Interrogation of Intermolecular Interactions in Liquid–liquid Phase Separation

By

Ruiwen Xu

A dissertation submitted in partial fulfillment of

The requirements for the degree of

Doctor of Philosophy

(Chemistry)

At the

UNIVERSITY OF WISCONSIN-MADISON

2025

Date of final oral examination: 07/24/2025

This dissertation is approved by the following members of the Final Oral Committee:

Samuel H. Gellman, Professor, Chemistry
Helen E. Blackwell, Professor, Chemistry
Aaron A. Hoskins, Professor, Biochemistry
Jeffery D. Martell, Assistant Professor, Chemistry

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor, Professor Sam Gellman, for being an exceptional mentor and role model, both in science and in life. Over the past five years, Sam has guided me in becoming an independent researcher and has inspired me to remain passionate and curious. I am especially grateful for his patience and dedication in helping me improve my writing—thank you for correcting my grammar and for sharing so many colorful and memorable English expressions.

I am also deeply thankful to everyone in the Gellman lab. Working with such a collaborative and supportive group has been one of the highlights of my time here. Special thanks to Dr. Jiani Niu, Dr. Zhen Yu, and Dr. Xinyu Liu for their constant support, advice, and encouragement throughout my Ph.D. I would also like to thank our LC team—Philip Lampkin, Dean Jarois, Aevi van der Stok, and Eden Xu—for keeping our HPLC and UPLC systems running smoothly. (Calvin is still my favorite.) I am excited to see the continuation of our work on liquid–liquid phase separation by Aevi van der Stok, Yucheng Jiang, and Rui Wang, and I look forward to the advances they will make in the field. I'm also grateful for all the thoughtful feedback, creative ideas, and lively discussions during group meetings and subgroups. A special thank-you to my undergrad mentee, Jiawen Huang. Your energy and enthusiasm made our time working together really enjoyable. I hope you love your Ph.D. experience at UIUC as much as I've loved mine here.

Finally, I want to thank all of our collaborators (listed below) who helped make this work possible, and the amazing instrument center staff who supported my MALDI and NMR experiments. None of this would have been possible without your help and contributions.

Collaborators:
Prof. Nick Abbott (Cornell University)
Cindy Qiu (Cornell University)
Prof. Yamuna Krishnan (University of Chicago)
Sangyoon Lee (University of Chicago)
Prof. Xuhui Huang (UW-Madison)
Dr. Ilona Chirsty Unarta (UW-Madison)

Gellman Lab Collaborators:
Dr. Zhen Yu
Dr. Ariel J. Kuhn
Yucheng Jiang
Rui Wang

TABLE OF CONTENTS

Chapter 1

Phase Separation Mediated by Tyrosine-Rich and Arginine-Rich Polypeptides Is Promoted by Hydrogen Bonding but Not Cation-π Interactions

Portions of this work have been published for publication in Journal of the American Chemical Society.

Graphics were prepared using Microsoft PowerPoint, GraphPad Prism, ChemDraw, and ImageJ.

1.1 Abstract

Liquid-liquid phase separation (LLPS) plays an important role in many cellular processes. The RNA-binding protein Fused in Sarcoma (FUS) has been shown to mediate phase separation through attractive interactions between the tyrosine-rich N-terminal domain and the arginine-rich C-terminal domain. Cation-π interactions between Arg and Tyr side chains have been proposed as a driving force for condensed phase formation involving FUS and related proteins. To explore this hypothesis, we developed a two-component system comprising a long Arg-rich FUS fragment, generated via heterologous expression, and a short Tyr-rich peptide, generated via chemical synthesis. Phase separation occurs when these components are mixed at low concentrations (5 μM each). Global replacement of Tyr with noncanonical residues allowed us to interrogate the role of the phenolic side chain in driving the polypeptide assembly that underlies condensed phase formation. The results suggest that cation-π interactions do not make a major contribution to inter-peptide attractions in our model system. Instead, the H-bond donor properties of the Tyr side chain hydroxyl appear to play a central role in driving phase separation.

1.2 Introduction

Spontaneous formation of condensed phases mediated by biopolymers is now increasingly recognized as a wide-spread phenomenon within cells[1–8] These processes, which include liquid-liquid phase separation (LLPS), facilitate the formation of highly concentrated compartments in which macromolecular assemblies serve as scaffolds. One defining feature of these condensed phases is their ability to recruit or exclude specific molecules.[8–10] This molecular sorting helps regulate cellular functions: clustering of specific proteins and/or nucleic acids appears to facilitate key physiological processes, such as ribosome assembly, RNA splicing or transcription.[1] The functionality of cellular condensates arises in part from the ease with which molecules can exchange between fluid “membraneless organelles” and the surrounding medium.[2,3] The absence of a defining lipid bilayer enables dynamics in another dimension as well: condensates can form and dissipate rapidly in response to diverse chemical or physical cues, allowing the cell to regulate its internal environment.[4]

Assembly of macromolecules in a manner that allows intra- and intermolecular dynamics requires a delicate balance of noncovalent forces. Formation of the macromolecular network that constitutes the condensate scaffold must be sufficiently favorable in terms of free energy to overcome the loss of translational entropy upon concentration, but interactions among the macromolecules should not be so strong as to cause formation of a gel or solid. Network formation can be stabilized by intrinsic intermolecular attractions, release of ordered water molecules and/or release of counterions. Some biomolecular condensation processes are mediated at least in part by interactions involving two or more polyvalent proteins that contain multiple modules with specific and complementary recognition properties.[1,8] Many condensation phenomena, however, arise from inter-actions involving much smaller molecular fragments, such as individual amino acid residue side chains.[11–16] The attractions that drive this latter form of assembly remain a subject of active inquiry.[10,17] The studies described here were intended to provide insight on these forces.

Efforts to understand the noncovalent forces that mediate formation of membraneless organelles are important because this type of intracellular assembly has been associated with various diseases.[6,7,18] Fundamental insights on the intermolecular interactions that drive LLPS and that influence the evolution of condensed liquid phases toward more ordered phases will lay a foundation for efforts to modulate biomolecular associations for therapeutic purposes.

The protein Fused in Sarcoma (FUS) has received considerable attention in studies of biopolymer-mediated LLPS.[18–20] Condensates scaffolded by FUS facilitate key physiological functions of RNA including translation and splicing. FUS is regarded as representative of a large family of proteins with related functions. Characteristics of these proteins include intrinsic disorder over much of the sequence, and specific domains that are rich in either tyrosine or arginine.[18] In the case of FUS, associations involving the C-terminal Arg-rich domain, and the N-terminal Tyr-rich domain appear to drive LLPS; phase separation is maintained even if these domains are not covalently linked. These associations have been attributed to cation-π interactions between guanidinium groups in Arg side chains and aromatic rings in Tyr side chains;[19,20] additional possible contributions of these side chains have been more recently suggested.[15]

The studies described here test the cation-π hypothesis in the context of phase separation mediated by Arg-rich and Tyr-rich polypeptides. Our experimental system was designed to allow evaluation of noncanonical side chain replacements for the phenol unit in tyrosine. This approach was inspired by incisive studies that have evaluated interactions between specific cation/aromatic pairs in discrete ligand-protein complexes.[21–29] In such cases, genetic code expansion via amber-suppression methods allows biosynthesis of proteins in which a single native aromatic side chain at a specific site is substituted with a non-canonical side chain. For example, Tyr, which bears an electron-rich aromatic ring, can be replaced with a noncanonical residue bearing an electron-poor aromatic ring, such as 3,4,5-trifluorophenylalanine (F_3Phe). An authentic cation-π interaction will be impaired or abolished upon Tyr → F_3Phe modification, but interactions arising from other noncovalent factors, such as hydrophobicity, will persist after this modification.[25–29]

We employed a phase-separation model system with two components, an Arg-rich polypeptide drawn from the C-terminal portion of FUS and a designed peptide that contains seven Tyr residues.

The latter component is accessible via solid-phase synthesis, which facilitated global replacement of Tyr with nonproteinogenic residues. Our findings indicate that cation-π interactions are not a major driving force for phase separation in this system, which contrasts with conclusions regarding LLPS mediated by FUS.[19,20,30] Our data suggest that the H-bond donor properties of the Tyr side chain are critical for phase separation mediated by the Arg-rich and Tyr-rich components. In addition, we find that replacing Tyr with highly hydrophobic side chains can support phase separation in this system, but neither cation-π attraction nor H-bond donation appears to be involved in these cases.

1.3 Results and Discussion

1.3.1 Design of the two-component LLPS

Our goal of using noncanonical side chains to evaluate LLPS required substitutions at multiple sites within a polypeptide, which is not readily accomplished via genetic code expansion methods. Some amino acids with non-canonical side chains similar to those of proteinogenic amino acids can be ribosomally incorporated without genetic code expansion.[30] However, the degree of incorporation can be variable within the mixture of polypeptides that is produced.[31] We therefore sought a two-component system in which phase separation could be induced by combining a relatively short Tyr-rich peptide with a longer Arg-rich polypeptide. Chemical synthesis of the Tyr-rich component would allow access to very pure analogues containing bespoke residues at any number of positions. Increasing polymer chain length generally enhances the propensity to undergo LLPS, and we anticipated that the length accessible for a ribosomally generated Arg-rich component would compensate for the relatively short Tyr-rich peptide to allow LLPS at concentrations in the low µM range, as is observed for FUS itself.[19,20]

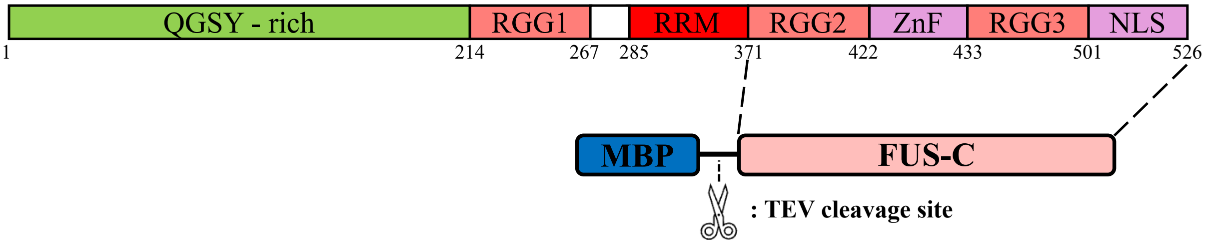


Figure 1.1. Domain structure of the protein FUS (1-526). The Arg-rich protein used in this work is based on residues 372-526 (FUS-C). This fragment was expressed in a construct that included maltose-binding protein, with a TEV protease cleavage site in-between.

A fragment comprising the C-terminal 154 residues of FUS (372-526) was selected as the Arg-rich component; this polypeptide is designated "FUS-C" below (Figure 1.1). FUS-C contains 24 Arg residues and is expected to have a net charge of approximately +13 near neutral pH. The 60 Gly residues in FUS-C should render this polypeptide largely disordered. In FUS itself, residues 422-453 constitute a zinc finger domain. For expression in *Escherichia coli*, the N-terminus of FUS-C was fused to maltose-binding protein (MBP) via a short linker that contained a tobacco etch virus (TEV) protease site (Figure 1). This fusion strategy is commonly employed for *E. coli* expression of intrinsically disordered proteins because the MBP module promotes solubility and facilitates isolation and purification.[32,33] FUS-C appeared to be fully soluble after TEV cleavage of 5 µM fusion protein at room temperature in 50 mM Tris, pH 7.5, containing 150 mM NaCl and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP). The phosphine was included to prevent disulfide formation between Cys residues.

Phase separation mediated by FUS alone is most avid for the full-length protein, but LLPS can occur when the LCD and RBD are expressed separately and then combined. The first design, peptide Test 1, corresponds to residues 33-66 of FUS, a segment within the LCD. A Glu residue was added at the N-terminus to give the peptide a negative charge near or above neutral pH for good water-solubility. When Test 1 was combined with the FUS-C, only solid aggregates were observed (Figure 1.2a). The solution mixture was collected and spun down to obtain a white precipitate, which was stable at high temperature and detergent sodium dodecyl-sulfate (SDS). Different concentration combinations of FUS-C and peptide Test 1 as well as different buffers and salt concentrations were extensively screened, but only aggregates were observed in the mixture. Careful examination of the peptide sequence led me to be concerned about the number of Gln residues in Test 1, because polyglutamine is known to form a type of aggregate known as a polyQ stack. This type of aggregate is formed by natural proteins that have expanded polyglutamine repeats, which are associated with human diseases.[34] I designed peptides, Test 2 and Test 3, based on the most repeated motifs in the FUS LCD: SYSSYG and GYGSYG. These new peptides did not contain Gln, but they nevertheless caused aggregate formation when combined with the FUS RBD (Figure 1.2b-c).

I noticed a difference in the amount of aggregate formed with Test 2 and Test 3 via microscopy, with Test 2 appearing to cause more aggregate formation than Test 3. After staining the aggregates with thioflavin T (ThT), a commonly used dye to detect the formation of amyloid fibril,[35,36] I was able to figure out that peptide Test 2 had more aggregates formed than peptide Test 3 when the peptide interacted with FUS-C.[37] I concluded that the glycine-containing peptide is more flexible, while glutamine and serine confer rigidity and promote aggregation formation, which is consistent with previous studies.[19] I decided to replace all Ser with Gly to enhance the flexibility of my peptides. Only Gly, Glu and Tyr were used in peptide Test 4 and Test 5. “GGGY” and “GGGGGY” repeats were installed to form a “sticker-and-spacer” pattern. Stickers are units that interact favorably with each other, while spacers separate the stickers from one another to impart flexibility and may modulate the phase separation behavior.[31,38] The tyrosine residues in Test 4 and Test 5 are intended to serve as the stickers that interact with arginine in FUS-C. Gly residues serve as the spacers to give conformational flexibility. The uniform pattering of aromatic residues is intended to promote LLPS while inhibiting aggregation. Very little aggregation was observed with peptide Test 4 and Test 5 (Figure 1.2 d-e).

Table 1. Evolution of the peptide sequence design.

Peptide No.	Sequence:	Observation
Test 1	`EYGQQSYSGYSQSTDTSGYGQSSYSSYGQSQNTGY`	Aggregates
Test 2	`ESYSSYGSYSSYGSYSSYGSYSSYGSYSSYG`	Aggregates
Test 3	`EGYGSYGGYGSYGGYGSYGGYGSYGGYGSYG`	Aggregates
Test 4	`EGGGYGGGYGGGYGGGYGGGY`	Few aggregates
Test 5	`EGGGGGYGGGGGYGGGGGYGGGGGYGGGGGY`	Few aggregates
1 (Tyr)	`GGGYEEEYGGGYGGGYEEEYGGGYGGGY`	Phase separation

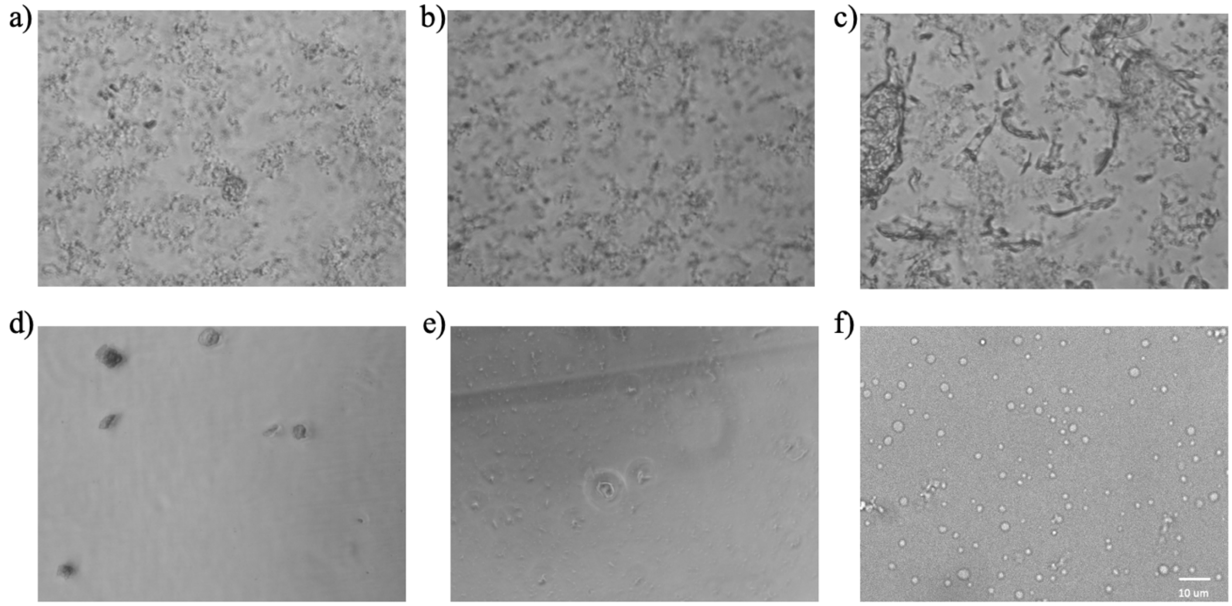


Figure 1.2. Formation of aggregates between 20 μM FUS-C with 40 μM of different peptides: **(a)** peptide Test 1; **(b)** peptide Test 2; **(c)** peptide Test 3; **(d)** peptide Test 4; **(e)** peptide Test 5; **(f)** peptide 1. The experiment was performed in in 50 mM Tris buffer, pH 7.5, 150 mM NaCl and 0.5 mM TCEP. Scale bar, 10 μm.

The solubility problems encountered with peptides Test 1 to Test 5 led us to design peptide 1(Tyr). Cakmak et al. have evaluated the impact of poly-ion multivalency on the functional performance of coacervates as compartments.[14] When one arginine interacted with glutamic acid, uniform solution was observed. However, when the polyanion $(Glu)_{10}$ mixed with polycation $(Arg)_{10}$, coacervates were observed. Increasing the number of Glu in peptide sequence leads to the finalized peptide model pep 1(Tyr). In addition to 7 evenly spaced Tyr residues, 1(Tyr) contains 15 Gly residues, which should ensure conformational disorder, and 6 Glu residues, to confer a net negative charge in the pH 7.5 buffer. Phase separation was observed when this new peptide model interacted with FUS-C (Figure 1.2f).

We observed that the removal of MBP tag from the protein construct MBP-FUS-C required carefully controlled conditions. When 20 μM MBP-FUS-C was treated with 2 mol% TEV protease at 37°C for 6 hours, a clear solution was initially observed after the tag was fully removed, and the protein solution was moved from the 37℃ incubator to room temperature. Aggregation was observed 30 min at room temperature. However, this aggregation was stopped by the addition of 1(Tyr) immediately after the protein sample was removed from the incubator. The lower temperature variation may cause FUS-C to form β-sheet-like structures, which lead to aggregation.[39] The ability of 1(Tyr) to halt of aggregation and induce phase separation makes the system extremely complicated to understand. We therefore developed a different cleavage protocol: 5 μM MBP-FUS-C was treated with 5 mol% TEV protease at room temperature for 6 hours. The protein solution remained clear (no aggregates under microscopy) for more than 12 hours. Combining 5 μM FUS-C and 50 μM 1(Tyr) at room temperature led to phase separation as detected via confocal microscopy. The sample of 1(Tyr) contained 1 mol% of the peptide bearing an N-terminal fluorescein unit, to enable visualization of the droplets (Figure 1.3A). Condensed phase droplets were observed to fuse on a short time scale, which provides qualitative evidence of the liquid nature of these droplets. Fluorescence microscopy studies were then conducted with samples in which each component contained 1 mol % of a fluorescently labeled derivative, 1(Tyr) bearing an N-terminal fluorescein unit or FUS-C with an extra Cys inserted at the C-terminus for tetramethylrhodamine (TAMRA) attachment. These studies established that both components were present in the condensed phase droplets (Figure 1.3B).

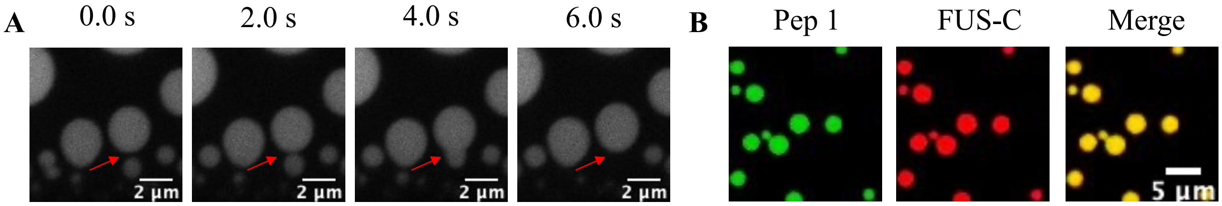


Figure 1.3. A. Time series of droplets formed by 50 µM 1(Tyr) (identified in Figure 2) and 5 µM FUS-C in 50 mM Tris, pH 7.5, containing 150 mM NaCl and 0.5 mM TCEP, showing droplet merger. Scale bar = 2 µm. **B.** Colocalization of 1(Tyr) (including 1 mol% FITC-labeled 1(Tyr), fluorescence image at 488 nm, false colored green) and FUS-C (including 1 mol% TAMRA-labeled FUS-C, fluorescence image at 561 nm, false colored red). The merged image is shown on the right. Scale bar = 5 µm.

Through the process of identifying and solving the challenges, I was able to develop 1(Tyr), which has a simple amino acid composition and supports LLPS when mixed with arginine-rich FUS-C. Peptide 1(Tyr) is readily prepared by solid-phase peptide synthesis (SPPS), which allowed me to replace the Tyr residues with noncanonical residues to evaluate the role of cation-π interactions in phase separation.

1.3.2 Evaluation of Charge Complementarity

To ask whether charge complementarity is necessary for LLPS that occurs when FUS-C and 1(Tyr) are combined, we examined analogues of 1(Tyr) in which three or all six of the Glu residues were replaced by Lys (Figure 1.4B). These peptides should have no net charge at pH 7.5 (1a-b) or a positive charge (1c), in contrast to the expected negative charge of 1(Tyr). Combining

50 µM 1a, 1b or 1c with 5 µM FUS-C did not lead to phase separation. Thus, the net negative charge of 1(Tyr) is necessary for co-assembly with cationic FUS-C to form the condensed phase, which means that the assembly can be considered a complex coacervate.[14] Charge complementarity should result in direct Coulombic attractions between 1(Tyr) and FUS-C as well as entropically favorable counterion release.[40–42]

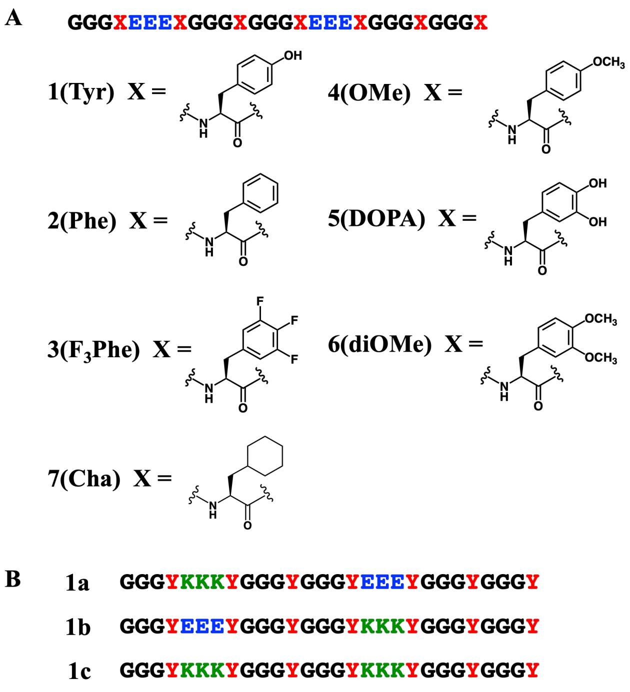
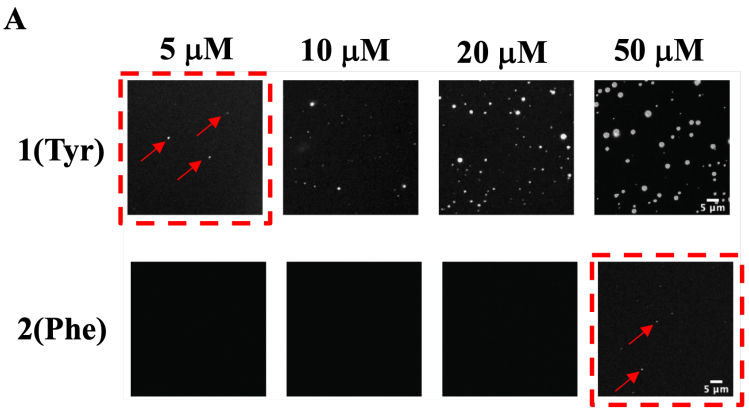


Figure 1.4. A. Sequence of designed peptides with different residues at X sites. **B**. Derivatives of **1(Tyr)** with different charge states.

To determine whether the negative charge of 1(Tyr) was sufficient for LLPS in the presence of FUS-C, we evaluated peptide 2(Phe), in which each of the seven Tyr residues was replaced with Phe. For this comparison, we evaluated a series of solutions containing 5 µM FUS-C and varying concentrations of the peptide in 50 mM Tris, pH 7.5, containing 150 mM NaCl and 0.5 mM TCEP. In an initial survey, samples containing 5, 10, 20 or 50 µM 1(Tyr) or 2(Phe), including 1 mol% of the fluorescein-labeled derivative, were examined by microscopy (Figure 1.5A). For solutions containing 1(Tyr) and FUS-C, a few droplets were detected with 5 µM peptide, and droplets were abundant at 20 µM. In contrast, droplets were observed only at 50 µM 2(Phe), and not at the lower concentrations.



B

Peptide	C_{min}
1(Tyr)	5
2(Phe)	>20
3(F_3Phe)	14
4(OMe)	14
5(DOPA)	2
6(diOMe)	>20
7(Cha)	4

Figure 1.5. **A**. Representative confocal images of droplets formed at different concentrations (μM) of **1(Tyr)** or **2(Phe)** with 5 μM FUS-C. Scale bar = 5 μm. **B**. The minimum peptide concentration (C_{min}; μM) for each peptide at which droplets were visible by fluorescence microscopy when mixed with 5 μM FUS-C. In each case, the peptide sample contained 1 mol% FITC-labeled peptide.

The dramatic difference between mixtures containing FUS-C and either 1(Tyr) or 2(Phe) that is evident in Figure 3A shows that Coulombic interactions are not the dominant factor driving phase separation in this system. Instead, this comparison suggests that the identity of the aromatic side chain plays a critical role in the molecular assembly required for condensed phase formation. These observations motivated experiments involving all of the 28-mers shown in Figure 1.4A to explore the noncovalent interactions that underlie phase separation.

1.3.3 Effects of Side Chain Identity on Phase Separation Propensity

For each of the peptides in Figure 1.4A, the propensity for phase separation upon mixing with 5 μM FUS-C was assessed in two ways. First, we identified the minimum pep-tide concentration (C_{min}) at which droplets were visible by fluorescence microscopy (each peptide sample contained 1 mol% of the fluorescein-labeled derivative). These values are summarized in Figure 1.5B, based on images shown in Figure 1.6. Second, for samples containing 5, 10, 20 or 50 µM peptide, microscopy images were used to determine the total area of the droplets within a 3000 µm^2 field of view (Figure 1.6). The area analysis was conducted with four independently prepared samples for each peptide + FUS-C pairing, with two independent measurements per sample. In each case, the microscopy images were obtained 30 minutes after the two components were mixed.

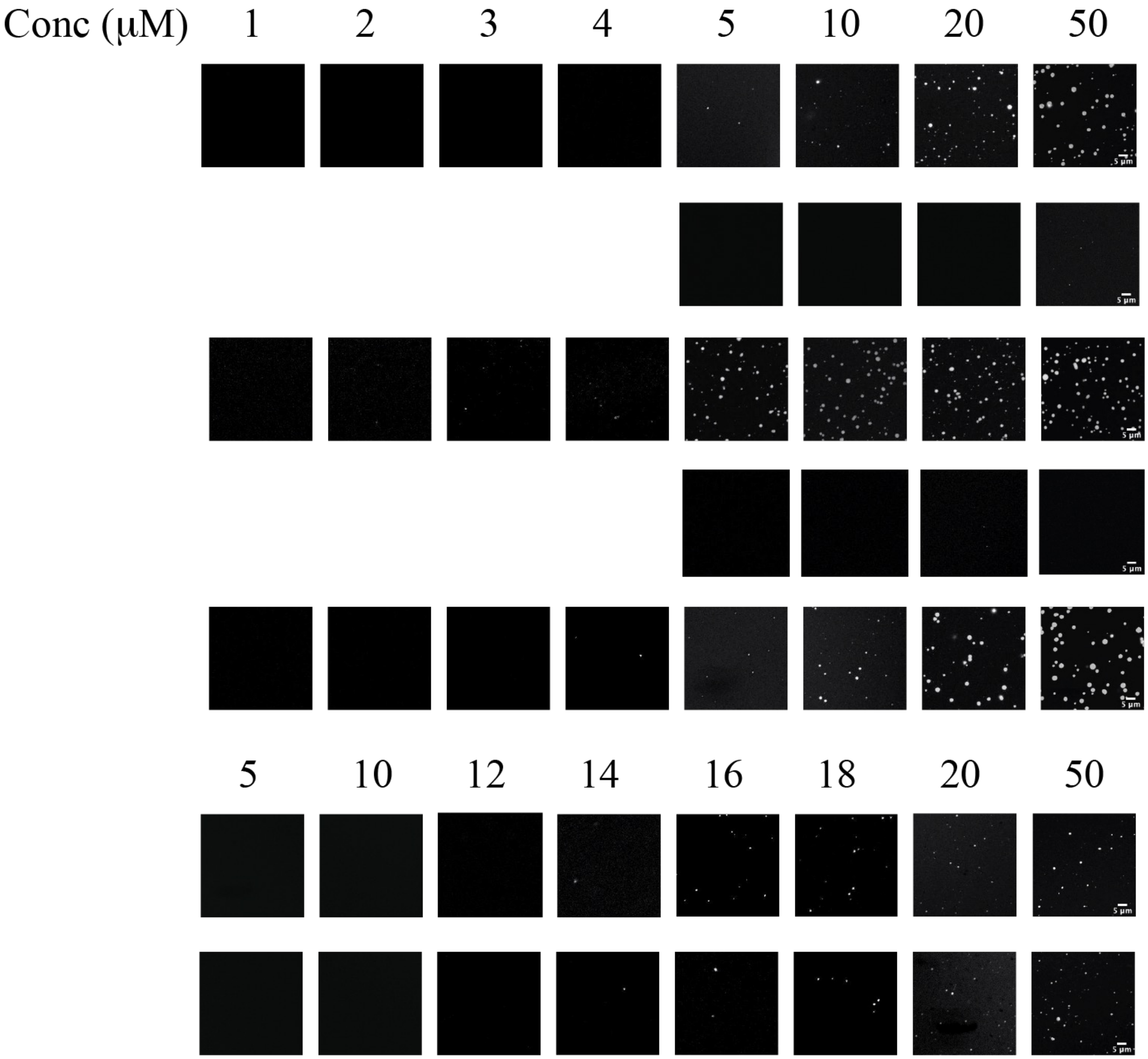


Figure 1.6. Representative confocal micrographs of coacervates at different concentrations of corresponding peptides with 5 µM FUS-C. These data were used to determine the C_{min} values shown in Fig. 3B. Scale bar = 5 µm. Buffer condition: 50 mM Tris, pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

For 1(Tyr), 5 μM was the lowest concentration at which droplets were visually detected under our conditions (Figure 1.5B), while the minimum concentration for 2(Phe) was >20 μM. Comparison of 1(Tyr) and 2(Phe) via droplet area at the four peptide concentrations reinforced the conclusion that these two peptide manifest profoundly different propensities for phase separation upon mixing with FUS-C (Figure 1.6). Collectively, these results indicate that a net negative charge is not sufficient to enable the avid phase separation observed when 1(Tyr) is combined with FUS-C and highlight the importance of the side chain difference between Tyr and Phe.

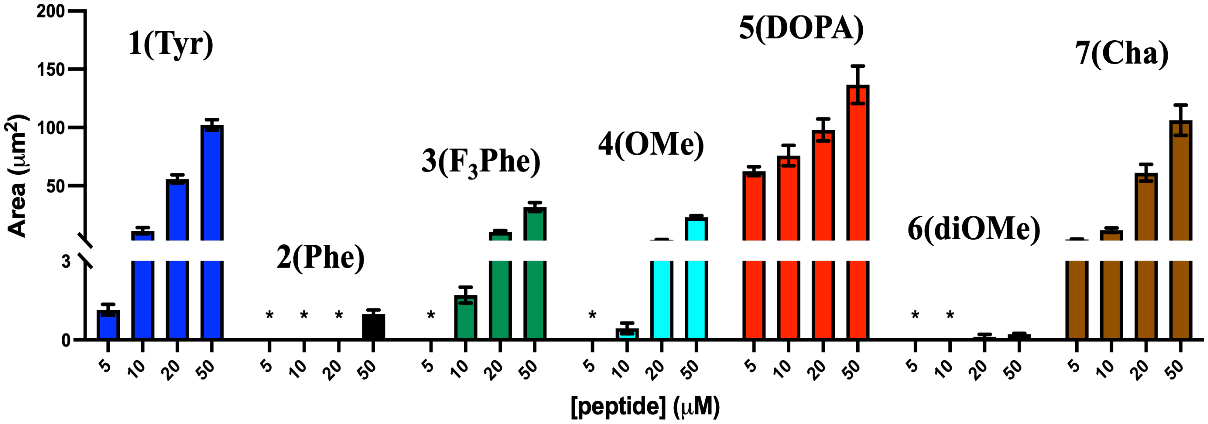


Figure 1.7. Area measurements of droplets observed in a 3000 μm^2 field of view. Phase separation condition: 5 μM FUS-C was treated with the indicated peptide at different concentrations (1 mol% FITC-labeled corresponding peptide). N = 4 independent replicates (two measurements per replicate). Error bar, SEM. * indicates no droplets observed.

The substantial decline in LLPS propensity resulting from Tyr-to-Phe replacement in our system is consistent with findings of Wang et al., who reported that global replacement of Tyr with Phe in the N-terminal domain of FUS led to a decrease in LLPS propensity.[19] However, Qamar et al. found that replacing a subset of these Tyr residues in FUS with Phe did not diminish LLPS propensity, and this observation was taken as evidence that cation-π interactions play an important role in driving FUS self-association in the condensed phase.[20]

The cation-π interaction arises from attraction between the positive charge and negative electrostatic potential of the aromatic ring (charge-quadrupole interaction).[21–23,27] Dougherty et al. have estimated this attraction based on quantum mechanical calculations that map out the electrostatic potential of the aromatic unit.[28] These calculations predict that the side chains of Tyr and Phe are energetically similar as participants in cation-π interactions. In contrast, the 1,2,3-trifluorophenyl group of F_3Phe is predicted to be a much poorer participant.[28] We therefore examined peptide 3(F_3Phe), the analogue of 1(Tyr) in which all seven Tyr residues were replaced by F_3Phe (Figure 1.4A). Droplets were detected at or above 14 μM 3(F_3Phe) (Figure 1.5B), which indicates a higher propensity for phase separation with FUS-C by this peptide relative to 2(Phe) (minimum concentration for droplets > 20 μM). Comparison of droplet area for 3(F_3Phe) vs. 2(Phe) (Figure 1.6) further supports the conclusion that 3(F_3Phe) has the higher propensity for phase separation in combination with FUS-C. This result is not consistent with the hypothesis that cation-π interactions drive phase separation in our system.

To understand the features of the Tyr side chain that underlie avid LLPS upon mixing 1(Tyr) and Arg-rich FUS-C, we examined peptides 4(OMe), 5(DOPA) and 6(diOMe), the analogues of 1(Tyr) in which all seven Tyr residues were replaced by 4-methoxyphenylalanine, 3,4-dihydroxyphenylalanine (DOPA) or 3,4-dimethoxyphenylalanine, respectively. These peptides showed substantial differences in their abilities to support LLPS in combination with FUS-C. Eliminating the side chain H-bond donor, by replacing 1(Tyr) with 4(OMe) (Figure 1.6), caused the minimum concentration for droplet formation to rise from 5 to 14 μM (Figure 1.5B). At each peptide concentration used for droplet area measurements, the value was lower in solutions containing 4(OMe) + FUS-C relative to solutions containing 1(Tyr) + FUS-C. Previously reported quantum mechanical calculations suggest that there is little difference between the electrostatic potentials of the Tyr and 4-methoxyphenylalanine side chains.[28] Therefore, the decline in LLPS propensity with FUS-C manifested by 4(OMe) relative to 1(Tyr) is not consistent with the hypothesis that phase separation in these systems is driven by attractions between cationic Arg side chains and aromatic side chain π systems.

Replacing 1(Tyr) with 5(DOPA) led to a substantial increase in LLPS propensity in combination with FUS-C. The minimum concentration for droplet formation declined to 2 µM for 5(DOPA) from 5 µM for 1(Tyr) (Figure 1.5B). This effect was evident also in droplet area measurements for solutions containing 5 µM peptide: the area value for 5(DOPA) was much higher than that for 1(Tyr) at this peptide concentration (Figure 1.6). These observations are consistent with the hypothesis that the H-bond donor properties of the aromatic side chain play a critical role in determining LLPS propensity upon combination with FUS-C. Replacing 5(DOPA) with 6(DiOMe) led to a dramatic decline in LLPS propensity: the minimum concentration for droplet formation was >20 µM for 6(diOMe), and droplet area for 5(DOPA) was much larger than for 6(diOMe) at all concentrations.

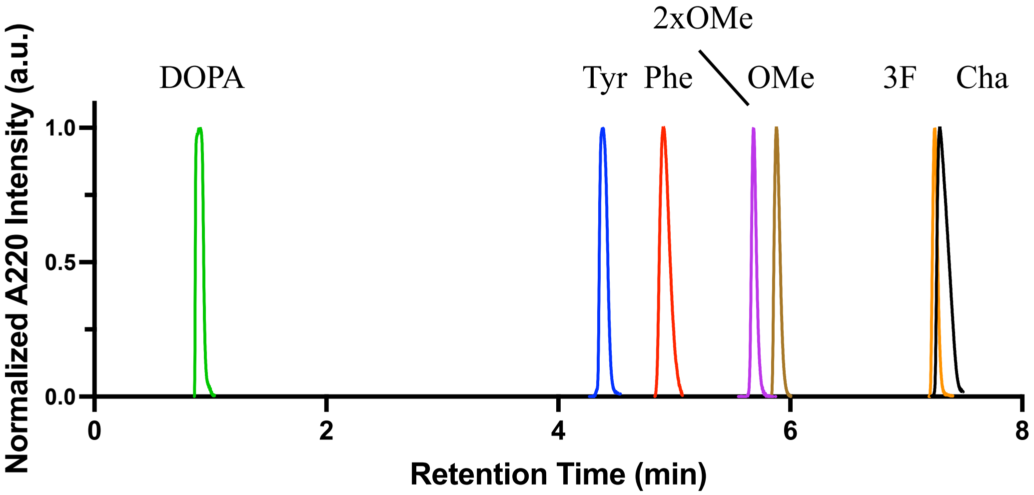


Figure 1.8. Reverse-phase high liquid chromatography retention time of Gly-Gly-Gly-Xxx tetramers, where is one of the seven residues shown in Figure 1.4A. 0%B 1min, then 0-60%B 15min; solvent A: 0.1% TFA in MilliQ water; solvent B: 0.1% TFA in ACN.

Taken together, the observations for peptides 1(Tyr), 2(Phe), 4(OMe), 5(DOPA) and 6(diOMe) are consistent with the hypothesis that aromatic ring substituents that serve as H-bond donors promote LLPS when the peptide is combined with FUS-C. In addition, these observations collectively are not consistent with the hypothesis that cation-π interactions play a significant energetic role in peptide + FUS-C LLPS.

The substantial LLPS observed with 3(F_3Phe) shows that H-bond donors on the aromatic side chains are not required for promotion of LLPS upon mixing with FUS-C. To ask whether side chain hydrophobicity could promote phase separation for at least some peptide + FUS-C pairings, we compared reverse-phase high-pressure liquid chromatography (RP-HPLC) elution profiles for short peptides containing Tyr or one of the replacement residues (Figure 1.8).[43,44] This comparison included cyclohexylalanine (Cha), which has a very hydrophobic and non-aromatic side chain. Tetrapeptides with the sequence Gly-Gly-Gly-Xxx were evaluated, where Xxx was the variable position. RP-HPLC elution profiles have commonly been used to com-pare hydrophobicity among peptides.[43–46] The stationary phase for our comparison, silica modified with C18 alkyl groups, is highly hydrophobic, and the mobile phase, a gradient of water-acetonitrile mixtures, is hydrophilic. With a consistent elution protocol, longer retention should directly correlate with greater hydrophobicity.

The tetramer GGG-DOPA was the least strongly retained, which is consistent with the expectation that the pair of hydroxyl groups on the aromatic ring should render DOPA very hydrophilic. GGG-Tyr had the second-shortest elution time, and GGG-Phe was slightly more strongly retained, i.e., more hydrophobic. Both GGG-MeOPhe and GGG-DiMeOPhe were more hydrophobic (longer elution times) relative to the tetrapeptides bearing side chains with hydroxyl groups or GGG-Phe. The RP-HPLC comparison suggested that the side chain of F_3Phe is similar to the side chain of Cha in terms of hydrophobicity.

The RP-HPLC results motivated us to evaluate peptide 7(Cha) for the ability to promote LLPS in combination with FUS-C in 50 mM Tris, pH 7.5, containing 150 mM NaCl and 0.5 mM TCEP. The minimum concentration for droplet detection was 4 μM (Figure 1.5B), which is significantly below the value for 3(F_3Phe). Droplet area was larger for 7(Cha) relative to 3(F_3Phe) at all four peptide concentrations evaluated (Figure 1.6). The Cha side chain is obviously incapable of participating in cation-π interactions or donating H-bonds. Thus, the observations with 7(Cha) are consistent with the hypothesis that LLPS resulting from peptide + FUS-C mixing can be promoted by side chain hydrophobicity.

Based on the assessments of droplet formation for peptide + FUS-C pairings in Figures 1.5 and 1.6, the overall LLPS propensity order among the 28-mer peptides is 5(DOPA) > 7(Cha) > 1(Tyr) > 3(F_3Phe) ~ 4(OMe) > 2(Phe) ~ 6(diOMe). This order is quite different from the RP-HPLC elution order among the GGG-Xxx tetramers [5(DOPA), 1(Tyr), 2(Phe), 6(diOMe), 4(OMe), 3(F_3Phe), 7(Cha)]. We interpret these data to suggest that at least two different noncovalent interaction profiles can promote the attractions among peptide and FUS-C molecules necessary to create the network that scaffolds a condensed phase. One noncovalent interaction profile feature H-bond donation by side chain groups, such as the hydroxyl groups of Tyr and DOPA, as a key factor. In the other profile, hydrophobic interactions of side chains are a key factor. Since residues with the two most hydrophobic side chains among those we evaluated, Cha and F_3Phe, are not found in proteins, this latter noncovalent interaction profile may not be relevant to LLPS mediated by natural proteins.

1.3.4 Fluorescence Recovery after Photobleaching (FRAP) with Selected Condensed Phases

FRAP studies were conducted to probe the physical properties of condensed phases formed by peptide + FUS-C pairs.[47] Three peptides were selected for these measurements, 1(Tyr) and 5(DOPA) to represent the H-bond-promoted mode of assembly, and 7(Cha) to represent the hydrophobically promoted mode of assembly. The data in Figures 1.5 and 1.6 indicate that these three peptides have the highest phase-separation propensities, when combined with FUS-C, among the set we evaluated. For the FRAP studies, LLPS was induced by combining 5 μM FUS-C with 50 μM peptide in 50 mM Tris, pH 7.5, with 150 mM NaCl and 0.5 mM TCEP at room temperature. Samples containing 1 mol % fluorescein-labeled peptide were used to monitor peptide dynamics, and samples containing 1 mol % TAMRA-labeled FUS-C were used to monitor the dynamics of this protein fragment. In each case, the droplets were relatively small (up to 1.5–2.5 μm in diameter), and the laser irradiation therefore bleached most of the droplet, even though these measurements focused on the largest droplets (Figure 1.9A). Fluorescence recovery in these cases results from peptide or FUS-C diffusion within the droplet and/or from peptide or FUS-C diffusion into the droplet from the surrounding dilute phase.

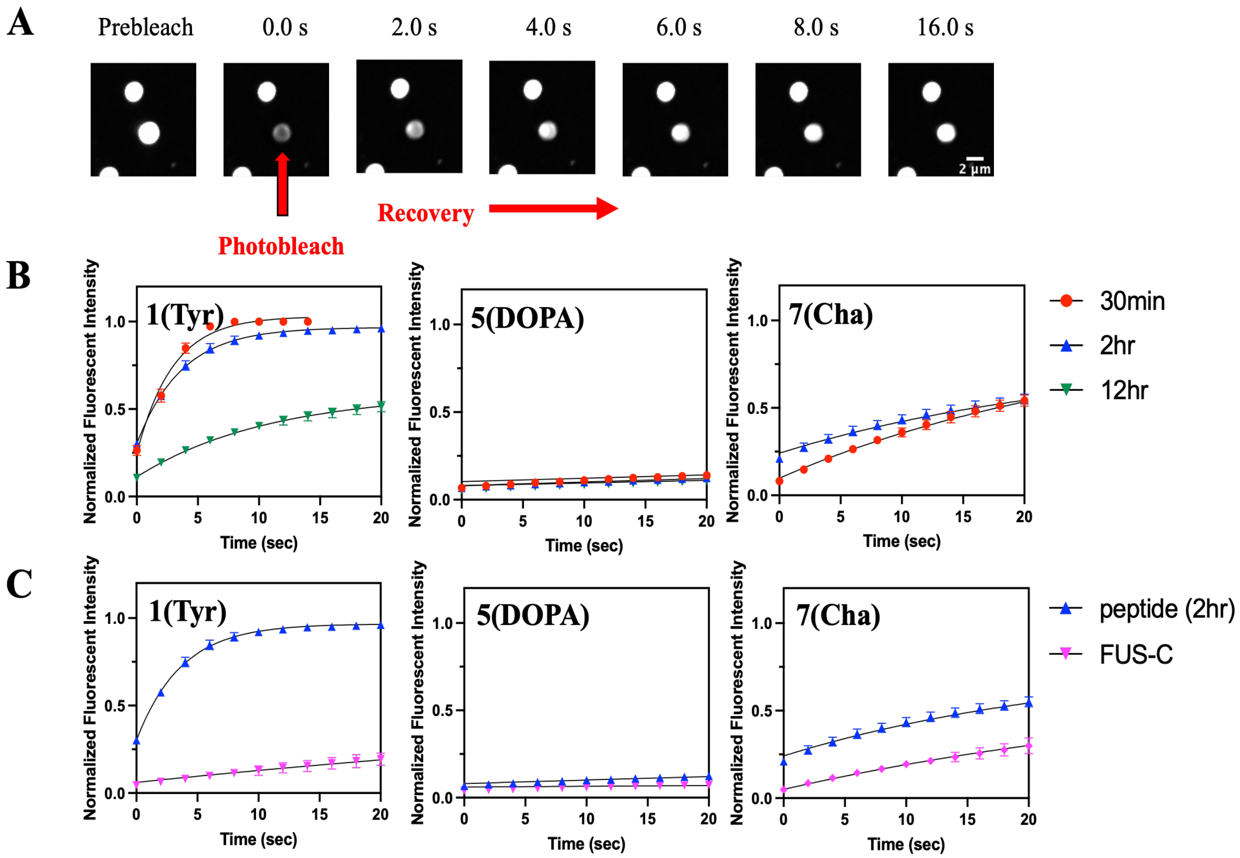


Figure 1.9. **A**. Representative micrographs for a FRAP experiment on condensates formed by 50 µM **1(Tyr)** (1mol% FITC-labeled) and 5 µM FUS-C. At 0.0 s, a region of interest within the droplet is photobleached (dark). Scale bar = 2 µm. **B**. FRAP comparison of 1**(Tyr)**, **5(DOPA)** and **7(Cha)** with 1mol% FITC-labeled corresponding peptide at different timepoints: 30 min (red), 2 hr (blue) and 12 hr (green). **C**. FRAP comparison between each of the three peptides (1mol% FITC-labeled) and FUS-C (1mol% TAMRA-labeled) at 2 hr timepoints: peptide (blue) and FUS-C (magenta). The blue data for each peptide in parts **B** and **C** are the same.

Based on UV-vis absorption data, the dilute phase for the three peptide/FUS-C pairs contained different peptide concentrations (Figure 1.10): ~1 µM 1(Tyr), ~13 µM 5(DOPA) and ~4 µM for 7(Cha). Thus, in each case, most the peptide was located in the condensed phase, although each dilute phase contained residual peptide that could exchange with peptide bleached by irradiation. Efforts to determine the concentration of FUS-C in the dilute phase were unsuccessful, which suggests that nearly all of this polypeptide was drawn into the three condensed phases under our experimental conditions.

In condensed phase droplets formed with 1(Tyr), the peptide was very mobile (Figure 1.9B). At 0.5 hr after mixing with FUS-C, droplets formed with 1(Tyr) fully regained the original fluorescence level with half-life ~2 sec after photobleaching; very similar behavior was observed 2 hr after mixing. By 12 hr after mixing, only ~50% of the original fluorescence level was recovered, and the half-life for recovery was somewhat longer, ~8 sec. The diminished recovery maximum at 12 hr suggests that there is a slow process by which peptide molecules become immobilized within the condensed phase and can no longer exchange with peptide molecules in the surrounding dilute phase, perhaps corresponding to gelation. Similar "aging" processes have been reported for condensed phases formed by FUS or FUS fragments.[39,48,49] Since the concentration of 1(Tyr) in the dilute phase was lower than for 5(DOPA) or 7(Cha), the rapid and complete fluorescence recovery observed over the first 2 hr for 1(Tyr) suggests that the timescale and extent of fluorescence recovery for all three peptides was not limited by the amount of peptide in the dilute phase.

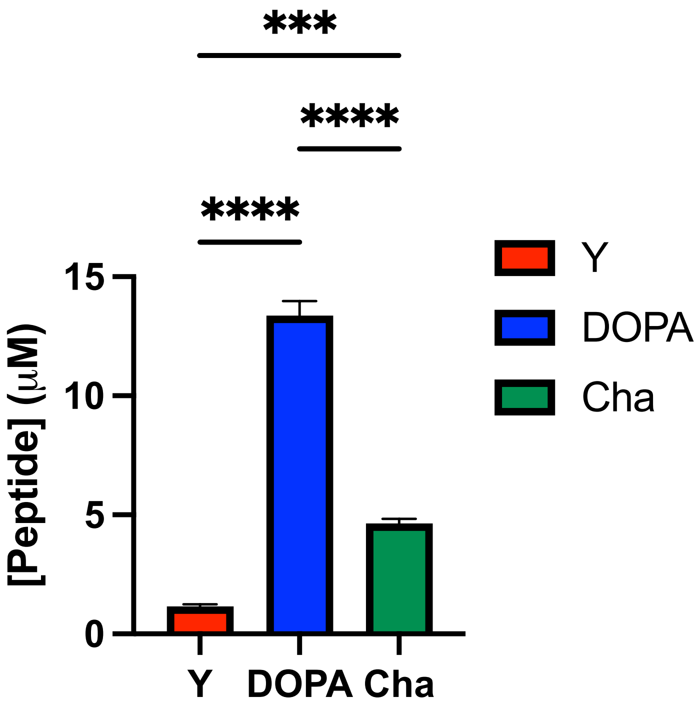


Figure 1.10. Concentration of peptide in the dilute phase after phase separation induced by mixing 5 μM FUS-C with 50 μM of corresponding peptide (10 mol% labeled): **1(Tyr), 5 (DOPA)** or **7(Cha)** in 50 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM TCEP. Samples were incubated at room temperature for 30 min. The coacervate sample was then centrifuged, and the supernatant was collected. The peptide concentration in the supernatant was determined by UV-spectroscopy. N = 4 independent replicates. Error bar, SEM. P values ($P < 0.001$ (***), $P < 0.0001$ (****))were calculated by one way ANOVA.

The FUS-C component appeared to be less dynamic than the peptide component in the condensed phase formed with 1(Tyr) (Figure 1.9C). FRAP data collected at 2 hr after mixing suggested a half-life of ~100 sec for recovery of the FUS-C component. The predicted extent of fluorescence recovery was about 50%. These values did not change substantially at 12 hr after mixing. Since the concentration of FUS-C in the dilute phase was apparently very low, the FRAP data may result from limitations in the amount of external FUS-C available to exchange with FUS-C bleached within the droplet. Alternatively, perhaps half of the FUS-C content in the droplets is immobilized upon phase separation. The FRAP data may suggest that the slow process that leads to immobilization of peptide 1(Tyr) within the network that spans the condensed phase does not operate on the FUS-C component, or that this process is not detected because of the rapid immobilization of a substantial fraction of the FUS-C.

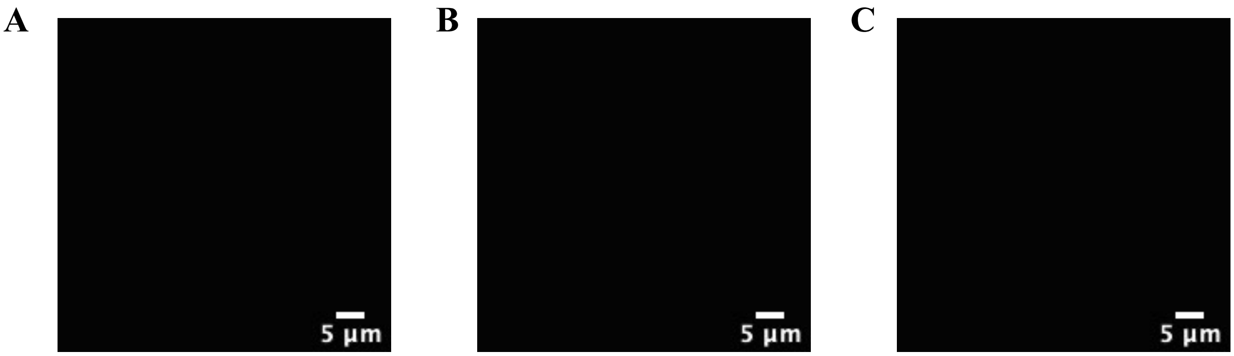


Figure 1.11. Representative confocal images after 6 M guanidine HCl treatment of droplets formed between 5 µM FUS-C and 50 µM **1(Tyr)** (1 mol% fluorescein-**1(Tyr)**, **A**), **5(DOPA)** (1 mol% fluorescein-**5(DOPA)**, **B**), or **7(Cha)** (1 mol% fluorescein-**7(Cha)**, **C**). Scale bar = 5 µm. No droplets were observed after the treatment with guanidine HCl. For the sample containing **5(DOPA)**, this result indicates a lack of large-scale covalent crosslinking. Buffer condition: 50 mM Tris, pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

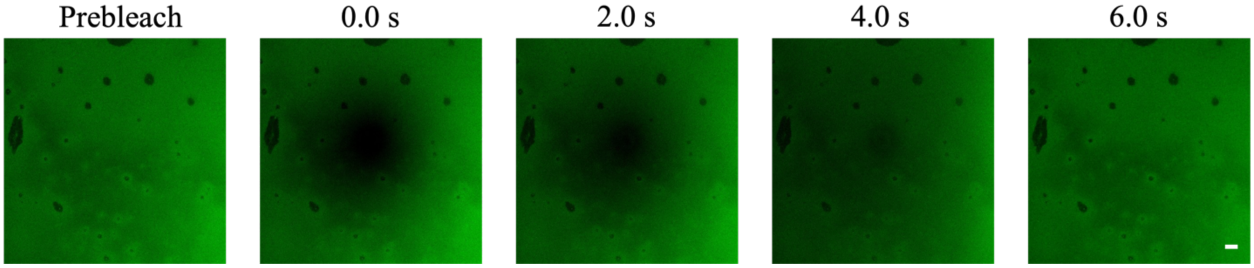


Figure 1.12. Representative confocal micrographs of coacervates formed with 5 µM FUS-C and 50 µM **7(Cha)** (1 mol% fluorescein-**7(Cha)**) at the 12-hour timepoint. No clear boundary for coacervate droplets was observed. Instead, the condensed phase appeared to have spread across the surface. These images show results of a FRAP experiment. Laser irradiation bleaches a region of the condensed phase covering the surface, and recovery of fluorescence is observed. Scale bar = 2 µm. Buffer condition: 50 mM Tris, pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

FRAP measurements involving droplets formed with 5(DOPA) indicated substantially reduced mobility of the peptide and FUS-C components, relative to the condensed phase formed by mixing 1(Tyr) and FUS-C. For both 5(DOPA) and FUS-C, fluorescence recovery was very slow, whether the measurement was made at 0.5 hr, 2 hr or 12 hr after mixing. The extent of recovery could not be accurately predicted from the FRAP data. These observations suggest that the condensed phase formed by 5(DOPA) + FUS-C may be a gel rather than a liquid. Treatment of this condensed phase with 6 M guanidine HCl led to complete dissolution, which indicates that condensate formation with 5(DOPA) does not involve extensive oxidative covalent crosslinking of DOPA side chains (Figure 1.11).[50] Condensates formed by 1(Tyr) + FUS-C and by 7(Cha) + FUS-C were also fully disrupted by 6 M guanidine HCl.

FRAP measurements with the condensate resulting from mixing peptide 7(Cha) and FUS-C revealed that the peptide component was dynamic at 0.5 hr and 2 hr after mixing. The estimated half-life for fluorescence recovery was 15-20 sec at both time points. The FUS-C component seemed to have a level of mobility similar to that of 7(Cha). FRAP data collected for FUS-C at 2 hr after mixing suggested a half-life in the range 10-50 sec for recovery. The predicted recovery was ~50%. Measurement at 12 hr, however, was impossible because the condensed phase droplets formed by 7(Cha) + FUS-C could no longer be observed via microscopy. The images suggested that after 12 hr the droplets had merged and formed a thin, continuous layer across the surface of the microscope slide (Figure 1.12). This type of "wetting" was not observed for condensed phases formed when either of the hydroxyl-bearing peptides 1(Tyr) or 5(DOPA) was combined with FUS-C.

We note that the surfaces of the glass slides used for these studies had been pretreated with bovine serum albumin (BSA) to minimize nonspecific adsorption. Our observation that the condensed phase formed by 7(Cha) + FUS-C slowly wets this surface is consistent with the recent report that another condensate formed with a hydrophobic protein pair wets BSA-treated glass surfaces.[51]

The droplet area comparisons (Figure 1.7) suggest that both 5(DOPA) and 7(Cha) have a higher propensity for co-assembly with FUS-C relative to any of the other 28-mer peptides we evaluated. For both 5(DOPA) and 7(Cha), condensed phase assemblies were very abundant at 5 μM peptide, while droplets were sparse for 1(Tyr) at 5 μM, and no droplets were observed for the other four peptides at this concentration. FRAP comparison, however, indicates that the condensed phase promoted by H-bonding (5(DOPA)) has material properties that differ from those of the condensed phase promoted by hydrophobicity (7(Cha)).

1.4 Conclusions

Evolution has harnessed cation-π interactions to stabilize complexes between proteins and positively charged ligands, as established by experimental studies conducted over the past three decades.[21–29,52] Close contacts between cationic and aromatic units in amino acid residue side chains observed in protein crystal structures suggest that cation-π interactions can stabilize native tertiary structure.[52,53] However, such contacts are much less frequent than the H-bonds or contacts between nonpolar side chains within tertiary structures,[54–56] which suggests that cation-π interactions are not a major driving force for protein folding. Both the hydrophobic effect and intramolecular H-bonding are believed to play key roles in determining specific protein tertiary structures,[54–58] although these roles have been debated.[55]

LLPS mediated by polypeptides (a single polypeptide or a combination of polypeptides, as in the studies described here) appears to require formation of a molecular network that spans the condensed phase and is held together by intermolecular attractions. For systems lacking defined recognition motifs (such as PDZ or SH3 domains and their peptide ligands[1,8]), the condensate-stabilizing interactions must involve small substructures such as side chains or the backbone atoms of individual residues.[11–16] The collective energetic threshold for formation of the condensate-stabilizing noncovalent interaction network must be lower (at appropriate concentrations) than the threshold for a tertiary structure-stabilizing (i.e., intramolecular) network, because most proteins or protein domains that mediate membraneless organelle formation in cells are intrinsically disordered. These features are exemplified by FUS and related proteins.[18–20] It is therefore important to elucidate noncovalent interaction networks that stabilize protein-mediated phase separation, which may differ from the networks that stabilize protein tertiary structures.[59]

Two highly cited studies published in 2018 concluded that cation-π interactions are essential for phase separation mediated by FUS and related proteins.[19,20] Nearly every paper that mentions biopolymer-mediated LLPS since 2018 has highlighted cation-π interactions as a driving force. Because, as noted above, cation-π interactions do not seem to be as common as H-bonds or nonpolar contacts within folded proteins, the cation-π interaction might be uniquely suited to formation of the dynamic intermolecular networks necessary to stabilize condensed phases.

Observations that have been interpreted to implicate cation-π interactions in FUS-mediated LLPS are not entirely consistent with the conclusion that cation-π interactions play a major role in this form of molecular assembly. For example, Qamar et al. showed that post-translational modification of Arg to generate asymmetric dimethylated Arg (ADMA) led to a decrease in phase separation propensity, and this observation was interpreted to support the hypothesis that cation-π interactions involving Arg and Tyr side chains provide a driving force for LLPS mediated by FUS and related proteins.[20] However, Waters et al. previously determined that ADMA forms a more stable cation-π interaction relative to Arg.[60] Wang et al. found that global replacement of Tyr with Phe in FUS led to a decrease in phase separation propensity, which led these authors to note that the role of Tyr-Arg interactions in promoting LLPS "cannot be solely due to generic cation-π interactions."[19] This caveat presumably reflects the fact that the Tyr and Phe side chains have similar propensities to engage in cation-π interactions.[28] Despite this caveat, these authors concluded that their results supported "the importance of cation-π interactions as drivers of phase separation."

One of the first studies to conclude that cation-π interactions are important for protein-mediated LLPS focused on Ddx4, a major constituent of germline granules.[31] The protein Ddx4 is largely disordered and rich in Phe and Arg. Nott et al. tested the cation-π hypothesis by heterologous expression of Ddx4 in the presence of racemic meta-fluoro-phenylalanine (mF-Phe), to generate a protein mixture with variable numbers of Phe → mF-Phe replacements. This material had a lower propensity for LLPS relative to native Ddx4 containing only L-Phe, which seems consistent with the hypothesis that cation-π interactions play an energetically significant role in this system. These authors also observed that conversion of Arg to ADMA in Ddx4 diminished LLPS propensity, which they interpreted to support the energetic importance of cation-π interactions in the assembly of Ddx4 molecules. However, this interpretation is not consistent with the finding that ADMA forms stronger cation-π interactions than does Arg,[60] as noted above.

Uncertainties associated with the studies summarized above motivated our experimental design. A phase-separating system comprised of two components, a long Arg-rich protein fragment generated via heterologous expression and a short Tyr-rich peptide synthesized chemically (1(Tyr)), freed us from the compositional constraints imposed by ribosomal synthesis for the Tyr-rich component. However, one feature of the noncovalent interaction network that controls LLPS induced by mixing FUS-C and 1(Tyr) differs from the noncovalent interaction network that controls LLPS mediated by full-length FUS itself. Charge complementarity between FUS-C (a polycation) and the peptide (a polyanion) is necessary to achieve phase separation at low μM concentrations. 1(Tyr) contains 6 Glu among the 28 residues. In contrast, the Tyr-rich domain of

FUS (residues 1-215) contains only five acidic residues. Our preliminary efforts to work with Tyr-rich fragments of < 40 residues derived from the N-terminal domain of FUS were hampered by low peptide solubility, presumably a result of the lack of charged side chains. These difficulties led us to design 1(Tyr).

The profound difference in LLPS propensity that results from global Tyr → Phe replacement (1(Tyr) → 2(Phe)) shows that the Tyr residues play an important role in promoting LLPS in combination with FUS-C. For FUS itself, global Tyr → Phe replacement suppresses LLPS.[19] This parallel suggests that our peptide+FUS-C system offers a valid basis for exploring the factors that control LLPS mediated by Arg-rich and Tyr-rich polypeptide components, including FUS and related proteins. However, we found that charge-neutral analogues of 1(Tyr) (i.e., peptides 1a and 1b) do not support phase separation when combined with FUS-C, at least in the concentration range we examined. Thus, it is possible that a noncovalent interaction network that includes attractions between anionic Glu side chains in 1(Tyr) and cationic Arg side chains in FUS-C limits the applicability of our conclusions to the behavior of full-length FUS.

Results obtained in our designed system are not consistent with the hypothesis that cation-π interactions are a major driver of phase separation induced by mixing Arg-rich FUS-C and 1(Tyr). This conclusion is evident not only from the decline in LLPS propensity upon Tyr → Phe replacement, but also from the significant decline in LLPS propensity upon replacing Tyr with O-methyl-tyrosine (4(OMe)). Ab initio calculations suggest that the methoxyphenyl group of the methoxyphenyl side chain has a slightly larger cation-binding energy than does the phenol group of the Tyr side chain.[28]

Our observation that replacing Tyr with DOPA (1(Tyr) → 5(DOPA)) substantially enhances phase separation propensity and results in a more ordered condensate suggests that H-bond donation by the Tyr and DOPA side chains plays a critical role in promoting phase separation in our system. This conclusion is consistent with analysis of Tyr → Phe mutations in folded proteins, which suggests that intramolecular H-bond donation by the Tyr hydroxyl can stabilize native tertiary structure.[57] Support for the proposed role of side chain H-bond donation in our system is provided by the substantial decline in LLPS propensity observed for 4(OMe) relative to (1(Tyr) and for 6(DiOMe) relative to 5(DOPA). Each of these comparisons involves removal of H-bond donors, presumably without diminishing the ability of the aromatic ring to engage in cation-π interactions.

Our data indicate that placing one methoxy group on the side chain aromatic ring (2(Phe) → 4(OMe)) causes an increase in phase separation propensity in combination with FUS-C, but placing a second methoxy group on the aromatic ring (4(OMe) → 6(diOMe)) causes a decrease in phase separation with FUS-C. An explanation for this trend is not clear to us. HPLC data (Figure 5) suggest that the side chain of O-methyl-tyrosine is slightly more hydrophobic than the side chain of di-O-methyl-DOPA, but this difference seems too small to explain a substantial variation in phase separation propensity.

The ability of the Tyr side chain to serve as an H-bond donor has received relatively little prior attention in the context of polypeptide-mediated LLPS. The most relevant studies, to our knowledge, have emerged from work inspired by mussel adhesive proteins.[61] Mussels anchor themselves to solid substrates via "foot proteins", many of which contain abundant DOPA residues. These proteins undergo LLPS as they are being secreted, exposure to sea water triggers solidification of the condensed phase. Some model studies have suggested that cation-π interactions involving DOPA side chains and Lys side chains, which are also abundant in mussel foot proteins, contribute to the self-adhesion of these proteins.[62] A different conclusion was reached, however, from recent studies of salt-triggered LLPS by a synthetic DOPA-rich peptide.[63] This peptide, designed based on a mussel foot protein, contained only DOPA, Lys and Gly and supported LLPS at high concentrations of peptide and salt. The analogous peptide with global DOPA → Tyr replacement did not undergo LLPS under the same conditions. Molecular dynamics simulations suggested that network formation by the DOPA-rich peptide was stabilized by intermolecular H-bonding between two DOPA side chains or between a DOPA side chain and the backbone of a Gly residue. Related studies on the mussel adhesive protein Pvfp-5β (83 residues) compared the version rich in Tyr with the analogue resulting from global Tyr → DOPA replacement.[64] LLPS occurred for solutions containing high concentrations of the DOPA version and salt; in contrast, the Tyr version formed solid aggregates under these conditions. Although Pvfp-5β contains eight Lys residues, a combination of 13C NMR measurements and quantum calculations led to the conclusion that cation-π interactions were not a driving force for assembly of either the Tyr or DOPA versions of this protein. Instead, the associations underlying LLPS by the DOPA version were attributed to H-bonding and π-π interactions of DOPA side chains.

Experimental and computational findings for mussel protein-based systems outlined above[63,64] offer an interesting context for our observation that 5(DOPA) manifests a very high phase-separation propensity when combined with FUS-C, which contains 60 Gly residues, even when each component is present at 5 μM. The network that scaffolds this condensed phase could be stabilized by transient H-bonding between DOPA side chains in 5(DOPA) and Gly residues in FUS-C. DOPA-DOPA H-bonding between different molecules of 5(DOPA) that are brought into proximity through interactions with FUS-C might also stabilize the network. It seems unlikely, however, that π-π interactions between DOPA side chains make significant contributions, because we found that replacing DOPA with the dimethyl derivative dramatically diminished LLPS propensity; π-π attractions should be possible in a mixture containing 6(DiOMe).

Our FRAP analysis of the condensed phase formed by 5(DOPA) and FUS-C indicates very low peptide mobility, which suggests that this condensate is a gel rather than a liquid. Formation of intracellular gel-like condensates in E. coli by a designed protein based on resilin has been recently described.[50] The protein contained multiple Tyr residues that could be replaced by DOPA residues when DOPA was provided in the growth medium. Formation of the intracellular gel required DOPA incorporation, and the gel was shown to contain chains that had been covalently crosslinked via intracellular oxidation. Our observation that the gel-like condensate formed by 5(DOPA) + FUS-C can be dissolved with guanidinium chloride suggests that this form of assembly does not depend upon extensive covalent crosslinking (Figure 1.18).

Testing the hypothesis that cation-π interactions promote LLPS observed for FUS-C in combination with 1(Tyr) led us to examine 3(F_3Phe). We found that 3(F_3Phe) had a much greater LLPS propensity relative to 2(Phe), which is contrary to the trend that would be predicted if cation-π interactions were energetically important for LLPS in this system. Hydrophobicity assessment via RP-HPLC indicated that F_3Phe is considerably more hydrophobic than Phe. This finding is consistent with an earlier report that pentafluorophenylalanine is more hydrophobic than phenylalanine.[65] Fluorine-bearing alkyl side chains are more hydrophobic than their natural alkyl counterparts.[66–68]

Our observations, in concert with previous studies highlighting the hydrophobicity of side chains bearing fluorine atoms in place of hydrogen,[65–68] raise the possibility that the condensate-spanning network formed by 3(F_3Phe) and FUS-C is stabilized by a web of noncovalent interactions that differs from the web stabilizing the condensate-spanning network formed by FUS-C and 1(Tyr) or 5(DOPA). Specifically, we hypothesize that hydrophobic interactions contribute to phase separation that results from mixing 3(F_3Phe) with FUS-C.

We explored the role of hydrophobicity in our LLPS system by examining 7(Cha). The cyclohexyl side chain lacks an aromatic π-system or H-bonding groups and is quite hydrophobic, but 7(Cha) nevertheless displays a very high LLPS propensity when combined with FUS-C. Since FUS-C contains few residues with nonpolar side chains, it seems unlikely that energetically significant hydrophobic interactions occur between FUS-C and 7(Cha). Instead, we propose that hydrophobically driven interactions occur between cyclohexyl side chains on different molecules of 7(Cha) that are brought into proximity through interactions with FUS-C, and that these hydrophobic interactions stabilize the condensate-spanning network. Coulombic interactions presumably attract 7(Cha) to FUS-C, and the resulting charge neutralization promotes proximity between different molecules of 7(Cha).

Microscopy observations suggest that the physical properties of condensates formed by mixing FUS-C with peptides that bear H-bond-donating side chains, i.e., 1(Tyr) or 5(DOPA), differ from the physical properties of those formed with peptides that bear hydrophobic side chains, such as 7(Cha). Condensate droplets formed with 1(Tyr) or 5(DOPA) persisted for at least 12 hr on the surface of a BSA-passivated microscope slide. In contrast, the condensate droplets formed with 7(Cha), evident after 2 hr on the surface, could no longer be observed after 12 hr. It appeared that these droplets had spread as a layer across the slide surface. Future studies will explore the possibility that changing the profile of non-covalent forces that stabilize condensed phases of this type can result in altered materials properties.

We have established an experimental system to explore the hypothesis that cation-π interactions play a major role in promoting LLPS mediated by Arg-rich and Tyr-rich polypeptides. This study was inspired by extensive analysis of LLPS mediated by FUS and related proteins, which contain an Arg-rich domain and a Tyr-rich domain.[19,20,69] Use of different polypeptides to provide the cationic and aromatic side chains, with the latter component short enough to be accessible via chemical synthesis, has enabled us to build upon earlier experimental designs that identified energetic contributions of isolated cation-π attractions to protein-ligand interactions.[21–28,53] Our results are not consistent with the hypothesis that cation-π interactions play a significant role in LLPS mediated by a designed 28-residue Tyr-rich peptide and an Arg-rich polypeptide corresponding to the last 154 residues of FUS. Instead, our data suggest that H-bond donation by the Tyr side chain hydroxyl group is an energetically important factor in the phase-separation process.

We speculate that the conclusions drawn from these studies are relevant to membraneless organelle formation within cells mediated by FUS and related proteins. Understanding the noncovalent interactions that stabilize the dynamic macrmolecular assemblies within membraneless organelles will provide a foundation for efforts directed toward targeting these assemblies for biomedical applications.[6,7,18,48,49,70,71]

1.5 Materials and Methods

1.5.1 Material Source

Fmoc-protected amino acids, ethyl (hydroxyimino) cyanoacetate (Oxyma), N,N’-diisopropylcarbodiimide (DIC),and hexafluorophosphate azabenzotriazole tetramethyl uronium

(HATU) were purchased from Chem-Impex international. Fluorescein isothiocyanate was purchased from Thermo Scientific. Diisopropylethylamine (DIEA), N,N-dimethylformamide (DMF), dichloromethane (DCM), piperidine, trifluoroacetic acid (TFA), triisopropylsilane (TIPS), thioanisole, acetonitrile, guanidinium chloride (GdmCl), sodium chloride (NaCl), Tris base, Dulbecco's PBS, RNA from torula yeast and BSA (7.5%) were purchased from Millipore Sigma. SPPS reaction vessel syringes and caps were purchased from Torviq. Rink Amide Resin (Low Loading) was purchased from CEM. Four-chamber glass bottom dishes were purchased from Cellvis.

1.5.2 Peptide synthesis, purification, and characterization

All peptides were prepared via CEM MARS6 microwave-assisted solid phase peptide synthesis. 50 μmole Rink Amide ProTide resin (LL) was added to a Torviq solid-phase peptide synthesis vessel with a stir bar. Resin was swelled in 1:1 v/v DMF:DCM for 10 minutes before the synthesis. Fmoc amino acids (4 eq., 0.1 M) were activated with ethyl cyano(hydroxyimino)acetate (Oxyma, 8 eq.) and N,N'-diisopropylcarbodiimide (DIC, 4 eq.), and this solution was added to the reaction vessel. Regular coupling cycles were done at 70°C for 4 minutes. Double extended coupling at 70°C for 12 min was performed for all noncanonical amino acids (TyrOMe, DOPA, F_3Phe, diOMe, Cha). Then the resin was washed 5 times with DMF. The Fmoc protecting group was removed by reacting with 20% v/v piperidine in DMF at 80°C for 2 minutes. For peptides bearing a fluorescein unit, a Fmoc-protected β-alanine residue was added to the N-terminus. After the Fmoc group was removed, the resin-bound peptide was mixed with fluorescein isothiocyanate (2 eq.) and DIEA (8 eq.) in DMF for 6 hours at room temperature in dark.

After all residues had been added, the resin was mixed with the cleavage cocktail (8 mL per 50 μmole), which is 94 % (v/v) TFA, 2.5 % (v/v) H_2O, 2.5 % (v/v) thioanisole, and 1% (v/v) TIPS, at room temperature for 3 hours. The solution containing the crude cleaved peptide was drained from the reaction vessel into a 50-mL falcon tube. Excess TFA was removed under a stream of nitrogen, and about 35 mL of cold diethyl ether was added to precipitate the crude peptide. The crude peptide pellet was isolated by filtration, washed with cold diethyl ether twice, and then dried under nitrogen.

The crude peptide was dissolved in DMSO (2 mL per 50 μmole) for HPLC purification. Dissolved peptide was filtered through a 0.22-micron Nylon filter and purified on an Agilent preparative HPLC system with a C18 CSH Prep column (5 μm, 19x250 mm, solvent A = H_2O + 10 mM NH4OH, solvent B = acetonitrile + 10 mM NH4OH, flow rate = 13 mL/min). Peptide masses were determined using MALDI-TOF-MS and ESI-MS, and purity was characterized by analytical UPLC (detection at 220 nm). HPLC fractions containing pure peptide were combined, and this solution was lyophilized (freeze-dried) for long-term storage. The lyophilized powder was dissolved in H_2O for all experiments.

1.5.3 Protein expression, and purification

A plasmid containing TEV protease gene pDZ2087 was a gift from David Waugh (Addgene plasmid # 92414). TEV protease was expressed and purified as previous reported. Purified TEV protease was aliquoted and stored at -20 °C for future use.[32]

The DNA fragments encoding two constructs: the wildtype FUS C-terminus (372-526) and FUS C-terminus (372-526) with a Cys inserted between R524 and P525, with a TEV protease cleavage site at the N-terminus (sequence: ENLYFQG) and an additional 6 histidine residues at the C-terminus (sequence: HHHHHH), were ordered from Twist Bioscience, and cloned into AvaI and EcoRI sites of vector pMAL-c5x (this vector contains the sequence for maltose binding protein). The resulting plasmids were transformed into E. coli strain BL21(DE3), and cells were grown at 37 °C in 2XYT with 100 μg/L ampicillin. At OD600 = 0.4, overexpression was induced using 1 mM IPTG at 18 °C for 12 hr. After cells were harvested through the centrifugation, cells were resuspended and lysed in 50 mM Tris buffer pH 7.5, 1000 mM NaCl, 0.5 mM TCEP, 1 mM MgCl2, 200 μg/mL lysozyme, 10 μg/mL DNase I, and cOmplete protease inhibitor using sonication. The protein was purified over a HisPrep FF column, and eluted with 50 mM Tris buffer pH 7.5, 1000 mM NaCl, 0.5 mM TCEP and 500 mM imidazole. To remove the excess imidazole, the protein solution was dialyzed against 10 volumes of buffer containing 50 mM Tris buffer pH 7.5, 150 mM NaCl, 0.5 mM TCEP. The protein was further purified using a Superdex 200 pg size exclusion column. The purified FUS C-terminus (372-526) with Cys inserted was further treated with 2 equivalents of maleimide-TAMRA and purified with another size-exclusion column to generate TMR-FUS.

1.5.4 Confocal Microscopy

Glass-bottom confocal dishes were passivated with 1% BSA in PBS and washed three times with MilliQ water. Peptide solutions were prepared by mixing 99 mol% unlabeled peptide and 1 mol% fluorophore-labeled peptide in 50 mM Tris buffer pH 7.5, 150 mM NaCl. The protein solution MBP-FUS-C was treated with 5 mol% TEV protease in 50 mM Tris buffer pH 7.5, 150 mM NaCl, 0.5 mM TCEP at room temperature for 6 hours. The cleavage product was then mixed with the peptide solution. Peptide and protein solutions were mixed (3 times pipetting) in the confocal dish, and the solution was incubated at room temperature. We found reproducibility of results depended on strict consistency in the mixing/pipetting protocol. Images were acquired on a Nikon AXR confocal microscope equipped with a 60x oil immersed objective at different timepoints. A 488-nm laser was used to excite the Fl-peptide samples; a 561-nm laser was used to excite the TMR-FUS samples. Images were processed in ImageJ.[69]

1.5.5 Fluorescence Recovery After Photobleaching (FRAP)

Peptide + FUS-C solutions containing 1 mol% labeled peptide or 1 mol% labeled protein were prepared as in section 4 above. The pre-and post-bleaching images were acquired on a Nikon AXR with the low laser power setting. A region of interest (ROI) with diameter = 1 μm was selected at the center of a relatively large droplet. Bleaching was performed with 100% 488-nm laser power for 8 seconds to reach >80% fluorescence reduction. The fluorescence recovery was monitored with the smallest time interval (2 sec).[70,71]

1.5.6 Determination of peptide concentration in the dilute phase

Phase separation samples were prepared by combining 50 μM peptide (10 mol% fluorescein-labeled) and 5 μM FUS-C in 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP. The total volume (V_0) was 0.4 mL. These samples were incubated for 30 minutes and then subjected to centrifugation (13,000 g, 30 min) to force the condensed phase to the bottom of the tube. A small aliquot (~ 10 μL) of the supernatant (assumed to be pure dilute phase) was then carefully removed from the top of the sample. Absorbance at 488 nm, which arises from the fluorescein moiety, was recorded for this sample via UV-Vis spectroscopy. Based on an extinction coefficient of 75,000 M^-

^{1}cm[72],5 the absorbance at 488 nm was used to calculate the concentration of the peptide in the dilute phase.

1.5.7 Measurement of droplet area

Phase separation samples were prepared by mixing appropriate amounts of peptide (1 mol% fluorescein-labeled) and 5 µM FUS-C in 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP. The mixture was transferred to the confocal plate (1% BSA coated). The sample was allowed to stand for 30 min to allow droplets to sediment. Images were taken afterwards. N = 4 independent replicates (two measurements per replicate). Images were processed using ImageJ.[69]

1.5.8 Circular Dichroism

CD experiments were conducted at 25 °C on a JASCO J-1500 CD spectrometer. Peptide samples were diluted to 50 µM in PBS, pH 7.4, and then transferred to a 1-mm quartz cuvette. The CD spectrum was measured from 260 to 200 nm with 0.1 nm intervals at 4 second digital integration time (D.I.T) at 100 nm/min scanning speed. The peptide concentrations were carefully remeasured for CD experiments, and the accurate values were used for molar ellipticity calculations. The data are presented as mean residue ellipticity in (deg•cm2•M^{-1}) averaged for ten spectral scans.

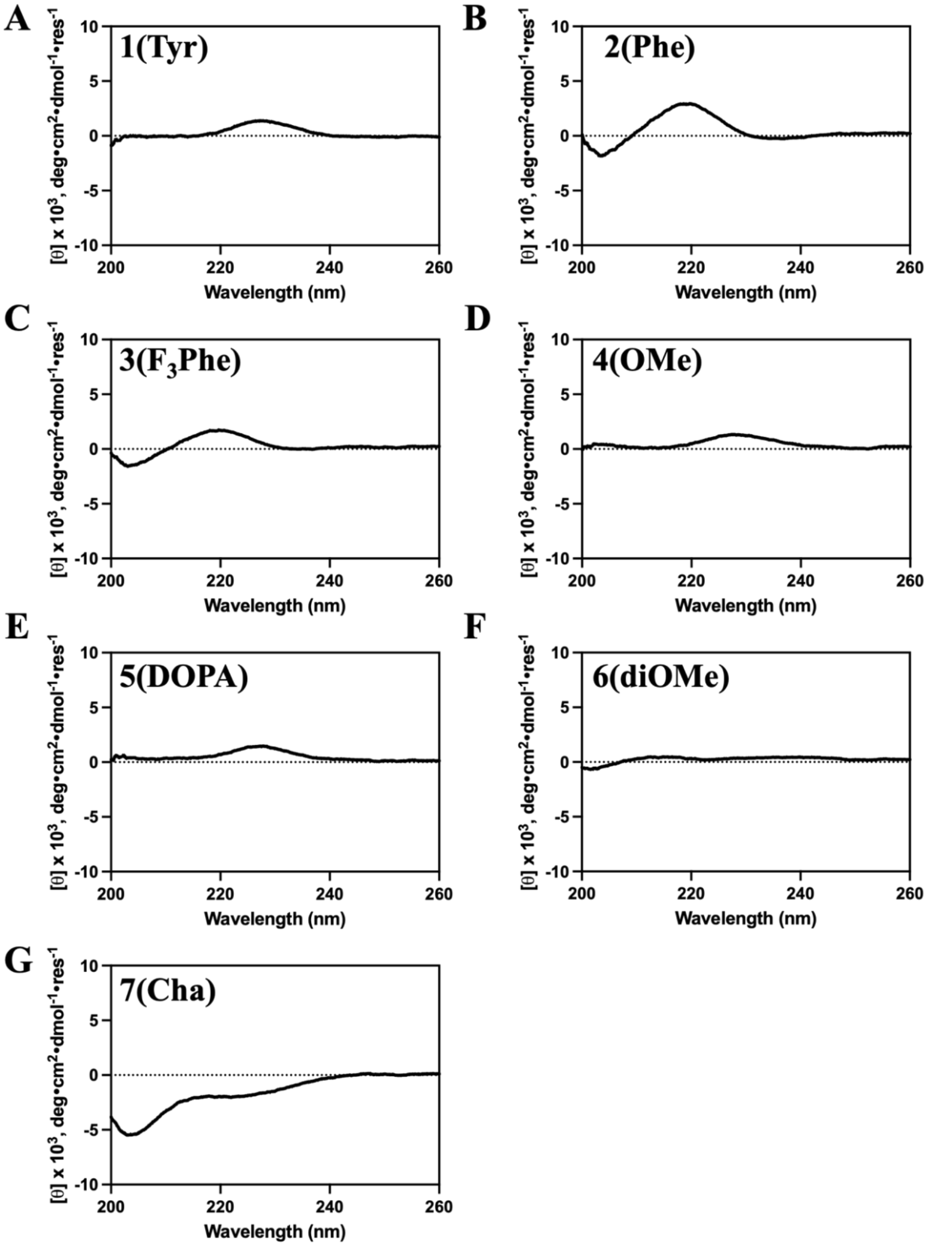


Figure 1.15. Far-UV CD for the seven peptides discussed in the main text, each at 50 μM in PBS, pH 7.4 at 25°C. The Tris buffer used for other experiments is not compatible with CD (very poor signal-to-noise ratio below 210 nm). Data points with high tension (HT) over 700 V have been excluded from the analysis. We conclude that each of the peptides in this study remains largely unfolded in a neutral pH buffer, as expected in light of the high glycine content (15 of 28 residues).

