

**Bioactivity and structural properties of chimeric analogs of
the starfish SALMFamide neuropeptides S1 and S2.**

Christopher E. Jones^{a, b*}, Claire B. Otara^{a*}, Nadine D. Younan^a,

John H. Viles^a and Maurice R. Elphick^a

^a. School of Biological & Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK. ^b. School of Science and Health, University of Western Sydney, Locked bag 1797, Penrith, 2751, Sydney, Australia.

* CEJ and CBO contributed equally

Running title: *Analysis of chimeric analogs of the SALMFamides S1 and S2*

To whom correspondence should be addressed:

Maurice R. Elphick, School of Biological & Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK. Tel: 0207 882 6664; Fax: 0207 882 7732; E-mail:

m.r.elphick@qmul.ac.uk

or

John H. Viles, School of Biological & Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK. Tel: 0207 882 8443; Fax: 0207 882 7732; E-mail:

j.viles@qmul.ac.uk

Abstract

The starfish SALMFamide neuropeptides S1 (GFNSALMFamide) and S2 (SGPYSFNSGLTFamide) are the prototypical members of a family of neuropeptides that act as muscle relaxants in echinoderms. Comparison of the bioactivity of S1 and S2 as muscle relaxants has revealed that S2 is ten times more potent than S1. Here we investigated a structural basis for this difference in potency by comparing the bioactivity and solution conformations (using NMR and CD spectroscopy) of S1 and S2 with three chimeric analogs of these peptides. A peptide comprising S1 with the addition of S2's N-terminal tetrapeptide (Long S1 or LS1; SGPYGFNSALMFamide) was not significantly different to S1 in its bioactivity and did not exhibit concentration-dependent structuring seen with S2. An analog of S1 with its penultimate residue substituted from S2 (S1(T); GFNSALTFamide) exhibited S1-like bioactivity and structure. However, an analog of S2 with its penultimate residue substituted from S1 (S2(M); SGPYSFNSGLMFamide) exhibited loss of S2-type bioactivity and structural properties. Collectively, our data indicate that the C-terminal regions of S1 and S2 are the key determinants of their differing bioactivity. However, the N-terminal region of S2 may influence its bioactivity by conferring structural stability in solution. Thus, analysis of chimeric SALMFamides has revealed how neuropeptide bioactivity is determined by a complex interplay of sequence and conformation.

1. Introduction

The SALMFamides are a family of neuropeptides that occur in species belonging to the phylum Echinodermata (e.g. starfish, sea cucumbers, sea urchins). There are two types of SALMFamides – L-type, which have the C-terminal motif SxLxFamide (where x is variable), and F-type, which have the C-terminal motif SxFxFamide. Furthermore, analysis of the pharmacological actions of SALMFamides has revealed that both L-type and F-type SALMFamides cause muscle relaxation in echinoderms [1, 2].

The first members of the SALMFamide neuropeptide family to be identified were the L-type SALMFamides S1 and S2, which were both isolated from the nervous system of the starfish species *Asterias rubens* and *Asterias forbesi* [3]. S1 is an octapeptide with the amino acid sequence H-Gly-Phe-Asn-Ser-Ala-Leu-Met-Phe-NH₂ and S2 is a dodecapeptide with the amino acid sequence H-Ser-Gly-Pro-Tyr-Ser-Phe-Asn-Ser-Gly-Leu-Thr-Phe-NH₂ (Fig. 1). Interestingly, injection of S1 or S2 into *A. rubens* triggers cardiac stomach eversion, a process that occurs naturally when starfish feed extra-orally on prey such as mussels [4]. Consistent with the effects of S1 and S2 *in vivo*, both peptides cause dose-dependent relaxation of cardiac stomach preparations *in vitro* [4, 5]. Furthermore, S1-immunoreactive and S2-immunoreactive nerve fibres are present in the innervation of the cardiac stomach, in close proximity to the muscle layer [6]. Therefore, it is thought that endogenous release of S1 and/or S2 may be responsible, at least in part, for mediating cardiac stomach eversion when starfish feed.

Comparison of the potency of S1 and S2 as cardiac stomach relaxants *in vitro* has revealed that S2 is approximately ten times more potent than S1 [2, 4, 7]. Similarly, when tested at the same concentration on other muscle preparations from *A. rubens* (tube feet and apical muscle) the relaxing effect of S2 is consistently greater than the effect of S1 [7, 8]. This difference in the potency/activity of S1 and S2 provided a basis for comparative analysis of the solution structures of S1 and S2 using circular dichroism (CD) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy [7]. Consistent with previous studies on small neuropeptides [9-12], CD and NMR data indicate that S1 does not have defined conformation in aqueous solution. In contrast, it was found that S2 has a

remarkably well-defined conformation in aqueous solution, with more than 220 NOEs identified in NMR NOESY data. However, the structuring of S2 is concentration dependent, with increasing concentration inducing a transition from an unstructured to a structured conformation. This indicates that at high concentrations oligomers of S2 are formed through self-association [7] whilst at physiological concentrations S2 remains unstructured.

The most striking difference in the sequences of S1 and S2 is the presence of the N-terminal SGPY tetrapeptide in S2 that is lacking in S1 (Fig. 1). Therefore, it was hypothesized that the N-terminal region of S2 may facilitate self-association of the S2 peptide at high concentrations. Consistent with this hypothesis, it was found that an N-terminally truncated analog of S2 lacking the SGPY tetrapeptide sequence (short S2 or SS2; SFNSGLTFamide) does not have a defined structure in aqueous solution [7]. However, comparative analysis of the bioactivity of S2 and SS2 yielded conflicting findings. S2 was more effective than SS2 as a muscle relaxant when tested at 1 μ M and 10 μ M on cardiac stomach preparations and when tested at 1 μ M on tube foot preparations. However, SS2 was more effective as a muscle relaxant than S2 when tested on tube feet at 10 μ M [7]. It is not clear, therefore, to what extent the presence of the N-terminal SGPY tetrapeptide and its effect in facilitating peptide self-association are important for the bioactivity of S2. Additional studies are now needed to further investigate the structure-activity relationships of S1 and S2.

Here we have analysed the solution structures and bioactivity of three novel chimeric analogs of S1 and S2. Firstly, Long S1 (LS1; SGPYGFNSALMFamide) is a dodecapeptide comprising S1 with the addition of the N-terminal four residues of S2 (SGPY). Analysis of this peptide enabled further investigation of the contribution of the N-terminal SGPY tetrapeptide in facilitating peptide self-association and for bioactivity. Secondly, S1(T), in which the penultimate residue of S1 (methionine) is replaced by the residue that occupies this position in S2 (threonine). Thirdly, S2(M), in which the penultimate residue of S2 (threonine) is replaced by the residue that occupies this position in S1

(methionine). Analysis of S1(T) and S2(M) enabled assessment of the contribution of C-terminal amino acid residues for SALMFamide structure and activity.

2. Materials and Methods

2.1 Materials

Peptides were custom-synthesized by the Advanced Biotechnology Centre at Imperial College London and purified using high performance liquid chromatography. All other chemicals used were obtained from VWR (Poole, Dorset, UK) with the exception of D₂O, which was obtained from Goss Scientific Instruments Ltd (Great Baddow, Essex, UK).

2.2 In vitro pharmacology

S1 and S2 cause dose-dependent relaxation of cardiac stomach and tube foot preparations from the starfish *A. rubens*. Therefore, these preparations were used here to assess the bioactivity of three chimeric analogs of S1 and S2 (LS1, S1(T) and S2(M)), employing the same methodology as reported previously [4, 7, 8]. Specimens of *A. rubens* were obtained from the Menai Straits (UK) and maintained in a circulating seawater aquarium in the School of Biological & Chemical Sciences at QMUL. Cardiac stomach and tube foot preparations were dissected, set up in organ baths containing seawater at 11°C and their contractility was measured using isotonic transducers (model 60-3001; Harvard, South Natick, MA, USA) linked to a chart recorder (Goerz Servogor 124). To enable assessment of the bioactivity of the chimeric peptides as muscle relaxants, sustained contracture of preparations was induced and maintained using seawater with 30 mM added KCl, as described previously [4, 7, 8].

LS1, S1(T) and S2(M) were tested on cardiac stomach preparations (n = 4, 8 and 3, respectively) and tube foot preparations (n = 9, 6 and 6, respectively) at three concentrations, 0.1 μM, 1.0 μM and 10 μM, and the effects of the peptides were expressed as a percentage of the relaxing effect of 10 μM S2, which was ascribed a value of 100%. To directly compare the bioactivity of the chimeric peptides with both S1 and S2, experiments were performed where S1, S2 and a chimeric peptide (LS1, S1(T) or S2(M)) were tested at a concentration of 1 μM on cardiac stomach preparations (n = 12, 7 and

10, respectively) and tube foot preparations (n = 13, 12 and 13, respectively). In these experiments, each peptide was tested twice and the order in which peptides were tested was randomised, with effects quantified by normalisation to the effect of 1 μ M S2.

2.3 Circular dichroism (CD) spectroscopy

CD spectra were recorded using a Chirascan CD spectrometer (Applied Photophysics Ltd.) equipped with a Peltier temperature controller. Spectra were the average of three scans recorded with a 1 nm bandwidth, a 0.5 nm step size, and a 5 s time constant. After background subtraction, the observed ellipticity (θ ; mdeg) was converted to a molar ellipticity (ΔE ; $M^{-1} \text{ cm}^{-1}$), using the formula: $\Delta E = \theta / (33000 lc)$ where l is the path length (cm) and c is the concentration (M). To examine the effect of temperature, CD spectra were obtained every 5°C from 10 to 90° using a 1 mm path length for ~ 0.1 mg/ml samples and a 1 cm path length for ~ 0.01 mg/ml samples.

2.4 NMR spectroscopy

Peptides were dissolved in 10% D₂O/90% H₂O to achieve a final concentration of 2 mM. The pH of the solutions was adjusted to pH 5.6 using 10 mM NaOH and 10 mM HCl. The peptide solutions were centrifuged to remove any suspended material and then the supernatant was transferred to 5 mm NMR tubes. Data were acquired using a Bruker Avance 600 MHz spectrometer and collected using Topspin software on a UNIX workstation. All experiments were performed using a 5 mm TXI, triple resonance probe equipped with a z-axis gradient. 2D NMR spectra were obtained at 303 K for LS1 and 283K for S1(T) and S2(M). Water suppression was achieved using a Water Gradient Tailored Excitation (WATERGATE) technique. 2D-Total Correlation Spectroscopy (TOCSY) experiments employed a DIPSI2 sequence for isotropic mixing, with a 65 ms mixing time. A 300 ms mixing time was used for Nuclear Overhauser Effect Spectroscopy (NOESY) and Rotating-frame Overhauser Effect

Spectroscopy (ROESY) experiments. All 2D experiments used STATES-TPPI phase cycling, and a spectral width of 14 ppm was applied in both dimensions with 2048 x 512 complex data points in the t2 and t1 dimensions, respectively. Prior to Fourier Transformation (FT), the data were linear predicted in the F1 dimension to 512 real points and then zero filled to produce a final matrix size of 4k x 1k, with a 90° phase-shifted sine squared window function applied to both dimensions. The chemical shifts were referenced to water at 4.7 ppm at 303K, pH 5.6. Spin-systems were manually assigned using the TOCSY data and sequential assignments were completed using the 300 ms NOESY or ROESY spectra.

3. Results

3.1 In vitro pharmacology

3.1.1 LongS1 (LS1)

An N-terminally extended analog of S1 (SGPYGFNSALMFamide or LS1) with the addition of the N-terminal tetrapeptide sequence (SGPY) from S2 was synthesized to investigate if addition of the N-terminal region would confer on S1 bioactivity comparable to that of S2. LS1 was tested on cardiac stomach and tube foot preparations at concentrations ranging from 0.1 μM to 10 μM , with the effect at each concentration expressed as a percentage of the effect of 10 μM S2. LS1 caused dose-dependent relaxation of cardiac stomach preparations, with 0.1 μM LS1 causing $38 \pm 10\%$ relaxation and 1 μM causing $67 \pm 8\%$ relaxation. At 10 μM , the effect of LS1 ($69 \pm 18\%$) was not significantly different to the effect of 1 μM LS1 ($p = 0.74$; t-test) (Fig. 2A). LS1 also caused dose-dependent relaxation of tube foot preparations. However, compared to the effects of LS1 on cardiac stomach, LS1 was less effective in causing relaxation of tube feet, with $14 \pm 5\%$ relaxation at 0.1 μM , $20 \pm 4\%$ relaxation at 1 μM and $37 \pm 8\%$ relaxation (Fig. 2C).

To compare the bioactivity of LS1 with S1 and S2, the effects of all three peptides on cardiac stomach (Fig. 2B) and tube foot (Fig. 2D) preparations were compared at a concentration of 1 μM , expressed as a percentage of the effect of 1 μM S2. The mean effect of LS1 on cardiac stomach preparations at 1 μM ($70 \pm 5\%$ relaxation) was significantly less than the mean effect of 1 μM S2 ($p = 4.65 \text{ E-}06$; t-test), but was not significantly different to the mean effect of 1 μM S1 ($p = 0.138$; t-test). The mean effect of LS1 on tube foot preparations at 1 μM ($21 \pm 5\%$ relaxation) was significantly less than the mean effect of 1 μM S2 ($p = 9.04 \text{ E-}10$; t-test) but was not significantly different to the mean effect of 1 μM S1 ($p = 0.14$; t-test).

Collectively, these data indicate that addition of the SGPY tetrapeptide sequence to the N-terminus of S1 does not confer on S1 bioactivity equivalent to that of S2. Furthermore, because the efficacy of LS1 as a muscle relaxant was found not to be significantly different to that of S1, differences in the potency/efficacy of S1 and S2 appear to be largely determined by differences in their C-terminal regions. To test this hypothesis, C-terminal analogs of S1 and S2 were synthesized and tested.

3.1.2 S1(T)

An analog of S1, GFNSALTFamide or S1(T), was synthesized wherein the residue in S1 that is penultimate to the C-terminal amide (methionine) was replaced with the residue that is in an equivalent position in S2 (threonine). S1(T) caused dose-dependent relaxation of cardiac stomach preparations, with 0.1 μM causing $30 \pm 6\%$ relaxation, 1 μM producing $52 \pm 7\%$ relaxation and 10 μM producing $95 \pm 5\%$ relaxation, in comparison with 10 μM S2 (Fig. 2E). S1(T) also caused dose-dependent relaxation of tube foot preparations, with 0.1 μM S1(T) causing $1 \pm 0.2\%$ relaxation, 1 μM S1(T) causing $10 \pm 3\%$ relaxation and 10 μM produced $83 \pm 16\%$ relaxation, in comparison with 10 μM S2 (Fig. 2G).

