

Application of Derandomisation to Uperin 3.x Peptides Circular Dichroism Spectra to Determine Their Secondary Structure Contents

Adewale Olamoyesan^{1,2,3*} , Alison Rodger¹ ¹Department of Molecular Sciences, Faculty of Science and Engineering, Macquarie University, Macquarie Park, Australia²Department of Chemistry, University of Lagos, Akoka, Lagos, Nigeria³College of Science and Computing, Wigwe University, Isiokpo, Rivers, Nigeria

ABSTRACT

Survival of frogs and toads in hostile environments depends on their ability to secrete antimicrobial or host-defence peptides. The Uperin, U3.x, peptides investigated were originally isolated from the skin secretions of a frog endemic to Australia. In this paper, we extend the automated *ad hoc* approach for derandomisation of proteins to U3.x, whose spectra are collected in water and after the addition of buffer to eventually understand how this affects the structure's goodness of fit and spectral prediction. Systematically with the "CD app' *disordered unit*" varying fractions of random coil, RC spectrum (10–90% in increments of 10%) were removed from U3.x's spectra baseline corrected (in molar extinction units) to produce derandomised spectra. Self-organising map spectroscopy, SOMSpec, analysis gives the secondary structure of the derandomised U3.x, afterward, the RC component was added back to determine the percentage of the secondary structure motifs in the original protein or peptide. Most of the applications to U3.x wild types, WTs, and their mutants were straightforward, though, a few of them were not. These exceptions required visual inspection of the spectra. We also found that as the percentage of derandomisation increases, the observed behaviour is like proteins, where folded structures decrease.

KEYWORDS

uperin, derandomised, normalised root mean square deviation, structure, circular dichroism

Received 03 September 2022, revised 29 December 2022, accepted 12 February 2023

INTRODUCTION

A vital component of the innate immune systems of most plants and animals is the antimicrobial peptide, AMP, or host defence peptide.¹ Suitable stress stimulus in the hostile environment where these organisms live, triggers the activity of these peptides against any invading microbe(s).² For the past three decades, the skin secretions of Australian frogs and toads were the source of several host defence peptides. These isolated peptides are rich in AMPs and, or they have one or more neuropeptide activities. Their prevalence in nature demonstrates AMP's effectiveness as a defence compound.^{3–5} They are attractive as starting points in developing new peptide-based antimicrobial agents due to their high activity and selectivity for microbes. The interaction of AMPs with microorganisms' lipid bilayer is always fatal; and subsequently leads to the disruption of their membranes in a specific, but not receptor-mediated, process.²

Uperin 3.x (U3.x) family is a class of peptides originally isolated from the skin secretions of *Uperoleia mjobergii* (a frog endemic to Australia). This type of peptide is known to have antimicrobial activity against Gram-positive bacteria and mammalian cell lines.^{6,7} All three U3.x used herein have 17 amino acid residues, as shown in Table 1.⁸ Among them, U3.4 WT and U3.5 WT have identical peptide sequences, with U3.4 WT being slightly more nonpolar. More than half of the individual positions of the amino acids of U3.6 WT differ from equivalent positions on the other two U3.x WT peptides. Except for U3.5 wt that self-aggregates in a saline buffer the rest are unstructured either in water or buffer.⁹ Ray and her group members¹⁰ observed that U3.x WT aggregates less than their corresponding seventh position variants. Based on this behaviour, Ray et al. rationalised that substituting nonpolar alanine residue for the positively charged amino acid at the seventh position increases fibril formation. In addition, this group conducted mechanistic studies of the initial stages of self-aggregation of these U3.x WT peptides and their variants using molecular dynamics, MD, simulations and

circular dichroism, CD, measurements. Their results show that the U3.x WT peptides demonstrated a lower propensity to β -sheet than the corresponding alanine variants. Likewise, the addition of salt increased hydrophobic interactions between peptides, which resulted in enhanced peptide aggregation.

A well-known assumption in structural biology is that the secondary structure of proteins is a prerequisite to their activity. To the best of our knowledge, there are quite a few methods to estimate the peptide secondary structure,^{11–13} therefore, there is a need for new or existing methods never used for peptides to be adapted for this purpose. Though the secondary structure determination for proteins is well-established,^{14–18} that for peptides is less successful, e.g., as demonstrated by San Miguel and coworkers.¹³ The authors estimated helical structure by assuming that 100% helical peptides have 208 nm and 222 nm CD signals of $-12 \text{ mol}^{-1} \text{ dm}^3 \text{ cm}^{-1}$ and the average 208 nm and 222 nm signal divided by -12 gave the fractional helix content. In general, a sample that looks partly folded on inspection of the CD spectrum will include a population of unfolded and folded peptides.

This work explores whether the *ad hoc* approach, systematically automated when complemented with self-organizing map spectroscopy, SOMSpec, is useful for the analysis of heterogeneous peptide populations.^{19–21} It further discusses how the extent of derandomisation on the spectra of three U3.x peptides and their seventh position alanine variants in water and after the addition of buffer over time affects the overall goodness of fit of the structure and spectral output from SOMSpec.

MATERIALS AND METHODS

Lisa Martin of Monash University supplied us with the CD spectra for this work. The materials and methods used are stated in Ray et al.¹⁰

SOMSpec Input Data Preparation

We ensured all the input data for SOMSpec analysis was saved in text format. Firstly, the CD spectra were saved in a comma-separated variable, CSV, format with each spectrum placed in a column, with the

*To whom correspondence should be addressed

Email: adewale.olamoyesan@hdr.mq.edu.au; adewale.ola-moyesan@gmail.com

Table 1. Amino acid sequences of U3.x WT peptides and corresponding seventh position alanine variants.

Peptide	Amino acid sequence																	C-ter
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
U3.4 WT	G	V	G	D	L	I	R	K	A	V	A	A	I	K	N	I	V	-NH ₂
U3.4 R7A	G	V	G	D	L	I	A	K	A	V	A	A	I	K	N	I	V	-NH ₂
U3.5 WT	G	V	G	D	L	I	R	K	A	V	S	V	I	K	N	I	V	-NH ₂
U3.5 R7A	G	V	G	D	L	I	A	K	A	V	S	V	I	K	N	I	V	-NH ₂
U3.6 WT	G	V	I	D	A	A	K	K	V	V	N	V	L	K	N	L	F	-NH ₂
U3.6 K7A	G	V	I	D	A	A	A	K	V	V	N	V	L	K	N	L	F	-NH ₂

reference spectra's structures forming five additional entries in each of its columns while that of unknown peptides (experimental spectra) lack structure entries. Secondly, the inspection of the worksheet named "SOMSpec" in the SP175 reference set file was carried out to ensure that all spectra have a wavelength range of 240–190 nm (the last five of the data points are for annotated structures).²⁰ From this original reference set, two additional reference sets were created with the wavelength ranges of 240–195 nm and 240–200 nm, respectively having 46 and 41 data points for each spectrum with the same number of structure entries as that in 240–190 nm annotated to it.

