (Figure 21 and Figure 23). The dependence of molar ellipticity at 222 nm on TFE % shows a slower saturations trend in the case of TSO8 (3-23) than TSO8 (1-23) (Figure 22 and Figure 24). Therefore, the GW sequence at N terminal which is lacking in TSO8 (3-23) contributes to the increase in helical content. This is confirmed by Bestsel calculations, as for TSO8 (3-23) is 28.7% of helical content in 50% TFE and 16.2% in SDS, smaller than the calculated values for TSO8 (1-23).

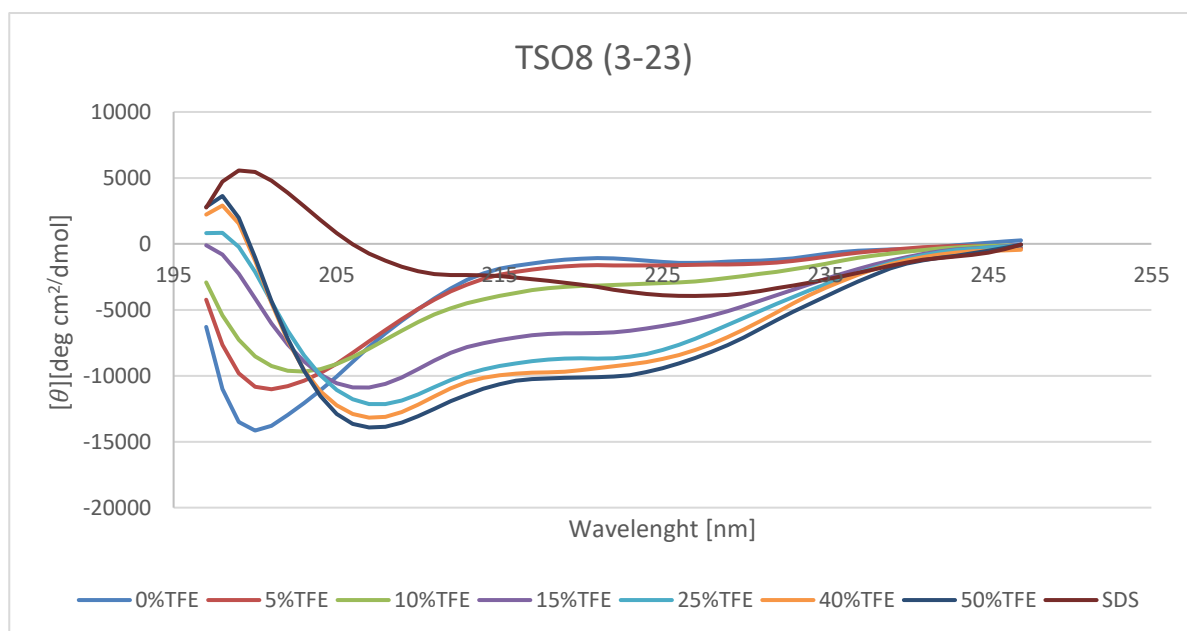


Figure 23 CD spectra of TSO8 (3-23) in SPB, TFE and SDS

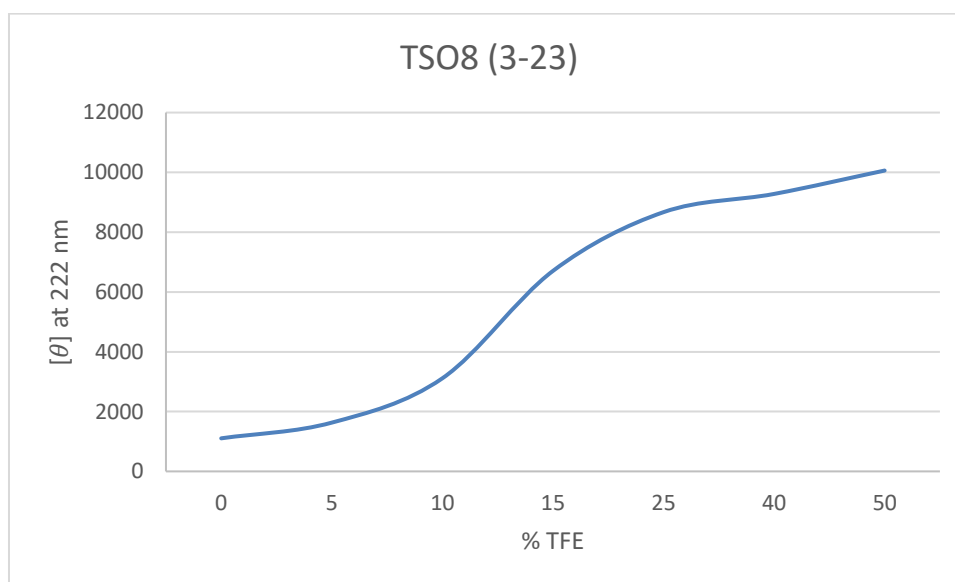


Figure 24 Molar ellipticity at 222 nm for different percentage of TFE

Finally, C-terminal fragment of TSO8 (23-39) has low % of helical structuring in all measured solutions, and Bestsel analysis shows about 3.8% of helical content in 50% TFE and 9.6% in SDS solution (see Figure 25 and Figure 26). Obviously, in membrane like environment, C-terminal fragment tends to increase helical content compared to aqueous environment.

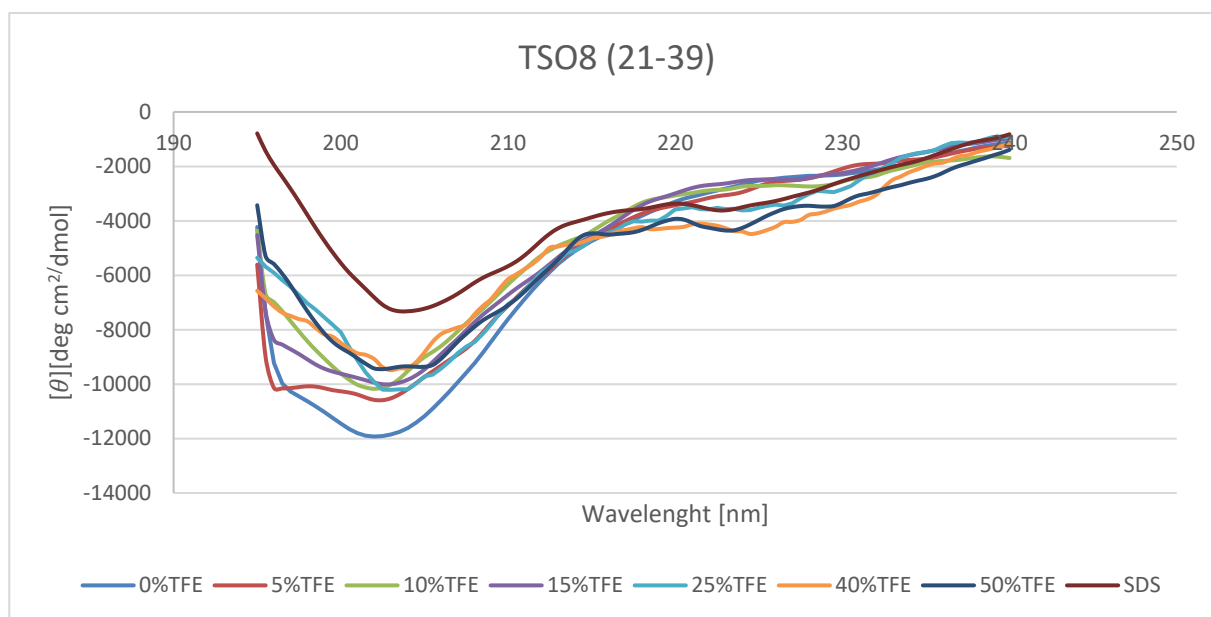


Figure 25 CD spectra of TSO8 (21-39) in SPB, TFE and SDS

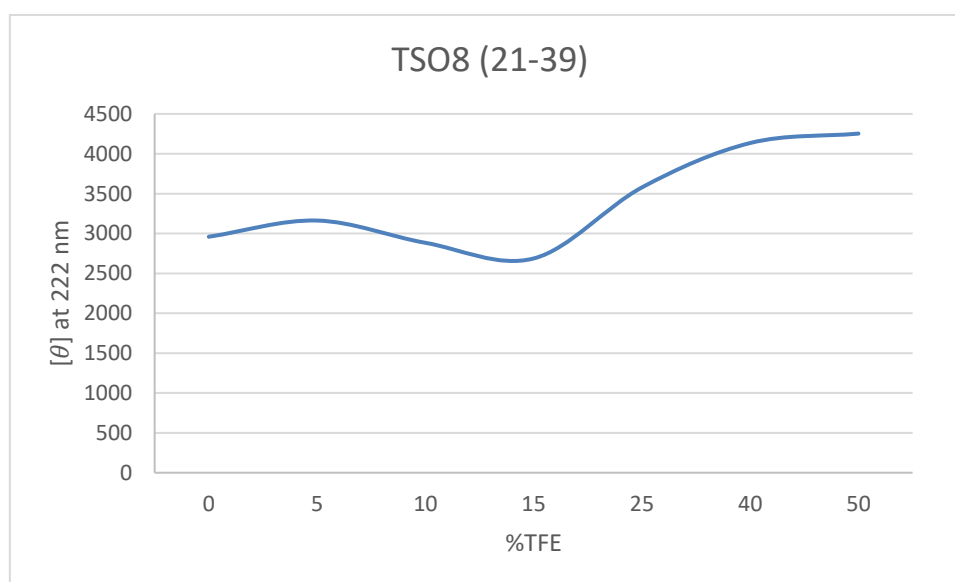


Figure 26 Molar ellipticity at 222 nm for different percentage of TFE

### 4.3 Antimicrobial activity

The antibacterial activity of the peptides was assessed by determining MIC and MBC values against two bacterial strains: Gram-positive *Staphylococcus aureus* and Gram-negative *Escherichia coli*. MESCO 2 and TSO8 showed similar activity against both tested strains with MIC 2  $\mu\text{M}$  against *E. coli* and slightly higher values against *S. aureus* (4 and 8  $\mu\text{M}$ , respectively) (see Table 2). MBC values were identical to MIC suggesting that the peptides are bactericidal rather than bacteriostatic. Looking at the TSO8 fragments, it can be seen that TSO8 (3-39) is the most active peptide with MIC/MBC values 1  $\mu\text{M}$  against *E. coli*. At the

same time, it is quite potent against *S. aureus* with MIC/MBC 4  $\mu\text{M}$ . In contrast to previously reported results by Sato et al. (34), conserved aromatic residue at position 2 (W) does not seem to be essential for antimicrobial activity. As for the smaller, N-terminal fragments, TSO8 (1-23) is slightly more potent (MIC values 2 and 4  $\mu\text{M}$  against *S. aureus* and *E. coli*) compared to TSO8 (3-23), with pronounced selectivity towards Gram-positive bacteria (see Table 2). Interestingly, TSO8 (1-23) contains the GW fragment suggesting that aromatic residue could have an effect on the shortened TSO8 sequence. Finally, TAEMU 1 was moderately active, with relatively high MIC/MBC values of 32  $\mu\text{M}$  towards *E. coli*, whilst antimicrobial activity was not determined for *S. aureus*, as it was higher than the highest tested concentration (64  $\mu\text{M}$ ).

