5-2019

Sequence-Selective Recognition of Methionine-Terminated Peptides by Cucurbit[8]uril

Hailey F. Taylor
Trinity University, taylor.hailey97@gmail.com

Sequence-Selective Recognition of Methionine-Terminated Peptides by Cucurbi[8]uril

Hailey Frances Taylor

A DEPARTMENT HONORS THESIS SUBMITTED TO THE
DEPARTMENT OF CHEMISTRY AT TRINITY UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR GRADUATION
WITH DEPARTMENTAL HONORS

APRIL 22, 2019

Adam R. Urbach, Ph.D. Thesis Advisor	Christopher J. Pursell, Ph.D. Chair, Department of Chemistry

Michael Soto, AVPAA

Acknowledgements

I would like to give special acknowledgement to my thesis advisor, Dr. Adam Urbach, not only for giving me the opportunity to take part in his research, but also for mentoring me in how to be an effective scientist. I have learned so much from this experience and will always be grateful that you accepted me into your lab. I could not have accomplished any of this without your constant encouragement and unfailing support.

My committee members, Dr. Urbach and Dr. Hunsicker-Wang, deserve many thanks for investing the time to closely read my thesis and help me to improve my writing. Thank you for your patience and support throughout this entire process.

Dr. Wooten's support is one of the reasons that I decided to become a chemistry major. Her constant reassurance and patience during general chemistry lab taught me that chemistry was not so overwhelming. Dr. Wooten, I will miss our impromptu conversations and all of the advice you have given me these past four years. Dr. Cooley fueled my interest in chemistry through her energetic lectures in organic chemistry. Her passion for science and teaching is something that I aspire to one day. I also would like to thank Dr. Maeder and Dr. Pursell for their constant encouragement throughout the past two years. The professors in the Trinity chemistry department have exceeded all of my expectations and I am so appreciative to have been a part of such a supportive group. I will always cherish my time spent here and all of the late nights in CSI.

The Urbach group members – Erin, Lois, Hayden, Emily, Elena, Tim, Will and Zoheb – have all been great to work with in lab over these past few years. Thank you for making all of our late nights in lab more enjoyable and memorable than I ever thought they could be. Zoheb Hirani, in particular, has been my diligent lab partner and an equal

contributor to the contents in Chapter 2. Thank you for your help with our scientific approach and careful thoughtfulness concerning our research.

I would also like to acknowledge the King lab: Dr. King, Heather, Gabrielle, and Gabby, for the food, coffee, and our many wonderful conversations. You all were always there to listen and always went out of your way to brighten my day.

I would especially like to thank my parents for always encouraging me to pursue my passions and approach life with positivity. I value tremendously all of your advice and support over the past four years – even when I call you late at night. Thank you for always being there to listen and always giving me clarity and guidance when I needed it most.

Thesis Abstract

Sequence-selective recognition is an important goal for synthetic receptors binding to proteins with a high affinity. This thesis describes the molecular recognition of methionine-terminated peptides by the synthetic receptor, cucurbit[8]uril (Q8) with submicromolar affinity in aqueous solution. Previous studies have demonstrated Q8 binding to N-terminal aromatic residues. In 2015 we discovered that Q8 can bind to certain dipeptide sequences with both side chains binding simultaneously with the cavity of Q8. The additional binding interface of the two residues suggested the possibility of targeting entirely nonaromatic peptides. To test this hypothesis, a peptide library was designed by qualitatively screening for Q8 binding though a fluorescent indicator displacement assay. The results revealed several leads from the Met-terminated series, suggesting strong binding which was characterized using isothermal titration calorimetry, ^{1}H NMR spectroscopy, and mass spectrometry. This discovery of Q8 binding to Met-terminated peptides, which contained non-aromatic sequences, incited a detailed investigation of 25 peptides to characterize the structural determinants of binding. Further work in inhibiting a non-selective protease using Q8 to bind these dipeptide sequences is ongoing.

Table of Contents

Chapter One

Introduction

Supramolecular chemistry works to understand the non-covalent binding interactions between molecules, which mediates many biological processes including enzyme activity, cell communication, and the assembly of multi-protein complexes like viruses or ribosomes. The field is interested in studying selective binding (Figure 1), which is determined by comparing the binding affinities of different guests for a given target and aids in determining potential binders[1]. The strength of such binding is measured either by an equilibrium association constant, K_a, or by a dissociation constant, K_d. Strong binding is indicated by either a large K_a or small K_d value. Factors that can affect the selectivity of binding include complementarity in the shape and electrostatic character of the host and guest, as well as how each are pre-organized prior to binding. The strength of binding is influenced by intermolecular forces, both coulombic and polarization forces. Coulombic forces involve molecules with permanent poles, i.e., ions and dipoles, and comprises ion-ion, ion-dipole, and dipole-dipole interactions, which includes hydrogen bonding. Polarization forces involve induced dipoles and include ion-induced dipole, dipole-induced dipole, and induced dipole-induced dipole interactions, i.e., van der Waals or London dispersion forces[2].

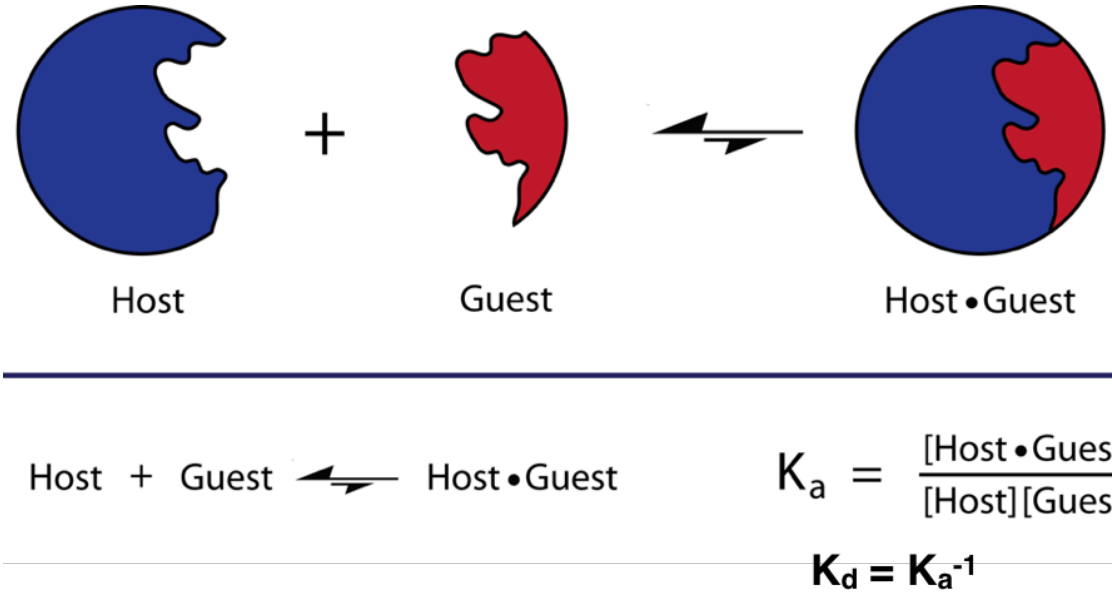


Figure 1. Schematic of host·guest equilibrium.

Non-Classical Hydrophobic Effect. Biological interactions occur in aqueous solution where the hydrophobic effect is an important factor. Water is unique in that it is a very small molecule compared to other solvents, is highly polar, and contains strong hydrogen bonding properties. Water molecules are able to create a network of hydrogen bonding while in solution. Upon dissolving a hydrophobic, concave host, water molecules are driven inside of the cavity. While in the cavity, they have a limited number of hydrogen bonds unlike in the solvent, and they become high in energy. The driving force behind binding is enthalpic due to the waters making hydrogen bonds with the bulk solvent after they are released from the cavity upon guest binding[3].