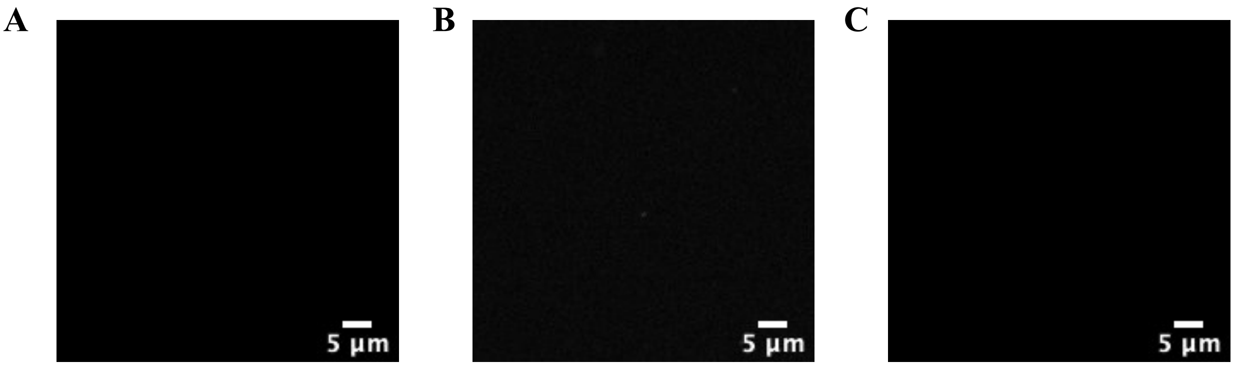


Figure 1.16. Representative confocal images of 50 µM **1a** (**A**), **1b** (**B**) or **1c** (**C**) when mixed with 5 µM FUS-C (1 mol% TAMRA-FUS-C). Scale bar = 5 µm. None of these three peptides induces phase separation under these conditions. Buffer condition: 50 mM Tris, pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

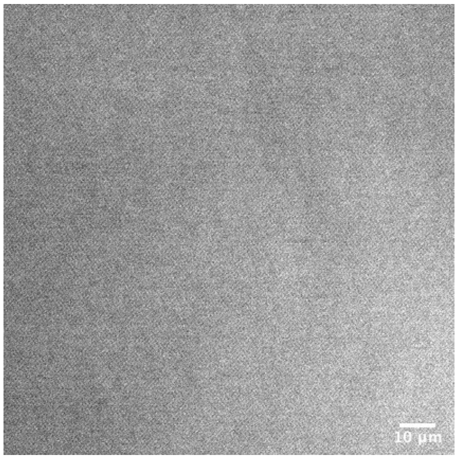


Figure 1.17. Representative DIC image of 5 µM maltose binding protein with 50 µM **1(Tyr)**. Scale bar = 10 µm. No coacervates or aggregates were observed. Solutions used for phase separation studies contained FUS-C along with the maltose binding protein (which had been detached by TEV protease treatment). We conclude that maltose binding protein does not participate in the phase separation behavior observed for peptide + FUS-C pairings. Buffer condition: 50 mM Tris, pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

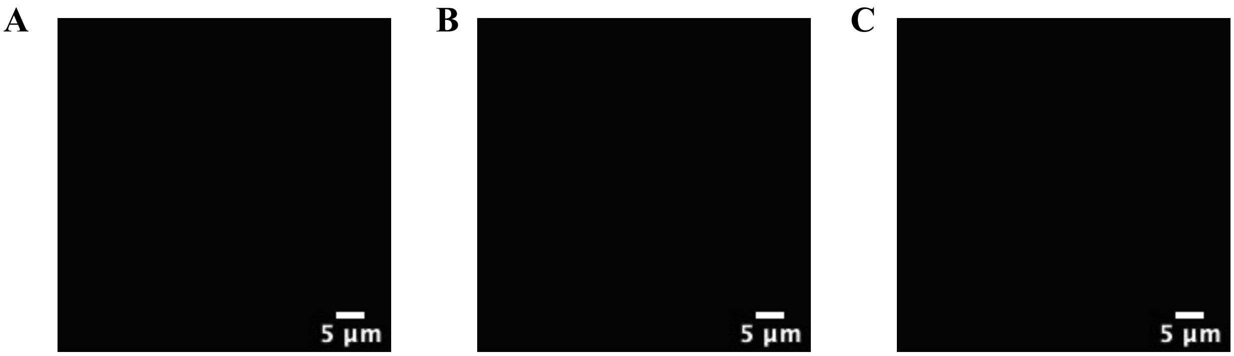


Figure 1.18. Representative confocal images after 6 M guanidine HCl treatment of droplets formed between 5 µM FUS-C and 50 µM **1(Tyr)** (1 mol% fluorescein-**1(Tyr)**, **A**), **5(DOPA)** (1 mol% fluorescein-**5(DOPA)**, **B**), or **7(Cha)** (1 mol% fluorescein-**7(Cha)**, **C**). Scale bar = 5 µm. No droplets were observed after the treatment with guanidine HCl. For the sample containing **5(DOPA)**, this result indicates a lack of large-scale covalent crosslinking. Buffer condition: 50 mM Tris, pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

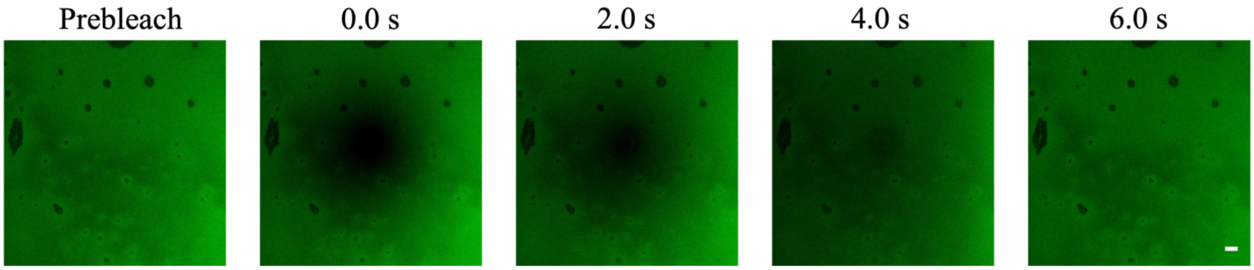
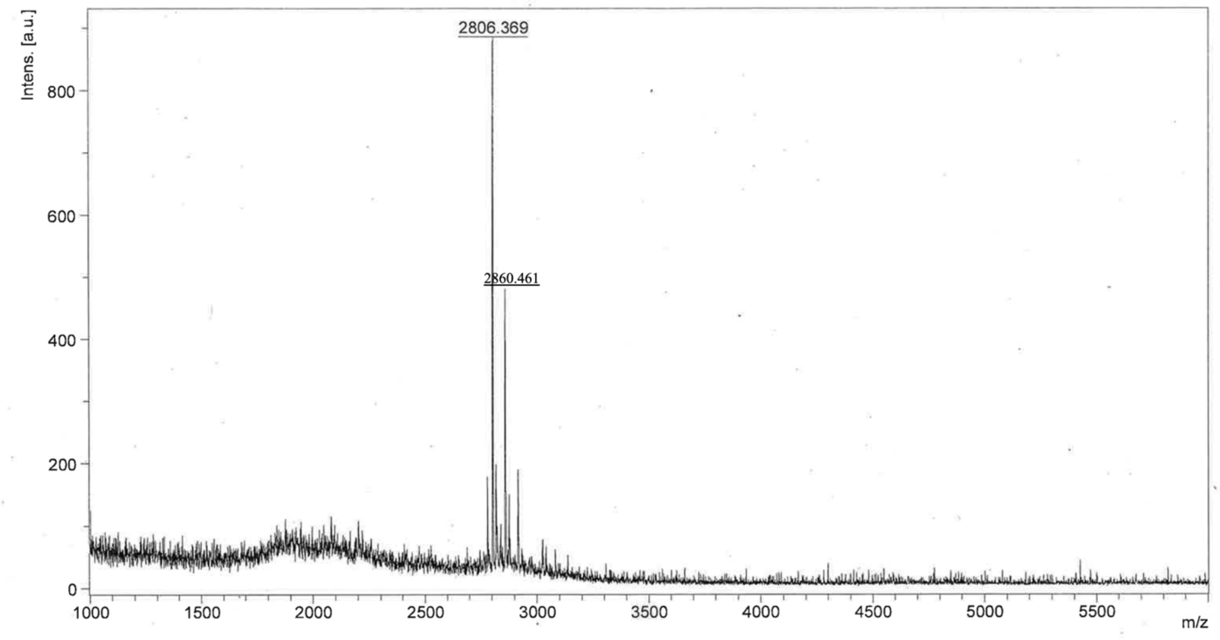
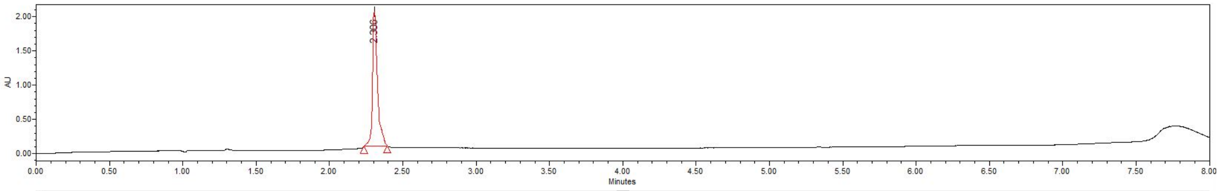


Figure 1.19. Representative confocal micrographs of coacervates formed with 5 µM FUS-C and 50 µM **7(Cha)** (1 mol% fluorescein-**7(Cha)**) at the 12-hour timepoint. No clear boundary for coacervate droplets was observed. Instead, the condensed phase appeared to have spread across the surface. These images show results of a FRAP experiment. Laser irradiation bleaches a region of the condensed phase covering the surface, and recovery of fluorescence is observed. Scale bar = 2 µm. Buffer condition: 50 mM Tris, pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

1.6 MALDI-TOF-MS Spectra and UPLC Data for Synthetic Peptides





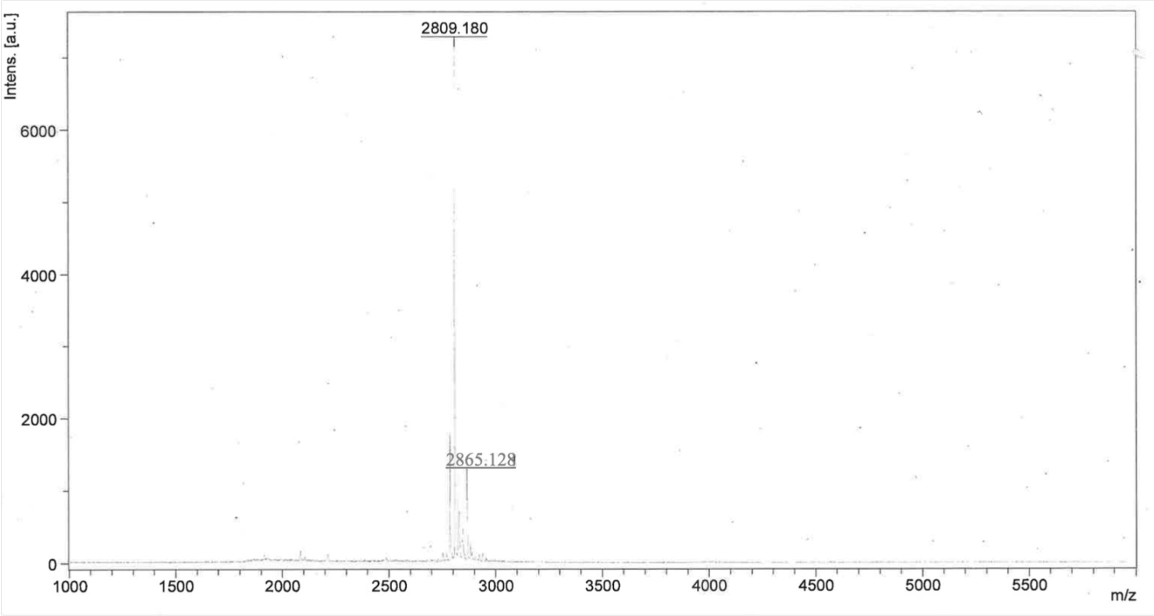
1 (Tyr): **`H₂N-GGGYEEEYGGGYGGGYEEEYGGGYGGGY-NH₂`**

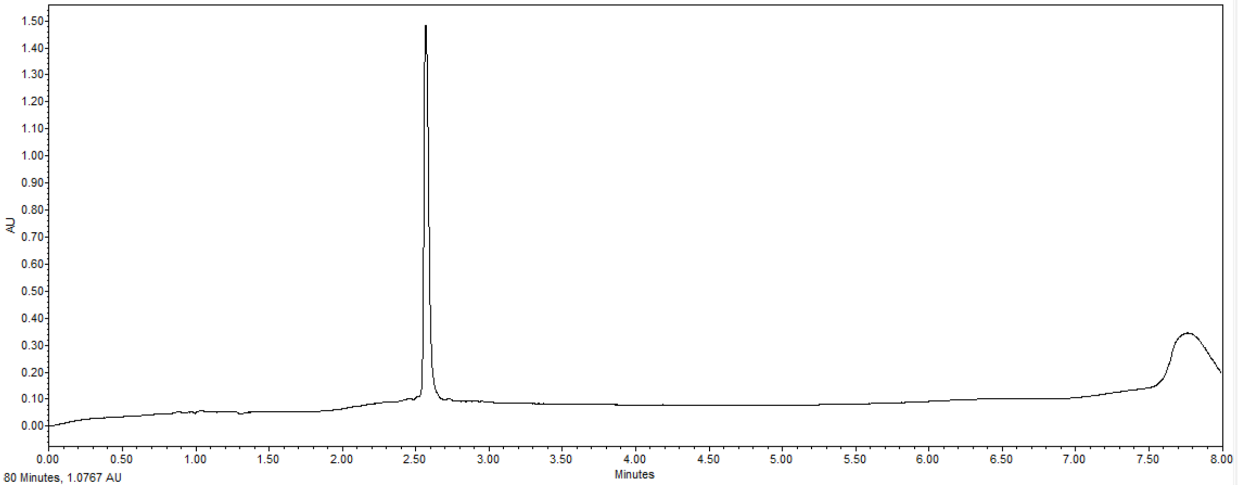
<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+NH_4]^+$ = 2806.137 $[M+4NH_4]^+$ = 2860.137

observed monoisotopic $[M+NH_4]^+$ = 2806.369 $[M+4NH_4]^+$ = 2860.461

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





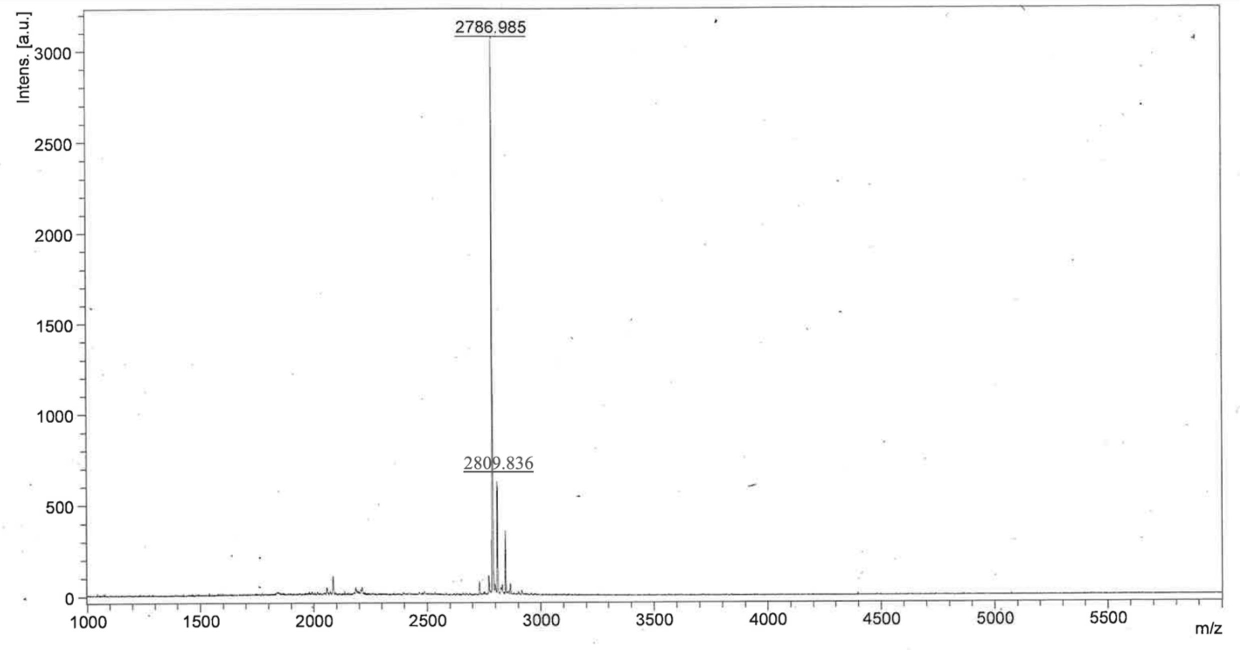
1a: **H_2N-GGGYKKKYGGGYGGGYEEEYGGGYGGGY-NH_2**

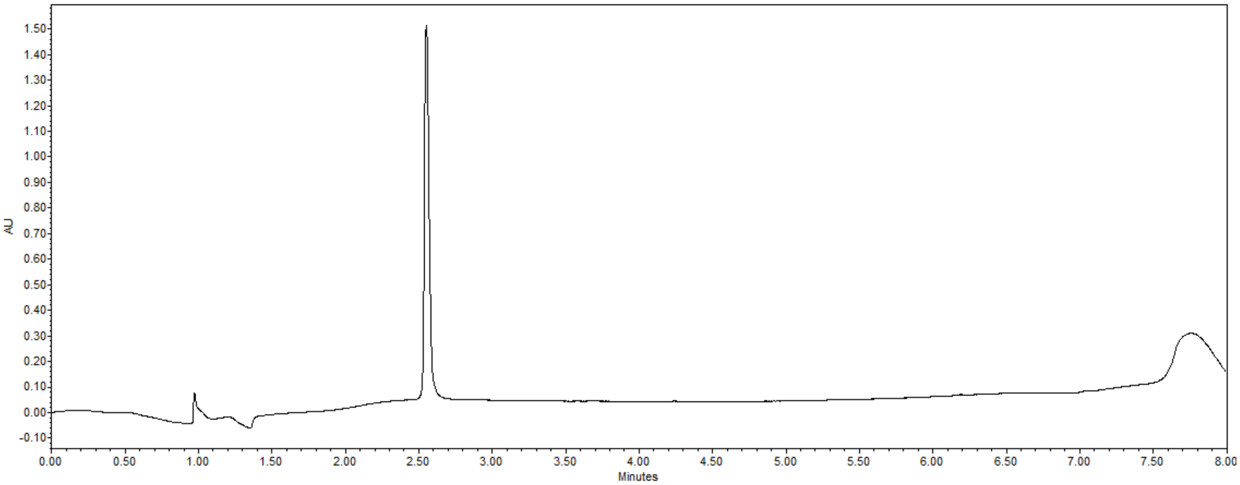
MALDI-TOF-MS: calculated monoisotopic $[M+Na]^+$ = 2809.886 $[M+DMSO+H]^+$ = 2865.886

observed monoisotopic $[M+Na]^+$ = 2809.180 $[M+DMSO+H]^+$ = 2865.128

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





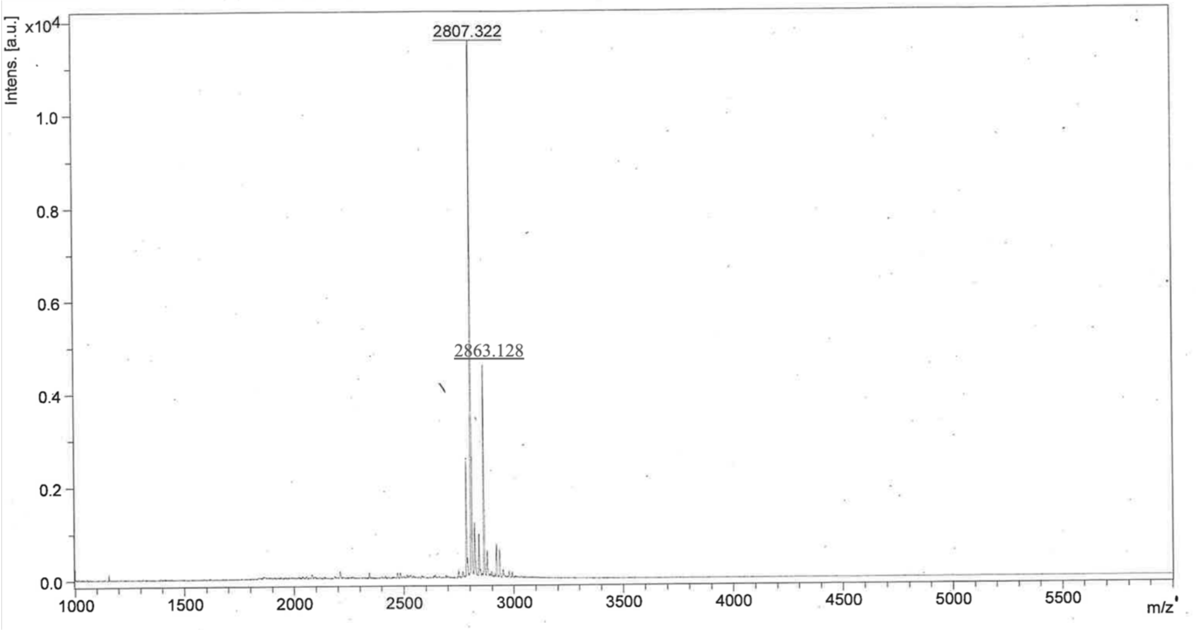
1b: **H_2N-GGGYEEEYGGGYGGGYKKKYGGGYGGGY-NH_2**

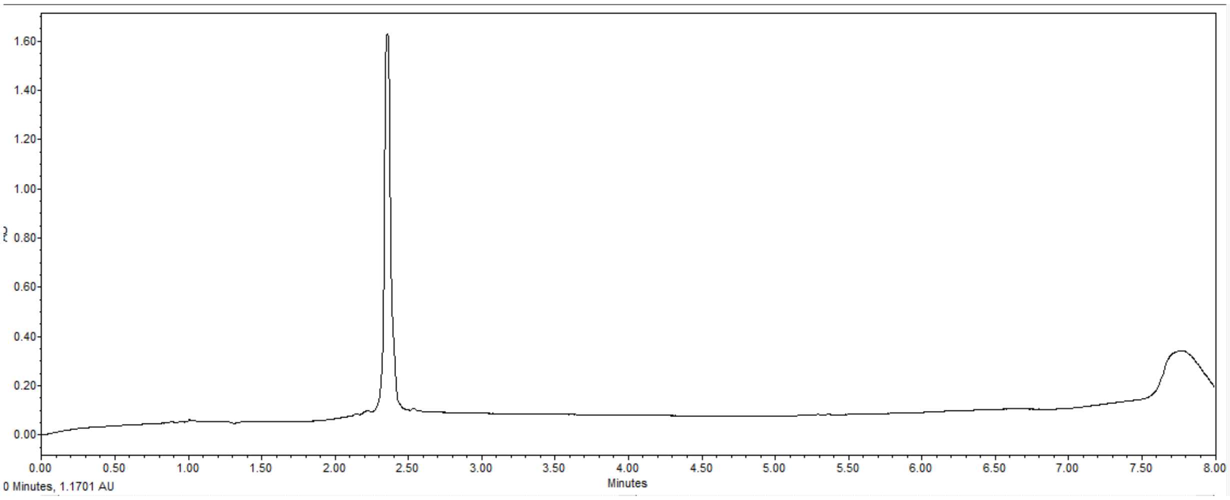
MALDI-TOF-MS: calculated monoisotopic $[M+H]^+ = 2786.886$ $[M+Na]^+ = 2809.886$

observed monoisotopic $[M+H]^+ = 2786.985$ $[M+Na]^+ = 2809.836$

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





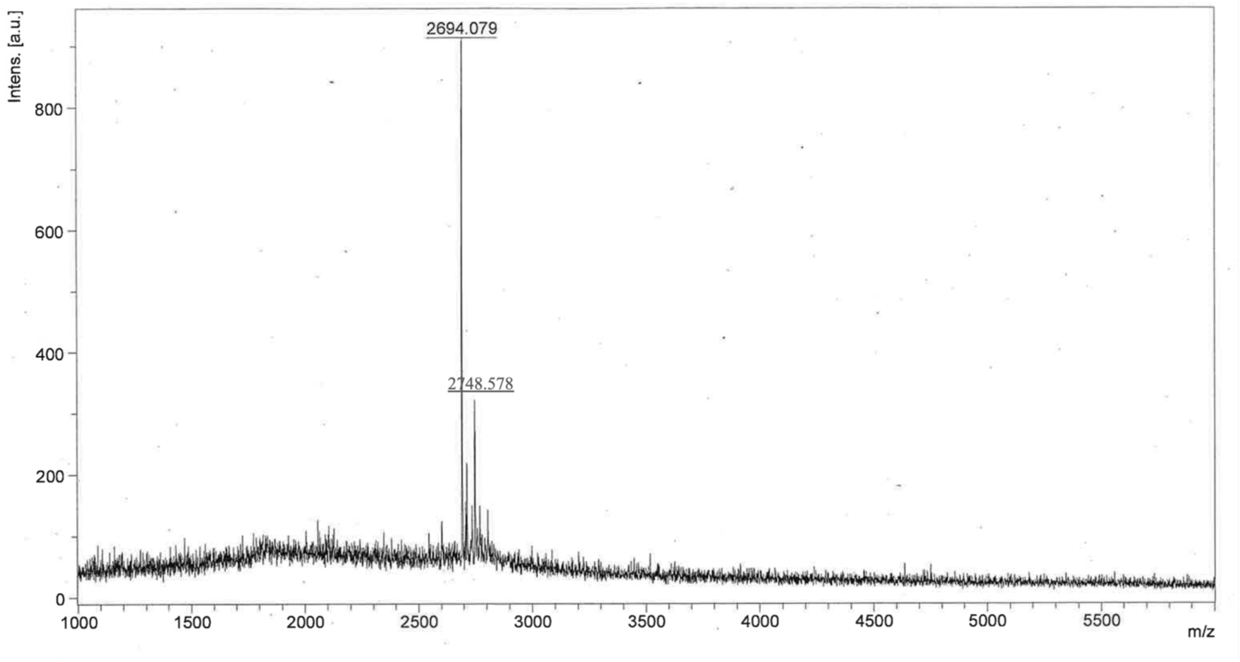
1c: **H_2N-GGGYKKKYGGGYGGGYKKKYGGGYGGGY-NH_2**

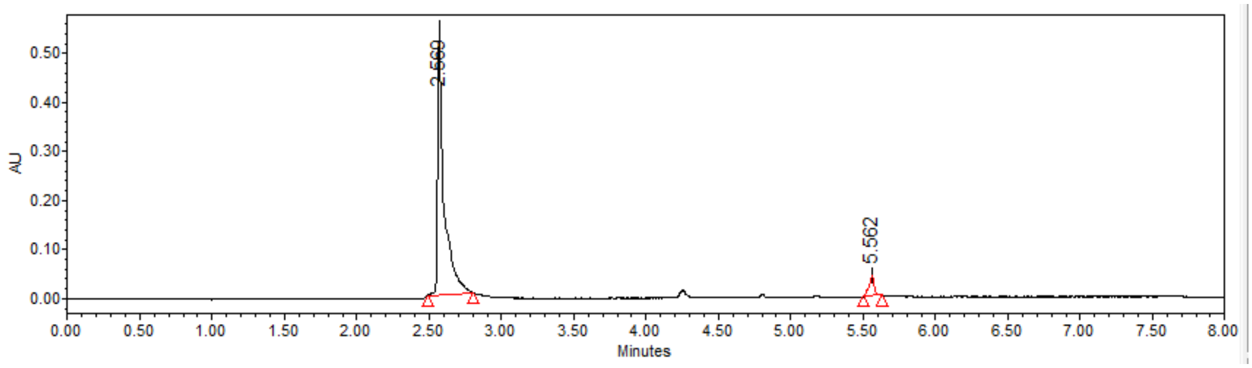
MALDI-TOF-MS: calculated monoisotopic $[M+Na]^+ = 2807.062$ $[M+DMSO+H]^+ = 2863.062$

observed monoisotopic $[M+Na]^+ = 2807.322$ $[M+DMSO+H]^+ = 2863.128$

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





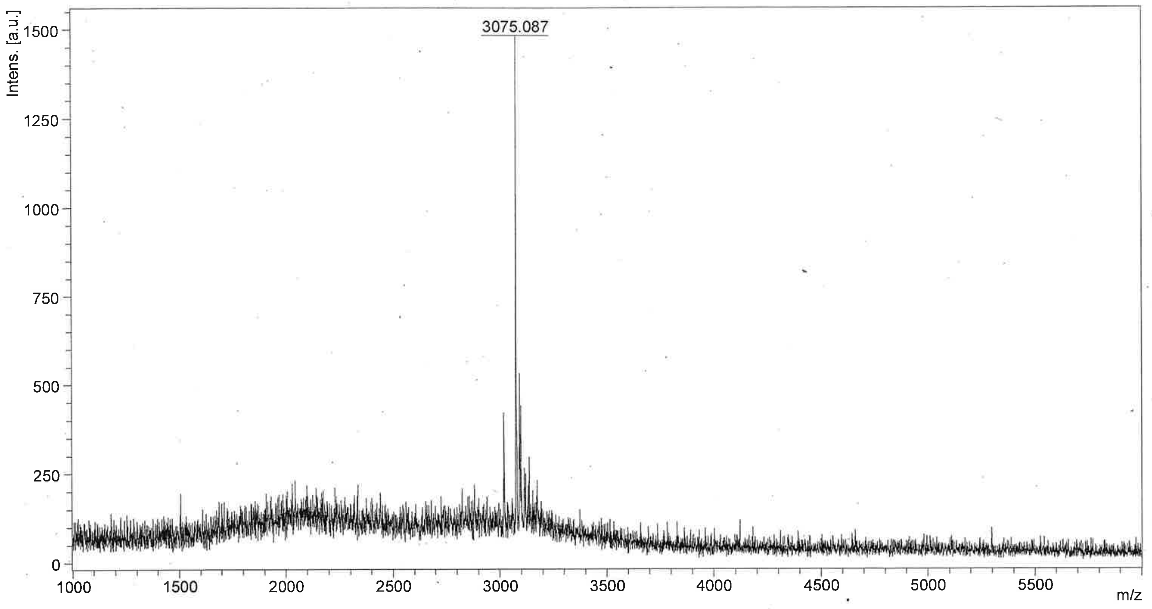
2(Phe): H_2N–GGGFEEEFGGGFGGGFEEEFGGGFGGGF–NH_2

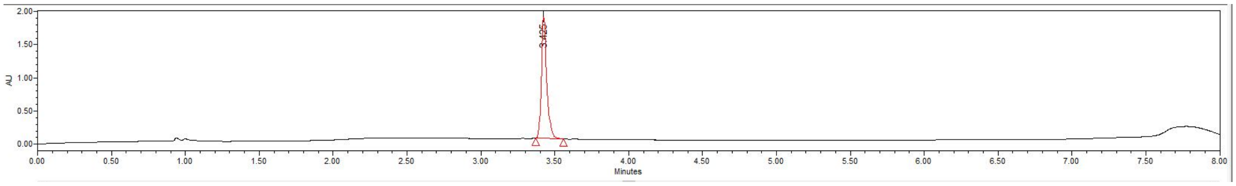
MALDI-TOF-MS: calculated monoisotopic $[M+NH_4]^+ = 2694.072$ $[M+4NH_4]^+ = 2748.072$

observed monoisotopic $[M+NH_4]^+ = 2694.079$ $[M+4NH_4]^+ = 2748.578$

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CCSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity = 94.1 %





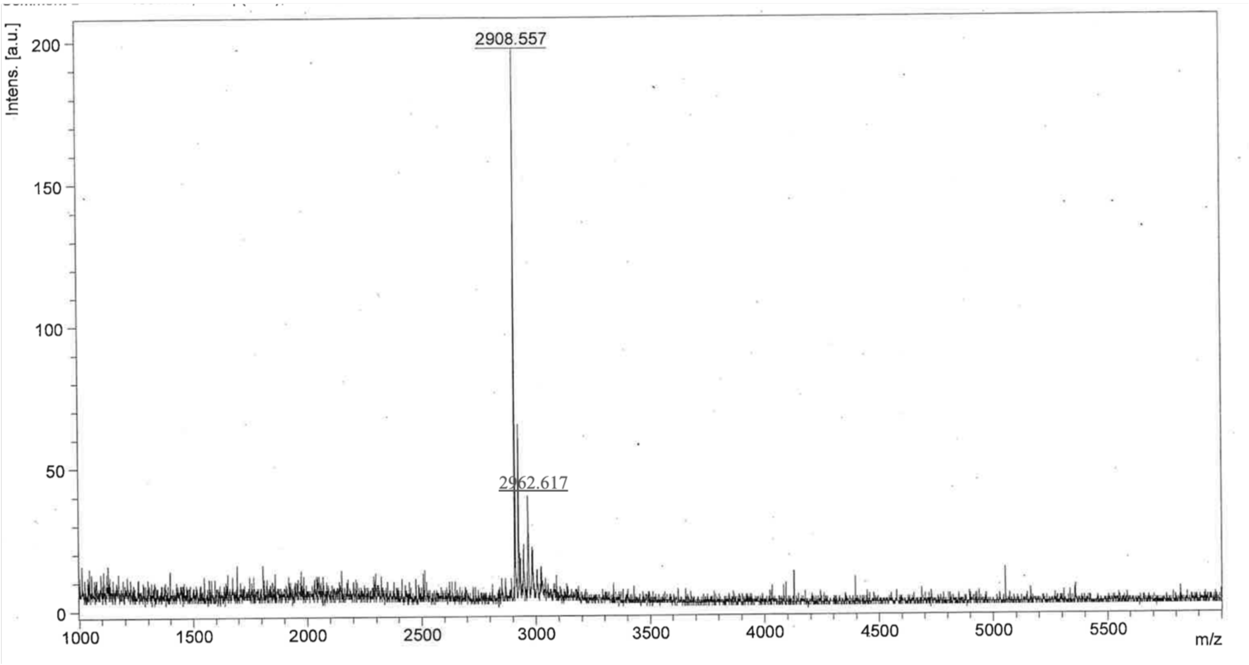
3(F_3Phe): `H₂N-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH₂  (X = F₃Phe)`

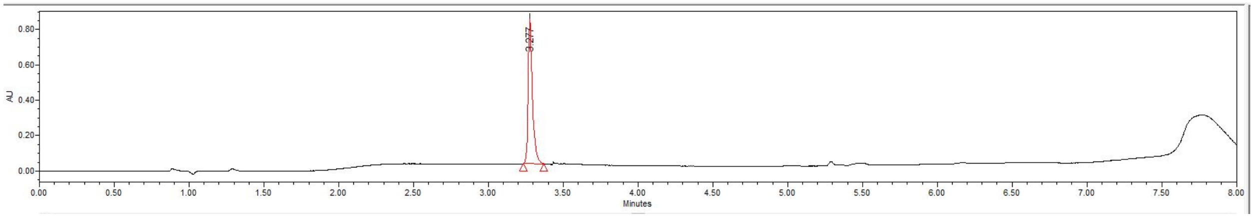
<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+NH_4]^+$ = 3073.640

observed monoisotopic $[M+NH_4]^+$ = 3075.087

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity = 95.09 %





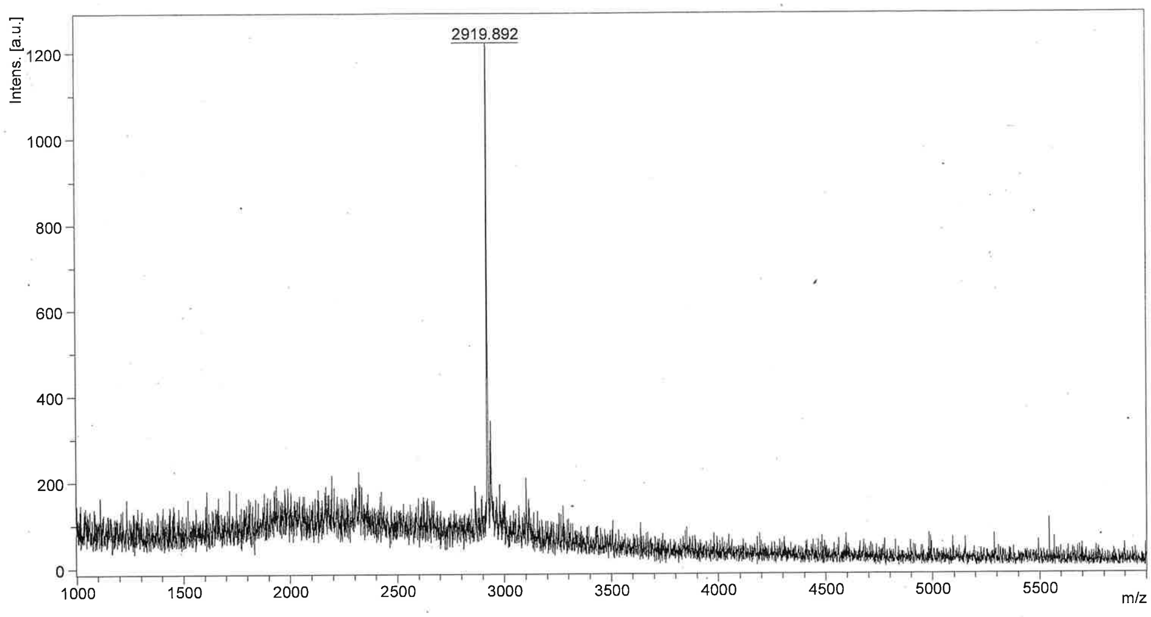
4(TyrOMe): `H2N-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH2  (X = TyrOMe)`

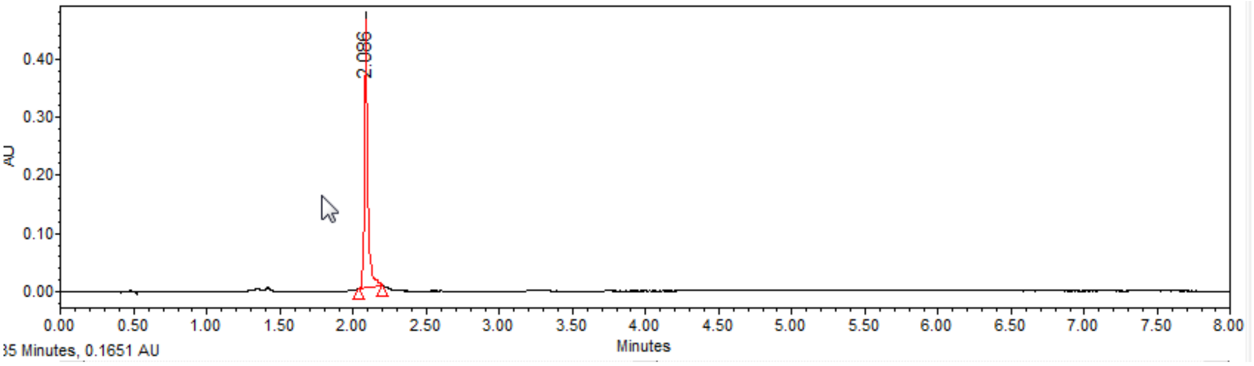
MALDI-TOF-MS: calculated monoisotopic $[M+NH_4]^+$ = 2906.020 $[M+4NH_4]^+$ = 2962.020

observed monoisotopic $[M+NH_4]^+$ = 2908.557 $[M+4NH_4]^+$ = 2962.617

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





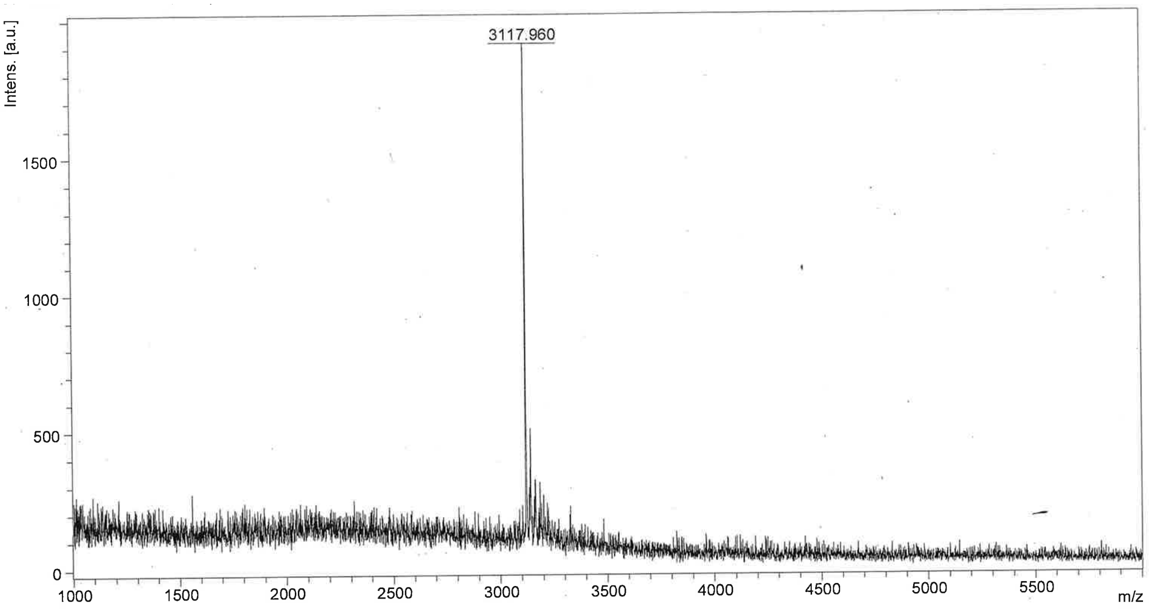
5(DOPA): **`H₂N-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH₂  (X = DOPA)`**

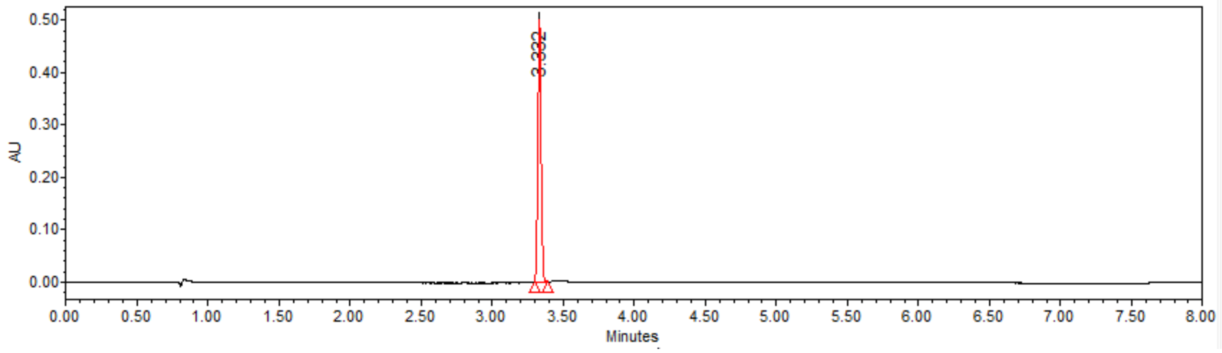
<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+NH_4]^+ = 2919.83$

observed monoisotopic $[M+NH_4]^+ = 2919.892$

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





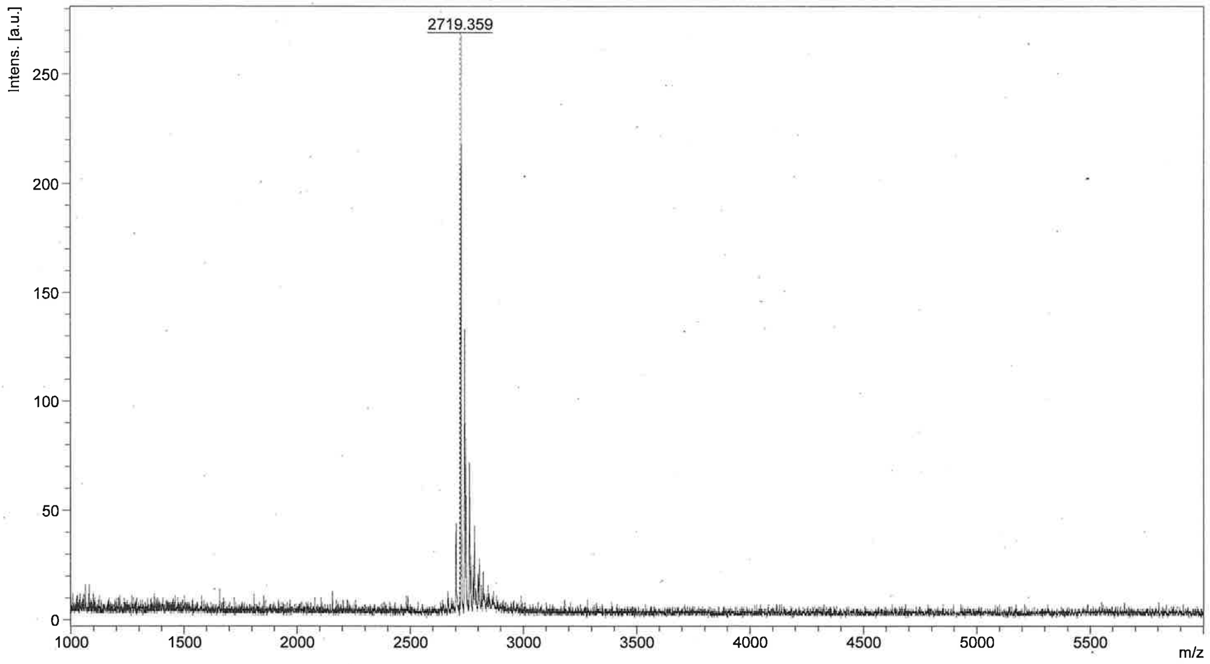
6(diOMe): **`H₂N-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH₂  (X = diOMe)`**

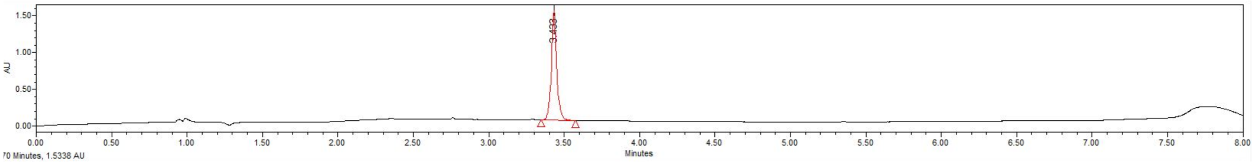
<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+NH_4]^+ = 3117.96$

observed monoisotopic $[M+NH_4]^+ = 3116.20$

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





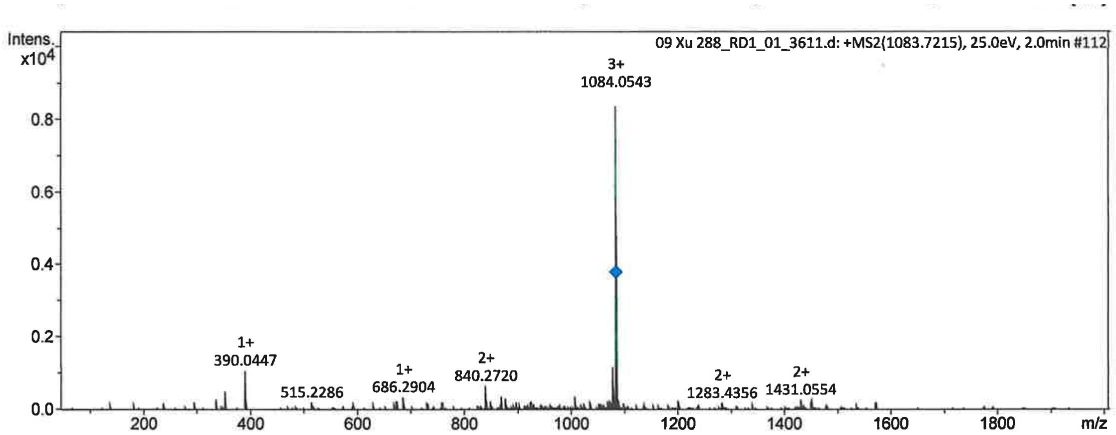
7(Cha): `H₂N-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH₂  (X = Cha)`

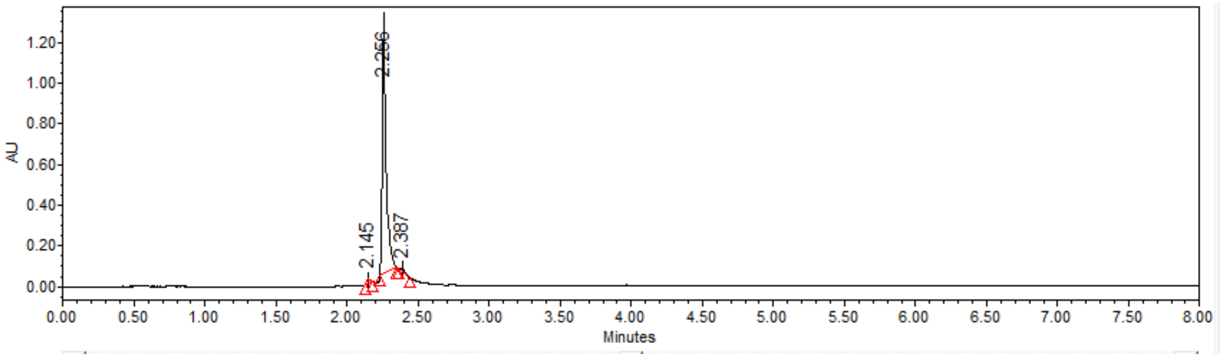
<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 2719.41

observed monoisotopic $[M+H]^+$ = 2719.359

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





FITC-1(Tyr): FITC-(βAla)-GGGYEEEYGGGYGGGYEEEYGGGYGGGY-NH$_2$

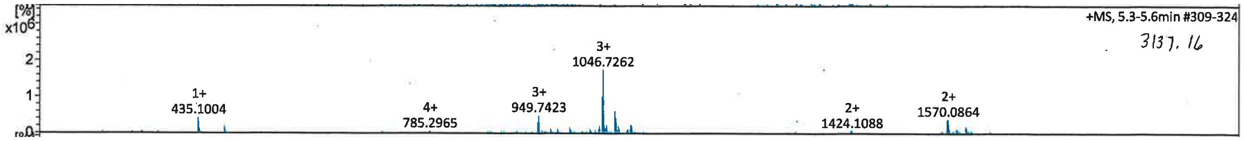
ESI-MS: calculated monoisotopic $[M+3H]^{3+}$ = 1084.063

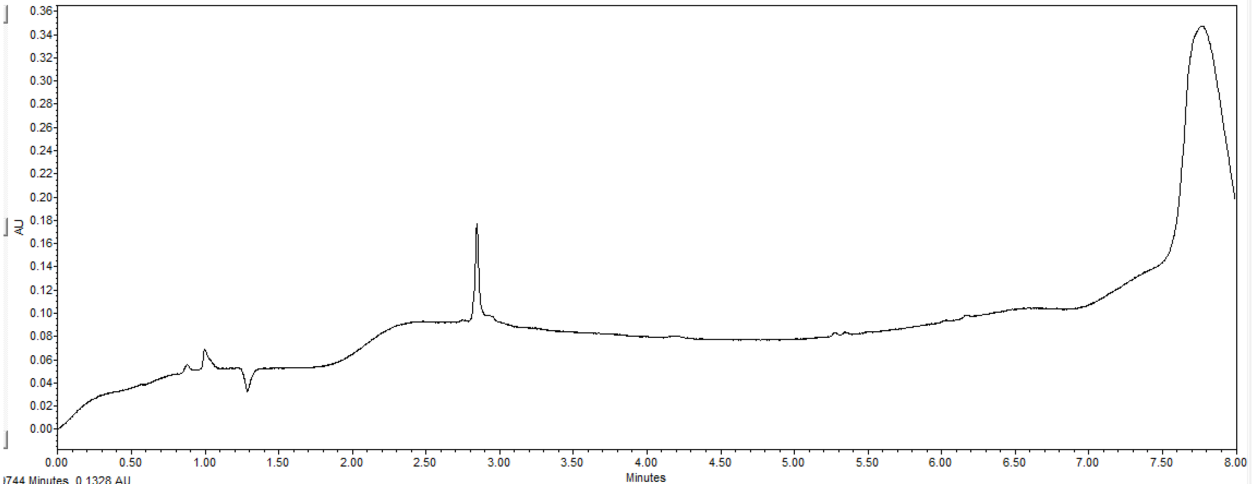
observed monoisotopic $[M+3H]^{3+}$ = 1084.054

(*ESI-MS was used for this peptide due to high laser fragmentation of the fluorescein ground under MALDI-TOF-MS condition)

UPLC: H$_2$O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity = 97.38 %





FITC-2(Phe): FITC-(βAla)-GGGFEEEFGGGFGGGFEEEFGGGFGGGF-NH$_2$

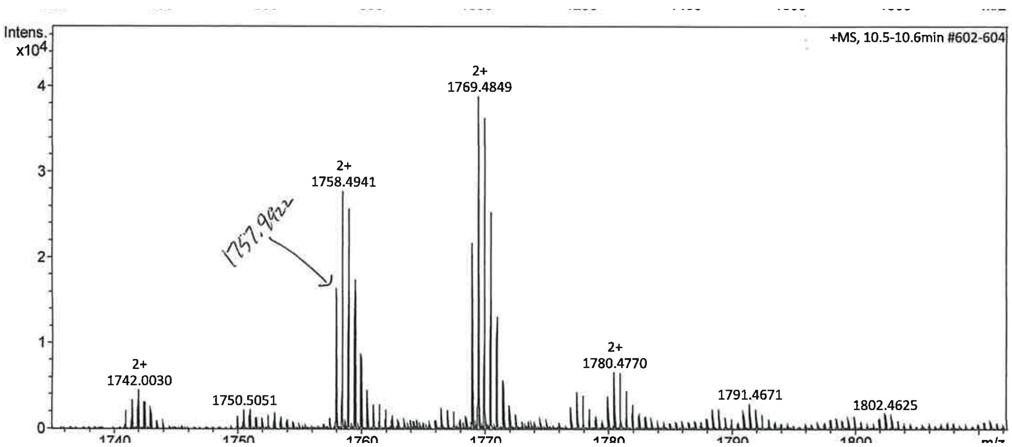
ESI-MS: calculated monoisotopic $[M+3H]^{3+} = 1046.0533$

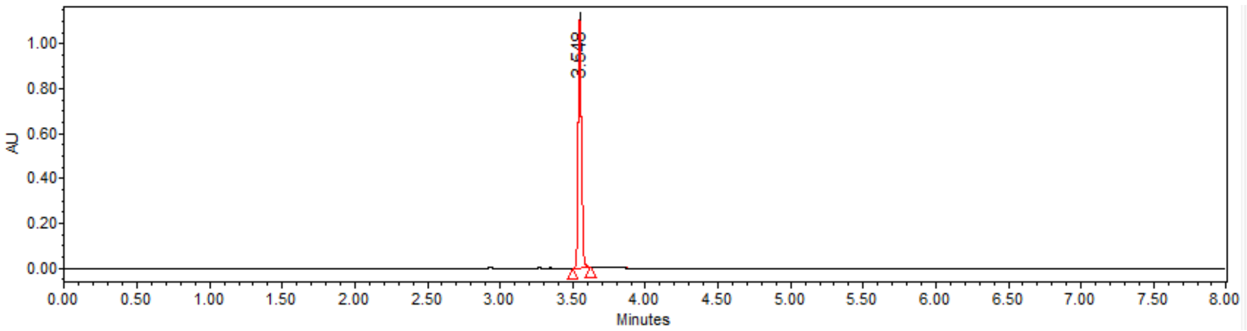
observed monoisotopic $[M+3H]^{3+} = 1046.7262$

(*ESI-MS was used for this peptide due to high laser fragmentation of the fluorescein ground under MALDI-TOF-MS condition)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity = 95.24 %





FITC-3(F_3Phe): `FITC-(βAla)-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH`$_2$ (X = F_3Phe)

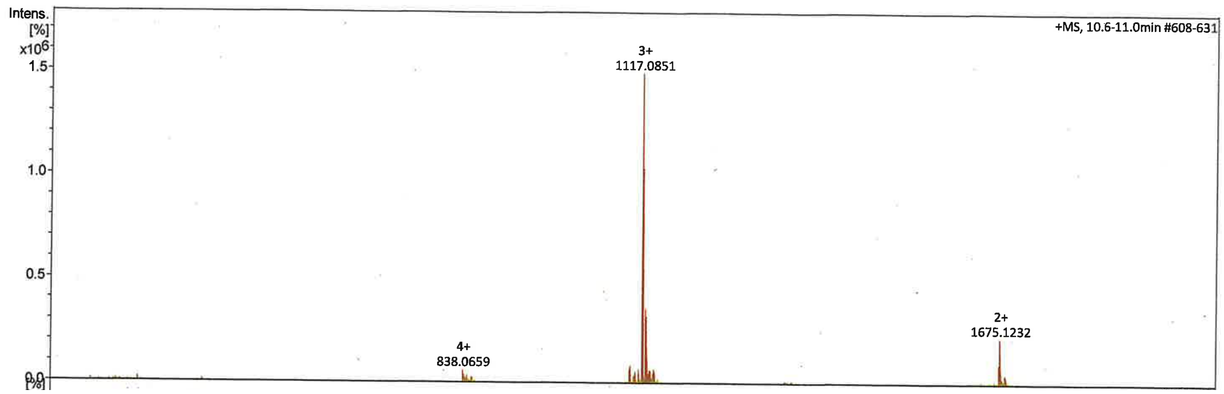
ESI-MS: calculated monoisotopic $[M+2H]^{2+}$ = 1757.983 $[M+Na+2H]^{2+}$ = 1769.483

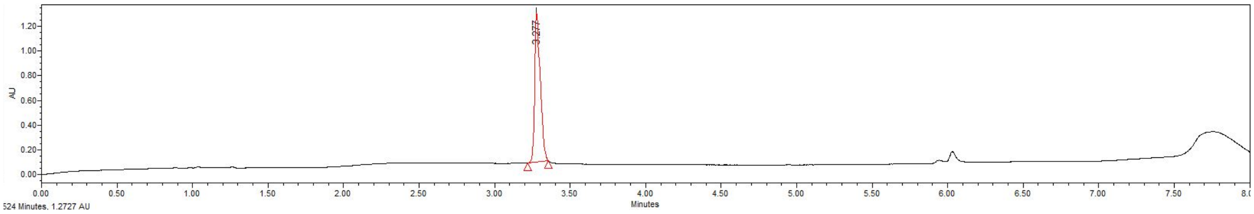
observed monoisotopic $[M+2H]^{2+}$ = 1757.992 $[M+Na+2H]^{2+}$ = 1769.485

(*ESI-MS was used for this peptide due to high laser fragmentation of the fluorescein ground under MALDI-TOF-MS condition)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å, 1.7 µm, 2.1 x 150 mm) column.

Purity > 99.0 %





FITC-4(TyrOMe): `FITC-(βAla)-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH`$_2$ `(X = TyrOMe)`

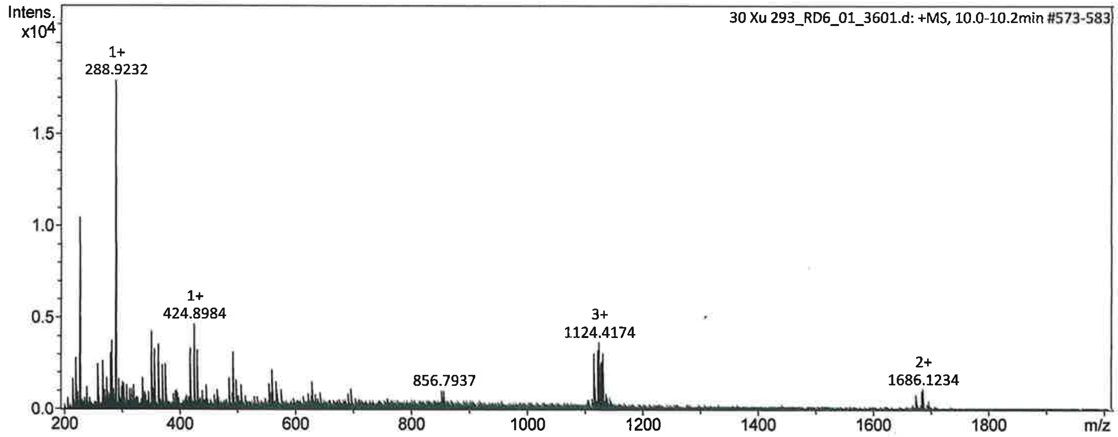
ESI-MS: calculated monoisotopic $[M+3H]^{3+} = 1116.412$ $[M+2H]^{2+} = 1674.118$

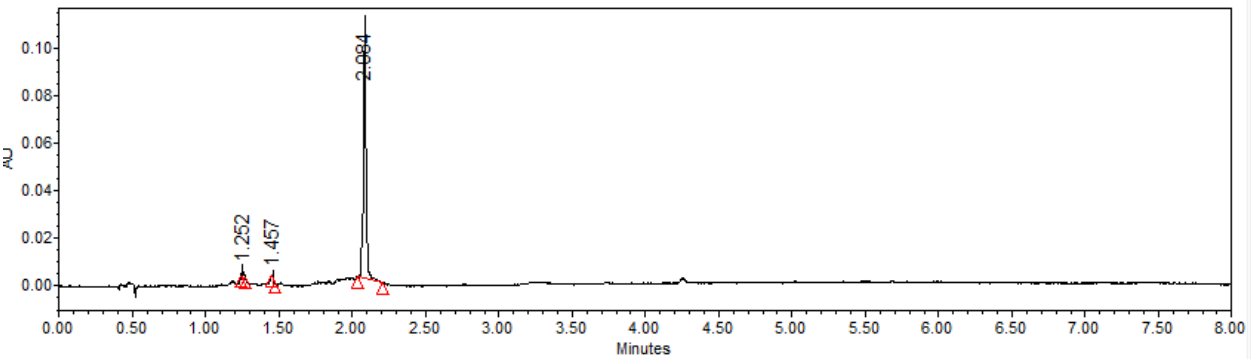
observed monoisotopic $[M+3H]^{3+} = 1117.085$ $[M+2H]^{2+} = 1675.123$

(*ESI-MS was used for this peptide due to high laser fragmentation of the fluorescein ground under MALDI-TOF-MS condition)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å, 1.7 µm, 2.1 x 150,mm) column.

Purity > 99.0 %





FITC-5(DOPA): FITC-(βAla)-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH_2 (X = DOPA)

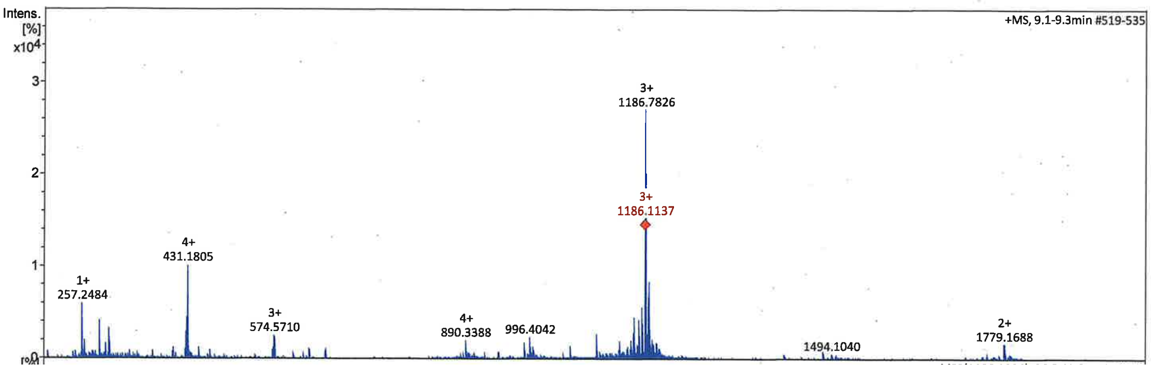
ESI-MS: calculated monoisotopic $[M+3H]^{3+}$ = 1123.748

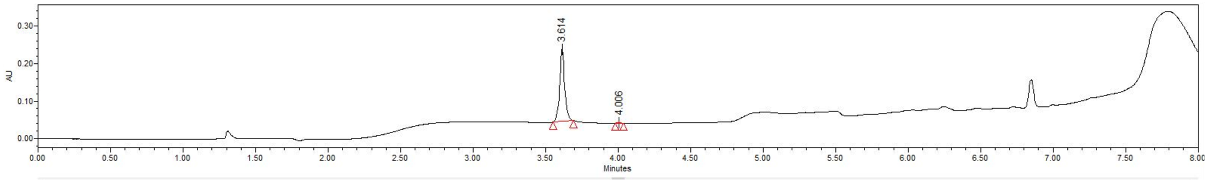
observed monoisotopic $[M+3H]^{3+}$ = 1124.417

(*ESI-MS was used for this peptide due to high laser fragmentation of the fluorescein ground under MALDI-TOF-MS condition)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å, 1.7 µm, 2.1 x 150mm) column.

Purity = 97.97 %





FITC-6(diOMe): `FITC-(βAla)-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-`NH_2 `(X = diOMe)`

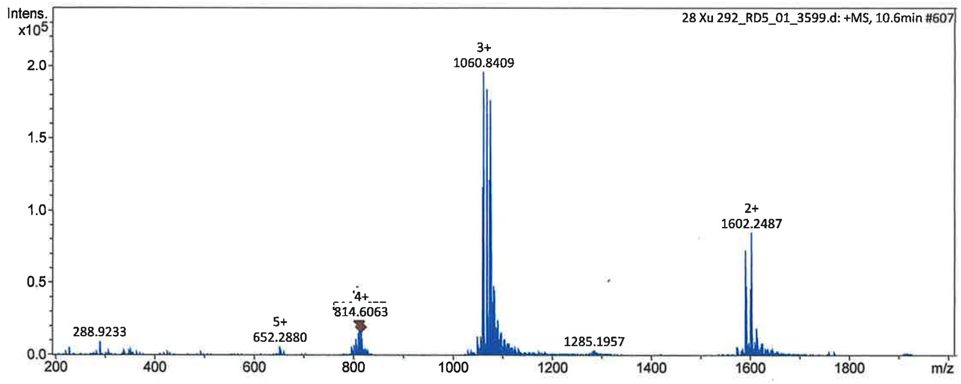
<u>ESI-MS</u>: calculated monoisotopic $[M+3H]^{3+}$ = 1186.100 $[M+4H]^{4+}$ = 889.825

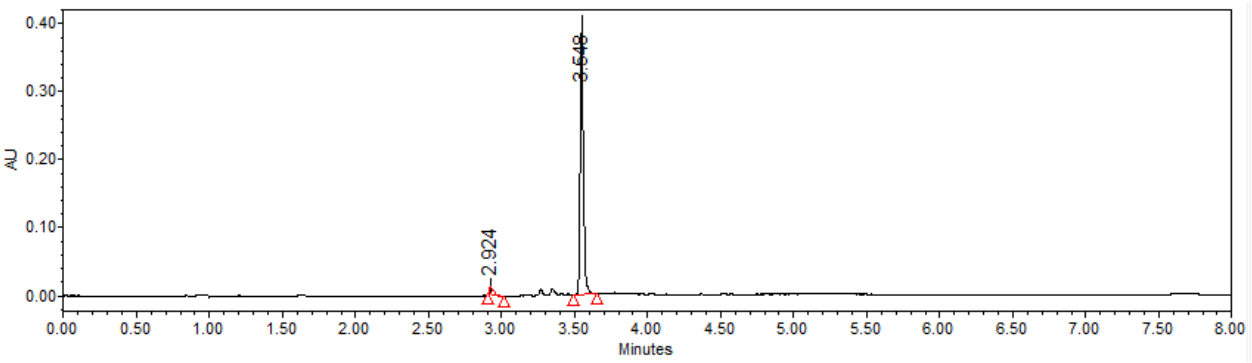
observed monoisotopic $[M+3H]^{3+}$ = 1186.114 $[M+3H]^{3+}$ = 890.339

(*ESI-MS was used for this peptide due to high laser fragmentation of the fluorescein ground under MALDI-TOF-MS condition)

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

Purity = 98.12 %





FITC-7(Cha): FITC-(βAla)-GGGXEEEXGGGXGGGXEEEXGGGXGGGX-NH$_2$ (X = Cha)

ESI-MS: calculated monoisotopic $[M+3H]^{3+}$ = 1060.497

observed monoisotopic $[M+3H]^{3+}$ = 1060.841

(*ESI-MS was used for this peptide due to high laser fragmentation of the fluorescein ground under MALDI-TOF-MS condition)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column.

1.7 References

(1) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular Condensates: Organizers of Cellular Biochemistry. *Nat Rev Mol Cell Biol* **2017**, *18* (5), 285–298. https://doi.org/10.1038/nrm.2017.7.

(2) Shin, Y.; Brangwynne, C. P. Liquid Phase Condensation in Cell Physiology and Disease. *Science (1979)* **2017**, *357* (6357). https://doi.org/10.1126/science.aaf4382.

(3) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu Rev Phys Chem* **2020**, *71*, 53–75. https://doi.org/10.1146/annurev-physchem-071819-113553.

(4) Roden, C.; Gladfelter, A. S. RNA Contributions to the Form and Function of Biomolecular Condensates. *Nat Rev Mol Cell Biol* **2021**, *22* (3), 183–195. https://doi.org/10.1038/s41580-020-0264-6.

(5) Lafontaine, D. L. J.; Riback, J. A.; Bascetin, R.; Brangwynne, C. P. The Nucleolus as a Multiphase Liquid Condensate. *Nat Rev Mol Cell Biol* **2021**, *22* (3), 165–182. https://doi.org/10.1038/s41580-020-0272-6.

(6) Wang, B.; Zhang, L.; Dai, T.; Qin, Z.; Lu, H.; Zhang, L.; Zhou, F. Liquid–Liquid Phase Separation in Human Health and Diseases. *Signal Transduct Target Ther* **2021**, *6* (1). https://doi.org/10.1038/s41392-021-00678-1.

(7) Kiebler, M. A.; Bauer, K. E. RNA Granules in Flux: Dynamics to Balance Physiology and Pathology. *Nat Rev Neurosci* **2024**. https://doi.org/10.1038/s41583-024-00859-1.

(8) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A Framework for Understanding the Functions of Biomolecular Condensates across Scales. *Nat Rev Mol Cell Biol* **2021**, *22* (3), 215–235. https://doi.org/10.1038/s41580-020-00303-z.

(9) Xing, W.; Muhlrad, D.; Parker, R.; Rosen, M. K. A Quantitative Inventory of Yeast P Body Proteins Reveals Principles of Composition and Specificity. *Elife* **2020**, *9*, 1–63. https://doi.org/10.7554/eLife.56525.

(10) Niu, J.; Qiu, C.; Abbott, N. L.; Gellman, S. H. Formation of versus Recruitment to RNA-Rich Condensates: Controlling Effects Exerted by Peptide Side Chain Identity. *J Am Chem Soc* **2022**, *144* (23), 10386–10395. https://doi.org/10.1021/jacs.2c02222.

(11) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V; Mittag, T. *Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains*; 2020; Vol. 367. https://www.science.org.

(12) Chang, L. W.; Lytle, T. K.; Radhakrishna, M.; Madinya, J. J.; Vélez, J.; Sing, C. E.; Perry, S. L. Sequence and Entropy-Based Control of Complex Coacervates. *Nat Commun* **2017**, *8* (1). https://doi.org/10.1038/s41467-017-01249-1.

(13) Lytle, T. K.; Chang, L. W.; Markiewicz, N.; Perry, S. L.; Sing, C. E. Designing Electrostatic Interactions via Polyelectrolyte Monomer Sequence. *ACS Cent Sci* **2019**, *5* (4), 709–718. https://doi.org/10.1021/acscentsci.9b00087.

(14) Cakmak, F. P.; Choi, S.; Meyer, M. C. O.; Bevilacqua, P. C.; Keating, C. D. Prebiotically-Relevant Low Polyion Multivalency Can Improve Functionality of Membraneless

Compartments. *Nat Commun* **2020**, *11* (1), 1–11. https://doi.org/10.1038/s41467-020-19775-w.

(15) Rekhi, S.; Garcia, C. G.; Barai, M.; Rizuan, A.; Schuster, B. S.; Kiick, K. L.; Mittal, J. Expanding the Molecular Language of Protein Liquid–Liquid Phase Separation. *Nat Chem* **2024**, *16* (7), 1113–1124. https://doi.org/10.1038/s41557-024-01489-x.

(16) Burke, K. A.; Janke, A. M.; Rhine, C. L.; Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules That Bind the C-Terminal Domain of RNA Polymerase II. *Mol Cell* **2015**, *60* (2), 231–241. https://doi.org/10.1016/j.molcel.2015.09.006.

(17) Akahoshi, Y.; Sugai, H.; Mimura, M.; Shinkai, Y.; Kurita, R.; Shiraki, K.; Tomita, S. Phase-Separation Propensity of Non-Ionic Amino Acids in Peptide-Based Complex Coacervation Systems. *Biomacromolecules* **2023**. https://doi.org/10.1021/acs.biomac.2c01148.

(18) Chen, C.; Ding, X.; Akram, N.; Xue, S.; Luo, S. Z. Fused in Sarcoma: Properties, Self-Assembly and Correlation with Neurodegenerative Diseases. *Molecules* **2019**, *24* (8). https://doi.org/10.3390/molecules24081622.

(19) Wang, J.; Choi, J. M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174* (3), 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006.

(20) Qamar, S.; Wang, G. Z.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Ströhl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.;

Tartaglia, G. G.; Farrer, L. A.; Kaminski Schierle, G. S.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **2018**, *173* (3), 720-734.e15. https://doi.org/10.1016/j.cell.2018.03.056.

(21) Ma, J. C.; Dougherty, D. A. The Cation−π Interaction. *Chem. Rev.* **1997**, *97*, 1303–1324.

(22) Dougherty, D. A. The Cation-π Interaction. *Acc Chem Res* **2013**, *46* (4), 885–893. https://doi.org/10.1021/ar300265y.

(23) Pieters, B. J. G. E.; Wuts, M. H. M.; Poater, J.; Kumar, K.; White, P. B.; Kamps, J. J. A. G.; Sherman, W.; Pruijn, G. J. M.; Paton, R. S.; Beuming, T.; Bickelhaupt, F. M.; Mecinović, J. Mechanism of Biomolecular Recognition of Trimethyllysine by the Fluorinated Aromatic Cage of KDM5A PHD3 Finger. *Commun Chem* **2020**, *3* (1). https://doi.org/10.1038/s42004-020-0313-2.

(24) Kumar, K.; Woo, S. M.; Siu, T.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. Cation-π Interactions in Protein-Ligand Binding: Theory and Data-Mining Reveal Different Roles for Lysine and Arginine. *Chem Sci* **2018**, *9* (10), 2655–2665. https://doi.org/10.1039/c7sc04905f.

(25) Beene, D. L.; Brandt, G. S.; Zhong, W.; Zacharias, N. M.; Lester, H. A.; Dougherty, D. A. Cation-π Interactions in Ligand Recognition by Serotonergic (5-HT3A) and Nicotinic Acetylcholine Receptors: The Anomalous Binding Properties of Nicotine. *Biochemistry* **2002**, *41* (32), 10262–10269. https://doi.org/10.1021/bi020266d.

(26) Pless, S. A.; Millen, K. S.; Hanek, A. P.; Lynch, J. W.; Lester, H. A.; Lummis, S. C. R.; Dougherty, D. A. A Cation-π Interaction in the Binding Site of the Glycine Receptor Is Mediated by a Phenylalanine Residue. *Journal of Neuroscience* **2008**, *28* (43), 10937–10942. https://doi.org/10.1523/JNEUROSCI.2540-08.2008.

(27) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Nicotine Binding to Brain Receptors Requires a Strong Cation- Interaction. *Nature* **2009**, *458* (7237), 534–537. https://doi.org/10.1038/nature07768.

(28) Lummis, S. C. R.; Beene, D. L.; Harrison, N. J.; Lester, H. A.; Dougherty, D. A. A Cation-π Binding Interaction with a Tyrosine in the Binding Site of the GABAC Receptor. *Chem Biol* **2005**, *12* (9), 993–997. https://doi.org/10.1016/j.chembiol.2005.06.012.

(29) Travis, C. R.; Kean, K. M.; Albanese, K. I.; Henriksen, H. C.; Treacy, J. W.; Chao, E. Y.; Houk, K. N.; Waters, M. L. Trimethyllysine Reader Proteins Exhibit Widespread Charge-Agnostic Binding via Different Mechanisms to Cationic and Neutral Ligands. *J Am Chem Soc* **2024**, *146* (5), 3086–3093. https://doi.org/10.1021/jacs.3c10031.

(30) Fang, K. Y.; Lieblich, S. A.; Tirrell, D. A. Incorporation of Non-Canonical Amino Acids into Proteins by Global Reassignment of Sense Codons. In *Methods in Molecular Biology*; Humana Press Inc., 2018; Vol. 1798, pp 173–186. https://doi.org/10.1007/978-1-4939-7893-9_13.

(31) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol Cell* **2015**, *57* (5), 936–947. https://doi.org/10.1016/j.molcel.2015.01.013.

(32) Raran-Kurussi, S.; Waugh, D. S. The Ability to Enhance the Solubility of Its Fusion Partners Is an Intrinsic Property of Maltose-Binding Protein but Their Folding Is Either Spontaneous or Chaperone-Mediated. *PLoS One* **2012**, *7* (11). https://doi.org/10.1371/journal.pone.0049589.

(33) Kapust, R. B.; Waugh, D. S. Escherichia Coli Maltose-binding Protein Is Uncommonly Effective at Promoting the Solubility of Polypeptides to Which It Is Fused . *Protein Science* **1999**, *8* (8), 1668–1674. https://doi.org/10.1110/ps.8.8.1668.

(34) Yushchenko, T.; Deuerling, E.; Hauser, K. Insights into the Aggregation Mechanism of PolyQ Proteins with Different Glutamine Repeat Lengths. *Biophys J* **2018**, *114* (8), 1847–1857. https://doi.org/10.1016/j.bpj.2018.02.037.

(35) Xue, C.; Lin, T. Y.; Chang, D.; Guo, Z. Thioflavin T as an Amyloid Dye: Fibril Quantification, Optimal Concentration and Effect on Aggregation. *R Soc Open Sci* **2017**, *4* (1), 160696. https://doi.org/10.1098/rsos.160696.

(36) Biancalana, M.; Koide, S. Molecular Mechanism of Thioflavin-T Binding to Amyloid Fibrils. *Biochim Biophys Acta* **2010**, *1804* (7), 1405–1412. https://doi.org/10.1016/j.bbapap.2010.04.001.

(37) Luo, F.; Gui, X.; Zhou, H.; Gu, J.; Li, Y.; Liu, X.; Zhao, M.; Li, D.; Li, X.; Liu, C. Atomic Structures of FUS LC Domain Segments Reveal Bases for Reversible Amyloid Fibril Formation. *Nat Struct Mol Biol* **2018**, *25* (4), 341–346. https://doi.org/10.1038/s41594-018-0050-8.

(38) Choi, J. M.; Holehouse, A. S.; Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu Rev Biophys* **2020**, *49*, 107–133. https://doi.org/10.1146/annurev-biophys-121219-081629.

(39) Emmanouilidis, L.; Bartalucci, E.; Kan, Y.; Ijavi, M.; Pérez, M. E.; Afanasyev, P.; Boehringer, D.; Zehnder, J.; Parekh, S. H.; Bonn, M.; Michaels, T. C. T.; Wiegand, T.; Allain, F. H. T. A Solid Beta-Sheet Structure Is Formed at the Surface of FUS Droplets during Aging. *Nat Chem Biol* **2024**, *20* (8), 1044–1052. https://doi.org/10.1038/s41589-024-01573-w.

(40) Anderson, C. F.; Record, M. T. Salt-Nucleic Acid Interactions. *Annu. Rev. Phys. Chern* **1995**, *46*, 657–700.

(41) Manning, G. S. Counterion Binding in Polyelectrolyte Theory. *Acc Chem Res* **1979**, *12*, 443–449.

(42) Mascotti, D. P.; Lohman, T. M. Thermodynamics of Oligoarginines Binding to RNA and DNA. *Biochemistry* **1997**, *36*, 7272–7279.

(43) Meek, J. L.; Rossetti, Z. L. Factors Affecting Retention and Resolution of Peptides in High-Performance Liquid Chromatography. *J Chromatogr* **1981**, *211*, 15–28.

(44) Meek, J. L. Prediction of Peptide Retention Times in High-Pressure Liquid Chromatography on the Basis of Amino Acid Composition. *Proc. Natl. Acad. Sci. USA* **1980**, *77* (3), 1632–1636.

(45) Guo, D.; Mant, C. T.; Taneja, A. K.; Parker, J. M. R.; Hodges, R. S. PREDICTION OF PEPTIDE RETENTION TIMES IN REVERSED-PHASE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY. *J Chromatogr* **1986**, *359*, 499–517.

(46) J Wilce, M. C.; Aguilar, M.-I.; W Heam, M. T. Physicochemical Basis of Amino Acid Hydrophobicity Scales: Evaluation of Four New Scales of Amino Acid Hydrophobicity Coefficients Derived from RP-HPLC of Peptides. *Anal Chem* **1995**, *67* (7), 1210–1219.

(47) Taylor, N. O.; Wei, M. T.; Stone, H. A.; Brangwynne, C. P. Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophys J* **2019**, *117* (7), 1285–1300. https://doi.org/10.1016/j.bpj.2019.08.030.

(48) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; Pozniakovski, A.; Poser, I.; Maghelli, N.; Royer, L. A.; Weigert, M.; Myers, E. W.; Grill, S.; Drechsel, D.; Hyman, A. A.; Alberti, S. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162* (5), 1066–1077. https://doi.org/10.1016/j.cell.2015.07.047.

(49) Murakami, T.; Qamar, S.; Lin, J. Q.; Schierle, G. S. K.; Rees, E.; Miyashita, A.; Costa, A. R.; Dodd, R. B.; Chan, F. T. S.; Michel, C. H.; Kronenberg-Versteeg, D.; Li, Y.; Yang, S. P.; Wakutani, Y.; Meadows, W.; Ferry, R. R.; Dong, L.; Tartaglia, G. G.; Favrin, G.; Lin, W. L.; Dickson, D. W.; Zhen, M.; Ron, D.; Schmitt-Ulms, G.; Fraser, P. E.; Shneider, N. A.; Holt, C.; Vendruscolo, M.; Kaminski, C. F.; St George-Hyslop, P. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **2015**, *88* (4), 678–690. https://doi.org/10.1016/j.neuron.2015.10.030.

(50) Zhu, Y. J.; Huang, S. C.; Xia, X. X.; Qian, Z. G. Noncanonical Amino Acid Incorporation Modulates Condensate States of Intrinsically Disordered Proteins in Escherichia Coli Cells. *Biomacromolecules* **2024**. https://doi.org/10.1021/acs.biomac.4c00864.

(51) Yao, R. W.; Rosen, M. K. Advanced Surface Passivation for High-Sensitivity Studies of Biomolecular Condensates. *Proc Natl Acad Sci U S A* **2024**, *121* (22). https://doi.org/10.1073/pnas.2403013121.

(52) Pinheiro, S.; Soteras, I.; Gelpí, J. L.; Dehez, F.; Chipot, C.; Luque, F. J.; Curutchet, C. Structural and Energetic Study of Cation-π-Cation Interactions in Proteins. *Physical Chemistry Chemical Physics* **2017**, *19* (15), 9849–9861. https://doi.org/10.1039/c6cp08448f.

(53) Gallivan, J. P.; Dougherty, D. A. Cation-π Interactions in Structural Biology. *Proc Natl Acad Sci U S A* **1999**, *96* (17), 9459–9464. https://doi.org/10.1073/pnas.96.17.9459.

(54) Baker, E. N.; Htjabardt, R. E. HYDROGEN BONDING IN GLOBULAR PROTEINS. *Prog Biophys. molec. Biol* **1984**, *44*, 97–179.

(55) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133–7155.

(56) Pace, C. N. Contribution of the Hydrophobic Effect to Globular Protein Stability. *J. Mol. Biol* **1992**, *226*, 29–35.

(57) Shirley, B. A.; Stanssens, P.; Hahn, U.; Pace, C. N. Contribution of Hydrogen Bonding to the Conformational Stability of Ribonuclease T1. *Biochemistry* **1992**, *31*, 4763–4769.

(58) Perutz, M. F.; Fersht, A. R.; Simon, S. R.; Roberts, G. C. K.; Ladner, J. E.; Ho, C. G.; Fermi, G.; Luisi, B.; Shaanan, B.; Liddington, R. C.; Riggs, A.; Simolo, K.; Korszun, Z.

R.; Stucky, G.; Moffat, K.; Mclendon, G.; Cantor, C. R.; Snyder, S. R.; Welty, E. V; Walder, R. Y.; Williams, L. A.; Walder, J. A.; Tomoda, A.; Takeshita, M.; Yoneyama, Y.; Amone, A.; White, S. L. *Contribution of Hydrogen Bonding to the Conformational Stability of Ribonuclease Tl1"9000 Gent*; 1992; Vol. 31.

(59) Das, S.; Lin, Y.-H.; Vernon, R. M.; Forman-Kay, J. D.; Chan, H. S. Comparative Roles of Charge, π, and Hydrophobic Interactions in Sequence-Dependent Phase Separation of Intrinsically Disordered Proteins. *Proc Natl Acad Sci U S A* **2020**, *117* (46), 28795–28805. https://doi.org/10.1073/pnas.2008122117/-/DCSupplemental.y.

(60) Hughes, R. M.; Waters, M. L. Arginine Methylation in a β-Hairpin Peptide: Implications for Arg-π Interactions, ΔCp°, and the Cold Denatured State. *J Am Chem Soc* **2006**, *128* (39), 12735–12742. https://doi.org/10.1021/ja061656g.

(61) Li, Y.; Cao, Y. The Molecular Mechanisms Underlying Mussel Adhesion. *Nanoscale Adv* **2019**, *1* (11), 4246–4257. https://doi.org/10.1039/c9na00582j.

(62) Gebbie, M. A.; Wei, W.; Schrader, A. M.; Cristiani, T. R.; Dobbs, H. A.; Idso, M.; Chmelka, B. F.; Herbert Waite, J.; Israelachvili, J. N. Tuning Underwater Adhesion with Cation-φ Interactions. *Nat Chem* **2017**, *9* (5), 473–479. https://doi.org/10.1038/nchem.2720.

(63) Guo, Q.; Zou, G.; Qian, X.; Chen, S.; Gao, H.; Yu, J. Hydrogen-Bonds Mediate Liquid-Liquid Phase Separation of Mussel Derived Adhesive Peptides. *Nat Commun* **2022**, *13* (1). https://doi.org/10.1038/s41467-022-33545-w.

(64) Deepankumar, K.; Guo, Q.; Mohanram, H.; Lim, J.; Mu, Y.; Pervushin, K.; Yu, J.; Miserez, A. Liquid–Liquid Phase Separation of the Green Mussel Adhesive Protein Pvfp-5

Is Regulated by the Post-Translated Dopa Amino Acid. *Advanced Materials* **2022**, *34* (25). https://doi.org/10.1002/adma.202103828.

(65) Woll, M. G.; Hadley, E. B.; Mecozzi, S.; Gellman, S. H. Stabilizing and Destabilizing Effects of Phenylalanine → F 5-Phenylalanine Mutations on the Folding of a Small Protein. *J Am Chem Soc* **2006**, *128* (50), 15932–15933. https://doi.org/10.1021/ja0634573.

(66) Lee, K. H.; Lee, H. Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. Fluorous Effect in Proteins: De Novo Design and Characterization of a Four-α-Helix Bundle Protein Containing Hexafluoroleucine. *Biochemistry* **2004**, *43* (51), 16277–16284. https://doi.org/10.1021/bi049086p.

(67) Buer, B. C.; Levin, B. J.; Marsh, E. N. G. Influence of Fluorination on the Thermodynamics of Protein Folding. *J Am Chem Soc* **2012**, *134* (31), 13027–13034. https://doi.org/10.1021/ja303521h.

(68) Salwiczek, M.; Nyakatura, E. K.; Gerling, U. I. M.; Ye, S.; Koksch, B. Fluorinated Amino Acids: Compatibility with Native Protein Structures and Effects on Protein–Protein Interactions. *Chem Soc Rev* **2012**, *41* (6), 2135–2171. https://doi.org/10.1039/c1cs15241f.

(69) Wake, N.; Weng, S.-L.; Zheng, T.; Wang, S.-H.; Kirilenko, V.; Mittal, J.; Fawzi, N. L. Expanding the Molecular Grammar of Polar Residues and Arginine in FUS Phase Separation. *Nat Chem Biol* **2025**. https://doi.org/10.1038/s41589-024-01828-6.

(70) Kilgore, H. R.; Mikhael, P. G.; Overholt, K. J.; Boija, A.; Hannett, N. M.; Van Dongen, C.; Lee, T. I.; Chang, Y. T.; Barzilay, R.; Young, R. A. Distinct Chemical Environments in Biomolecular Condensates. *Nat Chem Biol* **2024**, *20* (3), 291–301. https://doi.org/10.1038/s41589-023-01432-0.

(71) Ambadi Thody, S.; Clements, H. D.; Baniasadi, H.; Lyon, A. S.; Sigman, M. S.; Rosen, M. K. Small-Molecule Properties Define Partitioning into Biomolecular Condensates. *Nat Chem* **2024**. https://doi.org/10.1038/s41557-024-01630-w.

Chapter 2

Exploring the Effects of Sequence-Dependent Charge Patterning on Multiphase Separation of Intrinsically Disordered Proteins

Experiments were designed by Ruiwen Xu and Samuel H. Gellman.

Experiments were performed by Ruiwen Xu.

Graphics were prepared using Microsoft PowerPoint, GraphPad Prism, ImageJ.

2.1 Abstract

Previous studies of liquid-liquid phase separation (LLPS) in the Gellman lab focused on understanding the effects of peptide side chains and the underlying principles of biomolecular condensates in vitro. This chapter represents a shift from studying the effects of amino acid composition to studying the effects of phase separation behavior. More specifically, we sought to understand how the linear arrangement of oppositely charged residues, quantified by the κ (kappa) parameter, modulates miscibility in multiphase condensates. Using two phase-separating proteins, CBX5 and NPM1, we generated high κ (segregated charged residues) and low κ (mixed charged residues) synthetic variants by shuffling charged residues while maintaining identical amino acid compositions. Mixing experiments revealed that native CBX5 and NPM1 fully colocalize within single-phase condensates. However, combinations involving charge-patterned variants exhibit partial colocalization, with increased demixing observed for high-κ constructs. Bio-layer interferometry (BLI) confirms that κ modulates protein-protein interaction strengths. These findings suggest that charge patterning, independent of composition, can be a determinant of condensate miscibility, positioning κ as a predictive design parameter for engineering LLPS-based systems.

2.2 Introduction

Liquid-liquid phase separation (LLPS) is a fundamental organizing principle in cell biology, enabling the formation of membraneless organelles such as the nucleolus, stress granules, P-bodies, and heterochromatin domains.[1–5] These biomolecular condensates arise from dynamic, multivalent interactions among proteins and nucleic acids, allowing cells to spatially and temporally regulate biochemical reactions without relying on membrane encapsulation.[2,6–8] A central class of molecules involved in LLPS are proteins that contain intrinsically disordered regions (IDRs), which lack a fixed tertiary structure and are enriched in low-complexity sequences.[9–12] Their conformational flexibility and modular interaction motifs make them ideal scaffolds for phase separation.

Over the past decade, LLPS has gained interest in synthetic biology as a strategy to create programmable compartments for organizing and regulating biochemical reactions *in vitro*. One exciting area is the creation of multiphase condensates that contain multiple, coexisting liquid phases with distinct composition and properties.[13,14] Such multiphasic organization is observed in both natural and synthetic systems, as layered or spatially discrete condensates with unique compositions and material properties.[15,16] While the amino acid composition of IDRs, such as the proportions of charged, aromatic, or polar residues, has been widely studied in the context of LLPS,[17–28] we are interested in understanding the role of sequence patterning in governing phase behavior. One quantitative measure of charge patterning is the κ (kappa) parameter.[29,30] κ is a measurement of the extent of charge segregation in a sequence as defined Das and Pappu in 2013.[30] For κ, a value near 0 corresponds to a well-mixed distribution of oppositely charged residues, while a value near 1 indicates that like charges are spatially clustered into blocks. κ is calculated independent of net charge or overall amino acid content, focusing on the charged residue distribution alone. We are interested in understanding the role of κ in modulating phase compatibility between different proteins in a multiphasic system.

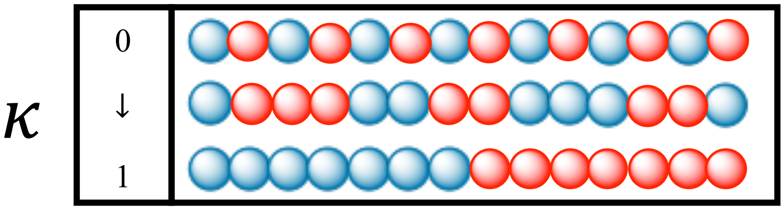


Figure 2.1. Schematic depiction of the implication of changing k values. Here, red and blue spheres represent positively charged residues and negatively charged residues, respectively. This image was regenerated based on Figure 1b from Holehouse et al.[30]

To explore this hypothesis, we studied two proteins with known phase separation behavior: human Chromebox homolog 5 (CBX5 below, also known as HP1α)[31,32] and human nucleophosmin 1 (NPM1 below).[33,34] Although unmodified human CBX5 alone does not undergo phase separation *in vitro*, previous studies have shown that phosphorylation of its N-terminal extension or binding to DNA can promote phase separation.[35,36] Mensah et al. found that the addition of PEG-8000 to the buffer can successfully induce phase separation of human CBX5, forming liquid-like condensates that contribute to heterochromatin organization.[37] Disordered regions of CBX5 are rich in charged and polar residues. NPM1 is a well-known scaffold of nucleolar sub-compartments and drives LLPS via interactions with RNA and other nucleolar proteins.[33,34] We generated both CBX5 and NPM1 using recombinant expression in *E. coli*, and confirmed that under LLPS-promoting conditions, the wild-type proteins are fully miscible—meaning they mix together to form uniform, single-phase condensates. To test whether charge patterning influences phase behavior, we generated a synthetic low-κ (charge-dispersed) variant of CBX5 and a high-κ (charge-clustered) variant of NPM1. Each variant was a sequence isomer of the original protein. That is, each variant had the same amino acid composition as the original protein. These sequences preserve amino acid composition and net charge, allowing us to isolate the effects of κ on phase behavior. Using confocal microscopy and bio-layer interferometry (BLI),[38–40] we characterized the effects of κ on phase separation behavior and protein-protein interaction strength. Our results revealed that charge patterning modulates protein-protein interactions and condensate miscibility. These findings demonstrate that κ can serve as a predictive and tunable design parameter for programming multiphase behavior in synthetic condensates, and underscore the broader principle that sequence context, not just content, governs the emergent properties of LLPS systems.