Comparison of the effects of S1(T), S1 and S2 on cardiac stomach preparations at 1 μM revealed that the effect of S1(T) at 1 μM ($68 \pm 8\%$ relaxation) was significantly less than the effect of 1 μM S2 ($p = 0.006$; t-test) and was not significantly different to the effect of 1 μM S1 ($p = 0.4$; t-test) (Fig. 2F). Comparison of the effects of S1(T), S1 and S2 on tube foot preparations at 1 μM revealed that the effect of S1(T) at 1 μM ($21 \pm 3\%$ relaxation) was significantly less than the effect of 1 μM S2 ($p = 7.856\text{E-}11$; t-test) but was not significantly different to the effect of S1 at 1 μM ($p = 0.1$; t-test) (Fig. 2H).

Thus, collectively the results of the experiments with S1(T) indicate that substitution of the C-terminal methionine residue in S1 with the equivalently positioned residue in S2 is not sufficient to confer on S1 the potency/efficacy of S2.

3.1.3 S2(M)

An analog of S2, SGPYSFN SGLMFamide or S2(M), was synthesized wherein the residue in S2 that is penultimate to the C-terminal amide (threonine) was replaced with the residue that is in an equivalent position in S1 (methionine). S2(M) caused dose-dependent relaxation of cardiac stomach preparations, with 0.1 μM S2(M) causing $18 \pm 8\%$ relaxation, 1 μM S2(M) causing $43 \pm 6\%$ relaxation and 10 μM causing $99 \pm 10\%$ relaxation, in comparison with the effect of 10 μM S2 (Fig. 2I). S2(M) also caused dose-dependent relaxation of tube foot preparations, with 0.1 μM S2(M) causing $5 \pm 4\%$ relaxation, 1 μM causing $15 \pm 4\%$ relaxation and 10 μM causing $40 \pm 10\%$ relaxation, in comparison with the effect of 10 μM S2 (Fig. 2K).

Comparison of the effects of S2(M), S1 and S2 at 1 μM on cardiac stomach preparations revealed that the effect of S2(M) was significantly less than the effect of 1 μM S2 ($47 \pm 5\%$ relaxation; $p = 1.648\text{E-}8$; t-test) but not significantly different to the effect of 1 μM S1 ($p = 0.012$; t-test) (Fig. 2J). Likewise, comparison of the effect of S2(M), S1 and S2 at 1 μM on tube foot preparations revealed that the effect of S2(M) was significantly less than the effect of S2 ($13 \pm 3\%$ relaxation; $p = 8.999\text{E-}12$; t-test) but not significantly different to the effect of S1 at 1 μM ($p = 0.241$; t-test) (Fig. 2L).

Collectively, these data indicate the C-terminally located threonine residue in S2 contributes to its superior potency/efficacy in comparison with S1 because substitution of this residue with the equivalently positioned residue from S1 (methionine) leads to a loss of bioactivity such that the effects of the S2 analog S2(M) are not significantly different to the effects of S1.

3.2 Structural analysis of LS1, S1(T) and S2(M) using NMR

3.2.1 Long S1 (LS1)

Previously we determined that loss of the four N-terminal amino acids from S2 resulted in a loss of concentration dependent structuring and loss of bioactivity [7]. However, addition of SGPY to the N-terminus of S1 does not confer S2-like bioactivity (Fig. 2). Therefore, removal of the N-terminal tetrapeptide from S2 (short S2 or SS2) results in loss of structure, but addition of the SGPY motif to S1 may not be sufficient to generate structure. To investigate this we obtained two-dimensional TOCSY and NOESY NMR spectra of LS1. From the 2D TOCSY we were able to assign all residues except Ser₁. When the ROESY spectrum of S1 (Fig. 3A) is compared to that of LS1 (Fig. 3B) it is clear that LS1 has far more NOE peaks than S1. The greater number of NOEs in LS1 (Fig. 3B) suggests that the addition of four N-terminal amino acids has resulted in the generation of some structure in LS1. However, the presence of substantial structure would result in a number of long-range NOE contacts. Inspection of the H_N-H_α region of the overlaid LS1 TOCSY and NOESY spectra shows that there are mostly only sequential connections apparent (Fig. 3C). Only one medium range NOE, a d_{αN}(*i*, *i*+2), is apparent between Leu₁₀ and Phe₁₂, however this may be an ambiguous assignment and will be discussed later. Also observed is a d_{NN}(*i*, *i*+2) between Ala₉ and Leu₁₁ as well as sequential d_{NN}(*i*, *i*+1) connections between Gly₅ and Phe₆ and Leu₁₁ and Phe₁₂ (not shown). This suggests that most of the apparent structure is located in the C-terminal region. This result is in contrast to S2, in which many medium and long range NOEs covering the entire peptide were apparent in the H_N-H_α region [7]. Overall, in the LS1 NOESY spectrum we were able to identify 77 NOEs compared to the more than 220 NOEs observed in S2 [7]. Thus, the N-terminal tetrapeptide can induce some conformational preference, but the structure is not as extensive as observed in S2 and in itself is not sufficient to promote S2-like bioactivity.

3.2.2 S1(T)

S1(T) contains a threonine residue in the penultimate position, as in S2, instead of the methionine residue present in S1. Other than the four N-terminal residues, S1 differs from S2 only at this position and at position five (S1 numbering) which is a glycine in S2 and an alanine in S1. We hypothesised that the threonine for methionine substitution may result in increased potency/efficacy but the pharmacological tests showed that bioactivity of S1(T) was not significantly different to S1 (Fig. 2E – H). Analysis of the NMR data supports this finding. The ROESY spectrum of S1(T) (Fig. 4A) is similar to that of S1 (Fig. 4B). The chemical shifts are all near their ‘random coil’ values and there are no ROEs present beyond sequential connectivities, indicating that S1(T), like S1, is largely unstructured. If the penultimate threonine is indeed important in receptor binding, perhaps by being involved in hydrogen-bonding via the side-chain hydroxyl group, then it must require the structuring induced by the N-terminal tetrapeptide in S2 in order to be placed in the correct position.

3.2.3 S2(M)

If Thr₁₁ is important for S2 bioactivity, then S2(M), which contains the penultimate methionine from S1 instead of Thr₁₁, should have reduced bioactivity. The pharmacology results have supported this, showing that S2(M) is no more effective than S1 as a muscle relaxant (Fig. 2I-L). Comparison of the H_N-H_α region of the ROESY spectrum of S2(M) (Fig. 5A) with the same region of S2 (Fig. 5B) shows that S2(M) also lacks the array of ROEs indicative of structure that are abundant in S2. Likely, a loss of structure has resulted in the lack of ROEs. Indeed, the number of ROEs in S2(M) is similar to LS1 as might be expected given that S2(M) and LS1 differ only by one amino acid at position 9 – alanine in LS1 and glycine in S2(M). Closer inspection of the H_N-H_α region of the overlaid TOCSY and ROESY spectra in Fig. 5C shows that S2(M) adopts multiple conformations. All long peptides (S2, LS1, S2(M)) that contain Pro₃ have *cis* and *trans* conformations due to the proline, however S2(M) also

has an additional two conformations around Met₁₁. Figure 5C shows that the sequential H_N-H_α connections can follow two divergent paths from Gly₉ and consequently Leu₁₀, Met₁₁ and Phe₁₂ have two distinct chemical shifts. This suggests that S2(M) can adopt at least four distinct conformations – Pro₃*Trans*Met_{11a}, Pro₃*Trans*Met_{11b}; Pro₃*Cis*Met_{11a} and Pro₃*Cis*Met_{11b}. Possibly, the splitting of the C-terminal residues is a consequence of the *cis/trans* Pro₃ isomerism and could suggest association of the N-terminal region with the C-terminus either within a folded monomer or possibly in a ‘head-to-tail’ dimer. However, it is notable that the *cis*-Proline is a minor component while the signals for the two sets of Met₁₁ have similar intensities. If the two conformations of Leu₁₀, Met₁₁ and Phe₁₂ were a result of the Pro₃ *cis/trans* isomerisation then it could reasonably be expected that the peak intensities would mimic the Pro₃ intensities, i.e. one set would have markedly reduced intensity. It is likely then that the splitting of Leu₁₀ – Phe₁₂ residues is not related to the *cis/trans* Pro₃ isomerism. The ambiguity in the assignment of the d_{αN}(*i*, *i*+2) connection between Leu₁₀ and Phe₁₂ in LS1 is because there may also be some minor doubling of the Met₁₁ peak in LS1, however any chemical shift difference is minor and the TOCSY does not provide sufficient evidence for a second peak. The clear presence of multiple conformations in the C-terminal region of S2(M) would undoubtedly impact on the ability of the peptide to bind to receptors and induce muscle relaxation. Taken together, the pharmacology and structural data indicate that the combination of the four N-terminal residues along with the presence of threonine in the penultimate position is necessary to form a peptide with optimal bioactivity. Possibly, the side-chain hydroxyl is required for hydrogen bonding with the S2 receptor.

3.3 Structural analysis of LS1 and S2(M) using CD spectroscopy

The number of ROEs observed in the ROESY spectra of LS1 and S2(M) (Fig. 3, 5) has suggested that both these peptides contain some structure. Previous studies have shown that structure in S2 was a consequence of self-association [7], and therefore we now wondered if LS1 and S2(M) also

exhibit S2-like concentration-dependent structuring. CD is a useful technique for the determination of protein and peptide structure. In the UV region of the CD spectrum the presence of stable structure results in optical activity due to the chirality around the amide bond. Initially, we obtained CD spectra of LS1 and S2(M) at low concentrations similar to those used in the pharmacology experiments and higher concentrations similar to the concentrations used in the NMR experiments (Fig. 6). For LS1 (Fig. 6A) the CD spectra obtained at both concentrations are very similar. At the lower concentration (0.011 mg/mL) the presence of a strong negative peak near 200 nm along with a positive peak at 225 nm suggests the presence of polyproline-II (PP-II) character [13]. PP-II is a highly extended conformation that does not contain backbone hydrogen bonds and is frequently found in ‘unstructured’ regions of proteins. At a ten-fold higher concentration, the UV CD spectrum of LS1 is similar to that at the lower concentration, the negative peak at 200 nm has got stronger (i.e. more negative) but a small peak at 225 nm is retained. This suggests that a similar structure is adopted at both concentrations.

The CD spectra of S2(M) at high and low concentrations (Fig. 6B) are very different to LS1 (Fig. 6A) and to S2 [7]. The spectra have a positive peak at 207 nm and negative peaks at 215 nm and 225 nm. In the higher concentration spectrum the peak at 215 nm that is readily apparent at the lower concentration appears as a shoulder on the more intense 225 nm peak. The S2(M) spectra are quite novel and don't resemble the spectra that would be obtained from any regular secondary structural element nor an unstructured peptide, which would feature a single negative band at ~198 nm. Furthermore, the lack of similarity to LS1 suggests that S2(M) also does not adopt a PPII conformation. Given that the NMR data suggest that S2(M) can adopt at least four different conformations, the CD spectrum at the higher concentration is likely a summation of all of these different conformations. The fact that the CD spectrum of the lower concentration does not resemble the spectrum of an unstructured peptide indicates that, in contrast to S2, S2(M) retains some structure at the concentrations used in the pharmacology experiments.

To further investigate the structure of LS1 and S2(M) we undertook temperature dependent unfolding experiments. As the temperature is increased from 10°C to 90°C the CD spectra of LS1 shows a loss of the positive band at 225 nm at both low (Fig. 7A) and high (Fig. 7B) concentrations. These data suggest that the PPII structure is lost at higher temperatures; it is notable that even at 90°C, and at both concentrations, a negative CD band is observed at 230 nm suggesting some self-association. As expected, the loss of PPII structure is not cooperative unfolding at either concentration (Fig. 7C, D). The rather unusual CD spectra of S2(M) are very different to those of LS1 - at the higher concentration (Fig. 8A) the negative CD signal at 225 nm remains almost unaffected by raising the temperature to 70°C, while there is a significant loss of positive signal at 207 nm. At the lower S2(M) concentration (Fig. 8B) the signal loss at 207 nm is more marked, accompanied by a negative band at 217 nm even at 70°C, which suggests an extended conformation due to self-association. It is notable that at 70°C with both concentrations of S2(M) the spectra do not show evidence for a completely unfolded peptide. The thermal denaturation curves (Fig. 8C, D) suggest a sharp transition in structure at 50°C with a marked loss of the positive CD band at 207 nm.

4. Discussion

The SALMFamide neuropeptides S1 and S2, which share 67% sequence identity (Fig. 1), act as muscle relaxants in starfish, but S2 is about ten times more potent than S1. In a previous study [7] a structural basis for this difference in the potency of S1 and S2 was investigated, focusing on the N-terminal SGPY tetrapeptide that is a feature of the S2 dodecapeptide (SGPYSFNSGLTFamide) but not the S1 octapeptide (GFNSALMFamide). It was found that the N-terminal SGPY motif confers on S2 the property of self-association induced structure at high concentrations. Given that S2 structuring is only apparent at high concentrations it is most likely not relevant to the physiological roles of S2 as a muscle relaxant in starfish. We note, though, that recent studies with bradykinin peptides have shown that conformation can be quite dependent on solution composition and it remains possible that other conditions will modify S2's oligomerization and conformation [14]. We have proposed that the ability of S2 to oligomerize may have implications for peptide packing in vesicles prior to release and possibly also for autocrine actions of the peptide directly after release when local concentrations are high. Analysis of the bioactivity of an N-terminally truncated form of S2 (SS2; SFNSGLTFamide), which like S1 does not exhibit self-association induced structure, yielded ambiguous findings - the bioactivity of SS2 was found to be either inferior or superior to that of S2, depending on the dose and muscle preparation tested. Thus, the contribution of the N-terminal SGPY sequence to the bioactivity of S2 remains unclear. Here we have further investigated this issue by analysis of the bioactivity and structural properties of chimeric analogs of S1 and S2.

Long S1 or LS1 (SGPYGFNSALMFamide), a chimeric peptide comprising the N-terminal tetrapeptide of S2 and the C-terminal octapeptide of S1, was tested and analysed to investigate if addition of the N-terminal SGPY sequence would confer S2-like bioactivity and structure on S1. It was found that the bioactivity of LS1 as a muscle relaxant was not significantly different to that of S1 and

that LS1 does not exhibit self-association induced structure. However, some structuring was detected in the C-terminal region of LS1 and the presence of a positive peak at 225 nm in CD spectra suggests the presence of a polyproline-II-type conformation. Thus, addition of the N-terminal SGPY motif does confer some structure on S1, but it is not equivalent to that seen with S2 and does not appear to enhance bioactivity. Collectively, these findings indicate that the superior bioactivity of S2 in comparison with S1 is not largely attributable to the presence of the N-terminal SGPY motif in S2. Therefore, we investigated the contribution of residues that differ in the C-terminal region of S1 and S2.