Understanding and even exploiting the enthalpic and entropic contributions to binding enables one to gain a deeper understanding on many biological functions. In doing so, catalysts, materials, and drugs could be designed to take advantage of these contributions. Drug design is of importance in order to combat disease, disorders, and drug resistance. By learning about the fundamentals of binding behind host•guest complexes

and how weak interactions are able to make a strong interaction, these complexes can allow for the opportunity for drugs to be specifically designed to a particular host.

Peptide and Protein Recognition. Typical approaches in targeting proteins with a high affinity and selectivity require either discovering ligands from screening combinatorial libraries or optimizing ligand design from a detailed knowledge of the desired protein's tertiary structure. However, protein structure is difficult to predict from the amino acid sequence. It would be particularly useful and powerful if protein recognition could be based on the amino acid sequence of the desired protein. Neighboring amino acid residues in the folded protein, however, are not necessarily neighbors in the amino sequence. To predict recognition based on sequence, the binding sites needs to be small (2-3 residues) so that the residues involved are adjacent in the sequence, such as in a small peptide. Therefore, each residue in the complex needs to contribute significantly to the binding energy in order to be considered efficient ligands. Aromatic residues are very efficient ligands due to their large, flat surfaces that provide significant hydrophobic, van der Waals, and electrostatic interactions. Therefore, it is no surprise that aromatic residues are important for stabilizing protein interactions[4-9]. Synthetic receptors that contain their own concave internal cavities are therefore well suited to bind to small sites on peptides and proteins by creating an extensive binding interface. Among the synthetic receptors reported to bind peptides non-covalently in aqueous solution, submicromolar equilibrium dissociation constants have been observed only with aromatic peptides for two classes of hosts: Fujita's coordination cages [9] and cucurbit[n]urils[10].

Fujita's coordination cages have shown sequence-selective recognition of aromatic amino acid containing peptides by the single binding pocket of their self-assembled cage. By having a large hydrophobic cavity made up of aromatic rings, they facilitate recognition through π - π stacking and CH - π interactions between the cage and the peptide (Figure 2). The cage is big enough to fit two aromatic residues within its cavity. The tripeptide sequence, WWA, is solely recognized. They then replaced one of the Trp with Tyr and varied its position; which revealed weak binding. When they extended the tripeptide into a tetrapeptide and hexapeptide, with the WWA tag in the middle of the sequence, Fujita and coworkers observed high affinity binding as was found with the tripeptide.

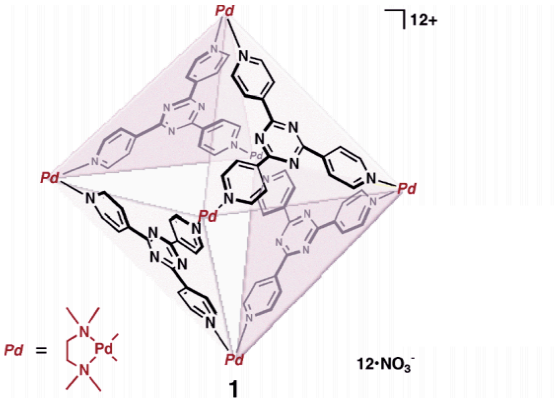


Figure 2. Chemical structure for a Fujita coordination complex. Reprinted with permission from JACS.[9]

Cucurbit[n]uril Peptide Recognition. Discovered in 1905[11], cucurbit[n]urils (Qn's) are synthetic organic macrocycles composed of repeating glycoluril units, with the cavity size depending on the number of units (Figures 3, 4). The carbonyl lined portals are approximately 2 Å narrower than the cavity itself, resulting in constrictive binding that generates a steric barrier for guest association and dissociation.[12] The carbonyl-lined portals create an electrostatic effect, having Qn exhibit a preference to interacting with

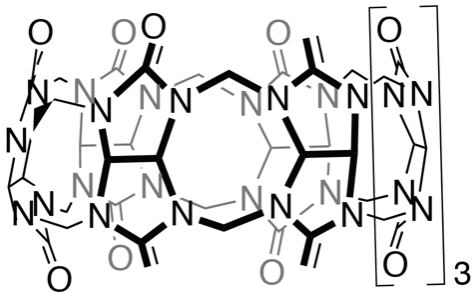


Figure 3. Chemical structure for Q8.

cationic guests. The hydrophobic cavity and the two restricted portals drive the binding of guests with nonpolar and cationic groups.

The size of the guest that can bind to Qn is dependent upon the number of repeating glycoluril units, which dictate cavity size. Q6 is big enough to fit an alkyl chain with an ammonium group at either end. Q7 is large enough to fit an aromatic guest within its cavity. Q8 is special amongst the cucurbit[n]uril family in that it is large enough to simultaneously bind two guests within its cavity.

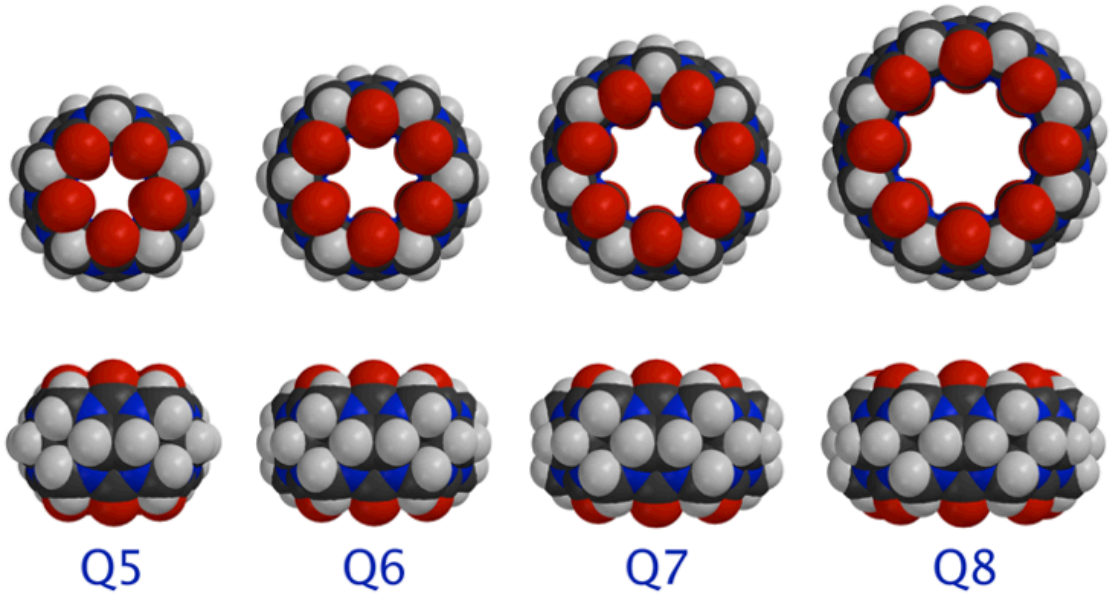


Figure 4. Structures of Qn (n= 5-8). Color code: oxygen, red; nitrogen, blue; carbon, black; hydrogen, grey.

The effects of electrostatic charge on the binding of Q8•MV to aromatic residues were studied to probe molecular recognition by Q8 by tracking the fluorescent changes.[13] MV denotes to methyl viologen and binds to Q8 with a K_d of 8.5 ($\pm$0.3) x 10^5 M^{-1}, which

was determined by isothermal titration calorimetry. The Urbach group examined the ability of Q8•MV to recognize specific amino acids and peptides in aqueous solution (Figure 5).