Table 4 MIC and MBC values for each peptide

Peptide name	<i>Escherichia coli</i>		<i>Staphylococcus aureus</i>	
	MIC [ $\mu\text{M}$ ]	MBC [ $\mu\text{M}$ ]	MIC [ $\mu\text{M}$ ]	MBC [ $\mu\text{M}$ ]
TAEMU_1	32	32	> 64	> 64
MESCO_2	2	2	4	4
TSO8_1_39	2	2	8	8
TSO8_3_39	1	1	4	4
TSO8_1-23	4	4	2	2
TSO8_3-23	8	8	4	4
TSO8_21-39	> 64	> 64	> 64	> 64

#### 4.4. Cytotoxicity assay

MEC1 cells (106/ml) were exposed to increasing peptide concentrations (0.1 - 1000  $\mu\text{M}$ ) in PBS buffer for 30min at 37°C and propidium iodide (PI) (10 $\mu\text{g/ml}$ ) added to inspect cell membrane integrity. Flow cytometry was performed immediately and in each run 10000 events were acquired with Attune NxT flow cytometer. Data were analyzed with FCS Express V7 and were expressed as % of PI positive cells (shown on Y axes, see Figure 27-Figure 29).

Comparing the effect of TSO8 (1-39), MESCO 2 and TAEMU 1 on MEC-1 cells at different concentrations, it can be seen that MESCO 2 is slightly less cytotoxic compared to TSO8 (1-

39), reaching saturation of PI positive cells at about 90% (see Figure 27). TAEMU 1, on the other hand is far less toxic, reaching saturation of PI positive cells at about 20%.

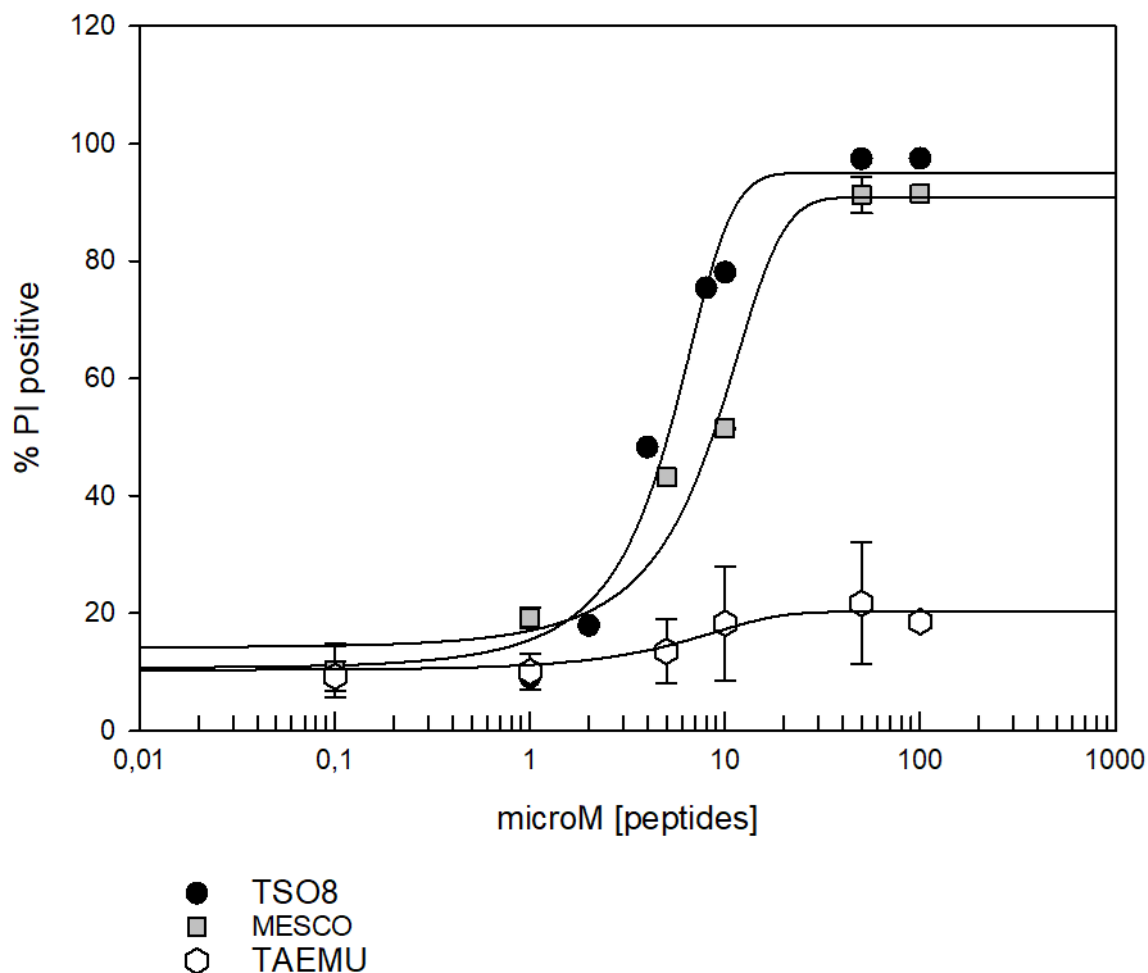


Figure 27 Percentage of PI positive (dead) MEC-1 cells over different concentration of TSO8 (1-39), MESCO 2 and TAEMU 1.

Comparing the effect of TSO8 and its fragments on MEC-1 cells, it can be concluded that all peptides are toxic over a certain threshold, except for the C-terminal fragment, TSO8 (21-39), which showed no, or negligible, toxicity even at the highest concentration used (see Figure 28). TSO8 (1-39) was slightly more toxic compared to than TSO8 (3-39) at concentrations higher than 10  $\mu$ M, which can most likely be attributed to short N-terminal GW sequence at N-terminal of TSO8 (1-39). TSO8 (1-23) behaves similarly as TSO8 (1-39) and TSO8 (3-39), whereas TSO8 (3-23) is less cytotoxic, reaching 70% PI positive cells at 100  $\mu$ M concentration.

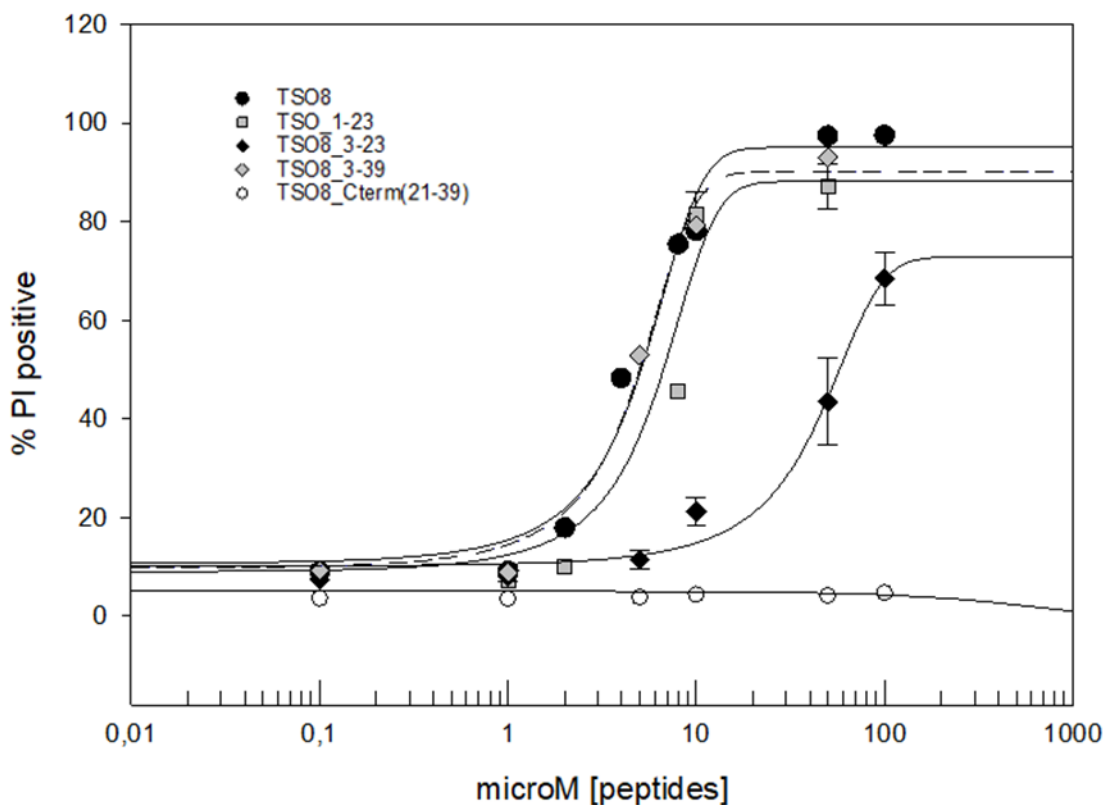


Figure 28 Percentage of PI positive (dead) MEC-1 cells at different concentrations of TSO8 (1-39), TSO8 (3-39), TSO8(1-23), TSO8(3-23) and TSO8 (21-39)

Finally, looking at the morphology of one specific peptide (TSO8), its cells are described with FSC-A and SSC-A signal. Vital cells are represented with Gate 2, dead cells with Gate 3, while Gate 4 contains all possible combination of signals. At concentration below 1  $\mu\text{M}$ , almost 90% of the cells are vital, while at 10  $\mu\text{M}$  concentration almost 80% of cells are dead. As the peptide concentration increases, there is a jump in PI signal (dead cells) and around 30  $\mu\text{M}$  saturation point is reached (see Figure 29).

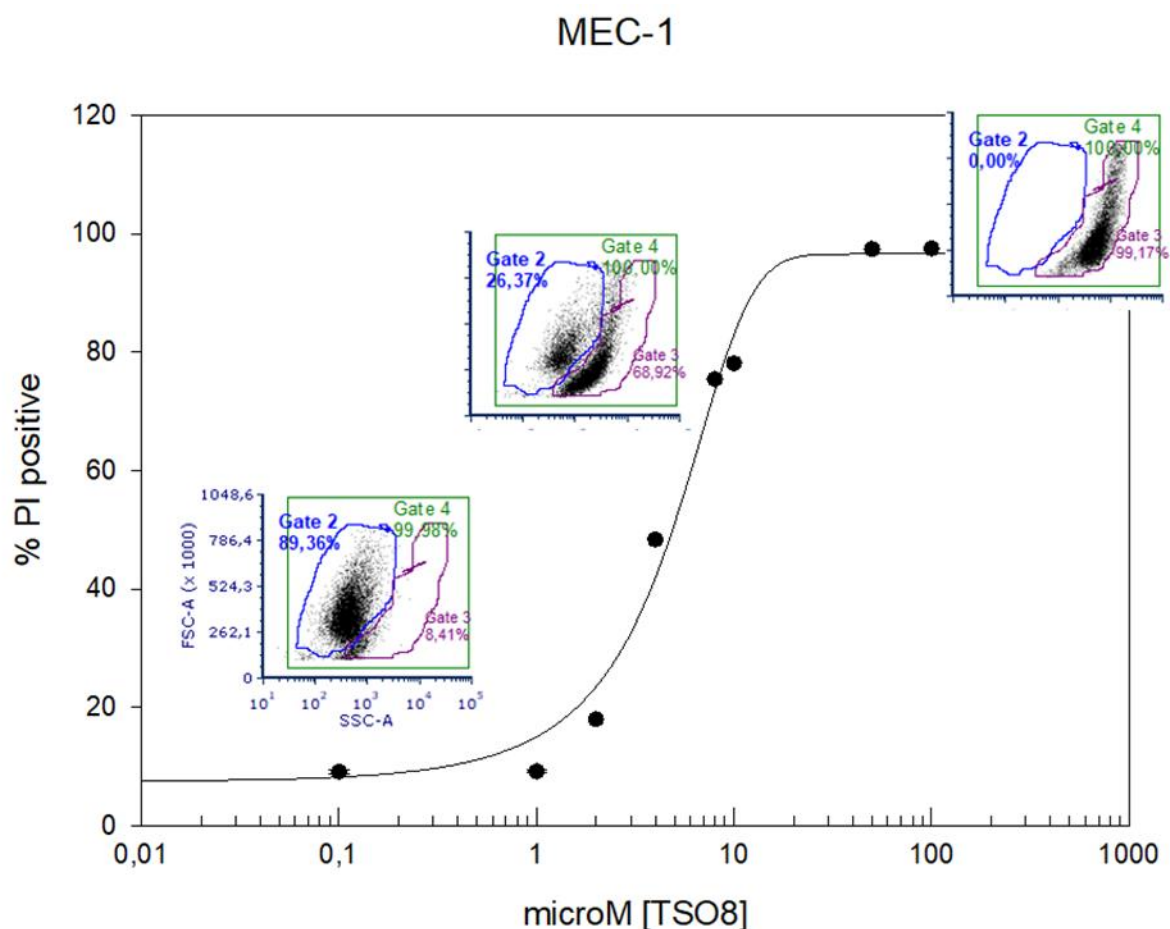


Figure 29 Percentage of PI positive (dead) MEC-1 cells at different TSO8 (1-39) concentrations.