2.3 Results and Discussion

2.3.1 Multiphase Separation Protein Construct Screen

At the beginning of this project, I was particularly interested in how phase separation relates to transcriptional regulation. In eukaryotic cells, phase separation plays two key roles: the formation of membraneless organelles (such as the nucleolus or stress granules),[15,41,42] and the regulation of transcriptional and post-transcriptional processes.[43–45] During transcription, intrinsically disordered regions (IDRs) within transcription factors, coactivators, and RNA-binding proteins can engage in multivalent interactions that lead to the formation of biomolecular condensates.[43–45] These condensates influence multiple stages of gene regulation, including transcription initiation, elongation, RNA splicing, and mRNA modification.[46–48] These processes often rely on weak, multivalent interactions that promote dynamic but selective phase-separated environments.[49]

To explore the possibility of generating multiphase protein condensates, I first reviewed relevant findings in the literature. One particularly influential study by Mensah et al. investigated how different IDR-containing proteins interact within model condensates.[37] They studied the partitioning of a wildtype mEGFP–HMGB1 protein and a mutant version (E186R) into preformed condensates of mCherry-MED1-IDR, mCherry-HP1α (CBX5), and mCherry-NPM1. Interestingly, the E186R mutant of HMGB1 showed preferential localization at the droplet periphery, suggesting that even small changes in IDR properties can dramatically affect miscibility.[37] This work prompted me to study similar systems using my own set of recombinant proteins.

Mediator subunit 1(MED1) is a key protein of the mediator complex, regulating transcription by bridging transcription factors and RNA polymerase II.[50] MED1 contains large IDRs that have been shown to undergo phase separation. High mobility group box 1 (HMGB1) is a chromatin-associated architectural protein involved in transcription regulation and cellular stress response.[51] It also contains disordered regions that promote phase separation. As a first step, I tested whether full-length mEGFP-HMGB1 and mCherry-MED1 could phase-separate independently under relevant buffer conditions in the presence of a crowding agent (10% PEG-8000). The mEGFP and mCherry fluorescent tags enabled visualization of the protein condensates under confocal microscopy. Both proteins successfully formed condensates, confirming their intrinsic ability to undergo LLPS (Figure 2.2A). When I combined 0.5 μM mEGFP-HMGB1 and 10 μM mCherry-MED1, I observed colocalization of green and red fluorescent signals within the same droplets under confocal microscopy (Figure 2.2B). This suggested that these two proteins are miscible and can form a single, mixed condensed phase.

To explore potential immiscibility, I next tested mixtures involving other known phase-separating proteins. I prepared condensates of MBP-FUS-C, a C-terminal low-complexity domain of FUS protein known for its strong tendency to phase separate (see Chapter1 for sequence of FUS-C). When MBP-FUS-C was mixed with mCherry-MED1, I observed the formation of two distinct populations of droplets: red droplets from mCherry-MED1, and colorless droplets from MBP-FUS-C (Figure 2.2D). There was little to no overlap between them, suggesting that these proteins do not co-partition and instead form immiscible condensates. A different outcome occurred when I combined eGFP-HMGB1 with MBP-FUS-C: only green droplets were visible, and no colorless droplets from MBP-FUS-C were observed under confocal microscopy (Figure 2.2E). These observations supported the idea that not all IDR-containing proteins readily mix.

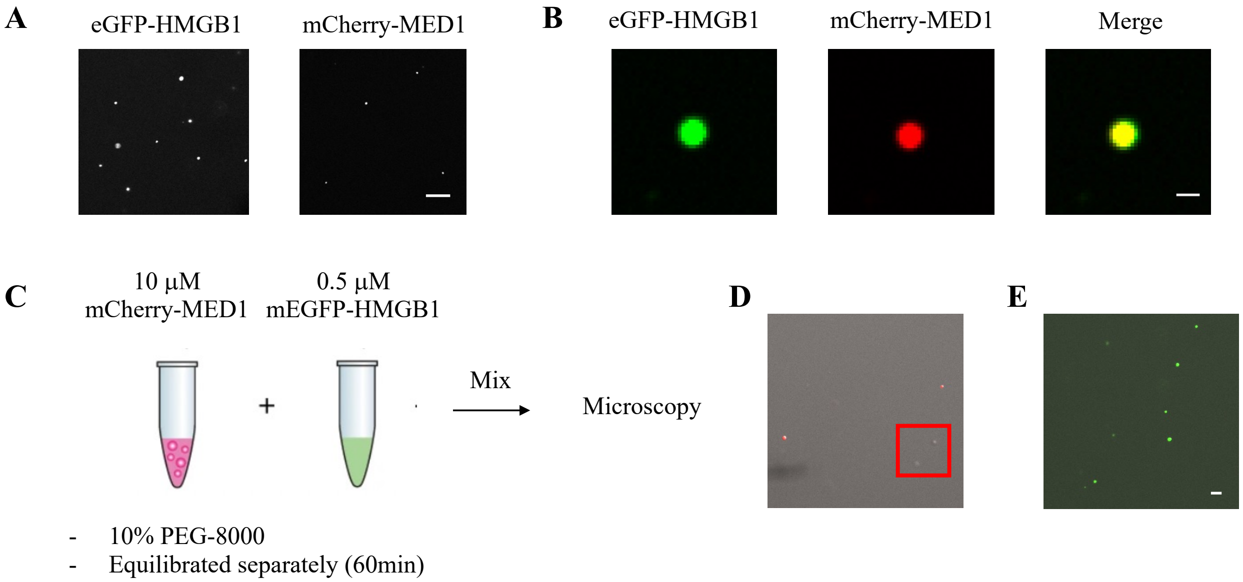


Figure 2.2. A. Single phase formed by 10 μM mEGFP-HMGB1 on BSA-coated confocal plate (left) and single phase formed by 10 μM mCherry-MED1 on BSA-coated confocal plate (right). Scale bar 10 μm. **B.** Colocalization of 0.5 μM eGFP-HMGB1 10 μM mCherry-MED1. Green channel: mEGFP-HMGB1; red channel: mCherry-MED1. Scale bar 2 μm. **C.** Scheme of co-droplet assays. **D.** Separated multiphase separation observed by 10 μM mCherry-MED1 and 1 μM MBP-FUS-C on BSA-coated confocal plate. **E.** Merged droplet between 10 μM mEGFP-HMGB1 and 1 μM MBP-FUS-C on BSA-coated confocal plate. Buffer: 50mM Tris pH 7.5, 150mM NaCl, 0.5mM TCEP, 10% PEG-8000. Scale bar 10 μm.

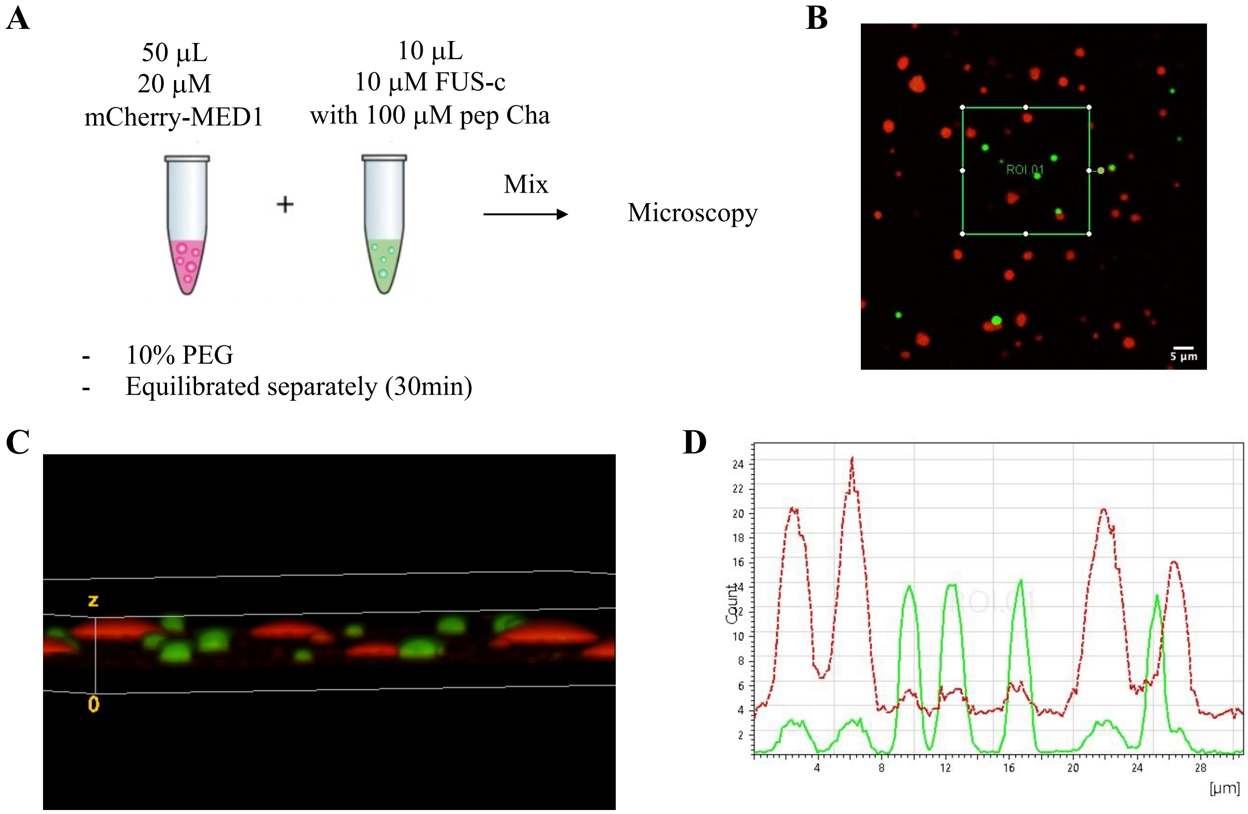


Figure 2.3. A. Scheme of co-droplet assays **B.** Colocalization of 20 μM mCherry-MED1 and droplets formed 10 μM FUS-C with 100 μM FITC-labeled pep Cha. Green channel: droplets formed by FUS-C/pep Cha; red channel: mCherry-MED1. Scale bar 5 μm. **C.** 3D image generated from B. **D.** Intensities of droplets in the region of interest from B. The green droplets do not colocalize with the red droplets. Buffer: 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 10% PEG-8000.

To further test the hypothesis that some IDR-containing proteins form immiscible condensates, I conducted a mixing experiment using pre-assembled mCherry-MED1 droplets and FITC-labeled pep Cha (a synthetic phase-separating peptide construct from Chapter 1) incorporated into FUS-C condensates. Again, I observed complete spatial segregation: red droplets of MED1 and green droplets of FUS-C/pep Cha, with no observable overlap (Figure 2.3B). These distinct compartments coexisted in the same field of view but did not coalesce over 2 hours, clearly demonstrating multiphasic behavior.

This raised the question of what drives the miscibility or immiscibility of phase-separated condensates. While factors such as overall charge, hydrophobicity, and amino acid composition are known to influence LLPS, I focused on one parameter, linear charge patterning, quantified by the κ (kappa) value.[30] The κ parameter captures how charged residues (arginine, lysine, glutamate, aspartate) are distributed along the protein sequence (Figure 2.1). A low κ value indicates a well-mixed distribution of oppositely charged residues, while a high κ value reflects clustering of like charges into blocks. This patterning can have a profound effect on the way proteins interact, particularly through electrostatic complementarity or repulsion.[26,52]

FUS-C had a κ value of 0.097, while MED1 had a slightly higher κ of 0.162. Although subtle, this difference may reflect an important design principle for condensate compatibility. These results led me to consider whether κ alone, independent of overall amino acid composition, could serve as a predictor of condensate compatibility. These results ultimately motivated the design of the CBX5 and NPM1 κ variants described in later sections. By directly tuning charge patterning (while keeping amino acid composition constant), we aimed to investigate how κ influences miscibility, condensate behavior, and protein–protein interaction strength.

2.3.2 Native CBX5 and NPM1 Form Fully Miscible Condensates

To establish a baseline for assessing phase compatibility, we first examined the phase separation behavior of wildtype human CBX5 and human NPM1. Here we used small-molecule fluorescent tags fluorescein (FITC) and tetramethylrhodamine (TAMRA) rather than fluorescent protein mEGFP and mCherry as the labeling strategy. Fluorescent proteins can be too big to interfere the protein-protein interactions thus influencing the phase separation behaviors. We observed phase separation under microscope for both 50 µM FITC-labeled CBX5 (Figure 2.4A) and 50 µM TMARA-labeled NPM1 (Figure 2.4B) in separate confocal dishes. However, when we incubated the two proteins for 30 minutes, we noticed an inconsistent results (Figure 2.4 C-D). In Figure 2.3C, CBX5 and NPM1 were completely immiscible. In Figure 2.4D, part of the NPM1 condensate was fused with one of the CBX5 droplets. We hypothesized that the hydrophilic nature of both proteins, prone to stick to the bovine serum albumin (BSA) coated confocal glass dish, slowing down the movement and merging of the condensates. We then changed the coating from hydrophilic (BSA) to hydrophobic, Pluronic F127 (PF-127).[53] With the PF-127 treated confocal dish, we combined FITC-labeled CBX5 and TAMRA-labeled NPM1 at equimolar concentrations. Both proteins formed well-mixed, spherical condensates with complete colocalization, as assessed by fluorescence microscopy (Figure 2.2E).

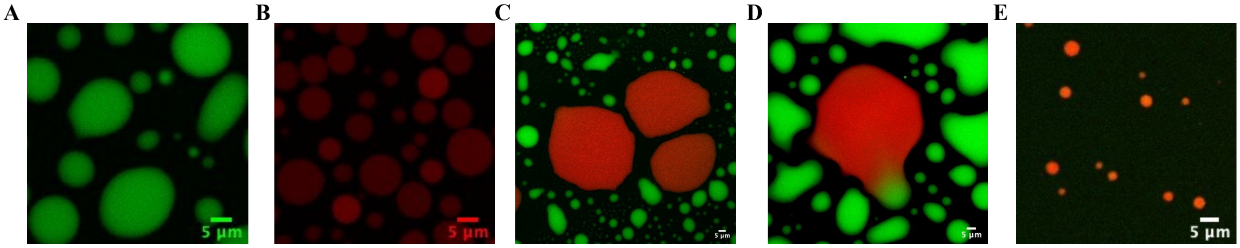


Figure 2.4. A. Single phase formed by 50 µM FITC-labeled wildtype CBX5 on BSA-coated confocal plate. **B.** Single phase formed by 50 µM TAMRA-labeled wildtype NPM1 on BSA-coated confocal plate. **C.** Separated multiphase separation observed by 50 µM FITC-labeled wildtype CBX5 and 50 µM TAMRA-labeled wildtype NPM1 on BSA-coated confocal plate. **D.** Merged droplet between 50 µM FITC-labeled wildtype CBX5 and 50 µM TAMRA-labeled wildtype NPM1 on BSA-coated confocal plate. **E.** Colocalization between 10 µM FITC-labeled wildtype CBX5 and 10 µM TAMRA-labeled wildtype NPM1 on a PF-127 coated confocal plate. Buffer: 50 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 10% PEG-8000. Scale bar 5 µm.

2.3.3 Charge Patterning Alters Miscibility in Condensates

To investigate the effect of charge patterning, we computationally designed low-κ (charge-dispersed) variant of CBX5 and high-κ (charge-clustered) of NPM1 using local CIDER, a tool developed by Pappu lab for analyzing IDP sequence properties.[54] Charged residues (K, R, D, E) were shuffled within each protein to modulate local charge density while keeping the amino acid composition unchanged. κ values were calculated using the method described by Das and Pappu 2013.[30] High κ sequences exhibit contiguous blocks of similarly charged residues, while low κ sequences feature well-mixed distributions of opposite charges. As summarized in Table 2.1, wild-type CBX5 and NPM1 exhibited intermediate κ values of 0.21 and 0.39, respectively. CBX5_low κ was engineered to have $\kappa = 0.06$ and NPM1_high κ was engineered to have $\kappa = 0.46$. The relative charge distribution of each construct was shown in Figure 2.5.

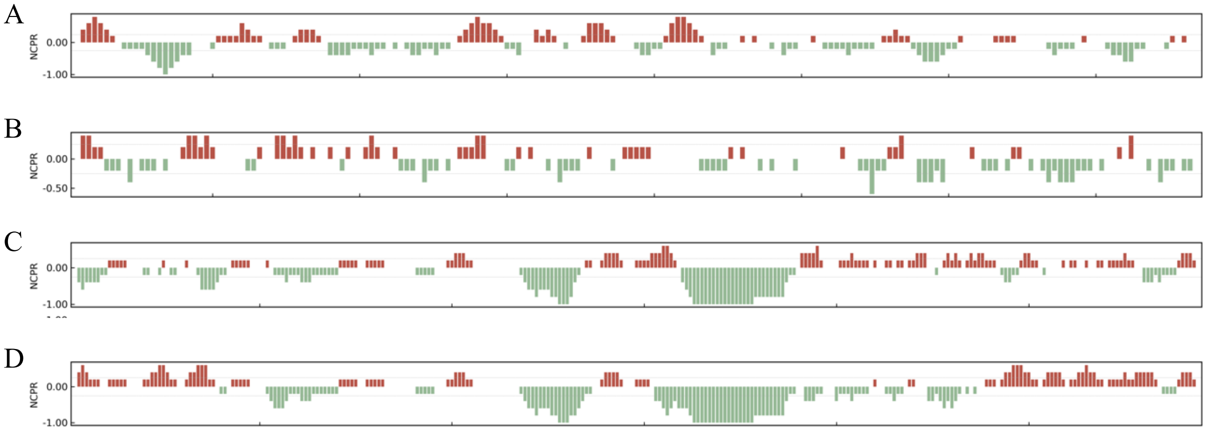


Figure 2.5. Net charge per residue (NCPR) lot for CBX5 (**A**), CBX5_low κ (**B**), NPM1 (**C**), NPM1_high κ (**D**). Red color bars indicated positively charged residues and green color bars indicated negatively charged residues.

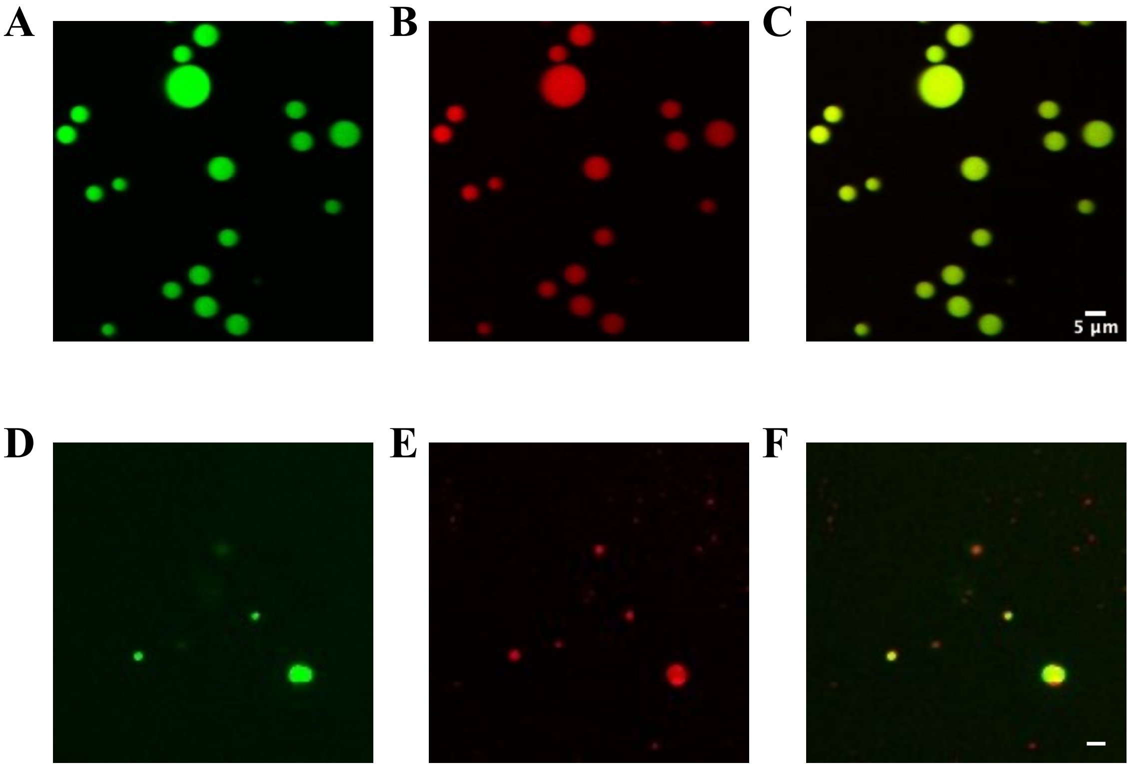


Figure 2.6. Colocalization of 25 µM FITC-labeled wildtype CBX5 and 25 µM TAMRA-labeled CBX5_low κ. Green channel: FITC-labeled wildtype CBX5 (**A**); red channel: TAMRA-labeled CBX5_low κ (**B**); merged (**C**). Colocalization of 10 µM FITC-labeled wildtype NPM1_ high κ and 10 µM TAMRA-labeled NPM1. Green channel: FITC-labeled wildtype NPM1_ high κ (**D**); red channel: TAMRA-labeled NPM1 (**E**); merged (**F**). Buffer: 50 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 10% PEG-8000. Scale bar 5 µm.

We next studied how κ value would affect phase separation behavior by performing pairwise mixing experiments using fluorescently labeled protein constructs. Mixtures of FITC-labeled wildtype CBX5 and TAMRA-labeled CBX5_low κ produced homogenous condensates (Figure 2.6 A-C). Similarly, TAMRA-labeled wildtype NPM1 and FITC-labeled NPM1_high κ also formed uniform phase-separated droplets (Figure 2.6 D-F). Surprisingly, when TAMRA-labeled CBX5_low κ and FITC-labeled NPM1_high κ were combined at equal molar ratios, in contrast to the native proteins, partially colocalized condensates was observed, with both merged droplets and protein-specific proteins (Figure 2.7). TAMRA-labeled CBX5_low κ protein formed both single-color red droplets (unmixed) and merged droplets containing FITC-labeled NPM1_high κ. NPM1_high κ protein was detected only in merged droplets with CBX5_low κ and did not form single-color green droplets. This asymmetric miscibility suggests that NPM1_high κ requires a binding partner to nucleate LLPS, whereas CBX5_low κ can self-condense but does not fully integrate with NPM1_high κ. This type of partial colocalization suggested that mismatched charge patterning reduces the favorability of heterotypic interactions, promoting partial demixing.

Table 2.1 Comparison of κ value of protein constructs.

Protein	κ value
CBX5	0.21
CBX5_low κ	0.06
NPM1	0.39
NPM1_high κ	0.46

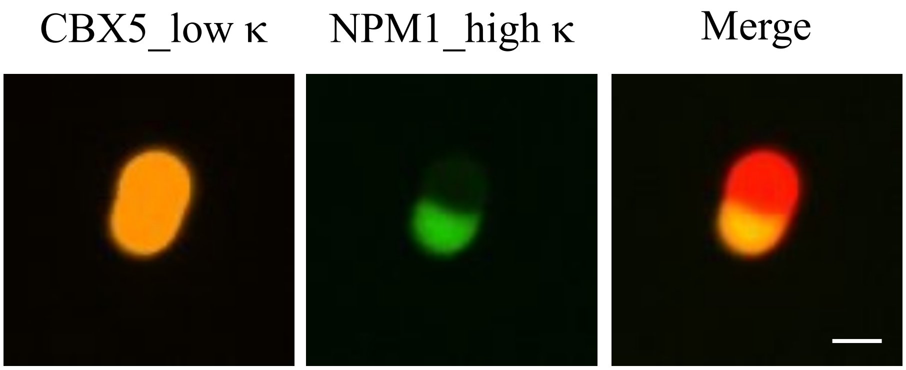


Figure 2.7. Partial colocalization between 10 μM TAMRA-labeled CBX5_low κ and 10 μM FITC-labeled NPM1_high κ and merged image on the right on a PF-127 coated confocal plate. Buffer: 50 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 10% PEG-8000. Scale bar = 2 μm.

2.3.4 κ Modulates Protein-Protein Interaction Strength

To directly probe how charge patterning influences intermolecular association, we used BLI to measure the binding affinities.[38–40] Wildtype CBX5 (no fluorescent tag) was used as the immobilized ligand. The binding affinity was tested against wildtype CBX5 and NPM1, CBX5_low κ and NPM1_high κ. CBX5 was covalently attached to the sensor surface using direct surface immobilization using amine-reactive coupling chemistry (EDC/NHS activation). The intrinsically disordered nature makes these proteins prone to nonspecific binding, making it hard to determine the number of substrates binding to the ligand. Here, we analyzed the binding data in two different fitting models: 1:1 binding model (Figure 2.8 A) and 2:1 heterogeneous binding model (Figure 2.8 B).[55–57] The 1:1 binding model assumes a simple interaction where one analyte molecule binds to a single site on the ligand, characterized by a single association and dissociation event. In contrast, the 2:1 heterogeneous ligand binding model accounts for more complex binding behaviors, including multiple binding sites or heterogeneous interactions.

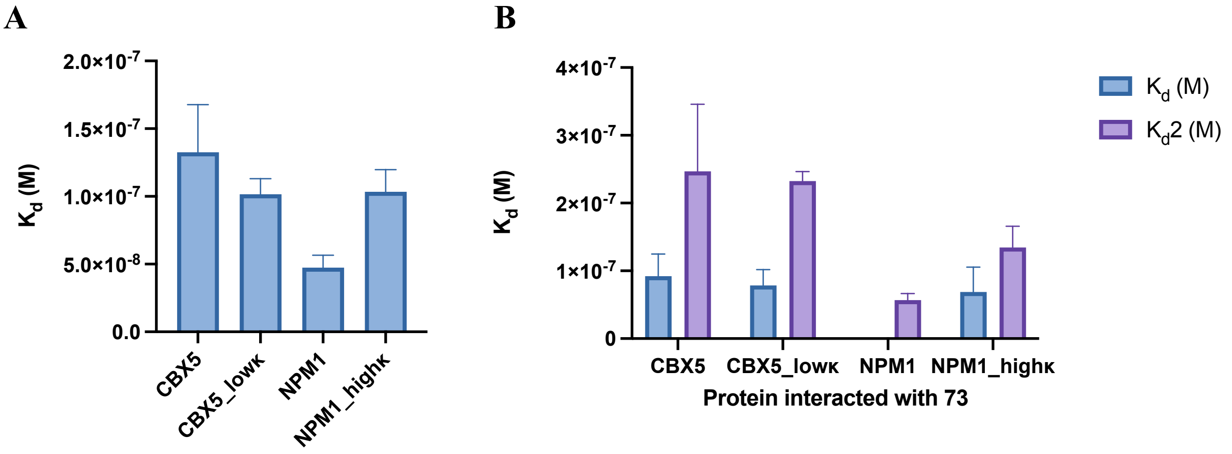


Figure 2.8. A. Comparison of K_d value calculated based on 1:1 binding model for different κ protein variants interacting with CBX5. N = 3, Error bar, SEM. **B.** Comparison of K_d1 and K_d2 value calculated based on 2:1 binding model for different κ protein variants interacting with CBX5. N = 3, Error bar, SEM.

The BLI data revealed that the binding strength depends not only on the amino acid composition, but also on how charged residues are distributed in the sequence. The affinity of immobilized CBX5 for CBX5 in the mobile phase was different from the affinity of immobilized CBX5 for CBX5_low κ in the mobile phase (Figure 2.8A). While all three of the protein constructs contain the same amino acid composition, wildtype CBX5 had K_d of 130 nM, and CBX5_low κ had K_d of 100 nM. These results indicate that charge patterning alone, independent of overall composition, significantly affects binding affinity. A similar trend was observed in Figure 2.8B when the 2:1 binding model was used for calculation. A different trend was observed when testing NPM1 variants as analytes. When using the 1:1 binding model, wildtype NPM1 had the stronger affinity for CBX5 (K_d = 47 nM), while NPM1_high κ showed progressively weaker binding (K_d = 103 nM). A similar trend was observed using the 2:1 binding model (Figure 2.8B).

2.4 Conclusion and Future Directions

In this study, we explored how linear charge patterning, quantified by the κ parameter, can influence the phase separation behavior and protein-protein interaction. We focused on four phase-separating proteins: wildtype CBX5 and NPM1, along with two engineered variants, CBX5_low κ (charge-dispersed) and NPM1_high κ (charge-clustered). These κ variants were designed to retain the same amino acid composition as the respective wildtype proteins, differing only in the distribution of charged residues. By altering the κ values, we generated "low κ" sequences with evenly mixed positive and negative charges and "high κ" sequences with extended clusters of similarly charged residues, allowing us to probe the role of charge patterning in condensate behavior and intermolecular binding.

We first tested wildtype CBX5 and NPM1 and noticed inconsistent results under confocal microscopy. On hydrophilic, BSA-coated glass, their droplets remained largely separate or showed partial fusion. Switching to a hydrophobic PF-127 coating enabled full colocalization of the two proteins. This outcome showed that wildtype CBX5 and NPM1 are compatible and can form homogenous condensates, but the surface coating can interfere with mixing. Our experiments showed that κ strongly influences how these proteins behave in solution. When we mixed proteins with similar or comparable κ values, such as wildtype CBX5 with CBX5_low κ or wildtype NPM1 with NPM1_high κ, they formed well-mixed and uniform phase-separated droplets. However, when we combined proteins with mismatched κ values, CBX5_low κ and NPM1_high κ, we observed partial colocalization. We observed both merged droplets and separate, single-color droplets (from CBX5_low κ), suggesting that the proteins did not fully interact with each other. This suggests that mismatched charge patterns reduce the strength of their interactions.

To further test how κ affects protein-protein interactions, we used BLI to measure binding strength between CBX5 and its variants or NPM1 variants. The results showed that κ had a clear effect on binding affinity. For example, CBX5_high κ bound more tightly to wildtype CBX5 than CBX5_low κ, even though the sequences had the same amino acid composition. On the other hand, NPM1_high κ showed weaker binding than wildtype NPM1. These differences confirm that it is not just the amino acid composition, but how the charges are arranged, that determines how strongly proteins interact.

Together, these results show that how charges are arranged in a protein plays a big role in both droplet formation and binding strength. Even if two proteins have the same composition, changing the layout of their charged residues can control whether they mix or stay separate, and how tightly they bind.

In the future, we can apply this approach to other disordered proteins to see if similar rules apply. It would also be useful to study how κ affects droplet behavior over time, such as fusion, viscosity, or exchange rates. Overall, κ is a simple but powerful tool for tuning protein phase separation behavior. Our findings demonstrate charge patterning, independent of sequence composition, can be another key determinant of condensate miscibility, positioning κ as a predictive design parameter for engineering phase-separation systems in both biological and synthetic contexts.

2.5 Experimental Methods

2.5.1 Protein Expression and Purification

The DNA fragments encoding four constructs: the wildtype CBX5, wildtype NPM1, CBX5_low κ, and NPM1_high κ, were ordered from Twist Bioscience, and cloned into NdeI and AvaI sites of vector pET-21b. The resulting plasmids were transformed into E. coli strain BL21(DE3). For both wildtype CBX5 and wildtype NPM1, the cells were grown at 37 °C in LB media with 100 μg/L ampicillin. For CBX5_low κ, and NPM1_high κ, the cells were grown at 37 °C in terrific broth with 100 μg/L ampicillin. At OD600 = 0.4, overexpression was induced using 1 mM IPTG at 18 °C for 12 hr. After cells were harvested through the centrifugation, cells were resuspended and lysed in 50 mM Tris buffer pH 7.5, 300 mM NaCl, 0.5 mM TCEP, 1 mM $MgCl_2$, 200 μg/mL lysozyme, 10 μg/mL DNase I, and cOmplete protease inhibitor using sonication. The protein was purified over a HisPrep FF column, and eluted with 50 mM Tris buffer pH 7.5, 300 mM NaCl, 0.5 mM TCEP and 500 mM imidazole. To remove the excess imidazole, the protein solution was dialyzed against 10 volumes of buffer containing 50 mM Tris buffer pH

7.5, 150 mM NaCl, 0.5 mM TCEP. The protein was further purified using a Superdex 200 pg size exclusion column. The purified proteins were further treated with 2 equivalents of maleimide FITC or maleimide-TAMRA and purified with another round of size-exclusion column to obtain the labeled proteins. To determine whether the proteins were labeled with fluorophores, I used a Nanodrop measurement of A280 and emission from the fluorescent dye and the results indicated 1:1 molar ratio of protein to dye. However, this bulk stoichiometry does not rule out the possibility that some proteins could have no dye modules, and some proteins could have two or three dye modules attached. Each protein has three Cys residues.

2.5.2 Confocal Microscopy

For glass-bottom confocal dishes coated with BSA, the glass dishes were incubated with 1% BSA in PBS for 1 hour and washed three times with MilliQ water and were then ready to use. For confocal dishes coated with PF-127, the glass dishes were incubated with 0.5% PF-127 in 50 mM Tris pH 7.5, 150 mM NaCl for 15 min at room temperature, and washed two times with the phase separation assay buffer. The slides or plates were then ready for use. It is important that once treated with PF127, the slides or plates must be kept under an aqueous solution. Drying irreversibly damages the self-assembled surfactant layer, which impacts the phase separation ability.

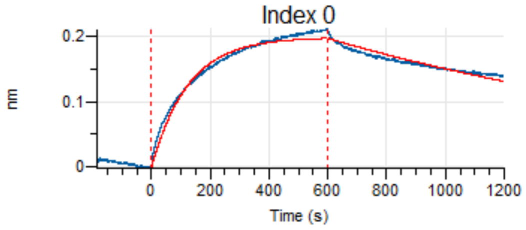
The labeled-protein was incubated with 50 mM Tris pH 7.5, 150 mM NaCl, 0.5mM TCEP, 10% PEG-8000 on treated confocal dishes for 30 minutes. Images were acquired on a Nikon AXR microscope using a 60X oil immersion objective. A 488-nm laser was used to excite the Fl-peptide samples; a 561-nm laser was used to excite the TMR-FUS samples. Images were processed in ImageJ.

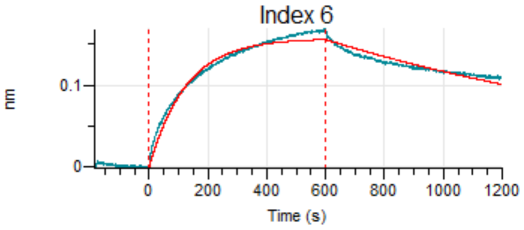
2.5.3 Bio-layer Interferometry

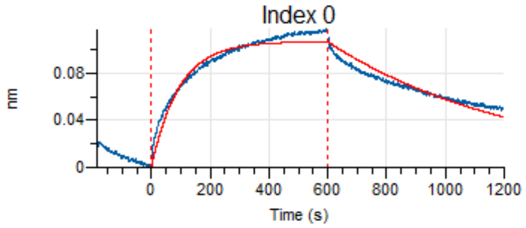
BLI experiments were performed on an Octet Red96 (Sartorius) instrument, with AR2G biosensors (Amine Reactive 2nd Gen). Before loading, 60 s baseline was recorded. 50 ug/mL CBX5 was loaded on the tips for 300 s, followed with 300 s quenching step. After loading, all biosensors underwent a 180 s baseline, and treated with 10 μM substrates of 600 s association and 600 s dissociation. Baseline measurements of loaded tips with no substrate were subtracted from their matched measurement of the loaded tip. Global kinetic fitting was used to determine K_ds across the dilution series.

2.5.3.1 Fitted Data Graph for Bio-layer Interferometry

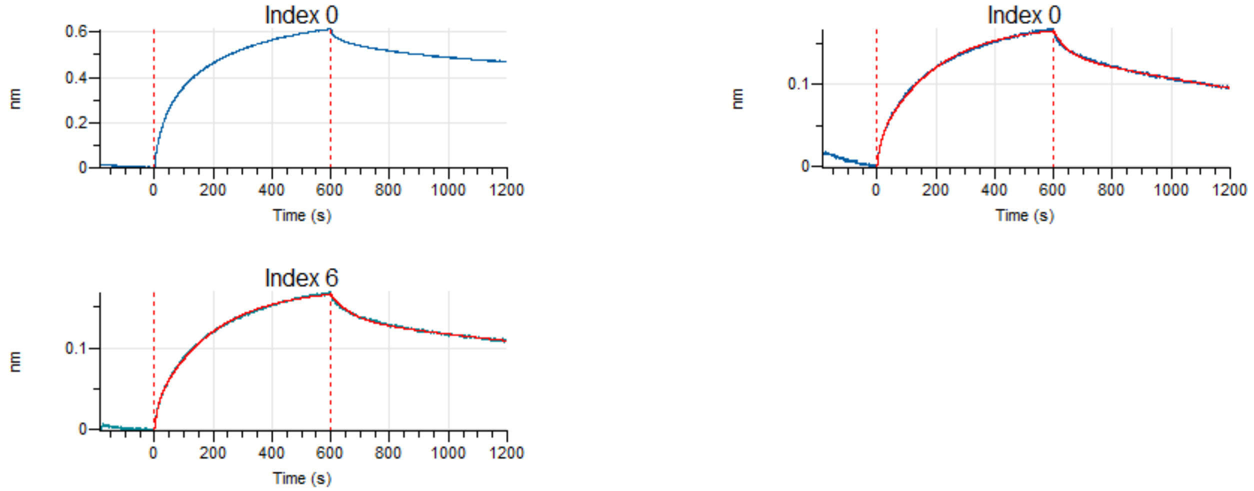
CBX5 (immobilized) interacted with CBX5 (substrate), 1:1 model



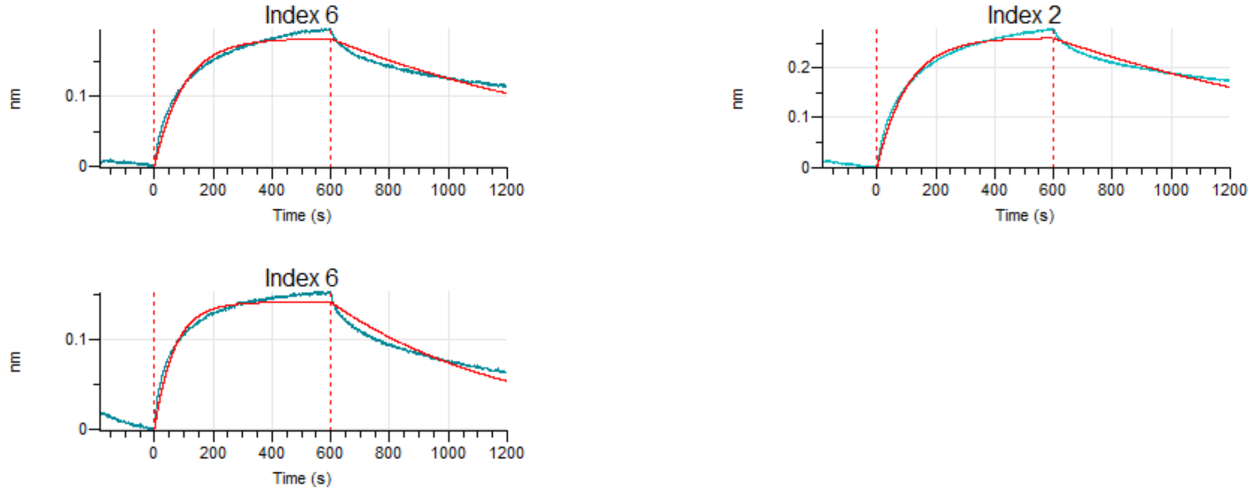




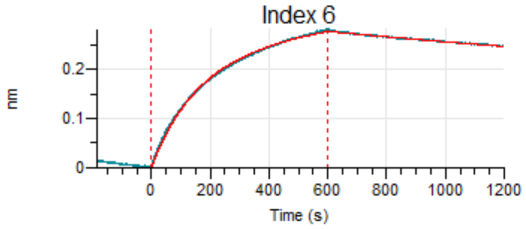
CBX5 (immobilized) interacted with CBX5 (substrate), 2:1 model

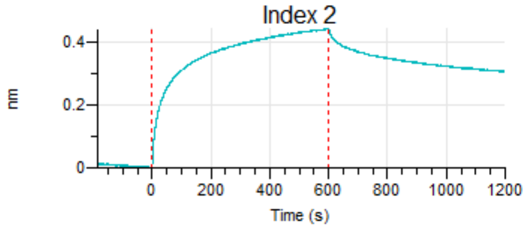


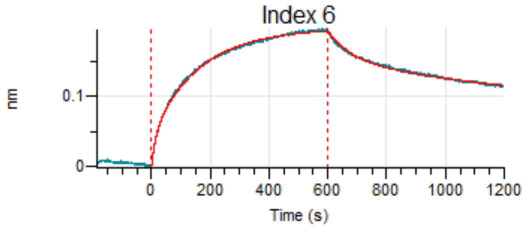
CBX5 (immobilized) interacted with CBX5_low κ (substrate), 1:1 model



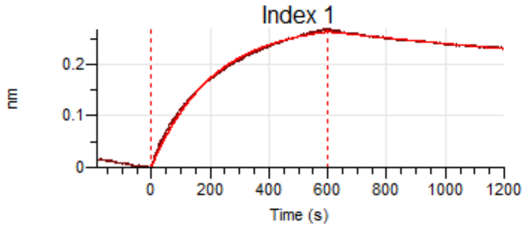
CBX5 (immobilized) interacted with CBX5_low κ (substrate), 2:1 model

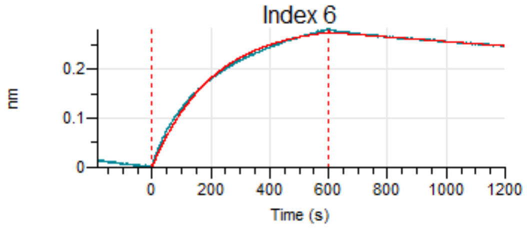


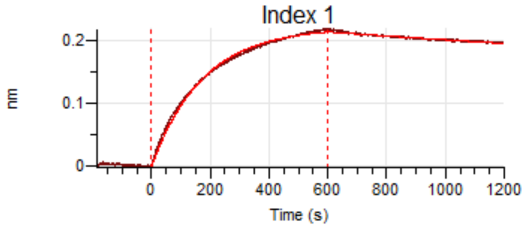




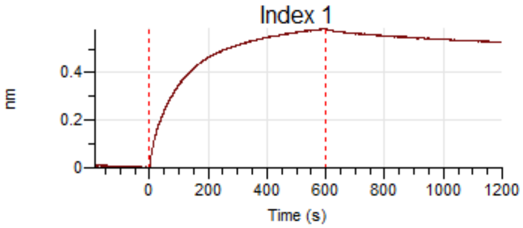
CBX5 (immobilized) interacted with NPM1 (substrate), 1:1 model

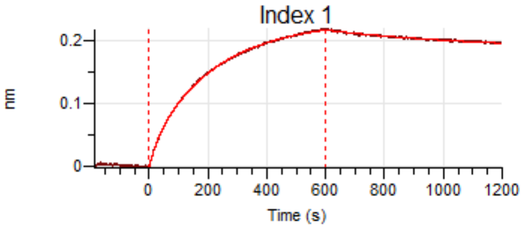


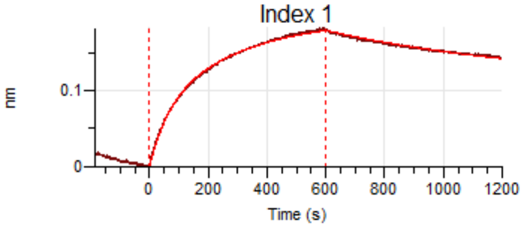




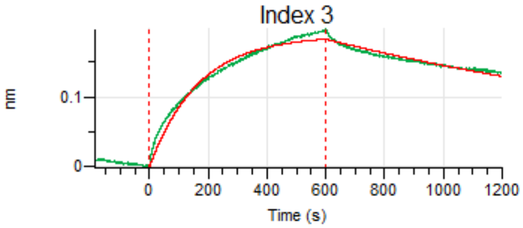
CBX5 (immobilized) interacted with NPM1 (substrate), 2:1 model

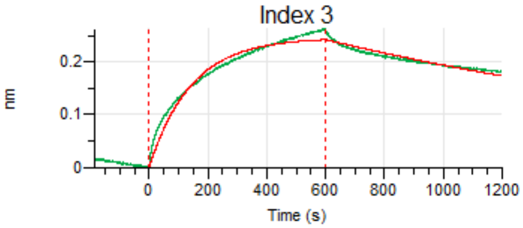


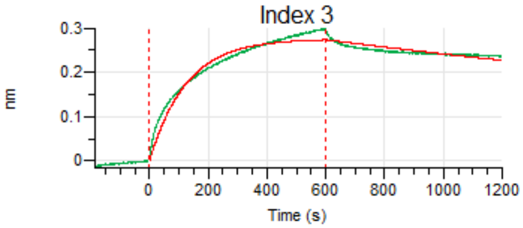




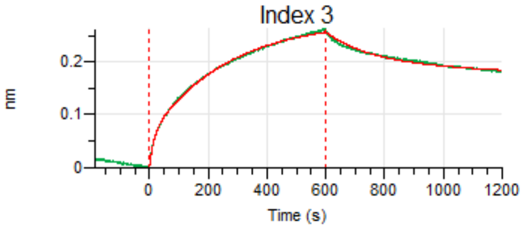
CBX5 (immobilized) interacted with NPM1_high κ (substrate), 1:1 model

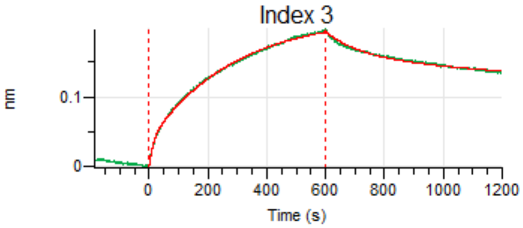


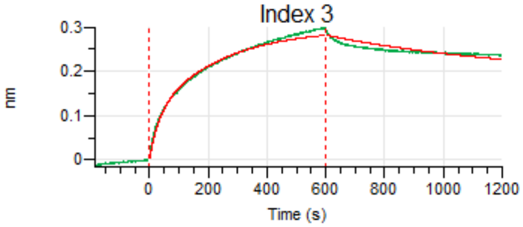




CBX5 (immobilized) interacted with NPM1_high κ (substrate), 2:1 model







2.5.4 Protein Constructs

(1) wildtype CBX5

Vector: pET-21b

Clone site: 5’: NdeI (CATATG)

Clone site: 3’: AvaI (CTCGAG)

Gene ordered: 599 bp ordered from Twist Bioscience

GAGATATA CATATG GGC AAG AAG ACG AAG CGT ACG GCT GAT AGC AGT TCG TCA
GAG GAT GAA GAG GAG TAC GTA GTG GAG AAA GTT TTG GAC CGT CGT GTT GTC
AAG GGC CAG GTA GAG TAC TTA TTG AAG TGG AAA GGT TTC AGC GAA GAG CAC
AAT ACT TGG GAG CCA GAA AAA AAC TTG GAC TGC CCC GAG CTG ATC TCA GAA
TTT ATG AAG AAA TAC AAA AAG ATG AAA GAA GGG GAG AAT AAC AAG CCA CGC
GAA AAG TCG GAG AGC AAT AAA CGT AAA AGC AAC TTC TCC AAC TCG GCG GAC
GAC ATC AAA TCG AAG AAA AAA CGT GAG CAG TCG AAC GAC ATC GCC CGC GGT
TTT GAA CGC GGT TTA GAA CCT GAG AAG ATT ATC GGC GCG ACA GAC AGT TGC
GGT GAT TTA ATG TTT TTA ATG AAA TGG AAA GAC ACT GAT GAG GCA GAC TTG GTT
CTG GCT AAA GAG GCT AAT GTC AAA TGC CCG CAG ATC GTC ATC GCA TTC TAT
GAA GAG CGC CTT ACA TGG CAT GCT TAC CCA GAG GAT GCA GAA AAT AAG GAG
AAG GAG ACA GCT AAA AGC CTCGAG CACCACCAC

Translation: (red – positively charged residues; blue – negatively charged residues)

MGKKTKRTADSSSSEDEEEYVVEKVLDRRVVKGQVEYLLKWKGFSEEHNTWEPEKNL
DCPELISEFMKKYKKMKEGENNKPREKSESNKRKSNFSNSADDIKSKKKREQSNDIARG
FERGLEPEKIIGATDSCGDLMFLMKWKDTDEADLVLAKEANVKCPQIVIAFYEERLTWH
AYPEDAENKEKETAKSLE-H6

(2) wildtype NPM1

Vector: pET-21b

Clone site: 5’: NdeI (CATATG)

Clone site: 3’: AvaI (CTCGAG)

Gene ordered: 905 bp ordered from Twist Bioscience

GAGATATA CATATG GAA GAT TCA ATG GAT ATG GAT ATG TCA CCA TTA CGC CCA
CAG AAT TAT TTA TTC GGG TGC GAA TTA AAG GCG GAC AAA GAC TAT CAT TTC AAG
GTT GAC AAT GAT GAG AAT GAA CAT CAA CTG TCT CTT CGC ACC GTT TCC CTG
GGC GCT GGC GCT AAA GAC GAA CTG CAC ATT GTG GAA GCC GAG GCA ATG AAC
TAT GAG GGC TCT CCG ATC AAA GTC ACC CTT GCC ACG TTA AAA ATG TCG GTA CAG
CCT ACA GTA TCC CTT GGT GGC TTT GAA ATC ACA CCT CCG GTG GTT TTA CGC TTG
AAG TGC GGG TCA GGG CCG GTT CAC ATC AGT GGC CAG CAC CTT GTC GCA GTT
GAG GAA GAT GCT GAG AGT GAA GAT GAA GAG GAG GAG GAT GTG AAA CTT TTG
AGT ATT AGT GGG AAG CGT TCT GCT CCT GGA GGC GGT TCC AAA GTC CCA CAA
AAG AAA GTG AAG CTT GCT GCT GAT GAG GAT GAT GAT GAC GAC GAC GAA GAA
GAT GAT GAT GAA GAC GAT GAT GAC GAT GAC TTC GAC GAT GAG GAA GCC GAA
GAA AAG GCT CCC GTA AAG AAG TCT ATT CGT GAT ACA CCA GCG AAA AAT GCA
CAA AAG TCA AAC CAA AAC GGA AAG GAT TCC AAA CCT AGC AGT ACC CCA CGT
TCG AAA GGT CAG GAG TCA TTT AAG AAA CAA GAA AAA ACT CCC AAG ACA CCA
AAA GGA CCC TCT TCG GTC GAA GAT ATT AAG GCA AAA ATG CAG GCG TCG ATT
GAG AAA GGA GGT AGC CTG CCC AAG GTC GAA GCA AAG TTC ATC AAC TAC GTT
AAA AAT TGT TTC CGT ATG ACC GAT CAG GAG GCT ATT CAA GAC CTT TGG CAA
TGG CGT AAA AGC CTCGAG CACCACCAC

Translation: (red – positively charged residues; blue – negatively charged residues)

MEDSMDMDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDEL
HIVEAEAMNYEGSPIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLV
AVEEDAESEDEEEEDVKLLSISGKRSAPGGGSKVPQKKVKLAADEDDDDDDEEDDDED
DDDDDFDDEEAEEKAPVKKSIRDTPAKNAQKSNQNGKDSKPSSTPRSKGQESFKKQEKT
PKTPKGPSSVEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTDQEAIQDLWQWRKS
LE-H6

(3) CBX5_low κ

Vector: pET-21b

Clone site: 5’: NdeI (CATATG)

Clone site: 3’: AvaI (CTCGAG)

Gene ordered: 599 bp ordered from Twist Bioscience

GAGATATA CATATG ACC AAA CGC GGT CCC GAC CGC GAC GGT CTT GTT GAG AAA
GAG AAC TGG TCA AGC GGG TGG AAG CGT AAT GAA CGT GCT TTG GTC GAA AAG
GGC GAG CTG AAG AAC ACA AAA AGC ATC AAG GAA TAT AAG GAG AAG ATC GAG
CAG AAG TCG TCG CTT AAA CTG GAG ATG AAA GAG TCT GAC AAA GAC GGC GTC
GCT CAT ACT AAG GCA AAT AAA GCC GAA ATG TGT GTG GTG AAG GAC AAG GAA
TTT CTT TAT GAA ATT CCT GTA CGT GGG GAA ACC ATT TAC AAA TGG ATG CTG GCG
AAA GAA TCT TCC GGC AAA GAA TTG TCG GAA AAA GAT GGT TAT CGT TCT GCG
GAC CTG ACC AAG GAC GTC AAT ATC TTT TTC TCC ATT CCA TTG AGC CTT AAG GAT
AAT GAA TTA GAG AAG TTT AAG CAG CAC GAC TCT GAG GCT AGT GTG GAA CGT
TTT AGC GCC GTT ACT GAC CTG TTT AAA GAA AAA AAT GAC TAT GCG GCA GAG
TGG GAA CCA AAC ATG GCA GAG CGT GAA AAA CCC CAG ACA AAA GAA ATT GAT
CGT GAG AAC CCG AAC CTCGAG CACCACCAC

Translation: (red – positively charged residues; blue – negatively charged residues)

MTKRGPDRDGLVEKENWSSGWKRNERALVEKGELKNTKSIKEYKEKIEQKSSLKLEMK
ESDKDGVAHTKANKAEMCVVKDKEFLYEIPVRGETIYKWMCAKESSGKELSEKDGYRS
ADCTKDVNIFFSIPLSLKDNELEKFKQHDSEASVERFSAVTDLFKEKNDYAAEWEPNMA
EREKPQTKEIDRENPNLE-H6

(4) NPM1_high κ

Vector: pET-21b

Clone site: 5’: NdeI (CATATG)

Clone site: 3’: AvaI (CTCGAG)

Gene ordered: 905 bp ordered from Twist Bioscience

GAGATATA CATATG AAA AAA TCC ATG AAG ATG CTG ATG TCT CCG TTG CGC CCA
CAG AAC TAT TTA TTC GGG TGT AAG CTG AAA GCG AAA AAG GAT TAC CAT TTT
AAA GTG AAA AAC AAA AAG AAC GAA CAT CAG CTT TCC TTG CGT ACT GTA TCA
TTA GGG GCC GGT GCG GAG GAT GAG TTA CAC ATC GTA GAA GCG GAG GCA ATG
AAT TAT GAG GGT TCT CCA ATC AAA GTT ACG CTG GCG ACG TTA AAA ATG AGT GTA
CAA CCC ACA GTT AGT CTT GGA GGC TTT GAG ATC ACG CCT CCG GTG GTT CTT
CGC CTG AAG TTG GGT TCG GGT CCC GTT CAT ATC TCT GGT CAG CAC TTG GTG
GCG GTC GAA GAA GAC GCT GAG TCT GAG GAT GAA GAG GAA GAG GAC GTG GAC
TTA CTG AGT ATT TCG GGA AAG CGT TCG GCA CCC GGT GGA GGC AGC AAG GTG
CCG CAA GAG GAC GTT GAT GAT GCT GCT GAC GAA GAC GAT GAT GAC GAC GAC
GAA GAA GAC GAC GAC GAG GAC GAC GAC GAT GAC GAT TTC GAT GAT GAG GAG
GCT GAA GAA AAA GCG CCC GTT GAG GAA TCC ATT CGT GAT ACC CCG GCT GAC
AAC GCA CAA GAT AGC AAT CAA AAT GGA AAA GAT AGT GAG CCT TCT TCC ACC
CCA CGT AGC GAG GGC CAA GAA TCG TTC GAC GAA CAA GAG AAG ACA CCA GAT
ACC CCA AAA GGA CCC TCG TCA GTG AAA AAG ATT AAA GCC AAA ATG CAA GCT
TCT ATT AAG AAG GGG GGC TCA CTG CCT AAG GTC AAA GCG AAA TTT ATC AAT
TAT GTG AAG AAT TTG TTT CGT ATG ACT AAG CAG AAA GCG ATT CAA GAC TTG
TGG CAA TGG CGC AAA AGC CTCGAG CACCACCAC

Translation: (red – positively charged residues; blue – negatively charged residues)

MKKSMKMLMSPLRPQNYLFGCKLKAKKDYHFKVKNKKNEHQLSLRTVSLGAGAEDEL
HIVEAEAMNYEGSPIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKLGSGPVHISGQHLVA
VEEDAESEDEEEEDVDLLSISGKRSAPGGGSKVPQEDVDDAADEDDDDDDEEDDDEDD
DDDDFDDEEAEEKAPVEESIRDTPADNAQDSNQNGKDSEPSSTPRSEGQESFDEQEKTPD
TPKGPSSVKKIKAKMQASIKKGGSLPKVKAKFINYVKNLFRMTKQKAIQDLWQWRKSL
E-H6

(5) mCherry-MED1 IDR

pET-28a-mCherry-MED1-IDR was a gift from Denes Hnisz (Addgene plasmid # 194545 ; http://n2t.net/addgene:194545 ; RRID:Addgene_194545).

Protein sequence: H_6-TEV-mCherry-MED1 IDR

MHHHHHHGIEENLYFQSGSGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEG
RPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWE
RVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMY
PEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTI
VEQYERAEGRHSTGGMDELYKGAPGSAGSAAGGSGEHHSGSQGPLLTTGDLGKEKTQK
RVKEGNGTSNSTLSGPGLDSKPGKRSRTPSNDGKSKDKPPKRKKADTEGKSPSHSSSNRP
FTPPTSTGGSKSPGSAGRSQTPPGVATPPIPKITIQIPKGTVMVGKPSSHSQYTSSGSVSSSG
SKSHHSHSSSSSSSASTSGKMKSSKSEGSSSSKLSSSMYSSQGSSGSSQSKNSSQSGGKPG
SSPITKHGLSSGSSSTKMKPQGKPSSLMNPSLSKPNISPSHSRPPGGSDKLASPMKPVPGT
PPSSKAKSPISSGSGGSHMSGTSSSSGMKSSSGLGSSGSLSQKTPPSSNSCTASSSSFSSSGS
SMSSSQNQHGSSKGKSPSRNKKPSLTAVIDKLKHGVVTSGPGGEDPLDGQMGVSTNSSS
HPMSSKHNMSGGEFQGKREKSDKDKSKVSTSGSSVDSSKKTSESKNVGSTGVAKIIISKH
DGGSPSIKAKVTLQKPGESSGEGLRPQMASSKNYGSPLISGSTPKHERGSPSHSKSPAYTP
QNLDSESESGSSIAEKSYQNSPSSDDGIRPLPEYSTEKHKKHKKEKKKVKDKDRDRDRD
KDRDKKKSHSIKPESWSKSPISSDQSLSMTSNTILSADRPSRLSPDFMIGEEDDDL

(6) mEGFP-HMGB1

pET-45b-mEGFP-HMGB1-WT was a gift from Denes Hnisz (Addgene plasmid # 194543 ; http://n2t.net/addgene:194543 ; RRID:Addgene_194543).

Protein sequence: H_6-TEV-mEGFP-HMGB1

MHHHHHHGIEENLYFQSGSGVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATN
GKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIS
FKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQK
NGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDH
MVLLEFVTAAGITLGMDELYTAMGKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASV
NFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGETKKKFKDPNA
PKRPPSAFFLFCSEYRPKIKGEHPGLSIGDVAKKLGEMWNNTAADDKQPYEKKAAKLKE
KYEKDIAAYRAKGKPDAAKKGVVKAEKSKKKKEEEEDEEDEEDEEEEEDEEDEDEEED
DDDE

(7) MBP-FUS-C

Vector: pMal-C5x

Protein sequence: MBP-TEV-FUS(372-526)-H_6

With MBP Tag:

ATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAAC
GGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTT
GAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGC
CCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGT
TGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGG
ATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATC
GCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCC
GGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGC
AAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTA
TGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAG
CGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCG
ATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACG
GCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTAC
TGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTA
TTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGC
TGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGC
TGAAGTCTTACGAGGAAGAGTTGGTGAAAGATCCGCGTATTGCCGCCACTATGGAAA
ACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGC
CGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCT
GAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACA
ACCTCGGGGAAAACTTATACTTTCAGGGGCGCGCGGACTTTAACCGCGGGGGCGGAA
ATGGCCGTGGGGGTCGTGGCCGTGGGGGTCCTATGGGCCGCGGAGGCTATGGCGGAG
GTGGTTCGGGTGGAGGCGGTCGCGGGGGCTTTCCTTCGGGAGGGGGCGGCGGTGGG
GGACAGCAGCGCGCAGGCGATTGGAAATGTCCGAATCCGACCTGCGAGAATATGAAT
TTTAGTTGGCGCAATGAATGCAATCAGTGTAAAGCCCCGAAACCGGATGGTCCAGGC
GGGGGTCCAGGTGGGAGCCACATGGGGGGAAATTATGGAGATGATCGCCGCGGCGGT
CGTGGGGGTTATGATCGTGGTGGTTATCGTGGGCGTGGGGGAGATCGTGGGGGATTTC
GTGGCGGGCGCGGTGGGGGTGATCGTGGCGGTTTTGGGCCTGGGAAAATGGATTCGC
GTGGCGAGCACCGCCAAGATCGCCGCGAGCGCCCGTATCATCATCACCATCATCATTA
AGAATTC

Translation:

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPD
IIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNK
DLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDI
KDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNID
TSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNK
DKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQ
TVDEALKDAQTNSSSNNNNNNNNNNLGENLYFQGRADFNRGGGNGRGGRGRGGPMG
RGGYGGGGSGGGGRGGFPSGGGGGGGQQRAGDWKCPNPTCENMNFSWRNECNQCKA
PKPDGPGGGPGGSHMGGNYGDDRRGGRGGYDRGGYRGRGGDRGGFRGGRGGGDRG
GFGPGKMDSRGEHRQDRRERPYHHHHHH

2.6 References

(1) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular Condensates: Organizers of Cellular Biochemistry. *Nat Rev Mol Cell Biol* 2017, *18* (5), 285–298. https://doi.org/10.1038/nrm.2017.7.

(2) Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu Rev Cell Dev Biol* 2014, *30*, 39–58. https://doi.org/10.1146/annurev-cellbio-100913-013325.

(3) Dimitrijević, D.; Bösenhofer, M.; Harasek, M. Liquid–Liquid Phase Separation of Two Non-Dissolving Liquids—A Mini Review. *Processes* 2023, *11* (4), 1145. https://doi.org/10.3390/pr11041145.

(4) McCarty, J.; Delaney, K. T.; Danielsen, S. P. O.; Fredrickson, G. H.; Shea, J. E. Complete Phase Diagram for Liquid-Liquid Phase Separation of Intrinsically Disordered Proteins. *Journal of Physical Chemistry Letters* 2019, *10* (8), 1644–1652. https://doi.org/10.1021/acs.jpclett.9b00099.

(5) Falahati, H.; Haji-Akbari, A. Thermodynamically Driven Assemblies and Liquid-Liquid Phase Separations in Biology. *Soft Matter* 2019, *15* (6), 1135–1154. https://doi.org/10.1039/c8sm02285b.

(6) Shin, Y.; Brangwynne, C. P. Liquid Phase Condensation in Cell Physiology and Disease. *Science (1979)* 2017, *357* (6357). https://doi.org/10.1126/science.aaf4382.

(7) Wang, B.; Zhang, L.; Dai, T.; Qin, Z.; Lu, H.; Zhang, L.; Zhou, F. Liquid–Liquid Phase Separation in Human Health and Diseases. *Signal Transduct Target Ther* 2021, *6* (1). https://doi.org/10.1038/s41392-021-00678-1.

(8) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A Framework for Understanding the Functions of Biomolecular Condensates across Scales. *Nat Rev Mol Cell Biol* 2021, *22* (3), 215–235. https://doi.org/10.1038/s41580-020-00303-z.

(9) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol Cell* 2015, *57* (5), 936–947. https://doi.org/10.1016/j.molcel.2015.01.013.

(10) Garcia Quiroz, F.; Li, N. K.; Roberts, S.; Weber, P.; Dzuricky, M.; Weitzhandler, I.; Yingling, Y. G.; Chilkoti, A. Intrinsically Disordered Proteins Access a Range of Hysteretic Phase Separation Behaviors. *Sci. Adv* 2019, *5*, 5177–5195.

(11) Harmon, T. S.; Holehouse, A. S.; Rosen, M. K.; Pappu, R. V. Intrinsically Disordered Linkers Determine the Interplay between Phase Separation and Gelation in Multivalent Proteins. *Elife* 2017, *6*, e30294. https://doi.org/10.7554/eLife.30294.001.

(12) Borcherds, W.; Bremer, A.; Borgia, M. B.; Mittag, T. How Do Intrinsically Disordered Protein Regions Encode a Driving Force for Liquid–Liquid Phase Separation? *Curr Opin Struct Biol* 2021, *67*, 41–50. https://doi.org/10.1016/j.sbi.2020.09.004.

(13) Lu, T.; Spruijt, E. Multiphase Complex Coacervate Droplets. *J Am Chem Soc* 2020, *142* (6), 2905–2914. https://doi.org/10.1021/jacs.9b11468.

(14) Wei, M.; Wang, X.; Qiao, Y. Multiphase Coacervates: Mimicking Complex Cellular Structures through Liquid-Liquid Phase Separation. *Chemical Communications* 2024, *60* (90), 13169–13178. https://doi.org/10.1039/d4cc04533e.

(15) Lafontaine, D. L. J.; Riback, J. A.; Bascetin, R.; Brangwynne, C. P. The Nucleolus as a Multiphase Liquid Condensate. *Nat Rev Mol Cell Biol* 2021, *22* (3), 165–182. https://doi.org/10.1038/s41580-020-0272-6.

(16) Feric, M.; Vaidya, N.; Harmon, T. S.; Mitrea, D. M.; Zhu, L.; Richardson, T. M.; Kriwacki, R. W.; Pappu, R. V.; Brangwynne, C. P. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* 2016, *165* (7), 1686–1697. https://doi.org/10.1016/j.cell.2016.04.047.

(17) Wang, J.; Choi, J. M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 2018, *174* (3), 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006.

(18) Wake, N.; Weng, S.-L.; Zheng, T.; Wang, S.-H.; Kirilenko, V.; Mittal, J.; Fawzi, N. L. Expanding the Molecular Grammar of Polar Residues and Arginine in FUS Phase Separation. *Nat Chem Biol* 2025. https://doi.org/10.1038/s41589-024-01828-6.

(19) Qamar, S.; Wang, G. Z.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Ströhl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.; Tartaglia, G. G.; Farrer, L. A.; Kaminski Schierle, G. S.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P.

FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* 2018, *173* (3), 720-734.e15. https://doi.org/10.1016/j.cell.2018.03.056.

(20) Kaur, T.; Raju, M.; Alshareedah, I.; Davis, R. B.; Potoyan, D. A.; Banerjee, P. R. Sequence-Encoded and Composition-Dependent Protein-RNA Interactions Control Multiphasic Condensate Morphologies. *Nat Commun* 2021, *12* (1), 872. https://doi.org/10.1038/s41467-021-21089-4.

(21) Niu, J.; Qiu, C.; Abbott, N. L.; Gellman, S. H. Formation of versus Recruitment to RNA-Rich Condensates: Controlling Effects Exerted by Peptide Side Chain Identity. *J Am Chem Soc* 2022, *144* (23), 10386–10395. https://doi.org/10.1021/jacs.2c02222.

(22) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V; Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains. *Science (1979)* 2020, *367*, 694–699.

(23) Madinya, J. J.; Chang, L. W.; Perry, S. L.; Sing, C. E. Sequence-Dependent Self-Coacervation in High Charge-Density Polyampholytes. *Mol Syst Des Eng* 2020, *5* (3), 632–644. https://doi.org/10.1039/c9me00074g.

(24) Das, S.; Lin, Y.-H.; Vernon, R. M.; Forman-Kay, J. D.; Chan, H. S. Comparative Roles of Charge, π, and Hydrophobic Interactions in Sequence-Dependent Phase Separation of Intrinsically Disordered Proteins. *Proc Natl Acad Sci U S A* 2020, *117* (46), 28795–28805. https://doi.org/10.1073/pnas.2008122117/-/DCSupplemental.y.

(25) Chang, L. W.; Lytle, T. K.; Radhakrishna, M.; Madinya, J. J.; Vélez, J.; Sing, C. E.; Perry, S. L. Sequence and Entropy-Based Control of Complex Coacervates. *Nat Commun* 2017, *8* (1). https://doi.org/10.1038/s41467-017-01249-1.

(26) Cakmak, F. P.; Choi, S.; Meyer, M. C. O.; Bevilacqua, P. C.; Keating, C. D. Prebiotically-Relevant Low Polyion Multivalency Can Improve Functionality of Membraneless Compartments. *Nat Commun* 2020, *11* (1), 1–11. https://doi.org/10.1038/s41467-020-19775-w.

(27) Krainer, G.; Welsh, T. J.; Joseph, J. A.; Espinosa, J. R.; Wittmann, S.; de Csilléry, E.; Sridhar, A.; Toprakcioglu, Z.; Gudiškytė, G.; Czekalska, M. A.; Arter, W. E.; Guillén-Boixet, J.; Franzmann, T. M.; Qamar, S.; George-Hyslop, P. S.; Hyman, A. A.; Collepardo-Guevara, R.; Alberti, S.; Knowles, T. P. J. Reentrant Liquid Condensate Phase of Proteins Is Stabilized by Hydrophobic and Non-Ionic Interactions. *Nat Commun* 2021, *12* (1), 1–14. https://doi.org/10.1038/s41467-021-21181-9.

(28) Akahoshi, Y.; Sugai, H.; Mimura, M.; Shinkai, Y.; Kurita, R.; Shiraki, K.; Tomita, S. Phase-Separation Propensity of Non-Ionic Amino Acids in Peptide-Based Complex Coacervation Systems. *Biomacromolecules* 2023, *24* (2), 704–713. https://doi.org/10.1021/acs.biomac.2c01148.

(29) Fedyukina, D. V.; Jennaro, T. S.; Cavagnero, S. Charge Segregation and Low Hydrophobicity Are Key Features of Ribosomal Proteins from Different Organisms. *Journal of Biological Chemistry* 2014, *289* (10), 6740–6750. https://doi.org/10.1074/jbc.M113.507707.

(30) Das, R. K.; Pappu, R. V. Conformations of Intrinsically Disordered Proteins Are Influenced by Linear Sequence Distributions of Oppositely Charged Residues. *Proc Natl Acad Sci U S A* 2013, *110* (33), 13392–13397. https://doi.org/10.1073/pnas.1304749110.

(31) Guthmann, M.; Burton, A.; Torres-Padilla, M. Expression and Phase Separation Potential of Heterochromatin Proteins during Early Mouse Development. *EMBO Rep* 2019, *20* (12), e47952. https://doi.org/10.15252/embr.201947952.

(32) Hansen, J. C.; Maeshima, K.; Hendzel, M. J. The Solid and Liquid States of Chromatin. *Epigenetics Chromatin* 2021, *14* (1), 50. https://doi.org/10.1186/s13072-021-00424-5.

(33) Mitrea, D. M.; Cika, J. A.; Guy, C. S.; Ban, D.; Banerjee, P. R.; Stanley, C. B.; Nourse, A.; Deniz, A. A.; Kriwacki, R. W. Nucleophosmin Integrates within the Nucleolus via Multi-Modal Interactions with Proteins Displaying R-Rich Linear Motifs and RRNA. *Elife* 2016, *5* (FEBRUARY2016), e13571. https://doi.org/10.7554/eLife.13571.001.

(34) Mitrea, D. M.; Cika, J. A.; Stanley, C. B.; Nourse, A.; Onuchic, P. L.; Banerjee, P. R.; Phillips, A. H.; Park, C. G.; Deniz, A. A.; Kriwacki, R. W. Self-Interaction of NPM1 Modulates Multiple Mechanisms of Liquid-Liquid Phase Separation. *Nat Commun* 2018, *9* (1), 842. https://doi.org/10.1038/s41467-018-03255-3.

(35) Strom, A. R.; Emelyanov, A. V.; Mir, M.; Fyodorov, D. V.; Darzacq, X.; Karpen, G. H. Phase Separation Drives Heterochromatin Domain Formation. *Nature* 2017, *547* (7662), 241–245. https://doi.org/10.1038/nature22989.

(36) Larson, A. G.; Elnatan, D.; Keenen, M. M.; Trnka, M. J.; Johnston, J. B.; Burlingame, A. L.; Agard, D. A.; Redding, S.; Narlikar, G. J. Liquid Droplet Formation by HP1α Suggests

a Role for Phase Separation in Heterochromatin. *Nature* 2017, *547* (7662), 236–240. https://doi.org/10.1038/nature22822.

(37) Mensah, M. A.; Niskanen, H.; Magalhaes, A. P.; Basu, S.; Kircher, M.; Sczakiel, H. L.; Reiter, A. M. V.; Elsner, J.; Meinecke, P.; Biskup, S.; Chung, B. H. Y.; Dombrowsky, G.; Eckmann-Scholz, C.; Hitz, M. P.; Hoischen, A.; Holterhus, P. M.; Hülsemann, W.; Kahrizi, K.; Kalscheuer, V. M.; Kan, A.; Krumbiegel, M.; Kurth, I.; Leubner, J.; Longardt, A. C.; Moritz, J. D.; Najmabadi, H.; Skipalova, K.; Snijders Blok, L.; Tzschach, A.; Wiedersberg, E.; Zenker, M.; Garcia-Cabau, C.; Buschow, R.; Salvatella, X.; Kraushar, M. L.; Mundlos, S.; Caliebe, A.; Spielmann, M.; Horn, D.; Hnisz, D. Aberrant Phase Separation and Nucleolar Dysfunction in Rare Genetic Diseases. *Nature* 2023, *614* (7948), 564–571. https://doi.org/10.1038/s41586-022-05682-1.

(38) Jug, A.; Bratkovič, T.; Ilaš, J. Biolayer Interferometry and Its Applications in Drug Discovery and Development. *Trends in Analytical Chemistry* 2024, *176*, 117741. https://doi.org/10.1016/j.trac.2024.117741.

(39) Desai, M.; Di, R.; Fan, H. Application of Biolayer Interferometry (BLI) for Studying Protein-Protein Interactions in Transcription. *J Vis Exp* 2019, *149*. https://doi.org/doi:10.3791/59687.

(40) Petersen, R. L. Strategies Using Bio-Layer Interferometry Biosensor Technology for Vaccine Research and Development. *Biosensors (Basel)* 2017, *7* (4), 49. https://doi.org/10.3390/bios7040049.

(41) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; Tompa, P.; Fuxreiter, M.

Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 2018, *28* (6), 420–435. https://doi.org/10.1016/j.tcb.2018.02.004.

(42) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu Rev Phys Chem* 2020, *71*, 53–75. https://doi.org/10.1146/annurev-physchem-071819-113553.

(43) Guo, Y. E.; Manteiga, J. C.; Henninger, J. E.; Sabari, B. R.; Dall'Agnese, A.; Hannett, N. M.; Spille, J. H.; Afeyan, L. K.; Zamudio, A. V.; Shrinivas, K.; Abraham, B. J.; Boija, A.; Decker, T. M.; Rimel, J. K.; Fant, C. B.; Lee, T. I.; Cisse, I. I.; Sharp, P. A.; Taatjes, D. J.; Young, R. A. Pol II Phosphorylation Regulates a Switch between Transcriptional and Splicing Condensates. *Nature* 2019, *572* (7770), 543–548. https://doi.org/10.1038/s41586-019-1464-0.

(44) Boija, A.; Klein, I. A.; Sabari, B. R.; Dall'Agnese, A.; Coffey, E. L.; Zamudio, A. V.; Li, C. H.; Shrinivas, K.; Manteiga, J. C.; Hannett, N. M.; Abraham, B. J.; Afeyan, L. K.; Guo, Y. E.; Rimel, J. K.; Fant, C. B.; Schuijers, J.; Lee, T. I.; Taatjes, D. J.; Young, R. A. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 2018, *175* (7), 1842-1855.e16. https://doi.org/10.1016/j.cell.2018.10.042.

(45) Peng, L.; Li, E. M.; Xu, L. Y. From Start to End: Phase Separation and Transcriptional Regulation. *BBA - Gene Regulatory Mechanisms* 2020, *1863* (12), 194641. https://doi.org/10.1016/j.bbagrm.2020.194641.

(46) Wagh, K.; Garcia, D. A.; Upadhyaya, A. Phase Separation in Transcription Factor Dynamics and Chromatin Organization. *Curr Opin Struct Biol* 2021, *71*, 148–155. https://doi.org/10.1016/j.sbi.2021.06.009.

(47) Stortz, M.; Presman, D. M.; Levi, V. Transcriptional Condensates: A Blessing or a Curse for Gene Regulation? *Commun Biol* 2024, *7* (1), 187. https://doi.org/10.1038/s42003-024-05892-5.

(48) Sharp, P. A.; Chakraborty, A. K.; Henninger, J. E.; Young, R. A. RNA in Formation and Regulation of Transcriptional Condensates. *RNA* 2022, *28*, 52. https://doi.org/10.1261/rna.

(49) Li, P.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q. X.; Nixon, B. T.; Rosen, M. K. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* 2012, *483* (7389), 336–340. https://doi.org/10.1038/nature10879.

(50) Sabari, B. R.; Dall'Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C.; Li, C. H.; Guo, Y. E.; Day, D. S.; Schuijers, J.; Vasile, E.; Malik, S.; Hnisz, D.; Lee, T. I.; Cisse, I. I.; Roeder, R. G.; Sharp, P. A.; Chakraborty, A. K.; Young, R. A. Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. *Science (1979)* 2018, *361* (6400), 379. https://doi.org/10.1126/science.aar3958.

(51) Ulloa, L.; Messmer, D. High-Mobility Group Box 1 (HMGB1) Protein: Friend and Foe. *Cytokine Growth Factor Rev* 2006, *17* (3), 189–201. https://doi.org/10.1016/j.cytogfr.2006.01.003.

(52) Pathak, J.; Priyadarshini, E.; Rawat, K.; Bohidar, H. B. Complex Coacervation in Charge Complementary Biopolymers: Electrostatic versus Surface Patch Binding. *Advances in Colloid and Interface Science*. Elsevier B.V. December 1, 2017, pp 40–53. https://doi.org/10.1016/j.cis.2017.10.006.

(53) Yao, R. W.; Rosen, M. K. Advanced Surface Passivation for High-Sensitivity Studies of Biomolecular Condensates. *Proc Natl Acad Sci U S A* 2024, *121* (22). https://doi.org/10.1073/pnas.2403013121.

(54) Holehouse, A. S.; Das, R. K.; Ahad, J. N.; Richardson, M. O. G.; Pappu, R. V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys J* 2017, *112* (1), 16–21. https://doi.org/10.1016/j.bpj.2016.11.3200.

(55) Tso, S. C.; Chen, Q.; Vishnivetskiy, S. A.; Gurevich, V. V.; Iverson, T. M.; Brautigam, C. A. Using Two-Site Binding Models to Analyze Microscale Thermophoresis Data. *Anal Biochem* 2018, *540–541*, 64–75. https://doi.org/10.1016/j.ab.2017.10.013.

(56) Jarmoskaite, I.; Alsadhan, I.; Vaidyanathan, P. P.; Herschlag, D. How to Measure and Evaluate Binding Affinities. *Elife* 2020, *9*, 1–34. https://doi.org/10.7554/ELIFE.57264.

(57) Orthwein, T.; Huergo, L. F.; Forchhammer, K.; Selim, K. A. Kinetic Analysis of a Protein-Protein Complex to Determine Its Dissociation Constant (Kd) and the Effective Concentration (Ec50) of an Interplaying Effector Molecule Using Bio-Layer Interferometry. *Bio Protoc* 2021, *11* (17), e4152. https://doi.org/10.21769/bioprotoc.4152.

Chapter 3

A New Two-component Phase Separation System that Features Coiled-coil Heterodimer Formation

Experiments were designed by Ruiwen Xu and Samuel H. Gellman.

Experiments were performed by Ruiwen Xu and Jiawen Huang.

Graphics were prepared using Microsoft PowerPoint, GraphPad Prism, ImageJ.

3.1 Abstract

Liquid–liquid phase separation (LLPS) enables cells to compartmentalize biochemical reactions without membranes. Engineering synthetic condensates offers new opportunities for organizing molecules and biochemical reactions spatially and temporally. Here, we designed a synthetic system that combines orthogonal coiled-coil (CC) peptide with phase-separating peptide to construct programmable condensates. We designed fusion constructs in which CC peptides act as specific dimerization tags, while Arg-rich and Glu-rich peptides drive phase separation through multivalent electrostatic interactions. We found that coiled-coil interactions were necessary to initiate phase separation under the tested conditions, and that backbone flexibility in phase-separating domains can influence the phase separation behavior. Constructs with flexible motifs formed condensates more readily and at lower concentrations compared to those with rigid motifs. This system enables the creation of coexisting and potentially interacting condensate compartments. In future applications, the introduction of bifunctional CC peptides could allow fusion of distinct condensates, enabling spatial-temporal control over synthetic biochemical reactions.

3.2 Introduction

A fundamental problem in cell biology is how the densely packed cellular space is organized to allow control over complex biochemical reactions in space and time. One way to achieve spatiotemporal control is to regulate the localization of reaction components.[1] While membrane-bound organelles, such as nucleus[2] or mitochondria,[3] provide membrane boundaries,[1,4] many crucial intracellular processes, such as RNA splicing, transcription and stress response,[5] are governed by membraneless organelles.[6–9] Previous studies have revealed that phase separation gives rise to membrane-less organelles within cells.[10–12] Under certain conditions, macromolecules, such as proteins and RNAs, within cells can be triggered to organize and condense into these liquid droplets that lack a surrounding membrane.[5,9,12]

Over the past decade, LLPS has gained interest in synthetic biology as a strategy to create programmable compartments for organizing and regulating biochemical reactions *in vitro*.[5,9,13] One exciting area is the creation of multiphase condensates that contain multiple, coexisting liquid phases with distinct compositions and properties.[14–19] Such multiphasic organization is observed in both natural and synthetic systems, as layered or spatially discrete condensates with unique compositions and material properties. A well-known natural example is the nucleolus, a membrane-less organelle found in the cell nucleus, which contains concentric layers enriched with different proteins and RNA molecules. These layers are thought to help organize ribosome assembly by separating steps of the process into different regions.[20–22] In synthetic systems, similar multiphase behavior can also be engineered. Banerjee et al. reported a three-component system involving a prion-like polypeptide (PLP), an Arg-rich polypeptide (RRP), and RNA. In this system, both PLP and RNA can bind to RRP, but they compete for this shared partner RRP. This competitive interaction drives the system to separate into two distinct coexisting condensates: one that is rich in PLP (homotypic PLP droplets) and another that contains both RRP and RNA (heterotypic RRP-RNA droplets).[23]

Despite these advances, engineering synthetic multiphase condensates still remains challenging due to the lack of structure of many phase-separating sequences[7,12,24] and the difficulty of independently tuning the assembly, stability, and identity of each phase. Here, we propose a strategy to separate phase separation ability from molecular interaction specificity by using bio-orthogonal coiled-coil (CC) peptides. Coiled-coils are well-characterized protein motifs that form stable heterodimers based on specific electrostatic and hydrophobic interactions.[25–28] Importantly, orthogonal CC pairs can be engineered to avoid cross-association, allowing multiple specific interactions to occur independently in the same system.[29–32] The ability to define interaction partners makes CC peptides ideal tools for assembling programmable interactions in complex mixtures.

In this work, we developed a modular system that uses a combination of phase-separating peptides and orthogonal CC peptides to build synthetic condensates (Figure 3.1.A). Each compartment contains a phase-separating domain responsible for droplet formation, fused to a CC domain that acts as a selective assembly module. This design allows us to separate the roles of phase separation and molecular specificity, while still allowing them to work together cooperatively.

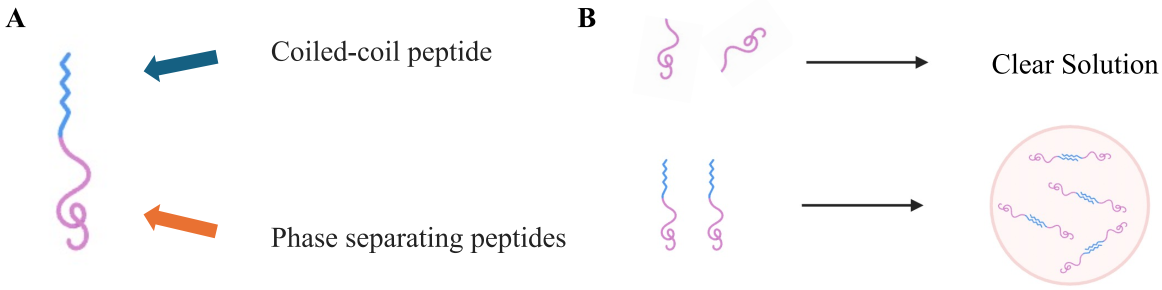


Figure 3.1. A. Peptide construct design: a coiled-coil peptide fused with a phase-separating peptide. **B.** Schematic graph showing that the phase-separating peptides, without the CC peptides, were not sufficient to drive phase separation formation; however, the dimerization of CC pairs can effectively bring the phase-separating peptides into proximity to promote phase separation.

We focused on Arg-rich and Glu-rich peptides, which interact through charge-charge interactions.[33–38] To explore how different structural features affect condensate formation, we designed the Arg-rich peptides with different motif regions in the sequence. In particular, we compared flexible motifs (GGSSGG) to rigid motifs (APAPAP).[39] To overcome peptide length limitations during synthesis, we developed a conjugation strategy using iodoacetyl- and cysteinetagged fragments that form covalent bonds through an SN2 reaction.[40] Our results show that CC-driven dimerization is sufficient to trigger phase separation even when the unfused phase-separating peptides do not form condensates on their own. Moreover, the linker flexibility plays an important role in the phase separation behavior.

Overall, this work shows that orthogonal coiled-coil peptides can be used as a tool to control phase separation and build synthetic condensates. This system gives us a new way to design programmable, self-assembling materials that can be used to organize reactions and build complex biochemical systems *in vitro*.

3.3 Results and Discussion

3.3.1 Design of Modular Coiled-Coil–Phase Separation Constructs

To investigate the feasibility of constructing protein condensates using orthogonal interaction modules, we designed a modular protein architecture in which phase-separating peptides were linked to CC peptides (Figure 3.1.A). The idea was to keep the two main functions, phase separation and molecular recognition, separate but cooperative. Phase separation was driven by phase-separating peptides, while coiled-coil pairs served as molecular tags to guide selective interactions between specific partners.

The phase separating peptides used in this study were selected to represent different internal environments in condensates caused by different intermolecular interactions. We were particularly interested in testing whether the chirality of amino acids and peptide backbone flexibility would change the internal environments of condensates. To do this, we designed a set of Arg-rich peptides (Table 3.1). These peptides contain four pairs of Ser-Arg repeats with three sets of two types of motifs in between: flexible motifs (GGSSGG, pep 1) and rigid motifs (APAPAP, pep 2).[39] This design allows us to assess how backbone flexibility affects the phase separation behavior and internal environment of the resulting condensates. The second set of peptides was designed to be Glu-rich with net negative charge (pep 3), similar to pep 1(Tyr) from Chapter 1. This peptide would be prone to interact with the Arg-rich peptides through charge complementarity.

Table 3.1. Sequence of peptides designed to promote phase separation and coiled-coil peptides.

Peptide	Sequence:	Note
1	**SRGGSSGGSRGGSSGGSRGGSSGGSRYGGC**	SR repeat with flexible motifs
2	**SRAPAPAPSRAPAPAPSRAPAPAPSRYGGC**	SR repeat with rigid motifs
3	**GGGYEEEYGGGYGGGYEEEYGGGYGGGYC**	Glu-rich with net negative charge
P1	**GGSPEDEIQALEEENAQLEQENAALEEEIAQLEYG**	CC pair with P2
P2	**GGSPEDKIAQLKEKNAALKEKNQQLKEKIQFALKYG**	CC pair with P1

The aforementioned peptides were fused to the heterodimeric CC peptides (P1, P2) selected from work of Lebar et al., which assemble with high specificity and minimal cross-reactivity.[30] P1 and P2 can form stable dimers with each other, making them ideal building blocks for creating programmable interaction networks in complex environment. Each CC peptide consists of approximately 35 amino acids, while the phase-separating domains are about 30 amino acids in length. With the microwave-assisted method, it is hard to synthesize peptides longer than 60 residues. To address this problem, we used a two-part synthetic strategy. The CC peptide was modified with iodoacetic acid at the N-terminus, and the phase separating peptide was attached with cysteine at the C-terminus. After individual purification, the two peptides were combined under mildly basic conditions. The thiol group on cysteine selectively attacks the iodoacetyl group, forming a stable covalent linkage between two peptide fragments (Figure 3.2).[40] With this strategy, we synthesized the full-length peptides with high purity and yield (Table 3.2).

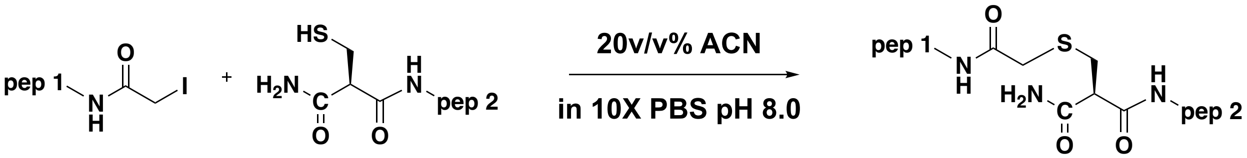


Figure 3.2. Synthesis of the two-part fused CC peptide and phase-separating peptide.

Table 3.2. Sequence of peptides designed to promote phase separation and coiled-coil peptides ("-" for covalent bond between iodoacetyl and thiol groups).

Peptide	Sequence:	Note
A	**SRGGSSGGSRGGSSGGSRGGSSGGSRYGGC-GGSPEDEIQALEEENAQLEQENAALEEEIAQLEYG**	P1 fused to pep 1
B	**SRAPAPAPSRAPAPAPSRAPAPAPSRYGGC-GGSPEDKIAQLKEKNAALKEKNQQLKEKIQFALKYG**	P1 fused to pep 2
C	**GGGYEEEYGGGYGGGYEEEYGGGYGGGYC-GGSPEDKIAQLKEKNAALKEKNQQLKEKIQFALKYG**	P2 fused to pep 3

3.3.2 Initial Assessment of this system

We evaluated whether orthogonal CC interactions could promote multivalent interactions in order to drive phase separation. We first tested the phase separation behavior of unfused phase-separating peptides. When 50 μM pep 1 (Arg-rich with flexible motif) was mixed with 50 μM pep 3 (Glu-rich with net negative charge), no phase separation was observed (Figure 3.3.A). The same outcome was observed with a pep 2 (Arg-rich with rigid motif) and pep 3 mixture (Figure 3.3.C). These results indicated that without the CC peptides, peptides 1-3 alone were not sufficient to drive phase separation formation under the given conditions. In contrast, when introduced the appropriate fusion peptides A-C were combined, phase separation was observed under microscope. When 50 μM pep A was mixed with 50 μM pep C, we observed formation of phase-separated coacervates via brightfield microscopy (Figure 3.3.B). Similarly, the combination of pep B with pep C led to visible condensate formation (Figure 3.3.D). This tells us that the dimerization of CC pairs can effectively bring the phase-separating peptides into proximity to promote phase separation.