Upon the four aromatic residues, Trp, Phe, Tyr, and His, binding to Q8•MV, the fluorescence intensity decreases. Among these residues, Trp bound to Q8•MV with the highest affinity ($K_d = 2.3 \times 10^{-5}$ M^{-1}) and with an 8-fold and 19-fold selectivity over Phe and Tyr respectively. The data suggested that relative binding affinity is dictated by the hydrophobicity of the second guest. Based off of this data, the Urbach group hypothesized that N-terminal Trp residues would bind with a higher affinity than C-terminal or internal Trp residues. By varying the location of the Trp residue, it would provide a basis for recognition to certain peptides.

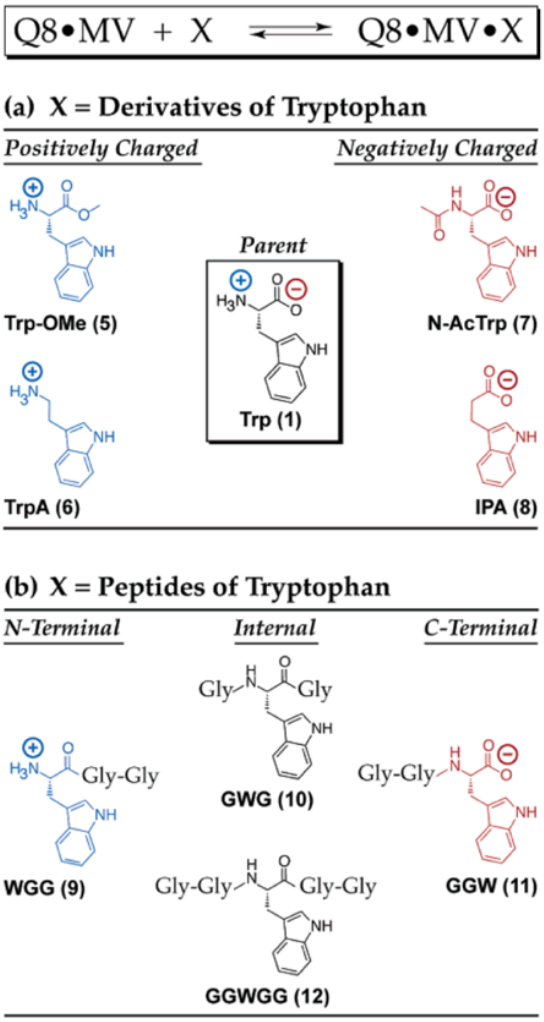


Figure 5. (a) Series of Trp derivatives (b) Trp containing peptides that were examined as guests for Q8•MV. Reprinted with permission from JACS.[13]

Q8 has the ability to bind to two aromatic guests, including peptides, and can dimerize N-terminal aromatic peptides. The tripeptides, Trp-Gly-Gly (YGG) and Phe-Gly-Gly (FGG), dimerize within the cavity of Q8 with a high affinity in aqueous solution[14]. The former binds with a negative cooperativity, while the latter binds with a positive cooperativity. Crystal structures of Q8•YGG and Q8•FGG_2 reveal a structural basis for

selective recognition with the inclusion of the hydrophobic aromatic side chain and the chelation of the N-terminal ammonium groups by the carbonyl oxygens of Q8 (Figure 6). The structures also illustrate ordered interactions of the adjacent peptide residues and aids in the discovery of superior binding-peptides. This extent of sequence-selective recognition of peptides has aided the design of synthetic hosts in aqueous solution and been useful to study dimer-mediated biochemical processes.

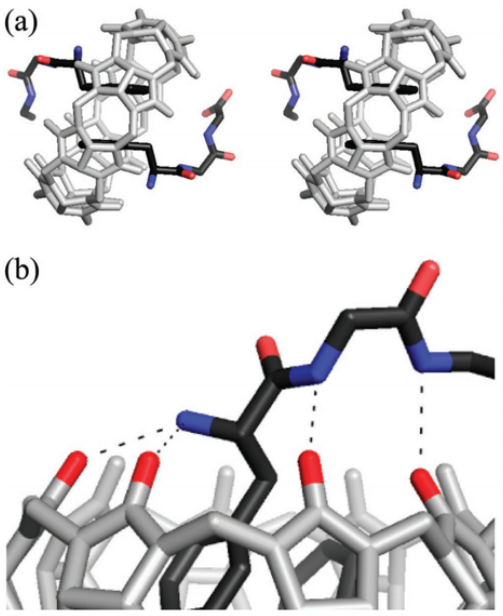


Figure 6. Crystal structure of Q8•FGG_2. (a) Cross-eyed stereoview; (b) dashed lines show key electrostatic interactions. Reprinted with permission from JACS.[14]

Unlike small molecules that bind inside the cavities of proteins,[15] Q7 can selectively recognize the N-terminal Phe on the B-chain of human insulin. Isothermal titration calorimetry and fluorescence spectroscopy experiments demonstrates Q7 binding to insulin with a K_d of 6.7 x 10^{-7} M^{-1} and with a 50-100-fold selectivity compared to other proteins that are larger but lack an N-terminal aromatic residue.[16] An insulin variant without the N-terminal Phe was tested and showed > 1000-fold selectivity. The crystal structure of the Q7•insulin complex complex revealed that the B-chain N-terminus unfolds in such a way to enable binding (Figure 7). This result suggested that site-selective recognition on proteins is inherent to the protein's terminus, including the unique epitope of the terminal residue and the greater freedom for the terminus to unfold in order to accommodate binding.

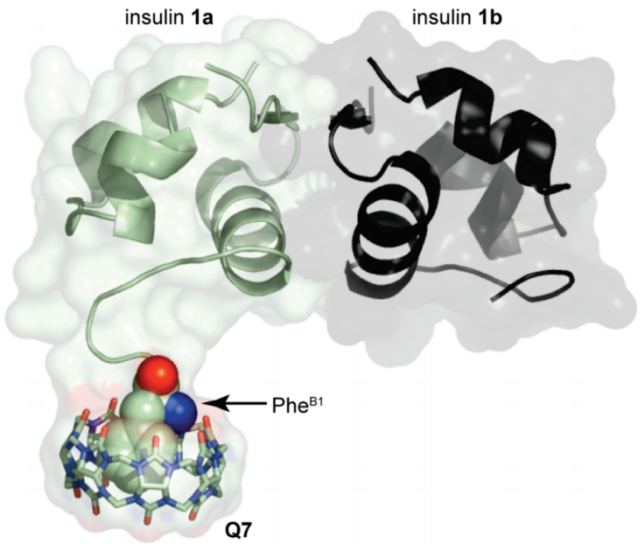


Figure 7. Crystal structure of Q8•4_2. (a) Cross-eyed stereoview; (b) dashed lines show key electrostatic interactions. Reprinted with permission from JACS.[16]

Even though this motif for Q8 only targets a single amino acid residue, it provides a great amount of selectivity due to the N-terminal residue being a unique epitope in the peptide chain. This degree of selectivity is sufficient to enable protein recognition both in simple and in complex mixtures. This motif, however, is limited to a single-residue binding site. In 2015, the Urbach group reported the discovery of Q8 binding with high affinity to the tripeptide, YLA, by including the side chains of the two neighboring residues, Y and L, within the cavity of Q8.[17] A fluorescent indicator displacement assay was used to qualitatively screen for peptide binding to Q8. A peptide library of 35 peptides was screened by measuring the change in fluorescence intensity of Q8•MBBI upon the addition of each peptide. The fluorescent indicator, tetramethylbenzobis(imidazolium) (MBBI), binds to Q8 with a $K_d = 1.8 \times 10^6$ M^{-1}.[18] Upon peptide binding, MBBI is released from the cavity and shows an increase of fluorescence intensity.