*Molecular modeling*

#### 4.4 Molecular modeling

Five hundred nanosecond long simulations were performed for TAEMU 1, MESCO 2, TSO8 (1-39) and TSO8\_lin in water solution, using molecular dynamics simulation package Gromacs 2021.2 (62). The analysis consists of calculating root mean square deviation (RMSD), radius of gyration, helicity per residue and the most likely secondary structure assignment for the peptides by using DSSP program (72). Results were plotted by using Gnuplot 5.2 software package (73), and VMD is used as visualization tool (74). The results are discussed for each peptide separately in the following paragraphs.

- **TAEMU 1**

The RMSD calculation over simulation time (Figure 30 a)) shows relatively small shifts from the initial structure, suggesting that the peptide is very stable in water, while radius of gyration over time suggests that the peptide is quite compact throughout the duration of the simulation (Figure 30 b)). The helical structure is characteristic for about thirteen N-terminal

residues (from 3<sup>rd</sup> to 16<sup>th</sup>) which are conformed to helical structure more than 70% of time. Unlike N-terminal, C-terminal residues prefer some other type of structuring (Figure 30 c)).

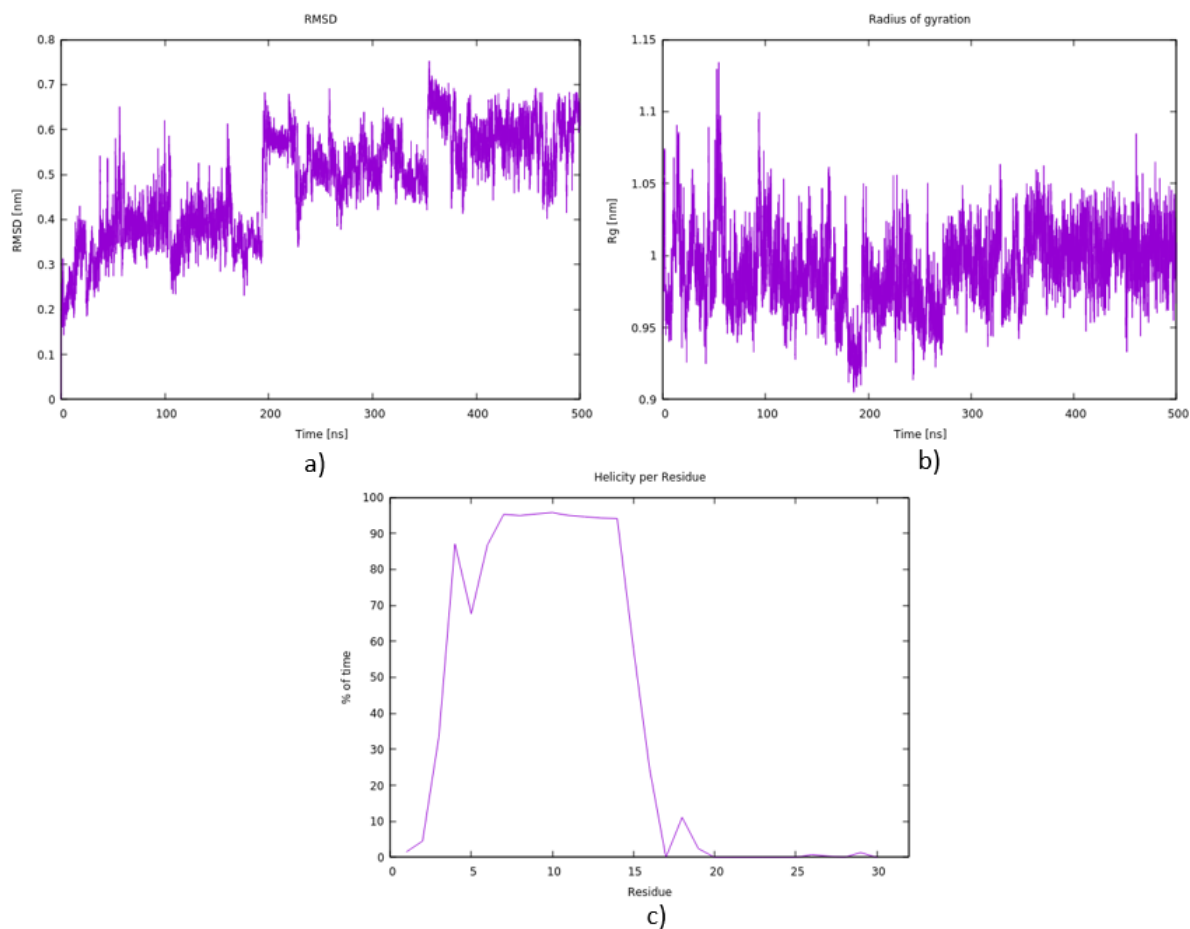


Figure 30 Simulation analysis for TAEMU 1 containing: a) RMSD, b) radius of gyration, c) helicity per residue.

As to the secondary structure, the N-terminal helix, from the 5<sup>th</sup> to the 15<sup>th</sup> amino acid residue is preserved during simulation time. Regarding the C-terminus, it is mostly unstructured, with transient formation of turn, bend, and B-bridge structure (see Figure 31).



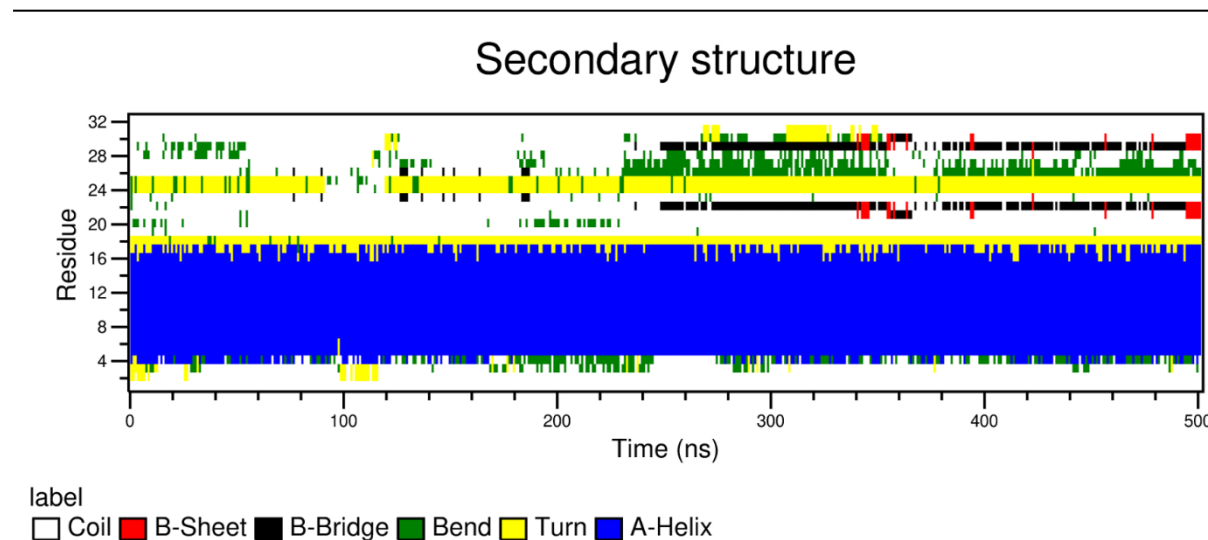


Figure 31 DSSP analysis of TAEMU 1 solvated in water for the simulation time of 500 ns

TEAMU 1 structure at the beginning (Figure 32 a)) and the one at the end of simulation (Figure 32 b)) differ only slightly.

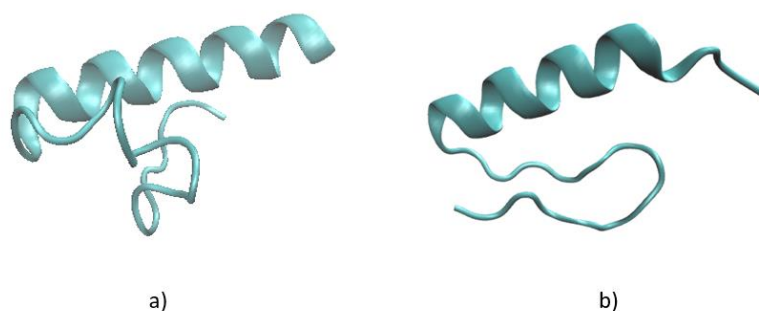


Figure 32 Structure of TAEMU 1 a) at 0 ns, b) after 500 ns of simulation.

- **MESCO 2**

Unlike TAEMU 1, the RMSD calculation for MESCO 2 (Figure 33 a)) shows a major shift from the initial structure, suggesting that the peptide is not stable in water. The radius of gyration of MESCO 2 seems to be quite compact during about first 300 ns of simulation, after which compactness is clearly disturbed (Figure 33 b)). Similarly, visualization by VMD shows that the unfolding starts after about 300 ns of the simulation time. Initial structure consisted of two helical parts (from 2<sup>nd</sup> -13<sup>th</sup> and 17<sup>th</sup> -27<sup>th</sup> residue) with a coil in between helices, followed by turn and coil at C-terminal, which are also visible as peaks in Figure 33

c), and the unfolding process contributes that these helical residues are helical for about 50% of the total simulation time.