The bioactivity of S1(T), an analog of S1 with a threonine residue replacing the methionine residue that is penultimate to the C-terminal amide in S1, was found not to be significantly different to that of S1. Furthermore, like S1 [7], the S1(T) octapeptide did not exhibit a stable structure in solution. Thus, conversion of the C-terminal region of S1 (LMFamide) to an S2-type C-terminal sequence (LTFamide) is not sufficient on its own to confer S2-like bioactivity on S1. However, conversion of the C-terminal region of S2 (LTFamide) to an S1-type C-terminal sequence (LMFamide) results in loss of S2-type bioactivity – thus, the relaxing effects of the S2 analog S2(M) are significantly less than the effects of S2 and not significantly different to the effects of S1. Interestingly, this amino acid conversion also results in loss of the self-associating property of S2 to form a stable conformation at high concentrations. However, like the LS1 peptide, S2(M) does exhibit some structuring but it lacks the stability seen with S2 at high concentrations and appears to adopt multiple conformations.

Collectively, the findings of our previous study [7] and the findings from the experiments reported here indicate that the superior bioactivity of S2 compared to S1 is largely attributable to differences in the C-terminal region of these peptides. In particular, the loss of bioactivity seen in the S2 analog S2(M) indicates that the C-terminal LTFamide motif of S2 appears to be a key determinant of its superior bioactivity in comparison with S1. This presumably reflects differences in the properties of threonine and methionine and hence how they may interact with residues that form receptor binding

sites for S2 - perhaps the side-chain hydroxyl group in the threonine residue is required for formation of hydrogen bonds. The presence of multiple conformations around Met₁₁ in S2(M) suggests that, in addition to Thr₁₁ potentially supporting interactions with the S2 receptor, it is also possible that this residue supports intra- or inter-molecular hydrogen-bonding. However, the contribution of the N-terminal SGPY sequence to the bioactivity of S2 should not be dismissed. The reduced bioactivity of the N-terminally truncated form of S2 (SS2) seen in some preparations [7] indicates that the SGPY motif contributes to S2's bioactivity. The SGPY motif clearly confers varying degrees of structural stability on the dodecapeptides S2, LS1 and S2(M) and we propose that it is this property in combination with the C-terminal residues of S2 that accounts for S2's superior bioactivity in comparison with S1. Further understanding of S2's bioactivity might be achieved if we could solve the structure of the high-concentration multimeric structure. However, solving the structure of monomeric S2 based on our NOESY data is very difficult because at NMR concentrations S2 is multimeric and lowering the concentration results in loss of structure. Under NMR conditions deconvoluting our NOE data to determine intra-monomer from inter-monomer contacts is not possible, especially for a peptide of only 12 amino-acids. Furthermore, it is difficult to assess how the multimeric structure has influenced NOE intensities and thus conversion of peak intensity to distant restraints is likely to result in incorrect distances. For example, a 'head-to-head' dimer could result in NOE contacts for the same proton pair that is a mixture of both intra- and inter-residue contacts. Diffusion Ordered Spectroscopy of S2 was undertaken to try to elucidate the arrangement of the oligomeric structure (data not shown), but given the small size of the peptide it is very difficult to deconvolute the diffusion coefficient to differentiate an unfolded monomer from a folded dimer, trimer or tetramer.

In conclusion, our analysis of chimeric SALMFamide neuropeptides has revealed how neuropeptide bioactivity is determined by a complex interplay of sequence and conformation. Further insights into the structural basis of the differences in the potency of S1 and S2 as muscle relaxants in starfish could be obtained if the receptor(s) that mediate the effects of these peptides is/are identified.

Based on sequence similarities that SALMFamide-type neuropeptides share with neuropeptides that have been identified in other phyla, candidate G-protein coupled receptors for SALMFamides have been proposed (e.g. GnIH/NPFF-type and QRFP-type receptors; [1, 15]). However, definitive insights will require cloning, heterologous expression and pharmacological characterisation of candidate receptors.

Another fascinating avenue for future research will be to extend our analysis of the bioactivity and structural properties of S1, S2 and their chimeric analogs to other SALMFamide-type neuropeptides that have been discovered in starfish and in other echinoderms. Analysis of the genome sequence of the starfish *Patiria miniata* has revealed that S1 is derived from a precursor protein that contains six other L-type SALMFamides. On the other hand, the S2-like neuropeptide in *P. miniata* (SNGPYSMSGLRSLTFamide) is derived from a different precursor protein, which also comprises eight F-type or F-type-like SALMFamides [16]. Recently, sequencing of the neural transcriptome of *A. rubens* has enabled identification of corresponding precursors in this species, with S1 derived from a precursor comprising six L-type SALMFamides and S2 derived from a precursor comprising seven F-type or F-type-like SALMFamides (D.C. Semmens & M.R. Elphick, unpublished data).

Immunocytochemical localisation of S1 and S2 in *A. rubens* has revealed that these two peptides, and hence their precursor-associated peptides, are expressed in different populations of neurons [6, 17]. Therefore, we can collectively infer from these observations that S1 is released from neurons as one constituent of a “cocktail” of seven L-type SALMFamides. On the other hand, it appears that S2 is released by a different population of neurons as one constituent of a “cocktail” of eight SALMFamides, with the seven other peptides being F-type or F-type-like SALMFamides. Therefore, in comparing the bioactivity and structural properties of S1 and S2, as reported previously [7] and here, we are in fact comparing two L-type SALMFamides that are derived from different precursors and different populations of neurons. Furthermore, because both S1 and S2 are derived from precursor proteins that contain other SALMFamides, testing and comparing the effects of S1 and S2 *in vitro* is not

representative of a physiological scenario. In future studies it will be of interest to compare the structural and functional properties of S1 with other L-type SALMFamides that are derived from the S1 precursor. Furthermore, analysis of the *in vitro* actions of a “cocktail” of the SALMFamides that are derived from the S1 precursor will reveal pharmacological actions that are representative of physiological processes *in vivo*. Similarly, it will be interesting to compare the effects of S2 with the effects of F-type and F-type-like SALMFamides that are derived from the S2 precursor and to compare the effects of peptides tested individually with the effects of the entire “cocktail” of SALMFamides that are derived from the S2 precursor. Combined with comparative analysis of the structural properties of all fifteen SALMFamides that are derived from the S1 and S2 precursors, it may then be possible to gain insights into the evolutionary and functional significance of the heterogeneous mixtures of SALMFamides that occur in starfish and in other echinoderms.

Our work on SALMFamides has broad relevance because the occurrence of multiple related neuropeptides derived from a single precursor protein is a common phenomenon, particularly in invertebrates [18]. However, the functional significance of these neuropeptide “cocktails” is not understood [19]. As an example, in the nematode *C. elegans*, a single precursor protein generates eight similar peptides belonging to the FMRFamide family of neuropeptides. Deletion of this gene results in five distinct and different behavioural defects but it is not clear how individual peptides contribute to a specific phenotype [20]. Similar to our findings in this work, it appears that conformation preferences of individual FMRFamide-type neuropeptides in molluscs play a role in modulating receptor binding. It is suggested that only a single conformation will bind to the receptor and therefore the effective neuropeptide concentration is less than the expressed concentration [21]. Thus, an understanding of the interplay of sequence, conformation and intra- and inter-peptide interactions may be the key to elucidation of the functional significance of neuropeptide “cocktails”.

Acknowledgements - We are grateful to Harold Toms (QMUL) for technical assistance with the NMR studies and Richard Melarange (QMUL) for technical assistance with *in vitro* pharmacology.

References

- [1] M.R. Elphick, SALMFamide salmagundi: The biology of a neuropeptide family in echinoderms, *Gen Comp Endocrinol*, (2014).
- [2] M.R. Elphick, R. Melarange, Neural control of muscle relaxation in echinoderms, *J Exp Biol*, 204 (2001) 875-885.
- [3] M.R. Elphick, D.A. Price, T.D. Lee, M.C. Thorndyke, The SALMFamides: a new family of neuropeptides isolated from an echinoderm, *Proc Biol Sci*, 243 (1991) 121-127.
- [4] R. Melarange, D.J. Potton, M.C. Thorndyke, M.R. Elphick, SALMFamide neuropeptides cause relaxation and eversion of the cardiac stomach in starfish., *Proc Biol Sci*, 266 (1999) 1785-1789.
- [5] M.R. Elphick, S.J. Newman, M.C. Thorndyke, Distribution and action of SALMFamide neuropeptides in the starfish *Asterias rubens*, *J Exp Biol*, 198 (1995) 2519-2525.
- [6] S.J. Newman, M.R. Elphick, M.C. Thorndyke, Tissue distribution of the SALMFamide neuropeptides S1 and S2 in the starfish *Asterias rubens* using novel monoclonal and polyclonal antibodies. 2. Digestive system, *Proc Biol Sci*, 261 (1995) 187-192.
- [7] C.B. Obara, C.E. Jones, N.D. Younan, J.H. Viles, M.R. Elphick, Structural analysis of the starfish SALMFamide neuropeptides S1 and S2: The N-terminal region of S2 facilitates self-association, *Bba-Proteom*, 1844 (2014) 358-365.
- [8] R. Melarange, M.R. Elphick, Comparative analysis of nitric oxide and SALMFamide neuropeptides as general muscle relaxants in starfish, *J Exp Biol*, 206 (2003) 893-899.
- [9] D.A. Keire, T.E. Solomon, J.R. Reeve, Jr., NMR evidence for different conformations of the bioactive region of rat CCK-8 and CCK-58, *Biochem Biophys Res Commun*, 293 (2002) 1014-1020.
- [10] G.Y. Xu, C.M. Deber, Conformations of neurotensin in solution and in membrane environments studied by 2-D NMR spectroscopy, *Int J Pept Protein Res*, 37 (1991) 528-535.
- [11] J.L. Nieto, M. Rico, J. Santoro, J. Herranz, F.J. Bermejo, Assignment and conformation of neurotensin in aqueous solution by ¹H NMR, *Int J Pept Protein Res*, 28 (1986) 315-323.
- [12] G. Chassaing, O. Convert, S. Lavielle, Preferential conformation of substance P in solution, *Eur J Biochem*, 154 (1986) 77-85.
- [13] A.A. Adzhubei, M.J. Sternberg, A.A. Makarov, Polyproline-II helix in proteins: structure and function, *Journal of molecular biology*, 425 (2013) 2100-2132.
- [14] N.A. Pierson, L. Chen, S.J. Valentine, D.H. Russell, D.E. Clemmer, Number of solution states of bradykinin from ion mobility and mass spectrometry measurements, *J Am Chem Soc*, 133 (2011) 13810-13813.
- [15] M.R. Elphick, From gonadotropin-inhibitory hormone to SIFamides: Are echinoderm SALMFamides the "missing link" in a bilaterian family of neuropeptides that regulate reproductive processes?, *Gen Comp Endocrinol*, (2013).
- [16] M.R. Elphick, S. Achhala, N. Martynyuk, The evolution and diversity of SALMFamide neuropeptides, *PLoS One*, 8 (2013) e59076.
- [17] S.J. Newman, M.R. Elphick, M.C. Thorndyke, Tissue distribution of the SALMFamide neuropeptides S1 and S2 in the starfish *Asterias rubens* using novel monoclonal and polyclonal antibodies. I. Nervous and locomotory systems, *Proc Biol Sci*, 261 (1995) 139-145.
- [18] C. Wegener, A. Gorbashov, Molecular evolution of neuropeptides in the genus *Drosophila*, *Genome Biol*, 9 (2008) R131.
- [19] R.S. Hewes, E.C. Snowdeal, 3rd, M. Saitoe, P.H. Taghert, Functional redundancy of FMRFamide-related peptides at the *Drosophila* larval neuromuscular junction, *J Neurosci*, 18 (1998) 7138-7151.
- [20] L.S. Nelson, M.L. Rosoff, C. Li, Disruption of a neuropeptide gene, *flp-1*, causes multiple behavioral defects in *Caenorhabditis elegans*, *Science*, 281 (1998) 1686-1690.
- [21] A.S. Edison, E. Espinoza, C. Zachariah, Conformational ensembles: the role of neuropeptide structures in receptor binding, *J Neurosci*, 19 (1999) 6318-6326.

Figure Legends

Figure 1. Comparison of the sequences of S1, S2, SS2 (short S2), LS1 (long S1), S1(T) and S2(M).

Residues that are identical in all six peptides are shown in black and residues that are variable or are not present in some peptides are shown in red.

Figure 2. Analysis of pharmacological effects of LS1, S1(T) and S2(M) on cardiac stomach and tube foot preparations. A. LS1 causes dose-dependent relaxation of cardiac stomach preparations. However, although the effect 1 μ M LS1 is significantly greater than 0.1 μ M LS1 ($p = 0.02$; t-test), the effect of 1 μ M LS1 is not significantly different to the effect of 10 μ M LS1 ($p = 0.74$; t-test). B. When tested at 1 μ M, LS1 is significantly less effective than S2 ($p = 4.65E-06$; ***, t-test) but is not significantly different to S1 ($p = 0.138$; t-test) as a relaxant of cardiac stomach preparations. C. LS1 causes dose-dependent relaxation of tube foot preparations. D. When tested at 1 μ M, LS1 is significantly less effective than S2 ($p = 9.04E-10$; ***, t-test) but is not significantly different to S1 ($p = 0.14$; t-test) as a relaxant of tube foot preparations. E. S1(T) causes dose-dependent relaxation of cardiac stomach preparations. F. When tested at 1 μ M, S1(T) is significantly less effective than S2 ($p = 0.0055$; ***, t-test) but is not significantly different to S1 ($p = 0.4$; t-test) as a relaxant of cardiac stomach preparations. G. S1(T) causes dose-dependent relaxation of tube foot preparations. H. When tested at 1 μ M, S1(T) is significantly less effective than S2 ($p = 7.856 E-11$; ***, t-test) but is not significantly different to S1 ($p = 0.129$; t-test) as a relaxant of the tube foot preparations. I. S2(M) causes dose-dependent relaxation of cardiac stomach preparations. J. When tested at 1 μ M, S2(M) is significantly less effective than S2 ($p = 1.648E8$; ***, t-test) but is not significantly different to S1 ($p = 0.012$; t-test) as a relaxant of cardiac stomach preparations. K. S2(M) causes dose-dependent relaxation of tube foot preparations. L. When tested at 1 μ M, S2(M) is significantly less effective than S2 ($p = 8.999E-12$;

***; t-test) but is not significantly different to S1 ($p = 0.241$; t-test) as a relaxant of the tube foot preparations.

Figure 3. Analysis by NMR (600 MHz) suggests LS1 contains some structure. The amide region of the ROESY spectrum ($\tau_m = 300$ ms) of S1 (A) shows no non-sequential cross-peaks which correlates with the lack of structure in this peptide. In contrast, the NOESY spectrum ($\tau_m = 300$ ms) of LS1 (B) (2mM, pH 5.6, 303K) shows non-sequential cross-peaks, especially to the side-chains of Phe₆/Phe₁₂ between 7.0 and 7.5 ppm, suggesting that LS1 contains some structure. Overlay of the NH-H _{α} region of the LS1 ROESY spectrum (red) and the TOCSY spectrum ($\tau_m = 65$ ms, blue) shows the sequential connections (solid lines) along with a possible $d_{\alpha N}(i, i+2)$ connection between L10 and F12 (dotted line).