In that study,[17] several tripeptides were found to bind to Q8 with high affinity through an interesting "pair inclusion" motif. The tripeptides contained an N-terminal Tyr

residue, with Leu, Lys, Tyr, or Phe at the second position, and Ala at the third position. These sequences were found to bind by a 1:1 "pair inclusion" motif, with binding affinities in the low micromolar to submicromolar range (Table 1). ^{1}H NMR studies revealed that the side chains of both the N-terminal Tyr and the neighboring amino acid residue bind within the cavity of Q8. Each peptide that binds to Q8 via this motif is noted by an asterisk (*) in Table 1. A K_d pf 7.2 nM was observed for the Q8•YLA complex in 10 mM sodium phosphate buffer. This is the highest reported affinity for a synthetic receptor binding to an unmodified peptide in aqueous solution. Separating the two residues resulted with an alanine (YAL) reduces the binding affinity by 4700-fold, demonstrating a notably high sequence specificity. A semiempirical model of the Q8•YLA complex is seen in Figure 8. This model shows both side chains, Y and L, are bound deeply with the cavity of Q8. An additional electrostatic stabilization is suggested from this structure by five hydrogen bonds between the positively charged N-terminus and NH groups on the peptide chain.

Table 1. Thermodynamic data for binding to Q8.[17]

Peptide	K_d (μM)[a]	ΔH[a] (kcal mol^{-1})	-TΔS[b] (kcal mol^{-1})
YLA*	0.0072 (±0.0003)	-15.5 (±0.4)	4.3 (±0.2)
YAL	34. (±7.)	-9.5 (±0.1)	3.4 (±0.1)
AYL*	3.1 (±0.5)	-14.1 (±0.5)	6.5 (±0.1)
YKA*	0.20 (±0.03)	-15.5 (±0.1)	6.3 (±0.2)
YYA*	0.70 (±0.06)	-16.1 (±0.2)	7.7 (±0.2)
YFA*	0.29 (±0.04)	-16.1 (±0.2)	7.2 (±0.1)

[a] Mean values measured from at least three ITC experiments at 300 K in 10 mM sodium phosphate, pH 7.0. [b] Entropic contributions to the free energy of binding were calculated from the K_d and ΔH values, with error propagated from those of K_d and ΔH.

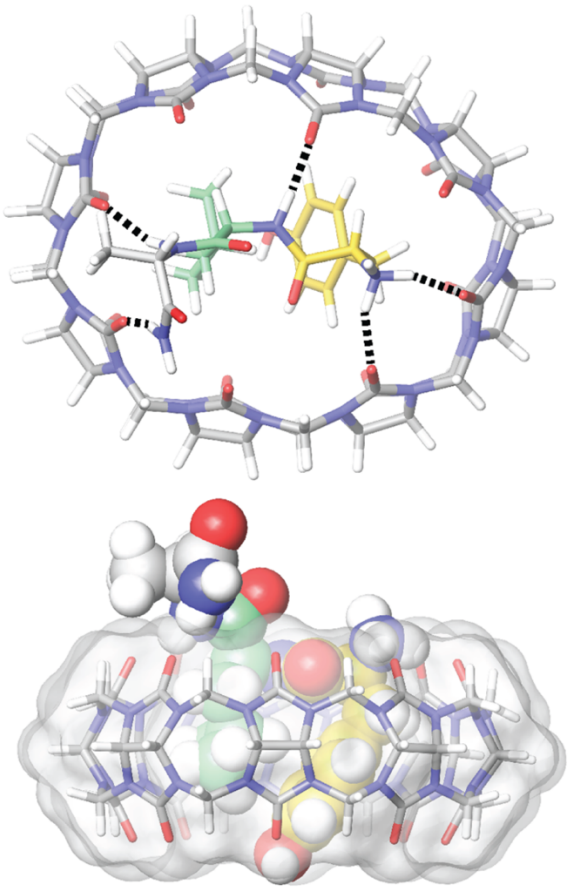


Figure 8. Semiempirical model of the Q8•YLA complex, which was built in Maestro by manually docking the peptide to fit the residues Tyr and Leu within the Q8 cavity and then subjecting the model to a geometry minimization in the OPLS 2005 molecular mechanics force field. Reprinted with permission from JACS.[17]

Content of this Thesis. The discovery of Q8 binding to two neighboring residues on a peptide chain lead us to consider what other pairs of sequences could be targeted through this "pair inclusion" binding motif. We hypothesized that other aliphatic sequence pairs could bind to Q8 in a similar way as aromatics. Chapter 2 investigates other sequence pairs that could be targeted by Q8 through a qualitative fluorescence binding assay with a library of 144 tripeptides and the fluorescence indicator, MBBI, against Q8. These results revealed outliers within the methionine-terminated series, on which we performed our quantitative, in-depth studies using isothermal titration calorimetry (ITC), electrospray ionization mass

spectrometry (ESI-MS), and proton nuclear magnetic resonance spectroscopy (^{1}H NMR). The results of our studies allowed us to establish a novel approach to molecular recognition of Met-terminated peptides and non-aromatic peptides with submicromolar affinity in aqueous solution, as well as establish the structural-determinants for binding. **Chapter 3** discusses our efforts to selectively digestion hen egg white lysozyme and a series of octapeptides using cucurbit[n]urils to selectively inhibit a non-selective exopeptidase.

References

(1) Smith, B. *Synthetic Receptors for Biomolecules: Design Principles and Applications*; Royal Society of Chemistry: London, UK, **2015**.

(2) Creighton, T. E. *Proteins*, W.H. Freeman, **1984**, 139-168.

(3) Biedermann, F.; Nau, W. M.; Schneider, H. J. *Angew. Chem. Intl. Ed.* **2014**, *53*, 11158-11171.

(4) Day, E. S.; Cote, S. M.; Whitty, A. *Biochemistry,* **2012**, *51*, 9124−9136.

(5) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210-1250.

(6) Salonen, L. M.; Ellermann, M.; Diederich, F. *Angew. Chem., Int. Ed.* **2011**, 50, 4808−4842.

(7) Martinez, C. R.; B. L. Iverson, B. L. *Chem. Sci*. **2012**, *3*, 2191-2201.

(8) Mahadevi, A. S.; Sastry, G. N. *Chem. Rev.* **2013**, *113*, 2100-2138.

(9) Tashiro, S.; Tominaga, M.; Kawano, M.; Therrien, B.; Ozeki, T.; Fujita, M. *J. Am. Chem. Soc.* **2005**, *127*, 4546-4547.

(10) Urbach, A. R.; Ramalingam, V. *Israel J. Chem.* **2011**, *51*, 664-678.

(11) Behrend, R.; Meyer, E.; Rusche, F. *Liebigs Ann. Chem.* **1905,** *339*, 1-37.

(12) Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. *Angew. Chem. Intl. Ed.* **2005,** *44*, 4844-4870.

(13) Bush, M. E.; Bouley, N. D; Urbach, A. R. *J. Am. Chem. Soc.* **2005,** *127*, 14511-14517.

(14) Heitmann, L. M; Taylor, A.B.; Hart, P. J.; Urbach, A. R. *J. Am. Chem. Soc.* **2006,** *128*, 12574-22581.

(15) Böhm, H. J.; Schneider, G.; Eds. *Protein-ligand interactions: from molecular recognition to drug design;* Wiley-VCH: Weinheim, 2003.

(16) Chinai, J. M.; Taylor A. B.; Ryno, L. M.; Hargreaves, N. D.; Morris, C. A.; Hart, P. J.; Urbach A. R. *J. Amer. Chem. Soc. Comm.* **2011,** *133*, 8810-8813.

(17) Smith, L. C. Smith, Leach, D. G.; Blaylock, B. E.; Ali, O. A.; Urbach, A. R. *J. Am. Chem. Soc.* **2015**, *137*, 3663-3669.