Thirdly, the raw CD spectra were created into test files truncated to a wavelength range of 240–190 nm, 240–195 nm, and 240–200 nm with a step size of 1 nm (so the data points are 51, 46, and 41 per spectrum for the wavelength ranges respectively). The test sets were converted to $\Delta\epsilon$ per molar residue by a scale factor of 3298, then baseline corrected. Lastly, derandomised spectra were automatically created by removing different fractions of the RC spectrum from each $\Delta\epsilon$ peptide spectrum.^{20,21}

SOMSpec Methodology

A square map with 2 500 nodes is created using map size (50 × 50), iterations (50 000), best matching units, BMUs (5). Before training, the weight vectors to populate the nodes within the map are initialised. Next, as the iteration begins, the input vectors are randomly presented to a cluster of the nodes in the map, and the weight vectors associated with each adapt them. At certain iterations associating weight vectors become more identical with the input one, and the BMU input vector is determined using Euclidean distance. Vectors near the BMU with remarkably identical coordinates in the weight map (neighbouring nodes) are brought closer to BMU as training is repeated by a factor L (learning rate, 0.1). The learning rate and neighbouring function used in SOMSpec are based on power series and Gaussian function respectively. Then the BMU, neighbourhood nodes, and other vectors will be assigned to structure labels that will show in the output SOM. Expectedly the five structural labels in the reference spectra must add up to 1, as well as the predicted structure types. Afterwards, the trained map with the secondary structure, SS, property is used to test the unknown spectrum by considering the user-defined number of BMUs (set as 5) in the map. With this value, the nodes very similar to BMUs in the test spectrum are iteratively identified by a distance-dependent model known as NRMSD²². The predicted spectrum is estimated by interpolating across the BMUs and is visualised in the plot accompanying the 2-D representation of SOM. Each trained map was used to predict the spectrum and structure of each peptide (test spectrum) in water only, and for which readings were taken at $t = 0$, $t = 1$, and $t = 24$ h after adding buffer with corresponding wavelength range.

RESULTS AND DISCUSSION

CD Spectra Tested Against Reference Set with a Wavelength Range of 240–190

The essence of taking measurements after adding buffer is to investigate the effect of buffer over time on peptides' structure

evolution or stability. Figure 1 shows the far-UV CD spectra for the aqueous solutions of U3.4 WT, U3.5 WT, U3.6 WT and their seventh position variants. All the peptides in water show a negative maximum at 198 nm, indicating they are mostly RC or unstructured peptides in solution. However, only U3.5 WT gained other spectral features after adding a buffer to the peptide solutions. The other two U3.x WT still maintain the negative maximum of the RC structure at 24 h (Figure 1a, c), while that for U3.5 WT shifted to 204 nm at 1 h. This is accompanied by the appearance of a broad negative maximum at 218 nm, which indicates a conformation transition from RC to α -helix and, or β -sheet. Finally, a positive maximum peak at 196 nm appeared together with a rise in the magnitude of the 218 nm peak, which seems to suggest a further increase in α -helix and, or β -sheet structure (from ~10% α -helix and ~26% β -sheet at $t = 0$ to ~15% α -helix and ~38% β -sheet at $t = 24$ (Figures 1b, d, f and S21)) at 24 h. The transition from a few percentages of folded structures to several percentages of folded structures is more for the seventh position variants than their respective U3.x wt.

The direct structure estimates and spectra fitting for the experimental spectra with SOMSpec using the augmented SP175 reference set were unsatisfactory for all the U3.x peptides in water and after the addition of the buffer (Figure S1–S18). Figure 2 shows the spectral NRMSD plots for U3.4 WT, U3.5 WT, U3.6 WT, and U3.4 R7A recorded in water and at $t = 0$, 1, and 24 h after adding buffer before derandomisation of the spectra. See Table 2 for the predicted α -helix and β -sheet at 24 h after adding the buffer to these U3.x peptides.

The spectra were then systematically derandomised by the "CD app' *disordered unit*" dedicated for the task as described by Olamoyesan et al.^{21,22} Equation 1, as stated in our previous work,²¹ mathematically expresses how varying fractions of RC are removed from the baseline corrected spectra in molar extinction units. The thirty-six spectra resulting from this process for each peptide were fitted with SOMSpec to generate thirty-six spectral predictions with NRMSDs and associate structure estimates. The difficulty for SOMSpec to find perfect places on the trained map are indicated by the NRMSDs, which mostly are above the nominated target value (0.03), except for U3.5 R7A –30%RC at $t = 1$ (0.025). We assessed multiple good fits by weighing the maxima, minima, and where the spectrum crosses zero while overlaying the predicted and the experimental spectra to determine the best of the fits.

Best Fit for U3.4 WT and U3.4 R7A Spectra Either Derandomised or Not

Assessment of the best options for these peptides was carried out using mainly their NRMSD values, except for two instances. Table 3 shows the best spectral NRMSDs and their associate structure estimates for an aqueous solution of U3.4 WT and U3.4 R7A only, after adding buffer, and subtracting (0.1–0.9) RC from the spectra. Almost all of the best choices of these peptides are obvious except for U3.4 R7A in water only with –60% RC and –70% RC, and at $t = 0$ h (immediately after adding buffer) with –20% RC and –40% RC. Though the NRMSDs were almost the same for the two options at each condition, the best choices were found to be –70% RC and –40% RC respectively before the addition of buffer and afterward, at $t = 0$ h (Figure 3b,c,e,