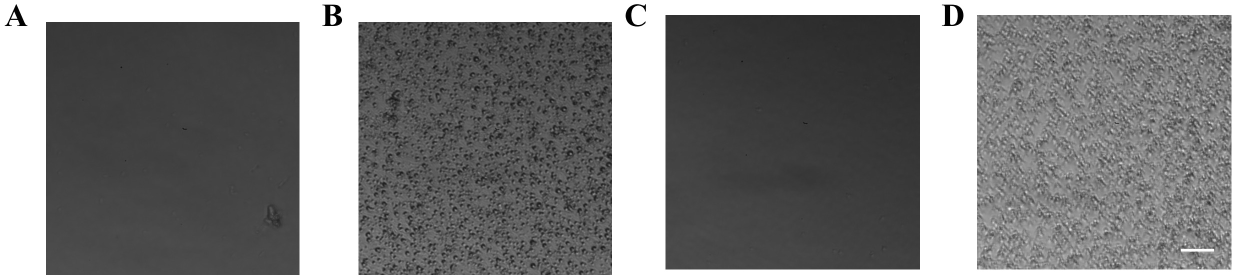


Figure 3.3. Effects of CC fusion on phase separation. **A.** 50 μM pep 1 with 50 μM pep 3. **B.** 50 μM pep A with 50 μM pep C. **C.** 50 μM pep 2 with 50 μM pep3. **D.** 50 μM pep B with 50 μM pep C. Buffer condition: 50mM Tris pH 7.5, scale bar 10 μm.

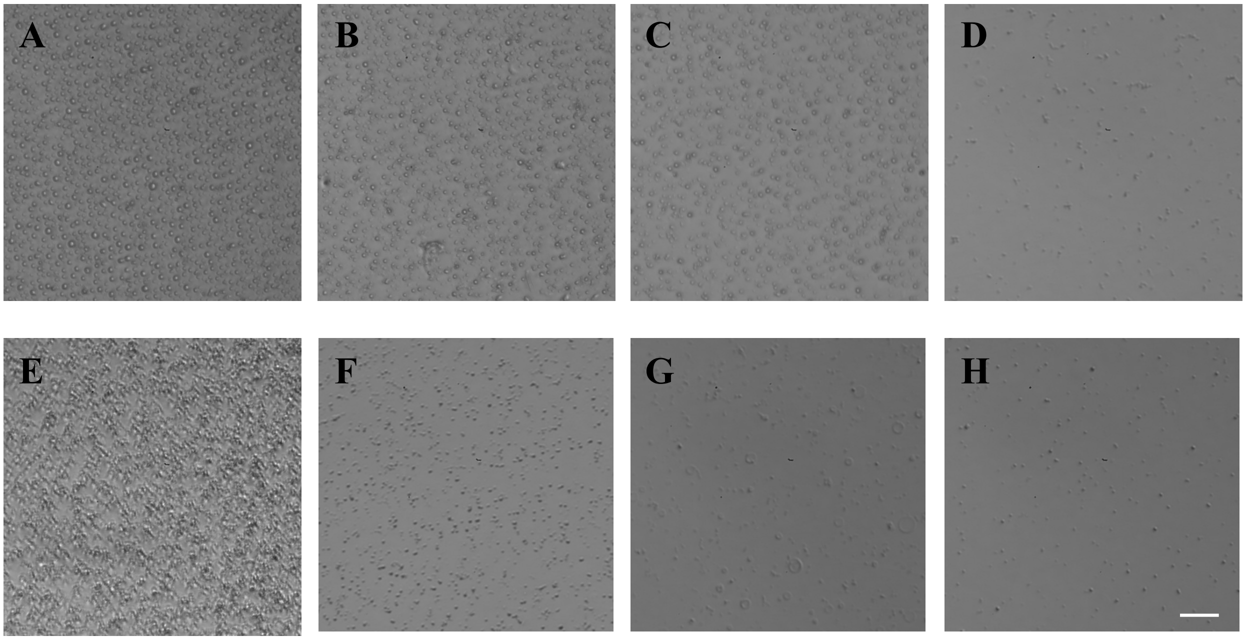


Figure 3.4. Comparison of phase separation behavior based on backbone flexibility. Equimolar pep A mixed with pep C at varying concentrations: **A**. 50 μM, **B**. 25 μM, **C**. 12.5 μM, **D**. 6.25 μM. Equimolar pep B mixed with pep C at varying concentrations: **E**. 50 μM, **F**. 25 μM, **G**. 12.5 μM, **H**. 6.25 μM. Buffer condition: 50mM Tris pH 7.5, scale bar 10 μm.

We next investigated whether the conformational flexibility of the motif in the phase-separating peptide could influence the onset and extent of LLPS. To probe this, we compared two versions of the Arg-rich fusion constructs: one with a flexible motif (GGSSGG, in pep A) and another with a rigid, proline-rich motif (APAPAP, in pep B). Both constructs were mixed with pep C at equimolar concentrations across a range of concentrations (50 µM, 25 µM, 12.5 µM, 6.25 µM). As shown in Figure 3.4 A-D, the pep A/pep C mixture with the flexible motif exhibited more robust phase separation, especially at 25 µM, and 12.5 µM. Condensates formed readily at 6.25 µM, and at 50 µM, extensive condensate formation was observed. In contrast, the pep B/pep C mixture with the rigid motif displayed significantly less condensate formation (Figure 3.4. E-H), particularly at lower concentrations. These results support the hypothesis that motif flexibility affects the phase separation onset and extent. Based on the preliminary results, we hypothesized that a flexible backbone may facilitate the alignment of interacting motifs within condensates, while rigid motifs may restrict this mobility. This leads to the differences between coacervate formation despite the presence of the same interaction motifs (Ser-Arg).

3.4 Conclusion and Future Directions

In this work, we developed a modular design strategy to build synthetic coacervate systems using orthogonal coiled-coil (CC) peptide interactions fused to low-complexity, phase-separating domains. The dimerization of CC pairs can effectively bring the phase-separating peptides into proximity to promote phase separation,[25,28–30,32] while the sequence composition and backbone flexibility of the phase-separating peptides can facilitate the formation of condensates.[7,41]

Our findings demonstrate that coiled-coil dimerization can effectively enhance multivalent interactions to bring the phase-separating peptides into proximity. Importantly, we observed that peptide backbone flexibility strongly affects the extent and robustness of LLPS. Constructs with flexible motifs (GGSSGG) exhibited more extensive condensate formation across a certain concentration range than those with rigid proline-rich motifs (APAPAP), suggesting that increased conformational freedom facilitates nucleation and internal packing within condensates.

This modular coiled-coil system allowed us to study how specific molecular features, such as the flexibility or rigidity of linker sequences, and the types of charged or aromatic residues, can influence how peptides come together to form phase-separated condensates. By keeping the recognition (via coiled-coil pairing) and phase separation (via phase-separating peptides) functions separate but connected, we were able to promote the multivalent interactions and decrease the energy barrier to form phase-separated condensates.

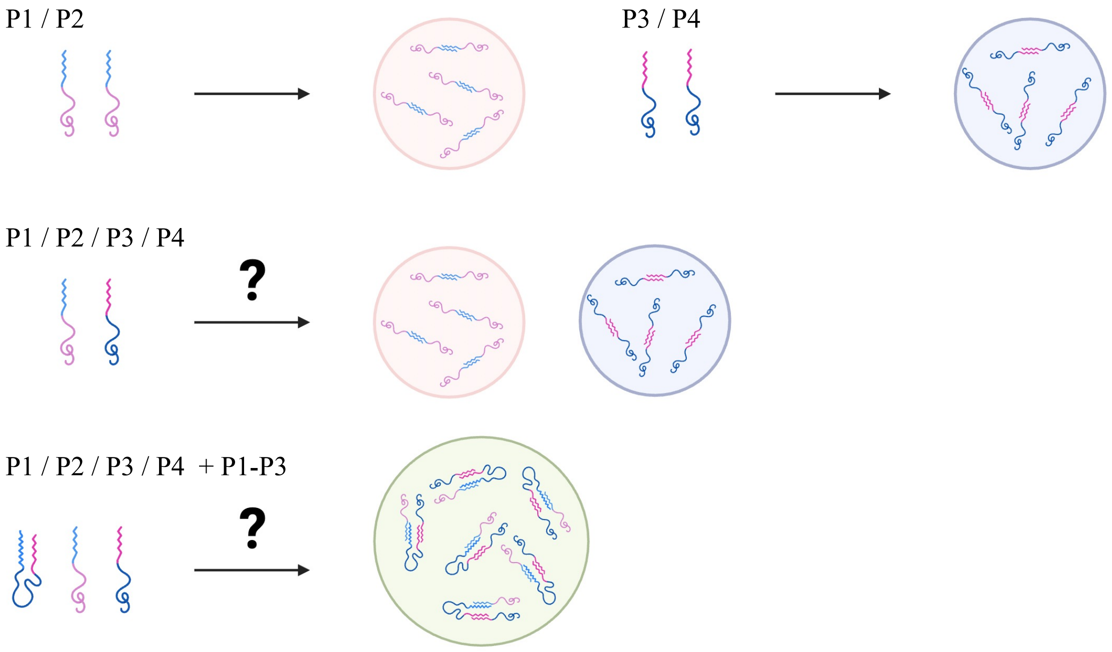


Figure 3.5. Scheme for phase-separating peptides fused with orthogonal CC peptides to form distinct coexisting condensates. Addition of fused bifunctional CC peptides may lead to the fusion of distinct condensates.

In future, we plan to expand this system to create multiple, coexisting condensate compartments. These compartments could be used to spatially and temporally organize biochemical reactions *in vitro*, similar to how organelles work in living cells. We will first finish testing how chirality (L-peptide vs. D-peptide) would affect the phase separation behavior using pep 4 and 5, and CC pair P3 and P4. Following that, we will explore whether chirality, or backbone flexibility affect the ability of condensates to recruit different molecules. If differences in recruitment are observed, we could use this property to selectively incorporate enzymes, catalytic scaffolds, or small molecules into distinct condensates. By introducing a fused CC peptide that links two orthogonal pairs (such as P1-P3), we could induce the fusion of two previously separated condensates. This controlled fusion could allow certain reactions to be activated only at specific times and locations. This would allow us to build synthetic systems that organize biochemical reactions in a controlled and programmable way. In the long term, this approach could help us create synthetic organelles, artificial cells, or new tools for managing complex reactions in labs or biotechnological applications.

3.5 Experimental Methods

3.5.1 Material Source

Fmoc-protected amino acids, ethyl (hydroxyimino) cyanoacetate (Oxyma), N,N'-diisopropylcarbodiimide (DIC),and hexafluorophosphate azabenzotriazole tetramethyl uronium

(HATU) were purchased from Chem-Impex international. Fluorescein isothiocyanate was purchased from Thermo Scientific. Diisopropylethylamine (DIEA), N,N-dimethylformamide (DMF), dichloromethane (DCM), piperidine, trifluoroacetic acid (TFA), triisopropylsilane (TIPS), thioanisole, acetonitrile, sodium chloride (NaCl), Tris base, and Dulbecco's PBS were purchased from Millipore Sigma. SPPS reaction vessel syringes and caps were purchased from Torviq. Rink Amide Resin (Low Loading) was purchased from CEM.

3.5.2 Peptide synthesis, purification, and characterization

All peptides were prepared via CEM MARS6 microwave-assisted solid phase peptide synthesis. 50 μmole Rink Amide ProTide resin (LL) was added to a Torviq solid-phase peptide synthesis vessel with a stir bar. Resin was swelled in 1:1 v/v DMF:DCM for 10 minutes before the synthesis. Fmoc amino acids (4 eq., 0.1 M) were activated with ethyl cyano(hydroxyimino)acetate (Oxyma, 8 eq.) and N,N'-diisopropylcarbodiimide (DIC, 4 eq.), and this solution was added to the reaction vessel. Regular coupling cycles were done at 70°C for 4 minutes. Then the resin was washed 5 times with DMF. The Fmoc protecting group was removed by reacting with 20% v/v piperidine in DMF at 80°C for 2 minutes. For peptides bearing an iodoacetic acid, iodoacetic acid was added to the N-terminus right below peptide cleavage.

After all residues had been added, the resin was mixed with the cleavage cocktail (8 mL per 50 μmole), which is 94 % (v/v) TFA, 2.5 % (v/v) H_2O, 2.5 % (v/v) thioanisole, and 1% (v/v) TIPS, at room temperature for 3 hours. The solution containing the crude cleaved peptide was drained from the reaction vessel into a 50-mL falcon tube. Excess TFA was removed under a stream of nitrogen, and about 35 mL of cold diethyl ether was added to precipitate the crude peptide. The crude peptide pellet was isolated by filtration, washed with cold diethyl ether twice, and then dried under nitrogen.

The crude product was dissolved in 600 μL 1:1 H_2O:acetonitrile and filtered through a 0.22 μm Nylon filter and purified on an Agilent preparative HPLC system with a C18 CSH Prep column (5 μm, 19x250 mm, solvent A = H_2O + 0.1 % TFA, solvent B = acetonitrile + 0.1 % TFA, flow rate = 13 mL/min). Peptide masses were determined using MALDI-TOF-MS. HPLC fractions containing pure peptide were combined, and this solution was lyophilized (freeze-dried) for long-term storage.

3.5.3 Synthesis of fused CC and phase-separating peptide

After purification, the lyophilized powders for phase-separating peptide (with a C-terminal cysteine) and the coiled-coil peptide (with an N-terminal iodoacetyl group) were weighed. A 1:1 molar ratio of phase-separating peptide and the coiled-coil peptide was dissolved in 10X PBS buffer with 20% ACN to enhance solubility. To promote the thiol–alkylation reaction, the pH of the solution was carefully adjusted to mildly basic conditions (pH ~8.0) using 1 M sodium hydroxide (NaOH). The reaction mixture was incubated at room temperature overnight. The mixture was then filtered through a 0.22 µm Nylon filter and purified on an Agilent preparative HPLC system with a C18 CSH Prep column (5 µm, 19x250 mm, solvent A = H_2O + 0.1 % TFA, solvent B = acetonitrile + 0.1 % TFA, flow rate = 13 mL/min).

3.6 MALDI-TOF-MS Spectra and UPLC Data for Synthetic Peptides

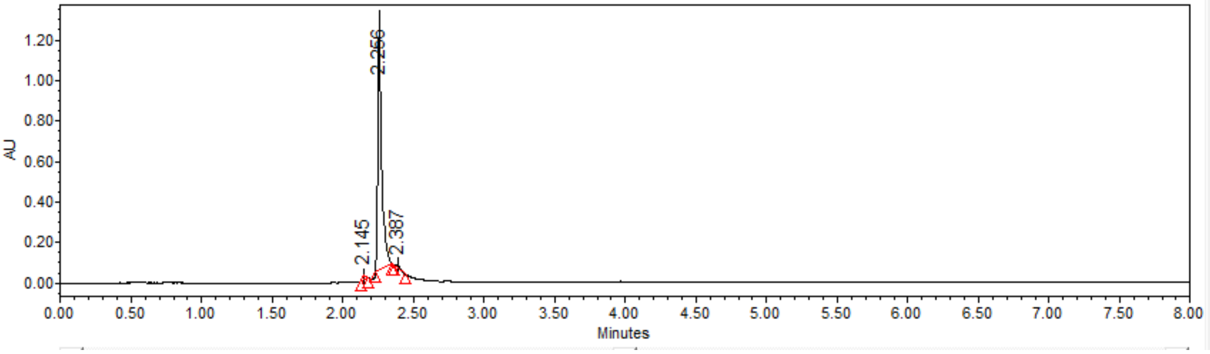
Iodo-P1: **`Iodoacetyl-GGSPEDEIQALEEENAQLEQENAALEEEIAQLEYG-CONH`$_2$**

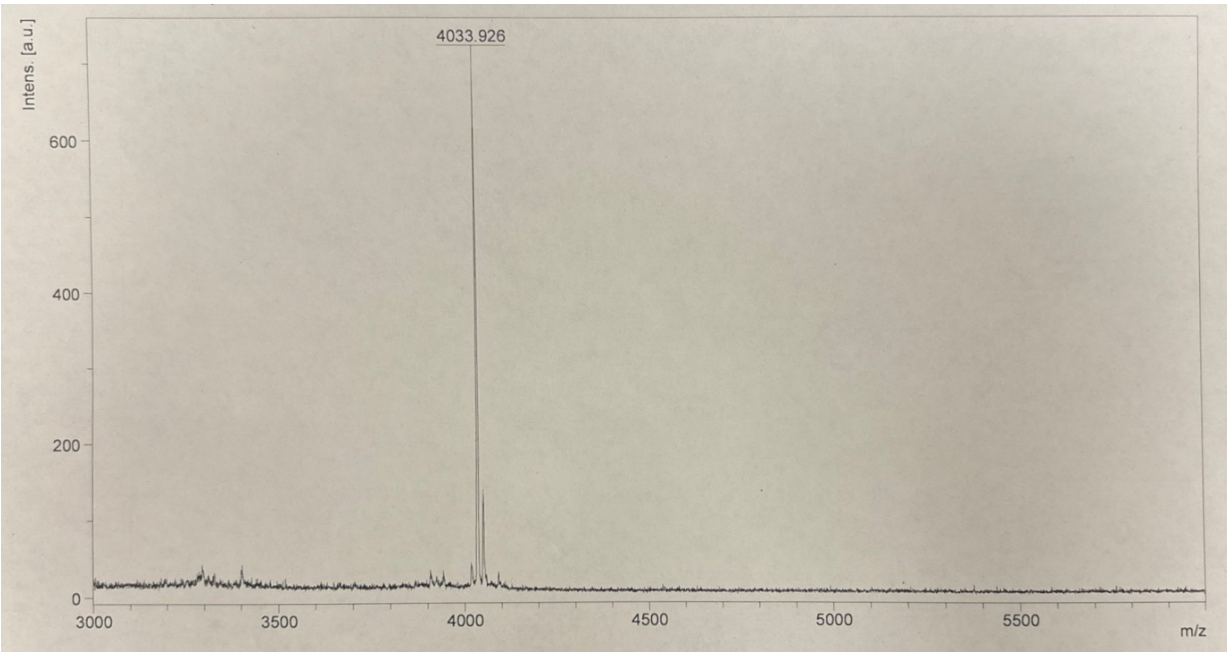
<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+Na]^+$ = 4034.65

observed monoisotopic $[M+Na]^+$ = 4033.926

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column. Wavelength at 220nm.

Purity > 99.0 %





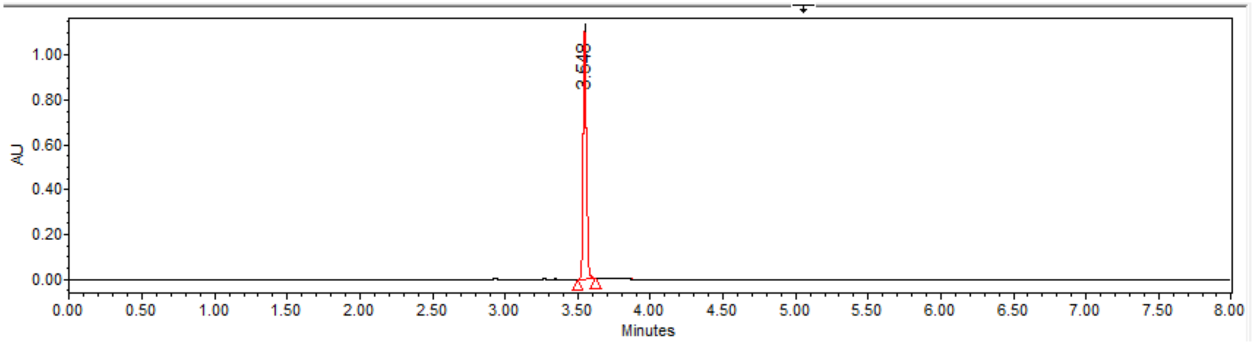
Iodo-P2: **Iodo-GGSPEDKIAQLKEKNAALKEKNQQLKEKIQFALKYG-CONH₂**

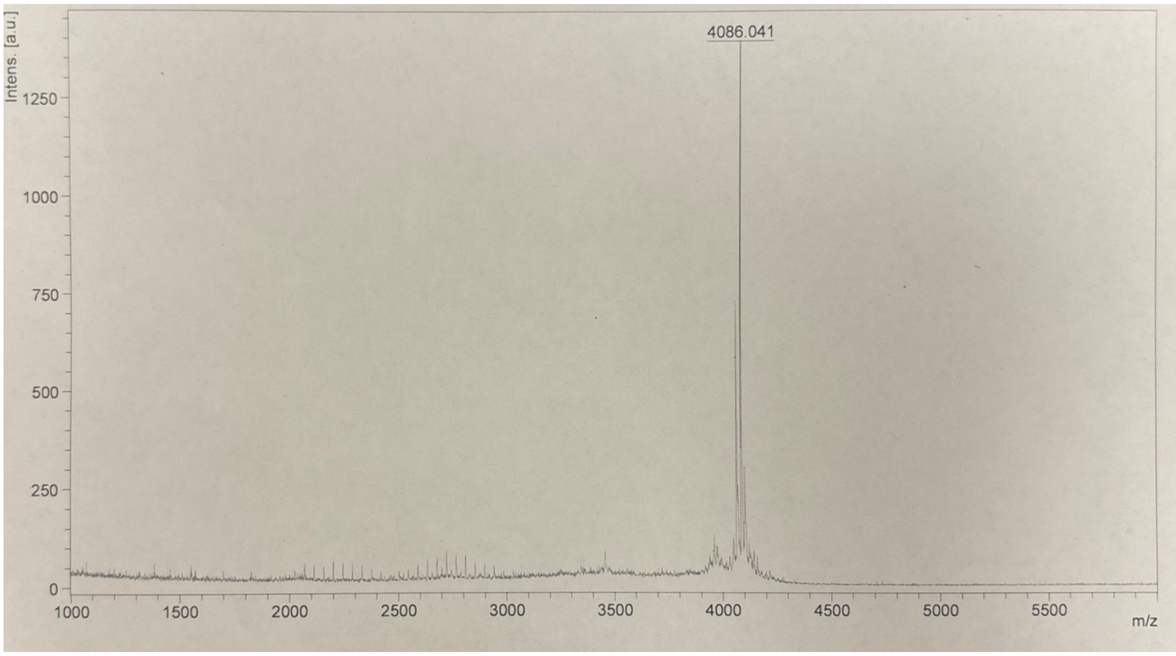
MALDI-TOF-MS: calculated monoisotopic $[M+Na]^+$ = 4085.07

observed monoisotopic $[M+Na]^+$ = 4086.041

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column. Wavelength at 220nm.

Purity = 97.38 %





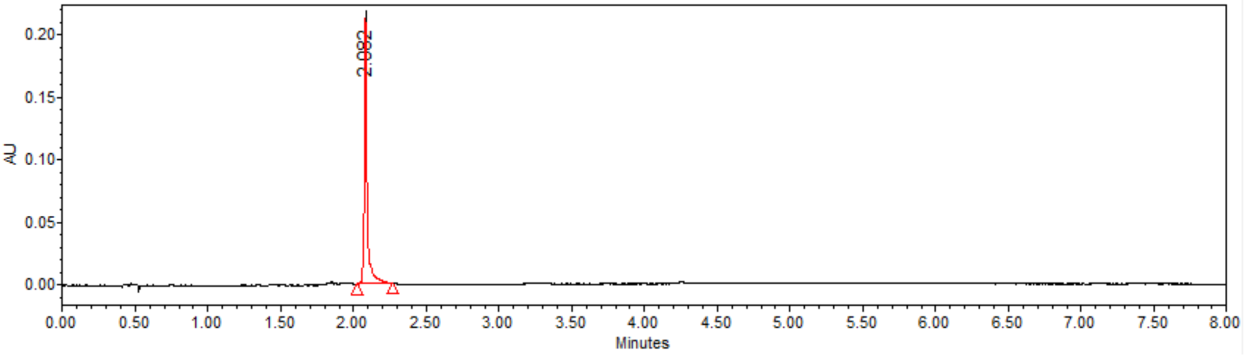
SRGSC: **H_2N-SRGGSSGGSRGGSSGGSRGGSSGGSRYGGC-$CONH_2$**

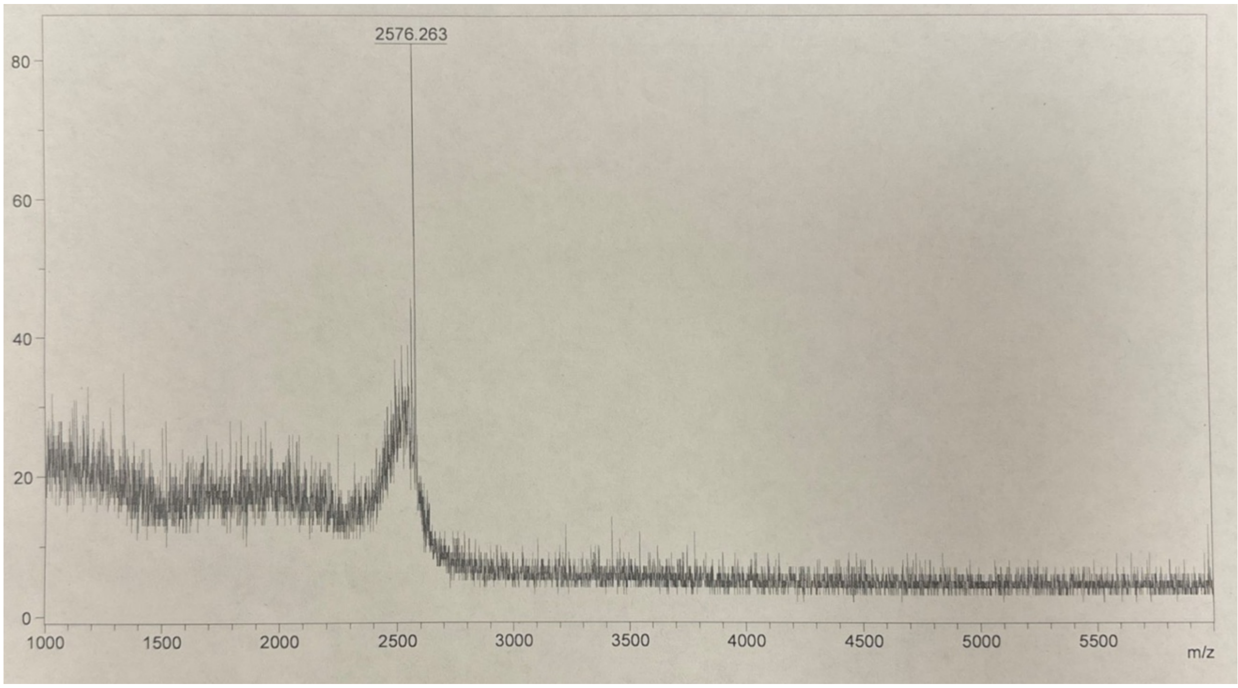
MALDI-TOF-MS: calculated monoisotopic $[M+H]^+$ = 2576.113

observed monoisotopic $[M+H]^+$ = 2576.263

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column. Wavelength at 220nm.

Purity > 99.0 %





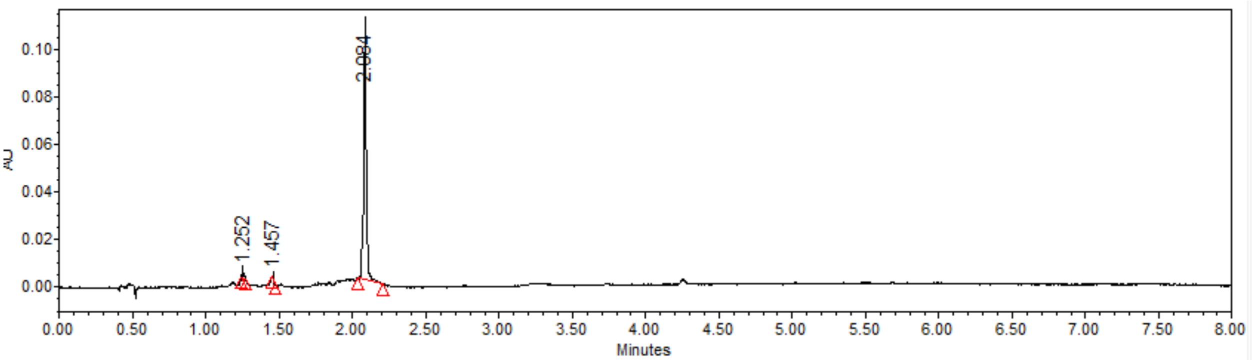
SRAPC: **$\mathrm{H_2N}$-SRAPAPAPSRAPAPAPSRAPAPAPSRYGGC-$\mathrm{CONH_2}$**

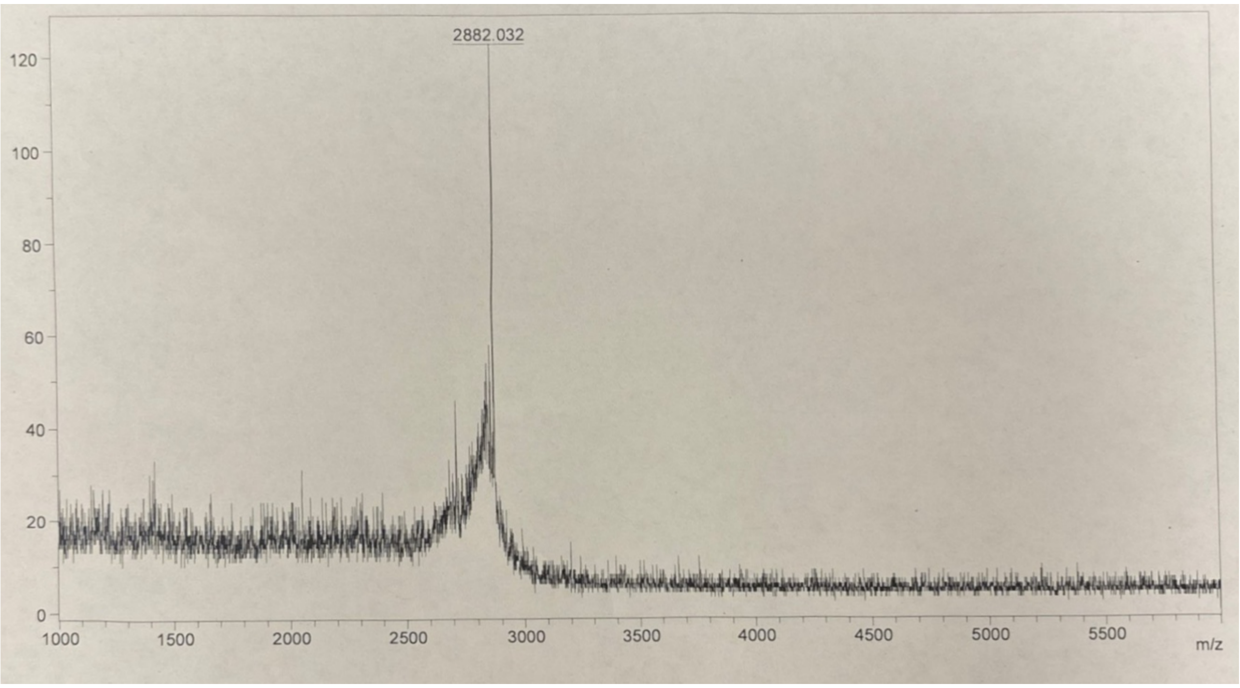
<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 2882.473

observed monoisotopic $[M+H]^+$ = 2882.032

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column. Wavelength at 220nm.

Purity > 99.0 %





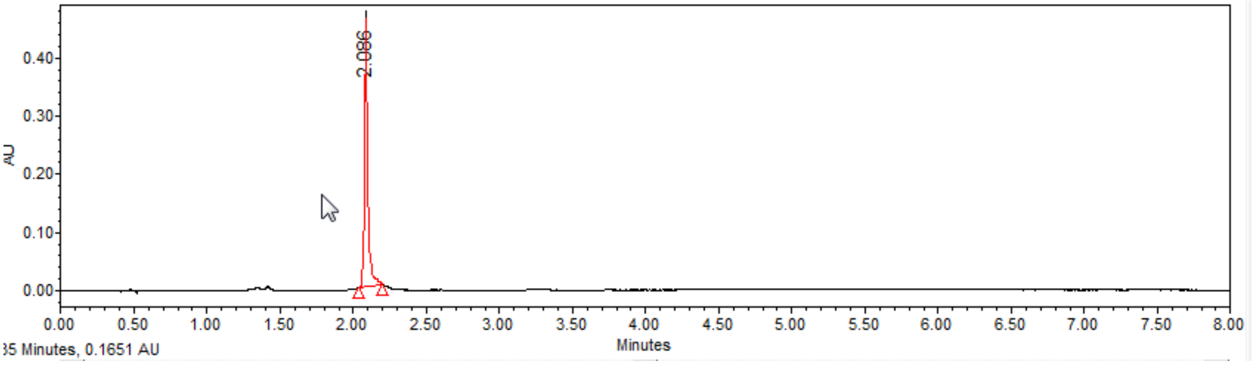
EEEYC: **H_2N-GGGYEEEYGGGYGGGYEEEYGGGYGGGYC-$CONH_2$**

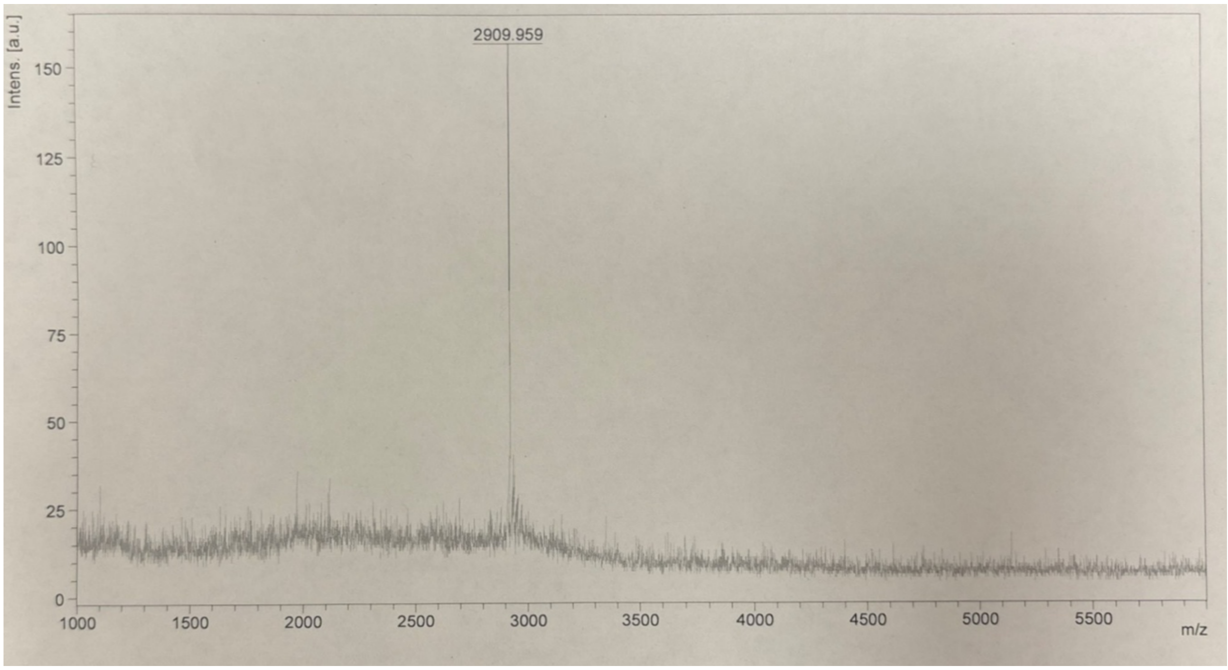
MALDI-TOF-MS: calculated monoisotopic $[M+NH_4]^+ = 2909.046$

observed monoisotopic $[M+NH_4]^+ = 2909.959$

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130 Å 1.7 µm, 2.1 x 150 mm) column. Wavelength at 220nm.

Purity > 99.0 %





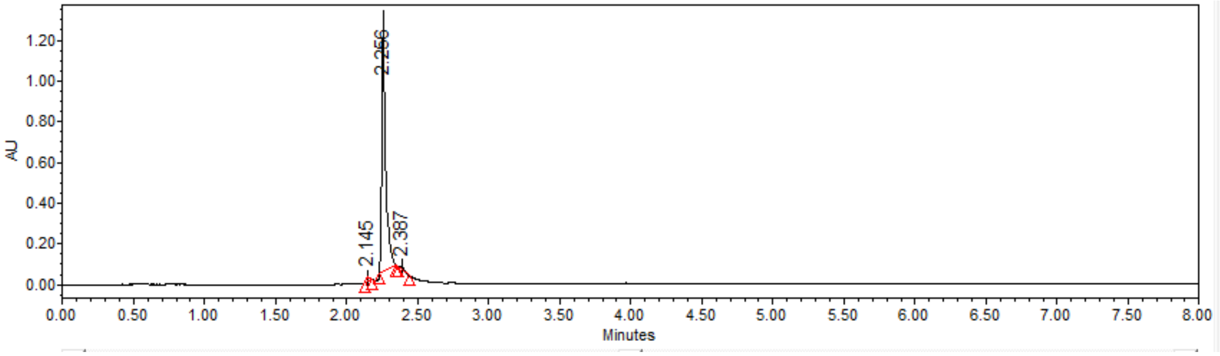
P1-SRGSC: **H_2N-SRGGSSGGSRGGSSGGSRGGSSGGSRYGGC-GGSPEDEIQALEEENAQLEQENAALEEEIAQLEYG-$CONH_2$ ("-" for covalent bond between iodoacetyl and thiol groups)**

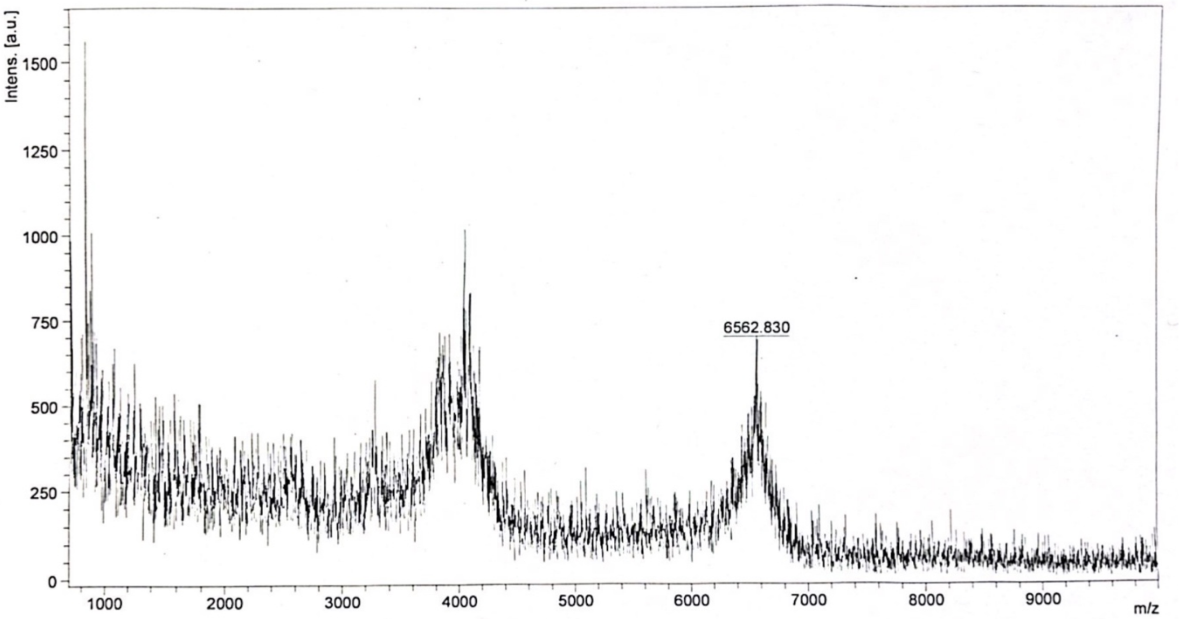
ESI-MS: calculated monoisotopic $[M+2K+2H]^+$ = 6562.863

observed monoisotopic $[M+2K+2H]^+$ = 6562.830

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column. Wavelength at 220nm.

Purity > 99.0 %





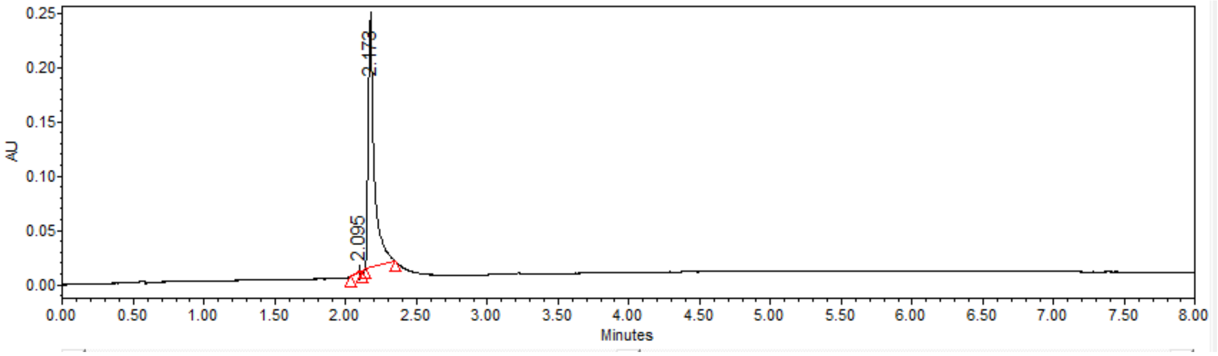
P1-SRGSC: **H_2N-SRAPAPAPSRAPAPAPSRAPAPAPSRYGGC-GGSPEDKIAQLKEKNAALKEKNQQLKEKIQFALKYG-CONH$_2$ ("-" for covalent bond between iodoacetyl and thiol groups)**

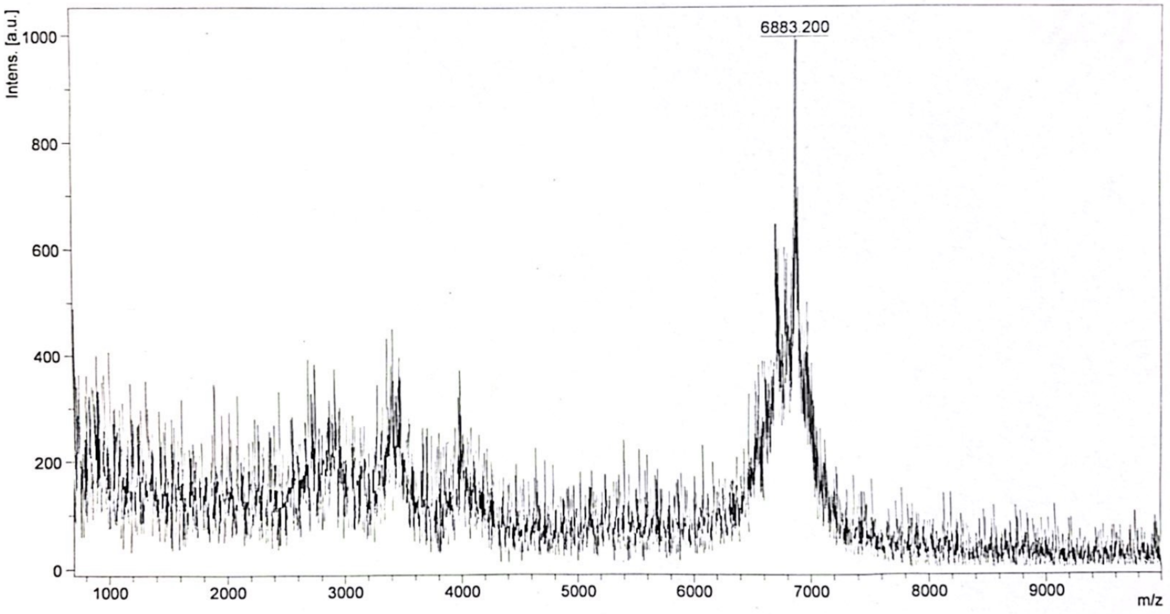
ESI-MS: calculated monoisotopic $[M+4N+2H]^+$ = 6883.223

observed monoisotopic $[M+4Na+2H]^+$ = 6883.200

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column. Wavelength at 220nm.

Purity > 99.0 %





3.7 References

(1) Atilla-Gokcumen, G. E.; Muro, E.; Relat-Goberna, J.; Sasse, S.; Bedigian, A.; Coughlin, M. L.; Garcia-Manyes, S.; Eggert, U. S. Dividing Cells Regulate Their Lipid Composition and Localization. *Cell* 2014, *156* (3), 428–439. https://doi.org/10.1016/j.cell.2013.12.015.

(2) Cascianelli, G.; Villani, M.; Tosti, M.; Marini, F.; Bartoccini, E.; Magni, M. V.; Albi, E. Lipid Microdomains in Cell Nucleus. *Mol Biol Cell* 2008, *19*, 5289–5295. https://doi.org/10.1091/mbc.E08.

(3) Schenkel, L. C.; Bakovic, M. Formation and Regulation of Mitochondrial Membranes. *Int J Cell Biol* 2014, *10`4*, 709828. https://doi.org/10.1155/2014/709828.

(4) Yang, Y.; Lee, M.; Fairn, G. D. Phospholipid Subcellular Localization and Dynamics. *Journal of Biological Chemistry* 2018, *293* (17), 6230–6240. https://doi.org/10.1074/jbc.R117.000582.

(5) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular Condensates: Organizers of Cellular Biochemistry. *Nat Rev Mol Cell Biol* 2017, *18* (5), 285–298. https://doi.org/10.1038/nrm.2017.7.

(6) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; Tompa, P.; Fuxreiter, M. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 2018, *28* (6), 420–435. https://doi.org/10.1016/j.tcb.2018.02.004.

(7) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu Rev Phys Chem* 2020, *71*, 53–75. https://doi.org/10.1146/annurev-physchem-071819-113553.

(8) Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu Rev Cell Dev Biol* 2014, *30*, 39–58. https://doi.org/10.1146/annurev-cellbio-100913-013325.

(9) Wang, B.; Zhang, L.; Dai, T.; Qin, Z.; Lu, H.; Zhang, L.; Zhou, F. Liquid–Liquid Phase Separation in Human Health and Diseases. *Signal Transduct Target Ther* 2021, *6* (1). https://doi.org/10.1038/s41392-021-00678-1.

(10) Gomes, E.; Shorter, J. The Molecular Language of Membraneless Organelles. *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc. May 3, 2019, pp 7115–7127. https://doi.org/10.1074/jbc.TM118.001192.

(11) Marnik, E. A.; Updike, D. L. Membraneless Organelles: P Granules in Caenorhabditis Elegans. *Traffic* 2019, *20* (6), 373–379. https://doi.org/10.1111/tra.12644.

(12) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol Cell* 2015, *57* (5), 936–947. https://doi.org/10.1016/j.molcel.2015.01.013.

(13) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A Framework for Understanding the Functions of Biomolecular Condensates across Scales. *Nat Rev Mol Cell Biol* 2021, *22* (3), 215–235. https://doi.org/10.1038/s41580-020-00303-z.

(14) Lu, T.; Spruijt, E. Multiphase Complex Coacervate Droplets. *J Am Chem Soc* 2020, *142* (6), 2905–2914. https://doi.org/10.1021/jacs.9b11468.

(15) Wang, J.; Chen, X.; Chen, E. Q.; Yang, S. Interfacial Tensions of Polyelectrolyte Multiphase Coacervation. *Macromolecules* 2024. https://doi.org/10.1021/acs.macromol.4c02017.

(16) Karoui, H.; Seck, M. J.; Martin, N. Self-Programmed Enzyme Phase Separation and Multiphase Coacervate Droplet Organization. *Chem Sci* 2021, *12* (8), 2794–2802. https://doi.org/10.1039/d0sc06418a.

(17) Rana, U.; Xu, K.; Narayanan, A.; Walls, M. T.; Panagiotopoulos, A. Z.; Avalos, J. L.; Brangwynne, C. P. Asymmetric Oligomerization State and Sequence Patterning Can Tune Multiphase Condensate Miscibility. *Nat Chem* 2024, *16* (7), 1073–1082. https://doi.org/10.1038/s41557-024-01456-6.

(18) Rai, S. K.; Khanna, R.; Avni, A.; Mukhopadhyay, S. Heterotypic Electrostatic Interactions Control Complex Phase Separation of Tau and Prion into Multiphasic Condensates and Co-Aggregates. *Proceedings of the National Academy of Sciences* 2023, *120* (2). https://doi.org/10.1073/pnas.2216338120.

(19) Wei, M.; Wang, X.; Qiao, Y. Multiphase Coacervates: Mimicking Complex Cellular Structures through Liquid-Liquid Phase Separation. *Chemical Communications* 2024, *60* (90), 13169–13178. https://doi.org/10.1039/d4cc04533e.

(20) Lafontaine, D. L. J.; Riback, J. A.; Bascetin, R.; Brangwynne, C. P. The Nucleolus as a Multiphase Liquid Condensate. *Nat Rev Mol Cell Biol* 2021, *22* (3), 165–182. https://doi.org/10.1038/s41580-020-0272-6.

(21) Lee, D. S. W.; Strom, A. R.; Brangwynne, C. P. The Mechanobiology of Nuclear Phase Separation. *APL Bioeng* 2022, *6* (2), 021503. https://doi.org/10.1063/5.0083286.

(22) Feric, M.; Vaidya, N.; Harmon, T. S.; Mitrea, D. M.; Zhu, L.; Richardson, T. M.; Kriwacki, R. W.; Pappu, R. V.; Brangwynne, C. P. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* 2016, *165* (7), 1686–1697. https://doi.org/10.1016/j.cell.2016.04.047.

(23) Kaur, T.; Raju, M.; Alshareedah, I.; Davis, R. B.; Potoyan, D. A.; Banerjee, P. R. Sequence-Encoded and Composition-Dependent Protein-RNA Interactions Control Multiphasic Condensate Morphologies. *Nat Commun* 2021, *12* (1), 872. https://doi.org/10.1038/s41467-021-21089-4.

(24) Bremer, A.; Farag, M.; Borcherds, W. M.; Peran, I.; Martin, E. W.; Pappu, R. V.; Mittag, T. Deciphering How Naturally Occurring Sequence Features Impact the Phase Behaviours of Disordered Prion-like Domains. *Nat Chem* 2022, *14* (2), 196–207. https://doi.org/10.1038/s41557-021-00840-w.

(25) Steinmetz, M. O.; Jelesarov, I.; Matousek, W. M.; Honnappa, S.; Jahnke, W.; Missimer, J. H.; Frank, S.; Alexandrescu, A. T.; Kammerer, R. A. Molecular Basis of Coiled-Coil Formation. *PNAS* 2007, *104* (17), 7062–7067.

(26) Jorgensen, M. D.; Chmielewski, J. Recent Advances in Coiled-Coil Peptide Materials and Their Biomedical Applications. *Chemical Communications* 2022, *58* (83), 11625–11636. https://doi.org/10.1039/d2cc04434j.

(27) Woolfson, D. N. Understanding a Protein Fold: The Physics, Chemistry, and Biology of α-Helical Coiled Coils. *Journal of Biological Chemistry* 2023, *299* (4), 104579. https://doi.org/10.1016/j.jbc.2023.104579.

(28) Plaper, T.; Rihtar, E.; Železnik Ramuta, T.; Forstnerič, V.; Jazbec, V.; Ivanovski, F.; Benčina, M.; Jerala, R. The Art of Designed Coiled-Coils for the Regulation of Mammalian Cells. *Cell Chem Biol* 2024, *31* (8), 1460–1472. https://doi.org/10.1016/j.chembiol.2024.06.001.

(29) Ramšak, M.; Ramirez, D. A.; Hough, L. E.; Shirts, M. R.; Vidmar, S.; Eleršič Filipič, K.; Anderluh, G.; Jerala, R. Programmable de Novo Designed Coiled Coil-Mediated Phase Separation in Mammalian Cells. *Nat Commun* 2023, *14* (1). https://doi.org/10.1038/s41467-023-43742-w.

(30) Lebar, T.; Lainšček, D.; Merljak, E.; Aupič, J.; Jerala, R. A Tunable Orthogonal Coiled-Coil Interaction Toolbox for Engineering Mammalian Cells. *Nat Chem Biol* 2020, *16* (5), 513–519. https://doi.org/10.1038/s41589-019-0443-y.

(31) Park, W. M.; Bedewy, M.; Berggren, K. K.; Keating, A. E. Modular Assembly of a Protein Nanotriangle Using Orthogonally Interacting Coiled Coils. *Sci Rep* 2017, *7* (1), 10577. https://doi.org/10.1038/s41598-017-10918-6.

(32) Aronsson, C.; Dånmark, S.; Zhou, F.; Öberg, P.; Enander, K.; Su, H.; Aili, D. Self-Sorting Heterodimeric Coiled Coil Peptides with Defined and Tuneable Self-Assembly Properties. *Sci Rep* 2015, *5*. https://doi.org/10.1038/srep14063.

(33) Wake, N.; Weng, S.-L.; Zheng, T.; Wang, S.-H.; Kirilenko, V.; Mittal, J.; Fawzi, N. L. Expanding the Molecular Grammar of Polar Residues and Arginine in FUS Phase Separation. *Nat Chem Biol* 2025. https://doi.org/10.1038/s41589-024-01828-6.

(34) Wang, J.; Choi, J. M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 2018, *174* (3), 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006.

(35) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular Interactions Underlying Liquid−liquid Phase Separation of the FUS Low-Complexity Domain. *Nat Struct Mol Biol* 2019, *26* (7), 637–648. https://doi.org/10.1038/s41594-019-0250-x.

(36) Qamar, S.; Wang, G. Z.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Ströhl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.; Tartaglia, G. G.; Farrer, L. A.; Kaminski Schierle, G. S.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* 2018, *173* (3), 720-734.e15. https://doi.org/10.1016/j.cell.2018.03.056.

(37) Murthy, A. C.; Tang, W. S.; Jovic, N.; Janke, A. M.; Seo, D. H.; Perdikari, T. M.; Mittal, J.; Fawzi, N. L. Molecular Interactions Contributing to FUS SYGQ LC-RGG Phase Separation and Co-Partitioning with RNA Polymerase II Heptads. *Nat Struct Mol Biol* 2021, *28* (11), 923–935. https://doi.org/10.1038/s41594-021-00677-4.

(38) Cakmak, F. P.; Choi, S.; Meyer, M. C. O.; Bevilacqua, P. C.; Keating, C. D. Prebiotically-Relevant Low Polyion Multivalency Can Improve Functionality of Membraneless

Compartments. *Nat Commun* 2020, *11* (1), 1–11. https://doi.org/10.1038/s41467-020-19775-w.

(39) Chen, X.; Zaro, J. L.; Shen, W. C. Fusion Protein Linkers: Property, Design and Functionality. *Adv Drug Deliv Rev* 2013, *65* (10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039.

(40) Boja, E. S.; Fales, H. M. Overalkylation of a Protein Digest with Iodoacetamide. *Anal Chem* 2001, *73* (15), 3576–3582. https://doi.org/10.1021/ac0103423.

(41) Schuster, B. S.; Dignon, G. L.; Tang, W. S.; Kelley, F. M.; Ranganath, A. K.; Jahnke, C. N.; Simpkins, A. G.; Regy, R. M.; Hammer, D. A.; Good, M. C.; Mittal, J. Identifying Sequence Perturbations to an Intrinsically Disordered Protein That Determine Its Phase-Separation Behavior. *Proc Natl Acad Sci U S A* 2020, *117* (21). https://doi.org/10.1073/pnas.2000223117.

Chapter 4

Comparative Analysis of Peptide Structure Prediction Using pepFOLD4, ESMFold, and AlphaFold3: Limitations in Predicting Reverse-Ordered Sequences

Experiments were designed by Ruiwen Xu and Samuel H. Gellman.

Experiments were performed by Ruiwen Xu and Jiawen Huang.

Graphics were prepared using Microsoft PowerPoint, GraphPad Prism, and Pymol.

4.1 Abstract

Advancements in computational protein structure prediction, particularly tools like pepFOLD4, ESMFold, and AlphaFold3, have significantly enhanced our ability to model peptide and protein structures. However, the reliability of these tools for accommodating sequence perturbations remains underexplored. In this study, we assessed the structural predictions of several miniproteins and their reverse-ordered variants using pepFOLD4, ESMFold, and AlphaFold3. While the forward sequence predictions closely matched the experimental crystal structure, all three tools failed to accurately model the reversed sequence. The reverse sequence predictions consistently favored α-helical conformations, conflicting with experimental data from circular dichroism (CD) spectroscopy, which indicated a lack of structured secondary elements. These findings suggest a systematic bias in current prediction algorithms towards over-stabilized secondary structures when dealing with atypical sequences and highlight the importance of integrating experimental validation in predictive workflows.

4.2 Introduction

The prediction of protein structure from its amino acid sequence has long been a central challenge in structural biology. Recent breakthroughs in protein structure prediction have dramatically changed the landscape of structural biology. Deep learning-based tools such as AlphaFold3[1,2], ESMFold[3], were trained on the vast datasets of experimentally determined protein structures. On contrary, fragment-based methods like pepFOLD4[4–6] predicts the structure from scratch based on the sequence information, without direct use of PDB structures[7]. These tools have not only accelerated the study of protein function and interaction but also opened new possibilities in synthetic biology, drug design, and peptide engineering.[8]

However, despite their success on naturally occurring sequences, there remain significant limitations in the generalization of these tools, especially non-natural or synthetically altered amino acid sequences.[9–12] Most predictive algorithms are trained or optimized using datasets composed of naturally occurring proteins and domains that conform to canonical folding principles. Whether these algorithms can accurately predict structures of synthetic peptides—especially when their sequences are modified, reversed, or deviate from natural evolutionary constraints—has not been comprehensively tested.

Childs et al., performed a systematic investigation on AlphaFold3's performance on D-peptides. While the model can technically accept D-chirality as input, it failed to preserve the correct stereochemistry in over 50% of cases.[13] Furthermore, AlphaFold3 frequently predicted incorrect folds and protein-peptide interaction modes, despite assigning high internal confidence scores to these biologically implausible models. This unreliability severely limits its utility for modeling D-peptides, which are increasingly used in drug development for their protease resistance and bioavailability.[14–16] In a related study, Silva et al., performed a large-scale analysis on lasso peptides, a class of peptides characterized by a threaded structure resembling a slipknot using AlphaFold2.[17] It showed that the model frequently generated topologically implausible motifs, such as highly pierced lassos, with low conservation across families, which suggests many predicted structures are artifacts rather than biologically relevant structures.[18] Similarly, Hou et al., systematically evaluated AlphaFold-Multimer's performance in predicting protein-protein complex structures, and discovered approximately 1.72% predicted structures contain topological links, such as interchain links that form knots or entanglements, which should generally not exist in native protein complex structures.[19] Together, these studies demonstrate that the advanced prediction tools can overpredict structure and struggle when dealing with exotic folds, stereochemical inversions, or sequences outside the training distribution.

In addition to the previously discussed limitations, our study investigates a related yet distinct challenge—sequence reversal.[20,21] Sequence reversal is a unique synthetic perturbation in which the amino acid composition remains unchanged, but the residue order is inverted from N- to C-terminus. This manipulation disrupts the directional context of side-chain interactions, backbone hydrogen bonding, and folding nucleation sites. Several previous studies have demonstrated that sequence reversal often leads to misfolding or structural disorder, even when the native sequence adopts a stable fold. Lacroix et al. provided experimental evidence that the reversed sequences of the SH3 domain (β-sheet for forward sequence), as well as the B domain of Staphylococcal protein A (α-helix bundle for forward sequence) and the B1 domain of Streptococcal protein G (a mixed α/β fold), all resulted in unfolded proteins.[22] In contrast, Zhang et al. reported that proteins with more flexible hydrophobic cores—such as α-helix bundle and Protein G—can adopt native-like folds in their reversed forms, suggesting that foldability may depend on the internal packing plasticity.[23] However, they also noted that smaller proteins, which require precise core packing and residue orientation, often lose their structure upon reversal, though in some cases, foldability could be partially recovered through targeted modifications, such as inserting residues in β-hairpin turns.

To investigate this question, we evaluated three widely used prediction tools—pepFOLD4, ESMFold, and AlphaFold3—across a representative set of peptides with experimentally known structures. We selected three experimentally characterized peptides, each belonging to a distinct structural class: (1) a peptide forming an α-helix-rich tertiary fold, Villin headpiece subdomain

(VHP)[24,25] (2) an α-helical bundle, GA module[26] and (3) a peptide that adopts a β-sheet conformation, human Pin1 WW domain ("WW" below)[27,28]. We synthesized reverse-ordered versions of each peptide (i.e., the amino acid sequence was flipped from N- to C-terminus): VHP_rev, GA_rev, and WW_rev. While this manipulation preserves amino acid composition, it disrupts the sequence directionality and residue-residue context that governs folding. CD spectroscopy was used to experimentally assess the secondary structure of both the native and reversed sequences.[29,30]

We found that all three tools accurately predicted the tertiary structure of native (forward) sequences as revealed by crystallography or NMR. However, with reverse-ordered sequences, the tools consistently predicted high α-helical content. In contrast, CD spectra for the reverse sequences revealed largely disordered conformations. These findings suggest that current peptide structure prediction tools may misrepresent structural features when faced with atypical sequences, reinforcing the need for caution and experimental validation when interpreting such predictions.

4.3 Results and Discussion

4.3.1 Peptide Design

In this study, three peptides with known structures from crystallography or NMR were selected to represent distinct structural classes: VHP (a tertiary structure with three α-helices, PDB entry 1YRF, Figure 4.1A), GA module (α-helical bundle, PDB entry 1PRB, Figure 4.1B), and WW domain (β-sheet rich peptide, PDB entry 4GWT, Figure 4.1C). Each of these peptides served as a forward-sequence control. For each native (forward) sequence, a reverse-order variant was synthesized by inverting the amino acid sequence from N-terminus to C-terminus, without modifying the residue identities. As such, these reverse variants represent non-canonical sequences that challenge the assumptions of existing prediction algorithms.

Figure 4.1 Crystal structure for VHP (**A**, PDB: 1YRF), NMR structure for GA module (**B**, PDB: 1PRB) and crystal structure for WW domain (**C**, PDB: 4GWT).

4.3.2 Peptide Structure Analysis using Circular Dichroism

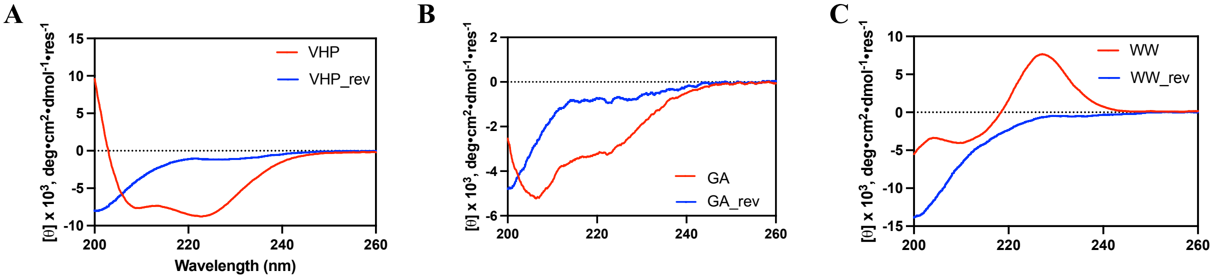


Figure 4.2 Far-UV CD for peptides and its corresponding reverse sequence. Measurements at 25°C. **A.** Measurements for VHP peptide and its reverse sequence peptide VHP_rev. Samples contained 50 µM polypeptide, 20 mM sodium acetate buffer, pH 5.0.[31] **B.** Measurements for GA module and its reverse sequence peptide GA_rev. Samples contained 50 µM polypeptide, 1X PBS buffer, pH 7.4.[26] **C.** Measurements for WW domain and its reverse sequence peptide WW_rev. Samples contained 50 µM polypeptide, 10 mM sodium phosphate buffer, pH 7.0.[27]

To test the computational predictions, all peptide variants were synthesized and analyzed using far UV CD spectroscopy. For the forward peptide sequences, CD spectra were consistent with the expected structural states and with previous reports on the CD characterization of these peptides. The α-helix-rich tertiary peptide VHP and the helical bundle GA module both exhibited double minima at approximately 208 and 222 nm. For the Pin1 WW domain, the CD signature features a highly characteristic maximum near 227 nm, which presumably reflects contributions from the two Trp side chains.

The CD spectra for the reverse sequences revealed a striking discrepancy with the predicted models. All three reversed peptides—regardless of the structure of the native peptides—produced spectra with a single minimum around 200 nm, a signature of disordered or unstructured polypeptides. No evidence of α-helical character was observed in the reversed versions of the α-helical peptides, VHP_rev and GA_rev, and the reversed β-sheet peptide, WW_rev, showed no features indicative of either helical or sheet-like structure. These results provide compelling evidence that, despite predictions of ordered structures by all three computational tools, the actual conformation of the reversed peptides in solution is predominantly unstructured.

4.3.3 Structure Prediction Tools

To assess how structural modeling tools handle both native and reverse-ordered sequences, we employed three structure prediction platforms: pepFOLD4, ESMFold, and AlphaFold3. Each tool uses a different approach. pepFOLD4 relies on a fragment-based approach and a coarse-grained force field to predict peptide structures from their amino acid sequences.[32,33] ESMFold, which stands for Evolutionary Scale Modeling Fold, uses a large protein language model (ESM-2) that has been trained on a massive dataset of known protein sequences to predict protein structures.[3] AlphaFold3 was trained on a massive dataset of experimentally determined protein structures from the Protein Data Bank (PDB) and uses a diffusion model to predict protein structures by iteratively denoising a random starting point.[2] All peptide sequences—both forward and reverse—were submitted to each tool, and the top-ranked model was analyzed in terms of predicted secondary structure and fold quality.

4.3.3.1 pepFOLD4 Prediction

For the forward peptide sequences, pepFOLD4 produced models consistent with known structures. The α-helix-rich tertiary peptide VHP (Figure 4.3A, RMSD = 0.943Å) and the helical bundle GA module (Figure 4.3C, RMSD = 1.822Å) were both accurately predicted, showing good agreement with expected helix packing and orientation. The β-sheet WW domain was also predicted correctly, as the tool captured the extended nature of the structure, strand pairing and hydrogen bonding patterns (Figure 4.3E, RMSD = 2.382Å).

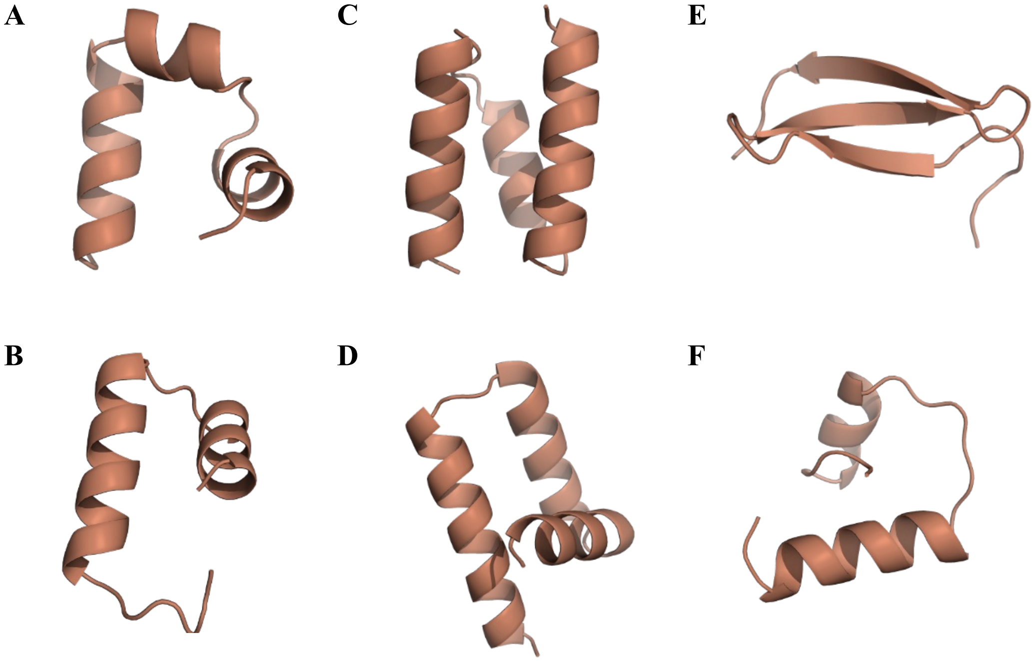


Figure 4.3 Structure predictions for peptides from pepFOLD4: VHP (**A**), VHP_rev (**B**), GA module (**C**), GA_rev (**D**), WW (**E**), and WW_rev (**F**).

The predicted structures for the reverse sequences revealed a consistent bias toward α-helical conformations. All three reversed peptides—including the reversed WW domain—were modeled as containing multiple α-helical segments (Figure 4.3 B, D, F). The reversed VHP peptide was predicted to form two short helices without any compact tertiary structure (Figure 4.3B). Similarly, the reversed GA module was modeled as three helices (Figure 4.3D). The reversed WW domain was predicted to contain two short α helices. However, CD data revealed that all reversed peptides were largely disordered in solution, showing no evidence of stable secondary structure (Figure 4.2). These results highlight the limitation of pepFOLD4's tendency to default to helical structures.

4.3.3.2 ESMFOLD Prediction

ESMFold performed exceptionally well on all three forward sequences. The predicted models for the α-helical peptides closely mirrored their crystallographic counterparts, with high confidence metrics (VHP, Figure 4.4A, RMSD = 0.587Å; GA module, Figure 4.4C, RMSD = 1.278Å). For the β-sheet WW domain, ESMFold successfully captured the antiparallel strand geometry and hydrogen bonding network, demonstrating its strength in modeling diverse secondary structures (Figure 4.4E, RMSD = 0.321Å).

Predictions for the reverse sequences revealed different behavior compared to other tools. For the reversed VHP peptide, ESMFold predicted the structure to be largely unstructured, with one short α-helical segment and an α-helical turn (Figure 4.4B). The reversed GA module was predicted to be two α-helices with a bend between (Figure 4.4D). The reversed WW domain was predicted to be largely unstructured with one short α-helical segment (Figure 4.4E). This prediction diverged from the fully helical models produced by other platforms and suggested that ESMFold partially recognized the disruption in foldability caused by sequence inversion. These predictions were in partial agreement with experimental CD data (Figure 4.2), which indicated a lack of defined secondary structure in all reversed peptides.

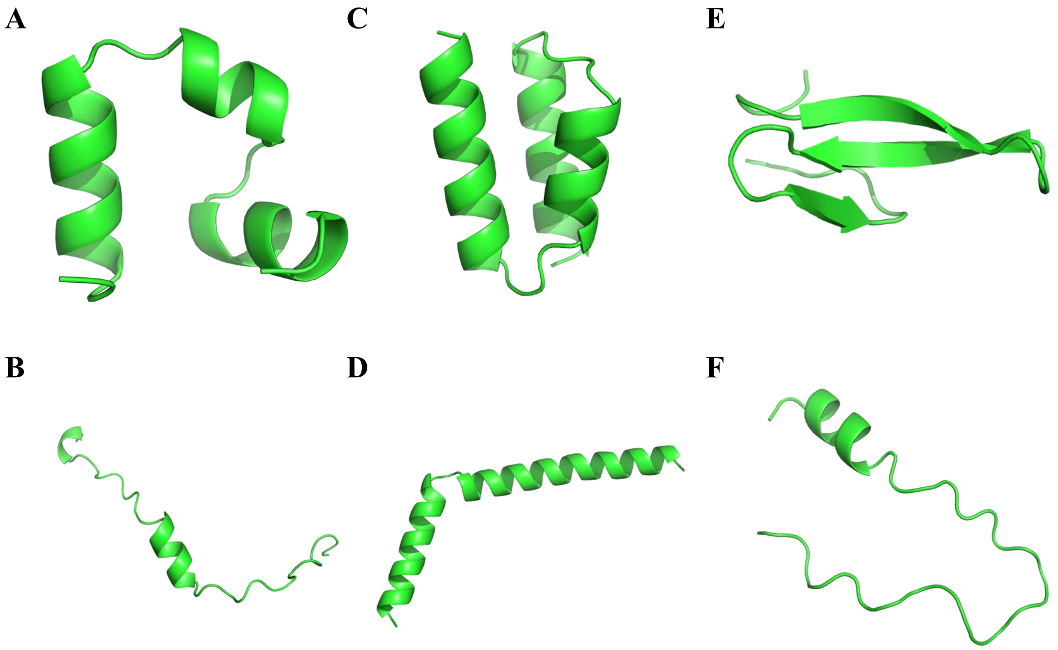


Figure 4.4. Structure predictions for peptides from ESMFOLD: VHP (**A**), VHP_rev (**B**), GA module (**C**), GA_rev (**D**), WW (**E**), and WW_rev (**F**).

4.3.3.3 AlphaFOLD3 Prediction

AlphaFold3 also generated accurate models for the forward peptides, achieving excellent agreement with experimental structures (VHP, Figure 4.5A, RMSD = 0.412Å; GA module, Figure 4.5C, RMSD = 1.675Å; WW domain, Figure 4.5E, RMSD = 0.388 Å). The predicted structures were assessed with the confidence scores, predicted Template Modeling (pTM).[34] A pTM score above 0.5 means the overall predicted fold for the complex might be similar to the true structure. The pTM scores were uniformly high (VHP, pTM = 0.64; GA module, pTM = 0.81; WW, pTM = 0.64), underscoring AlphaFold3's superiority in native sequence modeling.

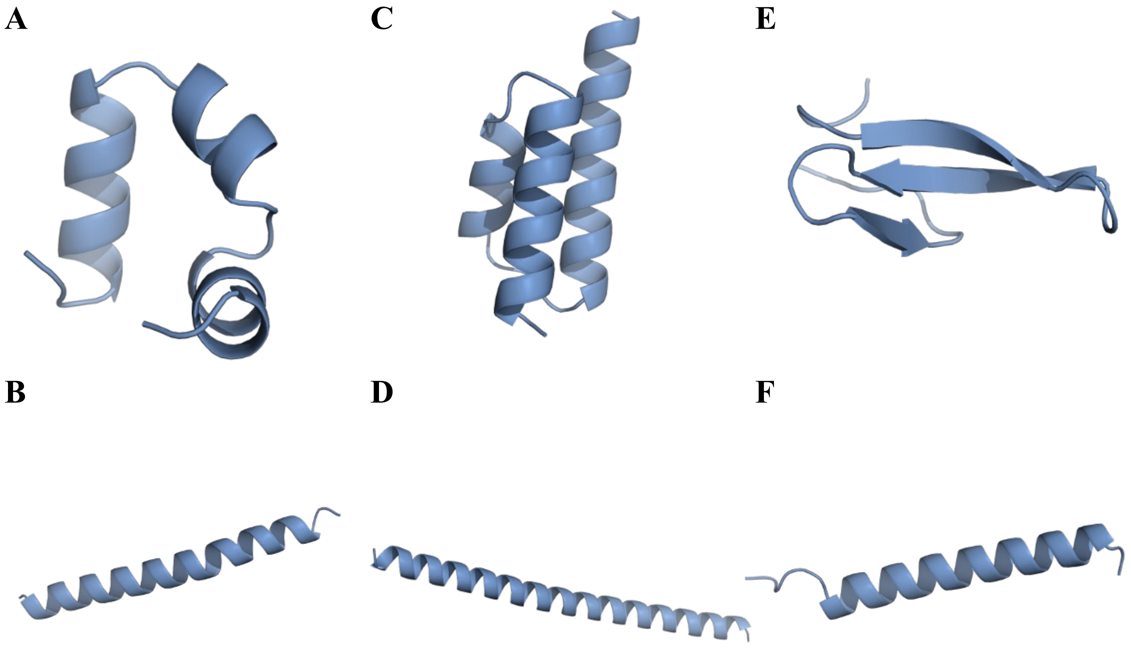


Figure 4.5 Structure predictions for peptides from AlphaFOLD3: VHP (**A**), VHP_rev (**B**), 1PRB (**C**), 1PRB_rev (**D**), WW (**E**), and WW_rev (**F**).

When reverse-ordered sequences were submitted, AlphaFold3 exhibited similar limitations as the other tools. The reversed VHP peptide (Figure 4.5B) and reversed GA module (Figure 4.5D) were both predicted to form overly regularized α-helices. The most significant discrepancy occurred with the reversed β-sheet WW peptide, which was modeled as a well-formed α-helix

(Figure 4.5F). However, CD data revealed that all reversed peptides were largely disordered in solution (Figure 4.2), contradicting with the predicted outcomes. These results indicate that AlphaFold3 shares the bias toward helical folding evident for other predictive tools, when presented with non-natural inputs.

Table 4.1. Comparison of structural predictions with CD results .

Peptide Type	**pepFOLD4 Prediction**	**ESMFold Prediction**	**AlphaFold3 Prediction**	**CD Result**	**Structural info**
VHP	Folded, native-like	Folded, native-like	Folded, native-like	Structured (α-helical)	a tertiary structure with three α-helices
VHP_rev	Multiple α-helices	Weak α-helical	Strong α-helical	Disordered	-
1PRB	Folded, native-like	Folded, native-like	Folded, native-like	Structured (α-helical)	α-helical bundle
1PRB_rev	Multiple α-helices	Strong α-helical	Strong α-helical	Disordered	-
WW	Folded, native-like	Folded, native-like	Folded, native-like	Structured (β-sheet)	β-sheet rich peptide
WW_rev	Multiple α-helices	Weak α-helical	Strong α-helical	Disordered	-

Together, these findings highlight a limitation of current peptide structure prediction tools: a tendency to overpredict α-helical content in reversed peptide sequences (Table 4.1). This behavior was likely from the insufficient modeling of sequence directionality. The results underscore the importance of experimental validation when applying predictive models to modified or synthetic peptides, particularly those are unnatural.

4.4 Conclusion

This study evaluated the performance of three widely used peptide structure prediction tools—pepFOLD4, ESMFold, and AlphaFold3—on both native (forward) and reversed peptide sequences. By using three structurally distinct peptides, a tertiary α-helix-rich, VHP, an α-helical bundle, GA module, and a β-sheet structure, WW domain, we assessed whether these tools could accurately model peptide conformations when sequence order is reversed.

Across all platforms, forward sequences were predicted with high fidelity, aligning closely with known structures and confirming the strength of these tools for native-like input. In contrast, reverse-ordered sequences were consistently mispredicted. Both pepFOLD4 and AlphaFold3 incorrectly modeled the reverse peptides as ordered, α-helical structures, regardless of their true native fold or secondary structure class. ESMFold partially recognized the disruption in foldability caused by sequence inversion, with predicted structures mostly unstructured with few α-helical contents. Circular dichroism spectroscopy revealed that the reversed peptides did not adopt stable secondary structures in solution and instead behaved as disordered, unstructured chains. This experimental evidence clearly contradicts the predicted α-helical conformations and underscores a key limitation in the current generation of peptide structure prediction models.

Collectively, our findings demonstrate that despite advances in deep learning and physics-based modeling, these algorithms remain sensitive when dealing with synthetic or unnatural inputs. The consistent bias toward α-helical predictions in reversed sequences likely reflects both training data composition and inductive modeling assumptions favoring order over disorder. These results emphasize the necessity of incorporating experimental validation, particularly for synthetic peptide variants or engineered sequences, when using prediction models for structure-based design. Future improvements in prediction algorithms should explicitly account for disordered states and sequence orientation, broadening their utility for non-natural peptides and novel protein engineering applications.

4.5. Materials and Methods

4.5.1. Peptide synthesis, purification and characterization

All peptides were prepared via microwave-assisted solid-phase synthesis using a LibertyBlue instrument. Fmoc-protected amino acids were purchased from Chem-Impex. Phenylalanine and glycine preloaded Wang resins were purchased from CEM. Coupling reagents ethyl cyano(hydroxyimino)acetate (Oxyma) and N,N'-diisopropylcarbodiimide (DIC) were purchased from Chem-Impex. N,N-dimethylformamide (DMF), dichloromethane (DCM), piperidine, trifluoroacetic acid (TFA), triisopropylsilane (TIPS), thioanisole, 1,2-ethanedithiol (EDT), and acetonitrile were purchased from Sigma-Aldrich.

All peptides were synthesized on a 50 µmole scale with 0.2 M Oxyma in DMF and 0.2 M DIC in DMF. Pre-loaded Wang resin was pre-swelled in 1:1 v/v DCM:DMF. Fmoc-amino acids (1.25 mL, 0.2 M) were activated with DIC (0.5 mL, 0.5 M) and Oxyma (0.5 mL, 1 M) in biotech grade DMF. Regular coupling cycles were performed at 90°C for 2 minutes. Histidine was coupled at 50°C to prevent racemization.

After all residues had been added, the resin was mixed with the cleavage cocktail (8 mL per 50 µmole) containing 91.5% (v/v) TFA, 2.5% (v/v) H_2O, 2.5% (v/v) EDT, 2.5% (v/v) thioanisole, and 1% (v/v) TIPS at room temperature for 3.5 hours for VHP, 1PRB and their reversed peptides or 5 hours for WW and WW_rev. Excess TFA was removed under a stream of nitrogen, and 35 mL of cold diethyl ether was then added to precipitate the crude peptide. The peptide was isolated by centrifugation, washed with cold ether twice, and then dried under nitrogen.

For VHP, VHP_rev, GA module and GA_rev, ~25 µmole of the crude product was dissolved in 600 µL 1:1 H_2O:ACN and filtered through a 0.22 µm Nylon filter for reversed-phase HPLC purification. VHP analogs were purified using a C18 HPLC column on an Agilent HPLC. H_2O and ACN with 0.1 % TFA were used as solvents A and B, respectively. A gradient from 10% to 60% B over 20 minutes with a flow rate of 12 mL/minute was used. For WW domain and WW_rev, ~25 µmole of the crude peptide was dissolved in 1500 µL 1:1 H_2O:ACN and then filtered for reversed-phase HPLC purification. WW domain analogs were first purified using a C8 column on an Agilent HPLC using 10% to 60% B over 20 minutes with a flow rate of 12 mL/minute. For the second purification, the WW domain analogs were purified with a C8 column using 20-50% B over 20 minutes with flow rate = 14mL/ min. Peptide masses were determined using MALDI-TOF-MS, and purity was checked using UPLC (at 220 nm).