(18) Biedermann, F.; Rauwald, U.; Cziferszky, M.; Williams, K. A.; Gann, L. D.; Guo, B. Y.; Urbach, A. R.; Bielawski, C. W.; Scherman, O. A. *Chem. Eur. J.* **2010**, *16*, 13716-22.

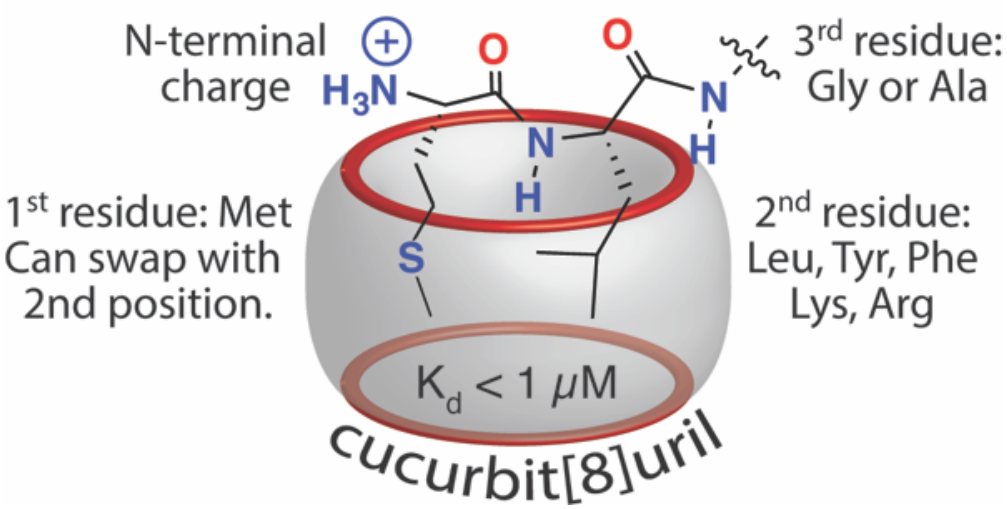
Chapter Two

Molecular Recognition of Methionine-Terminated Peptides by Cucurbit[8]uril

The contents of this thesis chapter are taken directly from the following publication: Hirani, Z.; Taylor, H. F.; Babcock, E. F.; Bockus, A. T.; Varnado, C. D. Jr.; Bielawski, C. W.; Urbach, A. R. *JACS*, **2018**, *140,* 12263-9.

Abstract. This chapter describes the molecular recognition of peptides containing an N-terminal methionine (Met) by the synthetic receptor, cucurbit[8]uirl (Q8), in aqueous solution and with submicromolar affinity. Prior work established that Q8 binds with high affinity to peptides containing aromatic amino acids either by simultaneous binding of two aromatic residues, one from each of two different peptides, or by simultaneous binding of an aromatic residue and its immediate neighbor on the same peptide. The additional binding interface of two neighboring residues suggested the possibility of targeting nonaromatic peptides, which have thus far bound only weakly to synthetic receptors. A peptide library designed to test this hypothesis was synthesized and screened qualitatively for Q8 binding using a fluorescent indicator displacement assay. The large fluorescence response observed for several Met-terminated peptides suggested strong binding, which was confirmed quantitatively by the determination of submicromolar equilibrium dissociation constant values for Q8 binding to MLA, MYA, and MFA using isothermal titration calorimetry (ITC). This discovery of high affinity binding to Met-terminated peptides and, more generally, to nonaromatic peptides prompted a detailed investigation of the determinants of binding in this system using ITC, electrospray ionization mass spectrometry, and ^{1}H NMR spectroscopy for 25 purified peptides. The studies establish the sequence determinants required for high-affinity binding of Met-terminated peptides and demonstrate that cucurbit[n]uril-mediated peptide recognition does not require an aromatic residue for high affinity. These results, combined with the known ability of cucurbit[n]urils to target N-termini and disordered loops in folded proteins, suggest that Q8 could be used

to target unmodified, Met-terminated proteins.



Graphical Abstract. Representation of Q8 binding to MLA and the determinants of binding found in this study. Reprinted with permission from JACS.[25]

Introduction. The development of synthetic compounds that bind target proteins with high affinity and selectivity drives many areas of basic and applied chemical biology.[1-7] Examples of successful approaches to protein recognition include small molecules that bind inside protein cavities,[4] aptamers and mimics of protein secondary structure that bind to large areas on protein surfaces,[1,6,8] and compounds that bind to relatively small areas (i.e., "hotspots") on protein surfaces,[1,6,9] These approaches typically require the selection of high-affinity ligands from combinatorial libraries or the optimization of ligand design based on detailed knowledge of the tertiary structure of the target protein. Protein structure remains difficult to predict from the sequence of amino acids, and thus we assert that design principles for protein recognition based primarily on the sequence of amino acids would be especially useful for their predictive power.

A strong and selective protein interaction requires a large binding interface,[10] which typically comprises several amino acid residues that are neighbors in the folded protein but not typically in the amino acid sequence. Therefore, in order to predict recognition based on sequence, the binding site should therefore be small enough (2-3 amino acid residues) so that the residues involved are adjacent in the sequence. In this situation, each amino acid residue in the complex would need to contribute significantly to the complexation energy—that is, to be efficient ligands. Aromatic ligands are particularly efficient due to their large, flat surfaces, which can provide substantial hydrophobic, van der Waals, and/or electrostatic interactions (e.g., cation-pi).[11-14] It is not surprising, therefore, that aromatic residues are commonly found to be important for stabilizing protein interactions.[9,15-17]

Synthetic receptors are well suited to binding small sites on proteins due to their ability to encapsulate their binding partners within a concave cavity and thus to create an extensive binding interface. Among the synthetic receptors reported to bind peptides noncovalently in aqueous solution,[18] submicromolar equilibrium dissociation constants have been observed only with aromatic peptides, and only for two classes of host: Fujita's coordination cages[19] and cucurbit[n]urils (Qn's, Figure 1). Qn's are particularly effective at targeting aromatic residues in peptides and proteins with sub-micromolar affinities.[20,21] The hydrophobic cavity and two constricted, C=O-lined portals drive the binding of guests with nonpolar and cationic groups. Cucurbit[7]uril (Q7) and cucurbit[8]uril (Q8) typically bind to a single aromatic residue at the N-terminal position of the polypeptide by including the aromatic side chain within the hydrophobic cavity of the Qn and interacting with the N-terminal ammonium group via ion-dipole interactions with the C=O groups at the Qn portal. Although this motif targets only a single residue, it affords an extraordinary degree of selectivity because the N-terminal residue is a unique binding epitope in the polypeptide chain comprising the N-terminal ammonium group proximal to the side chain of the first residue. Remarkably, this selectivity is sufficient to enable protein recognition in simple and complex mixtures.[22,23]

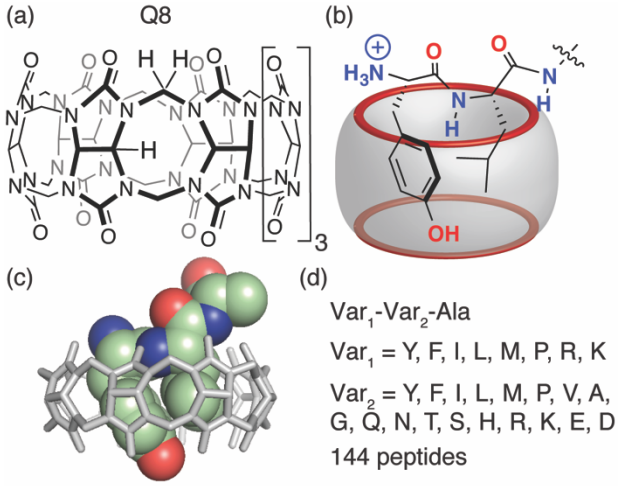


Figure 1. (a) Chemical formula of cucurbit[8]uril. (b) Schematic of the complex of Q8 with the peptide YLA. (c) Semiempirical model of the Q8•YLA complex based on 2D-NMR data. (d) Peptide library screened in this study. Reprinted with permission from JACS.[25]