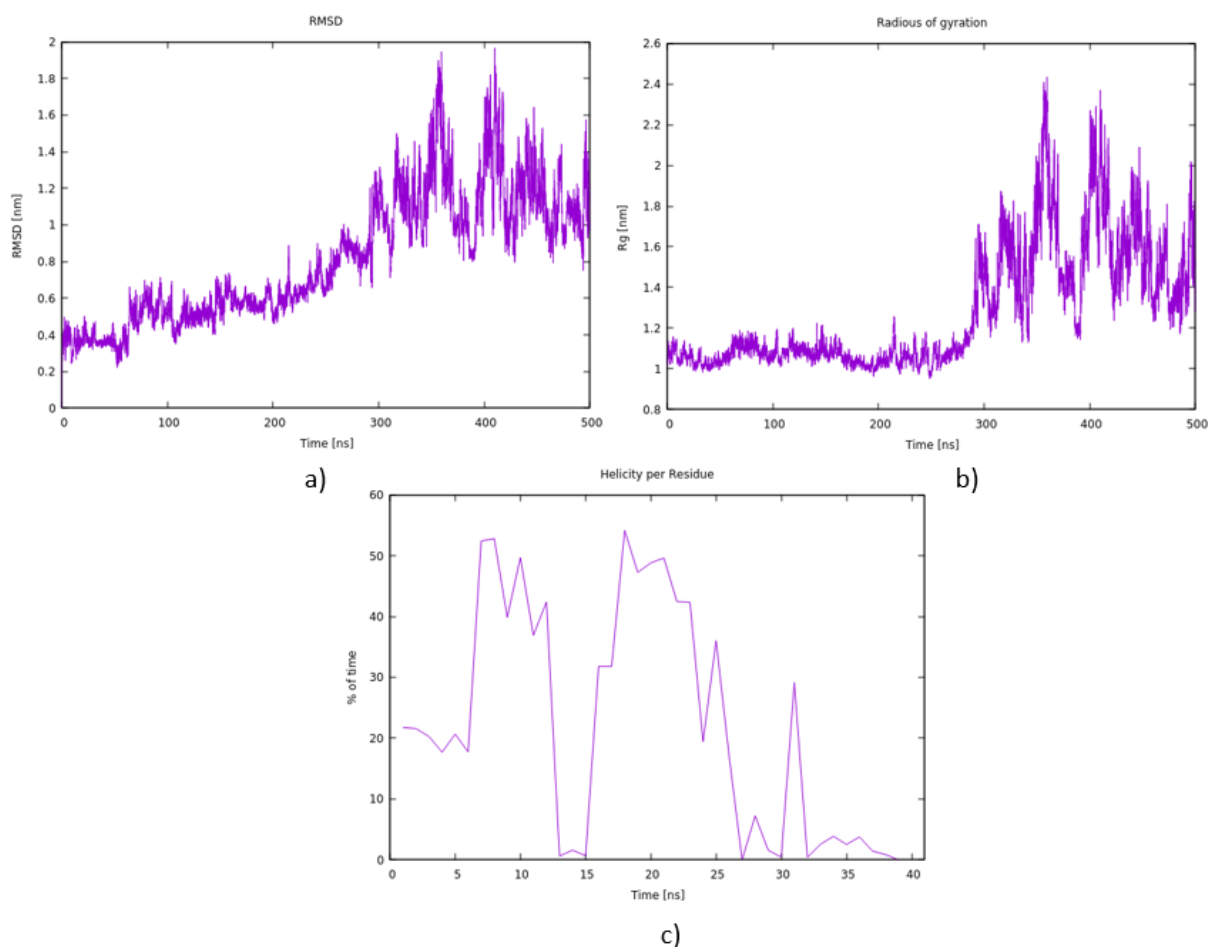


Figure 33 Simulation analysis for MESCO 2 containing: a) RMSD, b) radius of gyration, c) helicity per residue.

As to the secondary structure, two helical parts (from 2<sup>nd</sup> -13<sup>th</sup> and 17<sup>th</sup> -27<sup>th</sup>) starting from about 100 ns to 300 ns lose their initial helical structuring and remain unstructured for the last 200 ns simulation time. The unstructured C-terminal residues develop partly turn, band and B-bridge structure throughout last 250 ns simulation time (see Figure 34).

## Secondary structure

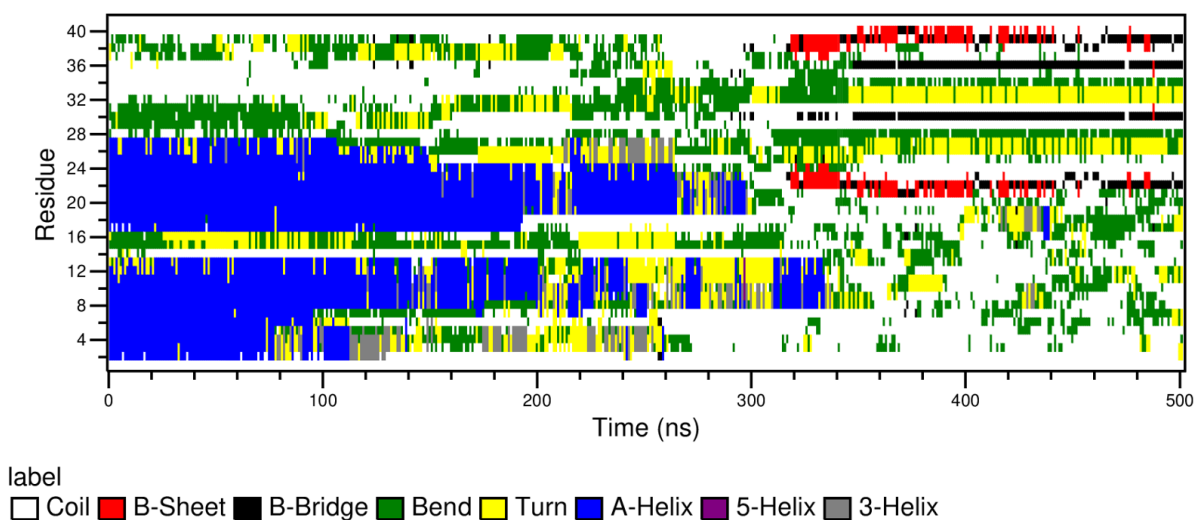


Figure 34 DSSP analysis of MESCO 2 solvated in water for the simulation time of 500 ns

The structured initial conformation and unstructured final one at the 500 ns simulation time are shown in Figure 35.

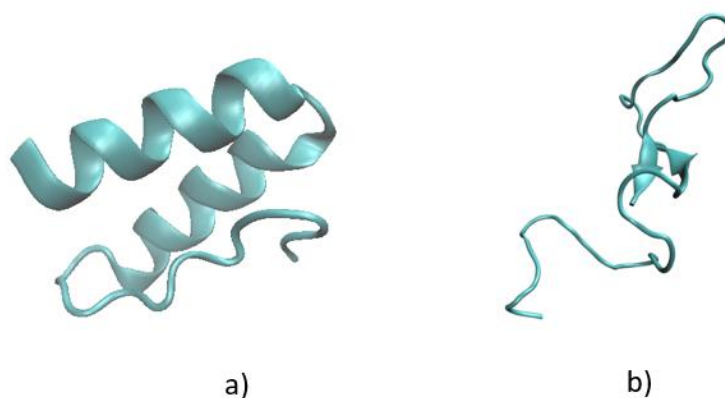


Figure 35 Structure of MESCO 2 a) at 0 ns, b) after 500 ns of simulation.

- **TSO8**

The cQUARK prediction did not include formation of the disulfide bridge between two cysteines residues for TSO8 (1-39) as these residues in the 3D model were spaced 6.25 Å spaced apart, while distance to form a bridge needs to be less than 2 Å. Therefore, incorporation of disulfide bond was done by with Disulfide by Design v 2.13 (75) (Appendix D). TSO8 (1-39) and TSO8\_lin differ in two amino acids, where 24<sup>th</sup> and 28<sup>th</sup> cysteines in

TSO8 (1-39) (Figure 36 a) are replaced by alanine to eliminate disulfide bridge in TSO8\_lin (Figure 36 b).

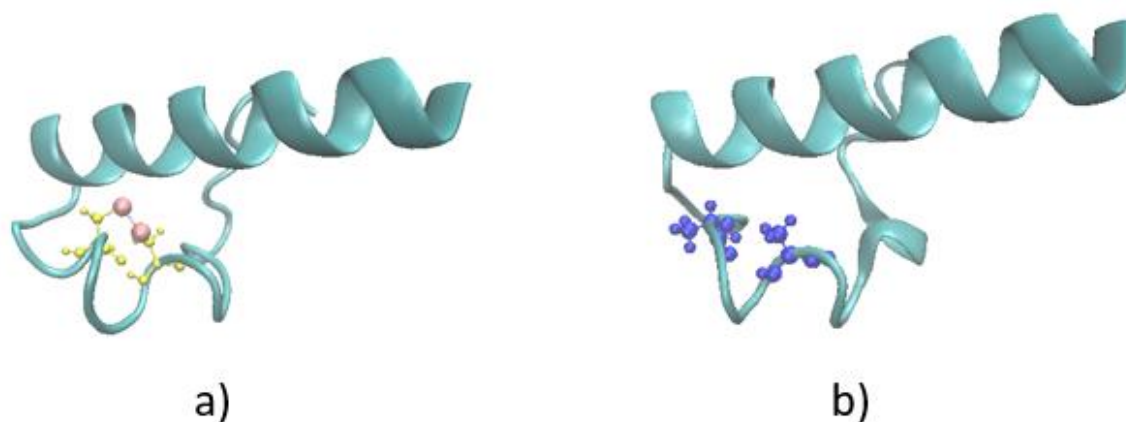


Figure 36 a) Structure of TSO8 (1-39) obtained by cQUARK and with added disulfide bridge by Disulfide by Design v 2.13, with highlighted cysteine residues (colored in yellow), sulfur atoms and disulfide bridge (colored in orange). b) Structure of TSO8\_lin obtained by cQUARK with highlighted alanine residues (colored in blue).

The RMSD calculation (Figure 37 a) shows relatively small shifts from the initial structure for the first 200 ns, after which value of RMSD is increased, reaching value of about 1.6 nm. Radius of gyration over time (Figure 37 b)), suggests TSO8 (1-39) is quite compact with the transient change about 200–300 ns. The N-terminal residues are helical more than 80% of the time, while other residues confirm other structures (Figure 37 c).

DSSP analysis confirms that the N-terminal helix is preserved during simulation, and also indicates that 19<sup>th</sup> to the 21<sup>st</sup>, residues may form 3-helices (see Figure 38).