Figure 4. S1T is not structured. The H_N-H _{α} region of 600 MHz ROESY spectrum ($\tau_m = 300$ ms) of S1(T) (2 mM, pH 5.6, 283K) (A) shows only sequential connectivities similarly to S1 (B) and both have a lack of medium and long range contacts that would indicate structure. The box labelled G₁ represents the intraresidue H_N-H _{α} peak of G₁ observed in the TOCSY but not in the ROESY.

Figure 5. S2(M) contains multiple conformations. The amide region of the 600 MHz NOESY spectrum ($\tau_m = 300$ ms) of S2(M) (2 mM, pH 5.4, 283K) (A) compared to the same region of the NOESY spectrum of S2 (B) (C) Overlay of the NH-H _{α} region of the S2(M) NOESY spectrum (red) and TOCSY spectrum (blue) shows sequential connections (solid lines). The dashed line highlights a second set of sequential connections from Leu₁₀ to Phe₁₂ that suggests the presence of an alternative conformation.

Figure 6. CD of LS1 and S2(M) at high and low concentrations. Panel A shows the CD spectra of LS1 at low (0.011 mg/mL, black trace) and high (0.13 mg/mL, red trace) concentrations. The positive peak at 225 nm, particularly at low concentration, suggests the presence of polyproline II conformation. The CD spectra of S2(M) (Panel B) at low (0.012 mg/mL, black trace) and high (0.18 mg/mL red trace) are quite unusual and not typical of regular secondary structure. The presence of the positive band at 207 nm and the negative band at 225 nm in both spectra suggests that the peptide has some similarity at both concentrations.

Figure 7. Thermal melting CD spectra of LS1. Panel A shows the CD spectra of LS1 (0.13 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 90°C (purple trace) in 5°C increments. Panel B shows the CD spectra of LS1 (0.011 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 90°C (purple trace). In both cases a spectrum was obtained at 10°C after cooling from 90°C and the unfolding was reversible. In neither high (Panel A) nor low concentration (Panel B) spectra is there evidence of a completely unfolded peptide, even at 90°C. Panels C and D show the ellipticity at 223 nm plotted as a function of temperature, and in neither case is there evidence of cooperative unfolding.

Figure 8. Thermal melting CD spectra of S2(M). Panel A shows the CD spectra of S2(M) (0.18 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 70°C (blue trace) in 5°C increments. Panel B shows the CD spectra of S2(M) (0.012 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 70°C (purple trace). In neither high (Panel A) nor low concentration (Panel B) spectra is there evidence of a completely unfolded peptide at the highest temperature tested. Panels C and D show the ellipticity at 207 nm, 217 nm and 225 nm plotted as a function of temperature.

**Bioactivity and structural properties of chimeric analogs of
the starfish SALMFamide neuropeptides S1 and S2.**

Christopher E. Jones^{a, b*}, Claire B. Otara^{a*}, Nadine D. Younan^a,

John H. Viles^a and Maurice R. Elphick^a

^a. School of Biological & Chemical Sciences, Queen Mary University of London, Mile End Road,
London, E1 4NS, UK. ^b. School of Science and Health, University of Western Sydney, Locked bag
1797, Penrith, 2751, Sydney, Australia.

* CEJ and CBO contributed equally

Running title: *Analysis of chimeric analogs of the SALMFamides S1 and S2*

To whom correspondence should be addressed:

Maurice R. Elphick, School of Biological & Chemical Sciences, Queen Mary University of London,
Mile End Road, London, E1 4NS, UK. Tel: 0207 882 6664; Fax: 0207 882 7732; E-mail:

m.r.elphick@qmul.ac.uk

or

John H. Viles, School of Biological & Chemical Sciences, Queen Mary University of London, Mile
End Road, London, E1 4NS, UK. Tel: 0207 882 8443; Fax: 0207 882 7732; E-mail:

j.viles@qmul.ac.uk

Abstract

The starfish SALMFamide neuropeptides S1 (GFNSALMFamide) and S2 (SGPYSFNSGLTFamide) are the prototypical members of a family of neuropeptides that act as muscle relaxants in echinoderms. Comparison of the bioactivity of S1 and S2 as muscle relaxants has revealed that S2 is ten times more potent than S1. Here we investigated a structural basis for this difference in potency by comparing the bioactivity and solution conformations (using NMR and CD spectroscopy) of S1 and S2 with three chimeric analogs of these peptides. A peptide comprising S1 with the addition of S2's N-terminal tetrapeptide (Long S1 or LS1; SGPYGFNSALMFamide) was not significantly different to S1 in its bioactivity and did not exhibit concentration-dependent structuring seen with S2. An analog of S1 with its penultimate residue substituted from S2 (S1(T); GFNSALTFamide) exhibited S1-like bioactivity and structure. However, an analog of S2 with its penultimate residue substituted from S1 (S2(M); SGPYSFNSGLMFamide) exhibited loss of S2-type bioactivity and structural properties. Collectively, our data indicate that the C-terminal regions of S1 and S2 are the key determinants of their differing bioactivity. However, the N-terminal region of S2 may influence its bioactivity by conferring structural stability in solution. Thus, analysis of chimeric SALMFamides has revealed how neuropeptide bioactivity is determined by a complex interplay of sequence and conformation.

1. Introduction

The SALMFamides are a family of neuropeptides that occur in species belonging to the phylum Echinodermata (e.g. starfish, sea cucumbers, sea urchins). There are two types of SALMFamides – L-type, which have the C-terminal motif SxLxFamide (where x is variable), and F-type, which have the C-terminal motif SxFxFamide. Furthermore, analysis of the pharmacological actions of SALMFamides has revealed that both L-type and F-type SALMFamides cause muscle relaxation in echinoderms [1, 2].

The first members of the SALMFamide neuropeptide family to be identified were the L-type SALMFamides S1 and S2, which were both isolated from the nervous system of the starfish species *Asterias rubens* and *Asterias forbesi* [3]. S1 is an octapeptide with the amino acid sequence H-Gly-Phe-Asn-Ser-Ala-Leu-Met-Phe-NH₂ and S2 is a dodecapeptide with the amino acid sequence H-Ser-Gly-Pro-Tyr-Ser-Phe-Asn-Ser-Gly-Leu-Thr-Phe-NH₂ (Fig. 1). Interestingly, injection of S1 or S2 into *A. rubens* triggers cardiac stomach eversion, a process that occurs naturally when starfish feed extra-orally on prey such as mussels [4]. Consistent with the effects of S1 and S2 *in vivo*, both peptides cause dose-dependent relaxation of cardiac stomach preparations *in vitro* [4, 5]. Furthermore, S1-immunoreactive and S2-immunoreactive nerve fibres are present in the innervation of the cardiac stomach, in close proximity to the muscle layer [6]. Therefore, it is thought that endogenous release of S1 and/or S2 may be responsible, at least in part, for mediating cardiac stomach eversion when starfish feed.

Comparison of the potency of S1 and S2 as cardiac stomach relaxants *in vitro* has revealed that S2 is approximately ten times more potent than S1 [2, 4, 7]. Similarly, when tested at the same concentration on other muscle preparations from *A. rubens* (tube feet and apical muscle) the relaxing effect of S2 is consistently greater than the effect of S1 [7, 8]. This difference in the potency/activity of S1 and S2 provided a basis for comparative analysis of the solution structures of S1 and S2 using circular dichroism (CD) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy [7]. Consistent with previous studies on small neuropeptides [9-12], CD and NMR data indicate that S1 does not have defined conformation in aqueous solution. In contrast, it was found that S2 has a

remarkably well-defined conformation in aqueous solution, with more than 220 NOEs identified in NMR NOESY data. However, the structuring of S2 is concentration dependent, with increasing concentration inducing a transition from an unstructured to a structured conformation. This indicates that at high concentrations oligomers of S2 are formed through self-association [7] whilst at physiological concentrations S2 remains unstructured.

The most striking difference in the sequences of S1 and S2 is the presence of the N-terminal SGPY tetrapeptide in S2 that is lacking in S1 (Fig. 1). Therefore, it was hypothesized that the N-terminal region of S2 may facilitate self-association of the S2 peptide at high concentrations. Consistent with this hypothesis, it was found that an N-terminally truncated analog of S2 lacking the SGPY tetrapeptide sequence (short S2 or SS2; SFNSGLTFamide) does not have a defined structure in aqueous solution [7]. However, comparative analysis of the bioactivity of S2 and SS2 yielded conflicting findings. S2 was more effective than SS2 as a muscle relaxant when tested at 1 μ M and 10 μ M on cardiac stomach preparations and when tested at 1 μ M on tube foot preparations. However, SS2 was more effective as a muscle relaxant than S2 when tested on tube feet at 10 μ M [7]. It is not clear, therefore, to what extent the presence of the N-terminal SGPY tetrapeptide and its effect in facilitating peptide self-association are important for the bioactivity of S2. Additional studies are now needed to further investigate the structure-activity relationships of S1 and S2.

Here we have analysed the solution structures and bioactivity of three novel chimeric analogs of S1 and S2. Firstly, Long S1 (LS1; SGPYGFNSALMFamide) is a dodecapeptide comprising S1 with the addition of the N-terminal four residues of S2 (SGPY). Analysis of this peptide enabled further investigation of the contribution of the N-terminal SGPY tetrapeptide in facilitating peptide self-association and for bioactivity. Secondly, S1(T), in which the penultimate residue of S1 (methionine) is replaced by the residue that occupies this position in S2 (threonine). Thirdly, S2(M), in which the penultimate residue of S2 (threonine) is replaced by the residue that occupies this position in S1

(methionine). Analysis of S1(T) and S2(M) enabled assessment of the contribution of C-terminal amino acid residues for SALMFamide structure and activity.

2. Materials and Methods

2.1 Materials

Peptides were custom-synthesized by the Advanced Biotechnology Centre at Imperial College London and purified using high performance liquid chromatography. All other chemicals used were obtained from VWR (Poole, Dorset, UK) with the exception of D₂O, which was obtained from Goss Scientific Instruments Ltd (Great Baddow, Essex, UK).

2.2 In vitro pharmacology

S1 and S2 cause dose-dependent relaxation of cardiac stomach and tube foot preparations from the starfish *A. rubens*. Therefore, these preparations were used here to assess the bioactivity of three chimeric analogs of S1 and S2 (LS1, S1(T) and S2(M)), employing the same methodology as reported previously [4, 7, 8]. Specimens of *A. rubens* were obtained from the Menai Straits (UK) and maintained in a circulating seawater aquarium in the School of Biological & Chemical Sciences at QMUL. Cardiac stomach and tube foot preparations were dissected, set up in organ baths containing seawater at 11°C and their contractility was measured using isotonic transducers (model 60-3001; Harvard, South Natick, MA, USA) linked to a chart recorder (Goerz Servogor 124). To enable assessment of the bioactivity of the chimeric peptides as muscle relaxants, sustained contracture of preparations was induced and maintained using seawater with 30 mM added KCl, as described previously [4, 7, 8].

LS1, S1(T) and S2(M) were tested on cardiac stomach preparations (n = 4, 8 and 3, respectively) and tube foot preparations (n = 9, 6 and 6, respectively) at three concentrations, 0.1 μM, 1.0 μM and 10 μM, and the effects of the peptides were expressed as a percentage of the relaxing effect of 10 μM S2, which was ascribed a value of 100%. To directly compare the bioactivity of the chimeric peptides with both S1 and S2, experiments were performed where S1, S2 and a chimeric peptide (LS1, S1(T) or S2(M)) were tested at a concentration of 1 μM on cardiac stomach preparations (n = 12, 7 and

10, respectively) and tube foot preparations (n = 13, 12 and 13, respectively). In these experiments, each peptide was tested twice and the order in which peptides were tested was randomised, with effects quantified by normalisation to the effect of 1 μ M S2.

2.3 Circular dichroism (CD) spectroscopy

CD spectra were recorded using a Chirascan CD spectrometer (Applied Photophysics Ltd.) equipped with a Peltier temperature controller. Spectra were the average of three scans recorded with a 1 nm bandwidth, a 0.5 nm step size, and a 5 s time constant. After background subtraction, the observed ellipticity (θ ; mdeg) was converted to a molar ellipticity (ΔE ; $M^{-1} \text{ cm}^{-1}$), using the formula: $\Delta E = \theta / (33000 lc)$ where l is the path length (cm) and c is the concentration (M). To examine the effect of temperature, CD spectra were obtained every 5°C from 10 to 90° using a 1 mm path length for ~ 0.1 mg/ml samples and a 1 cm path length for ~ 0.01 mg/ml samples.

2.4 NMR spectroscopy

Peptides were dissolved in 10% D₂O/90% H₂O to achieve a final concentration of 2 mM. The pH of the solutions was adjusted to pH 5.6 using 10 mM NaOH and 10 mM HCl. The peptide solutions were centrifuged to remove any suspended material and then the supernatant was transferred to 5 mm NMR tubes. Data were acquired using a Bruker Avance 600 MHz spectrometer and collected using Topspin software on a UNIX workstation. All experiments were performed using a 5 mm TXI, triple resonance probe equipped with a z-axis gradient. 2D NMR spectra were obtained at 303 K for LS1 and 283K for S1(T) and S2(M). Water suppression was achieved using a Water Gradient Tailored Excitation (WATERGATE) technique. 2D-Total Correlation Spectroscopy (TOCSY) experiments employed a DIPSI2 sequence for isotropic mixing, with a 65 ms mixing time. A 300 ms mixing time was used for Nuclear Overhauser Effect Spectroscopy (NOESY) and Rotating-frame Overhauser Effect

Spectroscopy (ROESY) experiments. All 2D experiments used STATES-TPPI phase cycling, and a spectral width of 14 ppm was applied in both dimensions with 2048 x 512 complex data points in the t₂ and t₁ dimensions, respectively. Prior to Fourier Transformation (FT), the data were linear predicted in the F1 dimension to 512 real points and then zero filled to produce a final matrix size of 4k x 1k, with a 90° phase-shifted sine squared window function applied to both dimensions. The chemical shifts were referenced to water at 4.7 ppm at 303K, pH 5.6. Spin-systems were manually assigned using the TOCSY data and sequential assignments were completed using the 300 ms NOESY or ROESY spectra.