Despite these successes, the utility of Qn-based peptide and protein recognition is inherently limited by the single-residue binding site. Recently, we reported the discovery

that Q8 can bind with high affinity and selectivity to certain peptides that fold to include the side chains of two neighboring residues within the Q8 cavity (Figure 1).[24] This "pair inclusion" motif enabled the submicromolar binding of the Tyr-containing peptides YLA, YYA, YFA, and YKA. The additional binding interface generated by the side chains of the N-terminal Tyr and its immediate neighbor suggests that the ligand efficiency of each residue can be reduced somewhat. Therefore, we hypothesized that an N-terminal aromatic residue may not be necessary for high-affinity binding if two sufficiently large residues could be included simultaneously. To test this hypothesis, we carried out a further exploration of the pair inclusion motif for peptide binding by Q8[25].

Results and Discussion

Design and Screen of a Peptide Library. A peptide library was designed to explore the scope and limitations of the pair inclusion motif. Past studies showed that only aromatic residues at the first position in the chain yielded high affinity binding to Q7 and Q8,[20,21] presumably due to their ligand efficiency. The discovery of the pair inclusion motif, however, inspired us to ask whether non-aromatic N-terminal residues may bind to Q8 with high affinity when paired with a second residue. The wealth of Qn literature teaches us that nonpolar and cationic functional groups are preferred for guest binding,[26,28] and we know that the displacement of high-energy water molecules from the Qn cavity is the dominant driving force for guest binding.[29] Therefore, the library of 144 tripeptides (Figure 1) of sequence X1-X2-Ala was designed to vary the first position (i.e., X1) among amino acids with large side chains that are hydrophobic or cationic, including Tyr, Phe, Ile, Leu, Met, Pro, Arg, and Lys. The second position in the chain was varied (i.e., X2) among 18

genetically encoded amino acids. Trp was omitted due to its incompatibility with the fluorescence assay. Cys was omitted due to its propensity to form disulfide bonds. All peptides contained Ala at the third position, a C-terminal primary amide group and a free amine at the N-terminus.

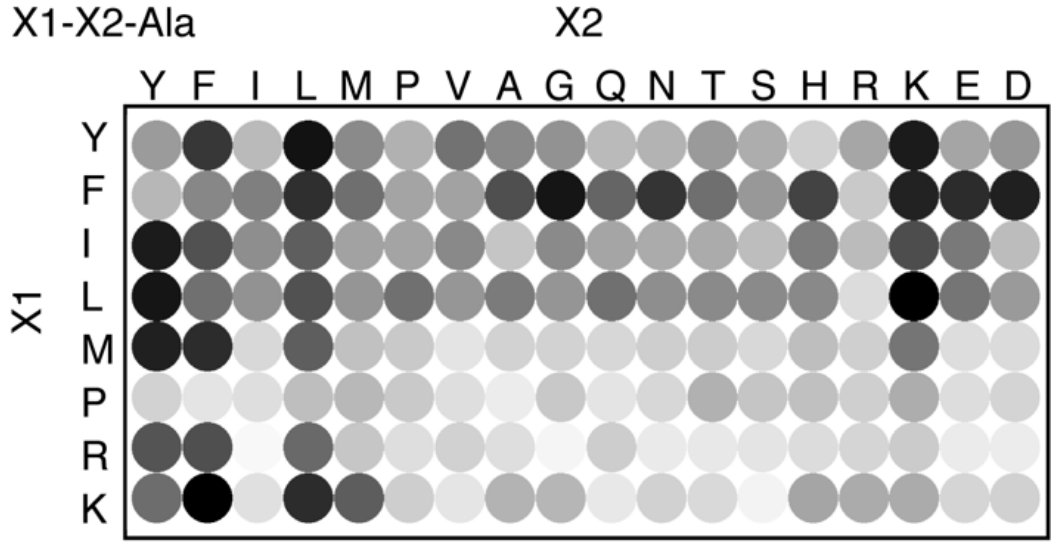


Figure 2. Fluorescence assay for the binding of Q8 to the library of 144 peptides of sequence Var_1-Var_2-Ala. The density of each circle correlates to the average degree of fluorescence enhancement observed upon adding 200 μM peptide to 40 uM Q8•MBBI complex at room temperature in 10 mM sodium phosphate, pH 7.0. Reprinted with permission from JACS.[25]

The peptide library was synthesized by Fmoc-scheme solid-phase synthesis on Rink amide resin using Synphase Lanterns to enable parallel synthesis, as described previously.[24,30] The library was screened for binding to Q8 using a fluorescent indicator displacement assay, as described previously.[30] The competitive displacement of MBBI from Q8 by a peptide analyte causes an increase in the observed fluorescence.[24] In all cases, we observed an increase in fluorescence intensity upon the addition of peptide to a mixture of Q8 and MBBI (Figure 2 and Table S1). Relative increases in fluorescence correlate to binding affinity and ranged from 11.6% to 44.3% at sample concentrations of 200 μM peptide, 40 μM Q8, and 40 μM MBBI in 10 mM sodium phosphate, pH 7.0. In general, the

observed changes in fluorescence intensity were greater for peptides containing the Tyr, Phe, Ile, and Leu at the N-terminus as well as peptides containing Tyr, Phe, Leu and Lys at the second position. It is important to note that peptides containing N-terminal Phe are known to form stable 2:1 peptide:Q8 complexes,[32] which can also displace MBBI and lead to a large increase in fluorescence.[24] The series containing N-terminal Pro showed a weak response overall. By contrast, the series containing N-terminal Met, Lys, and Arg displayed a large fluorescence increase for only a few sequences, which suggested significant sequence selectivity.

In-Depth Study of Met-Terminated Tripeptides. We were particularly excited about the potential to recognize Met-terminated peptides because eukaryotic proteins are expressed with N-terminal Met. Therefore, we conducted a detailed investigation of the thermodynamic and structural determinants of binding for a representative series of four purified Met-terminated tripeptides. MFA, MYA, MLA, and MKA were selected as leads from the fluorescence screen that represent aromatic, aliphatic, and basic side chains at the second position. The peptide MAA was included as a control containing an N-terminal Met and a minimal sidechain at the second position. The binding thermodynamics of Q8 to these peptides were determined by isothermal titration calorimetry (ITC) at 300 K in 10 mM sodium phosphate, pH 7.0 (Table 1 and Figures S1, S4, S7, S13, S14).

The ITC results show that Q8 binds to MFA, MYA, MLA, and MKA, with equilibrium dissociation constant (K_d) values ranging from 0.14-2.6 μM. MAA showed weak binding (K_d > 100 μM). The peptide:Q8 stoichiometry for complexes containing MFA, MYA, MLA, and MKA was determined to be 1:1 (Q8:pepide). All 1:1 complexes in this study

were confirmed by electrospray ionization time-of-flight mass spectrometry (ESI-TOF-MS, Figures S26-S50). Binding was enthalpically driven and entropically unfavorable. These results confirm the qualitative trends observed in the fluorescence screen, and they suggest a mode of binding similar to the Q8•YLA complex. Based on the significantly higher affinity of Q8 for MFA, MYA, MLA, and MKA versus MAA, we conclude that a second large residue is necessary for high affinity binding.