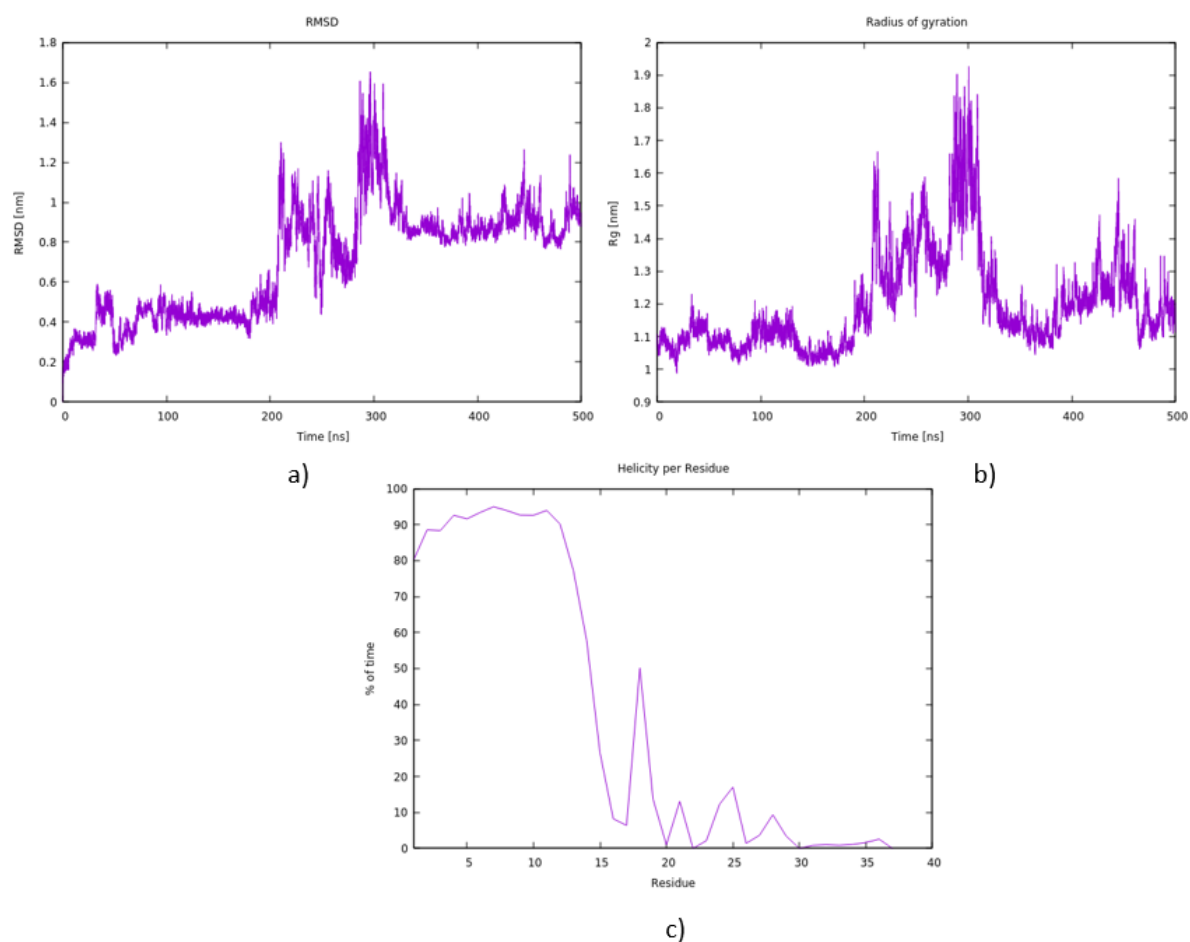


Figure 37 Simulation analysis for TSO8 (1-39) containing: a) RMSD, b) radius of gyration, c) helicity per residue.

## Secondary structure

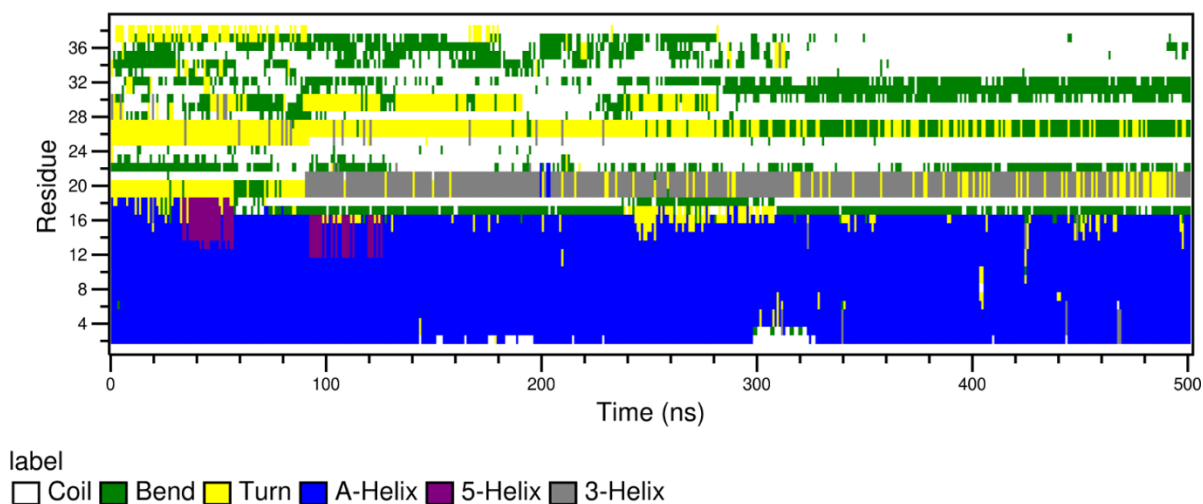
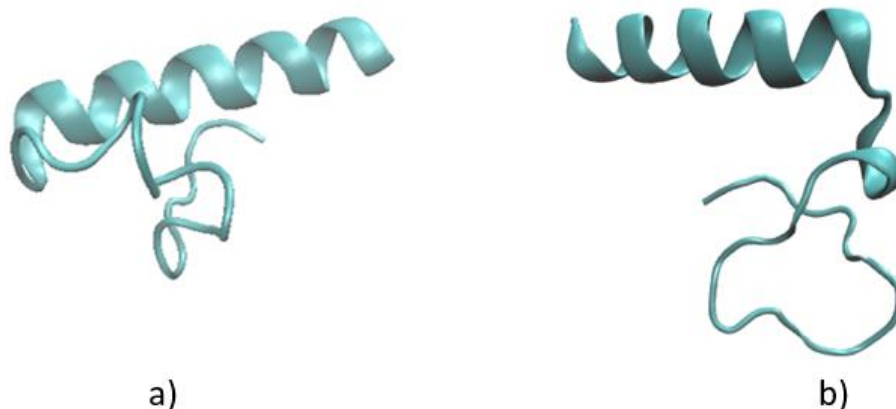


Figure 38 DSSP analysis of TSO8 (1-39) solvated in water for the simulation time of 500 ns.

The initial TSO8 (1-39) structure and the one at the end of simulation are presented in Figure 39, and the peptide seems to be very stable in water during simulation time of 500 ns.



*Figure 39 Structure of TSO8 (1-39) a) at 0 ns, b) after 500 ns of simulation.*

Unlike, TSO8 (1-39), the RMSD calculation for TSO8\_lin (Figure 40 a)) shows a shift from the initial structure in the first 50 ns, and oscillatory behavior for the rest of the simulation time. Radius of gyration indicates that peptide structure oscillates (see Figure 40 b)). The N-terminal residues are helical more than 80% of the time, and non-helical structure for C-terminal residue, behaves the same as TSO8 (1-39) (see Figure 40 c)).

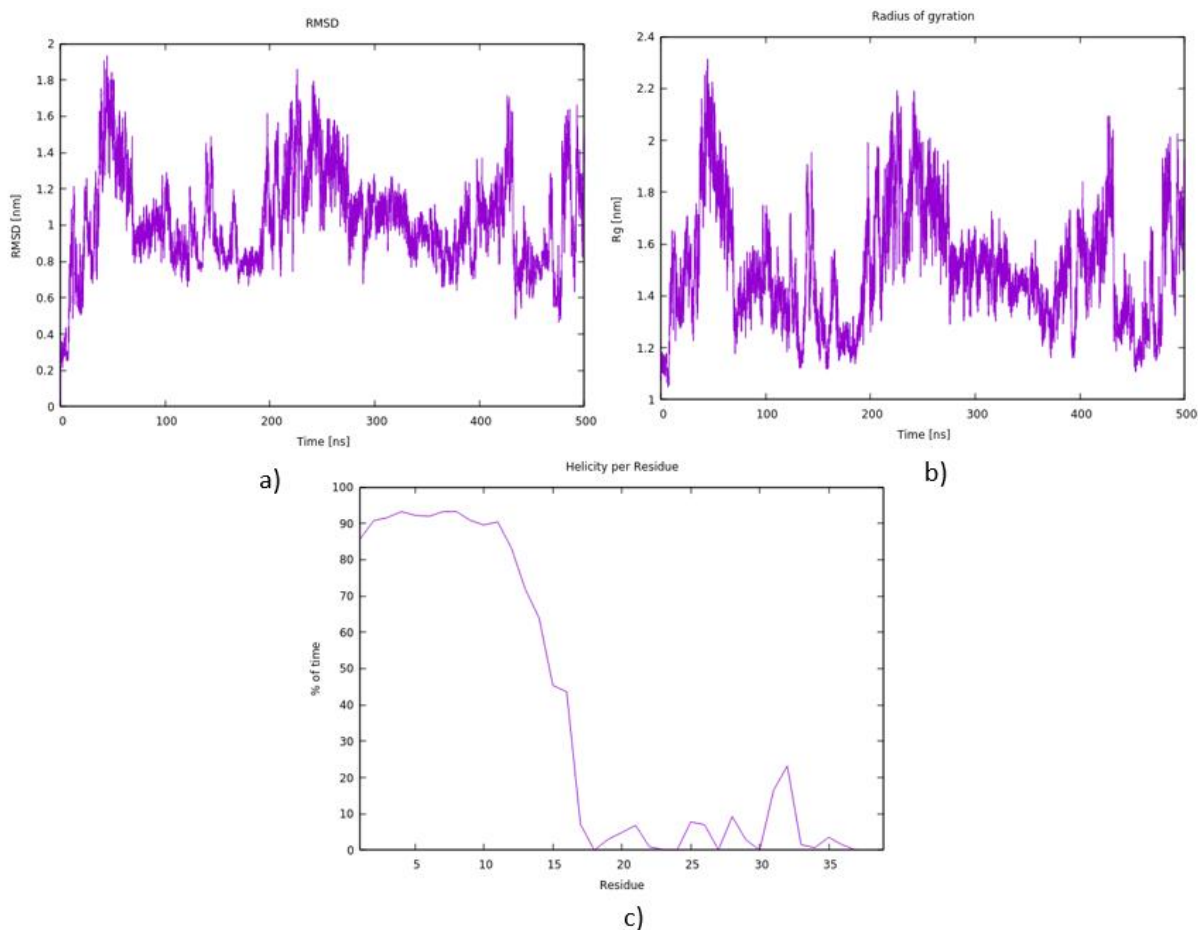


Figure 40 Simulation analysis for TSO8\_lin containing: a) RMSD, b) radius of gyration, c) helicity per residue.

Similar to TSO8 (1-39), DSSP analysis for TSO8\_lin confirms that N-terminal helix is preserved during simulation, while the other part of peptide is unstructured (see Figure 41). However, stability of N-terminal helix is reduced toward the middle of peptide compared to the TSO8 (1-39). Also, formation of 3-helical structure is not observed, but the C-terminal residues show a transient B-bridge, bend, turn or coil structure.

## Secondary structure

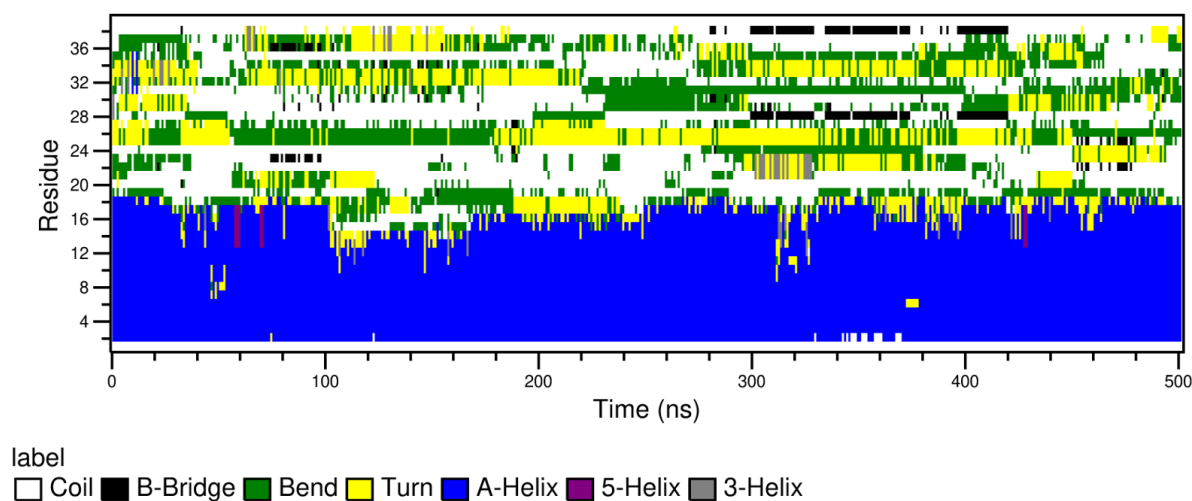


Figure 41 DSSP analysis of TSO8\_lin solvated in water for the simulation time of 500 ns.