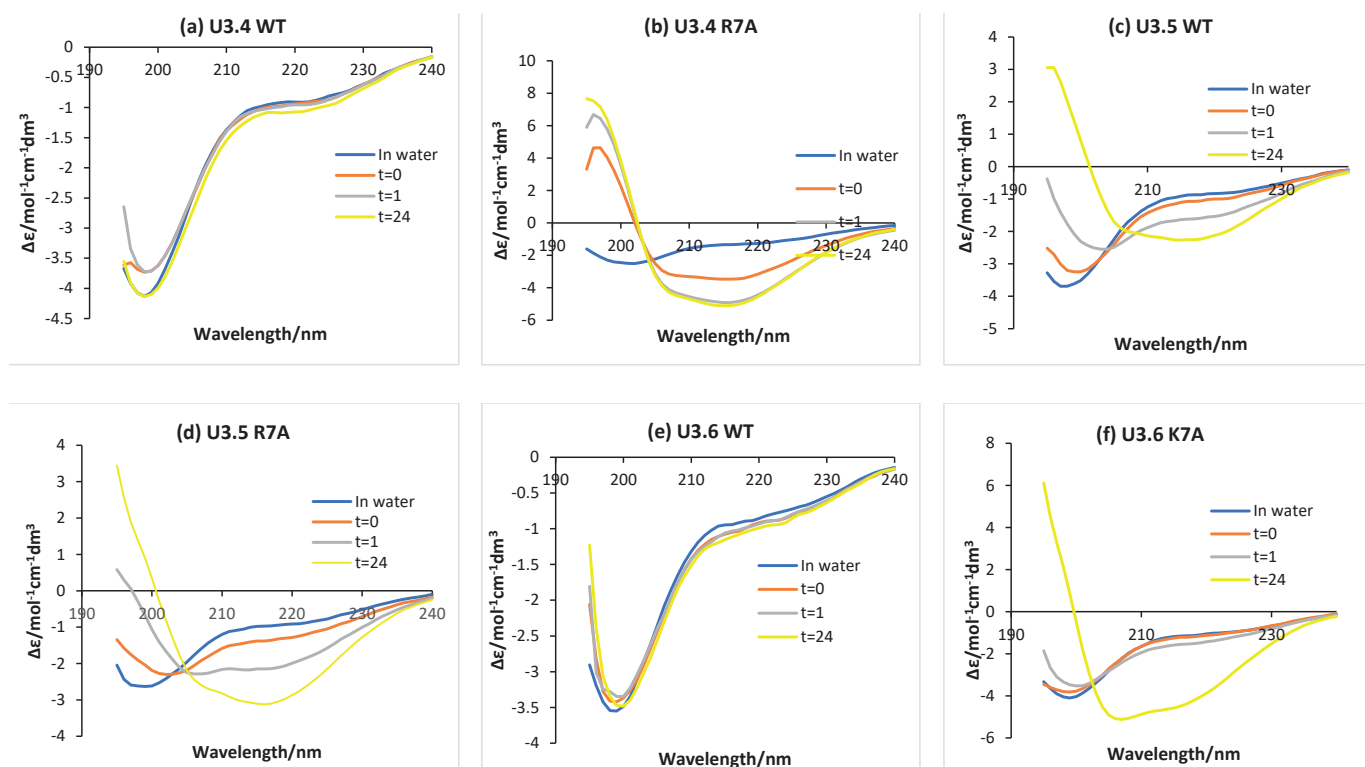


Figure 1. CD spectra of U3.x WT (a–c) and seventh position variants (d–f) in water (blue) and 0.1 M PBS (pH 7) buffer at $t=0$ (red), 1 h (grey), and 24 h (yellow). Spectra were measured in 0.1 mg/mL and converted to $\Delta\epsilon$.

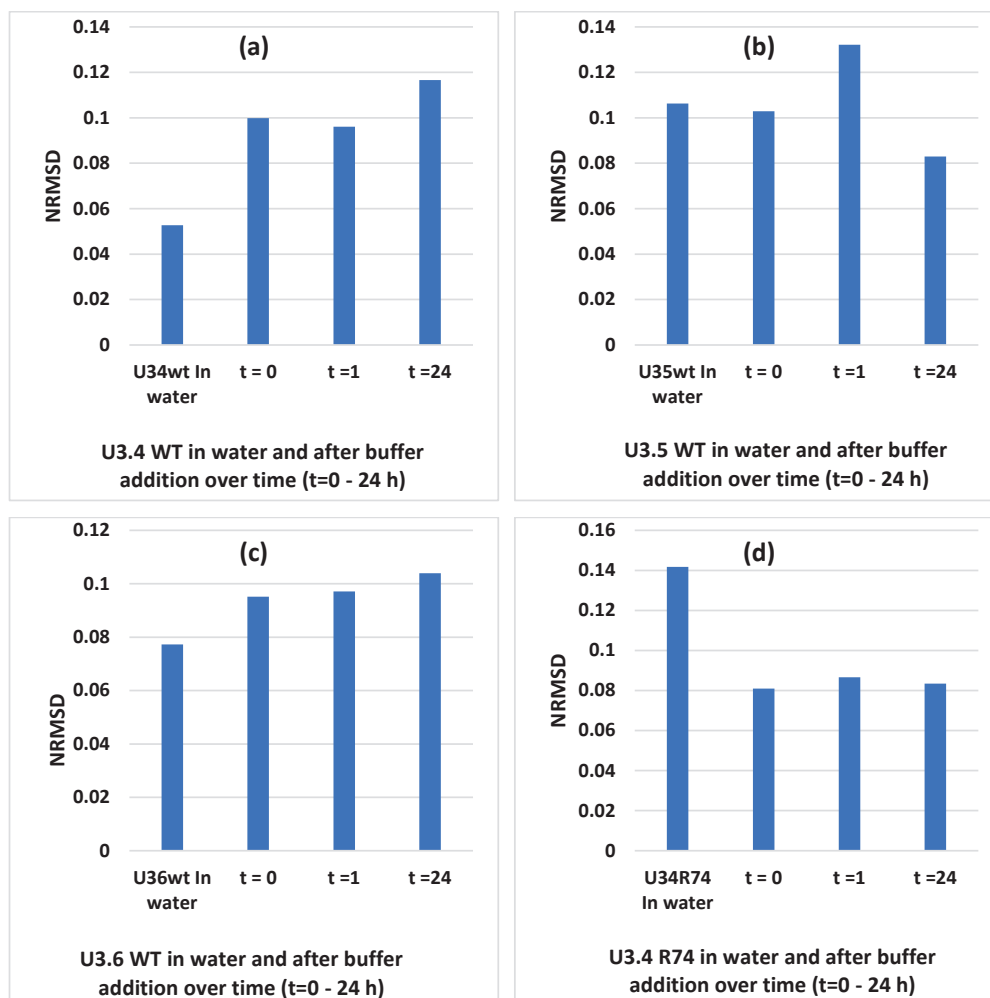
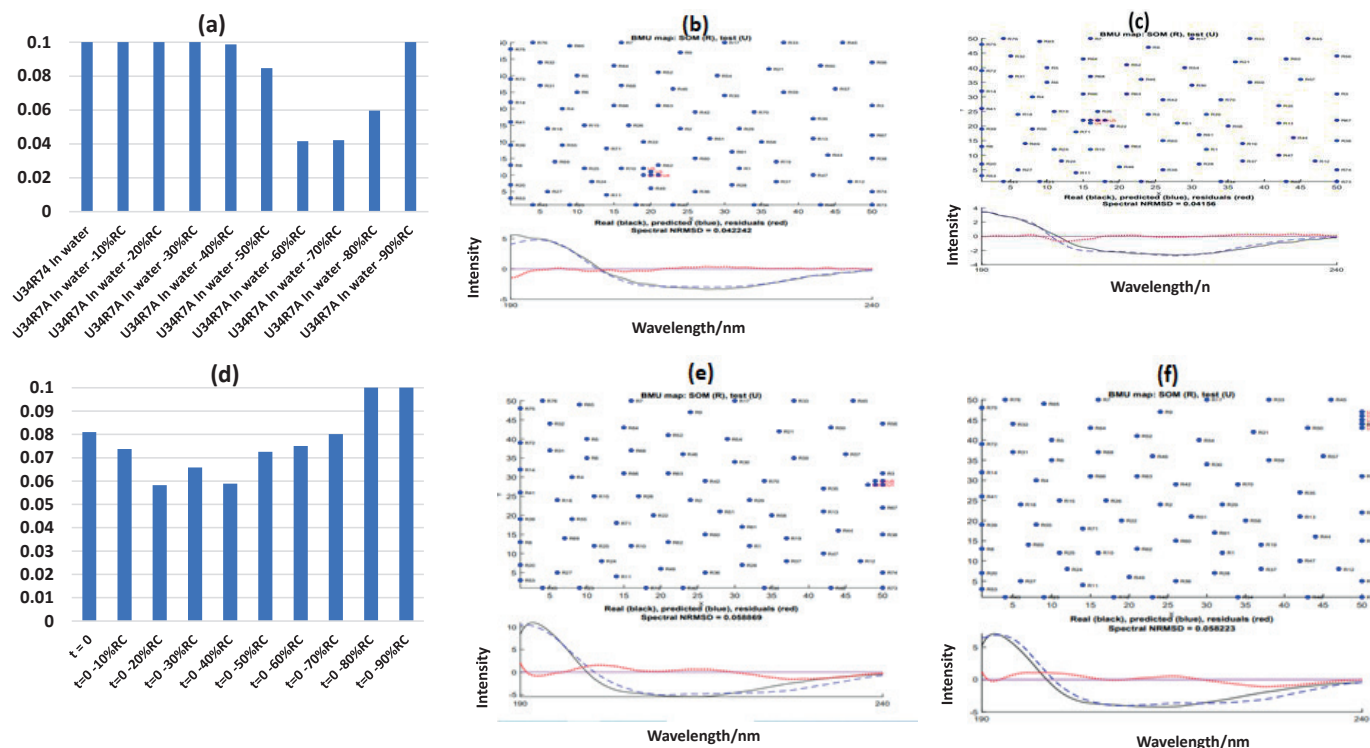