4.5.2. Circular Dichroism (CD) Spectroscopy

CD experiments were conducted at 25 °C on a JASCO J-1500 CD spectrometer. The VHP peptide and its reversed peptide VHP_rev were diluted to 50 µM using 20 mM sodium acetate buffer (pH = 5.0).[31] The GA module and its reversed peptide GA_rev were diluted to 50 µM using 1X PBS buffer, pH 7.4.[26] The WW domain and its reversed peptide WW_rev were diluted to 50 µM using 10 mM sodium phosphate buffer (pH = 7.0).[27] The peptide sample was then transferred to a 1-mm quartz cuvette. The CD spectrum was measured from 260 to 200 nm with 0.1 nm intervals at 4 second digital integration time (D.I.T) at 100 nm/min scanning speed. The data are presented as mean residue ellipticity in deg•cm2•M-1 calculated based on the equation shown below.

molar ellipticity (θ) =(mdeg value)/(32.98*concentration (mol/L)*path length (cm))

4.6. MALDI-TOF-MS Spectra and UPLC Data for Synthetic Peptides

VHP: **H_2N-LSDEDFKAVFGMTRSAFANLPLWKQQHLKKEKGLF-COOH**

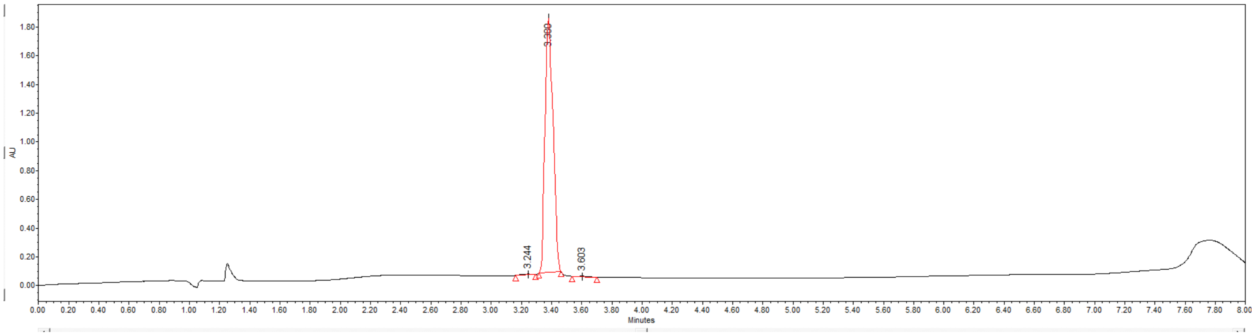
ESI-MS: calculated monoisotopic $[M+H]^+$ = 4080.1528

observed monoisotopic $[M+H]^+$ = 4080.153

(*ESI-MS/MS was used to obtain fragment ionized peptides and the measured the mass differences between the resulting fragment ions in MS/MS graphs were used to determine the amino acid sequence.)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity > 99.0 %



MS-MS fragmentation of peptide: LSDEDFKAVFGMTRSAFANLPLWKQQHLKKEKGLF

Elemental composition: C189 H291 N49 O50 S1

Mass: [Mo.] 4080.153 Da.(MH+) 2040.580 Da.(M2H+) 1360.722 Da.(M3H+) [Mass file: AA_MASS.MSS]

a	b	Res:	x	y'	z'	v'	w'
86.097	114.092	1 Leu 35	-	-	-	-	
173.129	201.124 ✓	2 Ser 34	3993.049	3967.069	3951.050	3936.050	3935.055
288.156	316.151 ✓	3 Asp 33	3906.017	3880.037	3864.018	3821.023	3820.028
417.199	445.193 ✓	4 Glu 32	3790.990	3765.010	3748.991	3691.981	3690.986
532.225	560.220 ✓	5 Asp 31	3661.947	3635.967	3619.948	3576.954	3575.959
679.294	707.289	6 Phe 30	3546.920	3520.940	3504.922	3429.885	3428.890
807.389	835.384 ✓	7 Lys 29	3399.852	3373.872	3357.853	3301.790	3300.795
878.426	906.421 ✓	8 Ala 28	3271.757	3245.777	3229.758	3230.753	-
977.494	1005.489 ✓	9 Val 27	3200.720	3174.740	3158.721	3131.685	3144.705
1124.563	1152.558 ✓	10 Phe 26	3101.651	3075.671	3059.653	2984.617	2983.621
1181.584	1209.579 ✓	11 Gly 25	2954.583	2928.603	2912.584	-	-
1312.625	1340.620	12 Met 24	2897.561	2871.581	2855.563	2796.555	2795.559
1413.672	1441.667	13 Thr 23	2766.521	2740.541	2724.522	2695.507	2708.527
1569.774	1597.768	14 Arg 22	2665.473	2639.493	2623.475	2539.406	2538.411
1656.806	1684.800	15 Ser 21	2509.372	2483.392	2467.373	2452.374	2451.379
1727.843	1755.838	16 Ala 20	2422.340	2396.360	2380.341	2381.337	-
1874.911	1902.906	17 Phe 19	2351.303	2325.323	2309.304	2234.268	2233.273
1945.948	1973.943	18 Ala 18	2204.234	2178.255	2162.236	2163.231	-
2059.991	2087.986 ✓	19 Asn 17	2133.197	2107.217	2091.199	2049.188	2048.193
2173.075	2201.070	20 Leu 16	2019.154	1993.175	1977.156	1936.104	1935.109
2270.128	2298.123	21 Pro 15	1906.070	1880.091 ✓	1864.072	-	-
2383.212	2411.207 ✓	22 Leu 14	1809.018	1783.038	1767.019	1725.967	1724.972
2569.291	2597.286	23 Trp 13	1695.933	1669.954	1653.935	1539.888	1538.893
2697.386	2725.381	24 Lys 12	1509.854	1483.874 ✓	1467.856	1411.793	1410.798
2825.445	2853.440	25 Gln 11	1381.759	1355.779	1339.761	1283.734	1282.739
2953.503	2981.498	26 Gln 10	1253.701	1227.721 ✓	1211.702	1155.676	1154.681
3090.562	3118.557	27 His 9	1125.642	1099.662 ✓	1083.644	1018.617	1017.622
3203.646	3231.641	28 Leu 8	988.583	962.603 ✓	946.585	905.533	904.538
3331.741	3359.736	29 Lys 7	875.499	849.519 ✓	833.501	777.438	776.443
3459.836	3487.831	30 Lys 6	747.404	721.424 ✓	705.406	649.343	648.348
3588.879	3616.874	31 Glu 5	619.309	593.329 ✓	577.311	520.300	519.305
3716.974	3744.969	32 Lys 4	490.267	464.287 ✓	448.268	392.205	391.210
3773.995	3801.990	33 Gly 3	362.172	336.192 ✓	320.173	-	-
3887.079	3915.074	34 Leu 2	305.150	279.170 ✓	263.152	222.100	221.105
-	-	35 Phe 1	192.066	166.086	150.068	75.031	74.036

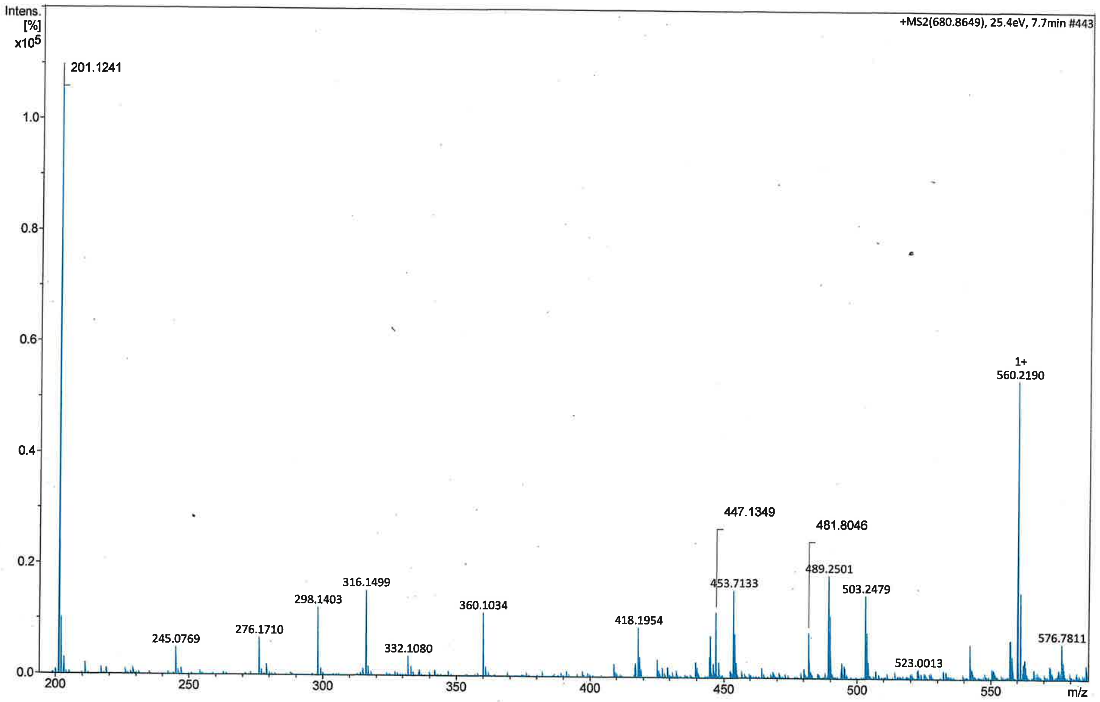
- Related immonium ions:
- 88.039[D] 87.055[N] 74.060[T] 60.044[S] 102.055[E] 101.071[Q] 70.065[P]
- 30.034[G] 44.049[A] 72.081[V] 104.053[M] 86.096[L] 120.081[F] 101.107[K]
- 110.071[H] 159.092[W] 129.113[R]

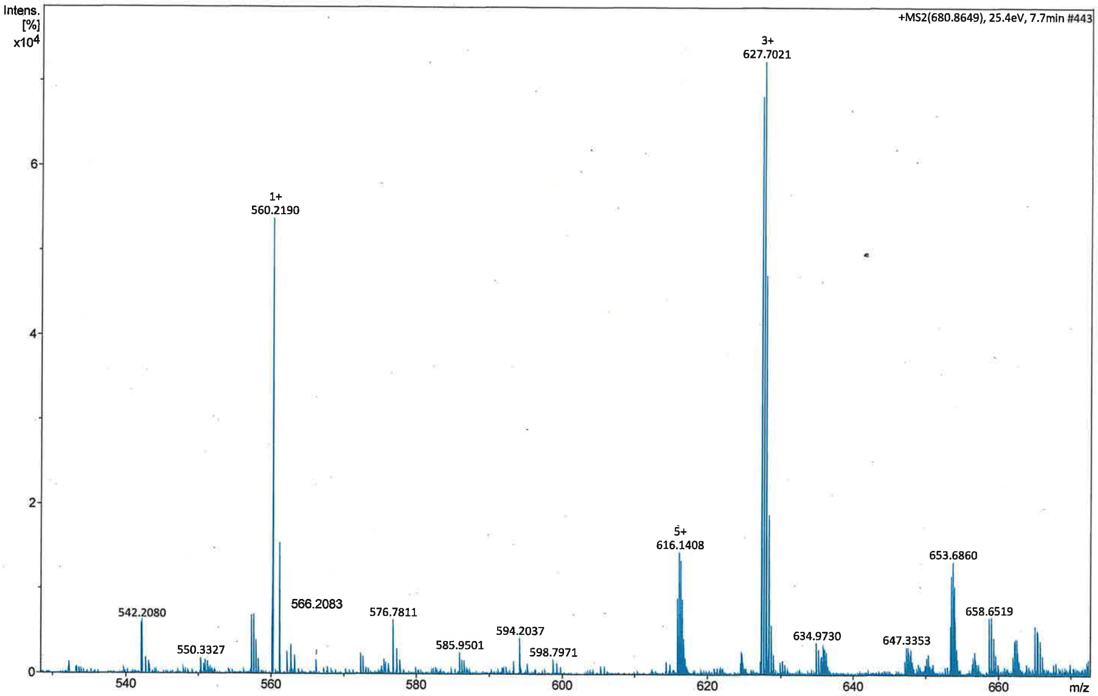
N-terminal: Hydrogen C-terminal: Free acid

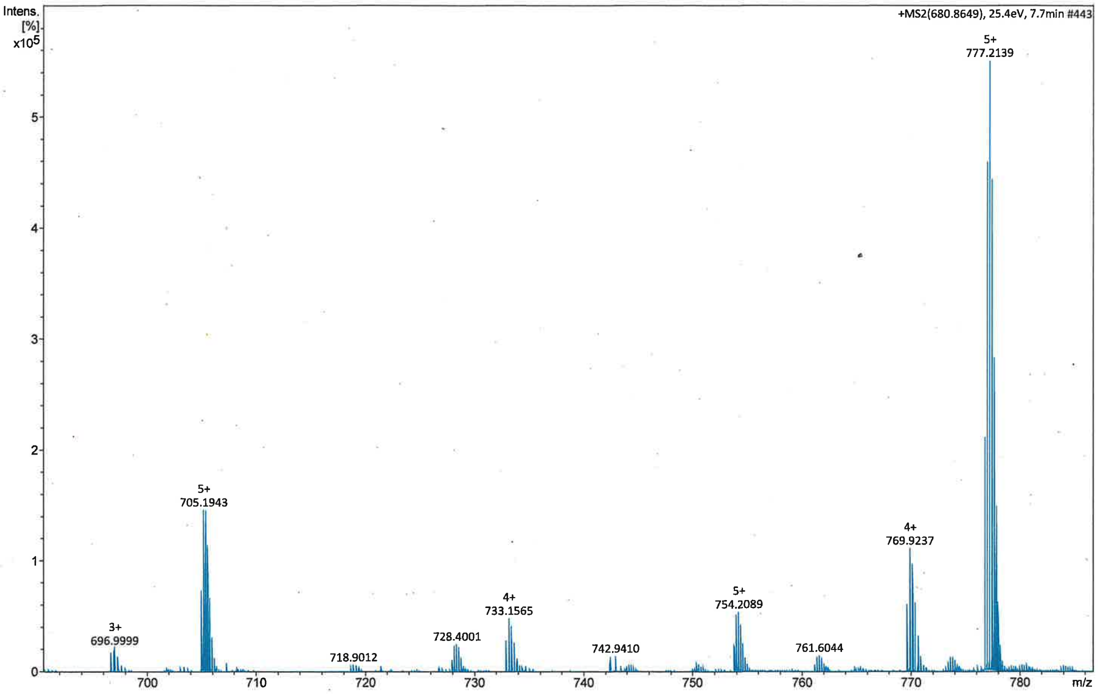
Sorted mass list:

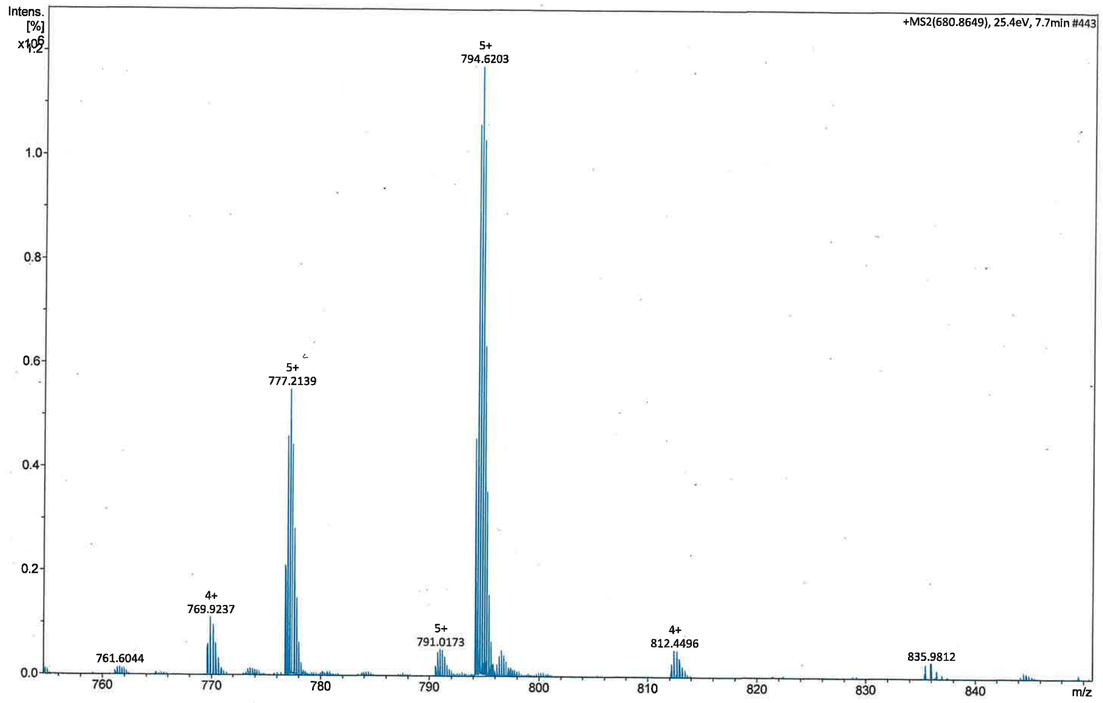
```
Mass Type          1227.721   y' 10  -2059.991   a 19  -2853.440   b 25  -3661.947   x 31
86.097    a  1  +  1253.701    x 10  -2087.986   b 19  -2855.563  z' 24  -3690.986  w' 32
114.092    b  1  +1282.739   w' 11  -2091.199  z' 17  -2871.581  y' 24  -3691.981  v' 32
173.129    a  2  +1283.734   v' 11  -2107.217  y' 17  -2897.561   x 24  -3748.991  z' 32
201.124    b  2  +1312.625    a 12  -2133.197   x 17  -2912.584  z' 25  -3765.010  y' 32
288.156    a  3  +1339.761   z' 11  -2162.236  z' 18  -2928.603  y' 25  -3790.990   x 32
316.151    b  3  +1340.620    b 12  -2163.231  v' 18  -2953.503   a 26  -3820.028  w' 33
417.199    a  4  +1355.779   y' 11  -2173.075   a 20  -2954.583   x 25  -3821.023  v' 33
445.193    b  4  +1381.759    x 11  -2178.255  y' 18  -2981.498   b 26  -3864.018  z' 33
532.225    a  5  +1410.798   w' 12  -2201.070   b 20  -2983.621  w' 26  -3880.037  y' 33
560.220    b  5  +1411.793   v' 12  -2204.234   x 18  -2984.617  v' 26  -3906.017   x 33
                                          page 1
```

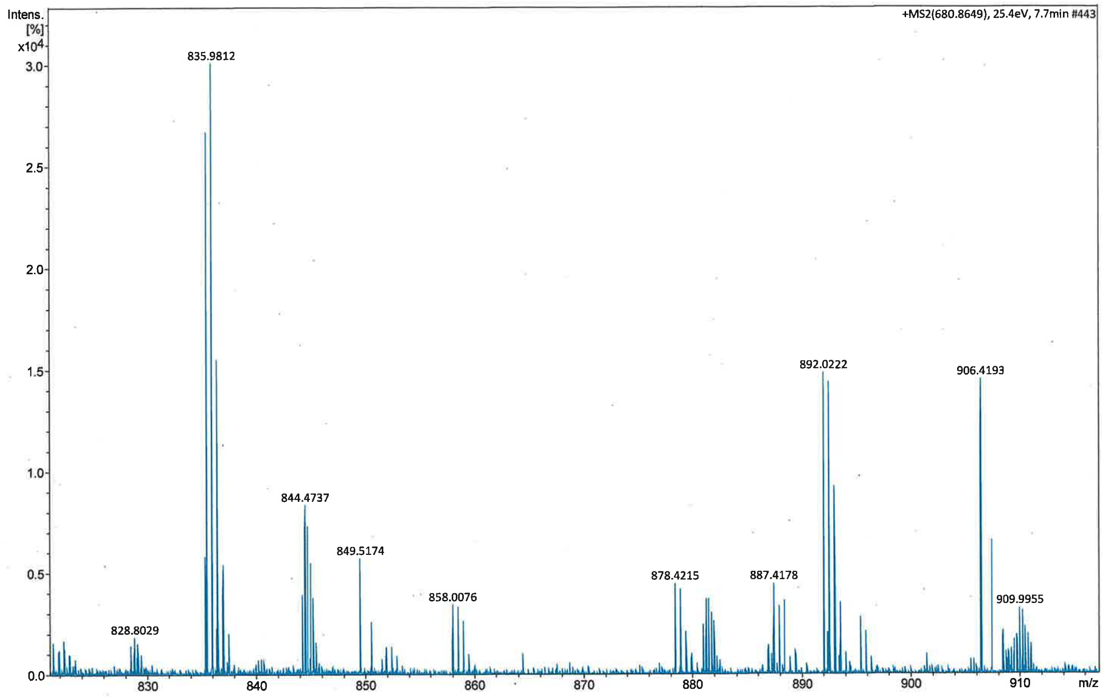
```
648.348    w'  6  +1413.672   a 13  -2233.273   w' 19  -3059.653   z' 26  -3935.055   w' 34  -
649.343    v'  6  +1441.667   b 13  -2234.268   v' 19  -3075.671   y' 26  -3936.050   v' 34  -
679.294     a  6  +1467.856  z' 12  -2270.128    a 21  -3090.562    a 27  -3951.050   z' 34  -
705.406    z'  6  +1483.874  y' 12  -2298.123    b 21  -3101.651    x 26  -3967.069   y' 34  -
707.289     b  6  +1509.854   x 12  -2309.304   z' 19  -3118.557    b 27  -3993.049    x 34  -
721.424    y'  6  +1538.893  w' 13  -2325.323   y' 19  -3131.685   v' 27  -
747.404     x  6  +1539.888  v' 13  -2351.303    x 19  -3144.705   w' 27  -
776.443    w'  7  +1569.774   a 14  -2380.341   z' 20  -3158.721   z' 27  -
777.438    v'  7  +1597.768   b 14  -2381.337   v' 20  -3174.740   y' 27  -577.311    z'  5  +
807.389     a  7  +1653.935  z' 13  -2383.212    a 22  -3200.720    x 27  -520.300    v'  5  +
833.501    z'  7  +1656.806   a 15  -2396.360   y' 20  -3203.646    a 28  -519.305    w'  5  +
835.384     b  7  +1669.954  y' 13  -2411.207    b 22  -3229.758   z' 28  -3716.974    a 32  -
849.519    y'  7  +1684.800   b 15  -2422.340    x 20  -3230.753   v' 28  -3744.969    b 32  -
875.499     x  7  +1695.933   x 13  -2451.379   w' 21  -3231.641    b 28  -490.267     x  4  +
878.426     a  8  +1724.972  w' 14  -2452.374   v' 21  -3245.777   y' 28  -464.287    y'  4  +
904.538    w'  8  +1725.967  v' 14  -2467.373   z' 21  -3271.757    x 28  -448.268    z'  4  +
905.533    v'  8  +1727.843   a 16  -2483.392   y' 21  -3300.795   w' 29  -392.205    v'  4  +
906.421     b  8  +1755.838   b 16  -2509.372    x 21  -3301.790   v' 29  -391.210    w'  4  +
946.585    z'  8  +1767.019  z' 14  -2538.411   w' 22  -3331.741    a 29  -3773.995    a 33  -
962.603    y'  8  +1783.038  y' 14  -2539.406   v' 22  -3357.853   z' 29  -3801.990    b 33  -
977.494     a  9  +1809.018   x 14  -2569.291    a 23  -3359.736    b 29  -362.172     x  3  +
988.583     x  8  +1864.072  z' 15  -2597.286    b 23  -3373.872   y' 29  -336.192    y'  3  +
1005.489    b  9  -1874.911   a 17  -2623.475   z' 22  -3399.852    x 29  -320.173    z'  3  +
1017.622   w'  9  -1880.091  y' 15  -2639.493   y' 22  -3428.890   w' 30  -3887.079    a 34  -
1018.617   v'  9  -1902.906   b 17  -2665.473    x 22  -3429.885   v' 30  -3915.074    b 34  -
1083.644   z'  9  -1906.070   x 15  -2695.507   v' 23  -3459.836    a 30  -305.150     x  2  +
1099.662   y'  9  -1935.109  w' 16  -2697.386    a 24  -3487.831    b 30  -279.170    y'  2  +
1124.563    a 10  -1936.104  v' 16  -2708.527   w' 23  -3504.922   z' 30  -263.152    z'  2  +
1125.642    x  9  -1945.948   a 18  -2724.522   z' 23  -3520.940   y' 30  -222.100    v'  2  +
1152.558    b 10  -1973.943   b 18  -2725.381    b 24  -3546.920    x 30  -221.105    w'  2  +
1154.681   w' 10  -1977.156  z' 16  -2740.541   y' 23  -3575.959   w' 31  -192.066     x  1  +
1155.676   v' 10  -1993.175  y' 16  -2766.521    x 23  -3576.954   v' 31  -166.086    y'  1  +
1181.584    a 11  -2019.154   x 16  -2795.559   w' 24  -3588.879    a 31  -150.068    z'  1  +
1209.579    b 11  -2048.193  w' 17  -2796.555   v' 24  -3619.948   z' 31  -75.031    v'  1  +
1211.702   z' 10  -2049.188  v' 17  -2825.445    a 25  -3635.967   y' 31  +
```

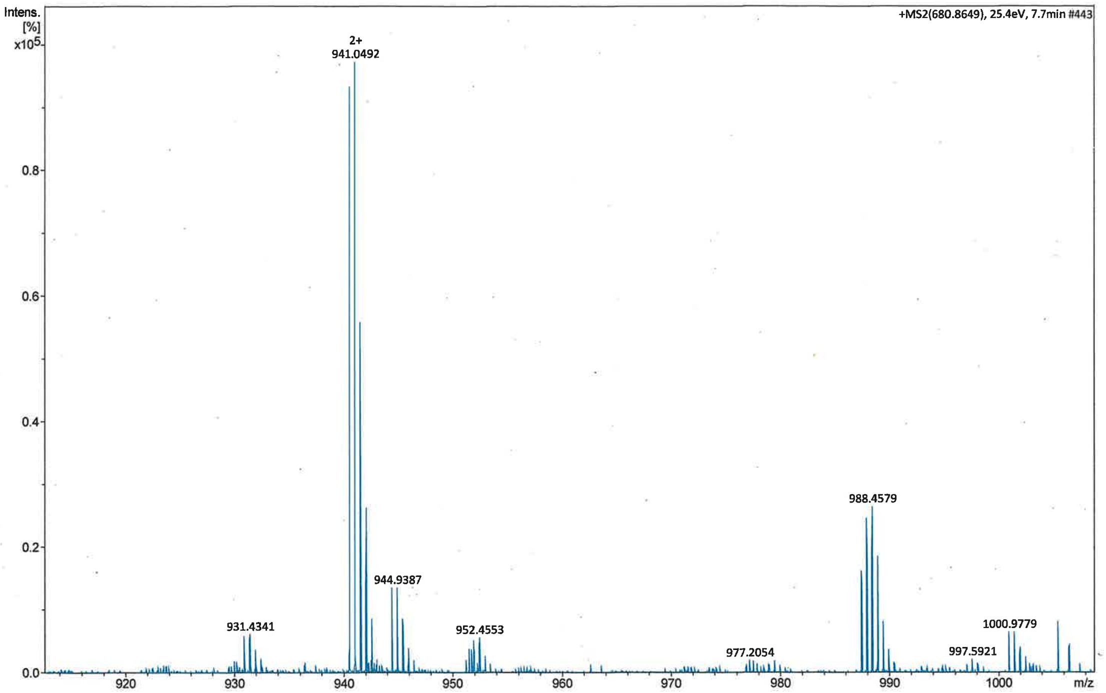


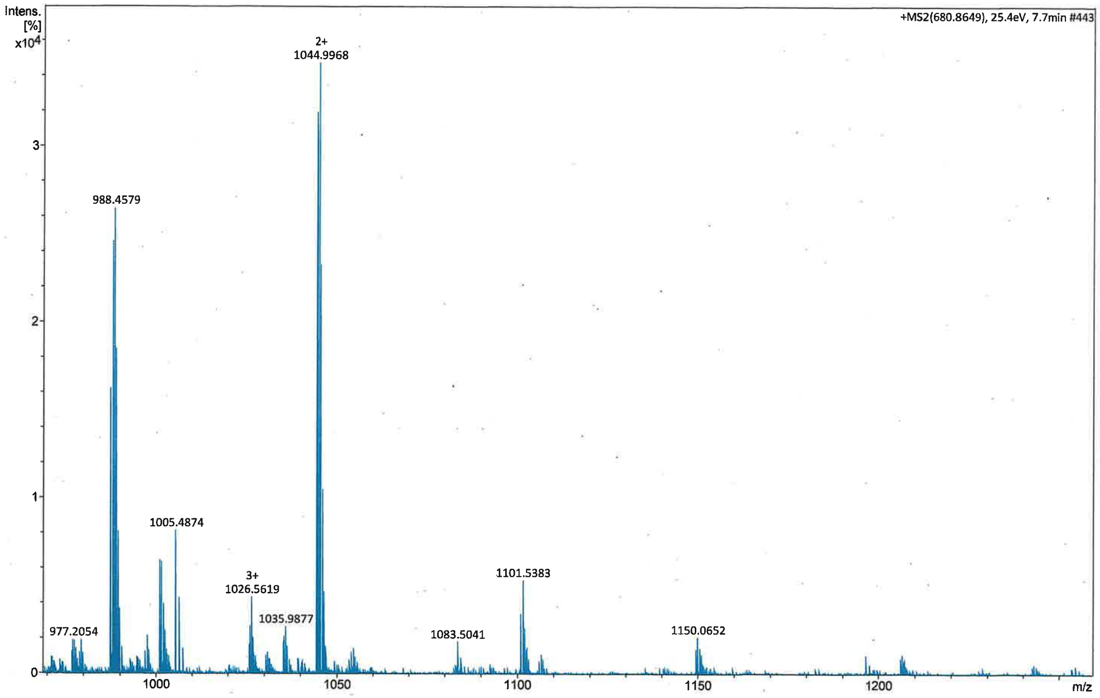

Intens.
[%]
x10^4
+MS2(680.8649), 25.4eV, 7.7min #443
3+
627.7021
1+
560.2190
5+
616.1408
653.6860
542.2080
566.2083
576.7811
658.6519
594.2037
634.9730
647.3353
585.9501
550.3327
598.7971
0
2
4
6
540
560
580
600
620
640
660
m/z


Intens.
[%]
x10^5
+MS2(680.8649), 25.4eV, 7.7min #443
5+
777.2139
5+
705.1943
4+
769.9237
4+
733.1565
5+
754.2089
3+
696.9999
728.4001
742.9410
761.6044
718.9012
0
1
2
3
4
5
700
710
720
730
740
750
760
770
780
m/z


Intens.
[%]
x10^6
+MS2(680.8649), 25.4eV, 7.7min #443
5+
794.6203
5+
777.2139
4+
769.9237
5+
791.0173
4+
812.4496
761.6044
835.9812
0.0
0.2
0.4
0.6
0.8
1.0
1.2
760
770
780
790
800
810
820
830
840
m/z


Intens.
[%]
x10^4
+MS2(680.8649), 25.4eV, 7.7min #443
835.9812
844.4737
849.5174
858.0076
828.8029
878.4215
887.4178
892.0222
906.4193
909.9955
0.0
0.5
1.0
1.5
2.0
2.5
3.0
830
840
850
860
870
880
890
900
910
m/z


Intens.
[%]
x10^5
+MS2(680.8649), 25.4eV, 7.7min #443
2+
941.0492
944.9387
931.4341
952.4553
977.2054
988.4579
997.5921
1000.9779
0.8
0.6
0.4
0.2
0.0
920
930
940
950
960
970
980
990
1000
m/z


Intens.
[%]
x10^4
+MS2(680.8649), 25.4eV, 7.7min #443
2+
1044.9968
988.4579
1005.4874
3+
1026.5619
1035.9877
977.2054
1083.5041
1101.5383
1150.0652
3
2
1
0
1000
1050
1100
1150
1200
m/z

VHP_rev: **H_2N-FLGKEKKLHQQKWLPLNAFASRTMGFVAKFDEDSL-COOH**

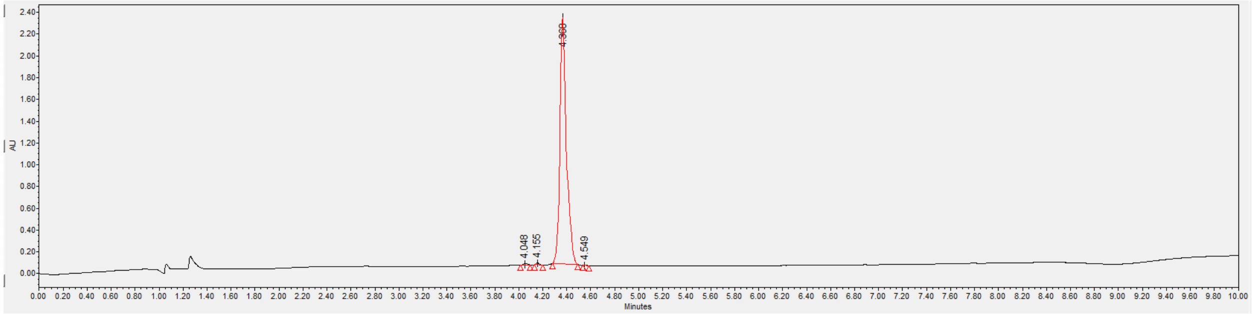
ESI-MS: calculated monoisotopic $[M+H]^+$ = 4080.1528

observed monoisotopic $[M+H]^+$ = 4080.153

(*ESI-MS/MS was used to obtain fragment ionized peptides and the measured the mass differences between the resulting fragment ions in MS/MS graphs were used to determine the amino acid sequence.)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity > 99.0 %



MS-MS fragmentation of peptide: FLGKEKKLHQQKWLPLNAFASRTMGFVAKFDEDSL

Elemental composition: C189 H291 N49 O50 S1

Mass: [Mo.] 4080.153 Da.(MH+) 2040.580 Da.(M2H+) 1360.722 Da.(M3H+) [Mass file: AA_MASS.MSS]

a	b	Res:	x	y'	z'	v'	w'
120.081	148.076	1 Phe 35	-	-	-	-	
233.165	261.160 ✓	2 Leu 34	3959.064	3933.084	3917.066	3876.014	3875.019
290.187	318.182 ✓	3 Gly 33	3845.980	3820.000	3803.982	-	-
418.282	446.277 ✓	4 Lys 32	3788.959	3762.979	3746.960	3690.898	3689.902
547.324	575.319 ✓	5 Glu 31	3660.864	3634.884	3618.865	3561.855	3560.860
675.419	703.414	6 Lys 30	3531.821	3505.841	3489.823	3433.760	3432.765
803.514	831.509	7 Lys 29	3403.726	3377.746	3361.728	3305.665	3304.670
916.598	944.593	8 Leu 28	3275.631	3249.651	3233.633	3192.581	3191.586
1053.657	1081.652 ✓	9 His 27	3162.547	3136.567	3120.549	3055.522	3054.527
1181.716	1209.711	10 Gln 26	3025.488	2999.508	2983.490	2927.463	2926.468
1309.774	1337.769	11 Gln 25	2897.430	2871.450	2855.431	2799.405	2798.410
1437.869	1465.864 ✓	12 Lys 24	2769.371	2743.391	2727.373	2671.310	2670.315
1623.949	1651.944	13 Trp 23	2641.276	2615.296	2599.278	2485.231	2484.235
1737.033	1765.028	14 Leu 22	2455.197	2429.217	2413.198	2372.147	2371.151
1834.086	1862.080	15 Pro 21	2342.113	2316.133	2300.114	-	-
1947.170	1975.165	16 Leu 20	2245.060	2219.080	2203.061	2162.010	2161.014
2061.213	2089.207	17 Asn 19	2131.976	2105.996	2089.977	2047.967	2046.972
2132.250	2160.245	18 Ala 18	2017.933	1991.953	1975.934	1976.930	-
2279.318	2307.313	19 Phe 17	1946.896	1920.916	1904.897	1829.861	1828.866
2350.355	2378.350	20 Ala 16	1799.827	1773.848	1757.829	1758.824	-
2437.387	2465.382	21 Ser 15	1728.790	1702.811	1686.792	1671.792	1670.797
2593.488	2621.483	22 Arg 14	1641.758	1615.778	1599.760	1515.691	1514.696
2694.536	2722.531	23 Thr 13	1485.657	1459.677	1443.659	1414.643	1427.664
2825.577	2853.571	24 Met 12	1384.609	1358.630	1342.611	1283.603	1282.608
2882.598	2910.593	25 Gly 11	1253.569	1227.589 ✓	1211.570	-	-
3029.666	3057.661	26 Phe 10	1196.548	1170.568 ✓	1154.549	1079.513	1078.518
3128.735	3156.730	27 Val 9	1049.479	1023.499 ✓	1007.481	980.445	993.465
3199.772	3227.767	28 Ala 8	950.411	924.431 ✓	908.412	909.407	-
3327.867	3355.862	29 Lys 7	879.374	853.394	837.375	781.312	780.317
3474.935	3502.930	30 Phe 6	751.279	725.299	709.280	634.244	633.249
3589.962	3617.957	31 Asp 5	604.210	578.230	562.212	519.217	518.222
3719.005	3747.000	32 Glu 4	489.183	463.203 ✓	447.185	390.175	389.179
3834.032	3862.027	33 Asp 3	360.141	334.161 ✓	318.142	275.148	274.152
3921.064	3949.059	34 Ser 2	245.114	219.134 ✓	203.115	188.116	187.120
-	-	35 Leu 1	158.082	132.102	116.083	75.031	74.036

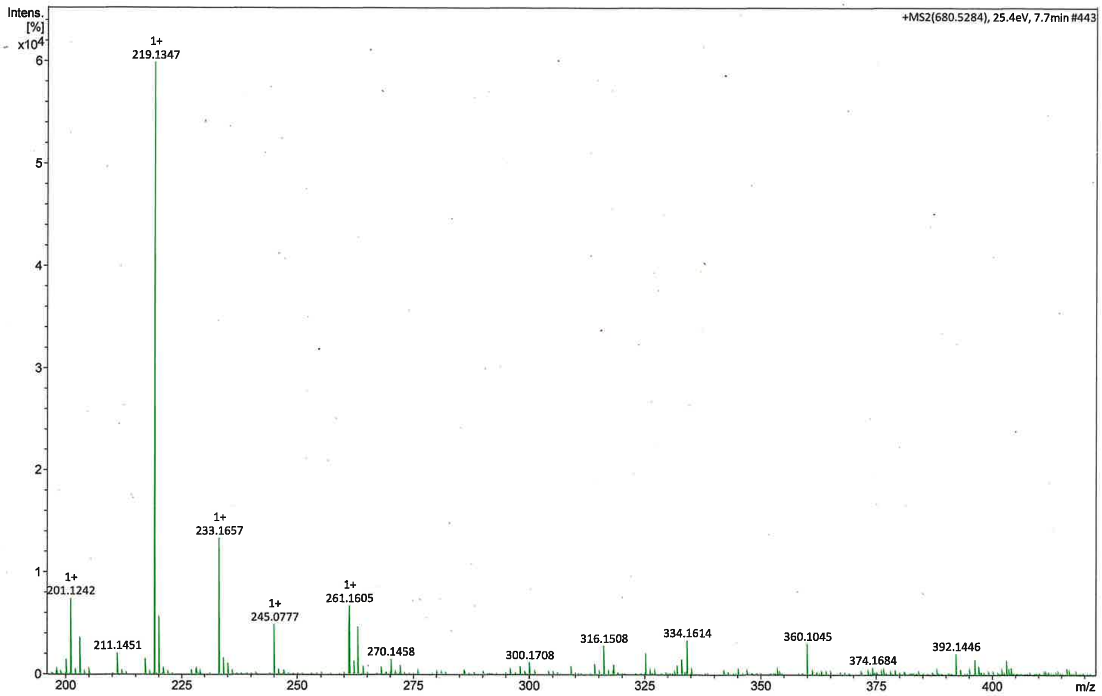
- Related immonium ions:
- 88.039[D] 87.055[N] 74.060[T] 60.044[S] 102.055[E] 101.071[Q] 70.065[P]
- 30.034[G] 44.049[A] 72.081[V] 104.053[M] 86.096[L] 120.081[F] 101.107[K]
- 110.071[H] 159.092[W] 129.113[R]

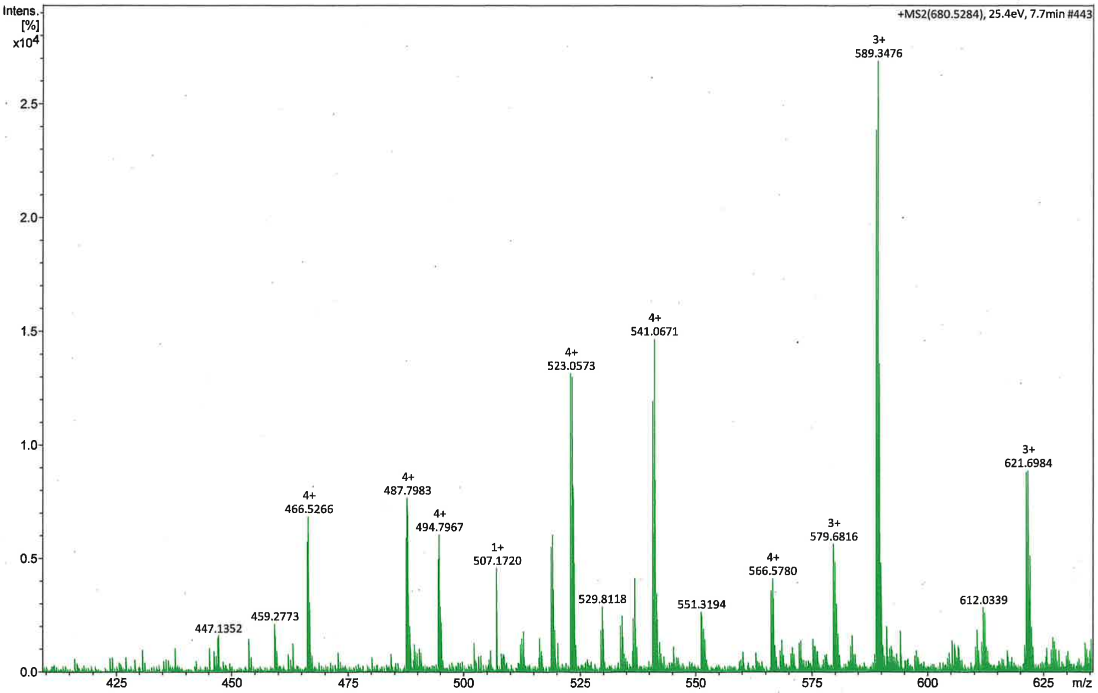
N-terminal: Hydrogen C-terminal: Free acid

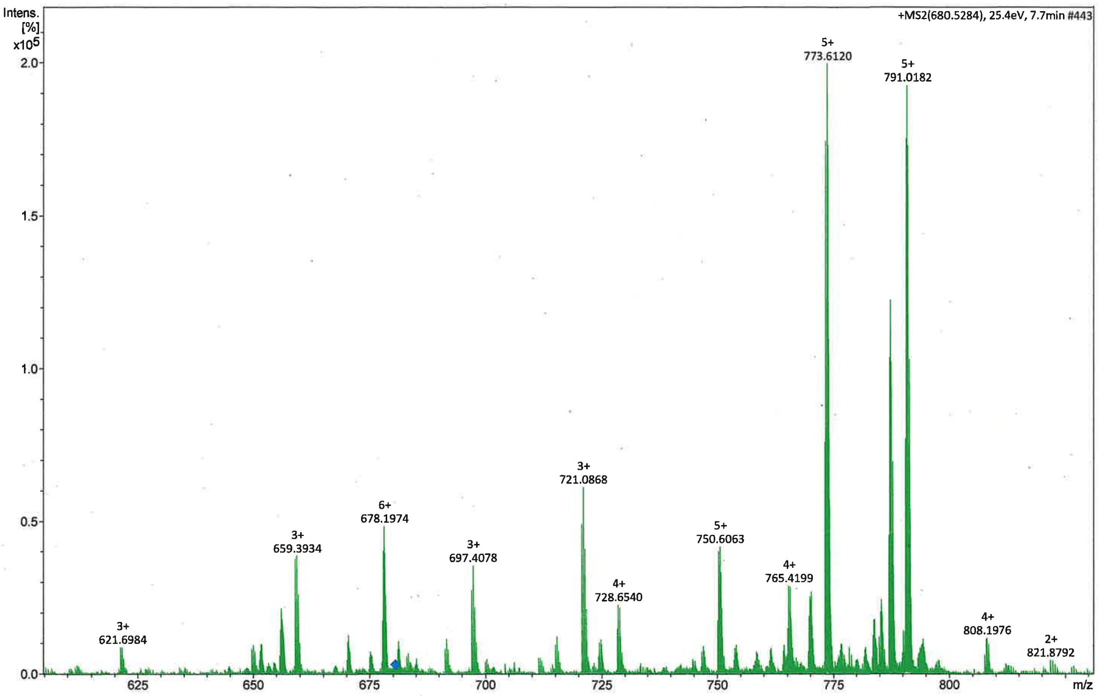
Sorted mass list:

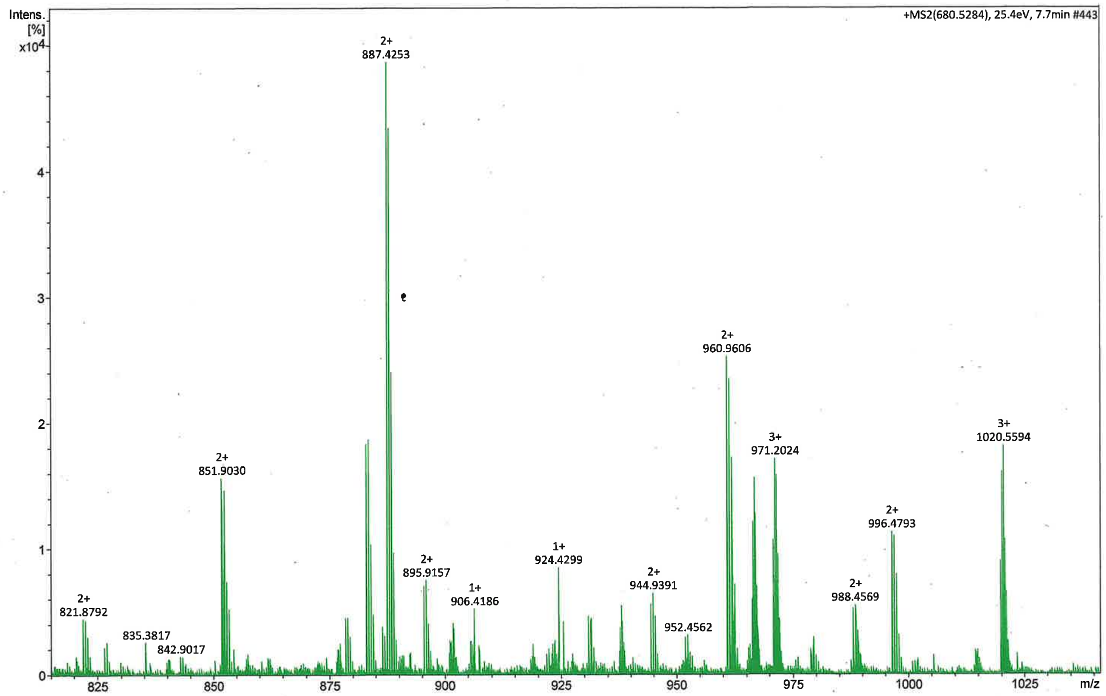
```
Mass Type              1211.570   z' 11   ·1991.953   y' 18   ·2825.577    a 24   ·3618.865   z' 31  ·
120.081    a  1   +1227.589   y' 11   ·2017.933    x 18   ·2853.571    b 24   ·3634.884   y' 31  ·
148.076    b  1   +1253.569    x 11   ·2046.972   w' 19   ·2855.431   z' 25   ·3660.864    x 31  ·
233.165    a  2   +1282.608   w' 12   ·2047.967   v' 19   ·2871.450   y' 25   ·3689.902   w' 32  ·
261.160    b  2   +1283.603   v' 12   ·2061.213    a 17   ·2882.598    a 25   ·3690.898   v' 32  ·
290.187    a  3   +1309.774    a 11   ·2089.207    b 17   ·2897.430    x 25   ·3746.960   z' 32  ·
318.182    b  3   +1337.769    b 11   ·2089.977   z' 19   ·2910.593    b 25   ·3762.979   y' 32  ·
418.282    a  4   +1342.611   z' 12   ·2105.996   y' 19   ·2926.468   w' 26   ·3788.959    x 32  ·
446.277    b  4   +1358.630   y' 12   ·2131.976    x 19   ·2927.463   v' 26   ·3803.982   z' 33  ·
547.324    a  5   +1384.609    x 12   ·2132.250    a 18   ·2983.490   z' 26   ·3820.000   y' 33  ·
575.319    b  5   +1414.643   v' 13   ·2160.245    b 18   ·2999.508   y' 26   ·3845.980    x 33  ·
```

```
604.210     x  5  +1427.664   w' 13  -2161.014   w' 20  -3025.488    x 26  -3875.019   w' 34  -
633.249    w'  6  +1437.869    a 12  -2162.010   v' 20  -3029.666    a 26  -3876.014   v' 34  -
634.244    v'  6  +1443.659   z' 13  -2203.061   z' 20  -3054.527   w' 27  -3917.066   z' 34  -
675.419     a  6  +1459.677   y' 13  -2219.080   y' 20  -3055.522   v' 27  -3933.084   y' 34  -
703.414     b  6  +1465.864    b 12  -2245.060    x 20  -3057.661    b 26  -3959.064    x 34  -
709.280    z'  6  +1485.657    x 13  -2279.318    a 19  -3120.549   z' 27  -
725.299    y'  6  +1514.696   w' 14  -2300.114   z' 21  -3128.735    a 27  -
751.279     x  6  +1515.691   v' 14  -2307.313    b 19  -3136.567   y' 27  -
780.317    w'  7  +1599.760   z' 14  -2316.133   y' 21  -3156.730    b 27  -518.222   w'  5  +
781.312    v'  7  +1615.778   y' 14  -2342.113    x 21  -3162.547    x 27  -3719.005    a 32  -
803.514     a  7  +1623.949    a 13  -2350.355    a 20  -3191.586   w' 28  -3747.000    b 32  -
831.509     b  7  +1641.758    x 14  -2371.151   w' 22  -3192.581   v' 28  -489.183    x  4  +
837.375    z'  7  +1651.944    b 13  -2372.147   v' 22  -3199.772    a 28  -463.203   y'  4  +
853.394    y'  7  +1670.797   w' 15  -2378.350    b 20  -3227.767    b 28  -447.185   z'  4  +
879.374     x  7  +1671.792   v' 15  -2413.198   z' 22  -3233.633   z' 28  -390.175   v'  4  +
908.412    z'  8  +1686.792   z' 15  -2429.217   y' 22  -3249.651   y' 28  -389.179   w'  4  +
909.407    v'  8  +1702.811   y' 15  -2437.387    a 21  -3275.631    x 28  -3834.032    a 33  -
916.598     a  8  +1728.790    x 15  -2455.197    x 22  -3304.670   w' 29  -3862.027    b 33  -
924.431    y'  8  +1737.033    a 14  -2465.382    b 21  -3305.665   v' 29  -360.141    x  3  +
944.593     b  8  +1757.829   z' 16  -2484.235   w' 23  -3327.867    a 29  -334.161   y'  3  +
950.411     x  8  +1758.824   v' 16  -2485.231   v' 23  -3355.862    b 29  -318.142   z'  3  +
980.445    v'  9  +1765.028    b 14  -2593.488    a 22  -3361.728   z' 29  -275.148   v'  3  +
993.465    w'  9  +1773.848   y' 16  -2599.278   z' 23  -3377.746   y' 29  -274.152   w'  3  +
1007.481    z'  9  -1799.827   x 16  -2615.296   y' 23  -3403.726    x 29  -3921.064    a 34  -
1023.499    y'  9  -1828.866  w' 17  -2621.483    b 22  -3432.765   w' 30  -3949.059    b 34  -
1049.479     x  9  -1829.861  v' 17  -2641.276    x 23  -3433.760   v' 30  -245.114    x  2  +
1053.657     a  9  -1834.086   a 15  -2670.315   w' 24  -3474.935    a 30  -219.134   y'  2  +
1078.518    w' 10  -1862.080   b 15  -2671.310   v' 24  -3489.823   z' 30  -203.115   z'  2  +
1079.513    v' 10  -1904.897  z' 17  -2694.536    a 23  -3502.930    b 30  -188.116   v'  2  +
1081.652     b  9  -1920.916  y' 17  -2722.531    b 23  -3505.841   y' 30  -187.120   w'  2  +
1154.549    z' 10  -1946.896   x 17  -2727.373   z' 24  -3531.821    x 30  -158.082    x  1  +
1170.568    y' 10  -1947.170   a 16  -2743.391   y' 24  -3560.860   w' 31  -132.102   y'  1  +
1181.716     a 10  -1975.165   b 16  -2769.371    x 24  -3561.855   v' 31  -116.083   z'  1  +
1196.548     x 10  -1975.934  z' 18  -2798.410   w' 25  -3589.962    a 31  -75.031   v'  1  +
1209.711     b 10  -1976.930  v' 18  -2799.405   v' 25  -3617.957    b 31  +
```




+MS2(680.5284), 25.4eV, 7.7min #443
Intens.
[%]
x10^4
2.5
2.0
1.5
1.0
0.5
0.0
3+
589.3476
4+
541.0671
4+
523.0573
3+
621.6984
4+
466.5266
4+
487.7983
4+
494.7967
1+
507.1720
3+
579.6816
4+
566.5780
529.8118
551.3194
612.0339
459.2773
447.1352
425
450
475
500
525
550
575
600
625
m/z


+MS2(680.5284), 25.4eV, 7.7min #443
Intens.
[%]
x10^5
2.0
1.5
1.0
0.5
0.0
5+
773.6120
5+
791.0182
3+
721.0868
6+
678.1974
3+
659.3934
3+
697.4078
5+
750.6063
4+
765.4199
4+
728.6540
4+
808.1976
2+
821.8792
3+
621.6984
625
650
675
700
725
750
775
800
m/z


Intens.
[%]
x10^4
+MS2(680.5284), 25.4eV, 7.7min #443
2+
887.4253
2+
960.9606
3+
971.2024
3+
1020.5594
2+
851.9030
2+
996.4793
1+
924.4299
2+
895.9157
2+
944.9391
2+
988.4569
1+
906.4186
2+
821.8792
952.4562
835.3817
842.9017
0
1
2
3
4
825
850
875
900
925
950
975
1000
1025
m/z

WW: **H_2N-KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG-COOH**

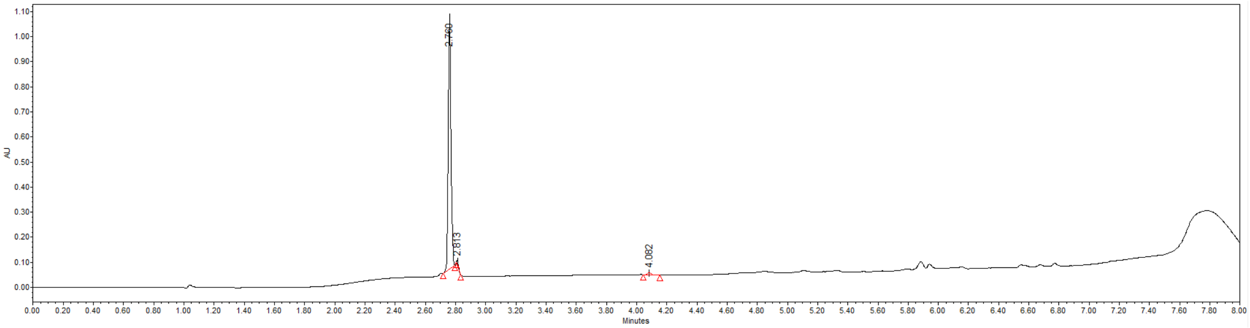
<u>ESI-MS</u>: calculated monoisotopic $[M+H]^+ = 4022.00$

observed monoisotopic $[M+H]^+ = 4021.97$

(*ESI-MS/MS was used to obtain fragment ionized peptides and the measured the mass differences between the resulting fragment ions in MS/MS graphs were used to determine the amino acid sequence.)

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity > 99.0 %



MS-MS fragmentation of peptide: KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG

Elemental composition: C179 H269 N55 O50 S1

Mass: [Mo.] 4021.999 Da.(MH+) 2011.503 Da.(M2H+) 1341.338 Da.(M3H+) [Mass file: AA_MASS.MSS]

a	b	Res:	x	y'	z'	v'	w'
101.108	129.103	1 Lys 34	-	-	-	-	
214.192	242.187 ✓	2 Leu 33	3919.884	3893.904	3877.885	3836.834	3835.838
311.245	339.240 ✓	3 Pro 32	3806.800	3780.820	3764.801	-	-
408.297	436.292 ✓	4 Pro 31	3709.747	3683.767	3667.749	-	-
465.319	493.314 ✓	5 Gly 30	3612.694	3586.715	3570.696	-	-
651.398	679.393 ✓	6 Trp 29	3555.673	3529.693	3513.674	3399.627	3398.632
780.441	808.436 ✓	7 Glu 28	3369.594	3343.614	3327.595	3270.585	3269.590
908.536	936.531	8 Lys 27	3240.551	3214.571	3198.552	3142.490	3141.495
1064.637	1092.632	9 Arg 26	3112.456	3086.476 ✓	3070.457	2986.389	2985.393
1195.677	1223.672	10 Met 25	2956.355	2930.375	2914.356	2855.348	2854.353
1282.709	1310.704	11 Ser 24	2825.314	2799.335	2783.316	2768.316	2767.321
1438.811	1466.805	12 Arg 23	2738.282	2712.303 ✓	2696.284	2612.215	2611.220
1525.843	1553.837	13 Ser 22	2582.181	2556.201	2540.183	2525.183	2524.188
1612.875	1640.870	14 Ser 21	2495.149	2469.169	2453.151	2438.151	2437.156
1669.896	1697.891	15 Gly 20	2408.117	2382.137	2366.119	-	-
1825.997	1853.992	16 Arg 19	2351.096	2325.116	2309.097	2225.028	2224.033
1925.066	1953.060	17 Val 18	2194.995	2169.015	2152.996	2125.960	2138.980
2088.129	2116.124	18 Tyr 17	2095.926	2069.946	2053.928	1962.897	1961.901
2251.192	2279.187	19 Tyr 16	1932.863	1906.883 ✓	1890.864	1799.833	1798.838
2398.261	2426.256	20 Phe 15	1769.800	1743.820 ✓	1727.801	1652.765	1651.770
2512.304	2540.298	21 Asn 14	1622.731	1596.751 ✓	1580.733	1538.722	1537.727
2649.362	2677.357	22 His 13	1508.688	1482.708 ✓	1466.690	1401.663	1400.668
2762.447	2790.441	23 Ile 12	1371.629	1345.650 ✓	1329.631	1288.579	1301.599
2863.494	2891.489	24 Thr 11	1258.545	1232.565 ✓	1216.547	1187.531	1200.552
2977.537	3005.532	25 Asn 10	1157.498	1131.518 ✓	1115.499	1073.488	1072.493
3048.574	3076.569	26 Ala 9	1043.455	1017.475 ✓	1001.456	1002.451	-
3135.606	3163.601	27 Ser 8	972.418	946.438 ✓	930.419	915.419	914.424
3263.665	3291.660	28 Gln 7	885.386	859.406 ✓	843.387	787.361	786.366
3449.744	3477.739	29 Trp 6	757.327	731.347 ✓	715.328	601.281	600.286
3578.787	3606.782	30 Glu 5	571.248	545.268 ✓	529.249	472.239	471.244
3734.888	3762.883	31 Arg 4	442.205	416.225 ✓	400.206	316.138	315.142
3831.941	3859.936	32 Pro 3	286.104	260.124 ✓	244.105	-	-
3918.973	3946.968	33 Ser 2	189.051	163.071	147.053	132.053	131.058
-	-	34 Gly 1	102.019	76.039	60.021	-	-

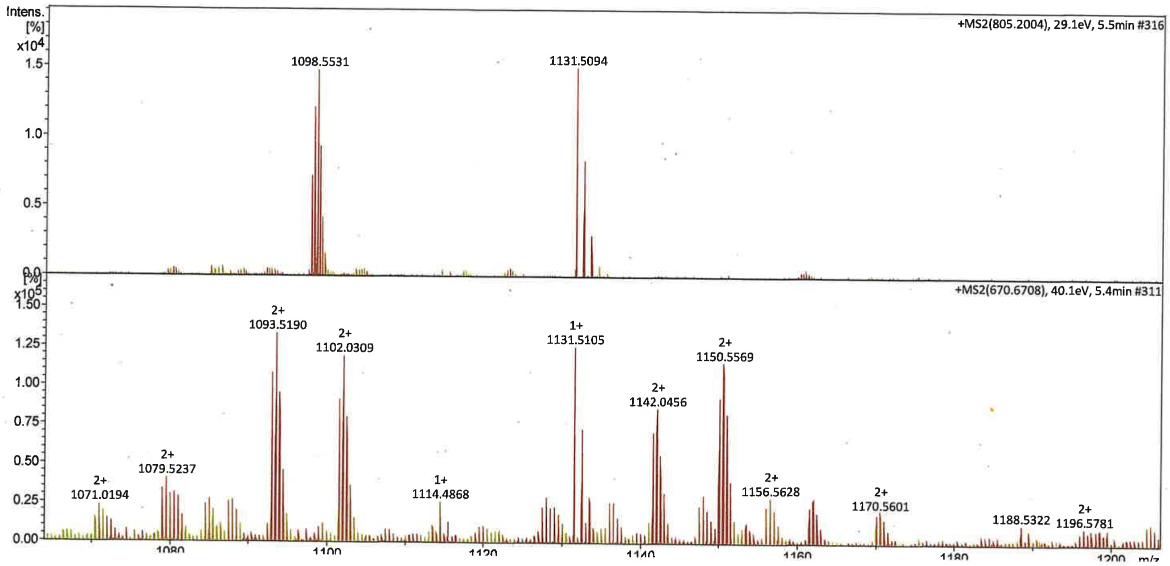
- Related immonium ions:
- 87.055[N] 74.060[T] 60.044[S] 102.055[E] 101.071[Q] 70.065[P] 30.034[G]
- 44.049[A] 72.081[V] 104.053[M] 86.096[I] 86.096[L] 136.076[Y] 120.081[F]
- 101.107[K] 110.071[H] 159.092[W] 129.113[R]

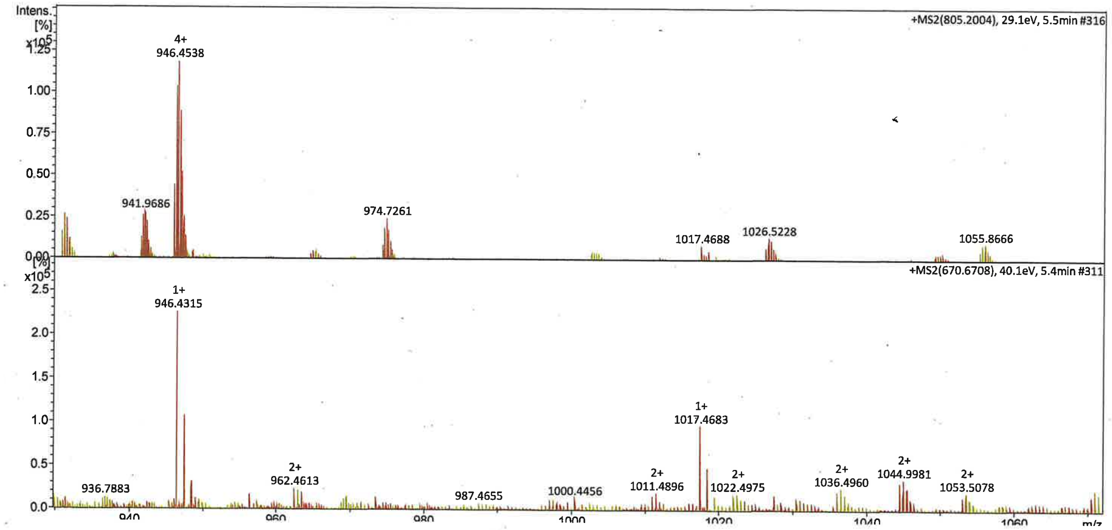
N-terminal: Hydrogen C-terminal: Free acid

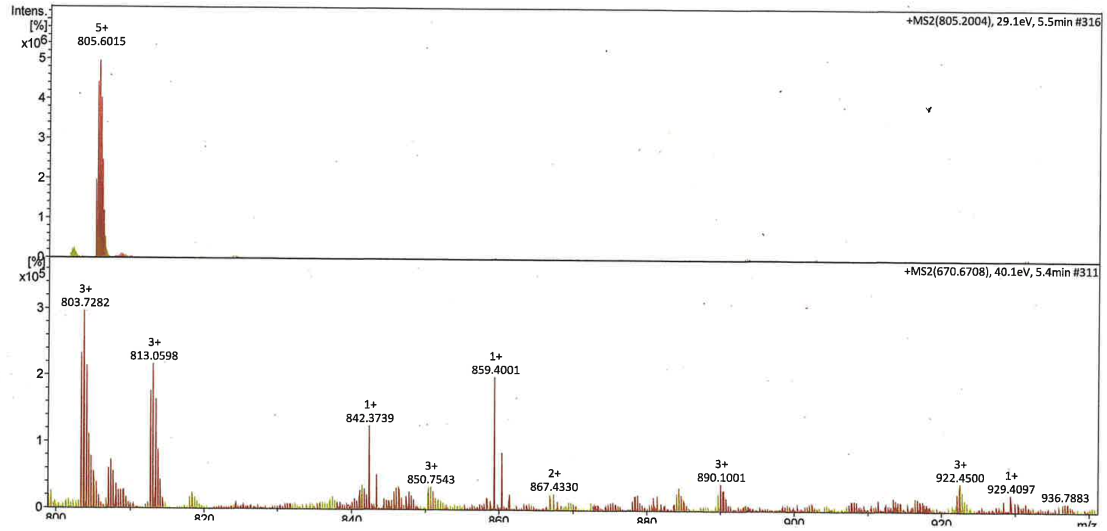
Sorted mass list:

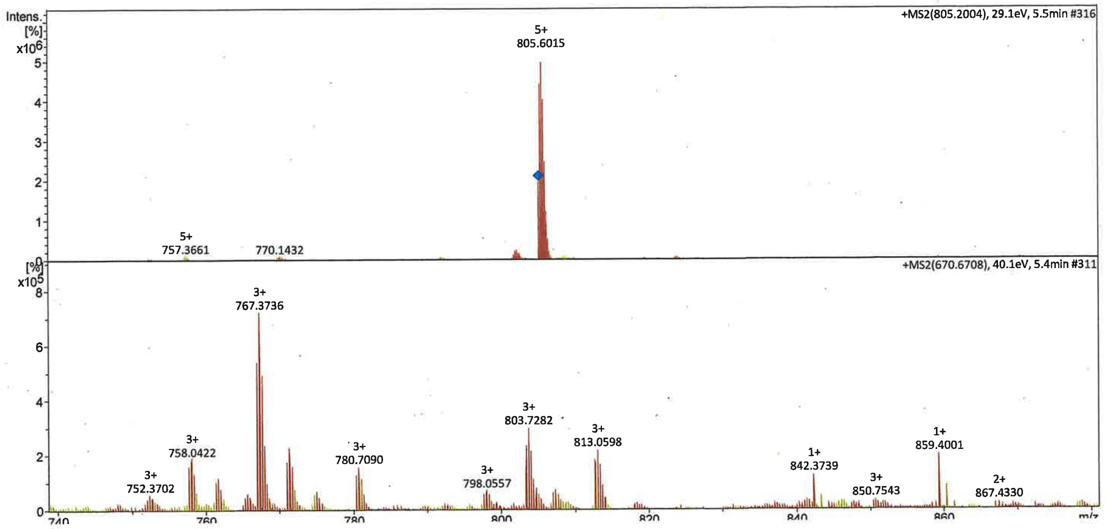
```
Mass Type           1073.488   v' 10  -1825.997    a 16  -2612.215   v' 23  -3399.627   v' 29
101.108    a  1   +1092.632    b  9  -1853.992    b 16  -2649.362    a 22  -3449.744    a 29
129.103    b  1   +1115.499   z' 10  -1890.864   z' 16  -2677.357    b 22  -3477.739    b 29
214.192    a  2   +1131.518   y' 10  -1906.883   y' 16  -2696.284   z' 23  -3513.674   z' 29
242.187    b  2   +1157.498    x 10  -1925.066    a 17  -2712.303   y' 23  -3529.693   y' 29
311.245    a  3   +1187.531   v' 11  -1932.863    x 16  -2738.282    x 23  -3555.673    x 29
339.240    b  3   +1195.677    a 10  -1953.060    b 17  -2762.447    a 23  -3570.696   z' 30
408.297    a  4   +1200.552   w' 11  -1961.901   w' 17  -2767.321   w' 24  -3578.787    a 30
436.292    b  4   +1216.547   z' 11  -1962.897   v' 17  -2768.316   v' 24  -3586.715   y' 30
442.205    x  4   +1223.672    b 10  -2053.928   z' 17  -2783.316   z' 24  -3606.782    b 30
465.319    a  5   +1232.565   y' 11  -2069.946   y' 17  -2790.441    b 23  -3612.694    x 30
471.244   w'  5   +1258.545    x 11  -2088.129    a 18  -2799.335   y' 24  -3667.749   z' 31
```

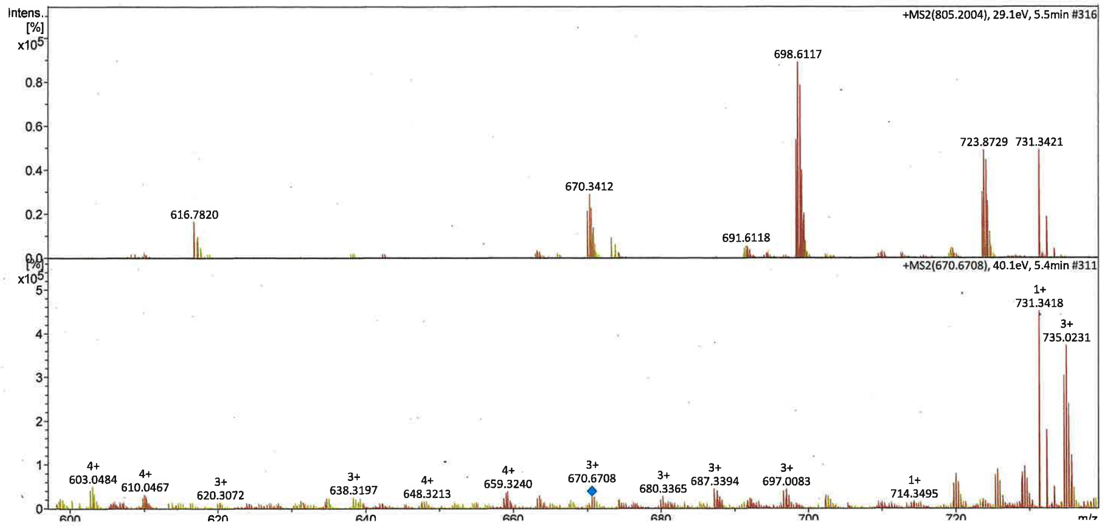
```
472.239    v'  5  +1282.709   a 11  -2095.926   x 17  -2825.314   x 24  -3683.767   y' 31  -
493.314     b  5  +1288.579  v' 12  -2116.124   b 18  -2854.353  w' 25  -3709.747    x 31  -
529.249    z'  5  +1301.599  w' 12  -2125.960  v' 18  -2855.348  v' 25  -3734.888    a 31  -
545.268    y'  5  +1310.704   b 11  -2138.980  w' 18  -2863.494   a 24  -3762.883    b 31  -
571.248     x  5  +1329.631  z' 12  -2152.996  z' 18  -2891.489   b 24  -3764.801   z' 32  -
600.286    w'  6  +1345.650  y' 12  -2169.015  y' 18  -2914.356  z' 25  -3780.820   y' 32  -
601.281    v'  6  +1371.629   x 12  -2194.995   x 18  -2930.375  y' 25  -3806.800    x 32  -
651.398     a  6  +1400.668  w' 13  -2224.033  w' 19  -2956.355   x 25  -3835.838   w' 33  -
679.393     b  6  +1401.663  v' 13  -2225.028  v' 19  -2977.537   a 25  -3836.834   v' 33  -
715.328    z'  6  +1438.811   a 12  -2251.192   a 19  -2985.393  w' 26  -3877.885   z' 33  -
731.347    y'  6  +1466.690  z' 13  -2279.187   b 19  -2986.389  v' 26  -3893.904   y' 33  -
757.327     x  6  +1466.805   b 12  -2309.097  z' 19  -3005.532   b 25  -3919.884    x 33  -
780.441     a  7  +1482.708  y' 13  -2325.116  y' 19  -3048.574   a 26  -
786.366    w'  7  +1508.688   x 13  -2351.096   x 19  -3070.457  z' 26  -
787.361    v'  7  +1525.843   a 13  -2366.119  z' 20  -3076.569   b 26  -
808.436     b  7  +1537.727  w' 14  -2382.137  y' 20  -3086.476  y' 26  -315.142    w'  4  +
843.387    z'  7  +1538.722  v' 14  -2398.261   a 20  -3112.456   x 26  -3831.941    a 32  -
859.406    y'  7  +1553.837   b 13  -2408.117   x 20  -3135.606   a 27  -3859.936    b 32  -
885.386     x  7  +1580.733  z' 14  -2426.256   b 20  -3141.495  w' 27  -286.104     x  3  +
908.536     a  8  +1596.751  y' 14  -2437.156  w' 21  -3142.490  v' 27  -260.124    y'  3  +
914.424    w'  8  +1612.875   a 14  -2438.151  v' 21  -3163.601   b 27  -244.105    z'  3  +
915.419    v'  8  +1622.731   x 14  -2453.151  z' 21  -3198.552  z' 27  -3918.973    a 33  -
930.419    z'  8  +1640.870   b 14  -2469.169  y' 21  -3214.571  y' 27  -3946.968    b 33  -
936.531     b  8  +1651.770  w' 15  -2495.149   x 21  -3240.551   x 27  -189.051     x  2  +
946.438    y'  8  +1652.765  v' 15  -2512.304   a 21  -3263.665   a 28  -163.071    y'  2  +
972.418     x  8  +1669.896   a 15  -2524.188  w' 22  -3269.590  w' 28  -147.053    z'  2  +
1001.456    z'  9  -1697.891   b 15  -2525.183  v' 22  -3270.585  v' 28  -132.053    v'  2  +
1002.451    v'  9  -1727.801  z' 15  -2540.183  z' 22  -3291.660   b 28  -131.058    w'  2  +
1017.475    y'  9  -1743.820  y' 15  -2540.298   b 21  -3327.595  z' 28  -102.019     x  1  +
1043.455     x  9  -1769.800   x 15  -2556.201  y' 22  -3343.614  y' 28  -76.039    y'  1  +
1064.637     a  9  -1798.838  w' 16  -2582.181   x 22  -3369.594   x 28  +
1072.493    w' 10  -1799.833  v' 16  -2611.220  w' 23  -3398.632  w' 29  +
```

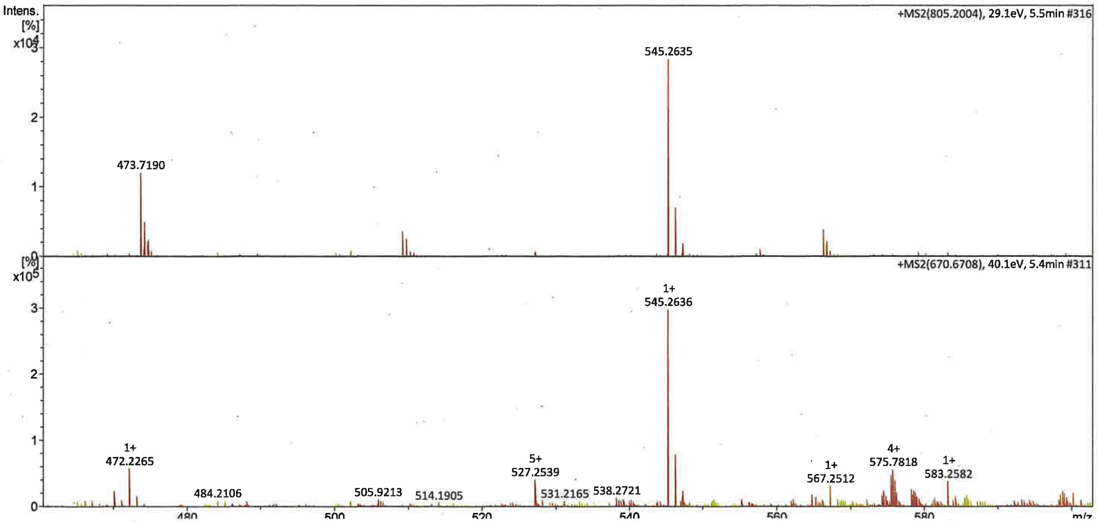


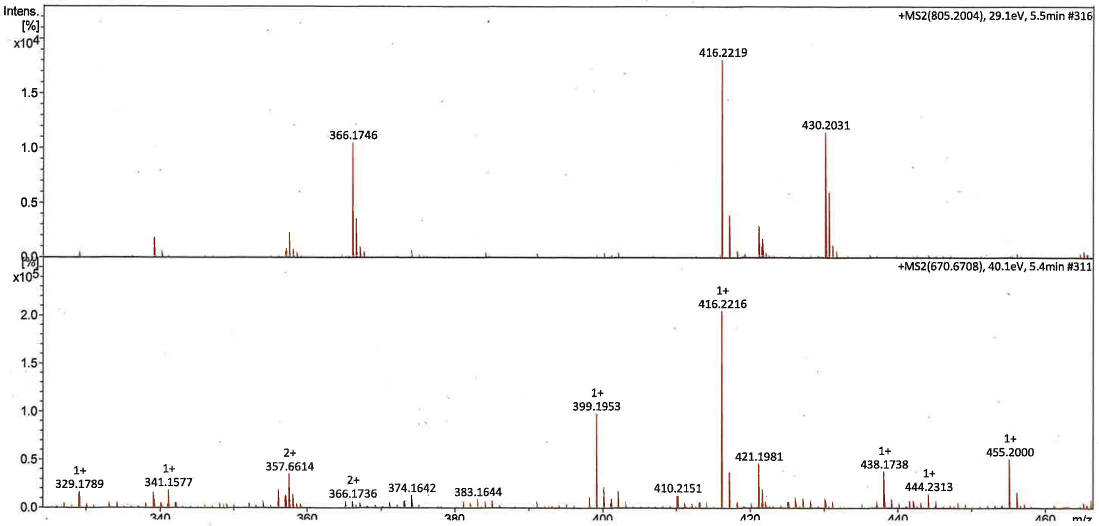

Intens.
[%]
x10^5
1.25
1.00
0.75
0.50
0.25
0.00
+MS2(805.2004), 29.1eV, 5.5min #316
4+
946.4538
941.9686
974.7261
1017.4688
1026.5228
1055.8666
x10^5
2.5
2.0
1.5
1.0
0.5
0.0
+MS2(670.6708), 40.1eV, 5.4min #311
1+
946.4315
936.7883
2+
962.4613
987.4655
1000.4456
2+
1011.4896
1+
1017.4683
2+
1022.4975
2+
1036.4960
2+
1044.9981
2+
1053.5078
940
960
980
1000
1020
1040
1060
m/z

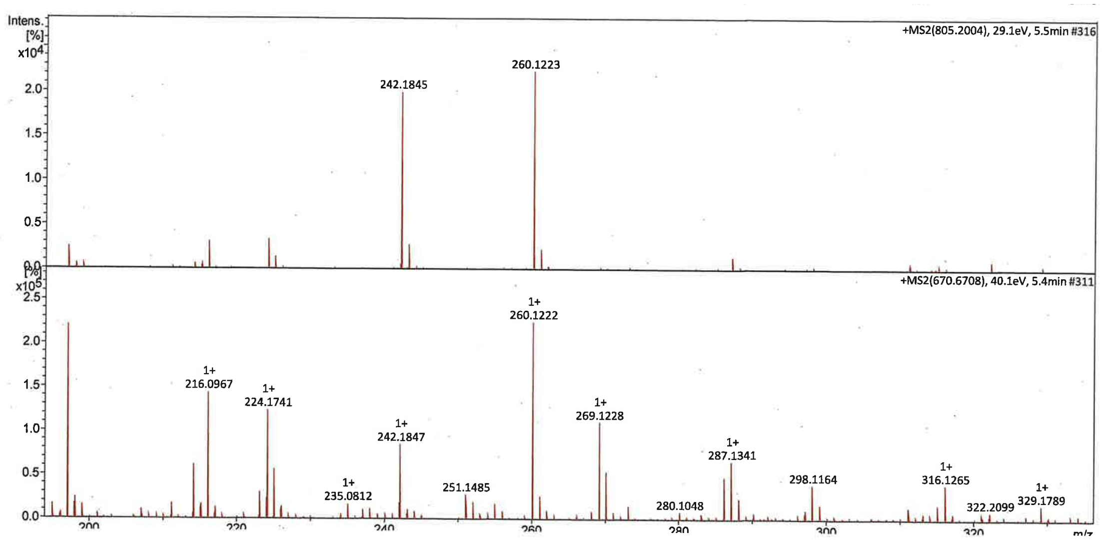

Intens.
[%]
x10^6
5
4
3
2
1
0
+MS2(805.2004), 29.1eV, 5.5min #316
5+
805.6015
[%]
x10^5
3
2
1
0
+MS2(670.6708), 40.1eV, 5.4min #311
3+
803.7282
3+
813.0598
1+
842.3739
3+
850.7543
1+
859.4001
2+
867.4330
3+
890.1001
3+
922.4500
1+
929.4097
936.7883
800
820
840
860
880
900
920
m/z

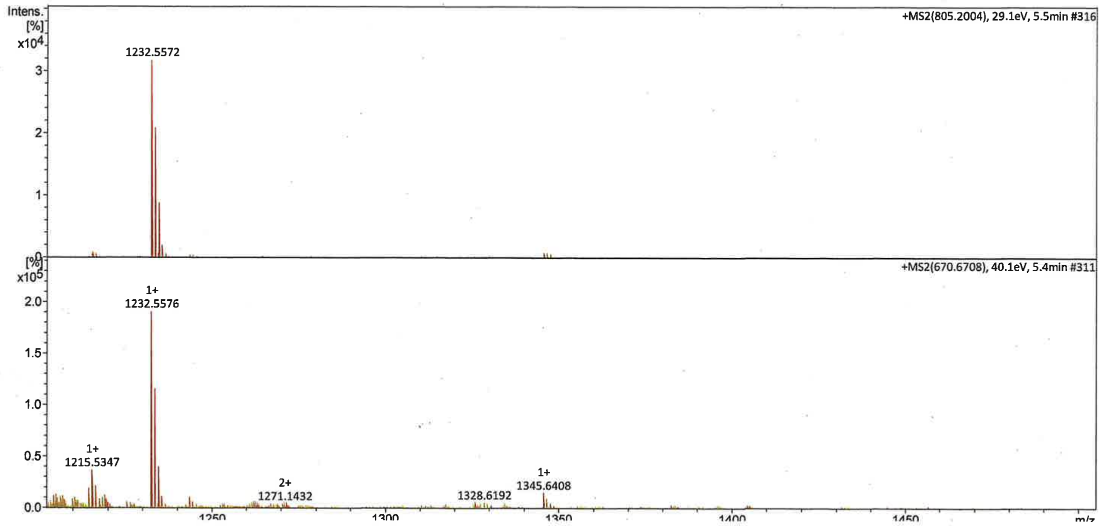

Intens.
[%]
x10^6
5
4
3
2
1
0
+MS2(805.2004), 29.1eV, 5.5min #316
5+
805.6015
5+
757.3661
770.1432
[%]
x10^5
8
6
4
2
0
+MS2(670.6708), 40.1eV, 5.4min #311
3+
767.3736
3+
758.0422
3+
752.3702
3+
780.7090
3+
798.0557
3+
803.7282
3+
813.0598
1+
842.3739
3+
850.7543
1+
859.4001
2+
867.4330
740
760
780
800
820
840
860
m/z


Intens.
[%]
x10^5
0.8
0.6
0.4
0.2
0.0
+MS2(805.2004), 29.1eV, 5.5min #316
616.7820
670.3412
691.6118
698.6117
723.8729
731.3421
[%]
x10^5
5
4
3
2
1
0
+MS2(670.6708), 40.1eV, 5.4min #311
4+ 603.0484
4+ 610.0467
3+ 620.3072
3+ 638.3197
4+ 648.3213
4+ 659.3240
3+ 670.6708
3+ 680.3365
3+ 687.3394
3+ 697.0083
1+ 714.3495
1+ 731.3418
3+ 735.0231
600
620
640
660
680
700
720
m/z


Intens.
[%]
x10^4
3
2
1
0
+MS2(805.2004), 29.1eV, 5.5min #316
473.7190
545.2635
[%]
x10^5
3
2
1
0
+MS2(670.6708), 40.1eV, 5.4min #311
1+ 472.2265
484.2106
505.9213
514.1905
5+ 527.2539
531.2165
538.2721
1+ 545.2636
1+ 567.2512
4+ 575.7818
1+ 583.2582
480
500
520
540
560
580
m/z


Intens.
[%]
x10^4
1.5
1.0
0.5
0.0
+MS2(805.2004), 29.1eV, 5.5min #316
366.1746
416.2219
430.2031
[%]
x10^5
2.0
1.5
1.0
0.5
0.0
+MS2(670.6708), 40.1eV, 5.4min #311
1+ 329.1789
1+ 341.1577
2+ 357.6614
2+ 366.1736
374.1642
383.1644
1+ 399.1953
410.2151
1+ 416.2216
421.1981
1+ 438.1738
1+ 444.2313
1+ 455.2000
340
360
380
400
420
440
460
m/z


Intens.
[%]
x10^4
2.0
1.5
1.0
0.5
0.0
+MS2(805.2004), 29.1eV, 5.5min #316
242.1845
260.1223
[%]
x10^5
2.5
2.0
1.5
1.0
0.5
0.0
+MS2(670.6708), 40.1eV, 5.4min #311
1+ 216.0967
1+ 224.1741
1+ 235.0812
1+ 242.1847
251.1485
1+ 260.1222
1+ 269.1228
280.1048
1+ 287.1341
298.1164
1+ 316.1265
322.2099
1+ 329.1789
200
220
240
260
280
300
320
m/z


Intens.
[%]
x10^4
3
2
1
0
+MS2(805.2004), 29.1eV, 5.5min #316
1232.5572
[%]
x10^5
2.0
1.5
1.0
0.5
0.0
+MS2(670.6708), 40.1eV, 5.4min #311
1+ 1215.5347
1+ 1232.5576
2+ 1271.1432
1328.6192
1+ 1345.6408
1250
1300
1350
1400
1450
m/z

WW_rev: **H₂N-GSPREWQSANTIHNFYYVRGSSRSMRKEWGPPLK-COOH**

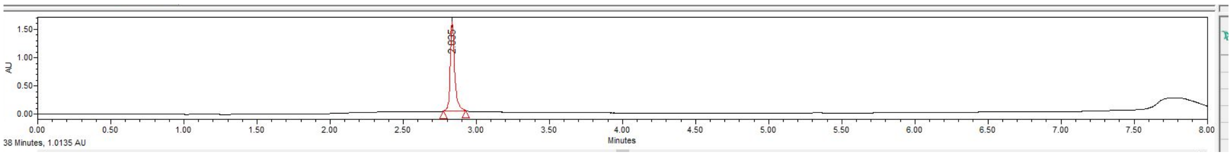
ESI-MS: calculated monoisotopic $[M+H]^+ = 4021.9991$

observed monoisotopic $[M+H]^+ = 4021.9999$

(*ESI-MS/MS was used to obtain fragment ionized peptides and the measured the mass differences between the resulting fragment ions in MS/MS graphs were used to determine the amino acid sequence.)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity > 99.0 %



MS-MS fragmentation of peptide: GSPREWQSANTIHNFYYVRGSSRSMRKEWGPPLK

Elemental composition: C179 H269 N55 O50 S1

Mass: [Mo.] 4021.999 Da.(MH+) 2011.503 Da.(M2H+) 1341.338 Da.(M3H+) [Mass file: AA_MASS.MSS]

a	b	Res:	x	y'	z'	v'	w'
30.034	58.029	1 Gly 34	-	-	-	-	
117.066	145.061	2 Ser 33	3990.957	3964.978	3948.959	3933.959	3932.964
214.119	242.114 ✓	3 Pro 32	3903.925	3877.946	3861.927	-	-
370.220	398.215 ✓ ✓	4 Arg 31	3806.873	3780.893	3764.874	3680.805	3679.810
499.263	527.258 ✓	5 Glu 30	3650.772	3624.792	3608.773	3551.763	3550.768
685.342	713.337 ✓	6 Trp 29	3521.729	3495.749	3479.730	3365.683	3364.688
813.401	841.396 ✓	7 Gln 28	3335.650	3309.670	3293.651	3237.625	3236.630
900.433	928.428 ✓	8 Ser 27	3207.591	3181.611	3165.593	3150.593	3149.598
971.470	999.465 ✓	9 Ala 26	3120.559	3094.579	3078.560	3079.556	-
1085.513	1113.508 ✓	10 Asn 25	3049.522	3023.542	3007.523	2965.513	2964.518
1186.561	1214.555 ✓	11 Thr 24	2935.479	2909.499	2893.480	2864.465	2877.486
1299.645	1327.639 ✓	12 Ile 23	2834.431	2808.451	2792.433	2751.381	2764.401
1436.703	1464.698	13 His 22	2721.347	2695.367 ✓	2679.349	2614.322	2613.327
1550.746	1578.741	14 Asn 21	2584.288	2558.309	2542.290	2500.279	2499.284
1697.815	1725.810	15 Phe 20	2470.245	2444.266	2428.247	2353.211	2352.216
1860.878	1888.873	16 Tyr 19	2323.177	2297.197	2281.178	2190.147	2189.152
2023.941	2051.936	17 Tyr 18	2160.114	2134.134	2118.115	2027.084	2026.089
2123.010	2151.005	18 Val 17	1997.050	1971.071	1955.052	1928.016	1941.036
2279.111	2307.106	19 Arg 16	1897.982	1872.002	1855.983	1771.915	1770.919
2336.132	2364.127	20 Gly 15	1741.881	1715.901	1699.882	-	-
2423.165	2451.159	21 Ser 14	1684.859	1658.880	1642.861	1627.861	1626.866
2510.197	2538.191	22 Ser 13	1597.827	1571.847	1555.829	1540.829	1539.834
2666.298	2694.293	23 Arg 12	1510.795	1484.815	1468.797	1384.728	1383.733
2753.330	2781.325	24 Ser 11	1354.694	1328.714	1312.696	1297.696	1296.701
2884.370	2912.365	25 Met 10	1267.662	1241.682 -	1225.664	1166.655	1165.660
3040.471	3068.466	26 Arg 9	1136.622	1110.642 ✓	1094.623	1010.554	1009.559
3168.566	3196.561	27 Lys 8	980.521	954.541 -	938.522	882.459	881.464
3297.609	3325.604	28 Glu 7	852.426	826.446 ✓	810.427	753.417	752.422
3483.688	3511.683	29 Trp 6	723.383	697.403 ✓	681.384	567.337	566.342
3540.710	3568.705	30 Gly 5	537.304	511.324 ✓	495.305	-	-
3637.762	3665.757	31 Pro 4	480.282	454.302 ✓	438.284	-	-
3734.815	3762.810	32 Pro 3	383.229	357.250 ✓	341.231	-	-
3847.899	3875.894	33 Leu 2	286.177	260.197 ✓	244.178	203.126	202.131
-	-	34 Lys 1	173.093	147.113	131.094	75.031	74.036

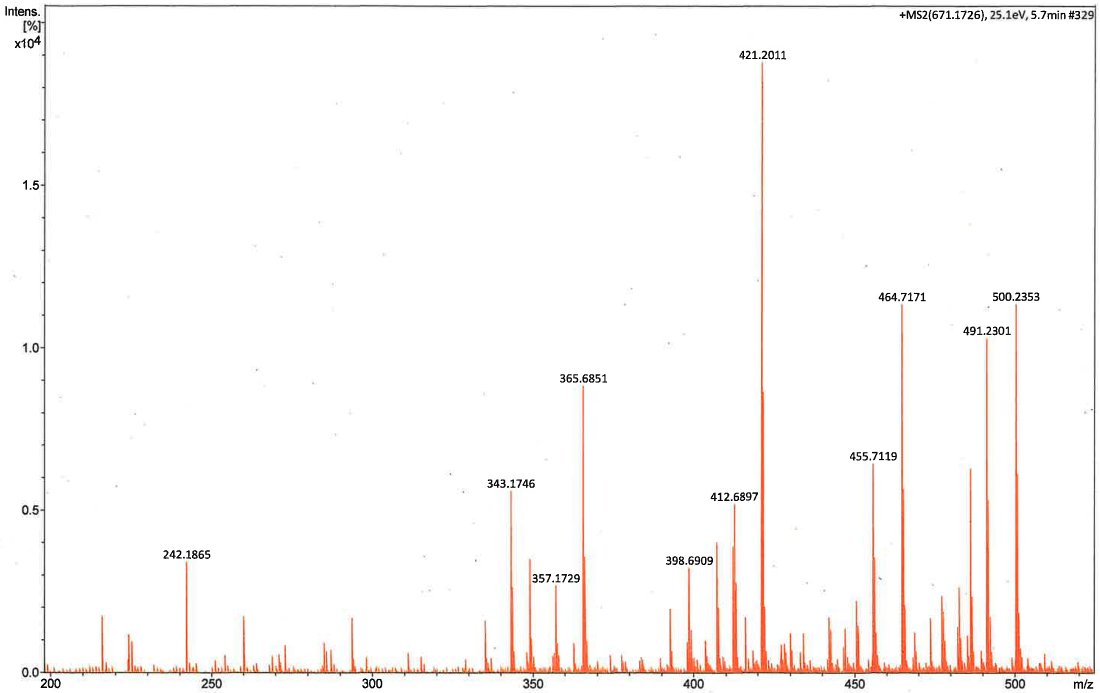
- Related immonium ions:
- 87.055[N] 74.060[T] 60.044[S] 102.055[E] 101.071[Q] 70.065[P] 30.034[G]
- 44.049[A] 72.081[V] 104.053[M] 86.096[I] 86.096[L] 136.076[Y] 120.081[F]
- 101.107[K] 110.071[H] 159.092[W] 129.113[R]

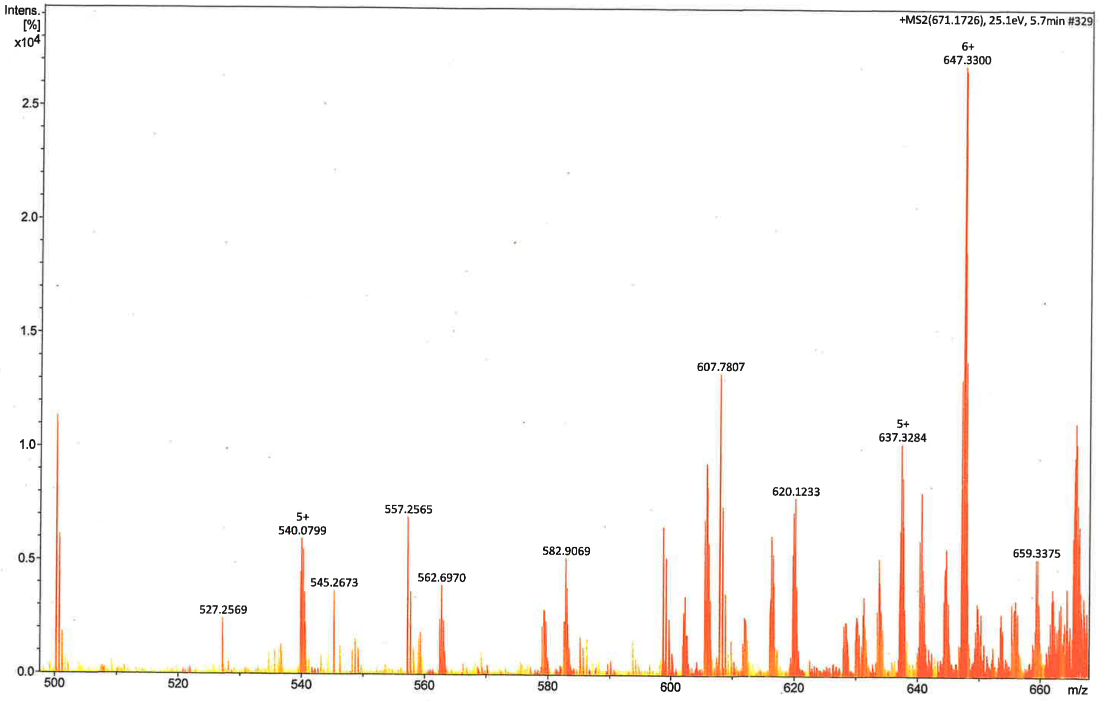
N-terminal: Hydrogen C-terminal: Free acid

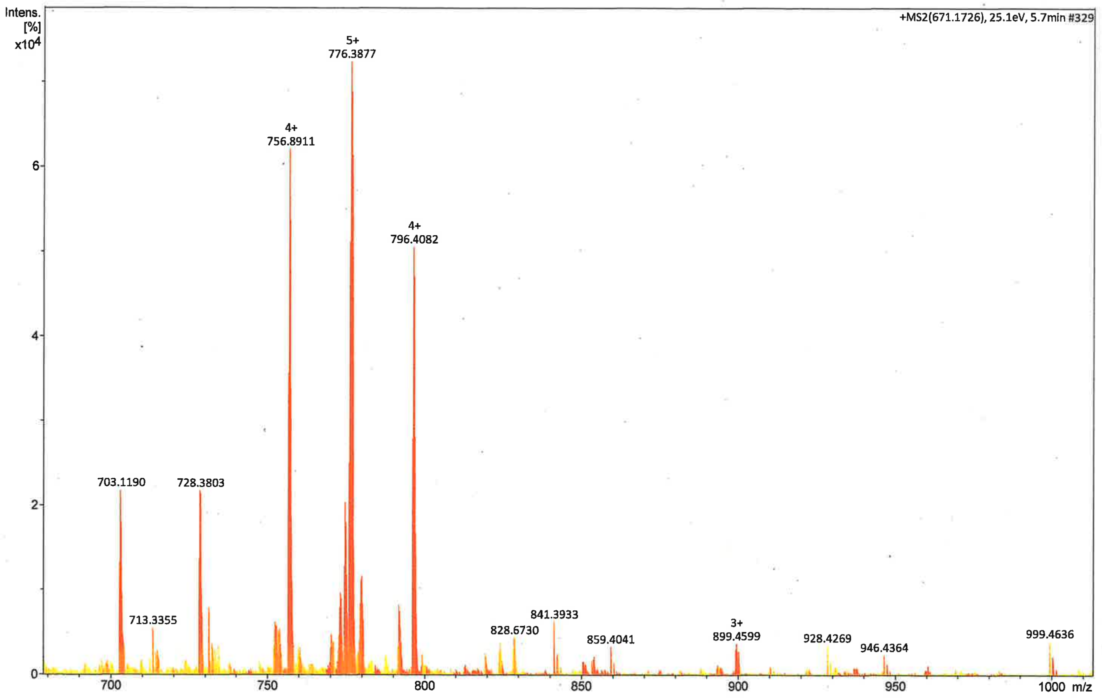
Sorted mass list:

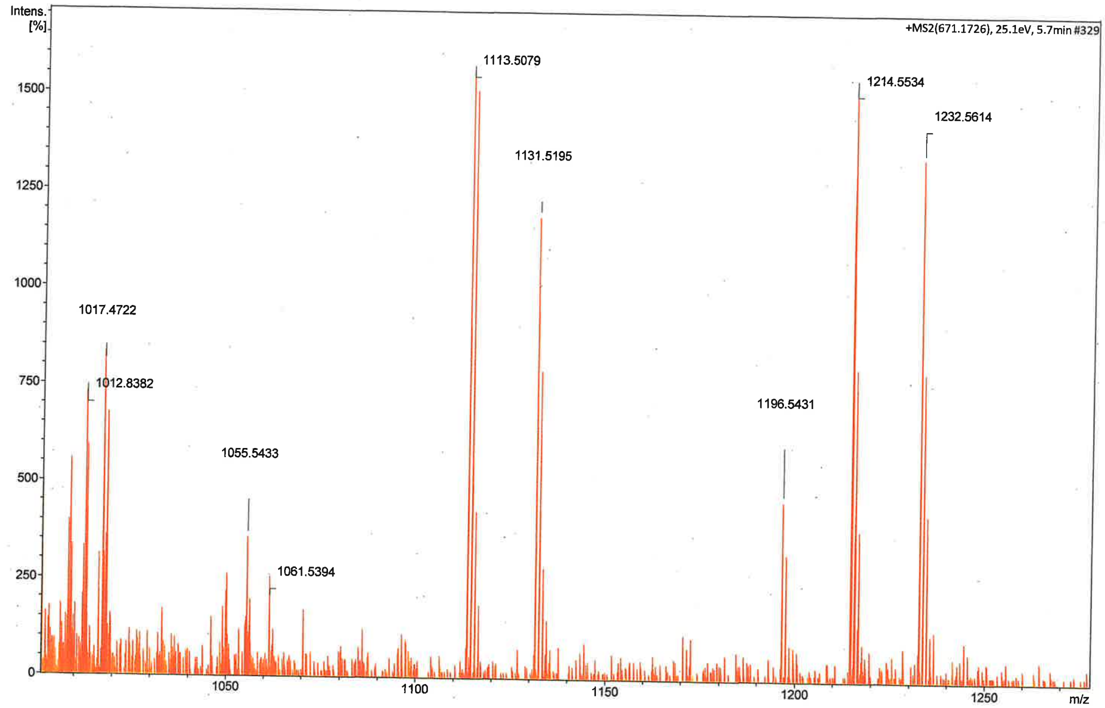
```
Mass Type          1166.655   v' 10  -1941.036   w' 17  -2764.401   w' 23  -3550.768   w' 30  -
30.034   a  1  +  1186.561    a 11  -1955.052   z' 17  -2781.325    b 24  -3551.763   v' 30  -
58.029   b  1  +  1214.555    b 11  -1971.071   y' 17  -2792.433   z' 23  -3568.705    b 30  -
117.066   a  2  +1225.664   z' 10  -1997.050    x 17  -2808.451   y' 23  -3608.773   z' 30  -
145.061   b  2  +1241.682   y' 10  -2023.941    a 17  -2834.431    x 23  -3624.792   y' 30  -
214.119   a  3  +1267.662    x 10  -2026.089   w' 18  -2864.465   v' 24  -3637.762    a 31  -
242.114   b  3  +1296.701   w' 11  -2027.084   v' 18  -2877.486   w' 24  -3650.772    x 30  -
370.220   a  4  +1297.696   v' 11  -2051.936    b 17  -2884.370    a 25  -3679.810   w' 31  -
398.215   b  4  +1299.645    a 12  -2118.115   z' 18  -2893.480   z' 24  -3680.805   v' 31  -
495.305  z'  5  +1312.696   z' 11  -2123.010    a 18  -2909.499   y' 24  -3764.874   z' 31  -
499.263   a  5  +1327.639    b 12  -2134.134   y' 18  -2912.365    b 25  -3780.893   y' 31  -
511.324  y'  5  +1328.714   y' 11  -2151.005    b 18  -2935.479    x 24  -3806.873    x 31  -
```

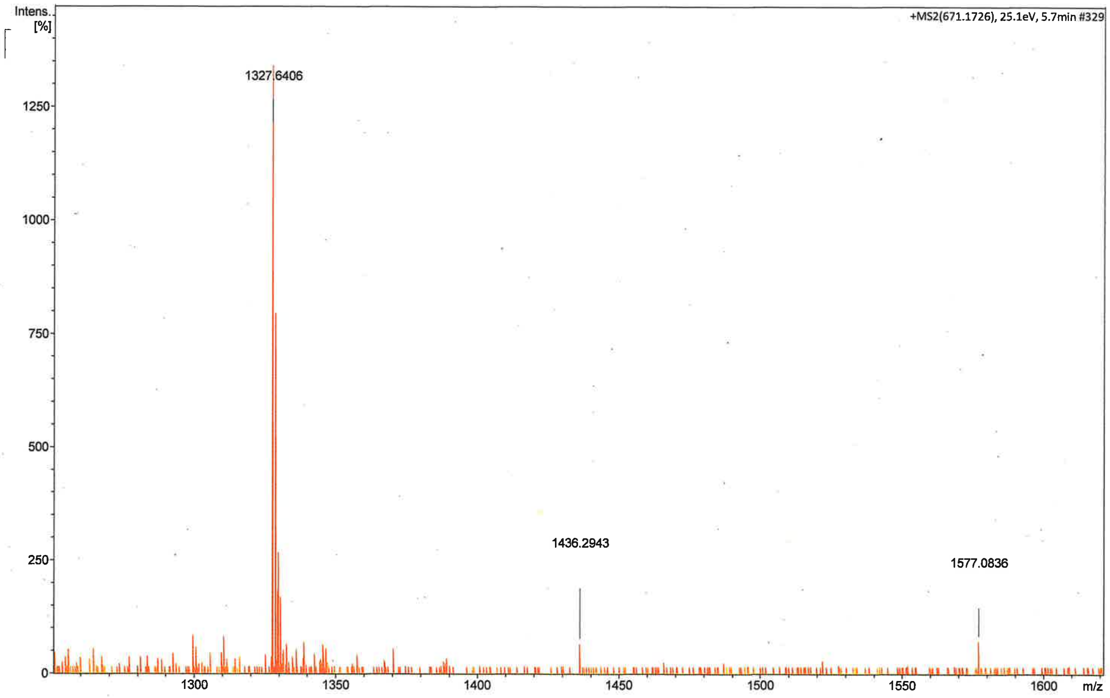
```
527.258    b  5  +1354.694   x 11  -2160.114   x 18  -2964.518   w' 25  -3861.927   z' 32  -
537.304    x  5  +1383.733  w' 12  -2189.152  w' 19  -2965.513   v' 25  -3877.946   y' 32  -
566.342   w'  6  +1384.728  v' 12  -2190.147  v' 19  -3007.523   z' 25  -3903.925    x 32  -
567.337   v'  6  +1436.703   a 13  -2279.111   a 19  -3023.542   y' 25  -3932.964   w' 33  -
681.384   z'  6  +1464.698   b 13  -2281.178  z' 19  -3040.471    a 26  -3933.959   v' 33  -
685.342    a  6  +1468.797  z' 12  -2297.197  y' 19  -3049.522    x 25  -3948.959   z' 33  -
697.403   y'  6  +1484.815  y' 12  -2307.106   b 19  -3068.466    b 26  -3964.978   y' 33  -
713.337    b  6  +1510.795   x 12  -2323.177   x 19  -3078.560   z' 26  -3990.957    x 33  -
723.383    x  6  +1539.834  w' 13  -2336.132   a 20  -3079.556   v' 26  -
752.422   w'  7  +1540.829  v' 13  -2352.216  w' 20  -3094.579   y' 26  -
753.417   v'  7  +1550.746   a 14  -2353.211  v' 20  -3120.559    x 26  -
810.427   z'  7  +1555.829  z' 13  -2364.127   b 20  -3149.598   w' 27  -438.284   z'  4  +
813.401    a  7  +1571.847  y' 13  -2423.165   a 21  -3150.593   v' 27  -3734.815    a 32  -
826.446   y'  7  +1578.741   b 14  -2428.247  z' 20  -3165.593   z' 27  -3762.810    b 32  -
841.396    b  7  +1597.827   x 13  -2444.266  y' 20  -3168.566    a 27  -383.229    x  3  +
852.426    x  7  +1626.866  w' 14  -2451.159   b 21  -3181.611   y' 27  -357.250   y'  3  +
881.464   w'  8  +1627.861  v' 14  -2470.245   x 20  -3196.561    b 27  -341.231   z'  3  +
882.459   v'  8  +1642.861  z' 14  -2499.284  w' 21  -3207.591    x 27  -3847.899    a 33  -
900.433    a  8  +1658.880  y' 14  -2500.279  v' 21  -3236.630   w' 28  -3875.894    b 33  -
928.428    b  8  +1684.859   x 14  -2510.197   a 22  -3237.625   v' 28  -286.177    x  2  +
938.522   z'  8  +1697.815   a 15  -2538.191   b 22  -3293.651   z' 28  -260.197   y'  2  +
954.541   y'  8  +1699.882  z' 15  -2542.290  z' 21  -3297.609    a 28  -244.178   z'  2  +
971.470    a  9  +1715.901  y' 15  -2558.309  y' 21  -3309.670   y' 28  -203.126   v'  2  +
980.521    x  8  +1725.810   b 15  -2584.288   x 21  -3325.604    b 28  -202.131   w'  2  +
999.465    b  9  +1741.881   x 15  -2613.327  w' 22  -3335.650    x 28  -173.093    x  1  +
1009.559   w'  9  -1770.919  w' 16  -2614.322  v' 22  -3364.688   w' 29  -147.113   y'  1  +
1010.554   v'  9  -1771.915  v' 16  -2666.298   a 23  -3365.683   v' 29  -131.094   z'  1  +
1085.513    a 10  -1855.983  z' 16  -2679.349  z' 22  -3479.730   z' 29  -75.031   v'  1  +
1094.623   z'  9  -1860.878   a 16  -2694.293   b 23  -3483.688    a 29  +
1110.642   y'  9  -1872.002  y' 16  -2695.367  y' 22  -3495.749   y' 29  +
1113.508    b 10  -1888.873   b 16  -2721.347   x 22  -3511.683    b 29  +
1136.622    x  9  -1897.982   x 16  -2751.381  v' 23  -3521.729    x 29  +
1165.660   w' 10  -1928.016  v' 17  -2753.330   a 24  -3540.710    a 30  +
```




Intens.
[%]
x10^4
+MS2(671.1726), 25.1eV, 5.7min #329
2.5
2.0
1.5
1.0
0.5
0.0
6+
647.3300
607.7807
5+
637.3284
620.1233
557.2565
5+
540.0799
582.9069
659.3375
545.2673
562.6970
527.2569
500
520
540
560
580
600
620
640
660
m/z


Intens.
[%]
x10^4
+MS2(671.1726), 25.1eV, 5.7min #329
6
4
2
0
5+
776.3877
4+
756.8911
4+
796.4082
703.1190
728.3803
713.3355
841.3933
828.6730
859.4041
3+
899.4599
928.4269
946.4364
999.4636
700
750
800
850
900
950
1000
m/z


Intens.
[%]
+MS2(671.1726), 25.1eV, 5.7min #329
1500
1250
1000
750
500
250
0
1017.4722
1012.8382
1055.5433
1061.5394
1113.5079
1131.5195
1196.5431
1214.5534
1232.5614
1050
1100
1150
1200
1250
m/z


Intens.
[%]
+MS2(671.1726), 25.1eV, 5.7min #329
1250
1000
750
500
250
0
1327.6406
1436.2943
1577.0836
1300
1350
1400
1450
1500
1550
1600
m/z

1PRB: **H_2N-LKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKNEILKAHA-COOH**

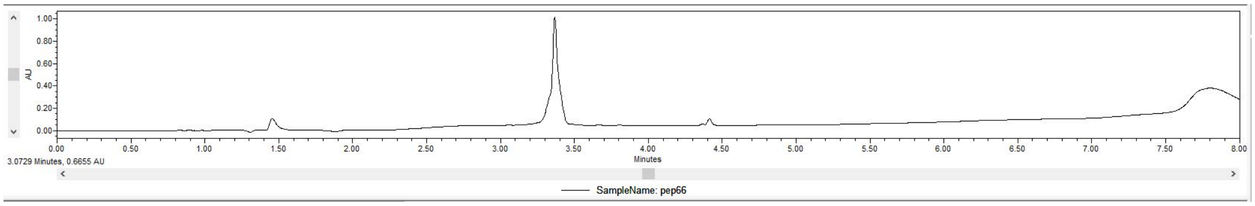
ESI-MS: calculated monoisotopic $[M+H]^+ = 5171.810$

observed monoisotopic $[M+H]^+ = 5171.829$

(*ESI-MS/MS was used to obtain fragment ionized peptides and the measured the mass differences between the resulting fragment ions in MS/MS graphs were used to determine the amino acid sequence.)

UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity > 99.0 %



MS-MS fragmentation of peptide: LKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKNEILKAHA

Elemental composition: C232 H380 N62 O71

Mass: [Mo.] 5171.810 Da.(MH+) 2586.409 Da.(M2H+) 1724.608 Da.(M3H+) [Mass file: AA_MASS.MSS]

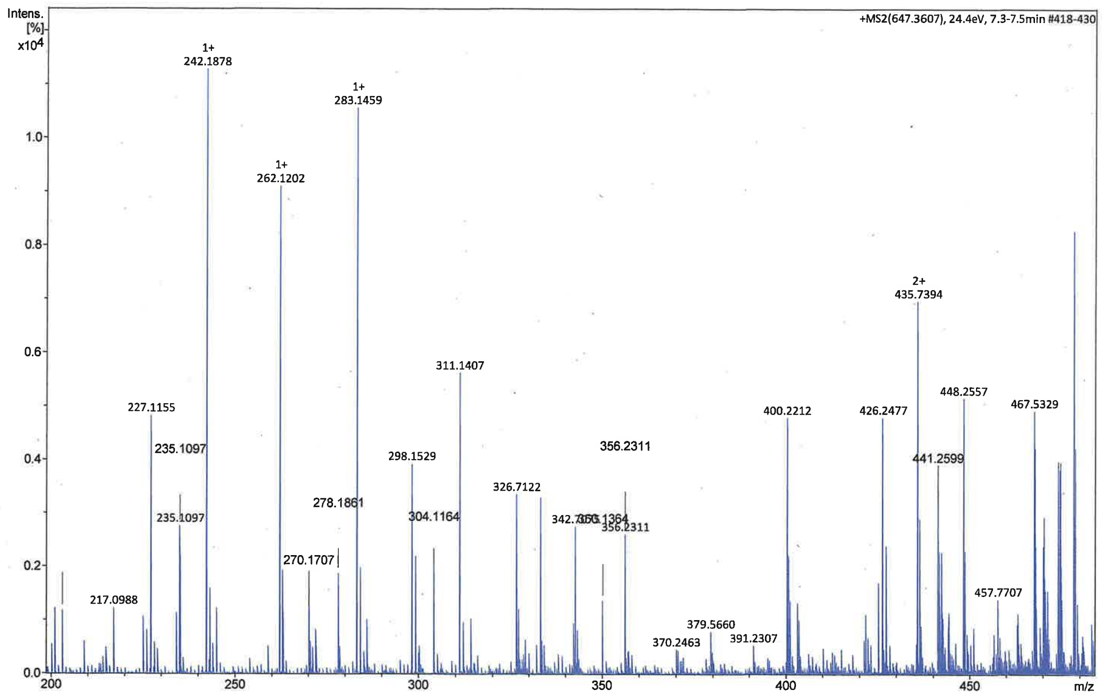
a	b	Res:	x	y'	z'	v'	w'
86.097	114.092	1 Leu 47	-	-	-	-	
214.192	242.187✓	2 Lys 46	5084.706	5058.726	5042.708	4986.645	4985.650
328.235	356.230✓	3 Asn 45	4956.611	4930.631	4914.613	4872.602	4871.607
399.272	427.267	4 Ala 44	4842.568	4816.588	4800.570	4801.565	-
527.367	555.362	5 Lys 43	4771.531	4745.551	4729.533	4673.470	4672.475
656.410	684.404	6 Glu 42	4643.436	4617.456	4601.438	4544.427	4543.432
771.436	799.431	7 Asp 41	4514.394	4488.414	4472.395	4429.400	4428.405
842.474	870.469✓	8 Ala 40	4399.367	4373.387	4357.368	4358.363	-
955.558✓	983.553✓	9 Ile 39	4328.329	4302.350	4286.331	4245.279	4258.300
1026.595	1054.590✓	10 Ala 38	4215.245	4189.266	4173.247	4174.242	-
1155.637	1183.632✓	11 Glu 37	4144.208	4118.228	4102.210	4045.200	4044.204
1268.721	1296.716✓	12 Leu 36	4015.166	3989.186	3973.167	3932.115	3931.120
1396.816	1424.811✓	13 Lys 35	3902.082	3876.102	3860.083	3804.020	3803.025
1524.911	1552.906	14 Lys 34	3773.987	3748.007	3731.988	3675.926	3674.930
1595.948	1623.943	15 Ala 33	3645.892	3619.912	3603.893	3604.888	-
1652.970	1680.965	16 Gly 32	3574.855	3548.875	3532.856	-	-
1766.054	1794.049	17 Ile 31	3517.833	3491.853	3475.835	3434.783	3447.803
1867.102	1895.097	18 Thr 30	3404.749	3378.769	3362.751	3333.735	3346.756
1954.134	1982.129	19 Ser 29	3303.701	3277.722✓	3261.703	3246.703	3245.708
2069.161	2097.156✓	20 Asp 28	3216.669	3190.690	3174.671	3131.676	3130.681
2216.229✓	2244.224✓	21 Phe 27	3101.642	3075.663✓	3059.644	2984.608	2983.613
2379.292✓	2407.287✓	22 Tyr 26	2954.574	2928.594✓	2912.575	2821.544	2820.549
2526.361	2554.356✓	23 Phe 25	2791.511	2765.531✓	2749.512	2674.476	2673.481
2640.404	2668.399	24 Asn 24	2644.442	2618.462✓	2602.444	2560.433	2559.438
2711.441	2739.436	25 Ala 23	2530.399	2504.420✓	2488.401	2489.396	-
2824.525	2852.520	26 Ile 22	2459.362	2433.382✓	2417.364	2376.312	2389.332
2938.568	2966.563	27 Asn 21	2346.278	2320.298✓	2304.280	2262.269	2261.274
3066.663	3094.658	28 Lys 20	2232.235	2206.255	2190.237	2134.174	2133.179
3137.700	3165.695	29 Ala 19	2104.140	2078.160	2062.142	2063.137	-
3265.795	3293.790	30 Lys 18	2033.103	2007.123	1991.105	1935.042	1934.047
3366.843	3394.837	31 Thr 17	1905.008	1879.028✓	1863.010	1833.994	1847.015
3465.911	3493.906	32 Val 16	1803.960	1777.981	1761.962	1734.926	1747.946
3594.954	3622.948	33 Glu 15	1704.892	1678.912✓	1662.894	1605.883	1604.888
3723.996	3751.991	34 Glu 14	1575.849	1549.870	1533.851	1476.841	1475.845
3823.065	3851.059	35 Val 13	1446.807	1420.827	1404.808	1377.772	1390.793
3937.108	3965.102	36 Asn 12	1347.738	1321.759✓	1305.740	1263.729	1262.734
4008.145	4036.140	37 Ala 11	1233.696	1207.716	1191.697	1192.692	-
4121.229	4149.224	38 Leu 10	1162.658	1136.679	1120.660	1079.608	1078.613
4249.324	4277.319	39 Lys 9	1049.574	1023.595✓	1007.576	951.513	950.518
4363.367	4391.361	40 Asn 8	921.479	895.500✓	879.481	837.470	836.475
4492.409	4520.404	41 Glu 7	807.436	781.457	765.438	708.428	707.432
4605.493	4633.488	42 Ile 6	678.394	652.414	636.395	595.344	608.364
4718.577	4746.572	43 Leu 5	565.310	539.330✓	523.311	482.260	481.264
4846.672	4874.667	44 Lys 4	452.226	426.246✓	410.227	354.165	353.169
4917.709	4945.704	45 Ala 3	324.131	298.151✓	282.132	283.128	-
5054.768	5082.763	46 His 2	253.094	227.114✓	211.095	146.069	145.073
-	-	47 Ala 1	116.035	90.055	74.036	75.031	-

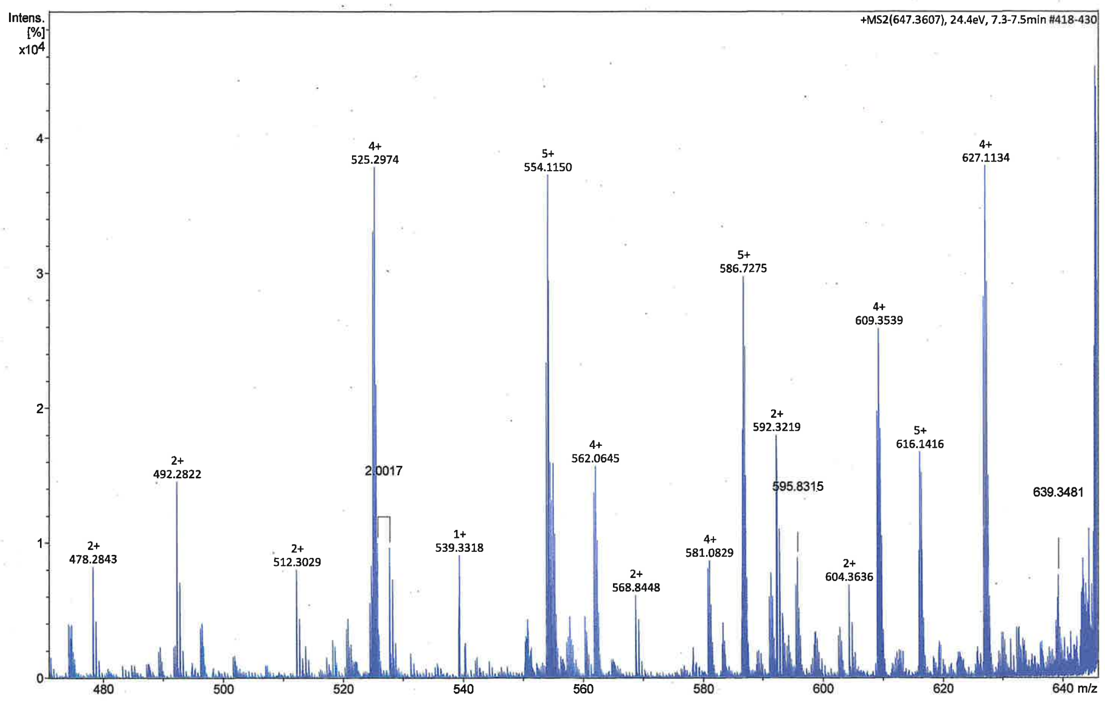
- Related immonium ions:
- 88.039[D] 87.055[N] 74.060[T] 60.044[S] 102.055[E] 30.034[G] 44.049[A]
- 72.081[V] 86.096[I] 86.096[L] 136.076[Y] 120.081[F] 101.107[K] 110.071[H]

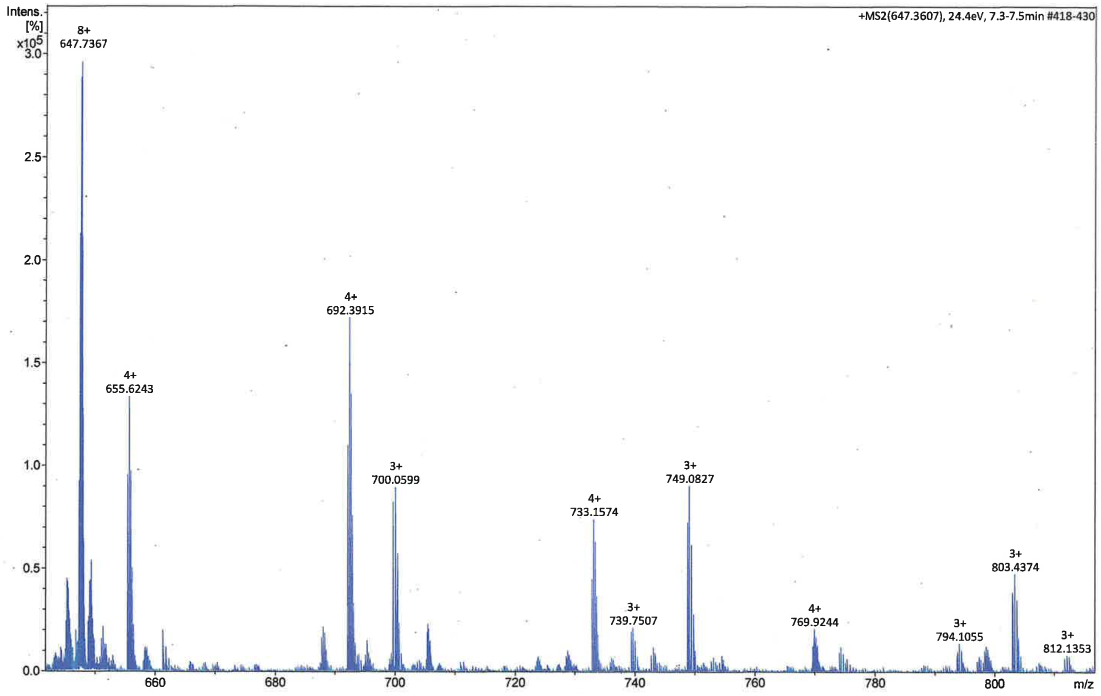
N-terminal: Hydrogen C-terminal: Free acid

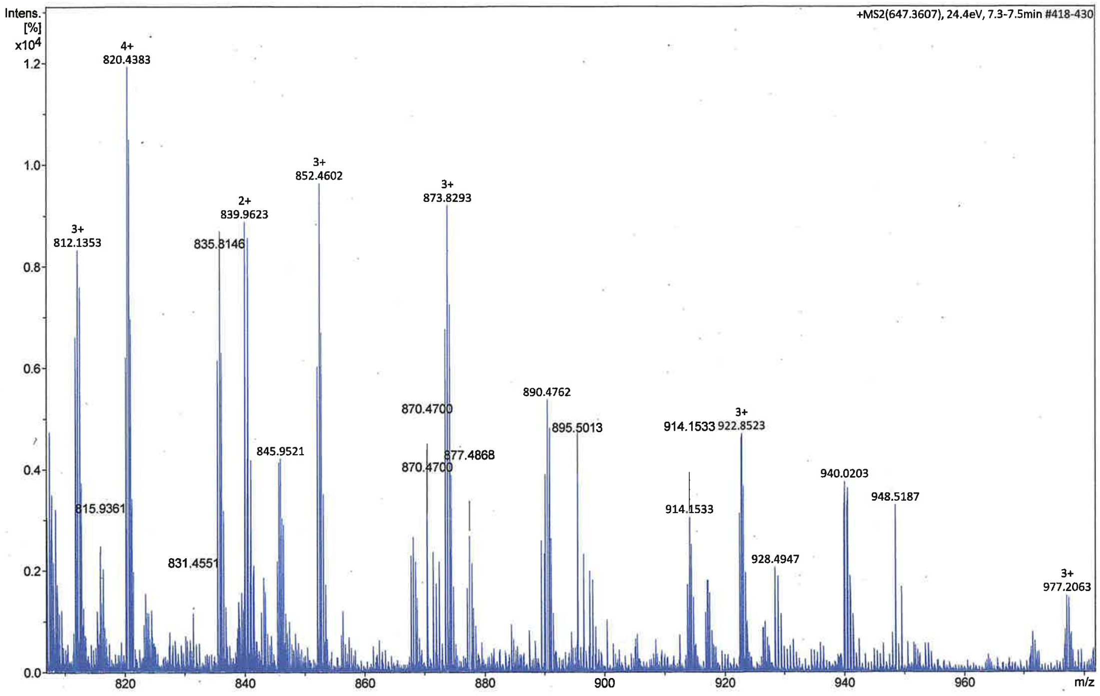
Sorted mass list:

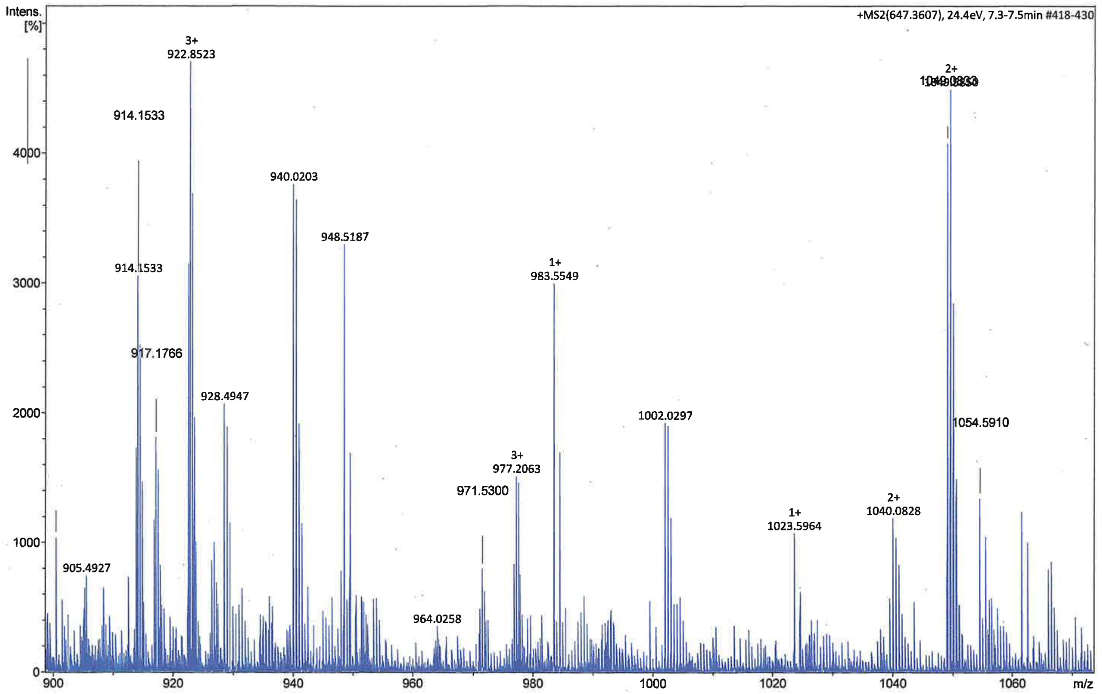
```
Mass Type          2379.292    a 22 -3434.783    v' 31 -4842.568    x 44 -1078.613   w' 10 -
86.097     a  1  + 2389.332   w' 22 -3447.803   w' 31 -4871.607   w' 45 -4249.324    a 39 -
114.092     b  1 +2407.287    b 22 -3475.835   z' 31 -4872.602   v' 45 -4277.319    b 39 -
214.192     a  2 +2417.364   z' 22 -3491.853   y' 31 -4914.613   z' 45 -1049.574    x  9 -
242.187     b  2 +2433.382   y' 22 -3517.833    x 31 -4930.631   y' 45 -1023.595   y'  9 -
328.235     a  3 +2459.362    x 22 -3532.856   z' 32 -4956.611    x 45 -1007.576   z'  9 -
356.230     b  3 +2488.401   z' 23 -3548.875   y' 32 -4985.650   w' 46 -951.513   v'  9 +
399.272     a  4 +2489.396   v' 23 -3574.855    x 32 -4986.645   v' 46 -950.518   w'  9 +
427.267     b  4 +2504.420   y' 23 -3603.893   z' 33 -5042.708   z' 46 -4363.367    a 40 -
527.367     a  5 +2526.361    a 23 -3604.888   v' 33 -5058.726   y' 46 -4391.361    b 40 -
555.362     b  5 +2530.399    x 23 -3619.912   y' 33 -5084.706    x 46 -921.479    x  8 +
656.410     a  6 +2554.356    b 23 -3645.892    x 33 -                  895.500   y'  8 +
684.404     b  6 +2559.438   w' 24 -3674.930   w' 34 -                  879.481   z'  8 +
771.436     a  7 +2560.433   v' 24 -3675.926   v' 34 -                  837.470   v'  8 +
799.431     b  7 +2602.444   z' 24 -3731.988   z' 34 -1833.994   v' 17 -836.475   w'  8 +
842.474     a  8 +2618.462   y' 24 -3748.007   y' 34 -1847.015   w' 17 -4492.409    a 41 -
870.469     b  8 +2640.404    a 24 -3773.987    x 34 -3465.911    a 32 -4520.404    b 41 -
955.558     a  9 +2644.442    x 24 -3803.025   w' 35 -3493.906    b 32 -807.436    x  7 +
983.553     b  9 +2668.399    b 24 -3804.020   v' 35 -1803.960    x 16 -781.457   y'  7 +
1026.595    a 10 -2673.481   w' 25 -3860.083   z' 35 -1777.981   y' 16 -765.438   z'  7 +
1054.590    b 10 -2674.476   v' 25 -3876.102   y' 35 -1761.962   z' 16 -708.428   v'  7 +
1155.637    a 11 -2711.441    a 25 -3902.082    x 35 -1734.926   v' 16 -707.432   w'  7 +
1183.632    b 11 -2739.436    b 25 -3931.120   w' 36 -1747.946   w' 16 -4605.493    a 42 -
1268.721    a 12 -2749.512   z' 25 -3932.115   v' 36 -3594.954    a 33 -4633.488    b 42 -
1296.716    b 12 -2765.531   y' 25 -3973.167   z' 36 -3622.948    b 33 -678.394    x  6 +
1396.816    a 13 -2791.511    x 25 -3989.186   y' 36 -1704.892    x 15 -652.414   y'  6 +
1424.811    b 13 -2820.549   w' 26 -4015.166    x 36 -1678.912   y' 15 -636.395   z'  6 +
1524.911    a 14 -2821.544   v' 26 -4044.204   w' 37 -1662.894   z' 15 -595.344   v'  6 +
1552.906    b 14 -2824.525    a 26 -4045.200   v' 37 -1605.883   v' 15 -608.364   w'  6 +
1595.948    a 15 -2852.520    b 26 -4102.210   z' 37 -1604.888   w' 15 -4718.577    a 43 -
1623.943    b 15 -2912.575   z' 26 -4118.228   y' 37 -3723.996    a 34 -4746.572    b 43 -
1652.970    a 16 -2928.594   y' 26 -4144.208    x 37 -3751.991    b 34 -565.310    x  5 +
1680.965    b 16 -2938.568    a 27 -4173.247   z' 38 -1575.849    x 14 -539.330   y'  5 +
1766.054    a 17 -2954.574    x 26 -4174.242   v' 38 -1549.870   y' 14 -523.311   z'  5 +
1794.049    b 17 -2966.563    b 27 -4189.266   y' 38 -1533.851   z' 14 -482.260   v'  5 +
1867.102    a 18 -2983.613   w' 27 -4215.245    x 38 -1476.841   v' 14 -481.264   w'  5 +
1895.097    b 18 -2984.608   v' 27 -4245.279   v' 39 -1475.845   w' 14 -4846.672    a 44 -
1934.047   w' 18 -3059.644   z' 27 -4258.300   w' 39 -3823.065    a 35 -4874.667    b 44 -
1935.042   v' 18 -3066.663    a 28 -4286.331   z' 39 -3851.059    b 35 -452.226    x  4 +
1954.134    a 19 -3075.663   y' 27 -4302.350   y' 39 -1446.807    x 13 -426.246   y'  4 +
1982.129    b 19 -3094.658    b 28 -4328.329    x 39 -1420.827   y' 13 -410.227   z'  4 +
1991.105   z' 18 -3101.642    x 27 -4357.368   z' 40 -1404.808   z' 13 -354.165   v'  4 +
2007.123   y' 18 -3130.681   w' 28 -4358.363   v' 40 -1377.772   v' 13 -353.169   w'  4 +
2033.103    x 18 -3131.676   v' 28 -4373.387   y' 40 -1390.793   w' 13 -4917.709    a 45 -
2062.142   z' 19 -3137.700    a 29 -4399.367    x 40 -3937.108    a 36 -4945.704    b 45 -
2063.137   v' 19 -3165.695    b 29 -4428.405   w' 41 -3965.102    b 36 -324.131    x  3 +
2069.161    a 20 -3174.671   z' 28 -4429.400   v' 41 -1347.738    x 12 -298.151   y'  3 +
2078.160   y' 19 -3190.690   y' 28 -4472.395   z' 41 -1321.759   y' 12 -282.132   z'  3 +
2097.156    b 20 -3216.669    x 28 -4488.414   y' 41 -1305.740   z' 12 -283.128   v'  3 +
2104.140    x 19 -3245.708   w' 29 -4514.394    x 41 -1263.729   v' 12 -5054.768    a 46 -
2133.179   w' 20 -3246.703   v' 29 -4543.432   w' 42 -1262.734   w' 12 -5082.763    b 46 -
2134.174   v' 20 -3261.703   z' 29 -4544.427   v' 42 -4008.145    a 37 -253.094    x  2 +
2190.237   z' 20 -3265.795    a 30 -4601.438   z' 42 -4036.140    b 37 -227.114   y'  2 +
2206.255   y' 20 -3277.722   y' 29 -4617.456   y' 42 -1233.696    x 11 -211.095   z'  2 +
2216.229    a 21 -3293.790    b 30 -4643.436    x 42 -1207.716   y' 11 -146.069   v'  2 +
2232.235    x 20 -3303.701    x 29 -4672.475   w' 43 -1191.697   z' 11 -145.073   w'  2 +
2244.224    b 21 -3333.735   v' 30 -4673.470   v' 43 -1192.692   v' 11 -116.035    x  1 +
2261.274   w' 21 -3346.756   w' 30 -4729.533   z' 43 -4121.229    a 38 -90.055    y'  1 +
2262.269   v' 21 -3362.751   z' 30 -4745.551   y' 43 -4149.224    b 38 -74.036    z'  1 +
2304.280   z' 21 -3366.843    a 31 -4771.531    x 43 -1162.658    x 10 +
2320.298   y' 21 -3378.769   y' 30 -4800.570   z' 44 -1136.679   y' 10 +
2346.278    x 21 -3394.837    b 31 -4801.565   v' 44 -1120.660   z' 10 +
2376.312   v' 22 -3404.749    x 30 -4816.588   y' 44 -1079.608   v' 10 +
```

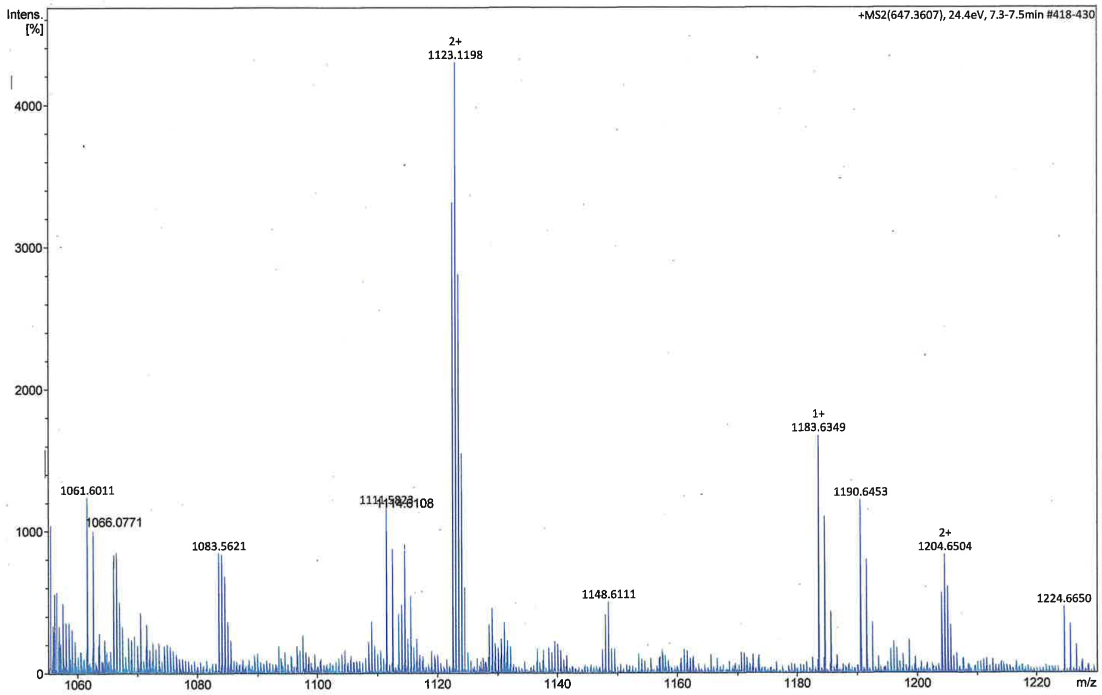

Intens.
[%]
x10^4
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
1.0
0.8
0.6
0.4
0.2
0.0
200
250
300
350
400
450
m/z
1+
242.1878
1+
283.1459
1+
262.1202
2+
435.7394
311.1407
448.2557
227.1155
400.2212
426.2477
467.5329
235.1097
356.2311
298.1529
441.2599
326.7122
278.1861
304.1164
235.1097
342.7[illegible]
350.1364
356.2311
270.1707
457.7707
217.0988
379.5660
370.2463
391.2307

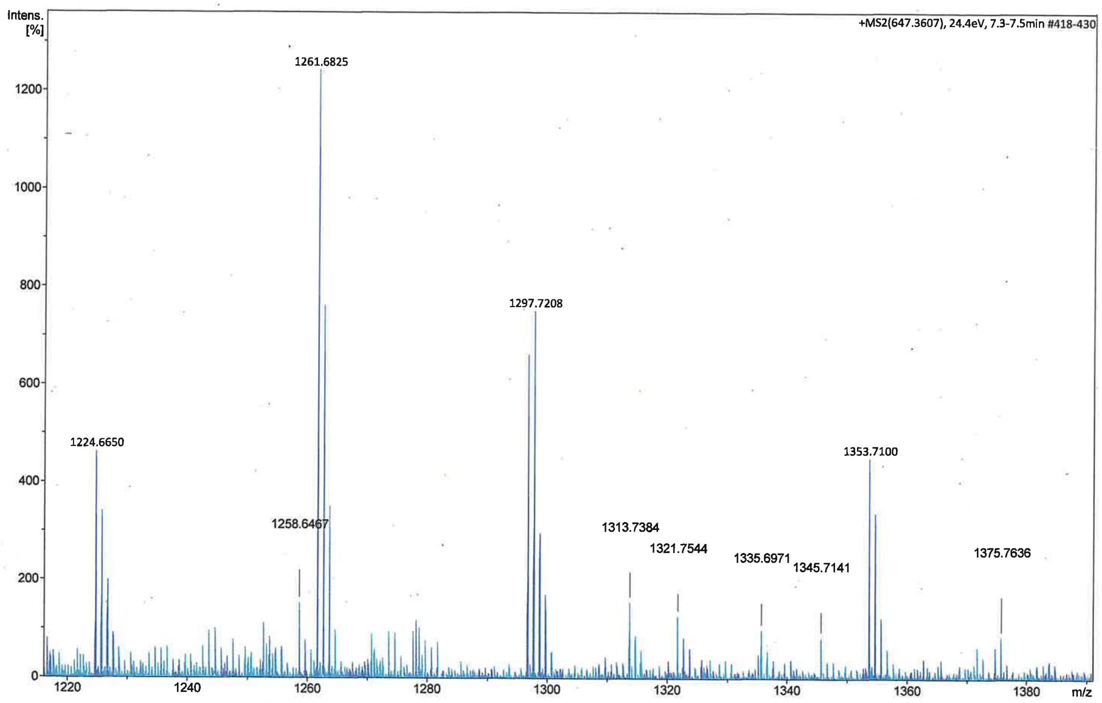

Intens.
[%]
x10^4
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
4
3
2
1
0
480
500
520
540
560
580
600
620
640 m/z
4+
525.2974
5+
554.1150
4+
627.1134
5+
586.7275
4+
609.3539
2+
592.3219
5+
616.1416
4+
562.0645
2+
492.2822
2.0017
595.8315
639.3481
1+
539.3318
2+
478.2843
2+
512.3029
4+
581.0829
2+
604.3636
2+
568.8448

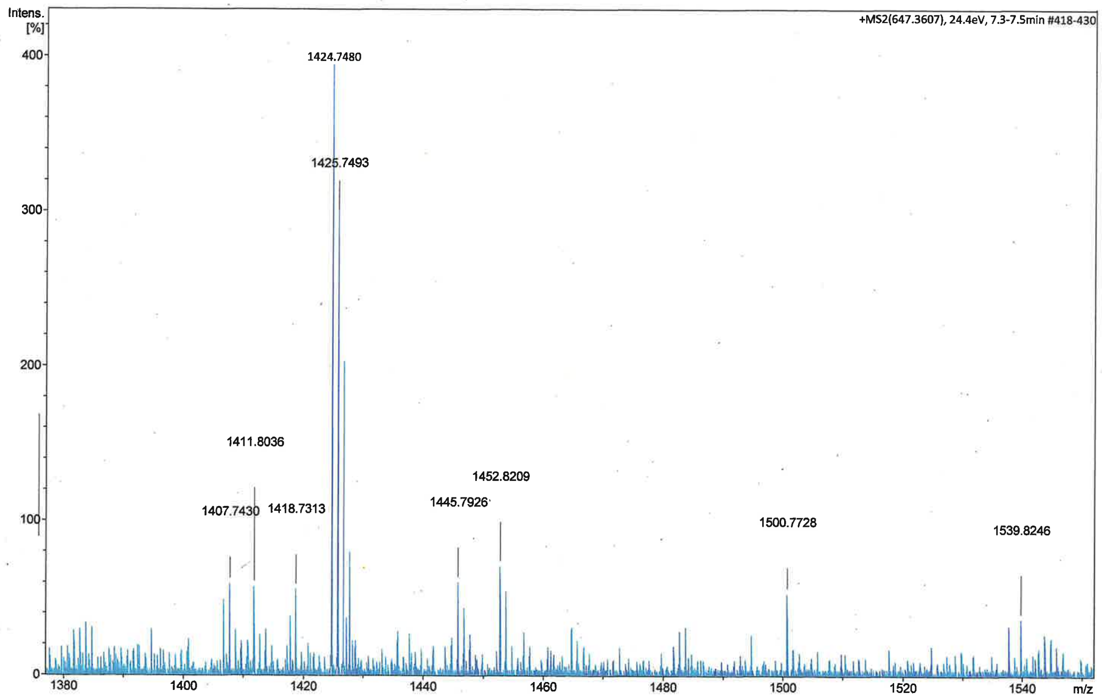

Intens.
[%]
x10^5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
8+
647.7367
4+
655.6243
4+
692.3915
3+
700.0599
4+
733.1574
3+
739.7507
3+
749.0827
4+
769.9244
3+
794.1055
3+
803.4374
3+
812.1353
660
680
700
720
740
760
780
800
m/z


Intens.
[%]
x10^4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
4+
820.4383
3+
812.1353
815.9361
831.4551
835.8146
2+
839.9623
845.9521
3+
852.4602
870.4700
870.4700
3+
873.8293
877.4868
890.4762
895.5013
914.1533
914.1533
3+
922.8523
928.4947
940.0203
948.5187
3+
977.2063
820
840
860
880
900
920
940
960
m/z


Intens.
[%]
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
3+
922.8523
914.1533
914.1533
917.1766
928.4947
940.0203
948.5187
1+
983.5549
1002.0297
3+
977.2063
971.5300
964.0258
905.4927
1+
1023.5964
2+
1040.0828
2+
1049.0833
1054.5910
0
1000
2000
3000
4000
900
920
940
960
980
1000
1020
1040
1060
m/z


Intens.
[%]
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
2+
1123.1198
1061.6011
1066.0771
1083.5621
1111.5823
1114.6108
1148.6111
1+
1183.6349
1190.6453
2+
1204.6504
1224.6650
0
1000
2000
3000
4000
1060
1080
1100
1120
1140
1160
1180
1200
1220
m/z


Intens.
[%]
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
1261.6825
1297.7208
1224.6650
1353.7100
1258.6467
1313.7384
1321.7544
1335.6971
1345.7141
1375.7636
1200
1000
800
600
400
200
0
1220
1240
1260
1280
1300
1320
1340
1360
1380
m/z


Intens.
[%]
+MS2(647.3607), 24.4eV, 7.3-7.5min #418-430
1424.7480
1425.7493
1411.8036
1407.7430
1418.7313
1452.8209
1445.7926
1500.7728
1539.8246
400
300
200
100
0
1380
1400
1420
1440
1460
1480
1500
1520
1540
m/z

1PRB_rev: **H_2N-AHAKLIENKLANVEEVTKAKNIANFYFDSTIGAKKLEAIADEKANKL-COOH**

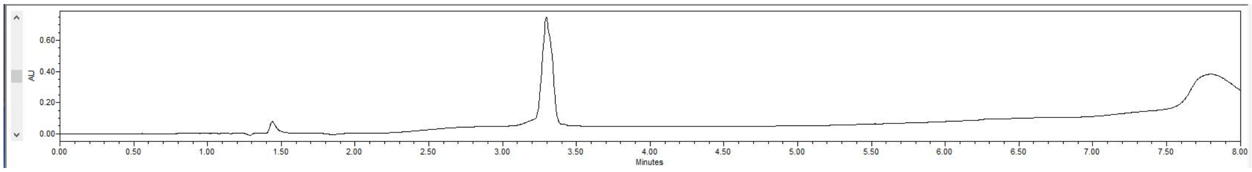
<u>ESI-MS</u>: calculated monoisotopic $[M+5H]^{5+} = 1035.1648$

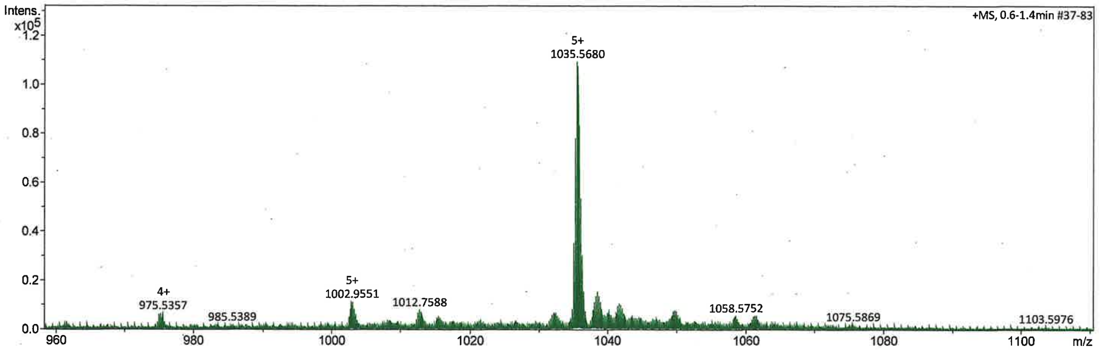
observed monoisotopic $[M+5H]^{5+} = 1035.5680$

(*ESI-MS/MS was used to obtain fragment ionized peptides and the measured the mass differences between the resulting fragment ions in MS/MS graphs were used to determine the amino acid sequence.)

<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity > 99.0 %





MS-MS fragmentation of peptide: AHAKLIENKLANVEEVTKAKNIANFYFDSTIGAKKLEAIADEKANKL

Elemental composition: C232 H380 N62 O71

Mass: [Mo.] 5171.810 Da.(MH+) 2586.409 Da.(M2H+) 1724.608 Da.(M3H+) [Mass file: AA_MASS.MSS]

a	b	Res:	x	y'	z'	v'	w'
44.050	72.045	1 Ala 47	-	-	-	-	
181.109	209.104	2 His 46	5126.753	5100.773	5084.754	5019.728	5018.733
252.146	280.141	3 Ala 45	4989.694	4963.714	4947.696	4948.691	-
380.241	408.236	4 Lys 44	4918.657	4892.677	4876.658	4820.596	4819.601
493.325	521.320	5 Leu 43	4790.562	4764.582	4748.564	4707.512	4706.517
606.409	634.404	6 Ile 42	4677.478	4651.498	4635.479	4594.428	4607.448
735.452	763.447	7 Glu 41	4564.394	4538.414	4522.395	4465.385	4464.390
849.495	877.490	8 Asn 40	4435.351	4409.372	4393.353	4351.342	4350.347
977.590	1005.585	9 Lys 39	4321.308	4295.329	4279.310	4223.247	4222.252
1090.674	1118.669	10 Leu 38	4193.213	4167.234	4151.215	4110.163	4109.168
1161.711	1189.706	11 Ala 37	4080.129	4054.150	4038.131	4039.126	-
1275.754	1303.749	12 Asn 36	4009.092	3983.112	3967.094	3925.083	3924.088
1374.822	1402.817	13 Val 35	3895.049	3869.070	3853.051	3826.015	3839.035
1503.865	1531.860	14 Glu 34	3795.981	3770.001	3753.982	3696.972	3695.977
1632.907	1660.902	15 Glu 33	3666.938	3640.958	3624.940	3567.930	3566.934
1731.976	1759.971	16 Val 32	3537.896	3511.916	3495.897	3468.861	3481.882
1833.023	1861.018	17 Thr 31	3438.827	3412.847	3396.829	3367.813	3380.834
1961.118	1989.113	18 Lys 30	3337.780	3311.800	3295.781	3239.718	3238.723
2032.156	2060.150	19 Ala 29	3209.685	3183.705	3167.686	3168.681	-
2160.250	2188.245	20 Lys 28	3138.648	3112.668	3096.649	3040.586	3039.591
2274.293	2302.288	21 Asn 27	3010.553	2984.573	2968.554	2926.543	2925.548
2387.377	2415.372	22 Ile 26	2896.510	2870.530	2854.511	2813.459	2826.480
2458.415	2486.409	23 Ala 25	2783.426	2757.446	2741.427	2742.422	-
2572.458	2600.452	24 Asn 24	2712.388	2686.409	2670.390	2628.379	2627.384
2719.526	2747.521	25 Phe 23	2598.346	2572.366	2556.347	2481.311	2480.316
2882.589	2910.584	26 Tyr 22	2451.277	2425.297	2409.279	2318.248	2317.252
3029.658	3057.653	27 Phe 21	2288.214	2262.234	2246.215	2171.179	2170.184
3144.685	3172.680	28 Asp 20	2141.145	2115.166	2099.147	2056.152	2055.157
3231.717	3259.712	29 Ser 19	2026.118	2000.139	1984.120	1969.120	1968.125
3332.764	3360.759	30 Thr 18	1939.086	1913.107	1897.088	1868.073	1881.093
3445.848	3473.843	31 Ile 17	1838.039	1812.059 ✓	1796.040	1754.989	1768.009
3502.870	3530.865	32 Gly 16	1724.955	1698.975 -	1682.956	-	-
3573.907	3601.902	33 Ala 15	1667.933	1641.953 -	1625.935	1626.930	-
3702.002	3729.997	34 Lys 14	1596.896	1570.916 -	1554.898	1498.835	1497.840
3830.097	3858.092	35 Lys 13	1468.801	1442.821 -	1426.803	1370.740	1369.745
3943.181	3971.176	36 Leu 12	1340.706	1314.726	1298.708	1257.656	1256.661
4072.224	4100.218	37 Glu 11	1227.622	1201.642	1185.624	1128.613	1127.618
4143.261	4171.256	38 Ala 10	1098.580	1072.600	1056.581	1057.576	-
4256.345	4284.340	39 Ile 9	1027.542	1001.563	985.544	944.492	957.513
4327.382	4355.377	40 Ala 8	914.458	888.479	872.460	873.455	-
4442.409	4470.404	41 Asp 7	843.421	817.441	801.423	758.428	757.433
4571.451	4599.446	42 Glu 6	728.394	702.414 ✓	686.396	629.386	628.390
4699.546	4727.541	43 Lys 5	599.352	573.372 ✓	557.353	501.291	500.295
4770.583	4798.578	44 Ala 4	471.257	445.277 ✓	429.258	430.253	-
4884.626	4912.621	45 Asn 3	400.220	374.240 ✓	358.221	316.211	315.215
5012.721	5040.716	46 Lys 2	286.177	260.197 ✓	244.178	188.116	187.120
-	-	47 Leu 1	158.082	132.102	116.083	75.031	74.036

- Ion charge state: 2

22.529	36.526	1 Ala 47	-	-	-	-	
91.058	105.056	2 His 46	2563.880	2550.890	2542.881	2510.368	2509.870
126.577	140.574	3 Ala 45	2495.351	2482.361	2474.351	2474.849	-
190.624	204.622	4 Lys 44	2459.832	2446.842	2438.833	2410.802	2410.304

247.166	261.164	5 Leu 43	2395.785	2382.795	2374.785	2354.260	2353.762
303.708	317.706	6 Ile 42	2339.243	2326.253	2318.243	2297.718	2304.228
368.230	382.227	7 Glu 41	2282.701	2269.711	2261.701	2233.196	2232.699
425.251	439.248	8 Asn 40	2218.179	2205.189	2197.180	2176.175	2175.677
489.298	503.296	9 Lys 39	2161.158	2148.168	2140.159	2112.127	2111.630
545.840	559.838	10 Leu 38	2097.110	2084.120	2076.111	2055.585	2055.088
581.359	595.356	11 Ala 37	2040.568	2027.578	2019.569	2020.067	-
638.381	652.378	12 Asn 36	2005.050	1992.060	1984.050	1963.045	1962.548
687.915	701.912	13 Val 35	1948.028	1935.038	1927.029	1913.511	1920.021
752.436	766.433	14 Glu 34	1898.494	1885.504	1877.495	1848.990	1848.492
816.957	830.955	15 Glu 33	1833.973	1820.983	1812.974	1784.468	1783.971
866.492	880.489	16 Val 32	1769.451	1756.462	1748.452	1734.934	1741.444
917.015	931.013	17 Thr 31	1719.917	1706.927	1698.918	1684.410	1690.921
981.063	995.060	18 Lys 30	1669.393	1656.404	1648.394	1620.363	1619.865
1016.581	1030.579	19 Ala 29	1605.346	1592.356	1584.347	1584.844	-
1080.629	1094.626	20 Lys 28	1569.827	1556.838	1548.828	1520.797	1520.299
1137.650	1151.648	21 Asn 27	1505.780	1492.790	1484.781	1463.775	1463.278
1194.192	1208.190	22 Ile 26	1448.758	1435.769	1427.759	1407.233	1413.744
1229.711	1243.708	23 Ala 25	1392.216	1379.227	1371.217	1371.715	-
1286.732	1300.730	24 Asn 24	1356.698	1343.708	1335.699	1314.693	1314.196
1360.267	1374.264	25 Phe 23	1299.676	1286.687	1278.677	1241.159	1240.661
1441.798	1455.796	26 Tyr 22	1226.142	1213.152	1205.143	1159.627	1159.130
1515.332	1529.330	27 Phe 21	1144.611	1131.621	1123.611	1086.093	1085.596
1572.846	1586.843	28 Asp 20	1071.076	1058.086	1050.077	1028.580	1028.082
1616.362	1630.359	29 Ser 19	1013.563	1000.573	992.564	985.064	984.566
1666.886	1680.883	30 Thr 18	970.047	957.057	949.048	934.540	941.050
1723.428	1737.425	31 Ile 17	919.523	906.533	898.524	877.998	884.508
1751.939	1765.936	32 Gly 16	862.981	849.991	841.982	-	-
1787.457	1801.455	33 Ala 15	834.470	821.480	813.471	813.969	-
1851.505	1865.502	34 Lys 14	798.952	785.962	777.952	749.921	749.423
1915.552	1929.550	35 Lys 13	734.904	721.914	713.905	685.874	685.376
1972.094	1986.092	36 Leu 12	670.857	657.867	649.857	629.332	628.834
2036.615	2050.613	37 Glu 11	614.315	601.325	593.315	564.810	564.313
2072.134	2086.131	38 Ala 10	549.793	536.803	528.794	529.292	-
2128.676	2142.673	39 Ile 9	514.275	501.285	493.276	472.750	479.260
2164.195	2178.192	40 Ala 8	457.733	444.743	436.734	437.231	-
2221.708	2235.705	41 Asp 7	422.214	409.224	401.215	379.718	379.220
2286.229	2300.227	42 Glu 6	364.701	351.711	343.702	315.196	314.699
2350.277	2364.274	43 Lys 5	300.179	287.190	279.180	251.149	250.651
2385.795	2399.793	44 Ala 4	236.132	223.142	215.133	215.630	-
2442.817	2456.814	45 Asn 3	200.613	187.624	179.614	158.609	158.111
2506.864	2520.862	46 Lys 2	143.592	130.602	122.593	94.561	94.064
-	-	47 Leu 1	79.544	66.555	58.545	38.019	37.522

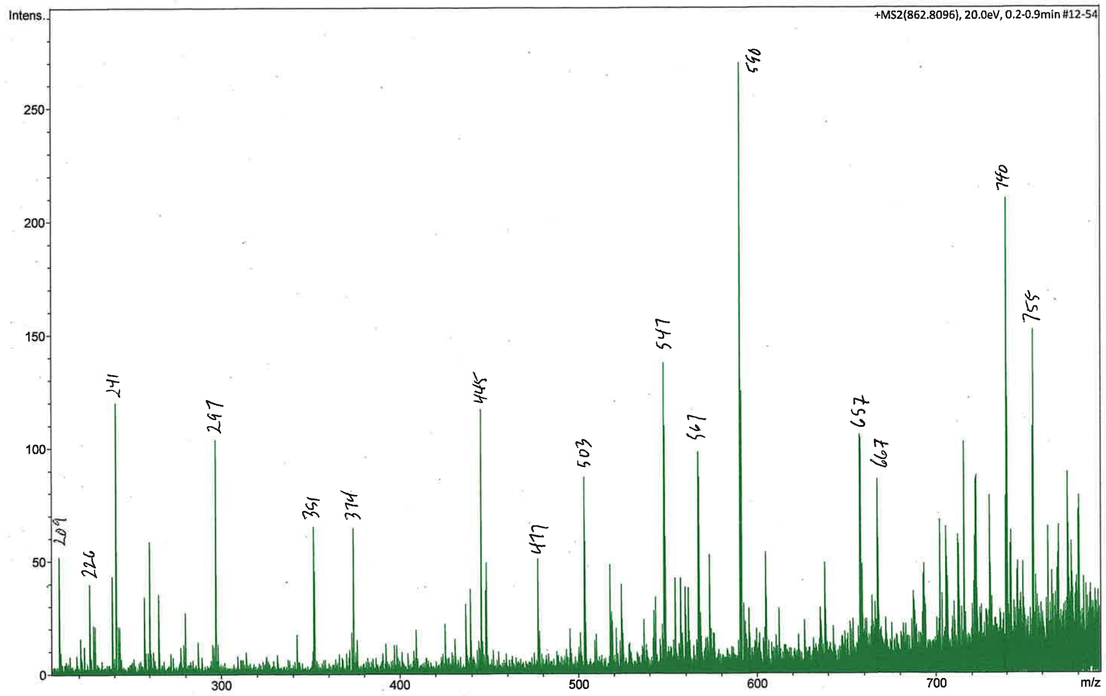
- Ion charge state: 3

15.355	24.686	1 Ala 47	-	-	-	-	
61.041	70.373	2 His 46	1709.589	1700.929	1695.590	1673.914	1673.582
84.720	94.052	3 Ala 45	1663.903	1655.243	1649.903	1650.235	-
127.419	136.750	4 Lys 44	1640.224	1631.564	1626.224	1607.537	1607.205
165.113	174.445	5 Leu 43	1597.526	1588.866	1583.526	1569.842	1569.510
202.808	212.140	6 Ile 42	1559.831	1551.171	1545.831	1532.147	1536.488
245.822	255.154	7 Glu 41	1522.136	1513.476	1508.137	1489.133	1488.801
283.836	293.168	8 Asn 40	1479.122	1470.462	1465.122	1451.119	1450.787
326.535	335.866	9 Lys 39	1441.108	1432.448	1427.108	1408.421	1408.089
364.229	373.561	10 Leu 38	1398.409	1389.749	1384.410	1370.726	1370.394
387.908	397.240	11 Ala 37	1360.715	1352.055	1346.715	1347.047	-
425.923	435.254	12 Asn 36	1337.036	1328.376	1323.036	1309.033	1308.701
458.946	468.277	13 Val 35	1299.021	1290.361	1285.022	1276.010	1280.350
501.960	511.291	14 Glu 34	1265.998	1257.339	1251.999	1232.996	1232.664
544.974	554.306	15 Glu 33	1222.984	1214.324	1208.985	1189.981	1189.650
577.997	587.328	16 Val 32	1179.970	1171.310	1165.971	1156.959	1161.299
611.679	621.011	17 Thr 31	1146.947	1138.287	1132.948	1123.276	1127.616

654.378	663.709	18 Lys 30	1113.265	1104.605	1099.265	1080.578	1080.246
678.057	687.388	19 Ala 29	1070.566	1061.906	1056.567	1056.899	-
720.755	730.087	20 Lys 28	1046.887	1038.227	1032.888	1014.200	1013.869
758.769	768.101	21 Asn 27	1004.189	995.529	990.190	976.186	975.854
796.464	805.796	22 Ile 26	966.175	957.515	952.175	938.491	942.831
820.143	829.475	23 Ala 25	928.480	919.820	914.481	914.812	-
858.157	867.489	24 Asn 24	904.801	896.141	890.801	876.798	876.466
907.180	916.512	25 Phe 23	866.787	858.127	852.787	827.775	827.443
961.535	970.866	26 Tyr 22	817.764	809.104	803.764	773.421	773.089
1010.557	1019.889	27 Phe 21	763.409	754.750	749.410	724.398	724.066
1048.900	1058.231	28 Asp 20	714.387	705.727	700.387	686.056	685.724
1077.910	1087.242	29 Ser 19	676.044	667.384	662.045	657.045	656.713
1111.593	1120.925	30 Thr 18	647.034	638.374	633.034	623.362	627.703
1149.288	1158.619	31 Ile 17	613.351	604.691	599.352	585.668	590.008
1168.295	1177.626	32 Gly 16	575.656	566.996	561.657	-	-
1191.974	1201.305	33 Ala 15	556.649	547.989	542.650	542.981	-
1234.672	1244.004	34 Lys 14	532.970	524.310	518.971	500.283	499.951
1277.370	1286.702	35 Lys 13	490.272	481.612	476.272	457.585	457.253
1315.065	1324.397	36 Leu 12	447.574	438.914	433.574	419.890	419.558
1358.079	1367.411	37 Glu 11	409.879	401.219	395.879	376.876	376.544
1381.758	1391.090	38 Ala 10	366.865	358.205	352.865	353.197	-
1419.453	1428.785	39 Ile 9	343.186	334.526	329.186	315.502	319.842
1443.132	1452.464	40 Ala 8	305.491	296.831	291.491	291.823	-
1481.474	1490.806	41 Asp 7	281.812	273.152	267.812	253.481	253.149
1524.489	1533.820	42 Glu 6	243.470	234.810	229.470	210.467	210.135
1567.187	1576.519	43 Lys 5	200.455	191.795	186.456	167.768	167.437
1590.866	1600.198	44 Ala 4	157.757	149.097	143.758	144.089	-
1628.880	1638.212	45 Asn 3	134.078	125.418	120.079	106.075	105.743
1671.579	1680.910	46 Lys 2	96.064	87.404	82.064	63.377	63.045
-	-	47 Leu 1	53.365	44.705	39.366	25.682	25.350

- Ion charge state: 4

11.768	18.767	1 Ala 47	-	-	-	-	
46.033	53.031	2 His 46	1282.444	1275.949	1271.944	1255.687	1255.439
63.792	70.791	3 Ala 45	1248.179	1241.684	1237.679	1237.928	-
95.816	102.814	4 Lys 44	1230.420	1223.925	1219.920	1205.904	1205.656
124.087	131.085	5 Leu 43	1198.396	1191.901	1187.896	1177.633	1177.385
152.358	159.356	6 Ile 42	1170.125	1163.630	1159.625	1149.362	1152.617
184.618	191.617	7 Glu 41	1141.854	1135.359	1131.354	1117.102	1116.853
213.129	220.128	8 Asn 40	1109.593	1103.098	1099.094	1088.591	1088.342
245.153	252.152	9 Lys 39	1081.083	1074.588	1070.583	1056.567	1056.318
273.424	280.423	10 Leu 38	1049.059	1042.564	1038.559	1028.296	1028.047
291.183	298.182	11 Ala 37	1020.788	1014.293	1010.288	1010.537	-
319.694	326.693	12 Asn 36	1003.029	996.534	992.529	982.026	981.777
344.461	351.460	13 Val 35	974.518	968.023	964.018	957.259	960.514
376.722	383.720	14 Glu 34	949.751	943.256	939.251	924.998	924.750
408.982	415.981	15 Glu 33	917.490	910.995	906.990	892.738	892.489
433.749	440.748	16 Val 32	885.229	878.734	874.730	867.971	871.226
459.011	466.010	17 Thr 31	860.462	853.967	849.963	842.709	845.964
491.035	498.034	18 Lys 30	835.200	828.705	824.701	810.685	810.436
508.794	515.793	19 Ala 29	803.177	796.682	792.677	792.926	-
540.818	547.817	20 Lys 28	785.417	778.922	774.918	760.902	760.653
569.329	576.328	21 Asn 27	753.394	746.899	742.894	732.391	732.143
597.600	604.599	22 Ile 26	724.883	718.388	714.383	704.120	707.375
615.359	622.358	23 Ala 25	696.612	690.117	686.112	686.361	-
643.870	650.869	24 Asn 24	678.853	672.358	668.353	657.850	657.601
680.637	687.636	25 Phe 23	650.342	643.847	639.842	621.083	620.834
721.403	728.401	26 Tyr 22	613.575	607.080	603.075	580.317	580.069
758.170	765.169	27 Phe 21	572.809	566.314	562.309	543.550	543.301
786.927	793.925	28 Asp 20	536.042	529.547	525.542	514.794	514.545
808.685	815.683	29 Ser 19	507.285	500.790	496.785	493.036	492.787
833.947	840.945	30 Thr 18	485.527	479.032	475.027	467.774	471.029


Intens.
+MS2(862.8096), 20.0eV, 0.2-0.9min #12-54
0
50
100
150
200
250
300
400
500
600
700
m/z
209
226
241
297
351
374
445
477
503
547
567
590
657
667
740
755

4.7. References

(1) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 2021, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(2) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C. C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* 2024, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(3) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. *Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model*; 2023; Vol. 379. https://www.science.org.

(4) Rey, J.; Murail, S.; De Vries, S.; Derreumaux, P.; Tuffery, P. PEP-FOLD4: A PH-Dependent Force Field for Peptide Structure Prediction in Aqueous Solution. *Nucleic Acids Res* 2023, *51* (W1), W432–W437. https://doi.org/10.1093/nar/gkad376.

(5) Lamiable, A.; Thevenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuffery, P. PEP-FOLD3: Faster Denovo Structure Prediction for Linear Peptides in Solution and in Complex. *Nucleic Acids Res* 2016, *44* (1), W449–W454. https://doi.org/10.1093/nar/gkw329.

(6) Binette, V.; Mousseau, N.; Tuffery, P. A Generalized Attraction-Repulsion Potential and Revisited Fragment Library Improves PEP-FOLD Peptide Structure Prediction. *J Chem Theory Comput* 2022, *18* (4), 2720–2736. https://doi.org/10.1021/acs.jctc.1c01293.

(7) Armstrong, D. R.; Berrisford, J. M.; Conroy, M. J.; Gutmanas, A.; Anyango, S.; Choudhary, P.; Clark, A. R.; Dana, J. M.; Deshpande, M.; Dunlop, R.; Gane, P.; Gáborová, R.; Gupta, D.; Haslam, P.; Koča, J.; Mak, L.; Mir, S.; Mukhopadhyay, A.; Nadzirin, N.; Nair, S.; Paysan-Lafosse, T.; Pravda, L.; Sehnal, D.; Salih, O.; Smart, O.; Tolchard, J.; Varadi, M.; Svobodova-Vařeková, R.; Zaki, H.; Kleywegt, G. J.; Velankar, S. PDBe: Improved Findability of Macromolecular Structure Data in the PDB. *Nucleic Acids Res* 2020, *48* (D1), D335–D343. https://doi.org/10.1093/nar/gkz990.

(8) Dill, K. A.; Maccallum, J. L. The Protein-Folding Problem, 50 Years On. *Science (1979)* 2012, *338* (6110), 1042–1046.

(9) Chakravarty, D.; Lee, M.; Porter, L. L. Proteins with Alternative Folds Reveal Blind Spots in AlphaFold-Based Protein Structure Prediction. *Curr Opin Struct Biol* 2025, *90*, 102973. https://doi.org/10.1016/j.sbi.2024.102973.

(10) Agarwal, V.; McShan, A. C. The Power and Pitfalls of AlphaFold2 for Structure Prediction beyond Rigid Globular Proteins. *Nat Chem Biol* 2024, *20* (8), 950–959. https://doi.org/10.1038/s41589-024-01638-w.

(11) Ruff, K. M.; Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of Molecular Biology*. Academic Press October 1, 2021. https://doi.org/10.1016/j.jmb.2021.167208.

(12) Terwilliger, T. C.; Liebschner, D.; Croll, T. I.; Williams, C. J.; McCoy, A. J.; Poon, B. K.; Afonine, P. V.; Oeffner, R. D.; Richardson, J. S.; Read, R. J.; Adams, P. D. AlphaFold Predictions Are Valuable Hypotheses and Accelerate but Do Not Replace Experimental Structure Determination. *Nat Methods* 2024, *21* (1), 110–116. https://doi.org/10.1038/s41592-023-02087-4.

(13) Childs, H.; Zhou, P.; Donald, B. R. Has AlphaFold 3 Solved the Protein Folding Problem for D-Peptides? *bioRxiv* 2025. https://doi.org/10.1101/2025.03.14.643307.

(14) Lander, A. J.; Jin, Y.; Luk, L. Y. P. D-Peptide and D-Protein Technology: Recent Advances, Challenges, and Opportunities**. *ChemBioChem* 2023, *24* (4). https://doi.org/10.1002/cbic.202200537.

(15) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Signal Transduct Target Ther* 2022, *7* (1). https://doi.org/10.1038/s41392-022-00904-4.

(16) Feng, Z.; Xu, B. Inspiration from the Mirror: D-Amino Acid Containing Peptides in Biomedical Approaches. *Biomolecular Concepts*. Walter de Gruyter GmbH June 1, 2016, pp 179–187. https://doi.org/10.1515/bmc-2015-0035.

(17) da Silva, F. B.; Perlinska, A. P.; Płonka, J.; Flapan, E.; Sulkowska, J. I. Universe of Lasso Proteins: Exploring the Limit of Entanglement and Folding Landscape of Proteins Predicted by AlphaFold. March 22, 2025. https://doi.org/10.1101/2025.03.21.644650.

(18) Chekan, J. R.; Koos, J. D.; Zong, C.; Maksimov, M. O.; Link, A. J.; Nair, S. K. Structure of the Lasso Peptide Isopeptidase Identifies a Topology for Processing Threaded Substrates. *J Am Chem Soc* 2016, *138* (50), 16452–16458. https://doi.org/10.1021/jacs.6b10389.

(19) Hou, Y.; Xie, T.; He, L.; Tao, L.; Huang, J. Topological Links in Predicted Protein Complex Structures Reveal Limitations of AlphaFold. *Commun Biol* 2023, *6* (1). https://doi.org/10.1038/s42003-023-05489-4.

(20) Olszewski, K. A.; Kolinski, A.; Skolnick, J. Does a Backwardly Read Protein Sequence Have a Unique Native State? *Protein Eng* 1996, *9*, 5–14.

(21) Arnittali, M.; Rissanou, A. N.; Kefala, A.; Kokkinidis, M.; Harmandaris, V. Structure of Amino Acid Sequence-Reversed WtRop Protein: Insights from Atomistic Molecular Dynamics Simulations. *J Biomol Struct Dyn* 2024, *42* (19), 9842–9856. https://doi.org/10.1080/07391102.2023.2252903.

(22) Lacroix, E.; Viguera, A. R.; Serrano, L. Reading Protein Sequences Backwards. *Fold Des* 1998, *3*, 79–85.

(23) Zhang, Y.; Weber, J. K.; Zhou, R. Folding and Stabilization of Native-Sequence-Reversed Proteins. *Sci Rep* 2016, *6*, 25138. https://doi.org/10.1038/srep25138.

(24) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. *High-Resolution x-Ray Crystal Structures of the Villin Headpiece Subdomain, an Ultrafast Folding Protein*; 2005; Vol. 24. www.pnas.orgcgidoi10.1073pnas.0502495102.

(25) Žoldák, G.; Stigler, J.; Pelz, B.; Li, H.; Rief, M. Ultrafast Folding Kinetics and Cooperativity of Villin Headpiece in Single-Molecule Force Spectroscopy. *Proc Natl Acad Sci U S A* 2013, *110* (45), 18156–18161. https://doi.org/10.1073/pnas.1311495110.

(26) Johansson, M. U.; De Cha Ã Teau, M.; Forse Ân, S.; Rn Drakenberg, T. È.; Bjo, L.; Rck, È. Communication: Solution Structure of the Albumin-Binding GA Module: A Versatile Bacterial Protein Domain. *J. Mol. Biol.* 1997, *266*, 859–865.

(27) Mortenson, D. E.; Kreitler, D. F.; Yun, H. G.; Gellman, S. H.; Forest, K. T. Evidence for Small-Molecule-Mediated Loop Stabilization in the Structure of the Isolated Pin1 WW Domain. *Acta Crystallogr D Biol Crystallogr* 2013, *69* (12), 2506–2512. https://doi.org/10.1107/S090744491302444X.

(28) Kowalski, J. A.; Liu, K.; Kelly, J. W. NMR Solution Structure of the Isolated Apo Pin1 WW Domain: Comparison to the x-Ray Crystal Structures of Pin1. *Biopolymers* 2002, *63* (2), 111–121. https://doi.org/10.1002/bip.10020.

(29) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat Protoc* 2006, *1* (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.

(30) Greenfield, N. J. Using Circular Dichroism Collected as a Function of Temperature to Determine the Thermodynamics of Protein Unfolding and Binding Interactions. *Nat Protoc* 2006, *1* (6), 2527–2535. https://doi.org/10.1038/nprot.2006.204.

(31) Kreitler, D. F.; Mortenson, D. E.; Forest, K. T.; Gellman, S. H. Effects of Single α-to-β Residue Replacements on Structure and Stability in a Small Protein: Insights from Quasiracemic Crystallization. *J Am Chem Soc* 2016, *138* (20), 6498–6505. https://doi.org/10.1021/jacs.6b01454.

(32) Thévenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P. PEP-FOLD: An Updated de Novo Structure Prediction Server for Both Linear and Disulfide Bonded Cyclic Peptides. *Nucleic Acids Res* 2012, *40* (W1), W288–W293. https://doi.org/10.1093/nar/gks419.

(33) Maupetit, J.; Derreumaux, P.; Tuffery, P. PEP-FOLD: An Online Resource for de Novo Peptide Structure Prediction. *Nucleic Acids Res* 2009, *37* (SUPPL. 2), W498–W503. https://doi.org/10.1093/nar/gkp323.

(34) Xu, J.; Zhang, Y. How Significant Is a Protein Structure Similarity with TM-Score = 0.5? *Bioinformatics* 2010, *26* (7), 889–895. https://doi.org/10.1093/bioinformatics/btq066.

Chapter 5

Effects of D-amino acid replacements on the conformational stability of miniproteins

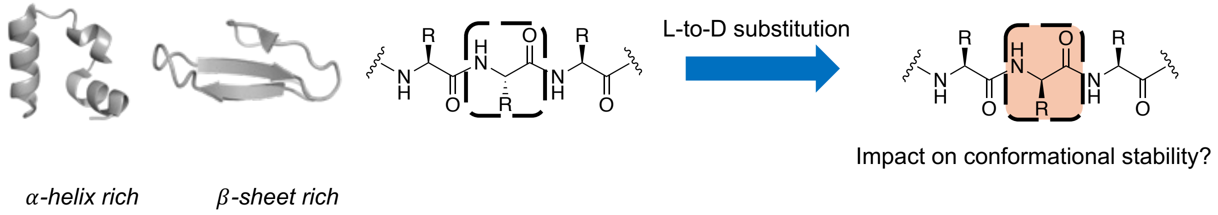
Portions of this work have been published for publication in: Effects of D-amino acid replacements on the conformational stability of miniproteins Xu, R.*; Huang, J*; Kuhn, A. J.; Gellman, S. *ChemBioChem* **2025**, 26, e202500085. (*These authors contributed equally to this work. Huang, J. was an undergraduate researcher under Xu, R.'s supervision.)

Most of the text and graphics are reprinted (adapted) with permission from {*ChemBioChem* **2025**, 26, e202500085}.

Graphics were prepared using Microsoft PowerPoint, GraphPad Prism, and ChemDraw.

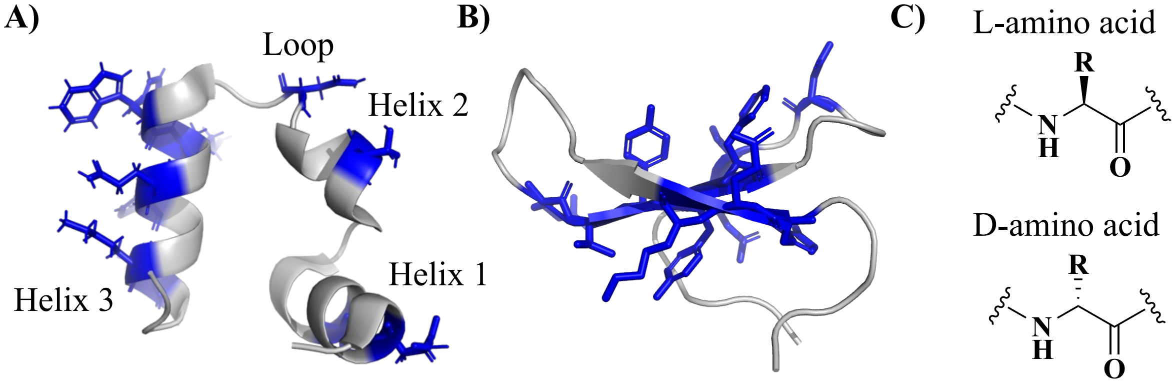
5.1. Abstract

For many proteins, proper function requires adoption of a specific tertiary structure. This study explores the effects of L-to-D amino acid substitutions on tertiary structure stability for two well-known miniproteins, a single-site variant of the chicken villin headpiece subdomain (VHP) and the human Pin1 WW domain (WW). For VHP, which features an α-helix-rich tertiary structure, substitutions led to significant destabilization, as detected by variable temperature circular dichroism (CD) measurements. For WW, which has a β-sheet-rich tertiary structure, most single L-to-D changes seemed to cause complete unfolding at room temperature, according to CD measurements. We probed the effects of L-to-D amino acid substitutions on protein stability using far-UV CD. Comparison of the α-helix-rich villin headpiece subdomain and the β-sheet-rich WW domain revealed that L-to-D substitutions generally destabilize proteins. These findings suggest that amino acid residue configuration changes at a single site will often prove to be deleterious in terms of tertiary structure stability, and in some cases dramatically destabilizing.



5.2. Introduction

Proteins are responsible for a wide range of functions that are necessary for life; no other class of biomolecules is so pervasive in terms of critical activities that include catalysis, energy capture and utilization, signal transduction and transport. From a chemical perspective, it seems remarkable that molecules constructed from only 20 different amino acid building blocks can display such a broad range of functions. The proteinogenic α-amino acids are defined not only by their side chains, but also, except for glycine, by their absolute configuration. The natural translational machinery strongly favors incorporation of L-α-amino acids into proteins. In contrast, prokaryotic nonribosomal synthetases often incorporate D-α-amino acid residues into short peptides.[1,2] Ribosomally produced proteins isolated from natural sources can contain low levels of D residues.[3] Some short, bioactive peptides of ribosomal origin contain D residues that are apparently generated as a post-translational modification.[4] Configurational change at a single site can exert profound effects on the biological activity of a peptide hormone.[5]



D)

VHP(1) H_2N-LSDEDFKAVFGMTRSAFANLPLWKQQHLKKEKGLF-COOH
WW(2) H_2N-KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG-COOH

Figure 5.1. A) Crystal structure of VHP (1YRF) with the positions of L-to-D substitution shown in blue. B) Structure of WW (4GWT) with the positions of L-to-D substitution shown in blue. C) Chemical structures of L-amino acid and D-amino acid. D) Sequences of VHP (**1**) and WW (**2**).

Incorporation of D-α-amino acid residues via chemical peptide synthesis is straightforward. Peptides containing D residues have been developed as drugs and drug candidates,[6,7] and have proven useful as research tools.[8–10] Inclusion of D residues in macrocyclic peptides can influence conformation and promote membrane permeability.[11–13] Because D residues suppress protease action at nearby positions[14], there is considerable interest in engineering translational machinery to promote D residue incorporation.[15,16]

Protein function often depends upon conformation, but our understanding of the impact of L-to-D changes on the folding behavior of linear polypeptides is limited. α-Helix stability is generally diminished when an L residue is replaced by the corresponding D residue. For alanine, this stereochemical change destabilizes the α-helix by ~1 kcal/mol[17,18]; however, energetic consequences of a configuration change appear to vary depending on the side chain[19]. Hecht et al. used modified ribosomal machinery to interrogate single L-to-D variants of two enzymes, *E. coli* dihydrofolate reductase (159 residues; three epimers examined) and firefly luciferase (550 residues; two epimers examined).[16] The resulting protein diastereomers were evaluated in terms of specific catalytic activity, which provides a very sensitive measure of the polypeptide's ability to achieve a native-like tertiary structure. Even when the L-to-D modification was made at a position known to be important in terms of catalytic function, substantial enzymatic activity was retained (≥ 10% of native specific activity). These findings suggest that although L-to-D modifications can be disruptive in terms of tertiary structure, the rest of the protein may be able to adjust to compensate at least partially for the disruption in a sufficiently large protein.

Here, we investigate the effects of L-to-D modifications on the conformational stability of two tertiary folding units that are small enough to be easily prepared via conventional solid-phase peptide synthesis (SPPS). The villin headpiece subdomain, 35 residues, forms a tertiary structure containing several α-helices[20,21]; below, "VHP" refers to the sequence from chicken with a single change (Asn27→His)[21] (Figure 5.1A). The human Pin1 WW domain ("WW" below), 34 residues, forms a β-sheet-rich tertiary structure[22] (Figure 5.1B). Neither of these "miniproteins" contains a disulfide or requires metal ion coordination for folding. Thus, each is a good vehicle for determining how single L-to-D substitutions, implemented at multiple positions, influence tertiary structure stability.

5.3. Results and Discussion

5.3.1. Evaluation of diastereomers of VHP.

High-resolution structures have been determined by x-ray crystallography or NMR for VHP, the Pin1 WW domain and sequence variants.[21,22] In each case, the native structure is associated with a distinctive circular dichroism (CD) signature.[23] For VHP, this signature has minima near 208 and 222 nm, as expected given that many residues participate in α-helical secondary structure. We used variable-temperature CD to evaluate the conformational stability of VHP and a set of diastereomers.[24] We designed an initial set of ten D-amino acid containing variants by systematically replacing the solvent-exposed L-amino acids on each of the three helices and in the loop between helix 2 and helix 3. Six L-to-D replacements were made in helix 1, one replacement in helix 2, two replacements in helix 3, and one replacement in the loop.

The protein concentration was set at 50 μM, and all CD measurements were conducted in 20 mM sodium acetate buffer, pH 5.0.[25] Each of the 10 single L-to-D VHP variants examined displayed a CD signature at room temperature that suggested a retention of α-helical secondary structure (Figure 5.2). However, in every case, the extent of α-helicity was diminished relative to VHP, with the Trp23 epimer most extreme. For variable-temperature measurements, we monitored mean residue ellipticity (MRE) at 223 nm. Upon heating, the solution containing VHP itself displays a sigmoidal thermal transition with a midpoint at 55.1 ± 1.2 °C. This transition was reversible, as indicated by the data collected as the sample was cooled (Figure 5.3). Solutions containing most of the single L-to-D variants also displayed a sigmoidal change in MRE at 223 nm upon heating, but in some of these cases the transition was not fully reversible (Figure 5.3, 5.6-5.15). These observations suggest that heating causes a portion of the protein to aggregate irreversibly. Therefore, the midpoints of the heating transitions are interpreted as an apparent melting temperature ($T_{m,app}$).

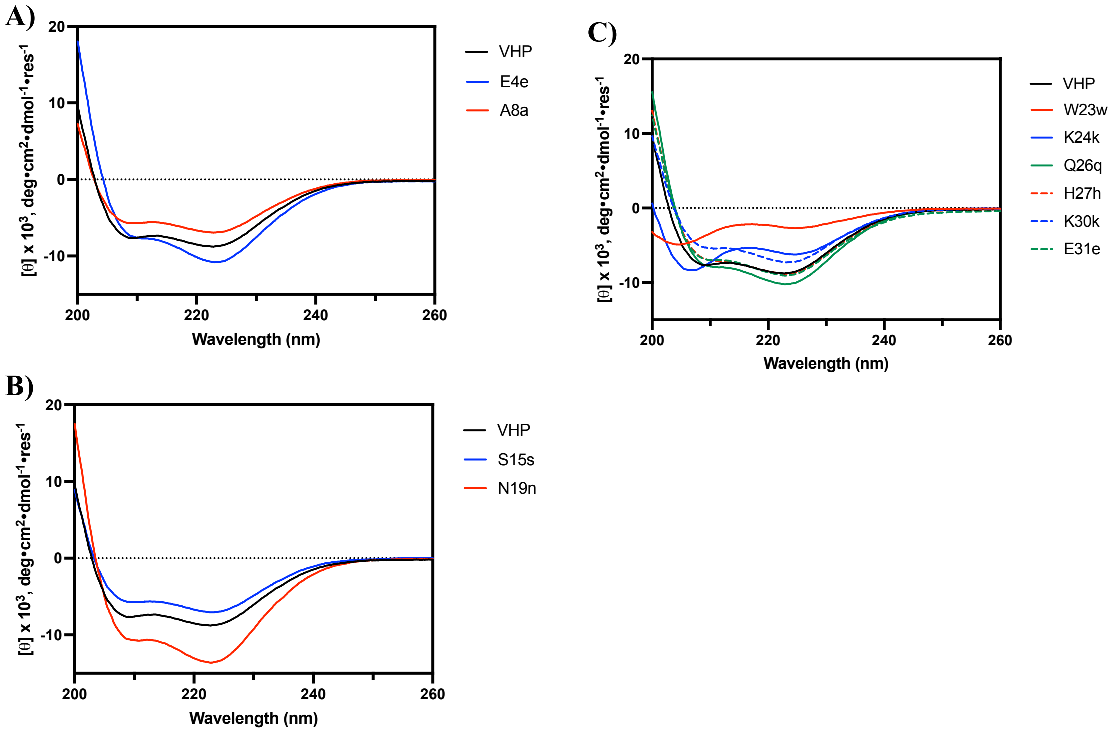


Figure 5.2. Far-UV CD for VHP and diastereomers containing a single L-to-D substitution. Samples contained 50 μM polypeptide, 20 mM sodium acetate buffer, pH 5.0. Measurements at 25°C. A) Substitutions in helix 1, B) substitutions in helix 2 and loop, C) substitutions in helix 3.