We may attribute the observed oscillations in RMSD and radius of gyration to the movement of the unstructured C-terminal residues closer or further apart from the N-terminal helix. Figure 42 compares initial structure and structure at the 500 ns simulation time.

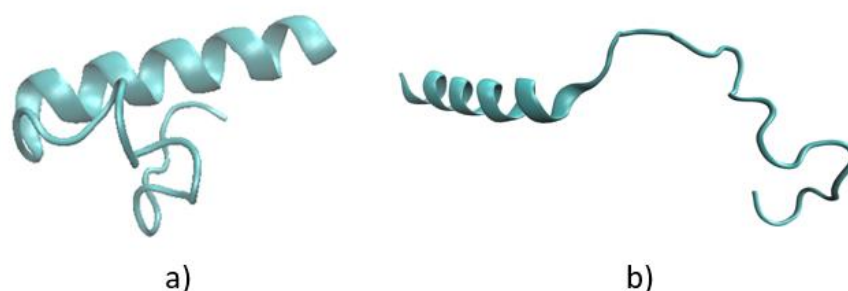


Figure 42 Structure of TSO8\_lin a) at 0 ns, b) after 500 ns of simulation.

The results as well as VMD visualization indicate that TSO8 (1-39) and TSO8\_lin behave similarly when solvated in water. For both peptides N-terminal helix remains preserved throughout simulation, which is expected because they are identical at N-terminal. However, the disulfide bridge present in TSO8 (1-39) reduced the flexibility of the C-terminal residue structure compared to the TSO8\_lin C-terminal residues structuring.



## 5 Conclusion

Peptide candidates for our study were obtained by genome mining methods from several different helminth organism genomes (Appendix A). In this manner, we were successful in identifying 31 potential AMP sequences. Following detailed bioinformatical analysis three peptides (TAEMU 1, MESCO 2 and TSO8) from *Taenia solium*, *Taenia multiceps* and *Mesocestoides corti* were selected for synthesis.

Certain peptides (TSO8\_full, TSO8\_3-39, MESCO\_2 and TAEMU) have been synthesized commercially by GenicBio (Shanghai, China), while most of the TSO8 fragments, used for structure-activity studies, have been synthesized in Prof. Alessandro Tossi's laboratory at the Department of Life Sciences, University of Trieste, Trieste, Italy.

Structural studies with CD spectroscopy were done for all selected peptides and TSO fragments in SPB buffer, different proportions of TFE and SDS. Peptides showed little structuring in aqueous buffer, while addition of TFE resulted in an increase in secondary structure content (mostly helix). In SDS environment, the peptides and fragments showed some content of regular secondary structure, in most cases similar to those obtained in the 50% TFE solution.

Peptides were screened for antibacterial activity against representatives of Gram-negative and Gram-positive bacteria. All peptides and fragments showed relatively high activity against both tested strains (*E. coli* and *S. aureus*), with the exception of TAEMU 1, for which MIC/MBC values were not determined against *S. aureus* as they were higher than the maximum concentration used for testing (64  $\mu\text{M}$ ).

Cell toxicity tests were performed by exposing human MEC-1 cells to peptides in the presence of the impermeant dye propidium iodide, followed by flow cytometry measurements. All peptides and fragments showed some cytotoxicity, with TAEMU 1 and TSO8 (21-39) showing a significantly lower cytotoxic activity than those obtained by other peptides and fragments.

The results of biological activity studies indicate TSO8 (3-23) as the best candidate for further development. Its antibacterial activity was appreciable (8  $\mu\text{M}$  and 4  $\mu\text{M}$  against *E. coli* and *S. aureus*, respectively) with moderate cytotoxicity against MEC1 cells (50% PI positive cells at peptide concentration of about 60  $\mu\text{M}$ ).

Molecular modelling consisted of 500 ns long MD simulations with TAEMU 1, MESCO 2, TSO8 (1-39) and TSO8\_lin in water. Detailed analysis included RMSD calculation, radius of gyration, helicity per residue and secondary structure content during simulation time and confirmed high stability during 500 ns simulation runs for all tested peptides, except MESCO 2.

Considering the full sequence of the tested peptides it can be seen that TSO8 and MESCO2 have an aromatic residue at position 2 (W and F, respectively). Given the fact that Sato et al. (76) have previously reported the importance of this residue at specific position in cecropin peptides found in insects, we have also explored the role of GW at N-terminus in TSO8 and its fragments.

Namely, the TSO8 (1-39) with GW at N-terminus proved to be less potent compared to TSO8 (3-39) (no GW segment). On the other, shorter, N-terminal fragments, TSO8 (1-23) and TSO (3-23) are in line with previous reports, with former peptide being slightly more potent compared to the fragment without GW at the beginning of the sequence. This would suggest that GW does play a role in biological activity of helminth peptides, but its contribution depends on a number of other factors (e.g., the overall length of the peptide and the presence or absence of disulfide bridge).

Further research needs to be done to determine mechanism of action and possible biomedical application of selected peptides.

## 6 Literature

1. Antimicrobial resistance [Internet]. [cited 2021 Sep 16]. Available from: <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
2. Cotter PD, Ross RP, Hill C. Bacteriocins — a viable alternative to antibiotics? *Nat Rev Microbiol*. 2013 Feb;11(2):95–105.
3. Bahar A, Ren D. Antimicrobial Peptides. *Pharmaceuticals*. 2013 Nov 28;6(12):1543–75.
4. Rončević T, Puizina J, Tossi A. Antimicrobial Peptides as Anti-Infective Agents in Pre-Post-Antibiotic Era? *Int J Mol Sci*. 2019 Nov 14;20(22):5713.
5. Gordon YJ, Romanowski EG, McDermott AM. A Review of Antimicrobial Peptides and Their Therapeutic Potential as Anti-Infective Drugs. *Curr Eye Res*. 2005 Jan;30(7):505–15.
6. Fjell CD, Hiss JA, Hancock REW, Schneider G. Designing antimicrobial peptides: form follows function. *Nat Rev Drug Discov*. 2012 Jan;11(1):37–51.
7. Wang G, Mishra B, Lau K, Lushnikova T, Golla R, Wang X. Antimicrobial Peptides in 2014. *Pharmaceuticals*. 2015 Mar 23;8(1):123–50.
8. Zasloff M. Antimicrobial peptides of multicellular organisms. *Nature*. 2002 Jan;415(6870):389–95.
9. Ilić N. Design and testing of Adepantins – functional artificial antibiotics. Dr Thesis. 2013;167.
10. Rončević T, Gajski G, Ilić N, Goić-Barišić I, Tonkić M, Zoranić L, et al. PGLa-H tandem-repeat peptides active against multidrug resistant clinical bacterial isolates. *Biochim Biophys Acta BBA - Biomembr*. 2017 Feb;1859(2):228–37.
11. Rončević T, Vukičević D, Krce L, Benincasa M, Aviani I, Maravić A, et al. Selection and redesign for high selectivity of membrane-active antimicrobial peptides from a dedicated sequence/function database. *Biochim Biophys Acta BBA - Biomembr*. 2019 Apr;1861(4):827–34.
12. Rončević T. IDENTIFICATION, REDESIGN AND CHARACTERIZATION OF ANURAN ANTIMICROBIAL PEPTIDES. Dr Thesis. 2019;94.
13. Rončević T, Vukičević D, Ilić N, Krce L, Gajski G, Tonkić M, et al. Antibacterial Activity Affected by the Conformational Flexibility in Glycine–Lysine Based  $\alpha$ -Helical Antimicrobial Peptides. *J Med Chem*. 2018 Apr 12;61(7):2924–36.
14. Nijnik A, Hancock R. Host defence peptides: antimicrobial and immunomodulatory activity and potential applications for tackling antibiotic-resistant infections. *Emerg Health Threats J*. 2009 Apr 11;2(1):7078.

15. Cederlund A, Gudmundsson GH, Agerberth B. Antimicrobial peptides important in innate immunity: Antimicrobial peptides important in innate immunity. *FEBS J.* 2011 Oct;278(20):3942–51.
16. Tossi A, Sandri L, Giangaspero A. Amphipathic,  $\alpha$ -helical antimicrobial peptides. *Biopolymers.* 2000;55(1):4–30.
17. Antimicrobial Peptide Database [Internet]. [cited 2021 Sep 23]. Available from: <https://aps.unmc.edu/home>
18. CAMP (Collection of Anti-Microbial Peptides) [Internet]. [cited 2021 Sep 24]. Available from: <http://www.camp.bicnirrh.res.in/>
19. DADP [Internet]. [cited 2021 Nov 30]. Available from: <http://split4.pmfst.hr/dadp/>
20. Defensins Knowledgebase [Internet]. [cited 2021 Nov 30]. Available from: <http://defensins.bii.a-star.edu.sg/>
21. Frontiers | Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields | Microbiology [Internet]. [cited 2021 Sep 27]. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.582779/full>
22. Zhang L, Gallo RL. Antimicrobial peptides. *Curr Biol.* 2016 Jan;26(1):R14–9.
23. Antimicrobial Peptides [Internet]. [cited 2021 Sep 27]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3873676/>
24. Nguyen LT, Haney EF, Vogel HJ. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends Biotechnol.* 2011 Sep;29(9):464–72.
25. Juretić D, Vukičević D, Ilić N, Antcheva N, Tossi A. Computational Design of Highly Selective Antimicrobial Peptides. *J Chem Inf Model.* 2009 Dec 28;49(12):2873–82.
26. Hansen PR, editor. Antimicrobial Peptides: Methods and Protocols [Internet]. New York, NY: Springer New York; 2017 [cited 2021 Sep 6]. (Methods in Molecular Biology; vol. 1548). Available from: <http://link.springer.com/10.1007/978-1-4939-6737-7>
27. Shah J, Shahidullah A. *Ascaris lumbricoides*: A Startling Discovery during Screening Colonoscopy. *Case Rep Gastroenterol.* 2018 May 31;12(2):224–9.
28. McKay DM, Shute A, Lopes F. Helminths and intestinal barrier function. *Tissue Barriers.* 2017 Jan 2;5(1):e1283385.
29. Helminths: Pathogenesis and Defenses - Medical Microbiology - NCBI Bookshelf [Internet]. [cited 2021 Nov 13]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK8191/>
30. Helminths: Structure, Classification, Growth, and Development - Medical Microbiology - NCBI Bookshelf [Internet]. [cited 2021 Nov 15]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK8282/>