Table 1. Thermodynamic data for the binding of Q8 to MXA peptides.[25]

Peptide	K_d (μM)[a]	ΔH[a] (kcal mol^{-1})	-TΔS[b] (kcal mol^{-1})
MFA	0.14 (±0.01)	-20.1 (±0.1)	10.6 (±0.1)
MYA	0.25 (±0.01)	-18.2 (±0.2)	9.3 (±0.1)
MLA	0.72 (±0.09)	-15.8 (±2.6)	7.4 (±2.5)
MKA	2.6 (±0.3)	-13.7 (±0.5)	6.0 (±0.6)
MAA	>100		

[a] Mean values measured from at least three ITC experiments at 300 K in 10 mM sodium phosphate, pH 7.0. [b] Entropic contributions to the free energy of binding were calculated from the K_d and ΔH values, with error propagated from those of K_d and ΔH.

We were surprised to observe such high affinities for these complexes due the lack of an N-terminal aromatic residues. To the best of our knowledge, this is the first report of submicromolar binding of a synthetic receptor to peptides without an aromatic residue or to peptides with an N-terminal Met residue. These results significantly expand the scope of peptide sequences that can be targeted predictively by synthetic agents. It also opens the door to targeting natural and recombinant eukaryotic proteins, which are all translated with an N-terminal Met. Given the potential utility of this discovery, the remainder of this chapter is concerned with determining the sequence-activity relationships for the binding of Q8 to Met-terminated peptides.

Sequence-Activity Relationships for Pair Inclusion in Tripeptides. Prior work on the Q8•YLA system demonstrated >100-fold reduction in binding affinity to Tyr-Leu when separating the Tyr and Leu residues by an Ala (e.g., YAL), or when moving the Tyr-Leu binding site away from the N-terminus (e.g., AYL).[24] Given the similar results observed thus far between the Q8•YLA and Q8•MXA complexes, we wanted to confirm that these sequence-activity relationships also apply to the Met-terminated peptides. We were also curious about the effects of reversing the sequence order of the inclusion pair (e.g., YLA vs. LYA), which had not been addressed previously. To answer these questions, we designed three series of peptides of sequence MAX, AMX, and XMA, as sequence variants of the parent MXA, with X chosen to be Tyr, Leu, and Lys in order to represent aromatic, aliphatic, and basic residues, respectively. Thermodynamic constants for the binding of these nine peptides to Q8 were determined by ITC at 300 K in 10 mM sodium phosphate, pH 7.0 (Table 2 and Figures S2, S3, S5, S6, S8-S12).

Table 2. Thermodynamic data for the binding of Q8 to MXA vs. MAX, AMX, and XMA peptides.[25]

Peptide	K_d (μM)[b]	ΔH[b] (kcal mol^{-1})	-TΔS[c] (kcal mol^{-1})
MYA[a]	0.25 (±0.01)	-18.2 (±0.2)	9.3 (±0.1)
MAY	>100		
AMY	6.1 (±0.5)	-13.8 (±0.2)	6.6 (±0.3)
YMA	1.3 (±0.2)	-16.6 (±0.8)	8.5 (±0.8)
MLA[a]	0.72 (±0.1)	-15.8 (±2.6)	7.4 (±2.5)
MAL	>100		
AML	>100		
LMA	0.60 (±0.11)	-12.1 (±0.2)	3.5 (±0.2)
MKA[a]	2.6 (±0.4)	-13.7 (±0.5)	6.0 (±0.6)
MAK	>100		
AMK	>100		
KMA	0.89 (±0.01)	-10.9 (±0.2)	2.6 (±0.2)

[a] Shown again for reference. [b] Mean values measured from at least three ITC experiments at 300 K in 10 mM sodium phosphate, pH 7.0. [c] Entropic contributions to the free energy of binding were calculated from the K_d and ΔH values, with error propagated from those of K_d and ΔH.

Compared to the MXA peptides, the MAX peptides retain the N-terminal Met but separate the target pair of residues by an Ala. The ITC data show that this change reduces binding affinity considerably, with K_d values >100 μM for all MAX peptides. Therefore, we conclude that the target residues need to be adjacent in sequence to achieve high affinity binding.

Compared to the MXA peptides, the AMX peptides keep the target pair of residues together but move the pair one residue away from the N-terminus. The ITC data show that this change reduces binding affinity considerably, with K_d values >100 μM for AML and AMK. The K_d value for AMY is 6.1 μM, which is 24-fold weaker affinity than MYA but still considerably stronger than AML or AMK. In prior work, we reported that Q8 binds AYL with a K_d value of 3.1 μM under identical conditions as reported here. These results show that Tyr can increase binding affinity for Q8 in several contexts.

Effects of the Third Residue. Given the preference of Q8 for binding a pair of neighboring residues at the N-terminus, we wanted to study the effects of the third residue on binding. We knew from the Q8•YLA study that the alanine methyl group experiences a downfield NMR chemical shift perturbation upon Q8 binding, suggesting that it could be interfering sterically with complex formation. Therefore, we designed a series of peptides that substitute Ala with Gly at the third position. In a tripeptide, however, the third position is also the C-terminus, and thus we extended the chain to five residues. Finally, we installed Tyr at the C-terminal position in order to facilitate quantifying the peptides by UV spectroscopy. As predicted from prior work[21] and confirmed by the ITC and NMR results described below, we know that this C-terminal Gly-Tyr sequence binds weakly to Q8 and does not compete effectively with pair inclusion.

Table 3. Thermodynamic data for the binding of Q8 to MXA vs. MXGGY and MXZGY peptides.[25]

Peptide	K_d (μM)[b]	ΔH[b] (kcal mol^{-1})	-TΔS[c] (kcal mol^{-1})
MYA[a]	0.25 (±0.01)	-18.2 (±0.2)	9.3 (±0.1)
MYGGY	0.16 (±0.01)	-19.4 (±0.4)	10.1 (±0.4)
MLA[a]	0.72 (±0.09)	-15.8 (±2.5)	7.4 (±2.5)
MLGGY	0.30 (±0.01)	-16.9 (±0.1)	7.9 (±0.1)
LMA[a]	0.60 (±0.11)	-12.1 (±0.2)	3.5 (±0.2)
LMGGY	0.16 (±0.02)	-22.9 (±0.2)	13.5 (±0.3)
MKA[a]	2.6 (±0.4)	-13.7 (±0.5)	6.0 (±0.6)
MKGGY	0.42 (±0.02)	-16.9 (±0.1)	8.2 (±0.1)
MKAGY	0.89 (±0.07)	-16.7 (±0.3)	8.3 (±0.4)
MKVGY	6.1 (±0.4)	-15.9 (±0.5)	8.8 (±0.5)

[a] Shown again for reference. [b] Mean values measured from at least three ITC experiments at 300 K in 10 mM sodium phosphate, pH 7.0. [c] Entropic contributions to the free energy of binding were calculated from the K_d and ΔH values, with error propagated from those of K_d and ΔH.

The ITC data (Table 3 and Figures S16-S21) show that the change from MXA to MXGGY increases binding affinity 2-4-fold, which we attribute to a reduction in steric hindrance by removing the methyl side chain. To further explore this effect, we designed two additional peptides, MKAGY and MKVGY, which systematically increase steric bulk at the third position within a pentapeptide. As expected, the binding affinity decreases as a function of the size of the side chain at the third position, with a 15-fold difference in affinity between MKGGY and MKVGY.