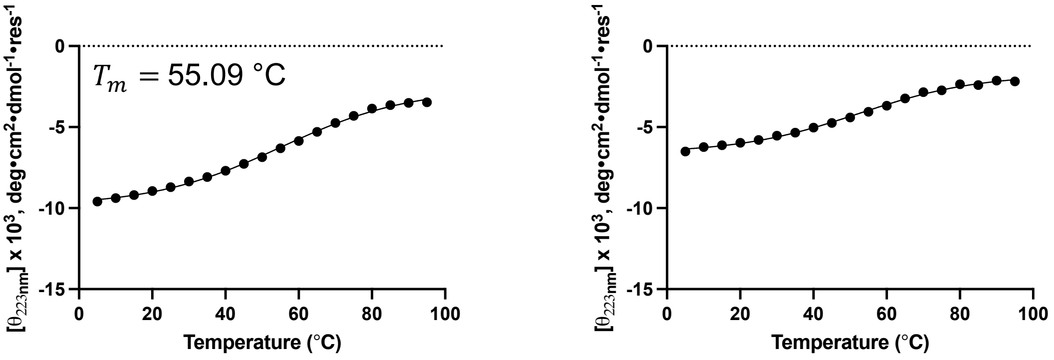


Figure 5.3. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP. Each data set was fit to a two-state folding model.

Table 5.1. Apparent melting temperatures derived from CD for VHP analogues with single L-to-D substitution.

Types	Variant	$T_{m,app}$ (°C)	$\Delta T_{m,app}$ (°C)
	VHP	55.09 (±1.2)	-
Helix	E4e	45.2 (±0.7)	-9.9 (±1.9)
	A8a	41.2 (±0.7)	-13.9 (±1.9)
Helix	S15s	50.2 (±0.7)	-4.9 (±1.9)
Loop	N19n	42.0 (±1.1)	-13.1 (±2.3)
Helix	W23w	Not determined	-
	K24k	26.6 (±1.2)	-28.5 (±2.4)
	Q26q	48.1 (±1.5)	-7.0 (±3.7)
	H27h	53.9 (±2.8)	-1.2 (±4.0)
	K30k	50.4 (±0.8)	-4.7 (±2.0)
	E31e	46.8 (±0.9)	-8.3 (±2.1)

Table 5.1 summarizes results for the 10 single L-to-D VHP variants we designed. The measured parameter in the thermal denaturation assays is defined as $T_{m,app}$ for apparent melting temperature. L-to-D substitution at Gln26, His27, Lys30 or Glu31 in helix 1 resulted modest destabilization relative to VHP, with His27 showing the smallest change ($\Delta T_{m,app}$ = -1.2 °C) and Lys24 the largest ($\Delta T_{m,app}$ = -8.3 °C). The trend among the Helix 1 substitutions was generally consistent with the trend reported for D-amino acid substitutions in an isolated α-helix[19], but the large destabilization observed for the Lys24 epimer ($\Delta T_{m,app}$ = -28.5 °C) deviated from this trend. The large difference between substitution at Lys24 and Lys30 shows that side chain identity is not the only factor that determines the impact of stereochemical inversion. The crystal structure of VHP reveals that the side chain of Lys24 makes more contacts with other side chains than does the side chain of Lys30 (Table 5.2). Presumably, greater involvement of the side chain in tertiary packing interactions will correlate with a large negative effect of L-to-D substitution on tertiary

structure stability. L-to-D substitution at Ser15 ($\Delta T_{m,app}$ = -4.9 °C) in Helix 2, Glu4 ($\Delta T_{m,app}$ = -9.9 °C), or Ala8 ($\Delta T_{m,app}$ = -13.9 °C) in Helix 1 or Asn19 ($\Delta T_{m,app}$ = -13.1 °C) in the loop between helices 2 and 3 exhibited modest destabilizing effects. In each case, CD at room temperature indicated retention of an α-helical signature.

Table 5.2. Accessible Surface Area (ASA) and Solvation energy effect (Δ^iG) calculated from VHP crystal structure (1YRF) using Protein Data Bank in Europe.[25–27]

##	Amino acid	ASA (Å^2)	Δ^iG (kcal/mol)
4	A:GLU 45	130.99	0.00
8	A:ALA 49	84.95	0.00
15	A:SER 56	84.39	0.23
19	A:ASN 60	134.12	0.00
23	A:TRP 64	164.43	0.15
24	A:LYS 65	72.21	0.00
26	A:GLN 67	104.17	-0.21
27	A:HIS 68	84.47	0.00
30	A:LYS 71	123.00	0.13
31	A:GLU 72	126.19	0.00

To further explore the tolerance for D residues in VHP, we designed double and triple L-to-D substituted analogs based on the most stable single substitutions: Ala8, Ser15, Asn19, His27, and Lys30. Data in Table 5.3 suggest that double L-to-D substitutions lead to a more significant destabilization than single substitutions. The $\Delta T_{m,app}$ values for the double substitutions are larger than the sum of $\Delta T_{m,app}$ values for the single substitutions. Thus, for example, the VHP stereoisomer with L-to-D substitution at His27 and Lys30 exhibited $\Delta T_{m,app}$ = -11.4 °C, while substitution at His27 alone resulted in $\Delta T_{m,app}$ = -1.2 °C, and substitution at Lys30 alone resulted $\Delta T_{m,app}$ = -4.7 °C. These findings suggest that multiple substitutions may cooperatively enhance tertiary structure destabilization.

Thermal destabilization trends observed for double substitutions were qualitatively consistent with those identified in single substitutions. Thus, for example, double-substituted stereoisomer substitution at His27 and Lys24 ($\Delta T_{m,app}$ = -31.8 °C) manifests a greater thermal destabilization relative to the stereoisomer with double substitution at His27 and Lys30 ($\Delta T_{m,app}$ = -11.4 °C), which correlates with the observation that single substitution at Lys24 ($\Delta T_{m,app}$ = -28.5 °C) is more destabilizing than single substitution at Lys30 ($\Delta T_{m,app}$ = -4.7 °C). Most VHP diastereomers with double L-to-D substitution maintained some degree of α-helicity at 25 °C, according to CD (Figure 5.3). The two VHP diastereomers with three L-to-D substitutions were highly destabilized, with $T_{m,app}$ values near or below room temperature. The diastereomer with four L-to-D substitutions exhibited a CD spectrum suggesting random coil at room temperature. Collectively, these results indicate that the VHP tertiary structure is readily destabilized by L-to-D substitution at a small number of positions. The destabilizing effect is not simply additive but appears to be exaggerated for multiple substitutions.

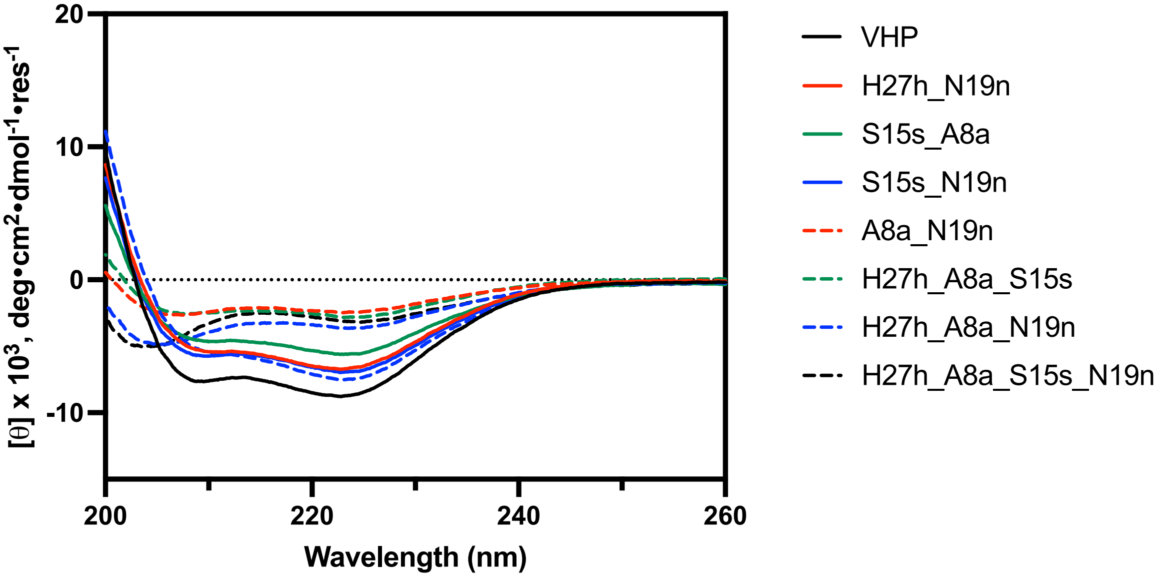


Figure 5.3. Far-UV CD for VHP and diastereomers containing multiple L-to-D substitutions. Samples contained 50 µM polypeptide, 20 mM sodium acetate buffer, pH 5.0. Measurements at 25°C.

Table 5.3. Apparent melting temperatures derived from CD for VHP analogues with multiple L-to-D substitutions.

Variant	$T_{m,app}$ (°C)	$\Delta T_{m,app}$ (°C)
VHP	55.1 (±1.2)	-
H27h_S15s	44.4 (±0.6)	-10.7 (±1.8)
H27h_K30k	43.5 (±1.4)	-11.6 (±2.6)
H27h_A8a	35.0 (±0.9)	-20.1 (±2.1)
H27h_N19n	34.1 (±2.1)	-21 (±3.3)
A8a_S15s	30.0 (±1.1)	-25.1 (±2.3)
S15s_N19n	28.8 (±1.6)	-26.3 (±2.8)
H27h_K24k	23.3 (±1.7)	-31.8 (±2.9)
A8a_N19n	20.7 (±3.9)	-34.4 (±5.1)
H27h_A8a_S15s	24.5 (±5.8)	-30.6 (±7.0)
H27h_A8a_N19n	16.7 (±3.9)	-38.4 (±5.1)
H27h_A8a_S15s_N19n	Not determined	-

5.3.2. Evaluation of diastereomers of WW.

For the Pin1 WW domain, the CD signature features a highly characteristic maximum near 227 nm, which presumably reflects contributions from the two Trp side chains. CD measurements involving the WW were conducted with 50 μM protein in 10 mM sodium phosphate buffer, pH 7.0 (Figure 5.4).[22] Comparisons involving eight single L-to-D variants revealed that only two of these diastereomers retained the characteristic maximum near 227 nm. One of these variants was the epimer at Arg16, in the loop between β-strands 1 and 2. The other was the epimer at Gln28, near the C-terminus of β-strand 3. The other six L-to-D replacements occurred in the core of the β-sheet that dominates the WW domain fold, and each seemed to disrupt tertiary structure formation at room temperature. Variable-temperature studies with WW, as monitored by MRE at 227 nm, revealed a sigmoidal transition. However, this thermal transition was not reversible, nor was the thermal transition for the Arg16 epimer or the Gln28 epimer. Table 5.4 presents the thermal denaturation data for both of these epimers. L-to-D substitution at position Arg16 led to $\Delta T_{m,app}$ = -13.4 °C, while L-to-D substitution at Gln28 exhibited a larger destabilization, $\Delta T_{m,app}$ = -18.6 °C. These findings show that placing a single D residue in WW is highly destabilizing to the tertiary fold.

Table 5.4. Apparent melting temperatures derived from CD for WW analogues with single L-to-D substitutions.

Variant	$T_{m,app}$ (°C)	$\Delta T_{m,app}$ (°C)
WW	53.6 (±0.6)	-
R16r	40.2 (±0.5)	-13.4 (±1.8)
Q28q	35.0 (±0.6)	-18.6

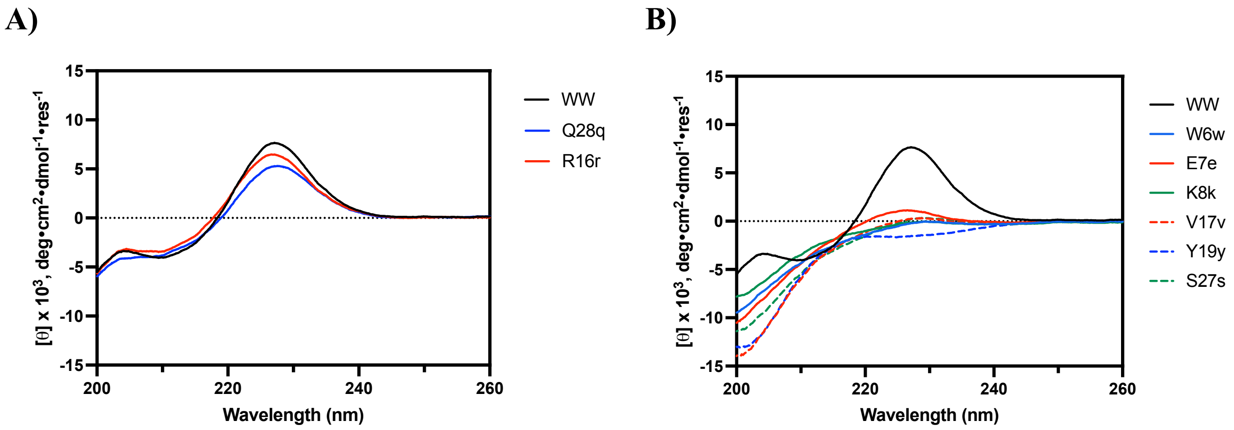


Figure 5.4. Far-UV CD *for WW and* derivatives containing a single L-to-D substitution. Samples contained 50 µM polypeptide, 10 mM sodium phosphate buffer, pH 7.0. Measurements at 25°C. A) WW and the two diastereomers that retain the native CD maximum at ~227 nm, B) WW and the highly destabilized diastereomers.

5.4 Conclusions

We have used two miniproteins that adopt a specific tertiary structure to evaluate the impact of stereoisomeric changes on conformational stability. The villin headpiece subdomain and the Pin1 WW domain are complementary in that most residues in VHP participate in α-helical secondary structure, while most residues in WW participate in β-sheet secondary structure. It is challenging to coax the ribosome to incorporate D-α-amino acid residues, but D incorporation is straightforward with chemical synthesis. The lengths of VHP and WW (35 and 34 residues) make them amenable to synthesis and evaluation of many diastereomers.

Our results indicate that the tertiary structures of both miniproteins are very sensitive to destabilization via L-to-D substitution. For VHP, CD data suggested that most of the single-site modifications we examined did not lead to complete loss of folding, as indicated by retention of some α-helicity at room temperature, although every substitution site led to some degree of destabilization. Multiple L-to-D substitutions resulted in greater destabilization; the one diastereomer we examined with four L-to-D substitutions seemed to be fully unfolded at room temperature. The WW domain appeared to be even more sensitive to disruption by L-to-D substitution than was VHP, since only two of the eight single substitutions we evaluated led to retention of the characteristic CD maximum near 227 nm. This greater sensitivity may arise because the WW domain tertiary structure is intrinsically less stable than the VHP tertiary structure ($T_{m,app}$ = 55 °C vs. 44 °C), or because α-helical secondary structure is more tolerant of L-to-D substitution than is β-sheet secondary structure, or a combination of these factors.

Our finding that L-to-D substitution is generally destabilizing in two small tertiary structures offers an interesting perspective on previous studies showing that replacing a native Gly residue with a D residue can occasionally stabilize peptide or protein folding patterns.[28–33] In these cases, the native Gly residue displays backbone torsion angles that would be unfavorable for L residues but are favorable for D residues. For Gly, of course, backbone torsion angles preferred by L residues and those preferred by D residues are energetically equivalent. The use of D residues to promote "mirror image" reverse turns has proven useful in the design of β-sheet-forming peptides.[34–36]

Polypeptides containing mixtures of L and D residues can adopt distinctive conformations that enable unique functions. Gramicidin A, for example, features an alternation of L and D residues. This natural antibiotic can form a double-stranded β-helix that functions as an ion channel.[37] For ribosomally produced polypeptides, however, evolutionary selection has necessarily operated on homochiral molecules. The results reported here are consistent with earlier studies[16] in showing that evolved protein tertiary structures are generally destabilized by L-to-D modification. This trend may explain why mammals devote energy to sustaining α-amino acid homochirality.[38]

5.5 Materials, Methods and Supplementary Figures

5.5.1. Peptide synthesis, purification and characterization

All peptides were prepared via microwave-assisted solid-phase synthesis using a LibertyBlue instrument. Fmoc-protected amino acids were purchased from Chem-Impex. Phenylalanine and glycine preloaded Wang resins were purchased from CEM. Coupling reagents ethyl cyano(hydroxyimino)acetate (Oxyma) and N,N'-diisopropylcarbodiimide (DIC) were purchased from Chem-Impex. N,N-dimethylformamide (DMF), dichloromethane (DCM), piperidine, trifluoroacetic acid (TFA), triisopropylsilane (TIPS), thioanisole, 1,2-ethanedithiol (EDT), and acetonitrile were purchased from Sigma-Aldrich.

All peptides were synthesized on a 50 µmole scale with 0.2 M Oxyma in DMF and 0.2 M DIC in DMF. Pre-loaded Wang resin was pre-swelled in 1:1 v/v DCM:DMF. Fmoc-amino acids (1.25 mL, 0.2 M) were activated with DIC (0.5 mL, 0.5 M) and Oxyma (0.5 mL, 1 M) in biotech grade DMF. Regular coupling cycles were performed at 90°C for 2 minutes. Double coupling at 90 °C was performed for all D-amino acid and arginine residues. The histidine residue was coupled at 50°C to prevent possible racemization.

After all residues had been added, the resin was mixed with the cleavage cocktail (8 mL per 50 µmole) containing 91.5% (v/v) TFA, 2.5% (v/v) H_2O, 2.5% (v/v) EDT, 2.5% (v/v) thioanisole, and 1% (v/v) TIPS at room temperature for 3.5 hours for VHP analogues or 5 hours for WW domain analogues. Excess TFA was removed under a stream of nitrogen, and 35 mL of cold diethyl ether was then added to precipitate the crude peptide. The peptide was isolated by centrifugation, washed with cold ether twice, and then dried under nitrogen.

For VHP peptide analogs, ~25 µmole of the crude product was dissolved in 600 µL 1:1 H_2O:acetonitrile and filtered through a 0.22 µm Nylon filter for reversed-phase HPLC purification.

VHP analogs were purified using a C18 HPLC column on an Agilent HPLC. H_2O and acetonitrile with 0.1 % TFA were used as solvents A and B, respectively. A gradient from 10% to 60% B over 20 minutes with a flow rate of 12 mL/minute was used. For WW domain analogs, ~25 µmole of the crude peptide was dissolved in 1500 µL 1:1 H_2O:acetonitrile and then filtered for reversed-phase HPLC purification. WW domain analogs were first purified using a C8 column on an Agilent HPLC using 10% to 60% B over 20 minutes with a flow rate of 12 mL/minute. For the second purification, the WW domain analogs were purified with a C8 column using 20-50% B over 20 minutes with flow rate = 14 mL/ min. Peptide masses were determined using MALDI-TOF-MS, and purity was checked using UPLC (detection at 220 nm).

5.5.2. Circular Dichroism (CD) Spectroscopy

CD experiments were conducted at 25 °C on a JASCO J-1500 CD spectrometer. The VHP peptide analogs were diluted to 50 µM using 20 mM sodium acetate buffer (pH = 5.0). The WW domain analogs were diluted to 50 µM using 10 mM sodium phosphate buffer (pH = 7.0). The peptide sample was then transferred to a 1-mm quartz cuvette. The CD spectrum was measured from 260 to 200 nm with 0.1 nm intervals at 4 second digital integration time (D.I.T) at 100 nm/min scanning speed. The data are presented as mean residue ellipticity in deg•cm2•M-1 calculated based on the equation shown below.

molar ellipticity (θ) =(mdeg value)/(32.98*concentration (mol/L)*path length (cm))

5.5.3. Thermal denaturation assays

A minimum at 223 nm was observed in the CD spectrum of each VHP analogue at room temperature, and a maximum at 227 nm was observed in the CD spectrum for each WW domain analogue at room temperature. Data for the thermal denaturation assays were collected at these two wavelengths. Thermal denaturation measurements were made from 5 to 95 °C with 2 °C/min increments with a 5-minute equilibration time at each temperature. The samples were held at 95 °C for 5 minutes before the temperature was decreased to 5 °C by 2 °C/min steps.

Each thermal denaturation data set was fitted into a 2-state folding model using the method Sigmoidal, 4PL, X is log(concentration) on GraphPad Prism 10. Previous study has shown that the folding thermodynamics of VHP can be accurately modeled as a 2-state system.[39] The measured parameter in the thermal denaturation assays is defined as $T_{m,app}$ for apparent melting temperature.

5.5.4. Additional data for CD thermal denaturation assays

VHP_E4e

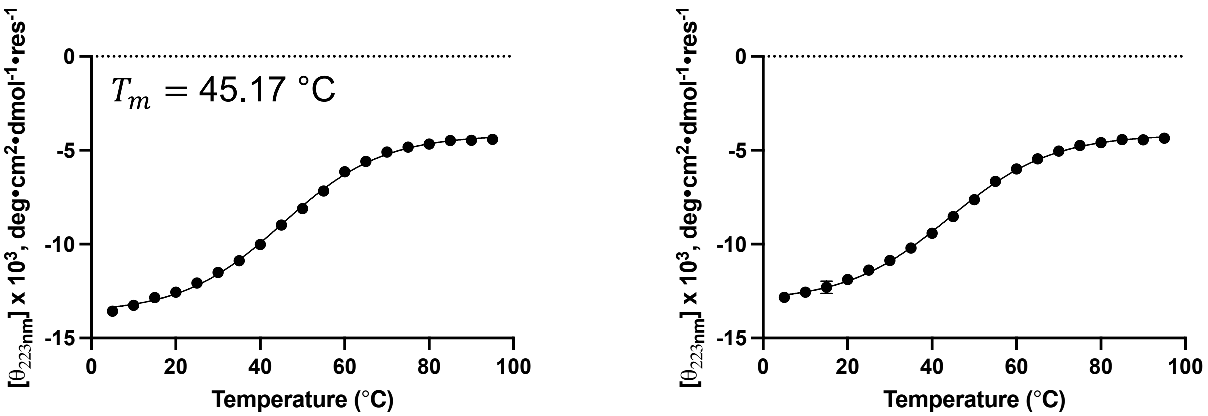


Figure 5.6. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP E4e mutant. Each data set was fit to a two-state folding model.

VHP_A8a

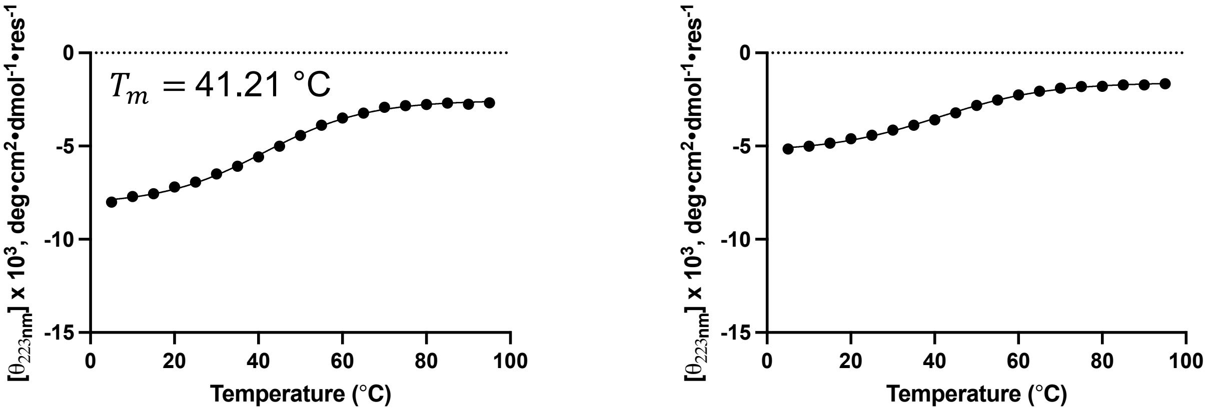


Figure 5.7. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP A8a mutant. Each data set was fit to a two-state folding model.

VHP_S15s

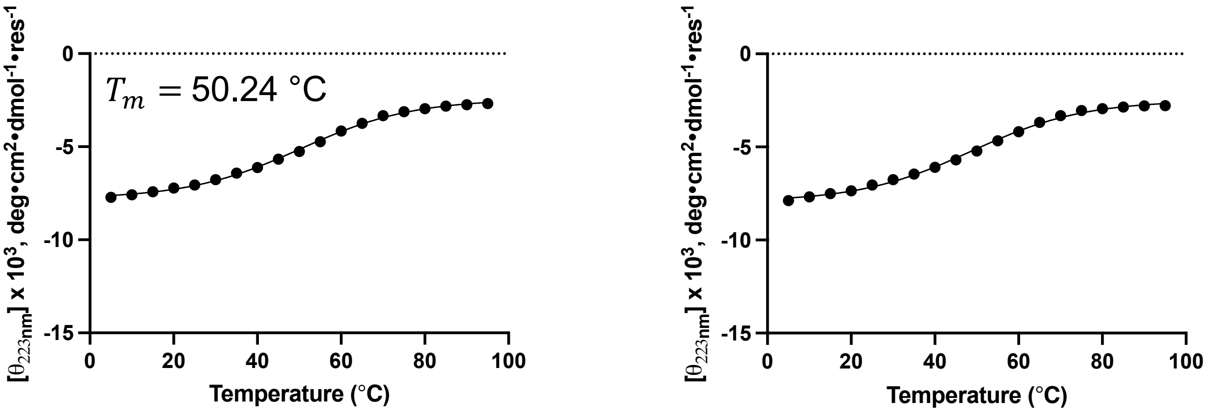


Figure 5.8. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP S15s mutant. Each data set was fit to a two-state folding model.

VHP_N19n

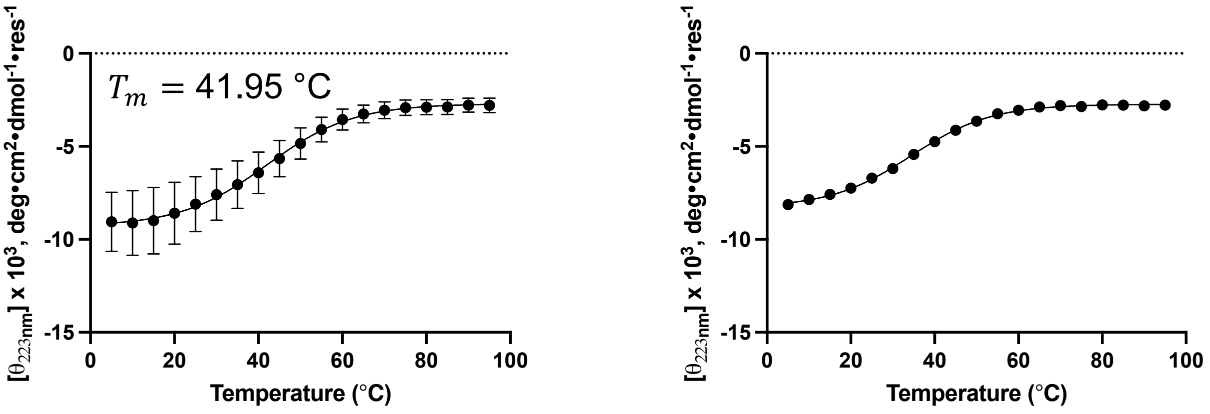


Figure 5.9. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP N19n mutant. Each data set was fit to a two-state folding model.

VHP_W23w

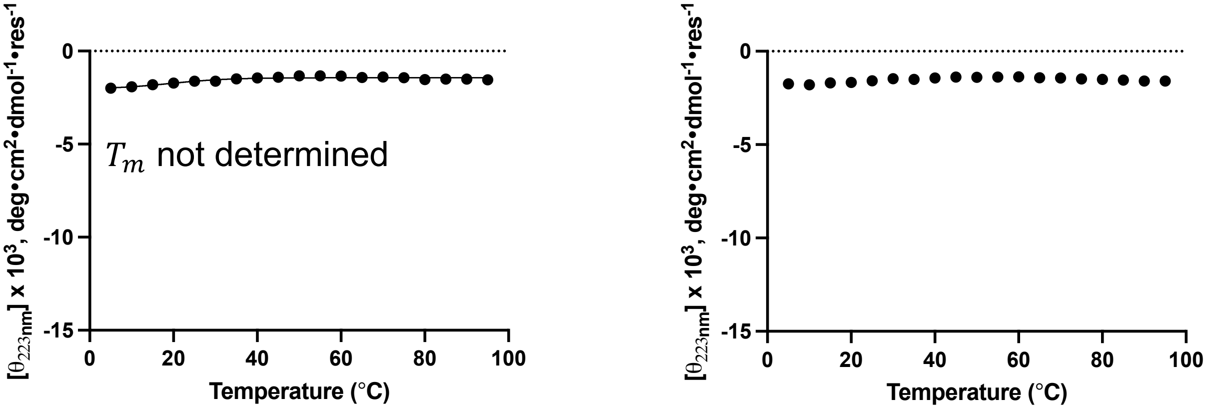


Figure 5.10. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP W23w mutant. Each data set was fit to a two-state folding model.

VHP_K24k

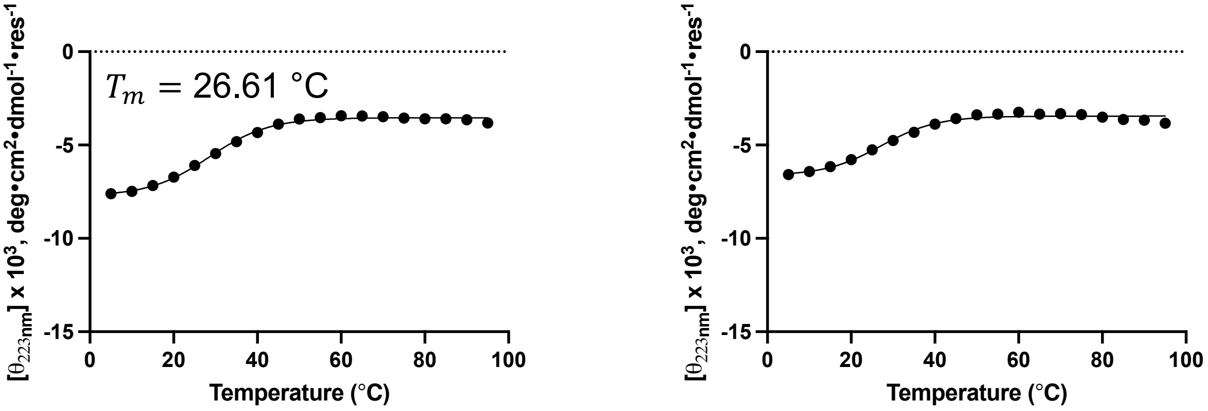


Figure 5.11. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP K24k mutant. Each data set was fit to a two-state folding model.

VHP_Q26q

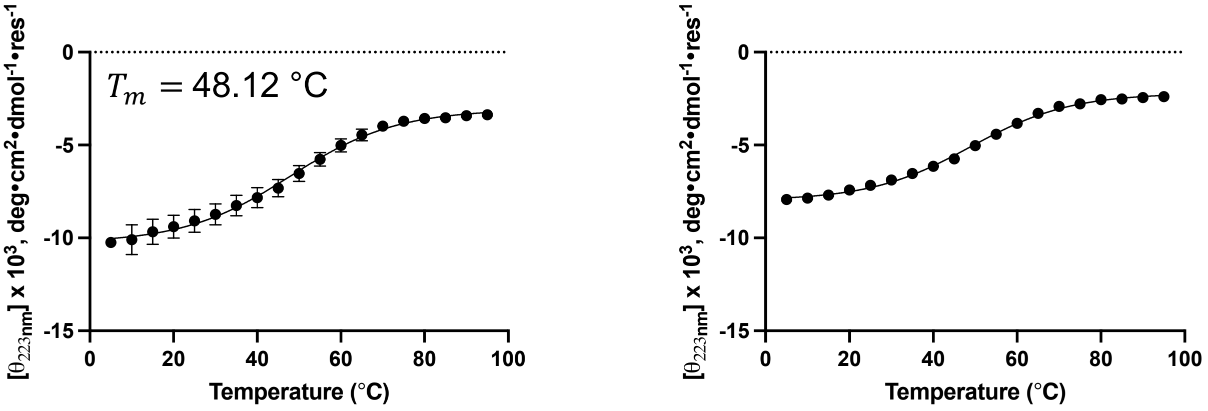


Figure 5.12. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP Q26q mutant. Each data set was fit to a two-state folding model.

VHP_H27h

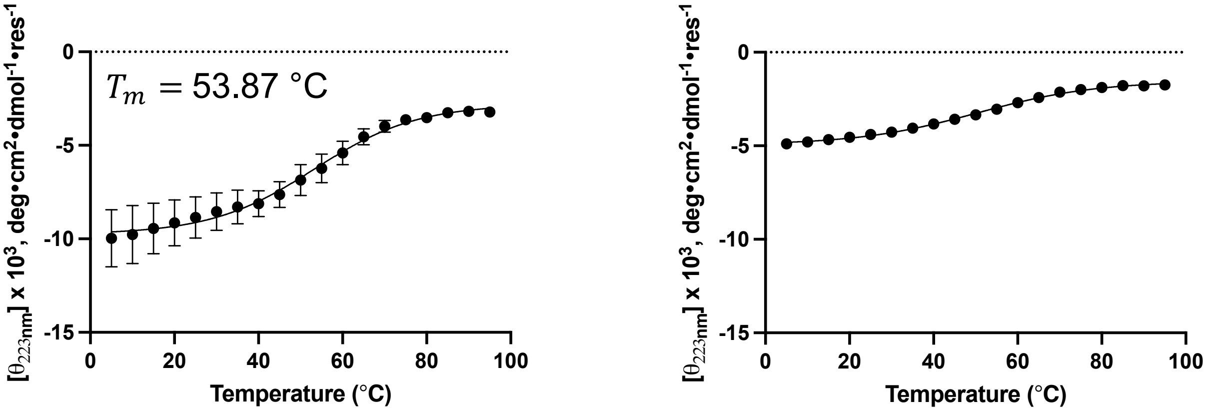


Figure 5.13. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP H27h mutant. Each data set was fit to a two-state folding model.

VHP_K30k

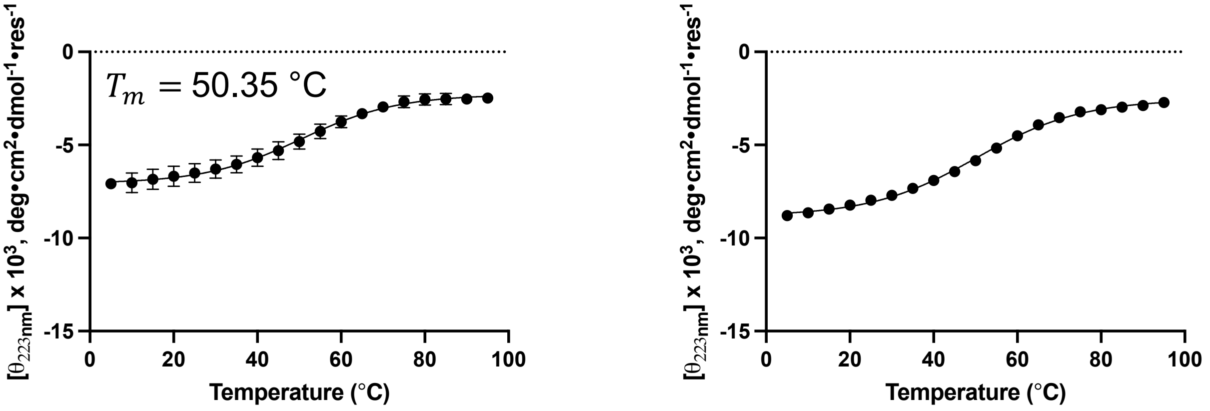


Figure 5.14. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP K30k mutant. Each data set was fit to a two-state folding model.

VHP_E31e

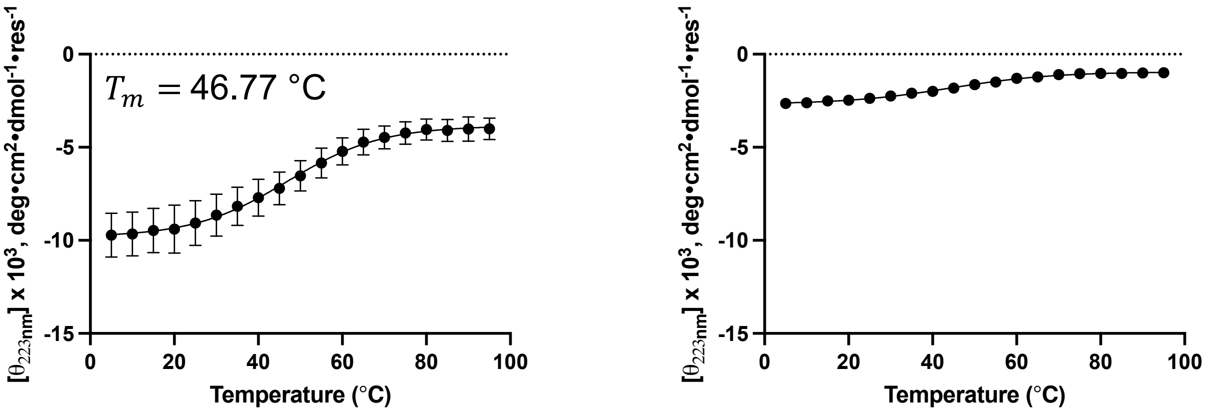


Figure 5.15. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP E31e mutant. Each data set was fit to a two-state folding model.

VHP_H27h_S15s

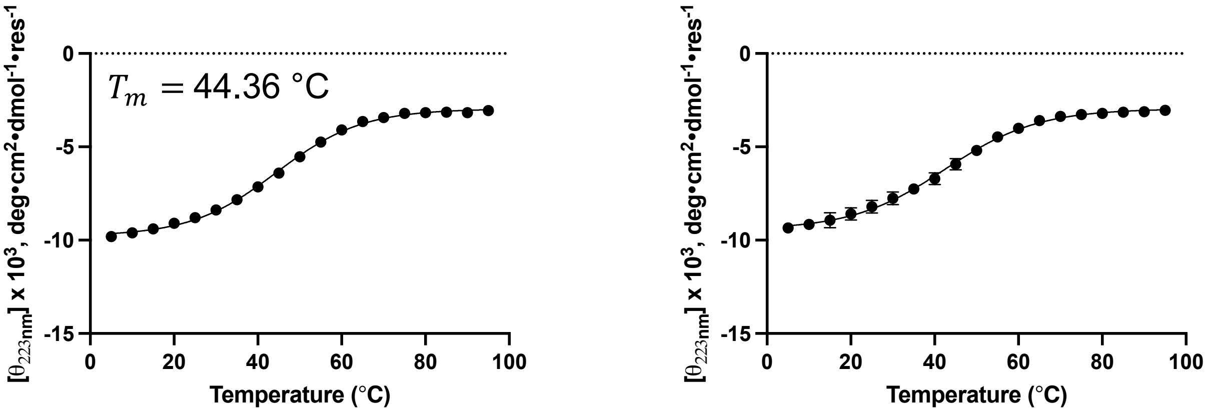


Figure 5.16. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP H27h S15s mutant. Each data set was fit to a two-state folding model.

VHP_H27h_K24k

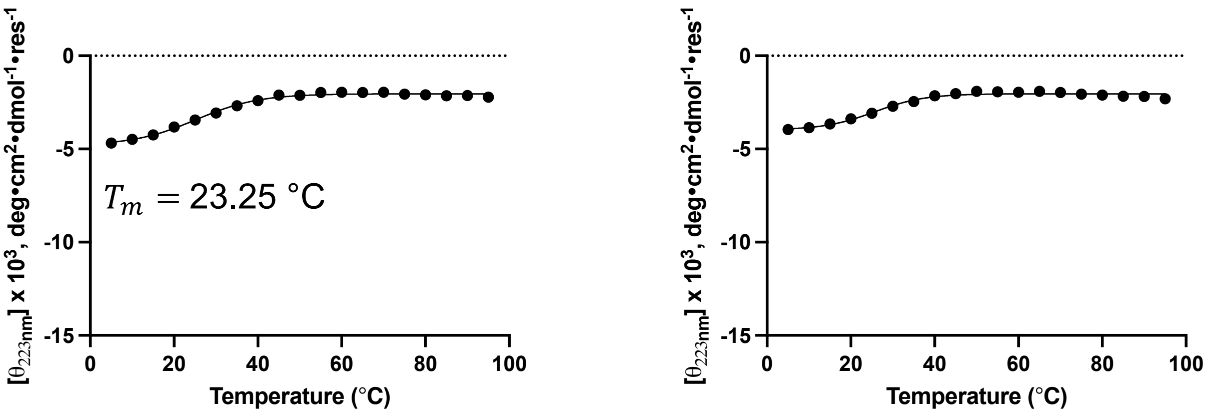


Figure 5.17. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP H27h K24k mutant. Each data set was fit to a two-state folding model.

VHP_H27h_K30k

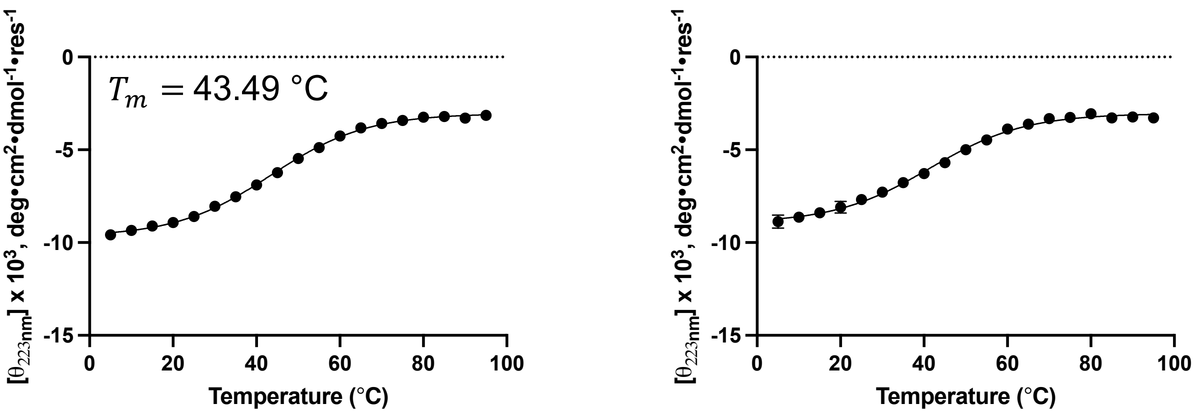


Figure 5.18. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP H27h K24k mutant. Each data set was fit to a two-state folding model.

VHP_H27h_A8a

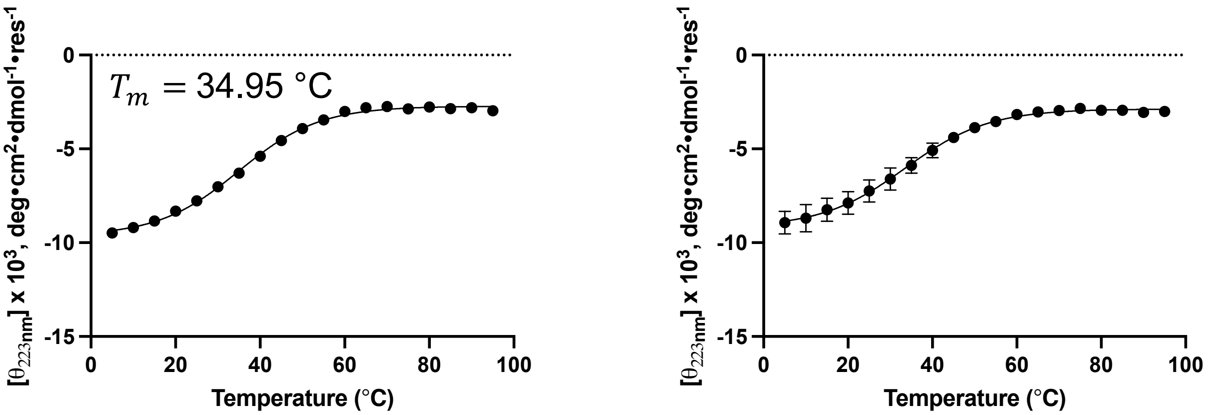


Figure 5.19. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP H27h A8a mutant. Each data set was fit to a two-state folding model.

VHP_A8a_S15s

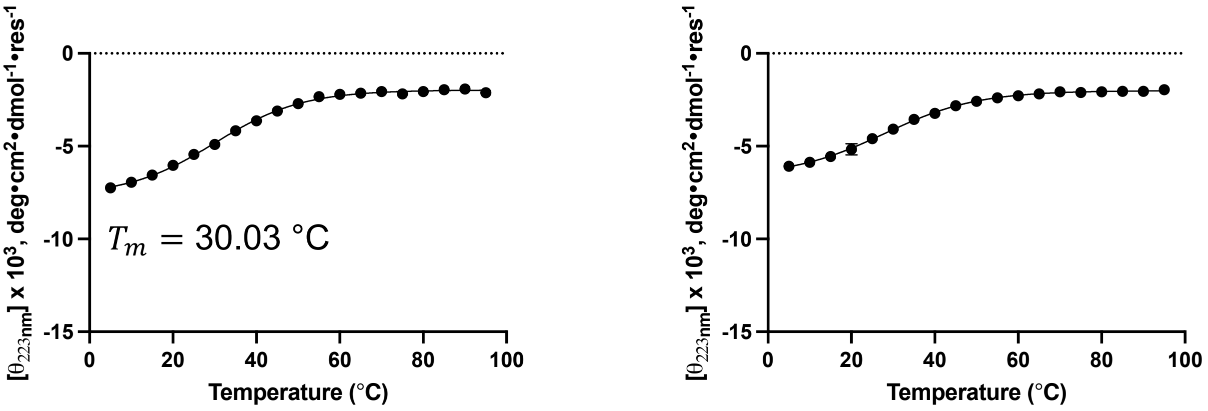


Figure 5.20. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP A8a S15s mutant. Each data set was fit to a two-state folding model.

VHP_A8a_N19n

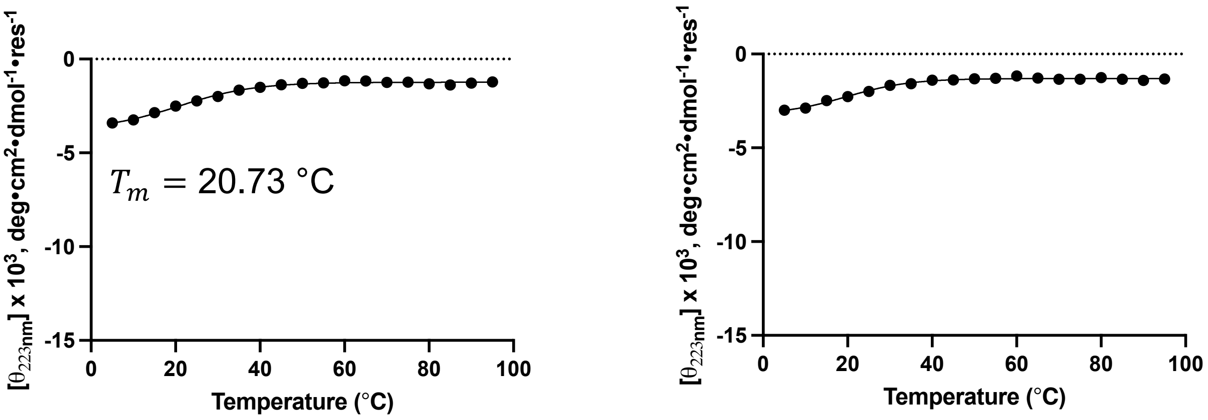


Figure 5.21. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP A8a N19n mutant. Each data set was fit to a two-state folding model.

VHP_S15s_N19n

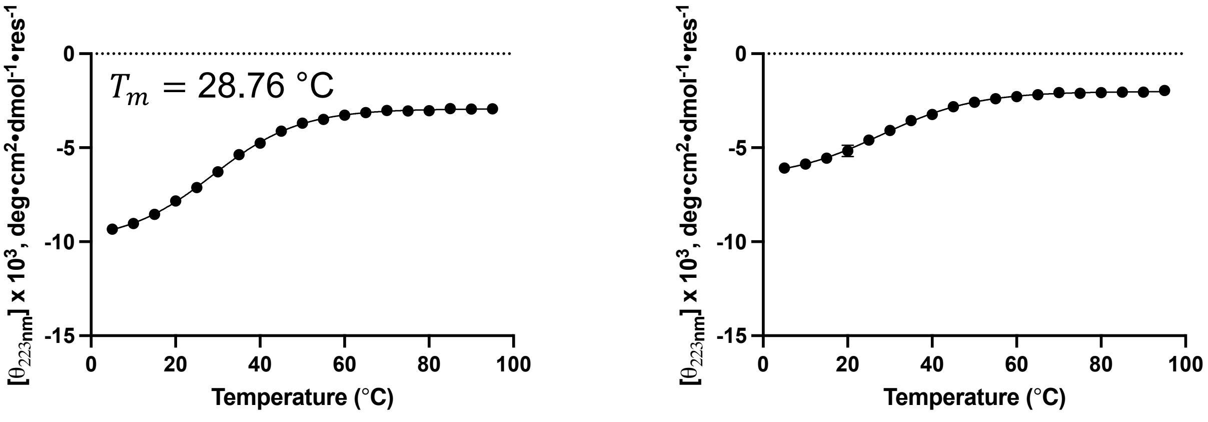


Figure 5.22. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP S15s N19n mutant. Each data set was fit to a two-state folding model.

VHP_H27h_N19n

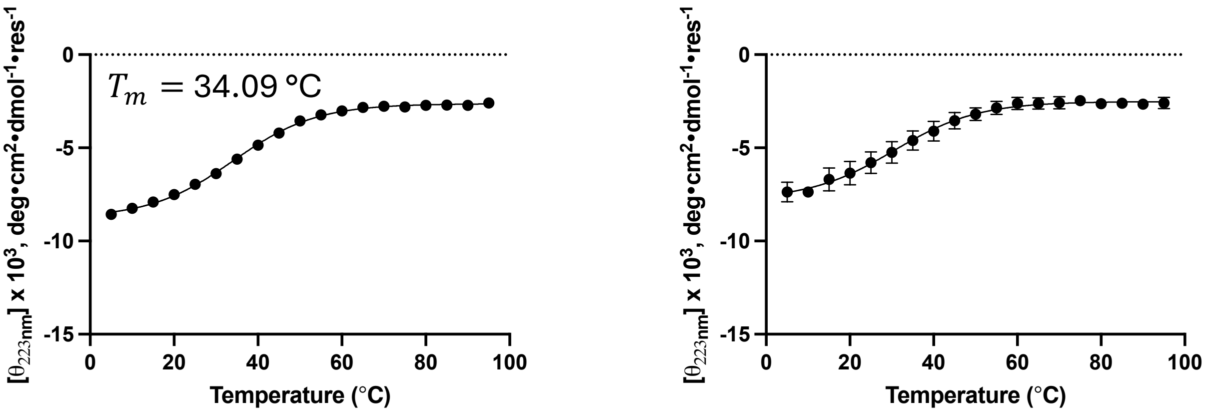


Figure 5.23. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP H27h N19n mutant. Each data set was fit to a two-state folding model.

VHP_A8a_S15s_H27h

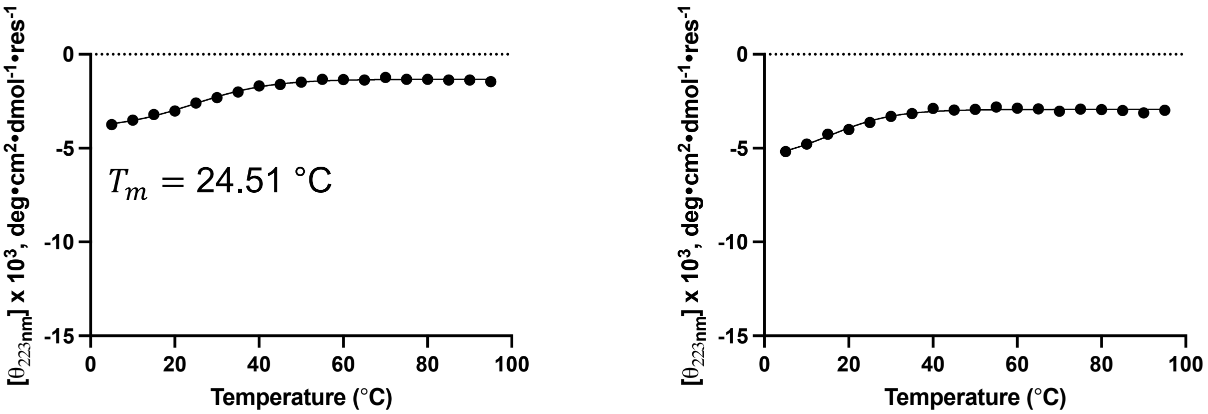


Figure 5.24. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP A8a S15s H27h mutant. Each data set was fit to a two-state folding model.

VHP_A8a_N19n_H27h

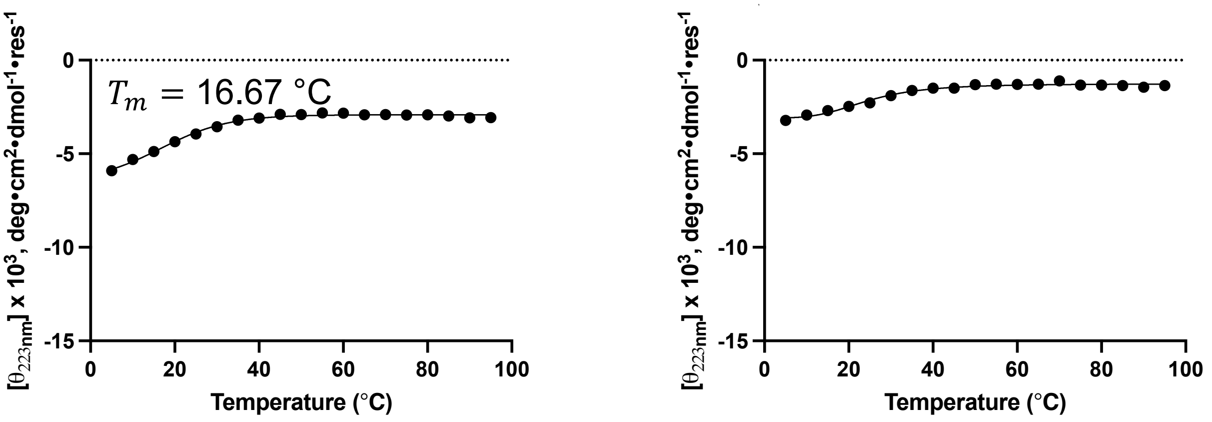


Figure 5.25. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP A8a N19n H27h mutant. Each data set was fit to a two-state folding model.

VHP_A8a_S15s_N19n_H27h

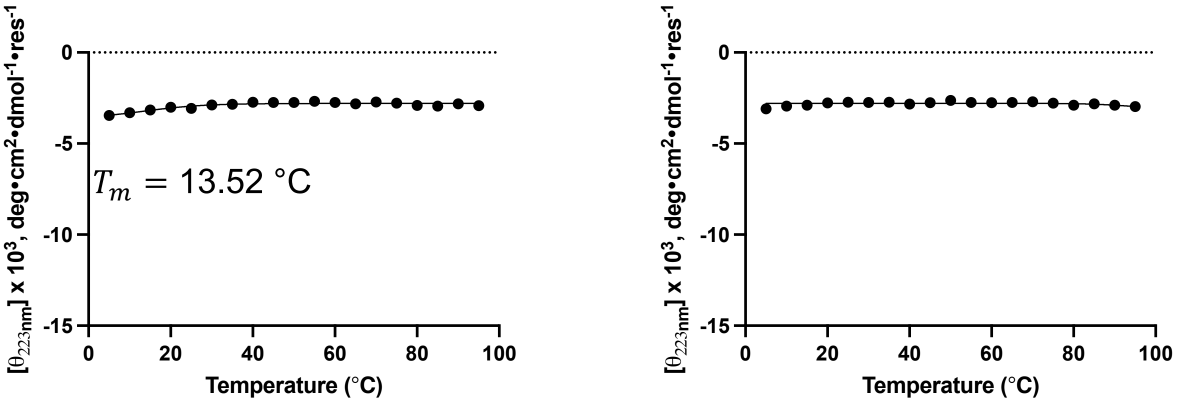


Figure 5.26. CD data obtained for increasing temperature (left) and decreasing temperature (right) for VHP A8a S15s N19n H27h mutant. Each data set was fit to a two-state folding model.

WW

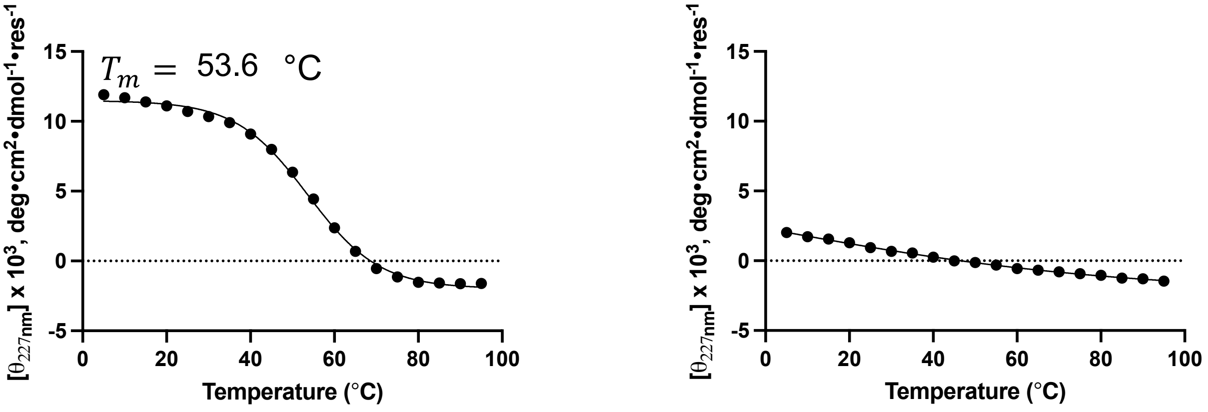


Figure 5.27. CD data obtained for increasing temperature (left) and decreasing temperature (right) for WW. Each data set was fit to a two-state folding model.

WW_R16r

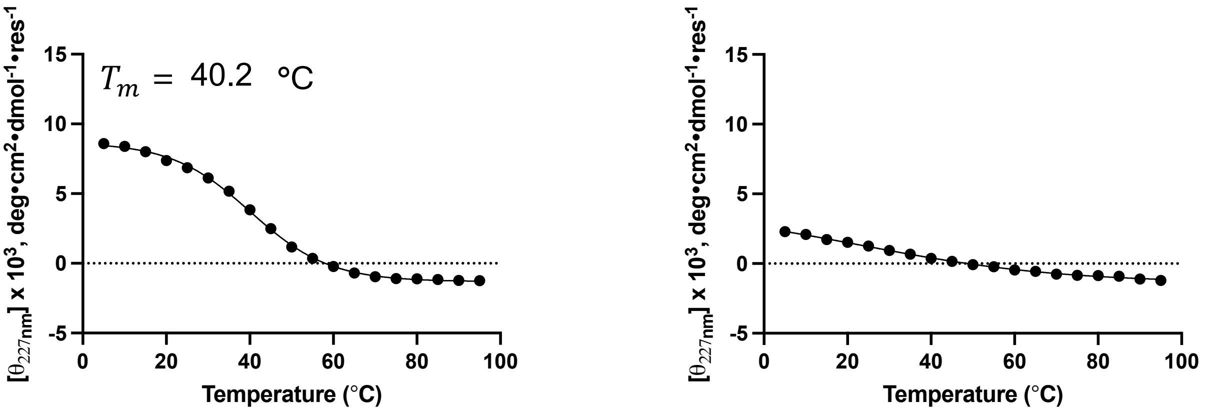


Figure 5.28. CD data obtained for increasing temperature (left) and decreasing temperature (right) for WW R16r mutant. Each data set was fit to a two-state folding model.

WW_Q28q

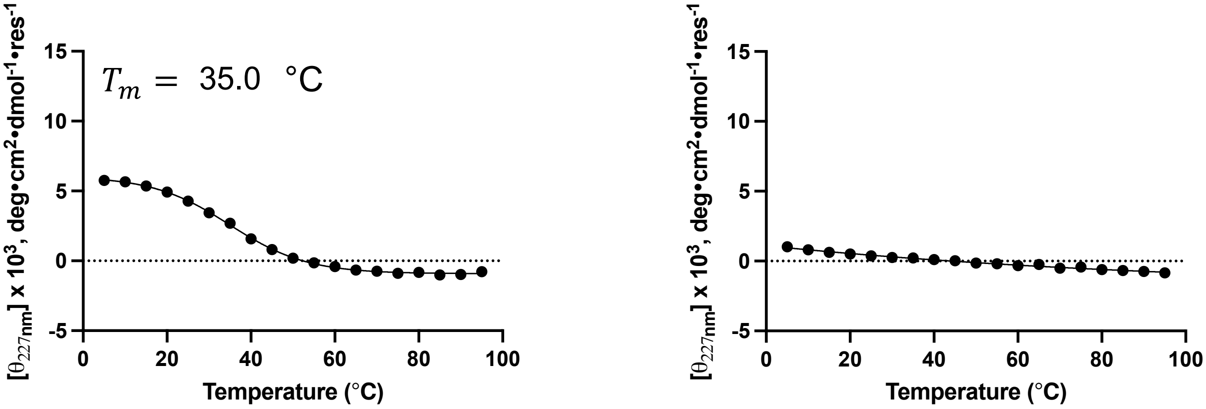
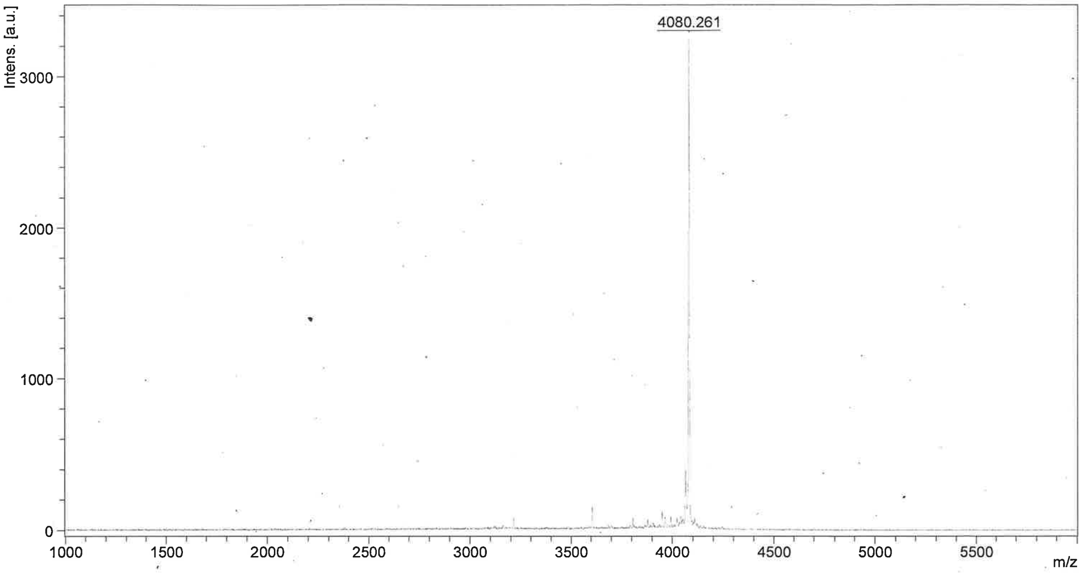


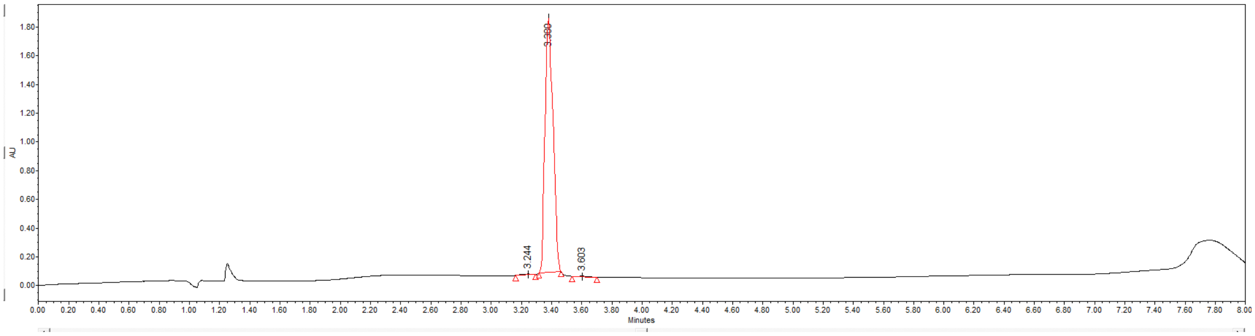
Figure 5.29. CD data obtained for increasing temperature (left) and decreasing temperature (right) for WW Q28q mutant. Each data set was fit to a two-state folding model.

5.6. MALDI-TOF-MS Spectra and UPLC Data for Synthetic Peptides

VHP

Sequence: **NH₂ — LSDEDFKAVFGMTRSAFANLPLWKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4079.135$

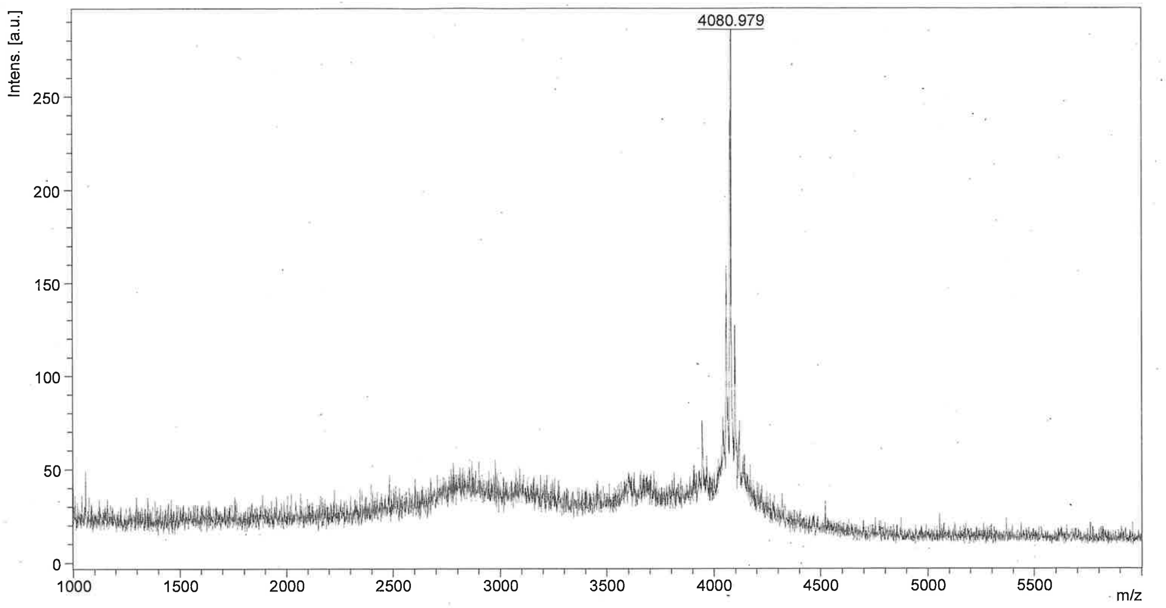
observed monoisotopic $[M+H]^+ = 4080.261$

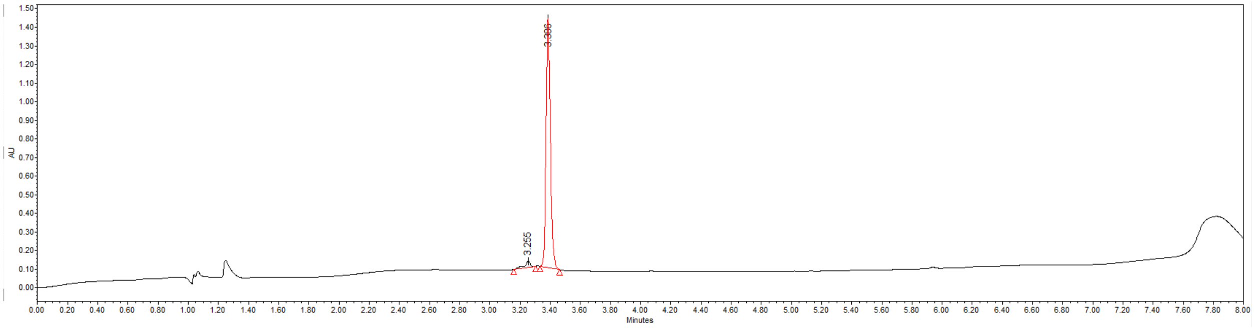
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.47 %

VHP_E4e

Sequence: **NH₂ — LSDeDFKAVFGMTRSAFANLPLWKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4079.135$

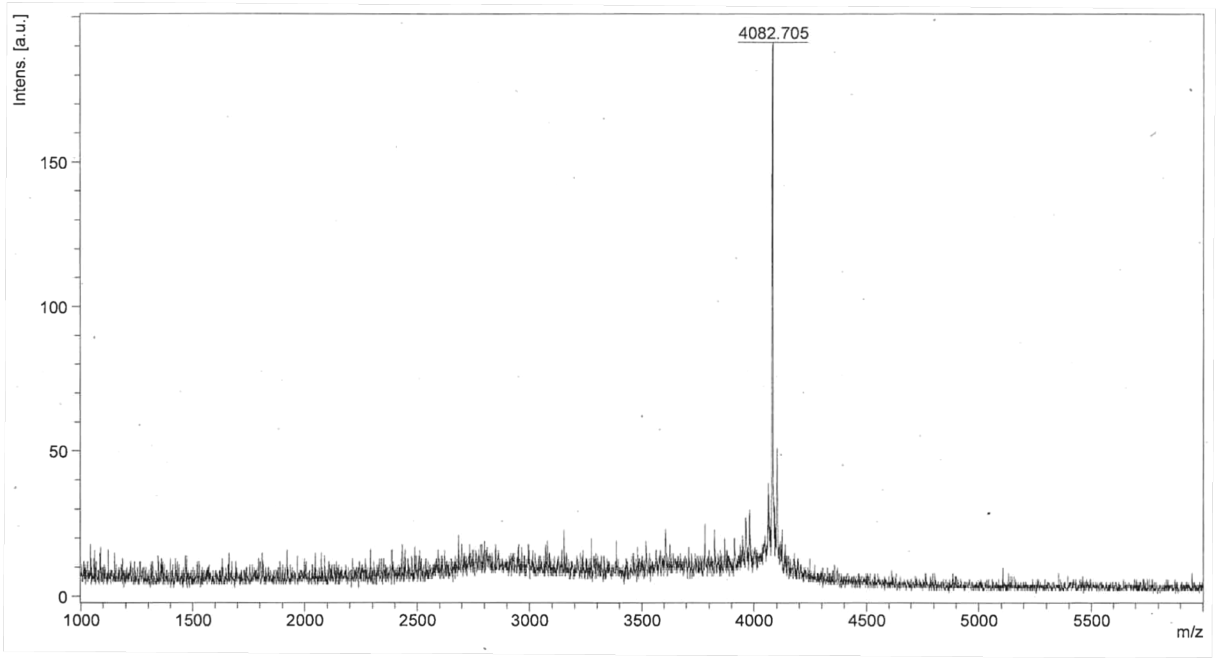
observed monoisotopic $[M+H]^+ = 4080.979$

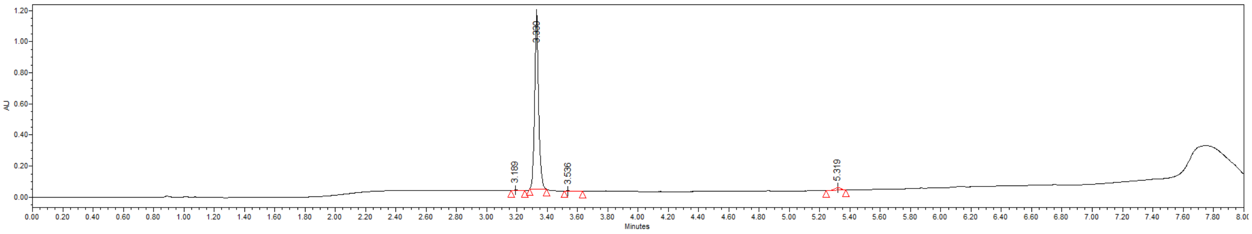
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 96.82 %

VHP_A8a

Sequence: **NH₂ — LSDEDFKaVFGMTRSAFANLPLWKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

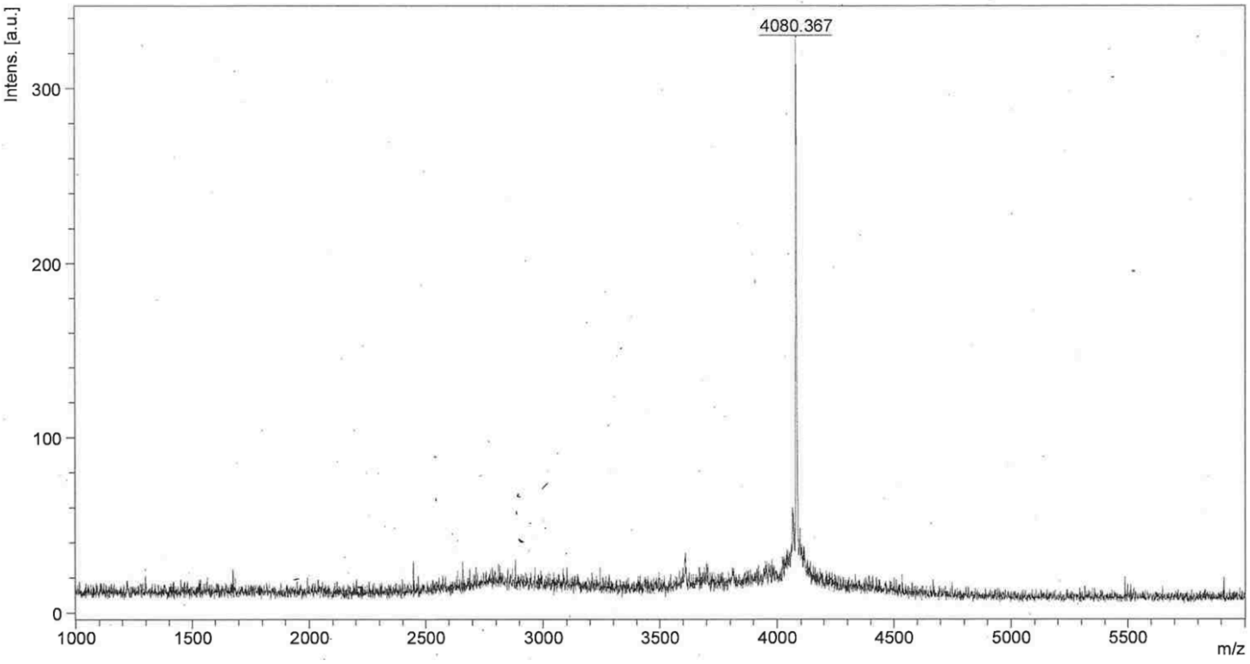
observed monoisotopic $[M+H]^+$ = 4082.705

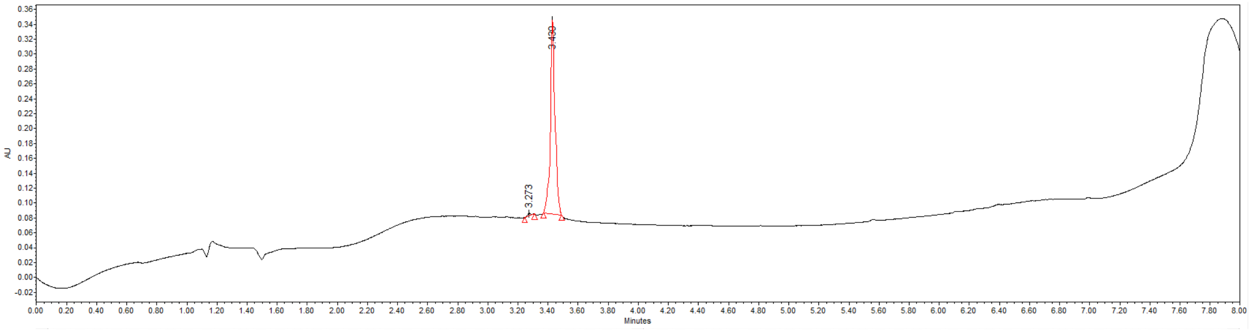
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 97.24 %

VHP_S15s

Sequence: **NH_2 — LSDEDFKAVFGMTRsAFANLPLWKQQHLKKEKGLF — COOH**





MALDI-TOF-MS: calculated monoisotopic $[M+H]^+$ = 4079.135

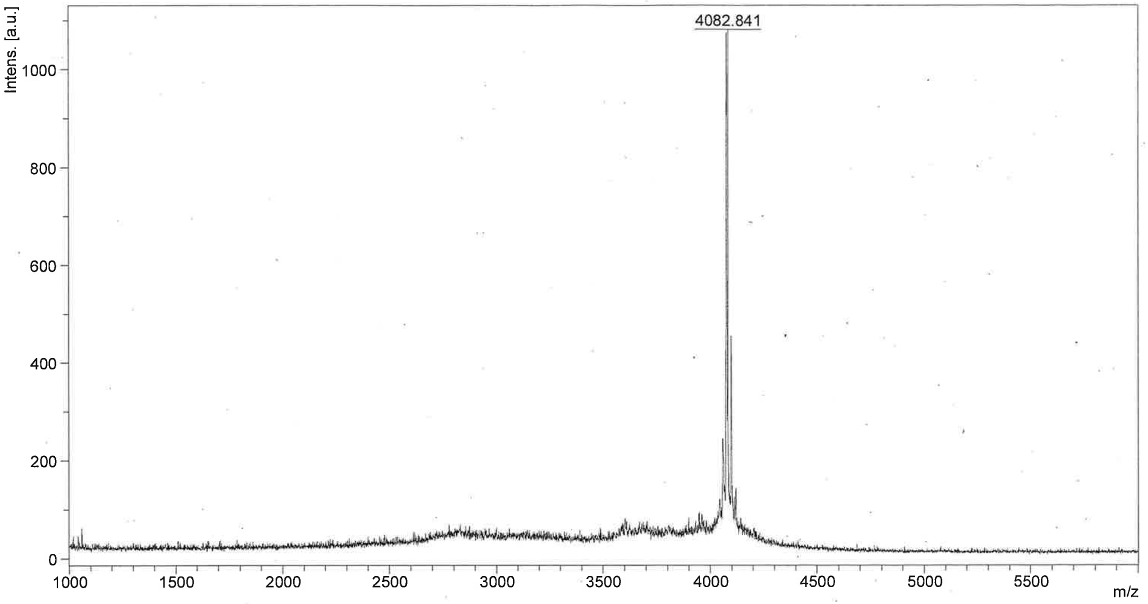
observed monoisotopic $[M+H]^+$ = 4080.367

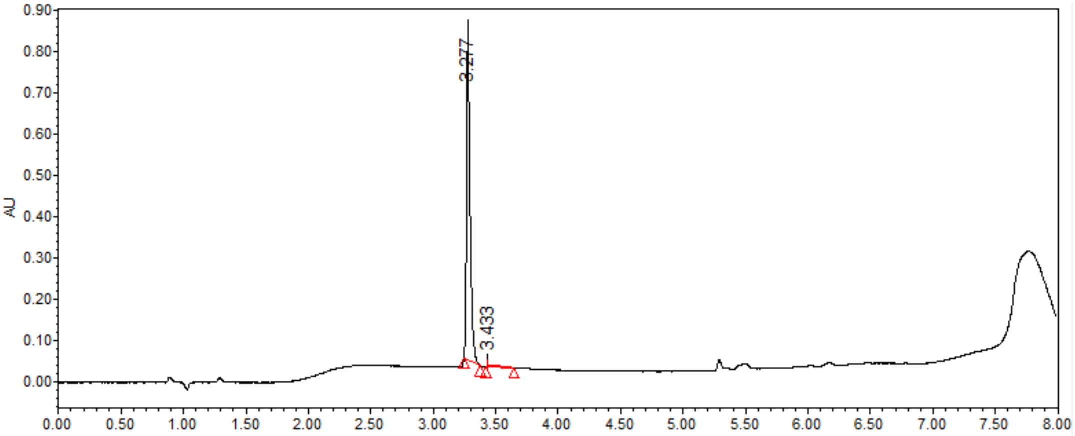
UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.07 %

VHP_N19n

Sequence: **NH$_2$ — LSDEDFKAVFGMTRSAFAnLPLWKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

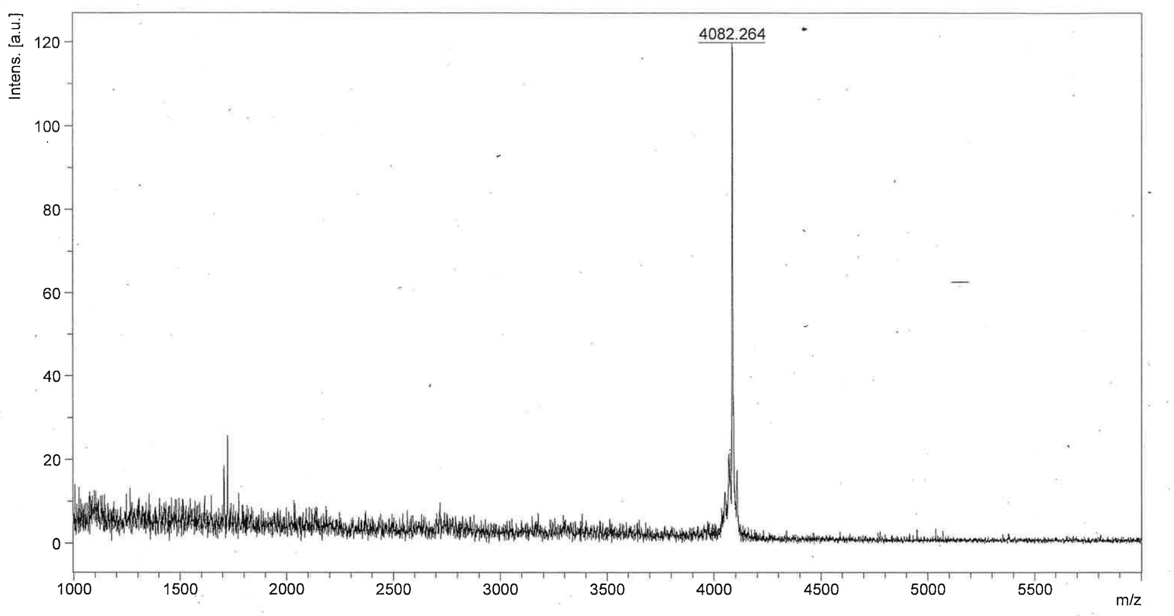
observed monoisotopic $[M+H]^+$ = 4082.841

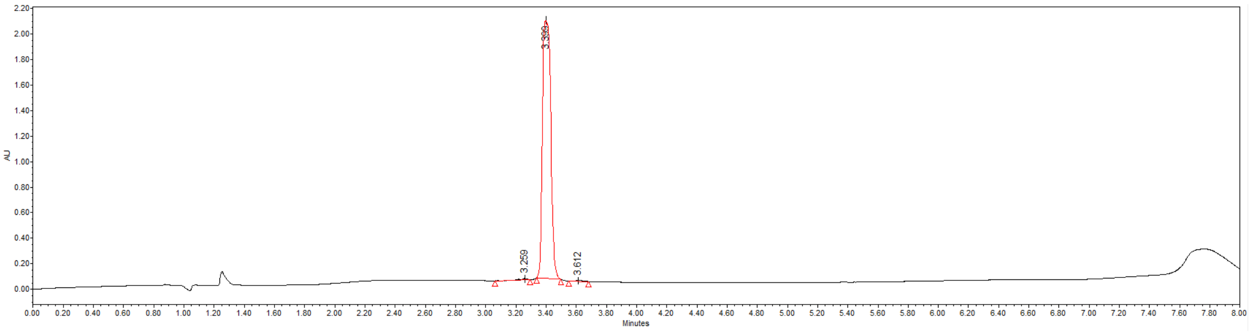
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 96.71 %

VHP_W23w

Sequence: **NH₂ — LSDEDFKAVFGMTRSAFANLPLwKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

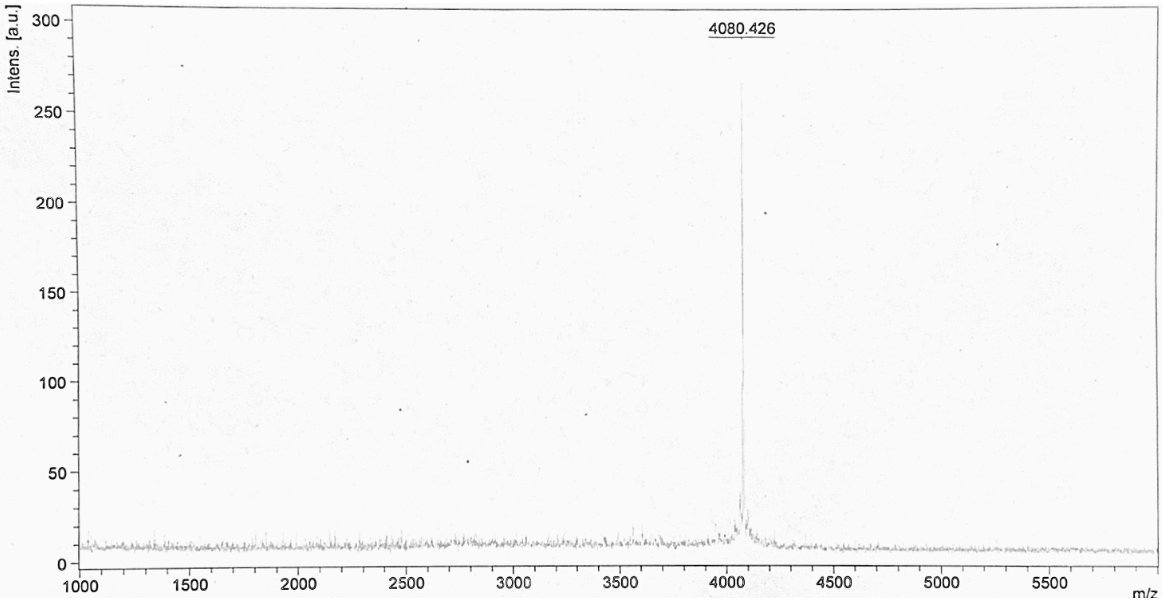
observed monoisotopic $[M+H]^+$ = 4082.264

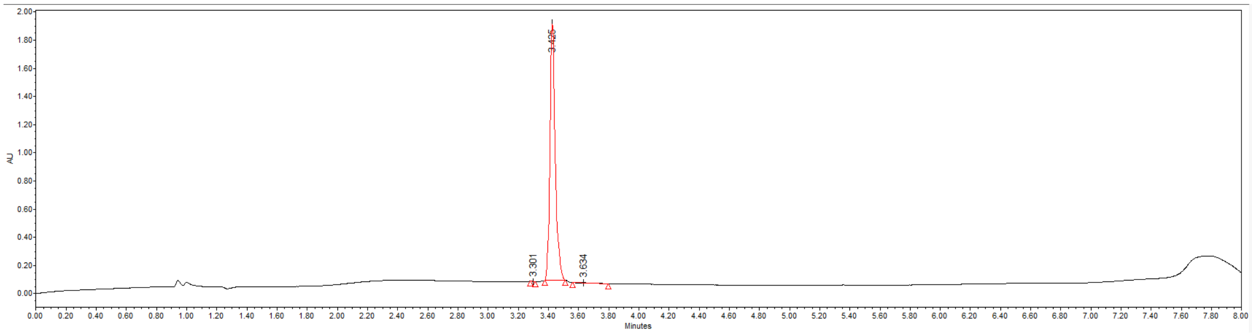
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.16 %

VHP_K24k

Sequence: **NH$_2$ — LSDEDFKAVFGMTRSAFANLPLWkQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

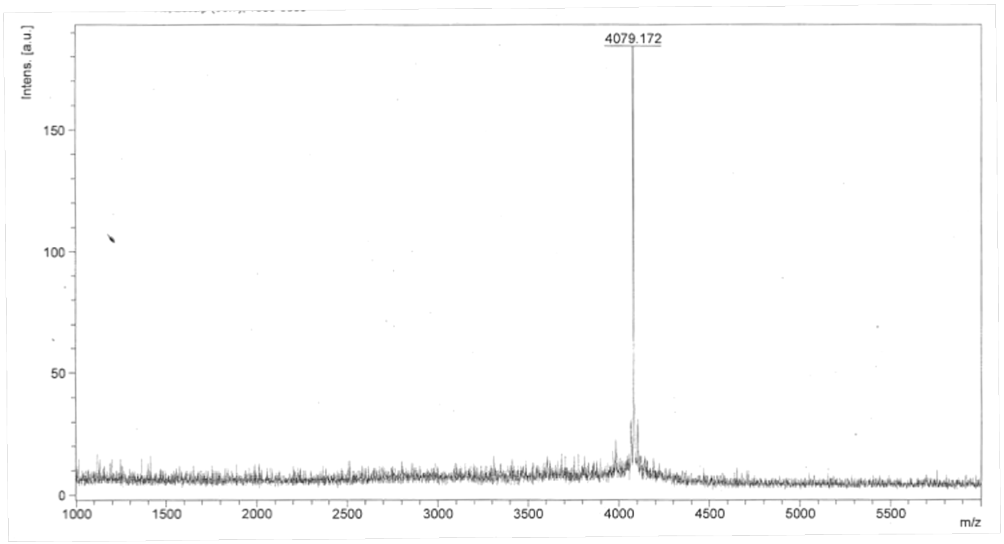
observed monoisotopic $[M+H]^+$ = 4080.426