Figure 2. NRMSDs from the fitting of the (a) U3.4 WT, (b) U3.5 WT, (c) U3.6 WT, and (d) U3.4 R7A without derandomisation.

Table 2. Structure estimates of uperin 3.x after addition of buffer at $t = 24$ h.

	α -helix	β -sheet		α -helix	β -sheet
U3.4 WT	10.01%	22.88%	U3.4 R7A	14.55%	37.99%
U3.5 WT	38.25%	17.17%	U3.5 R7A	16.56%	34.72%
U3.6 WT	10.28%	25.39%	U3.6 K7A	25.05%	21.62%

**Figure 3.** SOMSpec spectral NRMDSs for U3.4R7A (a) in water only with different degrees of randomisation and output, (b) with -70% RC, and (c) with -60% RC. The spectral NRMDSs (d) at $t = 0$ after the addition of buffer with varying degrees of randomisation and output (e) with -40% RC and (f) with -20% RC.

and f) based on spectrum shapes match. Alternatively, the best choice can be decided without necessarily inspecting the spectra by implementing a refinement with the percentage RC between the two observed options (lowest and very similar NRMDSs). There is a clear indication that at $t = 1$ and 24 h the same extent of derandomisation (-30% RC) gives best fits with similar structure estimates.

There is a gradual decrease in β -sheet, bend and turn contents as derandomisation increases in the mutant of U3.4 WT in water only and after adding buffer as well as in the WT. The mutation causes a significant increase in α -helix with, say 120.51%, 176.82%, and 282.53% respectively at 0, 1, and 24 h of adding buffer in comparison with its WT. Moreover, the percentage of other folded structures aside helix increased with a mean average of 26.92%, except β -sheet at $t = 1$, and 24 h after adding buffer which decreased respectively by 15.79% and 24.96% from the WT (Figure S19a and S20a). Although all the right answers for the WT were clear, the extent of derandomisation associated with the best fit increased from 0 to 0.9 after dropping by 0.1 and plummeting to 0 at $t = 24$ h.

Best Fit for U3.5 WT and U3.5 R7A Spectra Either Derandomised or Not

In general, there were no hidden challenges when assessing the best fit for U3.5 WT and U3.5 R7A spectral predictions and structure estimates for either derandomised or otherwise. The best fit for U3.5 WT in water only is with -90% RC (which results in 3.6% α -helix and 2.1% β -sheet (regenerated)). Immediately after adding the buffer (at $t = 0$ h), the derandomisation for best fit dropped to -80% RC, which gives rise to 4.3% α -helix and 6.9% β -sheet. This represents a 19.44% and 223.80% increment from the estimates of the initial condition. At 1 h the percentage derandomisation associated with the best fit fell to

-50% RC and remained the same at 24 h, which resulted in a further percentage increase of the most notable folded structures to 8.9% and 18.2% α -helix and 15.8% and 10.1% β -sheet, respectively from the initial conditions. Likewise, the structure evolution of the mutant of U3.5 WT follows a similar trend as the U3.5 WT respective to the change of extent of derandomisation associated with the best fit. We observed that the best choice for U3.5 R7A in water only has -80% RC removed from it with estimates of 4.7% α -helix, and 5.8% β -sheet. At $t = 0$ h after adding buffer the extent of derandomisation associated with the best fit is the same as in water only, and this resulted in a 20% increase of the α -helix while the β -sheet content dropped by 40%. However, at 1 h of adding the buffer, the best fit is with -40% RC which yielded an increment of 56.16% and 435% for each of the notable structures (11.4% α -helix and 21.4% β -sheet). Furthermore, the extent of deandomisation associated with best fit dropped to -10% RC at 24 h of placing the sample in water and buffer mixture; this gave rise to 16.6% α -helix and 31.2% β -sheet. Helical content of measurements taken in water only, at $t = 0$, and 1 h compared well, but moderately increased by 14.55 - 16.56% for both the WT and its mutant at $t = 24$ h after buffer addition (without no RC removed). Structure estimates of α -helix, β -sheet, bend and turn increase at 24 h after the addition of buffer. At the same time, unstructured content varies inversely to the β -sheet before derandomisation of the spectra for both U3.5 WT and U3.5 R7A. However, derandomisation of the spectra with varying fractions of RC causes a decrease in β -sheet, bend, and turn contents as the unstructured content increases. Overall, the drop or rise in the extent of derandomisation associated with the best answers is expected to cause a decrease or an increase in the unfolded structure contents (Table 4 and Figure S21–22).

Table 3. Best spectral NRMSDs and associate structure estimates for U3.4 WT and U3.4 R7A spectra were collected in the water only and after the addition of buffer at $t = 0$, $t = 1$, $t = 24$ h while derandomised by removing (0.1–0.9%) RC. Bold indicates preferred values where multiple options were good fits, as discussed in the text.