3. Results

3.1 In vitro pharmacology

3.1.1 LongS1 (LS1)

An N-terminally extended analog of S1 (SGPYGFNSALMFamide or LS1) with the addition of the N-terminal tetrapeptide sequence (SGPY) from S2 was synthesized to investigate if addition of the N-terminal region would confer on S1 bioactivity comparable to that of S2. LS1 was tested on cardiac stomach and tube foot preparations at concentrations ranging from 0.1 μM to 10 μM , with the effect at each concentration expressed as a percentage of the effect of 10 μM S2. LS1 caused dose-dependent relaxation of cardiac stomach preparations, with 0.1 μM LS1 causing $38 \pm 10\%$ relaxation and 1 μM causing $67 \pm 8\%$ relaxation. At 10 μM , the effect of LS1 ($69 \pm 18\%$) was not significantly different to the effect of 1 μM LS1 ($p = 0.74$; t-test) (Fig. 2A). LS1 also caused dose-dependent relaxation of tube foot preparations. However, compared to the effects of LS1 on cardiac stomach, LS1 was less effective in causing relaxation of tube feet, with $14 \pm 5\%$ relaxation at 0.1 μM , $20 \pm 4\%$ relaxation at 1 μM and $37 \pm 8\%$ relaxation (Fig. 2C).

To compare the bioactivity of LS1 with S1 and S2, the effects of all three peptides on cardiac stomach (Fig. 2B) and tube foot (Fig. 2D) preparations were compared at a concentration of 1 μM , expressed as a percentage of the effect of 1 μM S2. The mean effect of LS1 on cardiac stomach preparations at 1 μM ($70 \pm 5\%$ relaxation) was significantly less than the mean effect of 1 μM S2 ($p = 4.65 \text{ E-}06$; t-test), but was not significantly different to the mean effect of 1 μM S1 ($p = 0.138$; t-test). The mean effect of LS1 on tube foot preparations at 1 μM ($21 \pm 5\%$ relaxation) was significantly less than the mean effect of 1 μM S2 ($p = 9.04 \text{ E-}10$; t-test) but was not significantly different to the mean effect of 1 μM S1 ($p = 0.14$; t-test).

Collectively, these data indicate that addition of the SGPY tetrapeptide sequence to the N-terminus of S1 does not confer on S1 bioactivity equivalent to that of S2. Furthermore, because the efficacy of LS1 as a muscle relaxant was found not to be significantly different to that of S1, differences in the potency/efficacy of S1 and S2 appear to be largely determined by differences in their C-terminal regions. To test this hypothesis, C-terminal analogs of S1 and S2 were synthesized and tested.

3.1.2 S1(T)

An analog of S1, GFNSALTFamide or S1(T), was synthesized wherein the residue in S1 that is penultimate to the C-terminal amide (methionine) was replaced with the residue that is in an equivalent position in S2 (threonine). S1(T) caused dose-dependent relaxation of cardiac stomach preparations, with 0.1 μM causing $30 \pm 6\%$ relaxation, 1 μM producing $52 \pm 7\%$ relaxation and 10 μM producing $95 \pm 5\%$ relaxation, in comparison with 10 μM S2 (Fig. 2E). S1(T) also caused dose-dependent relaxation of tube foot preparations, with 0.1 μM S1(T) causing $1 \pm 0.2\%$ relaxation, 1 μM S1(T) causing $10 \pm 3\%$ relaxation and 10 μM produced $83 \pm 16\%$ relaxation, in comparison with 10 μM S2 (Fig. 2G).

Comparison of the effects of S1(T), S1 and S2 on cardiac stomach preparations at 1 μM revealed that the effect of S1(T) at 1 μM ($68 \pm 8\%$ relaxation) was significantly less than the effect of 1 μM S2 ($p = 0.006$; t-test) and was not significantly different to the effect of 1 μM S1 ($p = 0.4$; t-test) (Fig. 2F). Comparison of the effects of S1(T), S1 and S2 on tube foot preparations at 1 μM revealed that the effect of S1(T) at 1 μM ($21 \pm 3\%$ relaxation) was significantly less than the effect of 1 μM S2 ($p = 7.856\text{E-}11$; t-test) but was not significantly different to the effect of S1 at 1 μM ($p = 0.1$; t-test) (Fig. 2H).

Thus, collectively the results of the experiments with S1(T) indicate that substitution of the C-terminal methionine residue in S1 with the equivalently positioned residue in S2 is not sufficient to confer on S1 the potency/efficacy of S2.

3.1.3 S2(M)

An analog of S2, SGPYSFN SGLMFamide or S2(M), was synthesized wherein the residue in S2 that is penultimate to the C-terminal amide (threonine) was replaced with the residue that is in an equivalent position in S1 (methionine). S2(M) caused dose-dependent relaxation of cardiac stomach preparations, with 0.1 μM S2(M) causing $18 \pm 8\%$ relaxation, 1 μM S2(M) causing $43 \pm 6\%$ relaxation and 10 μM causing $99 \pm 10\%$ relaxation, in comparison with the effect of 10 μM S2 (Fig. 2I). S2(M) also caused dose-dependent relaxation of tube foot preparations, with 0.1 μM S2(M) causing $5 \pm 4\%$ relaxation, 1 μM causing $15 \pm 4\%$ relaxation and 10 μM causing $40 \pm 10\%$ relaxation, in comparison with the effect of 10 μM S2 (Fig. 2K).

Comparison of the effects of S2(M), S1 and S2 at 1 μM on cardiac stomach preparations revealed that the effect of S2(M) was significantly less than the effect of 1 μM S2 ($47 \pm 5\%$ relaxation; $p = 1.648\text{E-}8$; t-test) but not significantly different to the effect of 1 μM S1 ($p = 0.012$; t-test) (Fig. 2J). Likewise, comparison of the effect of S2(M), S1 and S2 at 1 μM on tube foot preparations revealed that the effect of S2(M) was significantly less than the effect of S2 ($13 \pm 3\%$ relaxation; $p = 8.999\text{E-}12$; t-test) but not significantly different to the effect of S1 at 1 μM ($p = 0.241$; t-test) (Fig. 2L).

Collectively, these data indicate the C-terminally located threonine residue in S2 contributes to its superior potency/efficacy in comparison with S1 because substitution of this residue with the equivalently positioned residue from S1 (methionine) leads to a loss of bioactivity such that the effects of the S2 analog S2(M) are not significantly different to the effects of S1.

3.2 Structural analysis of LS1, S1(T) and S2(M) using NMR

3.2.1 Long S1 (LS1)

Previously we determined that loss of the four N-terminal amino acids from S2 resulted in a loss of concentration dependent structuring and loss of bioactivity [7]. However, addition of SGPY to the N-terminus of S1 does not confer S2-like bioactivity (Fig. 2). Therefore, removal of the N-terminal tetrapeptide from S2 (short S2 or SS2) results in loss of structure, but addition of the SGPY motif to S1 may not be sufficient to generate structure. To investigate this we obtained two-dimensional TOCSY and NOESY NMR spectra of LS1. From the 2D TOCSY we were able to assign all residues except Ser₁. When the ROESY spectrum of S1 (Fig. 3A) is compared to that of LS1 (Fig. 3B) it is clear that LS1 has far more NOE peaks than S1. The greater number of NOEs in LS1 (Fig. 3B) suggests that the addition of four N-terminal amino acids has resulted in the generation of some structure in LS1. However, the presence of substantial structure would result in a number of long-range NOE contacts. Inspection of the H_N-H_α region of the overlaid LS1 TOCSY and NOESY spectra shows that there are mostly only sequential connections apparent (Fig. 3C). Only one medium range NOE, a d_{αN}(*i*, *i*+2), is apparent between Leu₁₀ and Phe₁₂, however this may be an ambiguous assignment and will be discussed later. Also observed is a d_{NN}(*i*, *i*+2) between Ala₉ and Leu₁₁ as well as sequential d_{NN}(*i*, *i*+1) connections between Gly₅ and Phe₆ and Leu₁₁ and Phe₁₂ (not shown). This suggests that most of the apparent structure is located in the C-terminal region. This result is in contrast to S2, in which many medium and long range NOEs covering the entire peptide were apparent in the H_N-H_α region [7]. Overall, in the LS1 NOESY spectrum we were able to identify 77 NOEs compared to the more than 220 NOEs observed in S2 [7]. Thus, the N-terminal tetrapeptide can induce some conformational preference, but the structure is not as extensive as observed in S2 and in itself is not sufficient to promote S2-like bioactivity.

3.2.2 S1(T)

S1(T) contains a threonine residue in the penultimate position, as in S2, instead of the methionine residue present in S1. Other than the four N-terminal residues, S1 differs from S2 only at this position and at position five (S1 numbering) which is a glycine in S2 and an alanine in S1. We hypothesised that the threonine for methionine substitution may result in increased potency/efficacy but the pharmacological tests showed that bioactivity of S1(T) was not significantly different to S1 (Fig. 2E – H). Analysis of the NMR data supports this finding. The ROESY spectrum of S1(T) (Fig. 4A) is similar to that of S1 (Fig. 4B). The chemical shifts are all near their ‘random coil’ values and there are no ROEs present beyond sequential connectivities, indicating that S1(T), like S1, is largely unstructured. If the penultimate threonine is indeed important in receptor binding, perhaps by being involved in hydrogen-bonding via the side-chain hydroxyl group, then it must require the structuring induced by the N-terminal tetrapeptide in S2 in order to be placed in the correct position.

3.2.3 S2(M)

If Thr₁₁ is important for S2 bioactivity, then S2(M), which contains the penultimate methionine from S1 instead of Thr₁₁, should have reduced bioactivity. The pharmacology results have supported this, showing that S2(M) is no more effective than S1 as a muscle relaxant (Fig. 2I-L). Comparison of the H_N-H_α region of the ROESY spectrum of S2(M) (Fig. 5A) with the same region of S2 (Fig. 5B) shows that S2(M) also lacks the array of ROEs indicative of structure that are abundant in S2. Likely, a loss of structure has resulted in the lack of ROEs. Indeed, the number of ROEs in S2(M) is similar to LS1 as might be expected given that S2(M) and LS1 differ only by one amino acid at position 9 – alanine in LS1 and glycine in S2(M). Closer inspection of the H_N-H_α region of the overlaid TOCSY and ROESY spectra in Fig. 5C shows that S2(M) adopts multiple conformations. All long peptides (S2, LS1, S2(M)) that contain Pro₃ have *cis* and *trans* conformations due to the proline, however S2(M) also

has an additional two conformations around Met₁₁. Figure 5C shows that the sequential H_N-H_α connections can follow two divergent paths from Gly₉ and consequently Leu₁₀, Met₁₁ and Phe₁₂ have two distinct chemical shifts. This suggests that S2(M) can adopt at least four distinct conformations – Pro₃*Trans*Met_{11a}, Pro₃*Trans*Met_{11b}; Pro₃*Cis*Met_{11a} and Pro₃*Cis*Met_{11b}. Possibly, the splitting of the C-terminal residues is a consequence of the *cis/trans* Pro₃ isomerism and could suggest association of the N-terminal region with the C-terminus either within a folded monomer or possibly in a ‘head-to-tail’ dimer. However, it is notable that the *cis*-Proline is a minor component while the signals for the two sets of Met₁₁ have similar intensities. If the two conformations of Leu₁₀, Met₁₁ and Phe₁₂ were a result of the Pro₃ *cis/trans* isomerisation then it could reasonably be expected that the peak intensities would mimic the Pro₃ intensities, i.e. one set would have markedly reduced intensity. It is likely then that the splitting of Leu₁₀ – Phe₁₂ residues is not related to the *cis/trans* Pro₃ isomerism. The ambiguity in the assignment of the d_{αN}(*i*, *i*+2) connection between Leu₁₀ and Phe₁₂ in LS1 is because there may also be some minor doubling of the Met₁₁ peak in LS1, however any chemical shift difference is minor and the TOCSY does not provide sufficient evidence for a second peak. The clear presence of multiple conformations in the C-terminal region of S2(M) would undoubtedly impact on the ability of the peptide to bind to receptors and induce muscle relaxation. Taken together, the pharmacology and structural data indicate that the combination of the four N-terminal residues along with the presence of threonine in the penultimate position is necessary to form a peptide with optimal bioactivity. Possibly, the side-chain hydroxyl is required for hydrogen bonding with the S2 receptor.

3.3 Structural analysis of LS1 and S2(M) using CD spectroscopy

The number of ROEs observed in the ROESY spectra of LS1 and S2(M) (Fig. 3, 5) has suggested that both these peptides contain some structure. Previous studies have shown that structure in S2 was a consequence of self-association [7], and therefore we now wondered if LS1 and S2(M) also

exhibit S2-like concentration-dependent structuring. CD is a useful technique for the determination of protein and peptide structure. In the UV region of the CD spectrum the presence of stable structure results in optical activity due to the chirality around the amide bond. Initially, we obtained CD spectra of LS1 and S2(M) at low concentrations similar to those used in the pharmacology experiments and higher concentrations similar to the concentrations used in the NMR experiments (Fig. 6). For LS1 (Fig. 6A) the CD spectra obtained at both concentrations are very similar. At the lower concentration (0.011 mg/mL) the presence of a strong negative peak near 200 nm along with a positive peak at 225 nm suggests the presence of polyproline-II (PP-II) character [13]. PP-II is a highly extended conformation that does not contain backbone hydrogen bonds and is frequently found in ‘unstructured’ regions of proteins. At a ten-fold higher concentration, the UV CD spectrum of LS1 is similar to that at the lower concentration, the negative peak at 200 nm has got stronger (i.e. more negative) but a small peak at 225 nm is retained. This suggests that a similar structure is adopted at both concentrations.