31. Rapin A, Harris NL. Helminth–Bacterial Interactions: Cause and Consequence. *Trends Immunol.* 2018 Sep;39(9):724–33.
32. Vemuri R, Shankar EM, Chieppa M, Eri R, Kavanagh K. Beyond Just Bacteria: Functional Biomes in the Gut Ecosystem Including Virome, Mycobiome, Archaeome and Helminths. *Microorganisms.* 2020 Mar 28;8(4):483.
33. Zaiss MM, Harris NL. Interactions between the intestinal microbiome and helminth parasites. *Parasite Immunol.* 2016 Jan;38(1):5–11.
34. Aguilar-Díaz H, Bobes RJ, Carrero JC, Camacho-Carranza R, Cervantes C, Cevallos MA, et al. The genome project of *Taenia solium*. *Parasitol Int.* 2006 Jan;55:S127–30.
35. Harris TW. WormBase: a multi-species resource for nematode biology and genomics. *Nucleic Acids Res.* 2004 Jan 1;32(90001):411D – 417.
36. Leoni G, De Poli A, Mardirossian M, Gambato S, Florian F, Venier P, et al. Myticalins: A Novel Multigenic Family of Linear, Cationic Antimicrobial Peptides from Marine Mussels (*Mytilus* spp.). *Mar Drugs.* 2017 Aug 22;15(8):261.
37. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011 Oct;8(10):785–6.
38. Zelezetsky I, Pag U, Antcheva N, Sahl H-G, Tossi A. Identification and optimization of an antimicrobial peptide from the ant venom toxin pilosulin. *Arch Biochem Biophys.* 2005 Feb;434(2):358–64.
39. Liu Y, Wang Z, Huang W, Pang S, Qian L, Zhang Y, et al. De Novo Sequencing and High-Contiguity Genome Assembly of *Moniezia expansa* Reveals Its Specific Fatty Acid Metabolism and Reproductive Stem Cell Regulatory Network. *Front Cell Infect Microbiol.* 2021 Jul 6;11:693914.
40. Cheng T, Liu G-H, Song H-Q, Lin R-Q, Zhu X-Q. The complete mitochondrial genome of the dwarf tapeworm *Hymenolepis nana*—a neglected zoonotic helminth. *Parasitol Res.* 2016 Mar;115(3):1253–62.
41. PSIPRED Workbench [Internet]. [cited 2021 Nov 5]. Available from: <http://bioinf.cs.ucl.ac.uk/psipred/>
42. C-QUARK: Contact Assisted Ab Initio Protein Structure Prediction [Internet]. [cited 2021 Nov 5]. Available from: <https://zhanggroup.org/C-QUARK/>
43. RPBS Web Portal [Internet]. [cited 2021 Nov 5]. Available from: <https://mobylye.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3>
44. PHYRE2 Protein Fold Recognition Server [Internet]. [cited 2021 Nov 5]. Available from: <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>
45. Peptide calculator [Internet]. [cited 2021 Dec 7]. Available from: <https://www.bachem.com/knowledge-center/peptide-calculator/>

46. BLAST: Basic Local Alignment Search Tool [Internet]. [cited 2021 Nov 5]. Available from: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
47. Amblard M, Fehrentz J-A, Martinez J, Subra G. Methods and protocols of modern solid phase peptide synthesis. *Mol Biotechnol*. 2006;33:16.
48. Peptide Companion - Excel Version – Community Project - European Peptide Society [Internet]. [cited 2021 Nov 30]. Available from: <https://www.eurpepsoc.com/peptide-companion-excel-version-community-project/>
49. Circular Dichroism - Chemistry LibreTexts [Internet]. [cited 2021 Nov 2]. Available from: [https://chem.libretexts.org/Bookshelves/Physical\\_and\\_Theoretical\\_Chemistry\\_Textbook\\_Maps/Supplemental\\_Modules\\_\(Physical\\_and\\_Theoretical\\_Chemistry\)/Spectroscopy/Electronic\\_Spectroscopy/Circular\\_Dichroism](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Electronic_Spectroscopy/Circular_Dichroism)
50. Micsonai A, Wien F, Kernya L, Lee Y-H, Goto Y, Réfrégiers M, et al. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci*. 2015 Jun 16;112(24):E3095–103.
51. Micsonai A, Wien F, Bulyáki É, Kun J, Moussong É, Lee Y-H, et al. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res*. 2018 Jul 2;46(W1):W315–22.
52. Micsonai A, Bulyáki É, Kardos J. BeStSel: From Secondary Structure Analysis to Protein Fold Prediction by Circular Dichroism Spectroscopy. In: Chen YW, Yiu C-PB, editors. *Structural Genomics* [Internet]. New York, NY: Springer US; 2021 [cited 2021 Nov 21]. p. 175–89. (Methods in Molecular Biology; vol. 2199). Available from: [http://link.springer.com/10.1007/978-1-0716-0892-0\\_11](http://link.springer.com/10.1007/978-1-0716-0892-0_11)
53. Reed J, Reed TA. A Set of Constructed Type Spectra for the Practical Estimation of Peptide Secondary Structure from Circular Dichroism. *Anal Biochem*. 1997 Dec;254(1):36–40.
54. Stacchini A, Aragno M, Vallario A, Alfarano A, Circosta P, Gottardi D, et al. MEC1 and MEC2: two new cell lines derived from B-chronic lymphocytic leukaemia in prolymphocytoid transformation. *Leuk Res*. 1999 Feb;23(2):127–36.
55. Becker OM, Karplus M. A guide to biomolecular simulations. Dordrecht, Netherlands: Springer; 2006. 220 p. (Focus on structural biology).
56. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol*. 2011 Dec;9(1):71.
57. Guvench O, MacKerell AD. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In: Kukol A, editor. *Molecular Modeling of Proteins* [Internet]. Totowa, NJ: Humana Press; 2008 [cited 2021 Dec 1]. p. 63–88. (Walker JM, editor. Methods in Molecular Biology; vol. 443). Available from: [http://link.springer.com/10.1007/978-1-59745-177-2\\_4](http://link.springer.com/10.1007/978-1-59745-177-2_4)
58. Bonomi M, Camilloni C, editors. *Biomolecular Simulations: Methods and Protocols* [Internet]. New York, NY: Springer New York; 2019 [cited 2021 Nov 30]. (Methods in

- Molecular Biology; vol. 2022). Available from: <http://link.springer.com/10.1007/978-1-4939-9608-7>
59. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004 Jul 15;25(9):1157–74.
  60. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983;4(2):187–217.
  61. Christen M, Hünenberger PH, Bakowies D, Baron R, Bürgi R, Geerke DP, et al. The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem*. 2005 Dec;26(16):1719–51.
  62. Welcome to the GROMACS documentation! — GROMACS 2021 documentation [Internet]. [cited 2021 Dec 8]. Available from: <https://manual.gromacs.org/documentation/2021/index.html#>
  63. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017 Jan;14(1):71–3.
  64. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983 Jul 15;79(2):926–35.
  65. CHARMM-GUI [Internet]. [cited 2021 Dec 11]. Available from: <https://charmm-gui.org/>
  66. Jo S, Kim T, Im W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. Yuan A, editor. *PLoS ONE*. 2007 Sep 12;2(9):e880.
  67. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*. 1981 Dec;52(12):7182–90.
  68. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 1984 Oct 15;81(8):3684–90.
  69. Hess B, Bekker H, Berendsen HJC. 3 LINCS: a linear constraint solver for molecular simulations. :16.
  70. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995 Nov 15;103(19):8577–93.
  71. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996 Feb;14(1):33–8.
  72. DSSP [Internet]. [cited 2022 Jan 18]. Available from: <https://swift.cmbi.umcn.nl/gv/dssp/>
-

73. gnuplot homepage [Internet]. [cited 2022 Jan 11]. Available from: <http://www.gnuplot.info/>
74. VMD - Visual Molecular Dynamics [Internet]. [cited 2021 Nov 30]. Available from: <https://www.ks.uiuc.edu/Research/vmd/>
75. Disulfide by Design [Internet]. [cited 2022 Jan 25]. Available from: <http://cptweb.cpt.wayne.edu/DbD2/index.php>
76. Sato H, Feix JB. Peptide–membrane interactions and mechanisms of membrane destruction by amphipathic  $\alpha$ -helical antimicrobial peptides. *Biochim Biophys Acta BBA - Biomembr.* 2006 Sep;1758(9):1245–56.



## 7 Appendix A

Table 5 consists of names and sequences of AMPs which were identified in different helminth species. Bioinformatical analysis of the sequences included, BLASTp search against non-redundant database, tBLASTn search against whole shotgun contigs (wgs), secondary structure prediction (PSIPRED), and tertiary structure predictions (cQUARK, PEPFOLD 3.5, *Phyre<sup>2</sup>*).

Table 5 Names and sequences of peptides candidates

Hydratigena_taeiformis_1	WGLRRRIGGAFRRVGDGIRRVFRGPWKICIPS CPKGGRRPRSWDA
Tsolium4_expressed	WGSFRRRISGAFRRVGDGIRRVFRKPWKICLGR CPGGTRRPGPLNP
Tsolium6	WGSFRRRISGAFRRVGGAFRRIGDGIRRAFRRGP WKICVPKCPEGRRPGPWTAA
Mesocestoides_corti	WRRLRRRISGGLRRIFRKPRRICFPYCPTGPRYP GPRPY
Tsolium1	WRRFRRRINRAFRRIGGGVRRVFIEPWKICLGK CLPGRKRPVLLTLQFLRID
Tsolium3	WRRFRRRINRAFRRIGGGIRRVFIEPWKICLGK LPGRKIPVLLTL
Tsolium2	WRRFRRRISRTFRRIGDGIRRVFREPWKICLGK CPHGRKRPVLLTL
Hydratigena_taeiformis_2	WRRFRRRLRRAFRRVRDGIIRRVFREPWRVCLG RCNHGRKPPVPLTF
Hydratigena_taeiformis_3	WRRFRRMLSRALRRVRDKFRKLFLEPWVCL GRCKHRRKPPVPLTL
Echinococcus_granulosus_2_expressed_conserved_prot ein_Echinococcus_granulosus	WRRIRRSIRRRIRSIFRKPRRICFPYCPKG- PKKGREN
Echinococcus_granulosus_3_hypothetical_protein_EGR _08647_Echinococcus_granulosus	WRRIRRSIRRRIRSIFRKPRRICFPYCPKGPCKG PYNENGLAIT
Echinococcus_granulosus	WRRIRRSIRRRIRSIFRKPRRICFPYCPKG- PKKGREN
Echinococcus_multilocularis	WRRIRRSIRRRIRNIFRKPRRICFPYCPKGPCKG REN
Tsolium8_expressed	WRRLRRSIRRRIRRRIFRKPRRICFPYCPKGPCKG RGDF