U3.4 WT	Best NRMSD	RC added	Derandomised			Regenerated original protein		
			α -helix	β -sheet	Other	α -helix	β -sheet	Other
In water	0.052	0%				0.101	0.253	0.646
$t = 0$	0.090	90%	0.331	0.180	0.489	0.033	0.033	0.934
$t = 1$	0.085	80%	0.197	0.296	0.507	0.039	0.059	0.902
$t = 24$	0.117	0%				0.100	0.229	0.671
U3.4 R7A								
In water	0.042 (0.042)	70% (60%)	0.265 (0.157)	0.265 (0.385)	0.470 (0.458)	0.079 (0.063)	0.080 (0.154)	0.841 (0.783)
$t = 0$	0.059 (0.058)	40% (20%)	0.371 (0.414)	0.190 (0.139)	0.439 (0.447)	0.223 (0.331)	0.114 (0.111)	0.663 (0.558)
$t = 1$	0.064	30%	0.382	0.179	0.439	0.268	0.125	0.607
$t = 24$	0.071	30%	0.392	0.173	0.432	0.274	0.121	0.605

Table 4. Best spectral NRMSDs and associate structure estimates for U3.5 WT and U3.5 R7A spectra collected in the water only and after the addition of buffer at $t = 0$, $t = 1$, $t = 24$ h while derandomised by removing (0–0.9%) RC. Bold indicates preferred values where multiple options were good fits, as discussed in the text.

U3.5 WT	Best NRMSD	RC added	Derandomised			Regenerated original protein		
			α -helix	β -sheet	Other	α -helix	β -sheet	Other
In water	0.065	90%	0.362	0.209	0.429	0.036	0.021	0.943
$t = 0$	0.048	80%	0.215	0.346	0.439	0.043	0.069	0.888
$t = 1$	0.054	50%	0.178	0.317	0.505	0.089	0.158	0.753
$t = 24$	0.067	50%	0.364	0.202	0.434	0.182	0.101	0.717
U3.5 R7A								
In water	0.063	80%	0.234	0.290	0.476	0.047	0.058	0.895
$t = 0$	0.042	80%	0.365	0.202	0.433	0.073	0.040	0.890
$t = 1$	0.040	40%	0.190	0.356	0.454	0.114	0.214	0.672
$t = 24$	0.045	10%	0.184	0.354	0.462	0.166	0.312	0.522

Best Fit for U3.6 WT and U3.6 K7A Spectra Either Derandomised or Not

The results indicative of the best fits for both U3.6 WT and its mutant are summarised in Table 5. As can be seen, the extent of derandomisation that can be credited for the best fit increased to -90% RC from -80% RC for the WT after the addition of buffer at 1 h and subsequently plummeted to -0%RC at 24 h. This drastic drop resulted in 10.3% α -helix and 25.4% β -sheet (Figure S23). For its mutant, almost all cases with no derandomisation resulted primarily as best fits, except at 1 h after adding the buffer to the peptide solution rose to -90% RC. The structure estimates for helical content are identical in three instances (in water only, at $t = 0$ and $t = 1$), while at 24 h of adding buffer, the α -helix increased to 25.05% (Figure S24). It appears that as we sequentially increase the fractions of RC subtracted from these peptides (e.g., from -50%–60% RC), this directly leads to an 8% average increase of unstructured content at each step. At later times than 0 h, the mutation causes an increase of 3–151% in α -helix, and for all the conditions bend increases gradually from 0.24–23.32% at $t = 24$ h (Figure S23 and S24) and yet β -sheet and turn increase in three of the measured conditions.

CD Spectra Tested Against Reference Set with the Same Wavelength Range of 240–195 nm

There is a feeling that the reduction of wavelength range down to 195 or 200 nm affects the quality of the structure information obtainable from the spectra. Thus, the aim in this section is to explore whether this was the case by creating a reference set and test file with spectra down to 195 nm, and then putting the spectra through SOMSpec. This research question will be answered by comparing the SOMSpec NRMSDs for spectra down to 190, 195, and 200 nm, with or without derandomisation.

Best Fit for U3.4 WT and U3.4 R7A Spectra Either Derandomised or Not from 240–195 nm

Table 6 displays the best NRMSDs and their structure estimates and percentage derandomisation added to U3.4 WT and its mutant spectra truncated to 195 nm that effect best fits. The best fit for U3.4 WT in water only has a lower extent of derandomisation (-30%RC) and its NRMSD value (less 0.05), suggests the structure estimates and fitted spectral are of reasonable fit. This resulted in 8.2% α -helix and 18% β -sheet. Immediately after adding the buffer to the peptide solution, the extent of derandomisation associated with the best fit dropped to -10%RC, to give 10.2% α -helix and 21.5% β -sheet. At $t = 1$ h, the extent of derandomisation associated with the best fit rose to 80%RC, expectedly with a drop of about 68% for both α -helix and α -sheet compared to previous condition's estimates. Two similar lowest NRMSDs were observed at $t = 24$ h; and the best choice was determined to be with -10%RC following the approach used in the previous occurrence. Therefore, the estimates for both α -helix and β -sheet appreciate by 203% and 232% respectively. We recognised for the mutant that its extent of derandomisation that caused the best fit, dropped immediately after adding the buffer and later rose moderately at 1 h and at the end of the reading rose back to the initial value (-60%RC). It appears that spectra whose best fits have no or lower derandomisation contain more folded structures than those with a higher degree of derandomisation. For example, at $t = 1$ h after adding buffer the estimates for α -helix rose by 171.26%.

Almost all the U3.4 WT spectra reduced down to 195 nm gave a better fit when compared to those down to 190 nm, except for three instances when dissolved in water only (with -0%RC and -80%RC) and immediately after the addition of buffer (with -90%RC). Expectedly, data down to 200 nm are worst, when compared to that down to 190 and 195 nm. In twenty-nine of the forty instances, data down to

Table 5. Best spectral NRMSDs and associate structure estimates for U3.6 WT and U3.6 K7A spectra collected in the water only and after the addition of buffer at $t = 0$, $t = 1$, $t = 24$ h while derandomised by removing (0.1–0.9%) RC. Bold indicates preferred values where multiple options were good fits as discussed in the text.

U3.6 WT	Best NRMSD	RC added	Derandomised			Regenerated original protein		
			α -helix	β -sheet	Other	α -helix	β -sheet	Other
In water	0.045	80%	0.163	0.380	0.457	0.033	0.076	0.891
$t = 0$	0.073	80%	0.195	0.339	0.466	0.039	0.068	0.893
$t = 1$	0.076	90%	0.375	0.187	0.438	0.037	0.019	0.944
$t = 24$	0.104	0%				0.103	0.254	0.643
U3.6 K7A								
In water	0.056	0%				0.096	0.246	0.658
$t = 0$	0.102	0%				0.100	0.231	0.669
$t = 1$	0.091	90%	0.429	0.138	0.433	0.043	0.014	0.943
$t = 24$	0.048	0%				0.251	0.216	0.533

Table 6. Spectral NRMSDs and associate structure estimates for U3.4 WT and U3.4 R7A spectra collected in the water only and after addition of buffer at $t = 0$, $t = 1$, $t = 24$ h while derandomised by removing (0–0.9) RC and then truncated to 195 nm. Bold indicates preferred values where multiple options were reasonable fits as discussed in the text.