The CD spectra of S2(M) at high and low concentrations (Fig. 6B) are very different to LS1 (Fig. 6A) and to S2 [7]. The spectra have a positive peak at 207 nm and negative peaks at 215 nm and 225 nm. In the higher concentration spectrum the peak at 215 nm that is readily apparent at the lower concentration appears as a shoulder on the more intense 225 nm peak. The S2(M) spectra are quite novel and don't resemble the spectra that would be obtained from any regular secondary structural element nor an unstructured peptide, which would feature a single negative band at ~198 nm. Furthermore, the lack of similarity to LS1 suggests that S2(M) also does not adopt a PPII conformation. Given that the NMR data suggest that S2(M) can adopt at least four different conformations, the CD spectrum at the higher concentration is likely a summation of all of these different conformations. The fact that the CD spectrum of the lower concentration does not resemble the spectrum of an unstructured peptide indicates that, in contrast to S2, S2(M) retains some structure at the concentrations used in the pharmacology experiments.

To further investigate the structure of LS1 and S2(M) we undertook temperature dependent unfolding experiments. As the temperature is increased from 10°C to 90°C the CD spectra of LS1 shows a loss of the positive band at 225 nm at both low (Fig. 7A) and high (Fig. 7B) concentrations. These data suggest that the PPII structure is lost at higher temperatures; it is notable that even at 90°C, and at both concentrations, a negative CD band is observed at 230 nm suggesting some self-association. As expected, the loss of PPII structure is not cooperative unfolding at either concentration (Fig. 7C, D). The rather unusual CD spectra of S2(M) are very different to those of LS1 - at the higher concentration (Fig. 8A) the negative CD signal at 225 nm remains almost unaffected by raising the temperature to 70°C, while there is a significant loss of positive signal at 207 nm. At the lower S2(M) concentration (Fig. 8B) the signal loss at 207 nm is more marked, accompanied by a negative band at 217 nm even at 70°C, which suggests an extended conformation due to self-association. It is notable that at 70°C with both concentrations of S2(M) the spectra do not show evidence for a completely unfolded peptide. The thermal denaturation curves (Fig. 8C, D) suggest a sharp transition in structure at 50°C with a marked loss of the positive CD band at 207 nm.

4. Discussion

The SALMFamide neuropeptides S1 and S2, which share 67% sequence identity (Fig. 1), act as muscle relaxants in starfish, but S2 is about ten times more potent than S1. In a previous study [7] a structural basis for this difference in the potency of S1 and S2 was investigated, focusing on the N-terminal SGPY tetrapeptide that is a feature of the S2 dodecapeptide (SGPYSFNSGLTFamide) but not the S1 octapeptide (GFNSALMFamide). It was found that the N-terminal SGPY motif confers on S2 the property of self-association induced structure at high concentrations. Given that S2 structuring is only apparent at high concentrations it is most likely not relevant to the physiological roles of S2 as a muscle relaxant in starfish. **We note, though, that recent studies with bradykinin peptides have shown that conformation can be quite dependent on solution composition and it remains possible that other conditions will modify S2's oligomerization and conformation [14].** We have proposed that **the ability of S2 to oligomerize** may have implications for peptide packing in vesicles prior to release and possibly also for autocrine actions of the peptide directly after release when local concentrations are high. Analysis of the bioactivity of an N-terminally truncated form of S2 (SS2; SFNSGLTFamide), which like S1 does not exhibit self-association induced structure, yielded ambiguous findings - the bioactivity of SS2 was found to be either inferior or superior to that of S2, depending on the dose and muscle preparation tested. Thus, the contribution of the N-terminal SGPY sequence to the bioactivity of S2 remains unclear. Here we have further investigated this issue by analysis of the bioactivity and structural properties of chimeric analogs of S1 and S2.

Long S1 or LS1 (SGPYGFNSALMFamide), a chimeric peptide comprising the N-terminal tetrapeptide of S2 and the C-terminal octapeptide of S1, was tested and analysed to investigate if addition of the N-terminal SGPY sequence would confer S2-like bioactivity and structure on S1. It was found that the bioactivity of LS1 as a muscle relaxant was not significantly different to that of S1 and

that LS1 does not exhibit self-association induced structure. However, some structuring was detected in the C-terminal region of LS1 and the presence of a positive peak at 225 nm in CD spectra suggests the presence of a polyproline-II-type conformation. Thus, addition of the N-terminal SGPY motif does confer some structure on S1, but it is not equivalent to that seen with S2 and does not appear to enhance bioactivity. Collectively, these findings indicate that the superior bioactivity of S2 in comparison with S1 is not largely attributable to the presence of the N-terminal SGPY motif in S2. Therefore, we investigated the contribution of residues that differ in the C-terminal region of S1 and S2.

The bioactivity of S1(T), an analog of S1 with a threonine residue replacing the methionine residue that is penultimate to the C-terminal amide in S1, was found not to be significantly different to that of S1. Furthermore, like S1 [7], the S1(T) octapeptide did not exhibit a stable structure in solution. Thus, conversion of the C-terminal region of S1 (LMFamide) to an S2-type C-terminal sequence (LTFamide) is not sufficient on its own to confer S2-like bioactivity on S1. However, conversion of the C-terminal region of S2 (LTFamide) to an S1-type C-terminal sequence (LMFamide) results in loss of S2-type bioactivity – thus, the relaxing effects of the S2 analog S2(M) are significantly less than the effects of S2 and not significantly different to the effects of S1. Interestingly, this amino acid conversion also results in loss of the self-associating property of S2 to form a stable conformation at high concentrations. However, like the LS1 peptide, S2(M) does exhibit some structuring but it lacks the stability seen with S2 at high concentrations and appears to adopt multiple conformations.

Collectively, the findings of our previous study [7] and the findings from the experiments reported here indicate that the superior bioactivity of S2 compared to S1 is largely attributable to differences in the C-terminal region of these peptides. In particular, the loss of bioactivity seen in the S2 analog S2(M) indicates that the C-terminal LTFamide motif of S2 appears to be a key determinant of its superior bioactivity in comparison with S1. This presumably reflects differences in the properties of threonine and methionine and hence how they may interact with residues that form receptor binding

sites for S2 - perhaps the side-chain hydroxyl group in the threonine residue is required for formation of hydrogen bonds. The presence of multiple conformations around Met₁₁ in S2(M) suggests that, in addition to Thr₁₁ potentially supporting interactions with the S2 receptor, it is also possible that this residue supports intra- or inter-molecular hydrogen-bonding. However, the contribution of the N-terminal SGPY sequence to the bioactivity of S2 should not be dismissed. The reduced bioactivity of the N-terminally truncated form of S2 (SS2) seen in some preparations [7] indicates that the SGPY motif contributes to S2's bioactivity. The SGPY motif clearly confers varying degrees of structural stability on the dodecapeptides S2, LS1 and S2(M) and we propose that it is this property in combination with the C-terminal residues of S2 that accounts for S2's superior bioactivity in comparison with S1. Further understanding of S2's bioactivity might be achieved if we could solve the structure of the high-concentration multimeric structure. However, solving the structure of monomeric S2 based on our NOESY data is very difficult because at NMR concentrations S2 is multimeric and lowering the concentration results in loss of structure. Under NMR conditions deconvoluting our NOE data to determine intra-monomer from inter-monomer contacts is not possible, especially for a peptide of only 12 amino-acids. Furthermore, it is difficult to assess how the multimeric structure has influenced NOE intensities and thus conversion of peak intensity to distant restraints is likely to result in incorrect distances. For example, a 'head-to-head' dimer could result in NOE contacts for the same proton pair that is a mixture of both intra- and inter-residue contacts. Diffusion Ordered Spectroscopy of S2 was undertaken to try to elucidate the arrangement of the oligomeric structure (data not shown), but given the small size of the peptide it is very difficult to deconvolute the diffusion coefficient to differentiate an unfolded monomer from a folded dimer, trimer or tetramer.

In conclusion, our analysis of chimeric SALMFamide neuropeptides has revealed how neuropeptide bioactivity is determined by a complex interplay of sequence and conformation. Further insights into the structural basis of the differences in the potency of S1 and S2 as muscle relaxants in starfish could be obtained if the receptor(s) that mediate the effects of these peptides is/are identified.

Based on sequence similarities that SALMFamide-type neuropeptides share with neuropeptides that have been identified in other phyla, candidate G-protein coupled receptors for SALMFamides have been proposed (e.g. GnIH/NPFF-type and QRFP-type receptors; [1, 15]). However, definitive insights will require cloning, heterologous expression and pharmacological characterisation of candidate receptors.

Another fascinating avenue for future research will be to extend our analysis of the bioactivity and structural properties of S1, S2 and their chimeric analogs to other SALMFamide-type neuropeptides that have been discovered in starfish and in other echinoderms. Analysis of the genome sequence of the starfish *Patiria miniata* has revealed that S1 is derived from a precursor protein that contains six other L-type SALMFamides. On the other hand, the S2-like neuropeptide in *P. miniata* (SNGPYSMSGLRSLTFamide) is derived from a different precursor protein, which also comprises eight F-type or F-type-like SALMFamides [16]. Recently, sequencing of the neural transcriptome of *A. rubens* has enabled identification of corresponding precursors in this species, with S1 derived from a precursor comprising six L-type SALMFamides and S2 derived from a precursor comprising seven F-type or F-type-like SALMFamides (D.C. Semmens & M.R. Elphick, unpublished data).

Immunocytochemical localisation of S1 and S2 in *A. rubens* has revealed that these two peptides, and hence their precursor-associated peptides, are expressed in different populations of neurons [6, 17]. Therefore, we can collectively infer from these observations that S1 is released from neurons as one constituent of a “cocktail” of seven L-type SALMFamides. On the other hand, it appears that S2 is released by a different population of neurons as one constituent of a “cocktail” of eight SALMFamides, with the seven other peptides being F-type or F-type-like SALMFamides. Therefore, in comparing the bioactivity and structural properties of S1 and S2, as reported previously [7] and here, we are in fact comparing two L-type SALMFamides that are derived from different precursors and different populations of neurons. Furthermore, because both S1 and S2 are derived from precursor proteins that contain other SALMFamides, testing and comparing the effects of S1 and S2 *in vitro* is not

representative of a physiological scenario. In future studies it will be of interest to compare the structural and functional properties of S1 with other L-type SALMFamides that are derived from the S1 precursor. Furthermore, analysis of the *in vitro* actions of a “cocktail” of the SALMFamides that are derived from the S1 precursor will reveal pharmacological actions that are representative of physiological processes *in vivo*. Similarly, it will be interesting to compare the effects of S2 with the effects of F-type and F-type-like SALMFamides that are derived from the S2 precursor and to compare the effects of peptides tested individually with the effects of the entire “cocktail” of SALMFamides that are derived from the S2 precursor. Combined with comparative analysis of the structural properties of all fifteen SALMFamides that are derived from the S1 and S2 precursors, it may then be possible to gain insights into the evolutionary and functional significance of the heterogeneous mixtures of SALMFamides that occur in starfish and in other echinoderms.

Our work on SALMFamides has broad relevance because the occurrence of multiple related neuropeptides derived from a single precursor protein is a common phenomenon, particularly in invertebrates [18]. However, the functional significance of these neuropeptide “cocktails” is not understood [19]. As an example, in the nematode *C. elegans*, a single precursor protein generates eight similar peptides belonging to the FMRFamide family of neuropeptides. Deletion of this gene results in five distinct and different behavioural defects but it is not clear how individual peptides contribute to a specific phenotype [20]. Similar to our findings in this work, it appears that conformation preferences of individual FMRFamide-type neuropeptides in molluscs play a role in modulating receptor binding. It is suggested that only a single conformation will bind to the receptor and therefore the effective neuropeptide concentration is less than the expressed concentration [21]. Thus, an understanding of the interplay of sequence, conformation and intra- and inter-peptide interactions may be the key to elucidation of the functional significance of neuropeptide “cocktails”.

Acknowledgements - We are grateful to Harold Toms (QMUL) for technical assistance with the NMR studies and Richard Melarange (QMUL) for technical assistance with *in vitro* pharmacology.

References

- [1] M.R. Elphick, SALMFamide salmagundi: The biology of a neuropeptide family in echinoderms, *Gen Comp Endocrinol*, (2014).
- [2] M.R. Elphick, R. Melarange, Neural control of muscle relaxation in echinoderms, *J Exp Biol*, 204 (2001) 875-885.
- [3] M.R. Elphick, D.A. Price, T.D. Lee, M.C. Thorndyke, The SALMFamides: a new family of neuropeptides isolated from an echinoderm, *Proc Biol Sci*, 243 (1991) 121-127.
- [4] R. Melarange, D.J. Potton, M.C. Thorndyke, M.R. Elphick, SALMFamide neuropeptides cause relaxation and eversion of the cardiac stomach in starfish., *Proc Biol Sci*, 266 (1999) 1785-1789.
- [5] M.R. Elphick, S.J. Newman, M.C. Thorndyke, Distribution and action of SALMFamide neuropeptides in the starfish *Asterias rubens*, *J Exp Biol*, 198 (1995) 2519-2525.
- [6] S.J. Newman, M.R. Elphick, M.C. Thorndyke, Tissue distribution of the SALMFamide neuropeptides S1 and S2 in the starfish *Asterias rubens* using novel monoclonal and polyclonal antibodies. 2. Digestive system, *Proc Biol Sci*, 261 (1995) 187-192.
- [7] C.B. Obara, C.E. Jones, N.D. Younan, J.H. Viles, M.R. Elphick, Structural analysis of the starfish SALMFamide neuropeptides S1 and S2: The N-terminal region of S2 facilitates self-association, *Bba-Proteins Proteom*, 1844 (2014) 358-365.
- [8] R. Melarange, M.R. Elphick, Comparative analysis of nitric oxide and SALMFamide neuropeptides as general muscle relaxants in starfish, *J Exp Biol*, 206 (2003) 893-899.
- [9] D.A. Keire, T.E. Solomon, J.R. Reeve, Jr., NMR evidence for different conformations of the bioactive region of rat CCK-8 and CCK-58, *Biochem Biophys Res Commun*, 293 (2002) 1014-1020.
- [10] G.Y. Xu, C.M. Deber, Conformations of neurotensin in solution and in membrane environments studied by 2-D NMR spectroscopy, *Int J Pept Protein Res*, 37 (1991) 528-535.
- [11] J.L. Nieto, M. Rico, J. Santoro, J. Herranz, F.J. Bermejo, Assignment and conformation of neurotensin in aqueous solution by ¹H NMR, *Int J Pept Protein Res*, 28 (1986) 315-323.
- [12] G. Chassaing, O. Convert, S. Lavielle, Preferential conformation of substance P in solution, *Eur J Biochem*, 154 (1986) 77-85.
- [13] A.A. Adzhubei, M.J. Sternberg, A.A. Makarov, Polyproline-II helix in proteins: structure and function, *Journal of molecular biology*, 425 (2013) 2100-2132.
- [14] N.A. Pierson, L. Chen, S.J. Valentine, D.H. Russell, D.E. Clemmer, Number of solution states of bradykinin from ion mobility and mass spectrometry measurements, *J Am Chem Soc*, 133 (2011) 13810-13813.
- [15] M.R. Elphick, From gonadotropin-inhibitory hormone to SIFamides: Are echinoderm SALMFamides the "missing link" in a bilaterian family of neuropeptides that regulate reproductive processes?, *Gen Comp Endocrinol*, (2013).
- [16] M.R. Elphick, S. Achhala, N. Martynyuk, The evolution and diversity of SALMFamide neuropeptides, *PLoS One*, 8 (2013) e59076.
- [17] S.J. Newman, M.R. Elphick, M.C. Thorndyke, Tissue distribution of the SALMFamide neuropeptides S1 and S2 in the starfish *Asterias rubens* using novel monoclonal and polyclonal antibodies. I. Nervous and locomotory systems, *Proc Biol Sci*, 261 (1995) 139-145.
- [18] C. Wegener, A. Gorbashov, Molecular evolution of neuropeptides in the genus *Drosophila*, *Genome Biol*, 9 (2008) R131.
- [19] R.S. Hewes, E.C. Snowdeal, 3rd, M. Saitoe, P.H. Taghert, Functional redundancy of FMRFamide-related peptides at the *Drosophila* larval neuromuscular junction, *J Neurosci*, 18 (1998) 7138-7151.
- [20] L.S. Nelson, M.L. Rosoff, C. Li, Disruption of a neuropeptide gene, *flp-1*, causes multiple behavioral defects in *Caenorhabditis elegans*, *Science*, 281 (1998) 1686-1690.
- [21] A.S. Edison, E. Espinoza, C. Zachariah, Conformational ensembles: the role of neuropeptide structures in receptor binding, *J Neurosci*, 19 (1999) 6318-6326.