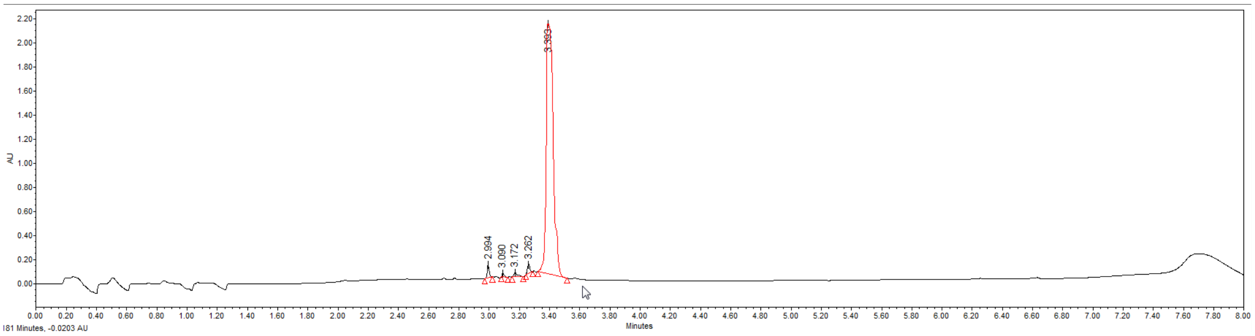
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.66 %

VHP_Q26q

Sequence: **NH₂ — LSDEDFKAVFGMTRSAFANLPLWKQqHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

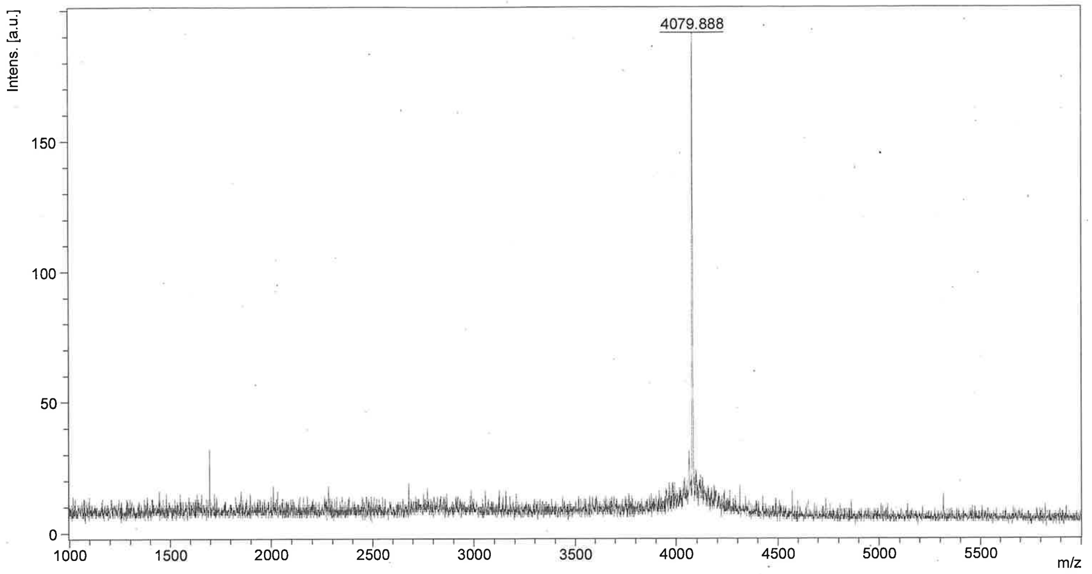
observed monoisotopic $[M+H]^+$ = 4079.172

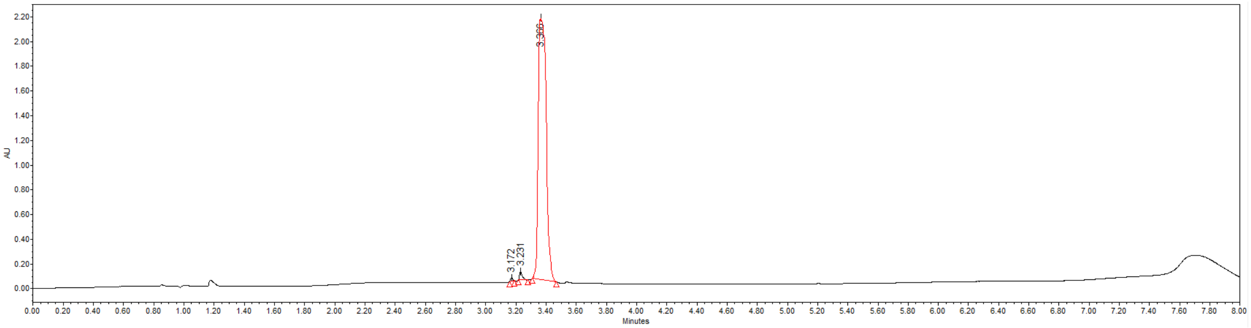
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 96.11 %

VHP_H27h

Sequence: **NH₂ — LSDEDFKAVFGMTRSAFANLPLWKQQhLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4079.135$

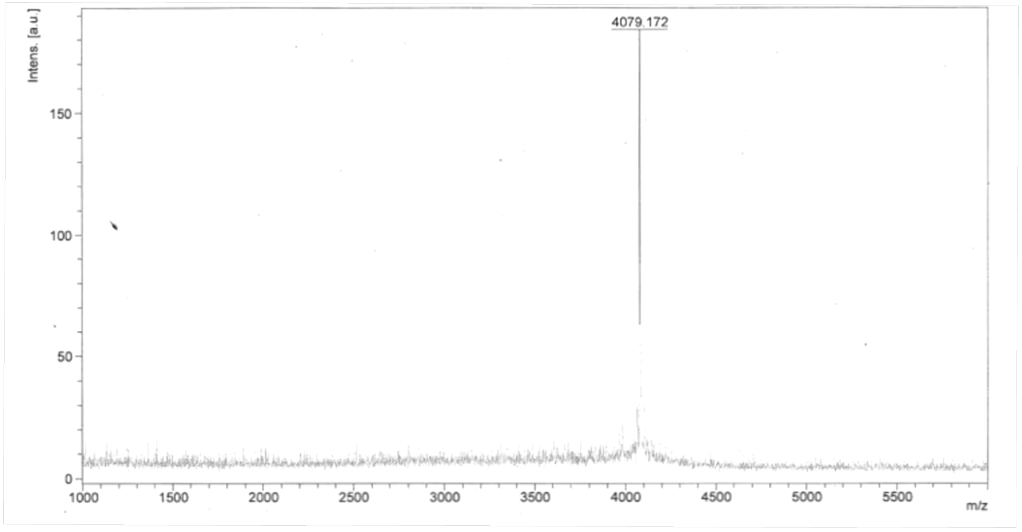
observed monoisotopic $[M+H]^+ = 4079.888$

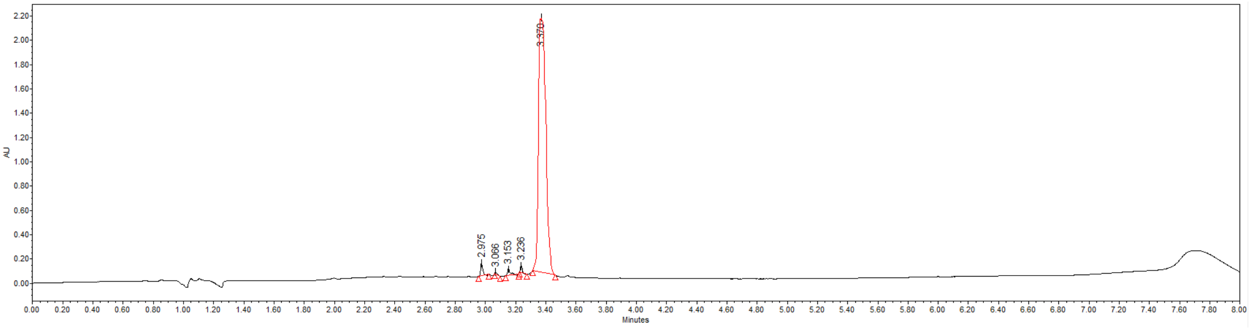
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 98.53 %

VHP_K30k

Sequence: **NH_2 — LSDEDFKAVFGMTRSAFANLPLWKQQHLKkEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4079.135$

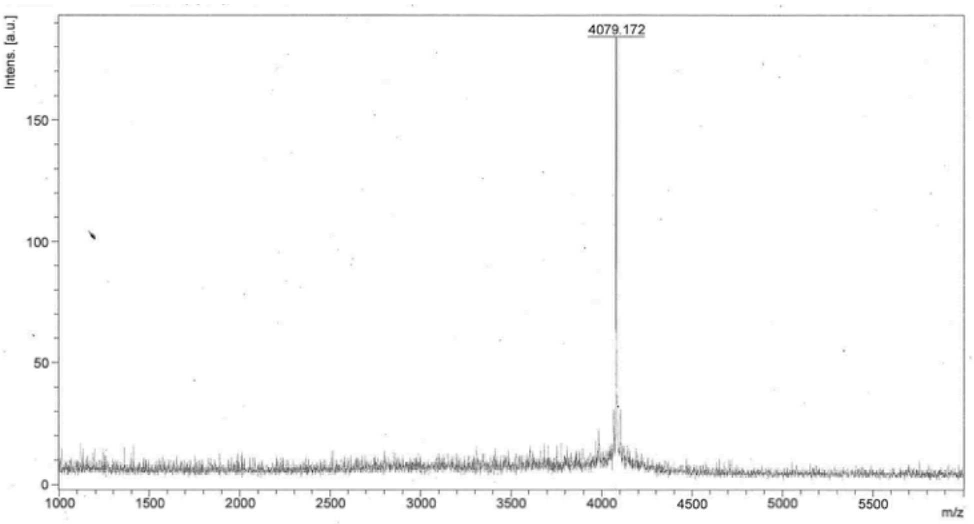
observed monoisotopic $[M+H]^+ = 4079.172$

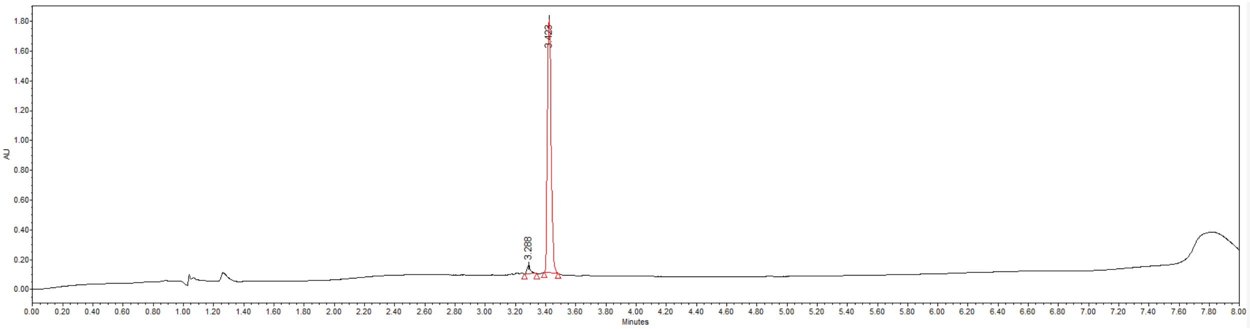
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 96.72 %

VHP_E31e

Sequence: **NH_2 — LSDEDFKAVFGMTRSAFANLPLWKQQHLKKeKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

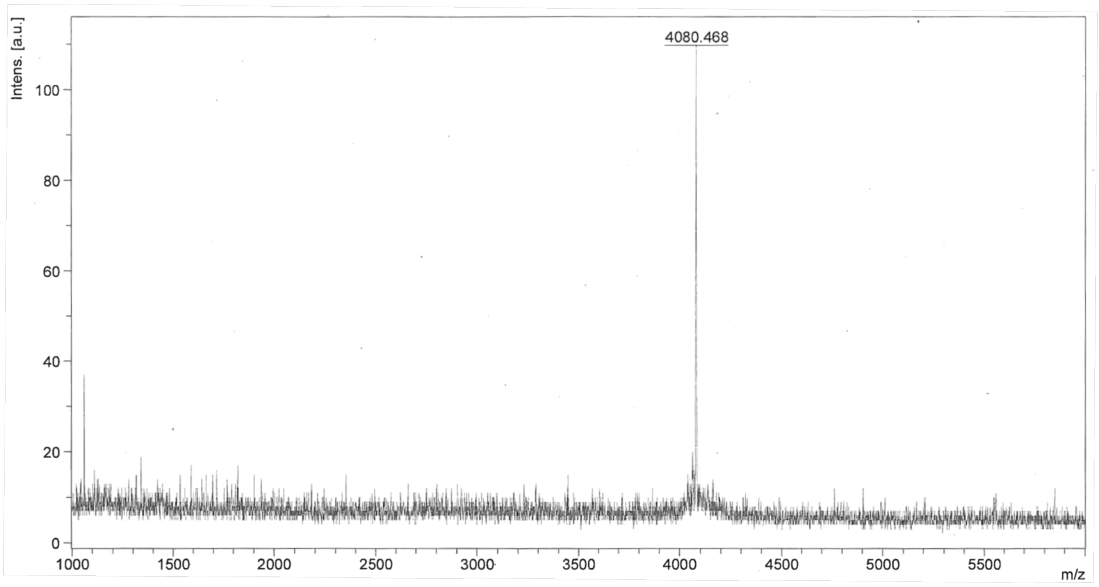
observed monoisotopic $[M+H]^+$ = 4079.172

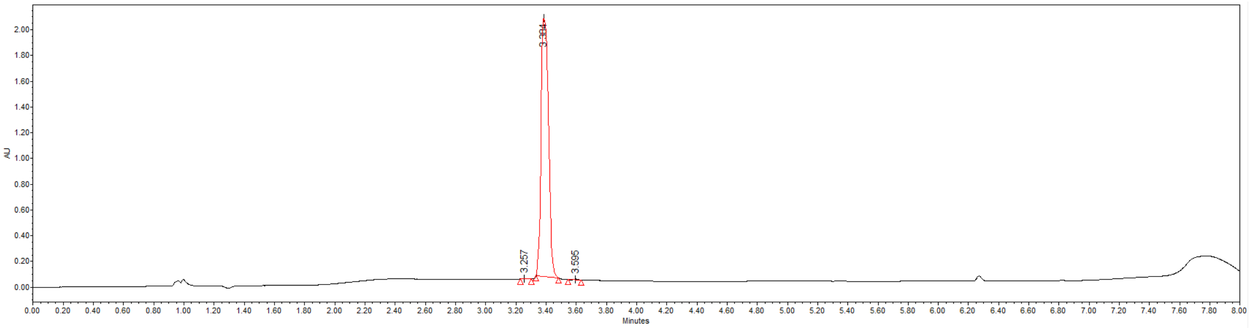
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 97.15 %

VHP_H27h_S15s

Sequence: **NH$_2$ — LSDEDFKAVFGMTRsAFANLPLWKQQhLKKEKGLF — COOH**





MALDI-TOF-MS: calculated monoisotopic $[M+H]^+ = 4079.135$

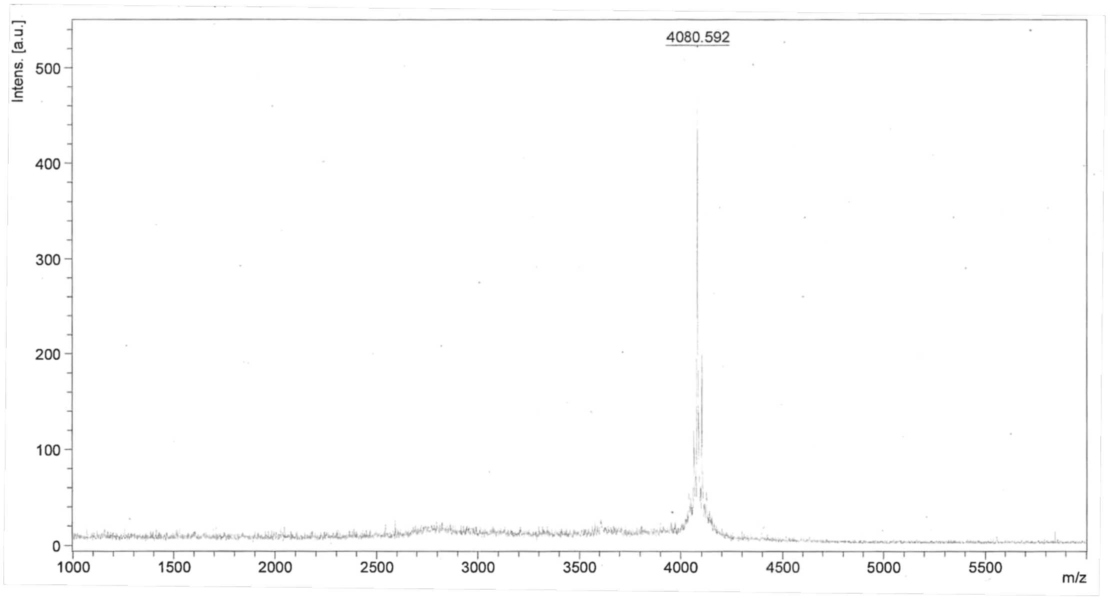
observed monoisotopic $[M+H]^+ = 4080.468$

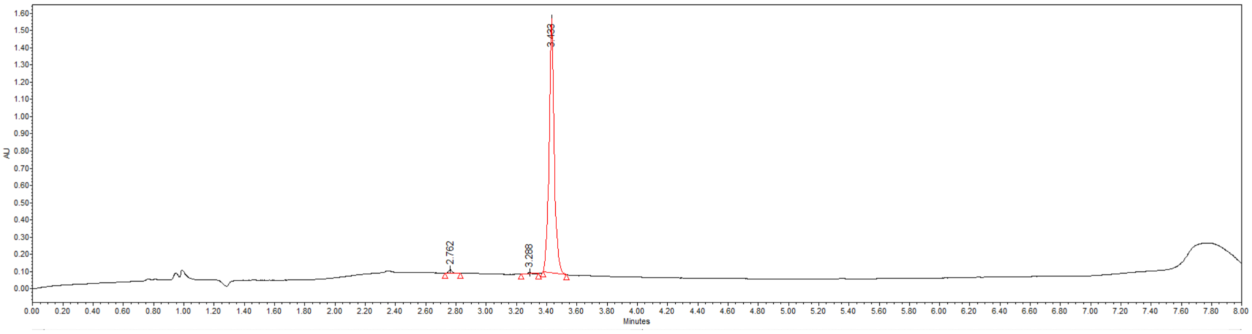
UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.68 %

VHP_H27h_K24k

Sequence: **NH₂ — LSDEDFKAVFGMTRSAFANLPLWkQQhLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4079.135$

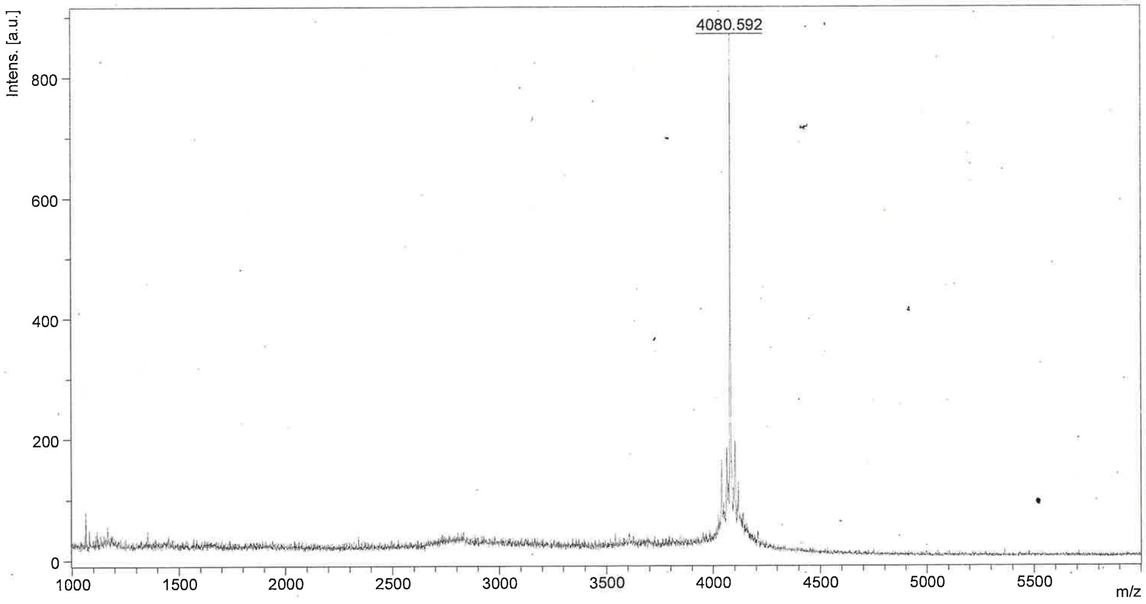
observed monoisotopic $[M+H]^+ = 4080.592$

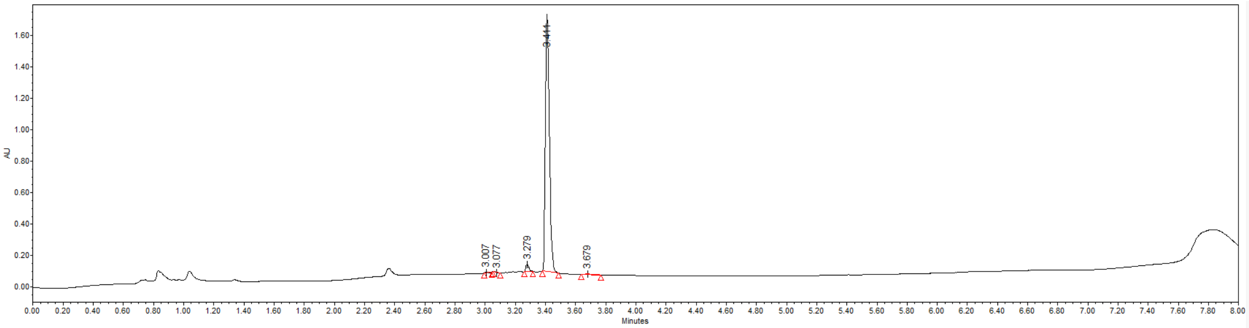
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 98.69 %

VHP_H27h_K30k

Sequence: **NH₂ — LSDEDFKAVFGMTRSAFANLPLWKQQhLKkEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

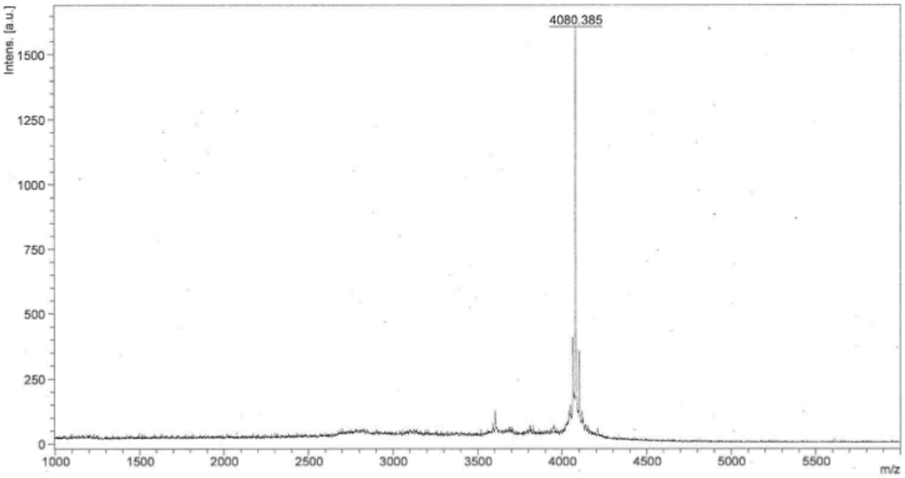
observed monoisotopic $[M+H]^+$ = 4080.592

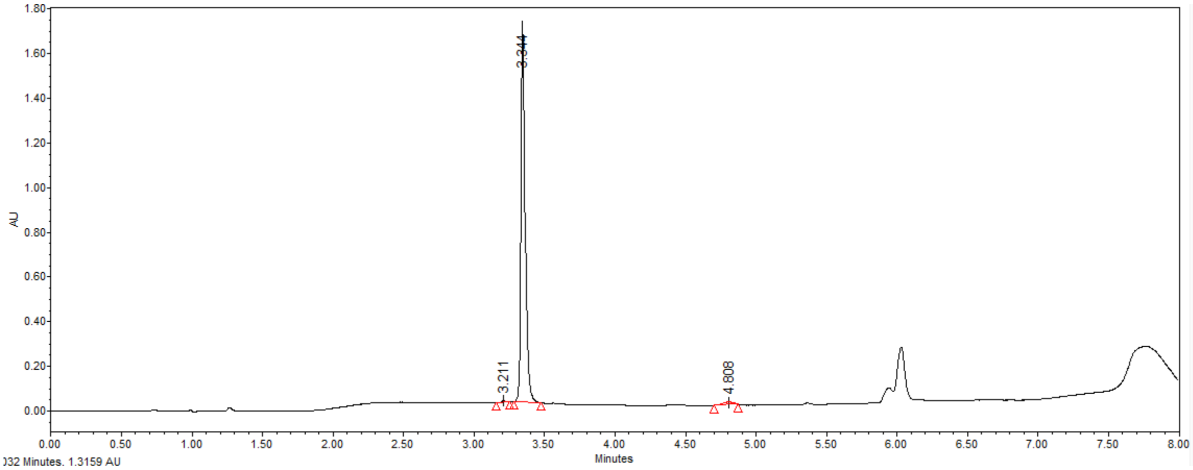
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 97.16 %

VHP_A8a_H27h

Sequence: **NH$_2$ — LSDEDFKaVFGMTRSAFANLPLWKQQhLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

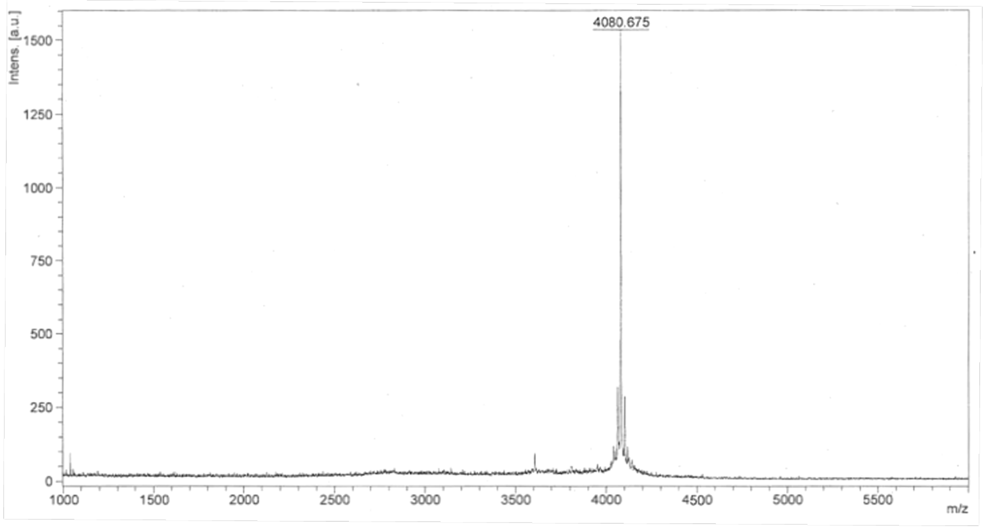
observed monoisotopic $[M+H]^+$ = 4080.385

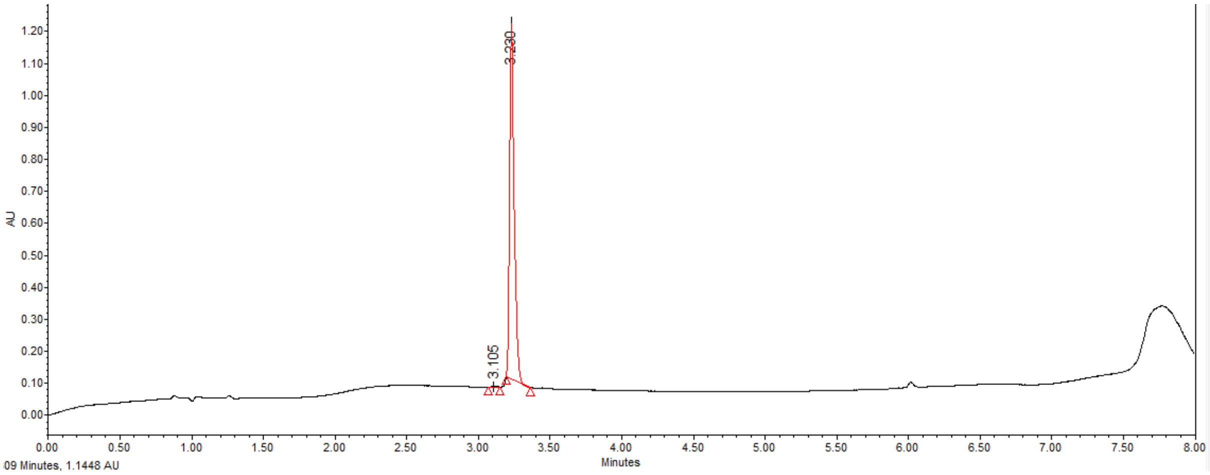
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 98.01 %

VHP_A8a_S15s

Sequence: **NH_2 — LSDEDFKaVFGMTRsAFANLPLWKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

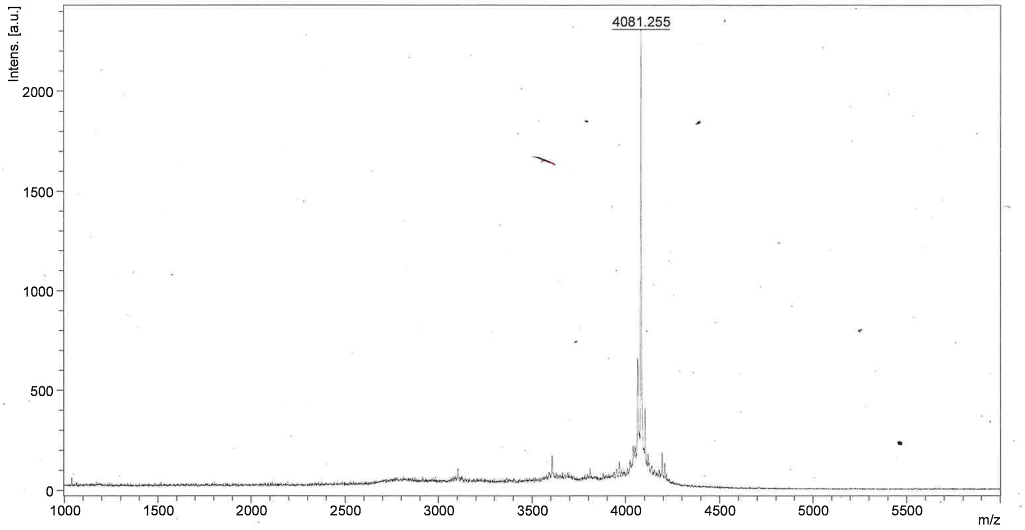
observed monoisotopic $[M+H]^+$ = 4080.675

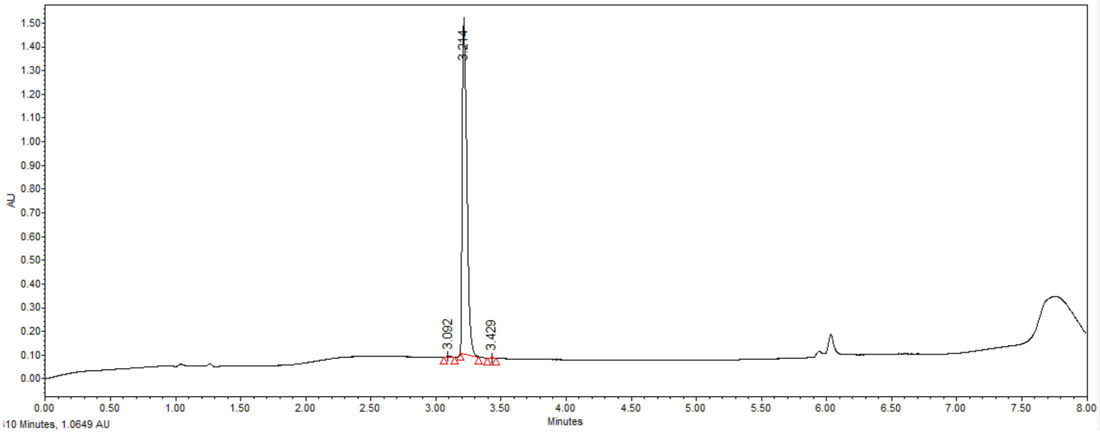
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.79 %

VHP_A8a_N19n

Sequence: **NH_2 — LSDEDFKaVFGMTRSAFAnLPLWKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic [M+H]$^+$ = 4079.135

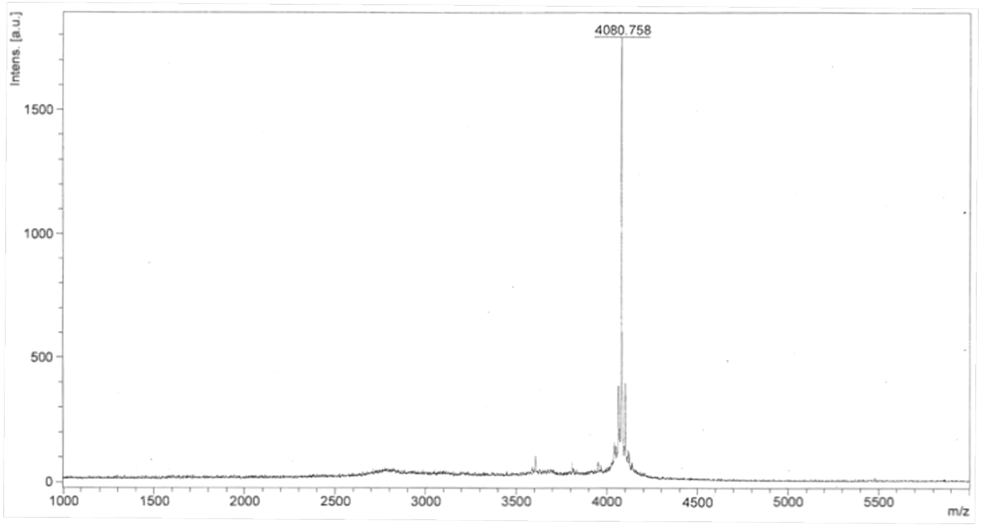
observed monoisotopic [M+H]$^+$ = 4081.255

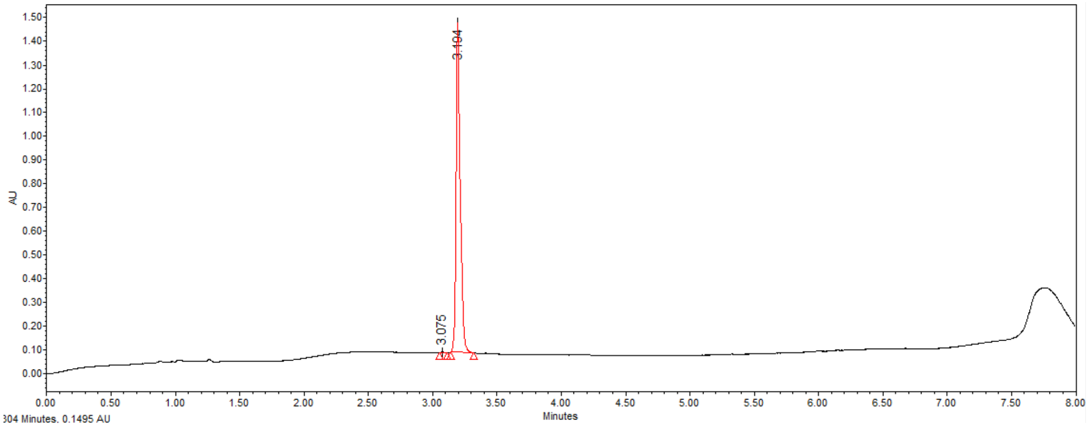
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.56 %

VHP_S15s_N19n

Sequence: **$\mathrm{NH_2}$ — LSDEDFKAVFGMTRsAFAnLPLWKQQHLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4079.135$

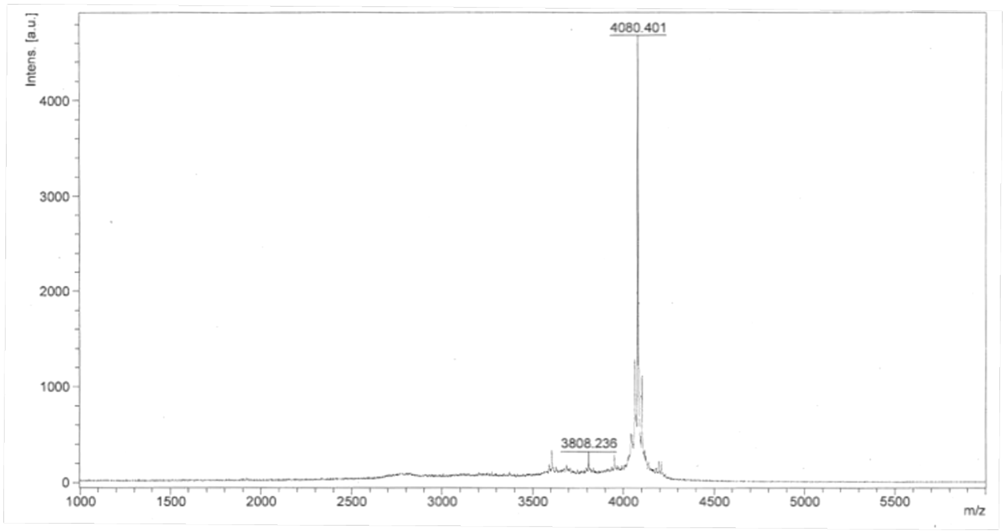
observed monoisotopic $[M+H]^+ = 4080.758$

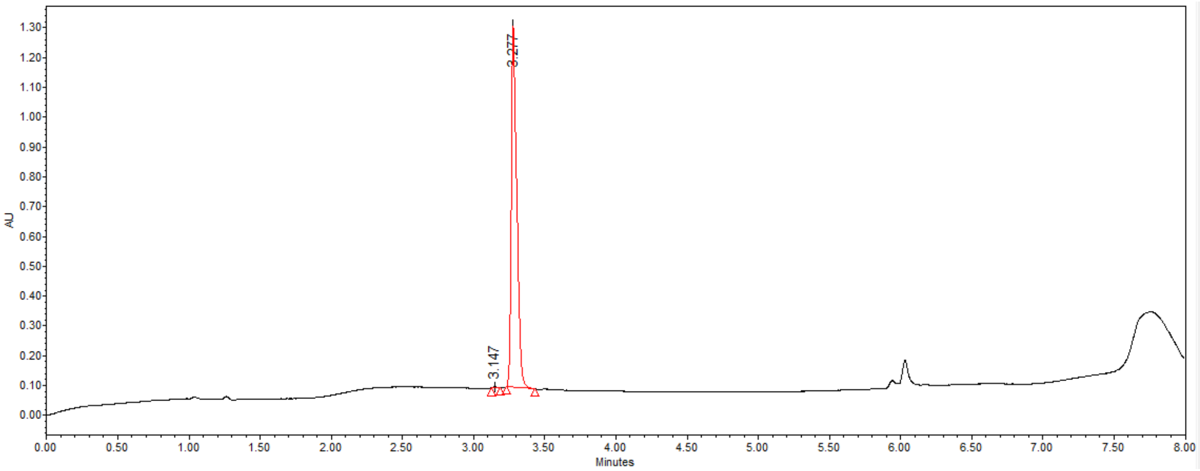
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.82 %

VHP_N19n_H27h

Sequence: **NH_2 — LSDEDFKAVFGMTRSAFAnLPLWKQQhLKKEKGLF — COOH**





MALDI-TOF-MS: calculated monoisotopic $[M+H]^+$ = 4079.135

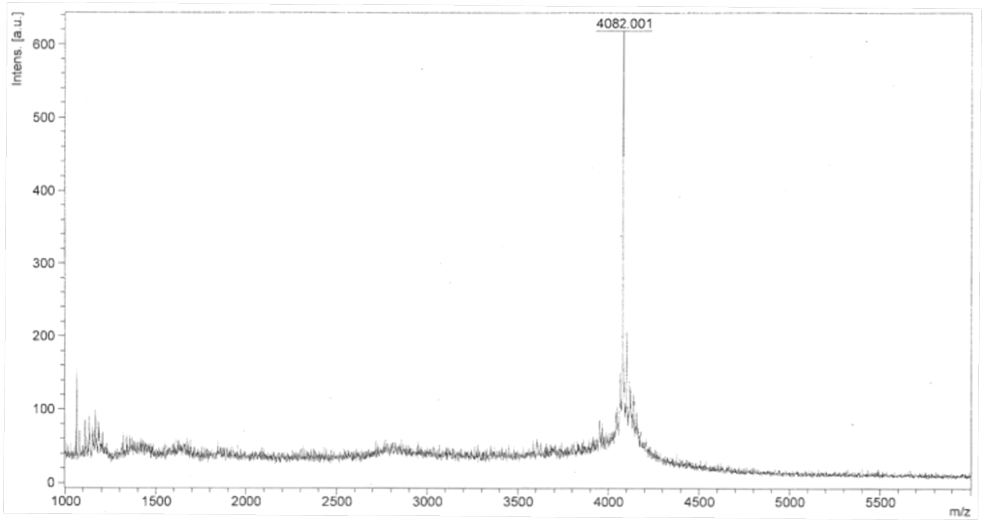
observed monoisotopic $[M+H]^+$ = 4080.401

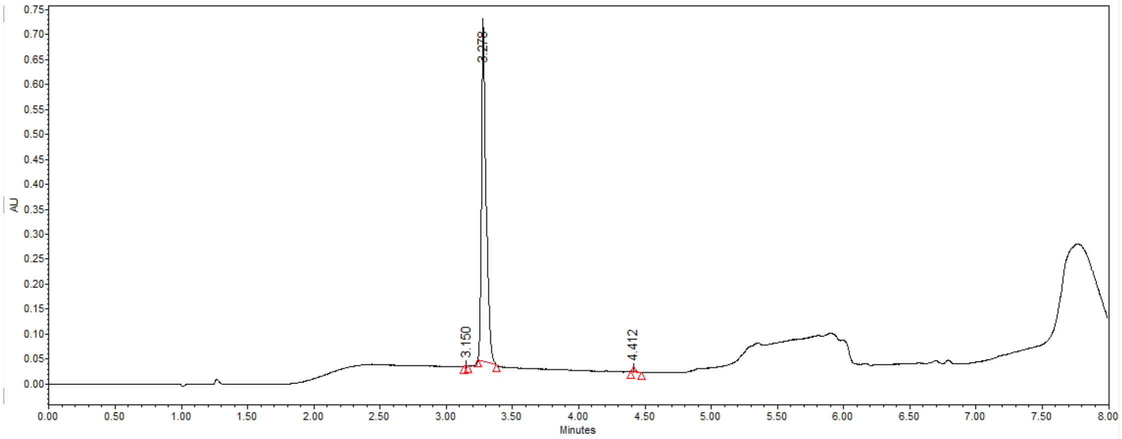
UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.73 %

VHP_A8a_S15s_H27h

Sequence: **NH_2 — LSDEDFKaVFGMTRsAFANLPLWKQQhLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

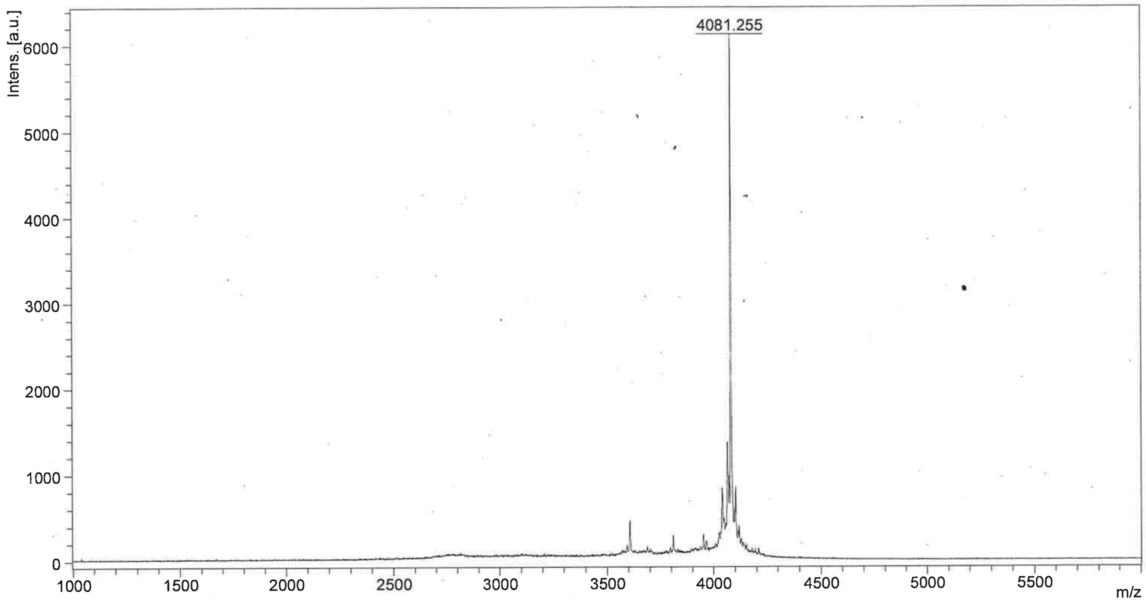
observed monoisotopic $[M+H]^+$ = 4082.001

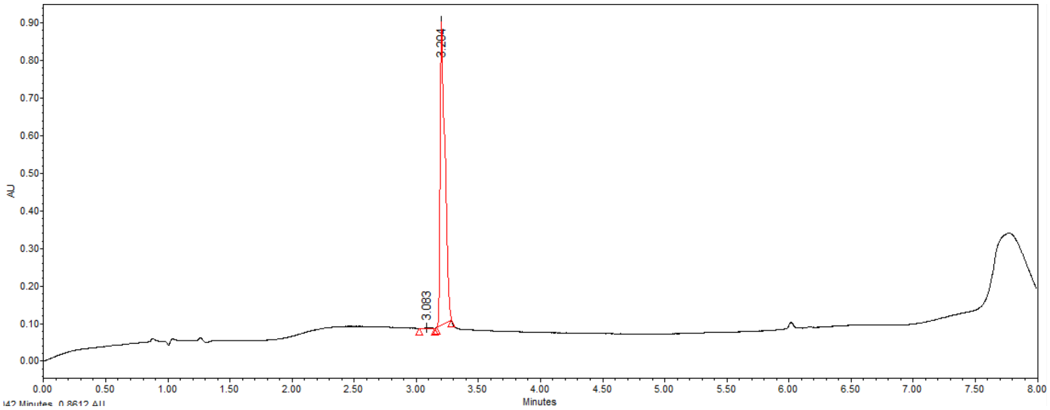
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.08 %

VHP_A8a_N19n_H27h

Sequence: **NH₂ — LSDEDFKaVFGMTRSAFAnLPLWKQQhLKKEKGLF — COOH**





MALDI-TOF-MS: calculated monoisotopic $[M+H]^+$ = 4079.135

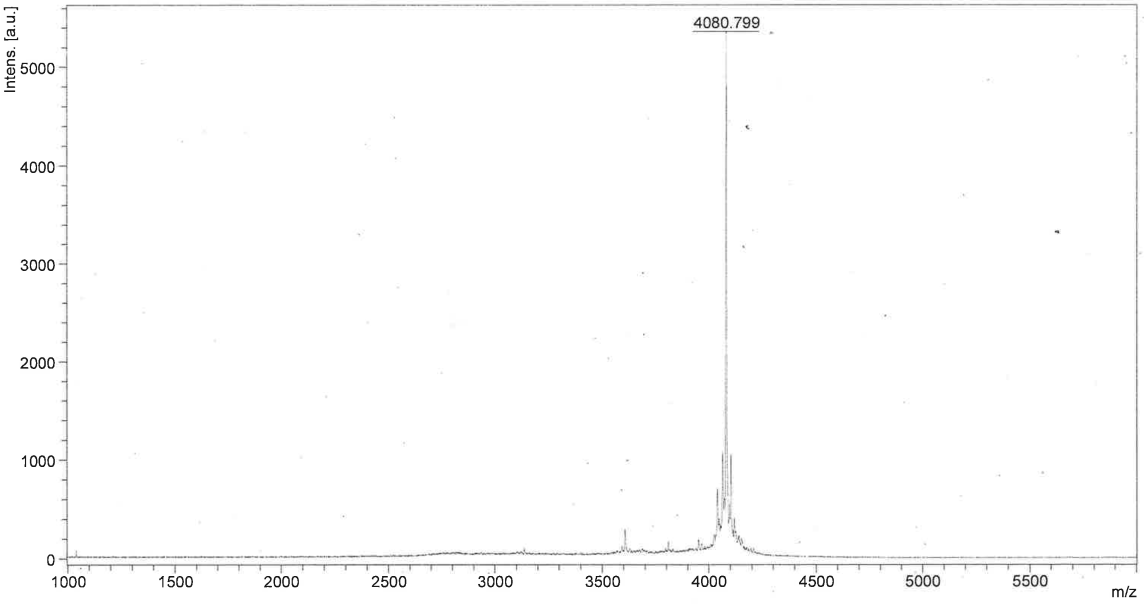
observed monoisotopic $[M+H]^+$ = 4081.255

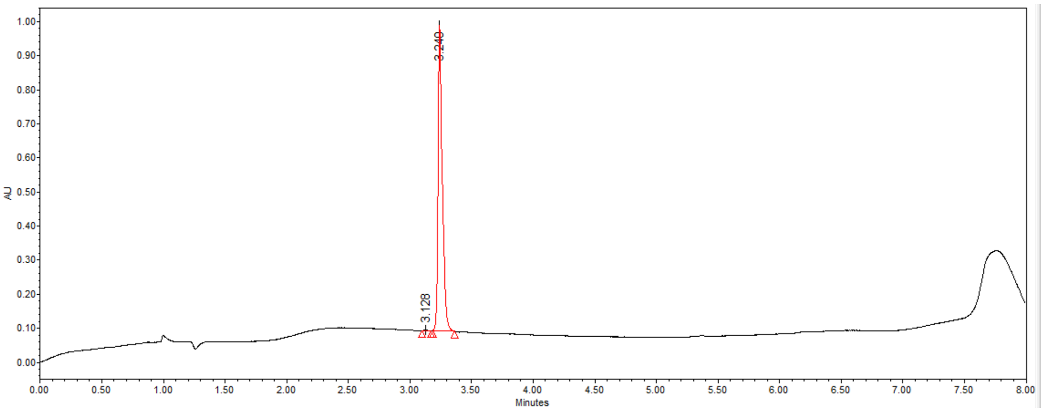
UPLC: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.81 %

VHP_A8a_S15s_N19n_H27h

Sequence: **NH_2 — LSDEDFKaVFGMTRsAFAnLPLWKQQhLKKEKGLF — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4079.135

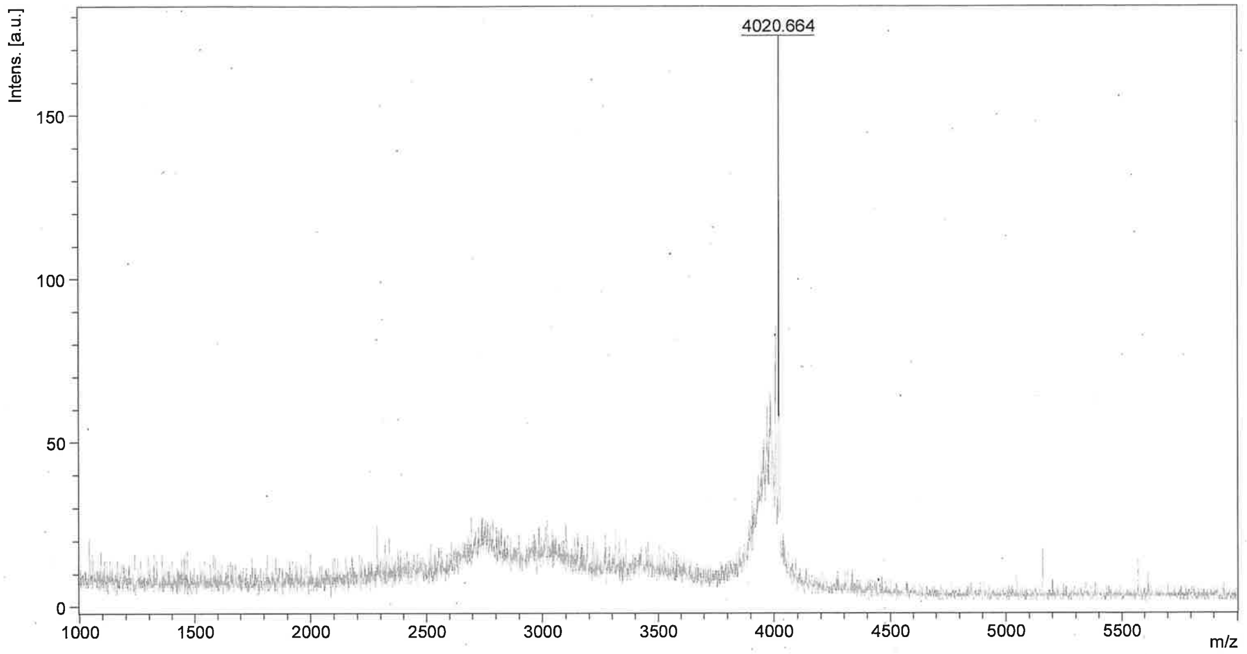
observed monoisotopic $[M+H]^+$ = 4080.799

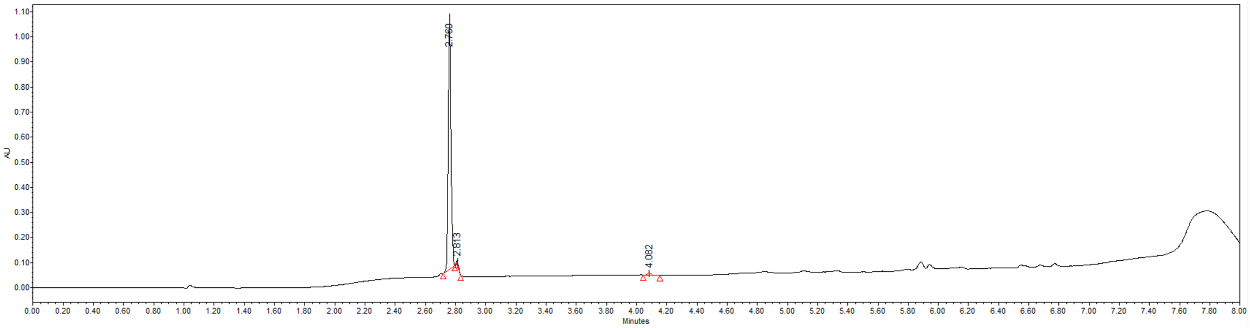
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 99.83 %

WW

Sequence: **NH$_2$ — KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4020.981$

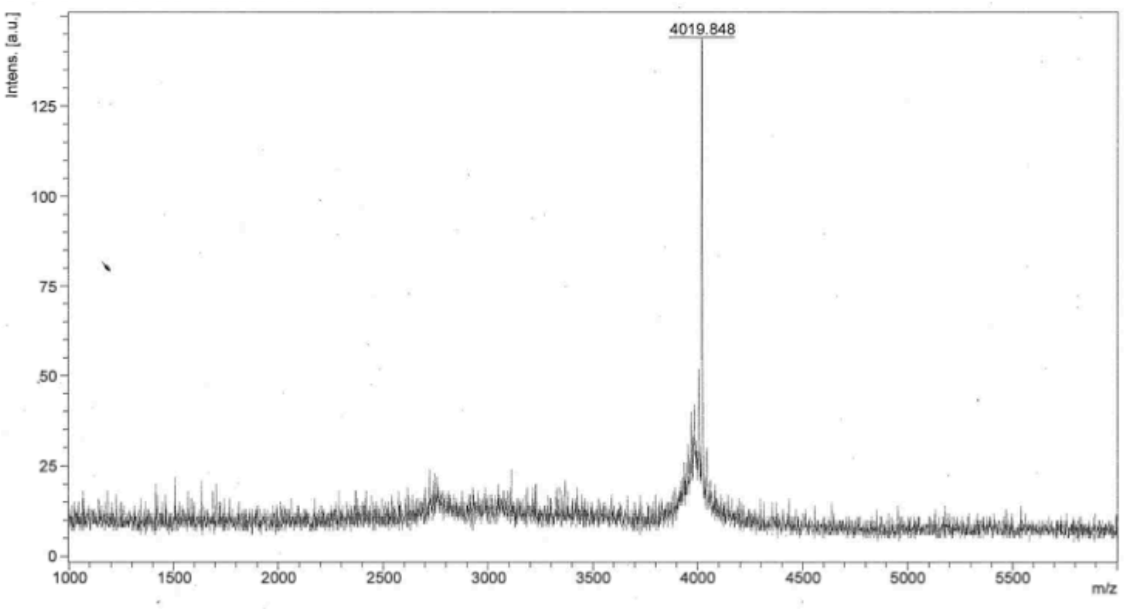
observed monoisotopic $[M+H]^+ = 4020.664$

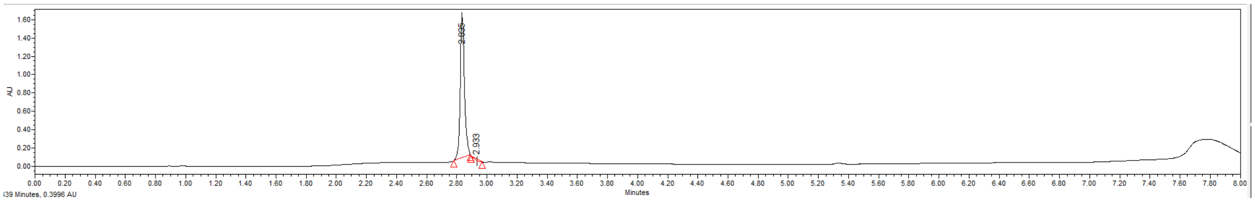
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 97.7 %

WW_W6w

Sequence: **NH₂ — KLPPGwEKRMSRSSGRVYYFNHITNASQWERPSG — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4020.981

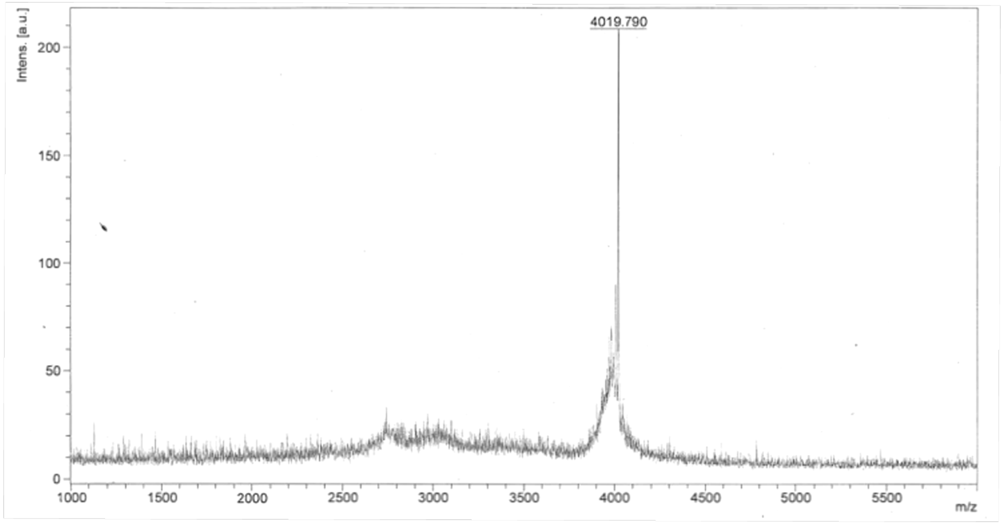
observed monoisotopic $[M+H]^+$ = 4019.848

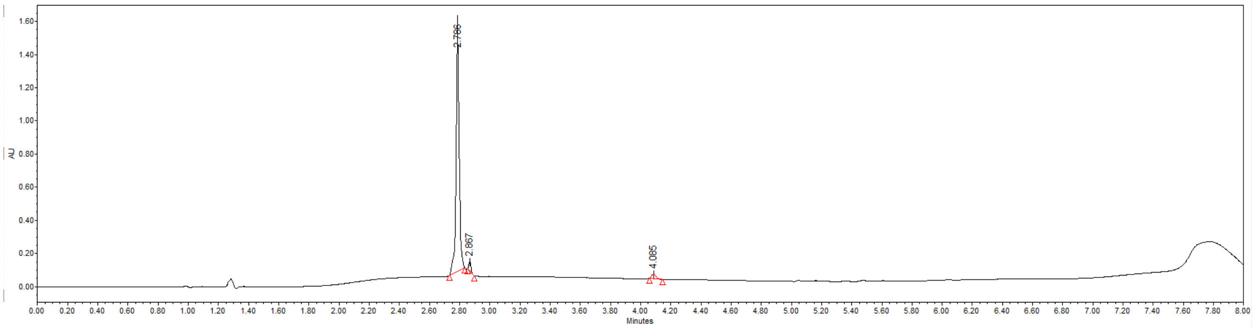
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 98.36 %

WW_E7e

Sequence: **NH$_2$ — KLPPGWeKRMSRSSGRVYYFNHITNASQWERPSG — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4020.981

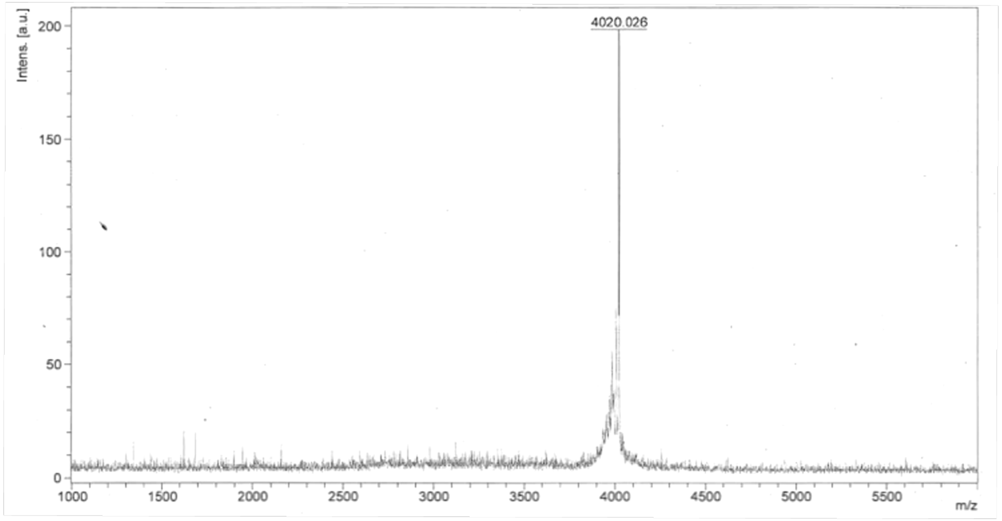
observed monoisotopic $[M+H]^+$ = 4019.790

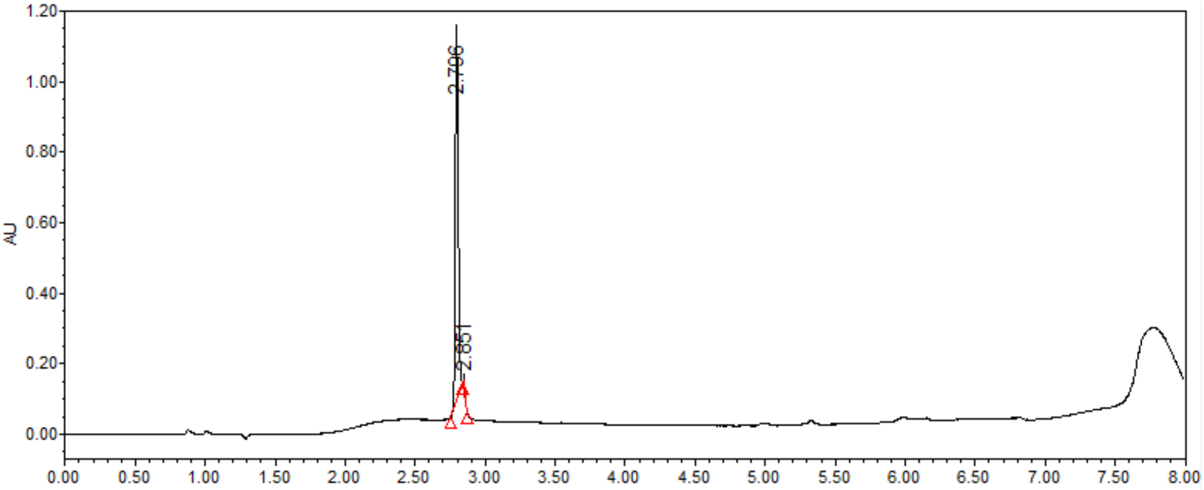
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 95.13 %

WW_K8k

Sequence: **NH₂ — KLPPGWEkRMSRSSGRVYYFNHITNASQWERPSG — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4020.981

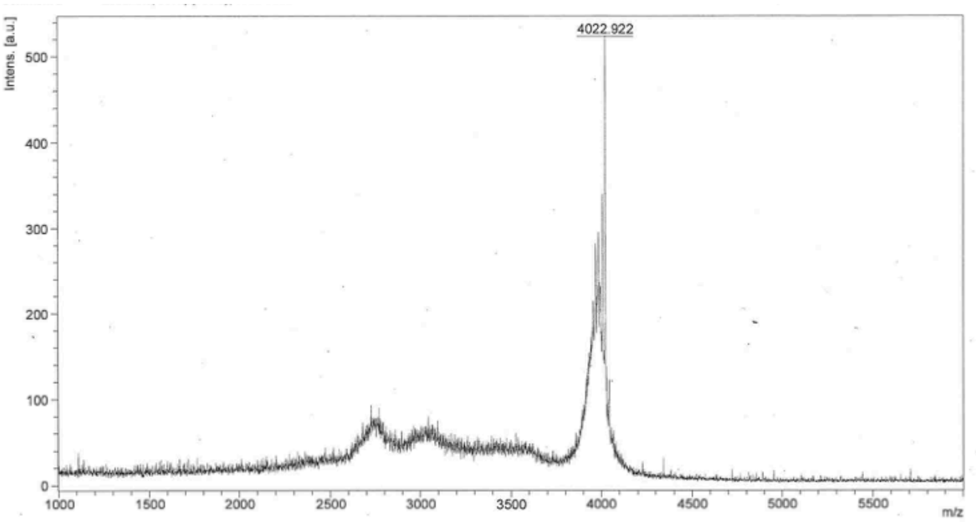
observed monoisotopic $[M+H]^+$ = 4020.026

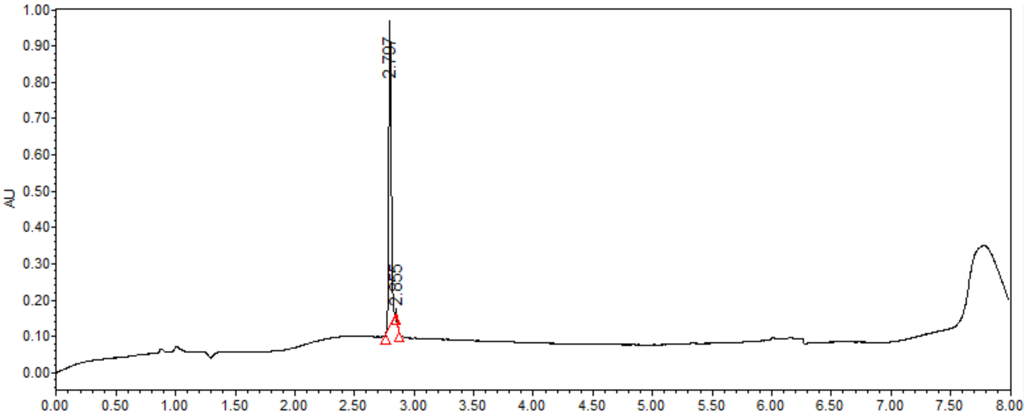
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 97.87 %

WW_R16r

Sequence: **NH₂ – KLPPGWEKRMSRSSGrVYYFNHITNASQWERPSG – COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4020.981

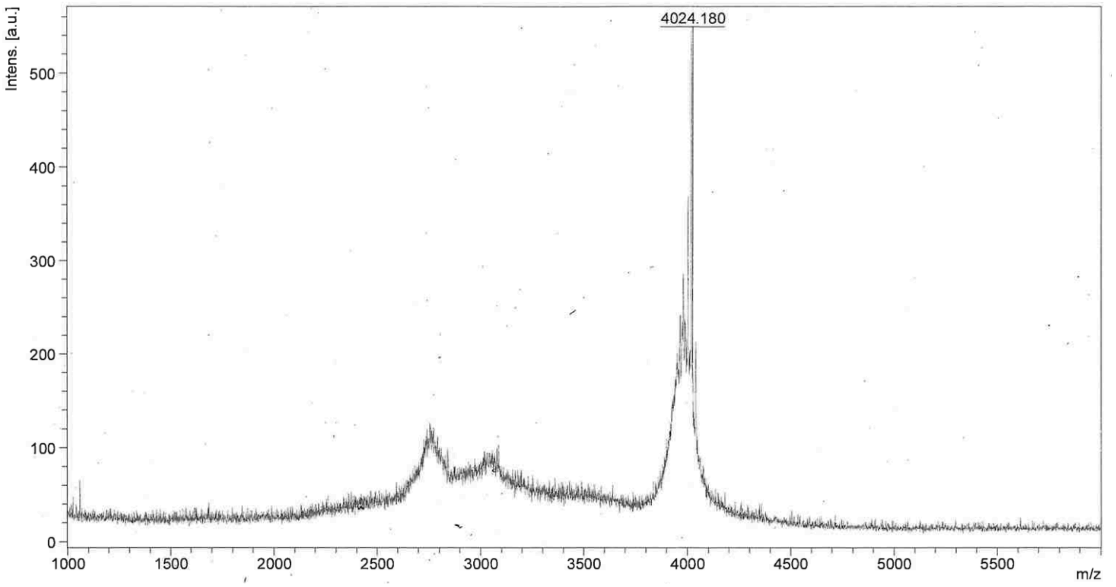
observed monoisotopic $[M+H]^+$ = 4022.922

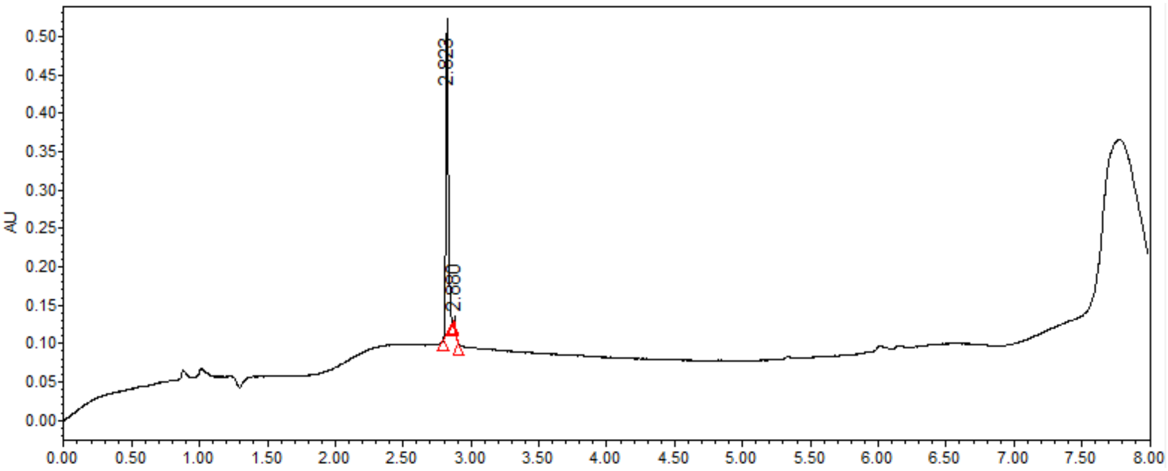
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 98.75 %

WW_Y19y

Sequence: **NH_2 — KLPPGWEKRMSRSSGRVYyFNHITNASQWERPSG — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4020.981$

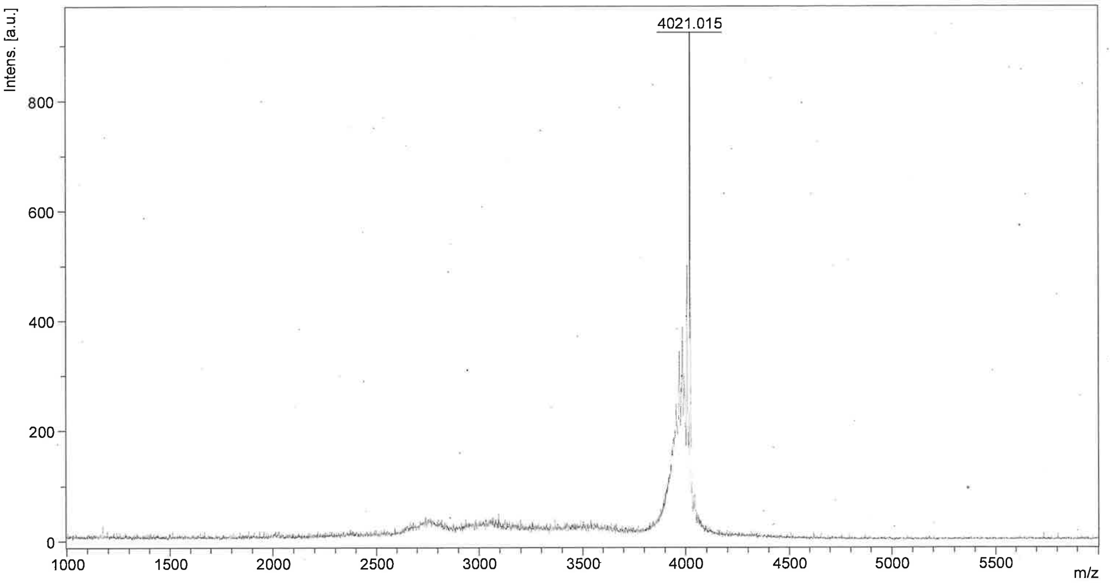
observed monoisotopic $[M+H]^+ = 4024.18$

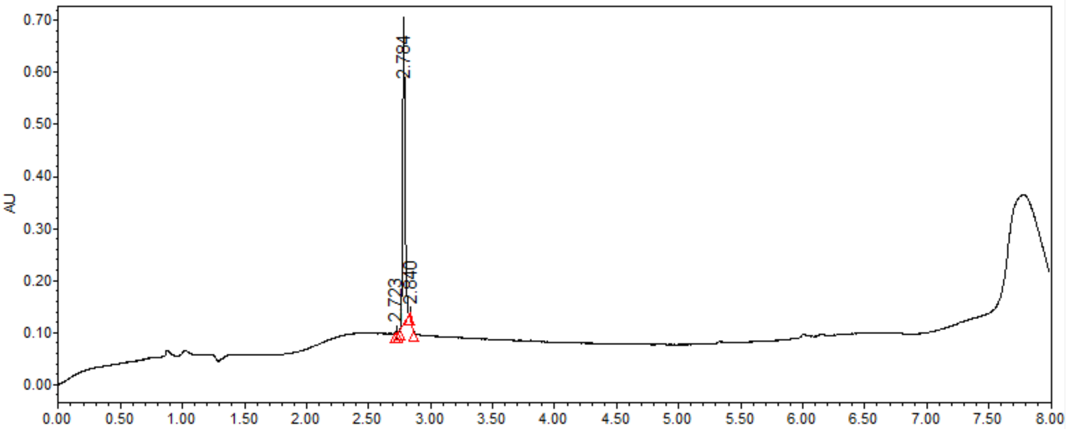
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 98.17 %

WW_S27s

Sequence: **NH₂ — KLPPGWEKRMSRSSGRVYYFNHITNAsQWERPSG — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+$ = 4020.981

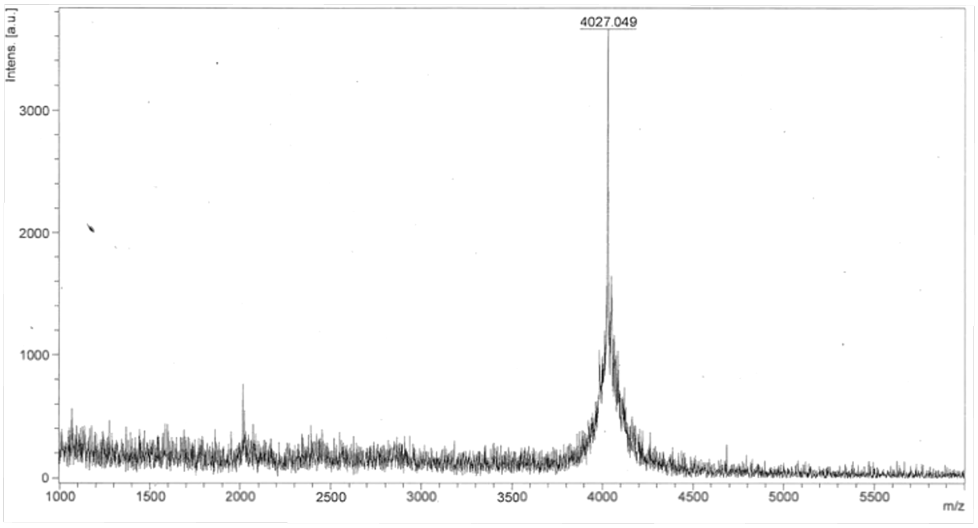
observed monoisotopic $[M+H]^+$ = 4021.015

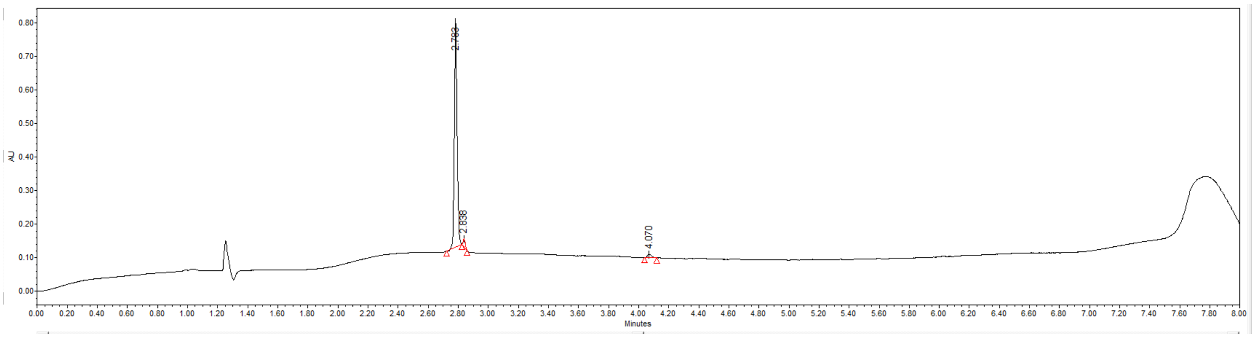
<u>UPLC</u>: H_2O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 97.66 %

WW_Q28q

Sequence: **NH$_2$ — KLPPGWEKRMSRSSGRVYYFNHITNASqWERPSG — COOH**





<u>MALDI-TOF-MS</u>: calculated monoisotopic $[M+H]^+ = 4020.981$

observed monoisotopic $[M+H]^+ = 4027.049$

<u>UPLC</u>: H$_2$O/MeCN + 0.1% TFA, 10-95% MeCN, 5 min, 0.3 mL/min on an ACQUITY Premier CSH C18 (130Å 1.7 µm, 2.1 x 150mm) column.

Purity = 95.14 %

5.7. References

(1) Marahiel, M. A. Working Outside the Protein-Synthesis Rules: Insights into Non-Ribosomal Peptide Synthesis. *Journal of Peptide Science*. December 2009, pp 799–807. https://doi.org/10.1002/psc.1183.

(2) Armstrong, D. W.; Berthod, A. Occurrence of D-Amino Acids in Natural Products. *Natural Products and Bioprospecting*. Springer November 7, 2023, p 47. https://doi.org/10.1007/s13659-023-00412-0.

(3) Miyamoto, T.; Sekine, M.; Ogawa, T.; Hidaka, M.; Homma, H.; Masaki, H. Detection of D-Amino Acids in Purified Proteins Synthesized in Escherichia Coli. *Amino Acids* **2010**, *38* (5), 1377–1385. https://doi.org/10.1007/s00726-009-0348-2.

(4) Checco, J. W.; Zhang, G.; Yuan, W. D.; Le, Z. W.; Jing, J.; Sweedler, J. V. Aplysia Allatotropin-Related Peptide and Its Newly Identified D-Amino Acid-Containing Epimer Both Activate a Receptor and a Neuronal Target. *Journal of Biological Chemistry* **2019**, *293* (43), 16862–16873. https://doi.org/10.1074/jbc.RA118.004367.

(5) Mroz, P. A.; Perez-Tilve, D.; Mayer, J. P.; DiMarchi, R. D. Stereochemical Inversion as a Route to Improved Biophysical Properties of Therapeutic Peptides Exemplified by Glucagon. *Commun Chem* **2019**, *2* (1). https://doi.org/10.1038/s42004-018-0100-5.

(6) È ller, Hr.; Ãpital Saint Antoine, H.; Ho Ãpitaux Pb, A. *The LHRH Antagonist Cetrorelix: A Review*.

(7) Eliasson, J.; Hvistendahl, M. K.; Freund, N.; Bolognani, F.; Meyer, C.; Jeppesen, P. B. Apraglutide, a Novel Glucagon-like Peptide-2 Analog, Improves Fluid Absorption in Patients with Short Bowel Syndrome Intestinal Failure: Findings from a Placebo-Controlled,

Randomized Phase 2 Trial. *Journal of Parenteral and Enteral Nutrition* **2022**, *46* (4), 896–904. https://doi.org/10.1002/jpen.2223.

(8) Chorev, M.; Goldman, M. E.; McKee, R. L.; Roubini, E.; Levy, J. J.; Gay, C. T.; Reagan, J. E.; Fisher, J. E.; Caporale, L. H.; Golub, E. E.; Caulfield, M. P.; Nutt, R. F.; Rosenblatt, M. Modifications of Position 12 in Parathyroid Hormone and Parathyroid Hormone Related Protein: Toward the Design of Highly Potent Antagonists. *Biochemistry* **1990**, *29* (6), 1580–1586. https://doi.org/10.1021/bi00458a032.

(9) White, A. D.; Peña, K. A.; Clark, L. J.; Santa Maria, C.; Liu, S.; Jean-Alphonse, F. G.; Young Lee, J.; Lei, S.; Cheng, Z.; Tu, C.-L.; Fang, F.; Szeto, N.; Gardella, T. J.; Xiao, K.; Gellman, S. H.; Bahar, I.; Sutkeviciute, I.; Chang, W.; Vilardaga, J.-P. *Spatial Bias in CAMP Generation Determines Biological Responses to PTH Type 1 Receptor Activation*; 2021; Vol. 14. https://www.science.org.

(10) Cary, B. P.; Deganutti, G.; Zhao, P.; Truong, T. T.; Piper, S. J.; Liu, X.; Belousoff, M. J.; Danev, R.; Sexton, P. M.; Wootten, D.; Gellman, S. H. Structural and Functional Diversity among Agonist-Bound States of the GLP-1 Receptor. *Nat Chem Biol* **2022**, *18* (3), 256–263. https://doi.org/10.1038/s41589-021-00945-w.

(11) Bockus, A. T.; Lexa, K. W.; Pye, C. R.; Kalgutkar, A. S.; Gardner, J. W.; Hund, K. C. R.; Hewitt, W. M.; Schwochert, J. A.; Glassey, E.; Price, D. A.; Mathiowetz, A. M.; Liras, S.; Jacobson, M. P.; Lokey, R. S. Probing the Physicochemical Boundaries of Cell Permeability and Oral Bioavailability in Lipophilic Macrocycles Inspired by Natural Products. *J Med Chem* **2015**, *58* (11), 4581–4589. https://doi.org/10.1021/acs.jmedchem.5b00128.

(12) Hickey, J. L.; Zaretsky, S.; St Denis, M. A.; Kumar Chakka, S.; Morshed, M. M.; Scully, C. C. G.; Roughton, A. L.; Yudin, A. K. Passive Membrane Permeability of Macrocycles Can

Be Controlled by Exocyclic Amide Bonds. *J Med Chem* **2016**, *59* (11), 5368–5376. https://doi.org/10.1021/acs.jmedchem.6b00222.

(13) Hosseinzadeh, P.; Bhardwaj, G.; Mulligan, V. K.; Shortridge, M. D.; Craven, T. W.; Pardo-Avila, F.; Rettie, S. A.; Kim, D. E.; Silva, D.-A.; Ibrahim, Y. M.; Webb, I. K.; Cort, J. R.; Adkins, J. N.; Varani, G.; Baker, D. *Comprehensive Computational Design of Ordered Peptide Macrocycles*; 2017; Vol. 358. https://www.science.org.

(14) Werner, H. M.; Cabalteja, C. C.; Horne, W. S. Peptide Backbone Composition and Protease Susceptibility: Impact of Modification Type, Position, and Tandem Substitution. *ChemBioChem* **2016**, *17* (8), 712–718. https://doi.org/10.1002/cbic.201500312.

(15) Hecht, S. M. Expansion of the Genetic Code Through the Use of Modified Bacterial Ribosomes. *Journal of Molecular Biology*. Academic Press August 20, 2022. https://doi.org/10.1016/j.jmb.2021.167211.

(16) Dedkova, L. M.; Fahmi, N. E.; Golovine, S. Y.; Hecht, S. M. Construction of Modified Ribosomes for Incorporation of D-Amino Acids into Proteins. *Biochemistry* **2006**, *45* (51), 15541–15551. https://doi.org/10.1021/bi060986a.

(17) Degrado, W. F.; Wasserman, Z. R.; Lear, J. D.; Neil, K. T. *UTC*; 1992; Vol. 114. https://pubs.acs.org/sharingguidelines.

(18) Fisher, B. F.; Hong, S. H.; Gellman, S. H. Helix Propensities of Amino Acid Residues via Thioester Exchange. *J Am Chem Soc* **2017**, *139* (38), 13292–13295. https://doi.org/10.1021/jacs.7b07930.

(19) Krause, E.; Bienert, M.; Schmieder, P.; Wenschuh, H. The Helix-Destabilizing Propensity Scale of D-Amino Acids: The Influence of Side Chain Steric Effects. *J Am Chem Soc* **2000**, *122* (20), 4865–4870. https://doi.org/10.1021/ja9940524.

(20) Žoldák, G.; Stigler, J.; Pelz, B.; Li, H.; Rief, M. Ultrafast Folding Kinetics and Cooperativity of Villin Headpiece in Single-Molecule Force Spectroscopy. *Proc Natl Acad Sci U S A* **2013**, *110* (45), 18156–18161. https://doi.org/10.1073/pnas.1311495110.

(21) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. *High-Resolution x-Ray Crystal Structures of the Villin Headpiece Subdomain, an Ultrafast Folding Protein*; 2005; Vol. 24. www.pnas.orgcgidoi10.1073pnas.0502495102.

(22) Kowalski, J. A.; Liu, K.; Kelly, J. W. NMR Solution Structure of the Isolated Apo Pin1 WW Domain: Comparison to the x-Ray Crystal Structures of Pin1. *Biopolymers* **2002**, *63* (2), 111–121. https://doi.org/10.1002/bip.10020.

(23) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat Protoc* **2006**, *1* (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.

(24) Greenfield, N. J. Using Circular Dichroism Collected as a Function of Temperature to Determine the Thermodynamics of Protein Unfolding and Binding Interactions. *Nat Protoc* **2006**, *1* (6), 2527–2535. https://doi.org/10.1038/nprot.2006.204.

(25) Anil, B.; Song, B.; Tang, Y.; Raleigh, D. P. Exploiting the Right Side of the Ramachandran Plot: Substitution of Glycines by D-Alanine Can Significantly Increase Protein Stability. *J Am Chem Soc* **2004**, *126* (41), 13194–13195. https://doi.org/10.1021/ja047119i.

(26) Valiyaveetil, F. I.; Sekedat, M.; Mackinnon, R.; Muir, T. W. Glycine as a D-Amino Acid Surrogate in the K-Selectivity Filter. *PNAS* **2004**.

(27) Hua, Q. X.; Nakagawa, S.; Hu, S. Q.; Jia, W.; Wang, S.; Weiss, M. A. Toward the Active Conformation of Insulin: Stereospecific Modulation of a Structural Switch in the B Chain. *Journal of Biological Chemistry* **2006**, *281* (34), 24900–24909. https://doi.org/10.1074/jbc.M602691200.

(28) Rodriguez-Granillo, A.; Annavarapu, S.; Zhang, L.; Koder, R. L.; Nanda, V. Computational Design of Thermostabilizing D-Amino Acid Substitutions. *J Am Chem Soc* **2011**, *133* (46), 18750–18759. https://doi.org/10.1021/ja205609c.

(29) Raskatov, J. A.; Teplow, D. B. Using Chirality to Probe the Conformational Dynamics and Assembly of Intrinsically Disordered Amyloid Proteins. *Sci Rep* **2017**, *7* (1). https://doi.org/10.1038/s41598-017-10525-5.

(30) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M. Ion Selectivity in a Semisynthetic K+ Channel Locked in the Conductive Conformation. *Science (1979)* **2006**, *314*, 1004–1007. https://doi.org/10.1126/science.1133601.

(31) Stanger, H.; Gellman, S. Rules for Antiparallel-Sheet Design: D-Pro-Gly Is Superior to L-Asn-Gly for-Hairpin Nucleation. *J Am Chem Soc* **1998**, *120* (17), 4236–4237.

(32) Schenck, H. L.; Gellman, S. H. Use of a Designed Triple-Stranded Antiparallel b-Sheet to Probe b-Sheet Cooperativity in Aqueous Solution. *J Am Chem Soc* **1998**, *120* (19), 4869–4870. https://doi.org/10.1021/ja973984.

(33) Culik, R. M.; Annavarapu, S.; Nanda, V.; Gai, F. Using D-Amino Acids to Delineate the Mechanism of Protein Folding: Application to Trp-Cage. *Chem Phys* **2013**, *422*, 131–134. https://doi.org/10.1016/j.chemphys.2013.01.021.

(34) Langs, D. A.; Smith, G. D.; Courseille, C.; Precigous, G.; Hospital, M. Monoclinic Uncomplexed Double-Stranded, Antiparallel, Left-Handed 5.6-Helix (Increases Decreases Beta 5.6) Structure of Gramicidin A: Alternate Patterns of helical Association and Deformation. *Proc Natl Acad Sci U S A* **1991**, *88*, 5345–5349.

(35) Gonda, Y.; Matsuda, A.; Adachi, K.; Ishii, C.; Suzuki, M.; Osaki, A.; Mita, M.; Nishizaki, N.; Ohtomo, Y.; Shimizu, T.; Yasui, M.; Hamase, K.; Sasabe, J. Mammals Sustain Amino Acid Homochirality against Chiral Conversion by Symbiotic Microbes. *Proc Natl Acad Sci U S A* **2023**, *120* (15). https://doi.org/10.1073/pnas.2300817120.