U3.4 WT	Best NRMSD	RC added	Derandomised			Regenerated original protein		
			α -helix	β -sheet	Other	α -helix	β -sheet	Other
In water	0.036	30%	0.117	0.252	0.631	0.082	0.180	0.812
$t = 0$	0.050	10%	0.112	0.225	0.663	0.102	0.215	0.683
$t = 1$	0.052	80%	0.174	0.343	0.483	0.035	0.069	0.866
$t = 24$	0.042 (0.042)	10% (20%)	0.118 (0.118)	0.255 (0.255)	0.470 (0.458)	0.106 (0.095)	0.230 (0.204)	0.664 (0.783)
U3.4 R7A								
In water	0.031	60%	0.123	0.411		0.049	0.164	0.787
$t = 0$	0.074	0%				0.167	0.380	0.453
$t = 1$	0.110	20%	0.567	0.101		0.453	0.081	0.466
$t = 24$	0.031	60%	0.576	0.094		0.231	0.038	0.731

195 nm gave the best fit when compared to the other wavelength ranges (190 nm and 200 nm) (Table S1). We hypothesised that these twenty-nine spectra are more similar to their BMUs in the training map

than that of wavelength ranges down to 190 nm and 200 nm. As the percentage RC removed increases a gradual decrease in the four structure types (namely helix, sheet, bend, and turn) was observed, while unstructured content increases (Figure S25 and S26).

Best Fit for U3.5 WT and U3.5 R7A Spectra Either Derandomised or Not from 240–195 nm

The best fits for U3.5 WT in water only and at 24 h of adding buffer have the same extent of derandomisation, however, the structure estimates for α -helix and β -sheet differ by 10.6% and 56%, respectively. There is a significant drop in the extent of derandomisation required for best fit immediately after the addition of a buffer but a slight fall to -50%RC was observed at 1 h. Though the spectral trends for U3.5 R7A were similar to that of U3.5 WT, its extent of derandomisation for best fit immediately after the addition of a buffer dropped and further dropped at 1 h and 24 h, respectively to -30%RC and 0%RC. We observed its β -sheet content increased over time after adding a buffer to 34.5% at $t = 24$ h (Table 7). Twenty-six of the forty spectra of U3.5 WT down to 195 nm showed improvement in their quality and estimate of secondary structure when compared to original data, whereas overall data down to 200 nm were worst. The behaviour of U3.5 R7A spectra down to 195 nm was similar to the case of its U3.5 WT when compared to original data and that down to 200 nm (Table S2).

Best Fit for U3.6 WT and U3.6 K7A Spectra Either Derandomised or Not from 240–195 nm

Other measurements taking aside that of U3.6 WT in water only have the same percentage of derandomisation for their best fits and

nearly exact structure estimates for all types (7.5% α -helix and 14.3% β -sheet). There is a significant drop in the extent of derandomisation associated with the mutant best fit immediately after the addition of buffer to 0%RC, as a result, the estimates appreciated by 234.78% and 64.04% respectively for α -helix and β -sheet. At 1 h, this dropped to -60%RC, and subsequently at 24 h further dropped to -30%RC resulting in 38.2% α -helix and 8.1% β -sheet. (Table 8). The exact numbers of structure elements that decrease as the unstructured content increases for both U3.6 wt and its mutant are found to be the same. (Figure S29 and S30). Of the three data sets those down to 195 nm were of the best quality while those down to 200 nm were of the poorest quality. However, in nine instances the predictions for spectrum down to 195 nm were of the worst quality of which four have the highest extent of derandomisation associated with them (Table S3).

CD Spectra Tested Against Reference Set with the Same Wavelength Range of 240–200 nm

To enable us to answer the question of whether the wavelength range 240–200 nm affects the quality of the spectral and structure predictions. We created spectra with a wavelength down to 200 nm by either removing 10 nm from the original data or 5 nm from the previous reference and test files (with a wavelength down to 195 nm). The reference set was trained with SOMSpec using the same input parameter as specified in section 2.2.5, the test file was tested against the trained map generated by SOMSpec to give the structure estimates and spectral fits.

Best Fit for U3.4 WT and U3.4 R7A Spectra Either Derandomised or Not from 240–200 nm

The best fits of the two initial conditions at which U3.4 WT spectra were collected have the same extent of derandomisation and their

Table 7. Best spectral NRMSDs and associate structure estimates for U3.5 WT and U3.5 R7A spectra collected in the water only and after addition of buffer at $t = 0$, $t = 1$, $t = 24$ h while derandomised by removing (0-0.9) RC and then truncated to 195 nm. Bold indicates preferred values where multiple options were reasonable fits as discussed in the text.

U3.5 WT	Best NRMSD	RC added	Derandomised			Regenerated original protein		
			α -helix	β -sheet	Other	α -helix	β -sheet	Other
In water	0.053	0%				0.113	0.264	0.623
$t = 0$	0.168	60%	0.176	0.364	0.460	0.070	0.146	0.784
$t = 1$	0.029	50%	0.213	0.323	0.464	0.107	0.162	0.731
$t = 24$	0.032	0%				0.125	0.412	0.463
U3.5 R7A								
In water	0.034	70%	0.080	0.465	0.455	0.231	0.038	0.731
$t = 0$	0.043	60%	0.123	0.414	0.463	0.049	0.165	0.786
$t = 1$	0.025	30%	0.161	0.372	0.467	0.112	0.260	0.628
$t = 24$	0.033	0%				0.191	0.345	0.464

Table 8. Best spectral NRMSDs and associate structure estimates for U3.6 WT and U3.6 K7A spectra collected in the water only and after addition of buffer at $t = 0$, $t = 1$, $t = 24$ h reduced to 195 nm and then derandomised by removing (0.1-0.9) RC. Bold indicates preferred values where multiple options were good fits as discussed in the text.

U3.6 WT	Best NRMSD	RC added	Derandomised			Regenerated original protein		
			α -helix	β -sheet	Other	α -helix	β -sheet	Other
In water	0.041	80%	0.113	0.421	0.466	0.023	0.084	0.893
$t = 0$	0.050	60%	0.192	0.354	0.454	0.077	0.142	0.781
$t = 1$	0.067	60%	0.185	0.359	0.456	0.074	0.143	0.783
$t = 24$	0.084	60%	0.189	0.358	0.453	0.075	0.143	0.782
U3.6 K7A								
In water	0.037	90%	0.587	0.083	0.330	0.059	0.008	0.933
$t = 0$	0.039	0%				0.118	0.257	0.625
$t = 1$	0.041	60%	0.251	0.297	0.452	0.101	0.119	0.780
$t = 24$	0.083	30%	0.545	0.116	0.339	0.382	0.081	0.537

structure estimates for all structure types. Subsequently, the extent of derandomisation of the best fit increased to -80%RC at 1 h but later at 24 h dropped to -30%RC after the addition of a buffer. This results in an increase of folded structures (13.5% α -helix and 18.1% β -sheet) within the formulated peptide at 24 h. U3.4 R7A best fits are expected to behave differently regarding the extent of derandomisation associated with them to that of WT under the same conditions. At first, this significantly dropped to a lower fraction (-20%RC) and then remained the same 1 h later and subsequently slightly dropped to -10%RC at 24 h. We observed irrespective of the percentage derandomisation associated with the best fit(s), α -helix content increases over time after the addition of the buffer (e.g., at $t = 24$ h resulted in 22% α -helix). For cases with -80%RC either WT or otherwise, their structure estimates are nearly the same (Table 9).