Figure Legends

Figure 1. Comparison of the sequences of S1, S2, SS2 (short S2), LS1 (long S1), S1(T) and S2(M).

Residues that are identical in all six peptides are shown in black and residues that are variable or are not present in some peptides are shown in red.

Figure 2. Analysis of pharmacological effects of LS1, S1(T) and S2(M) on cardiac stomach and tube foot preparations. A. LS1 causes dose-dependent relaxation of cardiac stomach preparations. However, although the effect 1 μ M LS1 is significantly greater than 0.1 μ M LS1 ($p = 0.02$; t-test), the effect of 1 μ M LS1 is not significantly different to the effect of 10 μ M LS1 ($p = 0.74$; t-test). B. When tested at 1 μ M, LS1 is significantly less effective than S2 ($p = 4.65E-06$; ***, t-test) but is not significantly different to S1 ($p = 0.138$; t-test) as a relaxant of cardiac stomach preparations. C. LS1 causes dose-dependent relaxation of tube foot preparations. D. When tested at 1 μ M, LS1 is significantly less effective than S2 ($p = 9.04E-10$; ***, t-test) but is not significantly different to S1 ($p = 0.14$; t-test) as a relaxant of tube foot preparations. E. S1(T) causes dose-dependent relaxation of cardiac stomach preparations. F. When tested at 1 μ M, S1(T) is significantly less effective than S2 ($p = 0.0055$; ***, t-test) but is not significantly different to S1 ($p = 0.4$; t-test) as a relaxant of cardiac stomach preparations. G. S1(T) causes dose-dependent relaxation of tube foot preparations. H. When tested at 1 μ M, S1(T) is significantly less effective than S2 ($p = 7.856 E-11$; ***, t-test) but is not significantly different to S1 ($p = 0.129$; t-test) as a relaxant of the tube foot preparations. I. S2(M) causes dose-dependent relaxation of cardiac stomach preparations. J. When tested at 1 μ M, S2(M) is significantly less effective than S2 ($p = 1.648E8$; ***, t-test) but is not significantly different to S1 ($p = 0.012$; t-test) as a relaxant of cardiac stomach preparations. K. S2(M) causes dose-dependent relaxation of tube foot preparations. L. When tested at 1 μ M, S2(M) is significantly less effective than S2 ($p = 8.999E-12$;

***; t-test) but is not significantly different to S1 ($p = 0.241$; t-test) as a relaxant of the tube foot preparations.

Figure 3. Analysis by NMR (600 MHz) suggests LS1 contains some structure. The amide region of the ROESY spectrum ($\tau_m = 300$ ms) of S1 (A) shows no non-sequential cross-peaks which correlates with the lack of structure in this peptide. In contrast, the NOESY spectrum ($\tau_m = 300$ ms) of LS1 (B) (2mM, pH 5.6, 303K) shows non-sequential cross-peaks, especially to the side-chains of Phe₆/Phe₁₂ between 7.0 and 7.5 ppm, suggesting that LS1 contains some structure. Overlay of the NH-H _{α} region of the LS1 ROESY spectrum (red) and the TOCSY spectrum ($\tau_m = 65$ ms, blue) shows the sequential connections (solid lines) along with a possible $d_{\alpha N}(i, i+2)$ connection between L10 and F12 (dotted line).

Figure 4. S1T is not structured. The H_N-H _{α} region of 600 MHz ROESY spectrum ($\tau_m = 300$ ms) of S1(T) (2 mM, pH 5.6, 283K) (A) shows only sequential connectivities similarly to S1 (B) and both have a lack of medium and long range contacts that would indicate structure. The box labelled G₁ represents the intraresidue H_N-H _{α} peak of G₁ observed in the TOCSY but not in the ROESY.

Figure 5. S2(M) contains multiple conformations. The amide region of the 600 MHz NOESY spectrum ($\tau_m = 300$ ms) of S2(M) (2 mM, pH 5.4, 283K) (A) compared to the same region of the NOESY spectrum of S2 (B) (C) Overlay of the NH-H _{α} region of the S2(M) NOESY spectrum (red) and TOCSY spectrum (blue) shows sequential connections (solid lines). The dashed line highlights a second set of sequential connections from Leu₁₀ to Phe₁₂ that suggests the presence of an alternative conformation.

Figure 6. CD of LS1 and S2(M) at high and low concentrations. Panel A shows the CD spectra of LS1 at low (0.011 mg/mL, black trace) and high (0.13 mg/mL, red trace) concentrations. The positive peak at 225 nm, particularly at low concentration, suggests the presence of polyproline II conformation. The CD spectra of S2(M) (Panel B) at low (0.012 mg/mL, black trace) and high (0.18 mg/mL red trace) are quite unusual and not typical of regular secondary structure. The presence of the positive band at 207 nm and the negative band at 225 nm in both spectra suggests that the peptide has some similarity at both concentrations.

Figure 7. Thermal melting CD spectra of LS1. Panel A shows the CD spectra of LS1 (0.13 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 90°C (purple trace) in 5°C increments. Panel B shows the CD spectra of LS1 (0.011 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 90°C (purple trace). In both cases a spectrum was obtained at 10°C after cooling from 90°C and the unfolding was reversible. In neither high (Panel A) nor low concentration (Panel B) spectra is there evidence of a completely unfolded peptide, even at 90°C. Panels C and D show the ellipticity at 223 nm plotted as a function of temperature, and in neither case is there evidence of cooperative unfolding.

Figure 8. Thermal melting CD spectra of S2(M). Panel A shows the CD spectra of S2(M) (0.18 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 70°C (blue trace) in 5°C increments. Panel B shows the CD spectra of S2(M) (0.012 mg/mL) obtained as the temperature is raised from 10°C (black trace) to 70°C (purple trace). In neither high (Panel A) nor low concentration (Panel B) spectra is there evidence of a completely unfolded peptide at the highest temperature tested. Panels C and D show the ellipticity at 207 nm, 217 nm and 225 nm plotted as a function of temperature.

Figures 1-8

S1 H-Gly-Phe-Asn-Ser-Ala-Leu-Met-Phe-NH₂
S2 H-Ser-Gly-Pro-Tyr-Ser-Phe-Asn-Ser-Gly-Leu-Thr-Phe-NH₂
SS2 H-Ser-Phe-Asn-Ser-Gly-Leu-Thr-Phe-NH₂
LS1 H-Ser-Gly-Pro-Tyr-Gly-Phe-Asn-Ser-Ala-Leu-Met-Phe-NH₂
S1 (T) H-Gly-Phe-Asn-Ser-Ala-Leu-Thr-Phe-NH₂
S2 (M) H-Ser-Gly-Pro-Tyr-Ser-Phe-Asn-Ser-Gly-Leu-Met-Phe-NH₂

Figure 1.

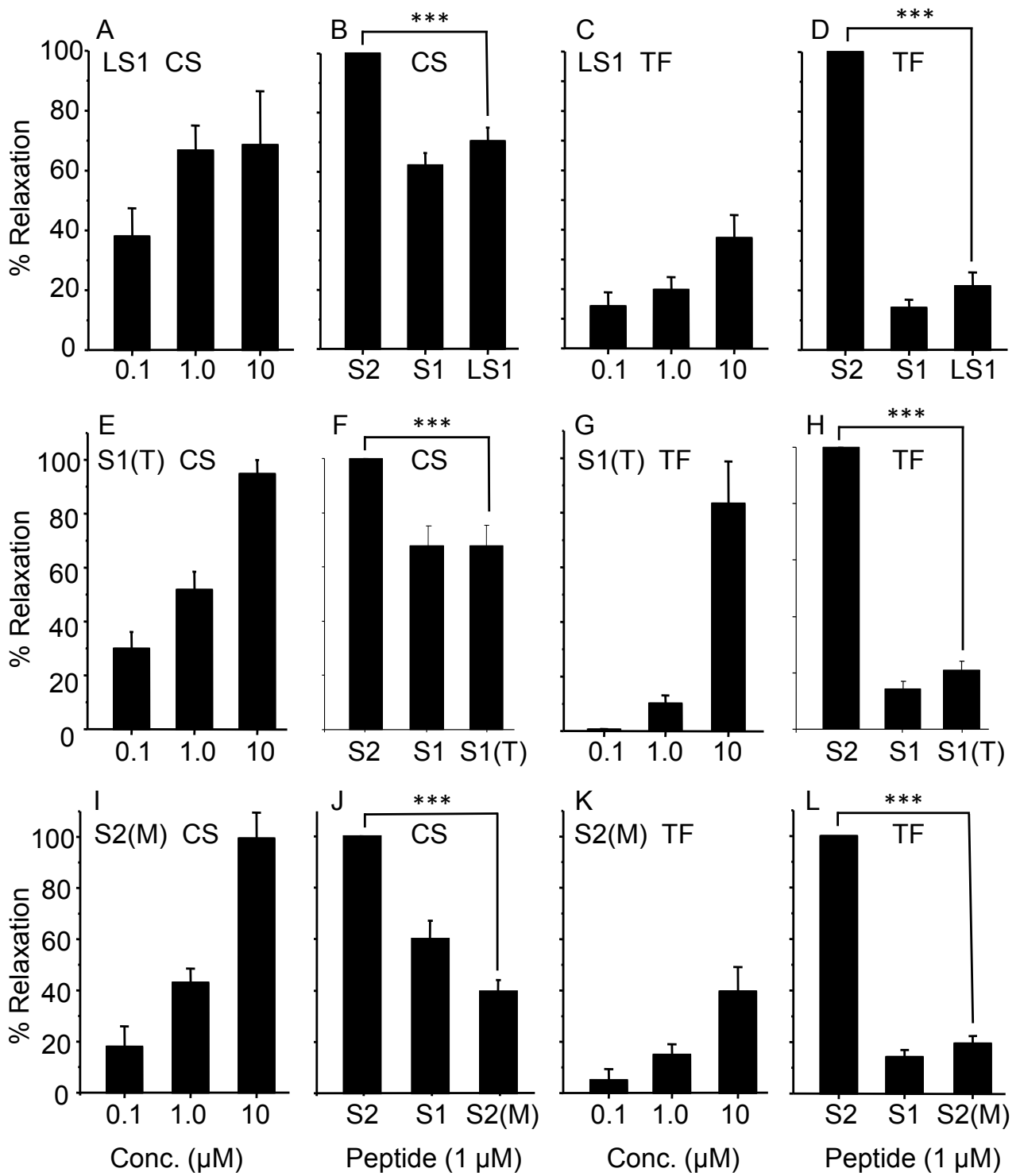


Figure 2.

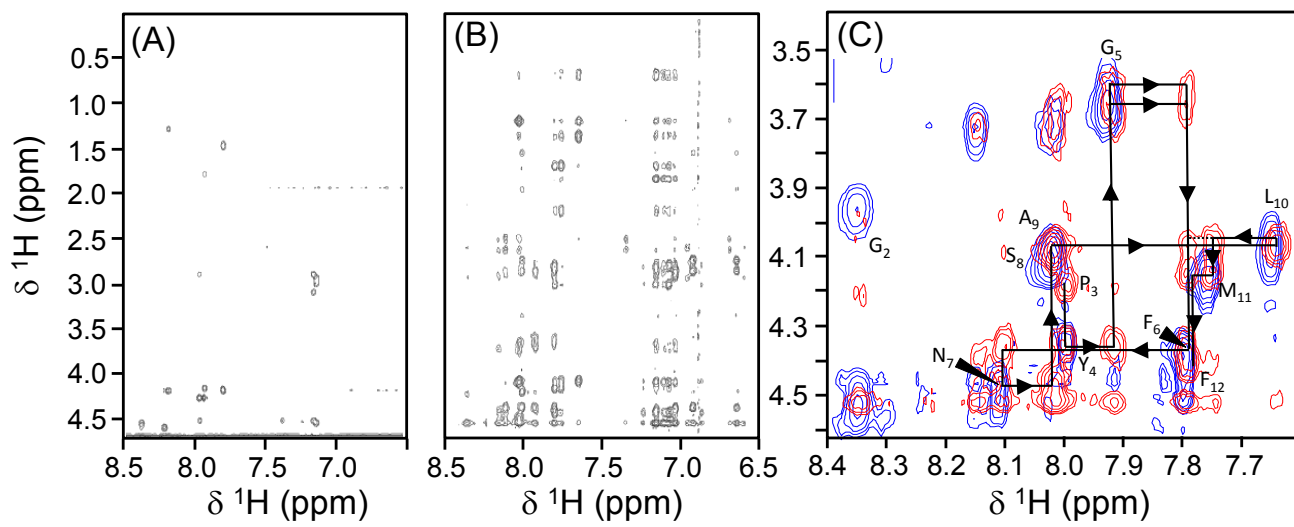


Figure 3.