Mesocestoides_corti_2_unnamed_protein_product_Mesocestoides_corti	FFRRIGRAFSRVGRGIGRGFRQLGRMLPRGNYK ICLGRCP
Mesocestoides_corti_3_unnamed_protein_product_Mesocestoides_corti	FLRRIGRAFSRVGRGIGRGFRQLGRMLPRGNYR ICLGRCPR
Hydratigena_taeiformis_4	WRRFRRRISGAFRRIGGGIRRLRGWRGPVSP HPYPGARFPYYAR
Tsolium5	WRRIRRRISGAFRRAGDGIRRVFRKPWKVFIGH LPHGGWNPGLLNA
Tsolium9	WRRFRQPRRAFRRIGNGIRRFRLRKHGDILIPIV KGVGRLRRRAV
Taenia_multiceps_2_type:internal_len:118_gc:universal_JR941162.1:2-352(+)	WSRFRRRLRRAFRRVGDVSRNFLRKHKDHLP
Hydratigera_taeiformis_5_unnamed_protein_product_Hydatigera_taeiaeformis	WCRFRRILSRALRRVREKI
Rodentolepsis_nana_unnamed_protein_product_Rodentolepis_nana	WRRFRSRIRSGLRIRRNIFRRVPIPLLIGILI
Hymenolepsis_diminuta_type:complete_len:113_gc:universal_GHNR01014249.1:479	WRRIKSRIRRGLRRLGNIYRRIPPIFVRVPI
moniezia_expansa_type:5prime_partial_len:125_gc:universal_JL280275.1:385-11(-)	WGRLRRRIGRAFRRLGNVFRIPAPVYAALLA KH
Echinococcus_multicularis_2_expressed_conserved_protein_Echinococcus_multilocularis	WRQSSLRRPRKVPIPTCPNGGRGPGPMTP
Tsolium7	WRRLRFRRPQKIPIPTCPNGGRSPGPMTP
Mesocestoides_corti_4_unnamed_protein_product_Mesocestoides_corti	WRRIRRRISREFRNAINLRRFPLHLIPMIARPG LIIP
Echinococcus_granulosus_4_expressed_protein_Echinococcus_granulosus	LRSRWRRINCGPRHGSRRSQRLPRPSHPPRLPQ RSFALA
Echinococcus_granulosus_5_hypothetical_protein_EGR_01725	LRSRWRRINCGPRHGSRRSQRLPRPSHPPRLPQ RSFALA
Echinococcus_multicularis_3_expressed_protein_Echinococcus_multilocularis	LRSRWRRINCGPRHGSRRSQRFPRPSHPPRLPR RSFALA
Taenia_multiceps_3_type:complete_len:106_gc:universal_JR943802.1:377-60(-)	SHRRWKRINCGPRRRIRNQMRFSRPAHPPRLPY RQLILA

## 8 Appendix B

Bioinformatical analysis included BLASTp search against non-redundant database and tBLASTn search against whole shotgun contigs (wgs). On the following figures are examples of the results. Figure 43 shows results of BLASTp search for TAEMU 1 against non-redundant database. No significant similarity was found.

Job Title: taemu1  
RID: 5VC9GGBK016 Search expires on 03-27 18:02 pm Download All  
Program: Citation  
Database: nr See details  
Query ID: Icl|Query\_67298  
Description: None  
Molecule type: amino acid  
Query Length: 32  
Other reports

Filter Results  
Percent Identity: [ ] to [ ]  
E value: [ ] to [ ]  
Query Coverage: [ ] to [ ]  
Filter Reset

No significant similarity found. For reasons why, [click here](#)

Figure 43 BLASTp search of TAEMU 1 against non-redundant database

Figure 44 shows results of BLASTp search for MESCO 2 against non-redundant database. Unlike TAEMU 1, MESCO 2 found three hits with significant E-value from *Mesocostoides corti*.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<a href="#">unnamed protein product (Mesocostoides corti)</a>	<a href="#">Mesocostoides corti</a>	88.6	88.6	100%	5e-21	100.00%	141	<a href="#">VDD75737.1</a>
<a href="#">unnamed protein product (Mesocostoides corti)</a>	<a href="#">Mesocostoides corti</a>	79.0	79.0	100%	3e-17	95.00%	143	<a href="#">VDD79841.1</a>
<a href="#">unnamed protein product (Mesocostoides corti)</a>	<a href="#">Mesocostoides corti</a>	50.4	50.4	97%	2e-06	60.87%	98	<a href="#">VDD75740.1</a>

Figure 44 BLASTp search of MESCO 2 against non-redundant database

Figure 45 show TSO8\_lin search for MESCO 2 against whole shotgun contigs (wgs). No significant similarity was found.

Your search is limited to records that include: insects (taxid:50557)

Job Title: Protein Sequence  
RID: UY9NMHJK013 Search expires on 12-08 16:17 pm Download All  
Program: Citation  
Database: nt See details  
Query ID: Icl|Query\_202163  
Description: None  
Molecule type: amino acid  
Query Length: 39  
Other reports

Filter Results  
Percent Identity: [ ] to [ ]  
E value: [ ] to [ ]  
Query Coverage: [ ] to [ ]  
Filter Reset

No significant similarity found. For reasons why, [click here](#)

Figure 45 tBLASTn search of TSO8\_lin against whole shotgun contigs (wgs)

## 9 Appendix C

On the following tables are presented results of CD analysis obtained by Bestsel program. Analysis was performed for each peptide and fragment. Results are presented in five columns, containing percentage of helix,  $\beta$ -sheet, turn or random coil for different % of TFE.

*Table 6 CD spectra analysis by Bestsel for TAEMU 1*

Concentration of TFE	(%) helix	(%) sheet	(%) turn	(%) coil
0%	2.0	32.7	15.9	49.4
5%	2.2	34.4	15.2	48.2
10%	5.6	34.9	14.4	45.1
15%	7.0	34.0	13.8	45.2
25%	11.3	31.8	13.8	43.1
40%	5.7	33.9	13.7	46.6
50%	13.3	30.6	13.3	42.7
SDS	12.9	28.3	14.1	44.7

*Table 7 CD spectra analysis by Bestsel for MESCO 2*

Concentration of TFE	(%) helix	(%) sheet	(%) turn	(%) coil
0%	0.3	32.2	17.3	50.2
5%	2.7	32.8	16.4	48.1
10%	4.6	31.5	16.5	47.4
15%	3.8	34.5	15.3	46.4
25%	8.9	30.0	15.4	45.6
40%	12.9	26.0	15.2	45.9
50%	9.5	29.2	15.8	45.5
SDS	10.9	32.2	13.8	43.8

*Table 8 CD spectra analysis by Bestsel for TSO8 (1-39)*

Concentration of TFE	(%) helix	(%) sheet	(%) turn	(%) coil
----------------------	-----------	-----------	----------	----------

0%	0.5	29.5	17.9	52.1
5%	4.5	30.9	16.5	48.1
10%	3.5	30.6	16.6	49.3
15%	7.8	31.2	15.4	45.6
25%	12.1	26.6	16.2	45.1
40%	12.1	24.7	15.7	47.4
50%	8.1	24.4	17.5	50.0
SDS	9.2	30.1	14.9	45.8

*Table 9 CD spectra analysis by Bestsel for TSO8 (3-39)*

Concentration of TFE	(%) helix	(%) sheet	(%) turn	(%) coil
0%	0.3	29.5	19.0	51.3
5%	1.0	30.4	18.6	50.0
10%	7.2	28.9	16.6	47.4
15%	12.2	27.9	15.5	44.4
25%	14.6	24.9	16.0	44.6
40%	15.6	25.8	14.2	45.3
50%	13.3	20.7	19.3	46.7
SDS	16.2	23.8	16.3	43.6

*Table 10 CD spectra analysis by Bestsel for TSO8 (1-23)*

Concentration of TFE	(%) helix	(%) sheet	(%) turn	(%) coil
0%	0.8	21.7	17.9	59.5
5%	6.0	23.7	17.8	52.5
10%	14.2	19.7	16.0	50.1
15%	30.3	18.5	13.0	38.3
25%	34.7	14.9	13.3	37.2
40%	29.9	21.2	11.9	37.0

50%	32.5	15.9	14.2	37.3
SDS	20.9	20.7	16.4	42.0

*Table 11 CD spectra analysis by Bestsel for TSO8 (3-23)*

Concentration of TFE	(%) helix	(%) sheet	(%) turn	(%) coil
0%	0.0	23.3	20.8	55.9
5%	0.0	26.0	19.5	54.4
10%	6.8	20.6	18.4	54.2
15%	21.6	18.0	15.9	44.6
25%	26.2	17.8	15.2	40.8
40%	25.3	15.2	14.2	45.3
50%	28.7	14.0	15.3	42.0
SDS	9.7	25.5	14.8	50.1

*Table 12 CD spectra analysis by Bestsel for TSO8 (23-39)*

Concentration of TFE	(%) helix	(%) sheet	(%) turn	(%) coil
0%	2.1	23.0	17.5	57.5
5%	1.7	22.9	16.8	58.6
10%	1.2	14.6	18.1	66.1
15%	1.8	26.2	16.1	55.8
25%	3.2	26.6	15.9	54.2
40%	3.6	18.8	17.5	60.1
50%	3.8	12.5	19.0	67.4
SDS	9.6	23.0	17.1	50.8

## 10 Appendix D

The cQUARK software prediction for TSO8 (1-39) didn't predict disulfide bond between cysteine residues, which were 6.25 Å spaced apart, while the residues need to be  $\leq 2$  Å to form a bridge. Therefore, incorporation of disulfide bond for TSO8 (1-39) was done by using program Disulfide by Design v 2.13 (75). Figure 46 shows TSO8 (1-39) before and after disulfide bond incorporation.

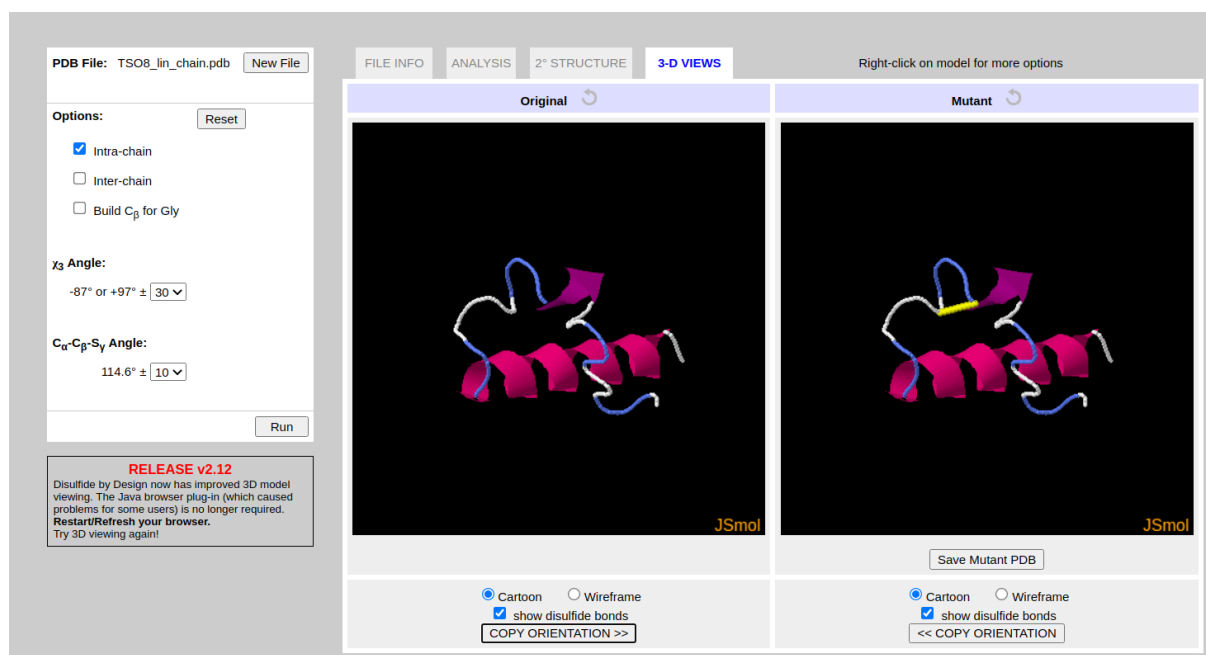


Figure 46 Incorporation of disulfide bond for TSO8 (1-39) by using program Disulfide by Design v 2.13 (72).