Best Fit for U3.5 wt and U3.5 R7A Spectra Either Derandomised or Not from 240–200 nm

Two of the four conditions for U3.5 wt considered have a higher extent of derandomisation associated with the best fits. As expected, the two notable structures' predictions for these are within a few percentages. There is a hidden challenge in determining the best fit at $t = 0$ h (immediately after the addition of the buffer to the peptide solution). Thus, two similar lowest NRMSDs were observed, and the best choice was determined to be with -10%RC by visually inspecting the spectra. For the mutant of U3.5 WT (U3.5 R7A), the extent of derandomisation associated with the best fit is the same in water only and immediately after the addition of buffer (-80%RC). Like with U3.5 WT, the notable structures' predictions for these are within a few percentages. Subsequently, as this dropped to -60%RC and further dropped to -40%RC over time the two notable

structures' estimates increased to 12.5% and 16.2% respectively, as this corresponds to 135.85% and 82.02% increase from estimates at $t = 1$ h (Table 10). We found that β -sheet, bend, and turn decrease as the percentage of RC removed from the U3.6K7A spectra increases, while α -helix with -20%RC to 50%RC deviates from this behaviour (Figure S33 and S34).

Best Fit for U3.6 WT and U3.6 R7A Spectra Either Derandomised or Not from 240–200 nm

All the best fits for U3.6 WT explored are associated with the same extent of derandomisation (which is of a higher fraction) and the same structure estimates for each element. For 3.6 K7A, two of the four conditions explored behaviour is similar to that of the WT. They have the same higher extent of derandomisation associated with the best fits and the same structure estimates for each element. Though in water only two similar lowest NRMSDs were observed at first, and then the best choice was determined to be with -0%RC. Later at $t = 1$ h the associated extent of derandomisation to the best fit markedly increased to -80%RC and then at $t = 24$ h significantly dropped to -30%RC. The latter case gives rise to 13.5% α -helix and 18.1% β -sheet which are respectively higher than the corresponding estimates for the former at $t = 1$ h by 221.43% and 260.78% (Table 11). Overall, as the percentage of derandomisation increases there is a gradual decrease in the four structure elements, viz., helix, sheet, bend and turn (Figure S35 and S36).

CONCLUSIONS

In this study, we have been able to show that the derandomisation approach previously reported by our group can also be used for peptides, and evaluate how this affects the goodness of fits of the

Table 9. Best spectral NRMDSs and associate structure estimates for U3.4 WT and U3.4 R7A spectra collected in the water only and after addition of buffer at t = 0, t = 1, t = 24 h reduced to 200 nm and then derandomised by removing (0-0.9) RC. Bold indicates preferred values where multiple options were good fits as discussed in the text.

U3.4 WT	Best NRMDS	RC added	Derandomised			Regenerated original protein		
			α-helix	β-sheet	Other	α-helix	β-sheet	Other
In water	0.041	0%				0.092	0.254	0.654
t = 0	0.050	0%				0.092	0.255	0.653
t = 1	0.067	80%	0.212	0.253	0.535	0.042	0.051	0.907
t = 24	0.084	30%	0.192	0.259	0.549	0.135	0.181	0.684
U3.4 R7A								
In water	0.219	80%	0.499	0.102	0.399	0.049	0.047	0.904
t = 0	0.007	20%	0.176	0.295	0.529	0.141	0.236	0.623
t = 1	0.078	20%	0.246	0.232	0.522	0.197	0.186	0.617
t = 24	0.076	10%	0.245	0.234	0.521	0.220	0.210	0.570

Table 10. Best spectral NRMDSs and associate structure estimates for U3.5 WT and U3.5 R7A spectra collected in the water only and after addition of buffer at t = 0, t = 1, t = 24 h reduced to 200 nm and then derandomised by removing (0-0.9) RC. Bold indicates preferred values where multiple options were good fits as discussed in the text.

U3.5 WT	Best NRMDS	RC added	Derandomised			Regenerated original protein		
			α-helix	β-sheet	Other	α-helix	β-sheet	Other
In water	0.088	90%	0.255	0.245	0.500	0.026	0.021	0.953
t = 0	0.123	10%	0.092	0.263	0.643	0.084	0.237	0.679
	0.118	20%	0.091	0.281	0.628	0.076	0.250	0.674
t = 1	0.082	70%	0.176	0.296	0.528	0.053	0.089	0.858
t = 24	0.071	40%	0.209	0.269	0.522	0.125	0.162	0.713
U3.5 R7A								
In water	0.042	80%	0.160	0.387	0.453	0.032	0.077	0.891
t = 0	0.063	80%	0.267	0.213	0.520	0.053	0.043	0.904
t = 1	0.063	60%	0.232	0.250	0.518	0.093	0.100	0.807
t = 24	0.072	40%	0.228	0.251	0.521	0.137	0.151	0.712

Table 11. Best spectral NRMDSs and associate structure estimates for U3.6 WT and U3.6 K7A spectra collected in the water only and after addition of buffer at t = 0, t = 1, t = 24 h reduced to 200 nm and then derandomised by removing (0-0.9) RC. Bold indicates preferred values where multiple options were good fits as discussed in the text.

U3.6 WT	Best NRMDS	RC added	Derandomised			Regenerated original protein		
			α-helix	β-sheet	Other	α-helix	β-sheet	Other
In water	0.074	90%	0.259	0.225	0.516	0.026	0.023	0.951
t = 0	0.101	90%	0.258	0.220	0.522	0.028	0.022	0.950
t = 1	0.094	90%	0.259	0.219	0.522	0.026	0.022	0.952
t = 24	0.125	90%	0.252	0.224	0.524	0.025	0.022	0.953
U3.6 K7A								
In water	0.149	0%				0.090	0.250	0.660
	0.182	10%	0.090	0.250	0.660	0.081	0.224	0.695
t = 0	0.177	0%				0.090	0.250	0.660
t = 1	0.145	80%	0.212	0.253	0.535	0.042	0.051	0.907
t = 24	0.812	30%	0.192	0.259	0.549	0.135	0.181	0.684

spectral and structure outputs as a function of interaction time with the buffer. The U3.x peptides derandomised are discovered on the skin of toadlet endemic to Australia and in their pristine state they exhibit antimicrobial properties. Almost all the U3.x peptides were straightforward to analyse, except in five instances where visual inspection of the spectra was required to determine the right answers. We found that the extent of derandomisation associated with the best fit could drop or jump or remain the same all through the conditions assessed. In general, it is recognised that an increase or a decrease in the two notable folded structures could have been caused by a drop or a rise of the percentage RC responsible for the right answer, as

well as a gradual decrease in four to three structure elements as the percentage RC removed from peptide spectra increases. Our findings on whether the wavelength range affects the quality of spectral and structure predictions extend those of Spencer et al.²³, confirming that spectra down to 195 nm are of the best quality while that down to 200 nm are of the poorest quality. We could infer that the structural information is less in the spectra whose wavelength is down to 200 nm, and the quality of the buffer data is only satisfactory down to 195 nm due to the chloride absorbance. There is a need for experiments to prove the underlying approach of derandomisation and its effect on the structure of the proteins or peptides in some processes is real.