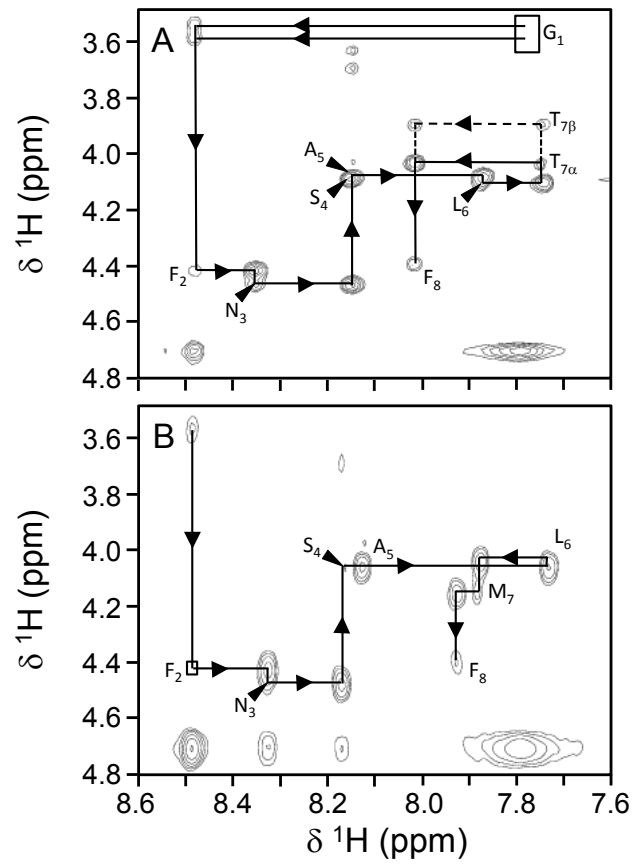


Figure 4.

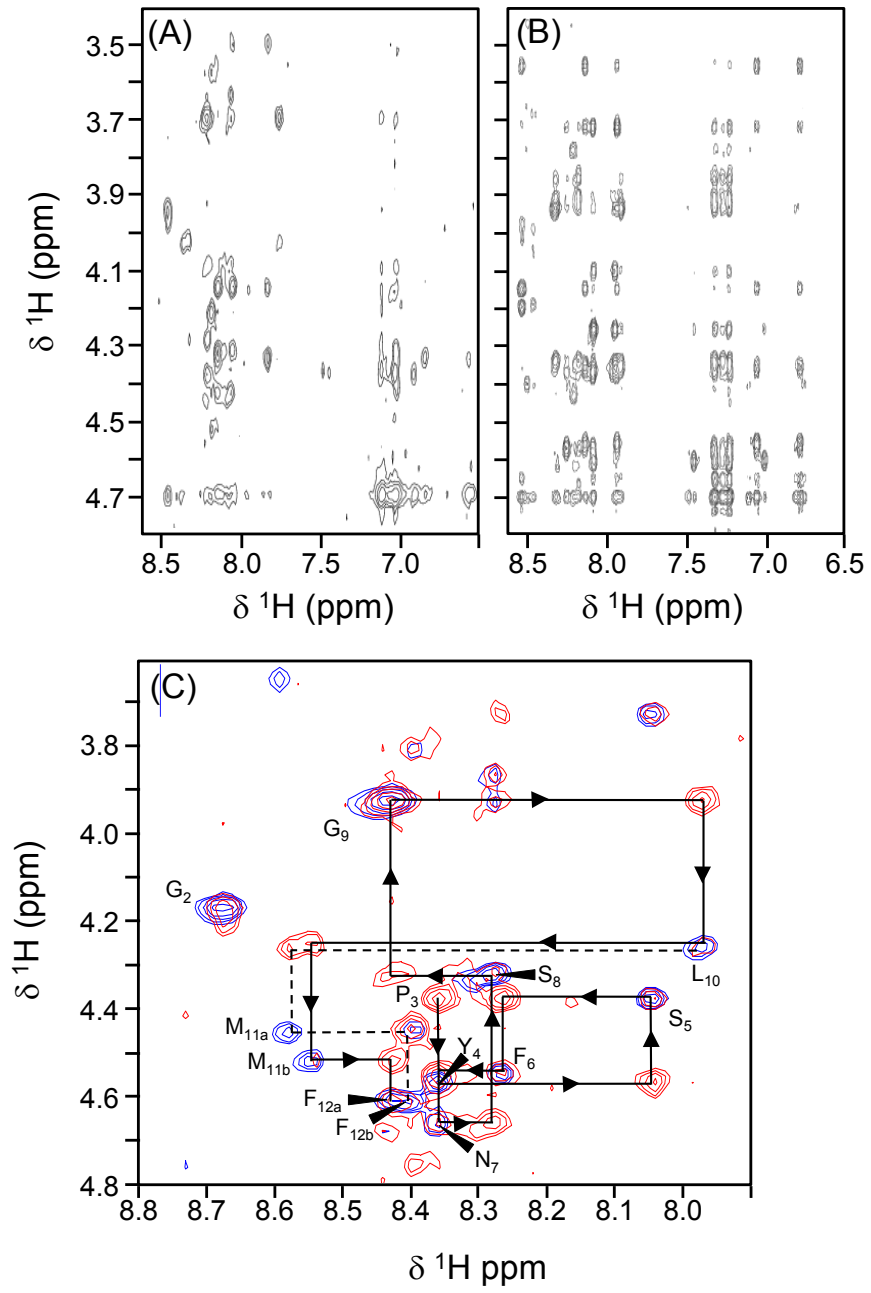


Figure 5.

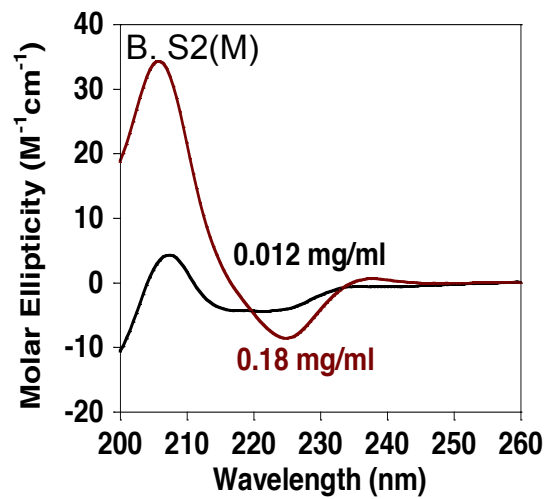
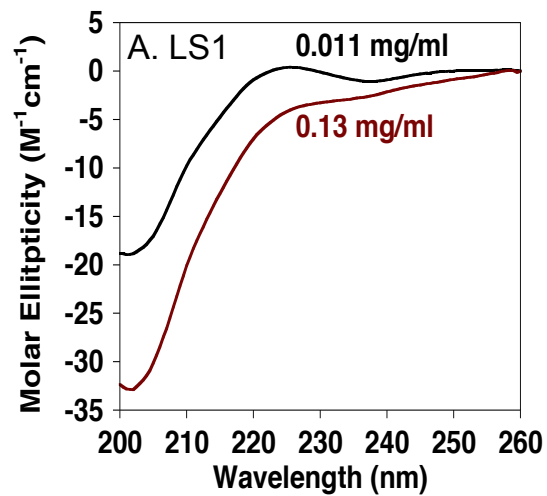


Figure 6.

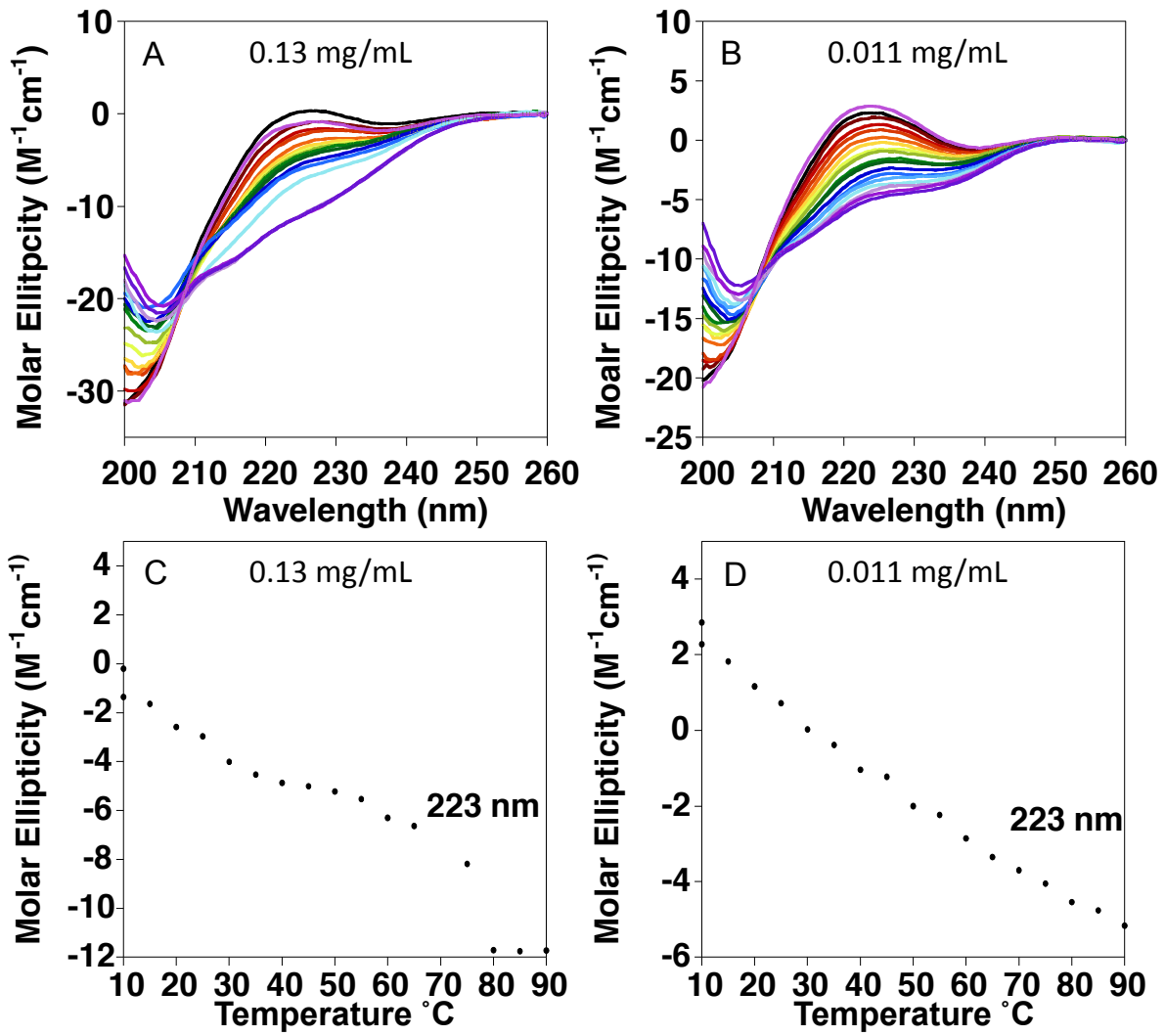


Figure 7.

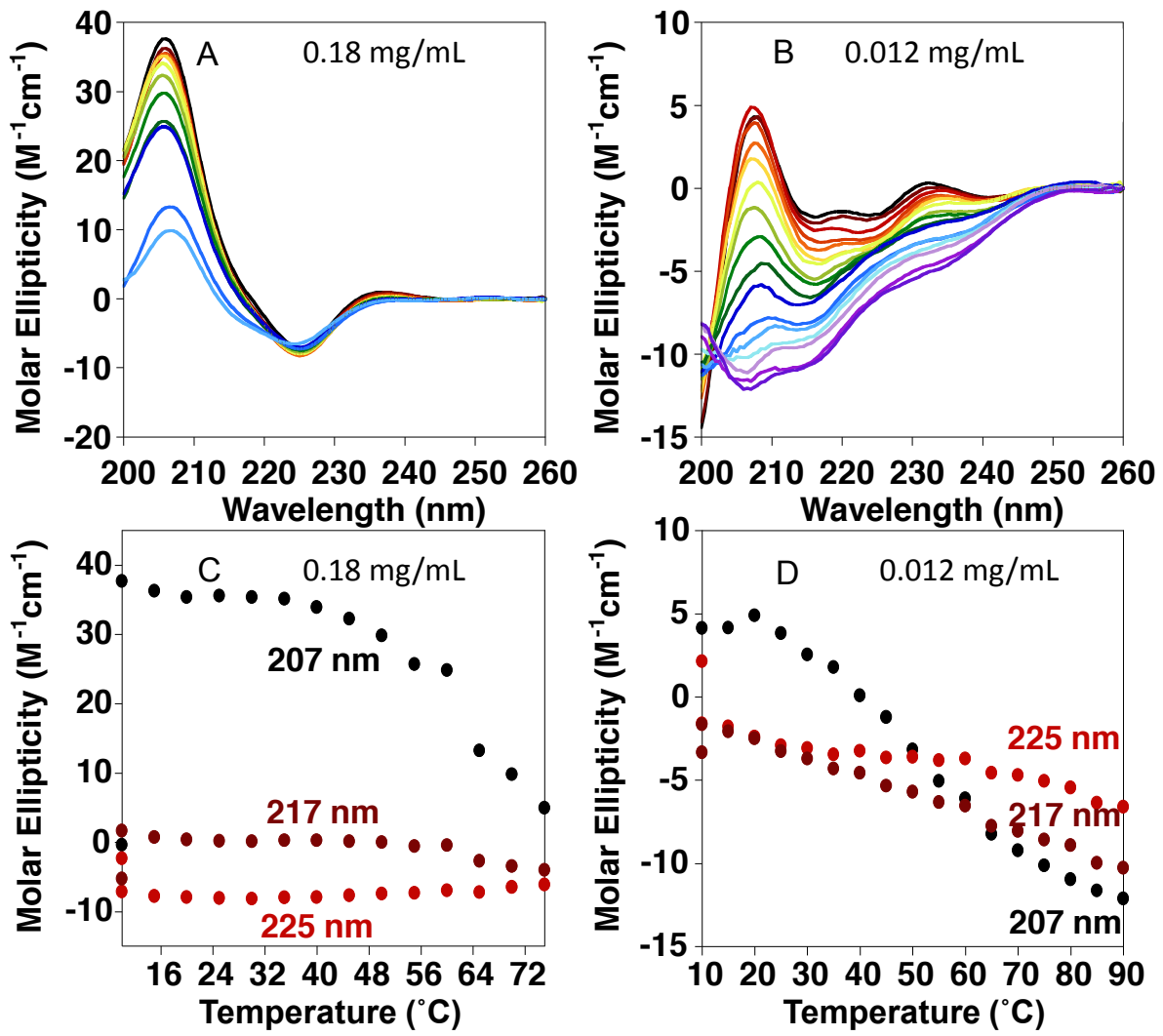


Figure 8.