ACKNOWLEDGEMENTS

The support from the international Macquarie University Research Excellence Scholarship (iMQRES) is gracefully acknowledged.

ORCID IDS

Adewale Olamoyesan: <https://orcid.org/0000-0002-9879-6726>

Alison Rodger: <https://orcid.org/0000-0001-8817-4651>

SUPPLEMENTARY MATERIAL

Supplementary tables and figures are available with this article.

REFERENCES

1. Nylén F, Miraglia E, Cederlund A, Ottosson H, Stromberg R, Gudmundsson GH, Agerberth B. Boosting innate immunity: development and validation of a cell-based screening assay to identify LL-37 inducers. *Innate Immun.* 2014;20(4):364–376. <https://doi.org/10.1177/1753425913493338>.
2. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol.* 2005;3(3):238–250. <https://doi.org/10.1038/nrmicro1098>.
3. Apponyi MA, Pukala TL, Brinkworth CS, Maselli VM, Bowie JH, Tyler MJ, Booker GW, Wallace JC, Carver JA, Separovic F, et al. Host-defence peptides of Australian anurans: structure, mechanism of action and evolutionary significance. *Peptides.* 2004;25(6):1035–1054. <https://doi.org/10.1016/j.peptides.2004.03.006>.
4. Bowie JH, Separovic F, Tyler MJ. Host-defense peptides of Australian anurans. Part 2. Structure, activity, mechanism of action, and evolutionary significance. *Peptides.* 2012;37(1):174–188. <https://doi.org/10.1016/j.peptides.2012.06.017>.
5. Pukala TL, Bowie JH, Maselli VM, Musgrave IF, Tyler MJ. Host-defence peptides from the glandular secretions of amphibians: structure and activity. *Nat Prod Rep.* 2006;23(3):368–393. <https://doi.org/10.1039/b512118n>.
6. Bowie JH, Wegner KL, Wabnitz PA, Chia BCS, Carver JA, Wallace JC, Tyler MJ. Host defense antibacterial peptides from the skin secretions of Australian amphibians. *Protein Pept Lett.* 1999;6:259–269. <https://doi.org/10.2174/092986650605221117113359>.
7. Chia BCS, Bowie JH, Carver JA, Mulhern TD. The solution structure of uperin 3.6, an antibiotic peptide from the granular dorsal glands of the Australian toadlet, *Uperoleia mjobergii*. *J Pept Res.* 1999;54(2):137–145. <https://doi.org/10.1034/j.1399-3011.1999.00095.x>.
8. Bradford AM, Bowie JH, Tyler MJ, Wallace JC. New antibiotic uperin peptides from the dorsal glands of the Australian toadlet *Uperoleia mjobergii*. *Aust J Chem.* 1996;49(12):1325–1331. <https://doi.org/10.1071/CH9961325>.
9. Calabrese AN, Liu Y, Wang T, Musgrave IF, Pukala TL, Tabor RF, Martin LL, Carver JA, Bowie JH. The amyloid fibril-forming properties of the amphibian antimicrobial peptide uperin 3.5. *ChemBioChem.* 2016;17(3):239–246. <https://doi.org/10.1002/cbic.201500518>.
10. Ray S, Holden S, Prasad AK, Martin LL, Panwar AS. Exploring the role of peptide helical stability in the propensity of uperin 3. x peptides toward beta-aggregation. *J Phys Chem B.* 2020;124(51):11659–11670. <https://doi.org/10.1021/acs.jpcc.0c10000>.
11. Juban MM, Javadpour MM, Barkley MD. Circular dichroism studies of secondary structure of peptides. In: Shafer WM, editor. *Antibacterial Peptide Protocols*. New Jersey: Humana Totowa; 1997. p. 73–78. <https://doi.org/10.1385/0-89603-408-9:73>.
12. Chakrabartty A, Kortemme T, Padmanabhan S, Baldwin RLJB. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry.* 1993;32(21):5560–5565. <https://doi.org/10.1021/bi00072a010>.
13. San Miguel M, Marrington R, Rodger PM, Rodger A, Robinson C. An *Escherichia coli* twin-arginine signal peptide switches between helical and unstructured conformations depending on the hydrophobicity of the environment. *Eur J Biochem.* 2003;270(16):3345–3352. <https://doi.org/10.1046/j.1432-1033.2003.03710.x>.
14. Johnson WC. Analysing protein circular dichroism spectra for accurate secondary structures. *Proteins: Struct Funct Bioinf.* 1999;35:307–312.
15. Andrade M, Chacon P, Merelo JJ, Moran F. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng Des Sel.* 1993;6(4):383–390. <https://doi.org/10.1093/protein/6.4.383>.
16. Bulheller BM, Hirst JD. DichroCalc—circular and linear dichroism online. *Bioinformatics.* 2009;25(4):539–540. <https://doi.org/10.1093/bioinformatics/btp016>.
17. Hall V, Nash A, Rodger A. SSNN, a method for neural network protein secondary structure fitting using circular dichroism data. *Anal Methods.* 2014;6(17):6721–6726. <https://doi.org/10.1039/C3AY41831F>.
18. Hall V, Sklepari M, Rodger A. Protein secondary structure prediction from circular dichroism spectra using a self-organising map with concentration correction. *Chirality.* 2014;26(9):471–482. <https://doi.org/10.1002/chir.22338>.
19. Olamoyesan A. Protein Structure Fitting by a Self-Organising Map Spectroscopy (SOMSpec): Circular Dichroism and Infrared Spectroscopy; 2022. PhD thesis. Macquarie University, Sydney, Australia. 2022
20. Olamoyesan A, Ang D, Rodger A. Circular dichroism for secondary structure determination of proteins with unfolded domains using a self-organising map algorithm SOMSpec. *RSC Adv.* 2021;11(39):23985–23991. <https://doi.org/10.1039/D1RA02898G>.
21. Olamoyesan A. Automation of the ad hoc approach for derandomization of proteins: a tutorial for undergraduates in molecular sciences. *J Chem Educ.* Forthcoming.
22. Natita W, Wiboonsak W, Dusadee S. Appropriate learning rate and neighborhood function of self-organising map (SOM) for specific humidity pattern classification over Southern Thailand. *Int. J. Model. Optim.* 2016;6:61–65.
23. Spencer SE, Rodger A. Bayesian inference assessment of protein secondary structure analysis using circular dichroism data—how much structural information is contained in protein circular dichroism spectra? *Anal Methods.* 2021;13(3):359–368. <https://doi.org/10.1039/D0AY01645D>.