Additional Sequence Determinants at the Second Position. To further elucidate the sequence-activity relationships in this binding motif, we asked whether other hydrophobic and basic residues could be tolerated at the second position. As analogues of MLGGY, the peptides MIGGY and MVGGY were tested for binding to Q8, and surprisingly showed low affinity by ITC (Table 4 and Figures S15 and S22-S25). We believe this is due to branching at the beta carbon that occurs in Val and Ile residues but not in Leu. Based on the published semiempirical model of the Q8•YLA complex,[24] that branching would force an alkyl group directly into the portal of Q8, thereby causing an unfavorable steric interaction. As an analogue to MKGGY, the peptide MRGGY was tested for binding to Q8 by ITC and found to bind 5-fold less tightly than MKGGY. We obtained MHGGY commercially but were unable to solubilize it. We speculate that His should bind poorly because its cationic side chain would be destabilized upon encapsulation within the hydrophobic Q8 cavity, whereas side chains of Lys and Arg are long enough to thread the cationic groups all the way through Q8. Ser and Gly residues were also installed at the second position, and the resulting MSGGY and MGGGY peptides showed low affinity for

Q8 by ITC. These results support the need for a large side chain at the second position that is either hydrophobic or basic but not branched at the beta carbon.

Table 4. Thermodynamic data for the binding of Q8 to MXGGY peptides.[25]

Peptide	K_d (μM)[b]	ΔH[b] (kcal mol^{-1})	-TΔS[c] (kcal mol^{-1})
MLGGY[a]	0.30 (±0.01)	-16.9 (±0.4)	7.9 (±0.4)
MIGGY	>100		
MVGGY	>100		
MKGGY[a]	0.42 (±0.02)	-16.9 (±0.1)	8.2 (±0.1)
MRGGY	2.1 (±0.2)	-12.4 (±0.4)	4.6 (±0.5)
MSGGY	>100		
MGGGY	>100		

[a] Shown again for reference. [b] Mean values measured from at least three ITC experiments at 300 K in 10 mM sodium phosphate, pH 7.0.
[c] Entropic contributions to the free energy of binding were calculated from the K_d and ΔH values, with error propagated from those of K_d and ΔH.

Structural Characterization. ^{1}H NMR spectroscopy was used to characterize the 25 purified peptides described above (Figures 3a and S51-S108). For each peptide, we include an overlay of one-dimensional spectra of the peptide in the absence of Q8, at a mole ratio of 0.5:1 Q8:peptide (which, importantly, is also 2:1 peptide:Q8 ratio), and at a mole ratio of 1:1 Q8:peptide. The mole ratios were determined by comparing peak integrations in the NMR spectra. When necessary, two-dimensional ^{1}H-^{1}H correlation spectroscopy (COSY) was used to assign the signals of bound and unbound peptides. Through-space interactions were characterized using two dimensional ^{1}H-^{1}H nuclear Overhauser effect spectroscopy (NOESY) for some of the pure peptide samples and two-dimensional ^{1}H-^{1}H rotating frame nuclear Overhauser effect spectroscopy (ROESY) for some of the Q8•peptide complexes.

The NMR data were analyzed for changes to the peptide spectra induced by the addition of Q8, including: 1) perturbation in chemical shift, in which upfield perturbation indicates

the positioning of that group within the cavity of Q8, and downfield perturbation indicates the positioning of that group proximal to carbonyl groups whether at the portal or in the peptide backbone;[32] and 2) resolution of the signals corresponding to pairs of geminal protons in the bound state. In the unbound state, the peptide chain is mobile, and the magnetic environment at both protons is similar on the NMR timescale. In the bound state, however, restricted mobility can lead to resolution of the two signals if the protons are in different magnetic environments. These spectral characteristics are discussed here for each the following categories of peptide structure: 1) the Met residue, which is present in all peptides; 2) the residue adjacent to Met for peptides that bind with high affinity (e.g., Leu in MLA or Lys in MKVGY; 3) the residue adjacent to the target pair (e.g., Ala in MLA or Val in MKVGY); and 4) C-terminal Tyr residues in the pentapeptides.

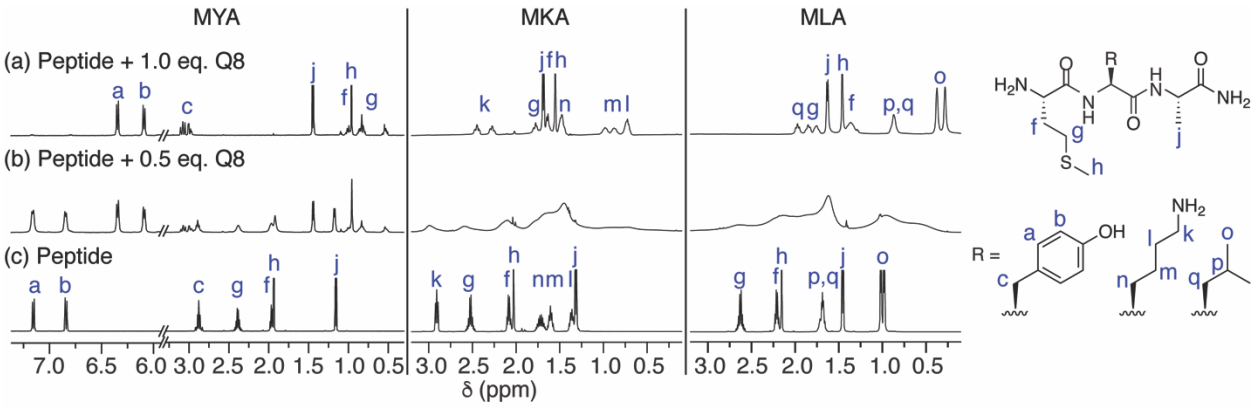


Figure 3a. 500 MHz ^{1}H NMR titrations of samples containing (a) a 1:1 mixture of Q8:peptide, (b) a 0.5:1 mixture of Q8:peptide, and (c) peptide. Spectra were acquired at 25 °C in D_2O. Reprinted with permission from JACS.[25]

For the Met residues in all peptides, we observe an upfield perturbation in chemical shifts of the signals corresponding to the side chain protons. This indicates that the Met side chain of all peptides is included within the Q8 cavity. For tripeptides that bind with

measurable affinity to Q8, *vide supra*, we observe increased resolution of signals corresponding to the beta and gamma geminal proton pairs in the Met side chain upon Q8 binding. When the neighboring residue is Tyr or Phe, then the separation between geminal proton pairs on Met is greater than 0.3 ppm, and when the neighboring residue is Lys or Leu, the separation is less than 0.3 ppm. We believe this effect is due to restricted motion of the Met side chain upon binding, which would be exacerbated by a larger neighboring side chain. It is interesting to note that the resolution of geminal signals is the same when the residue neighboring the inclusion pair is Ala or Gly (e.g., MYGGY vs. MYA).

There are differences in the kinetics of chemical exchange on the NMR time scale for the various mixtures of Q8 and peptide. Exchange is slow for all peptides containing the MY pair, but fast for MFA. Exchange is slow for MKGGY, MKAGY, and KMA but fast for MKVGY and MKA. Exchange is slow for AML but fast for MLA. In addition, all weak binders (K_d > 100 μM) as well as the peptides MRGGY, LMA, and LMGGY are fast exchange. The compounds in this study have fairly similar binding constants. For high-affinity binding, the relative kinetics of chemical exchange on the NMR timescale are dictated by the relationship between the unimolecular dissociation rate constant and the difference in chemical shift between the bound and unbound states ($\Delta\delta$). Therefore, for functional groups with similar $\Delta\delta$ values, as observed for many complexes here, the observed differences in chemical exchange kinetics reflect differences in the association rate constants. Unfortunately, the NMR data alone are insufficient to draw more in-depth conclusions about the effects of peptide structure on the binding kinetics.

For peptides that bind tightly to Q8, we characterized the structural effects of the residue neighboring Met upon addition of Q8. We observe upfield perturbation in the

chemical shifts for signals corresponding to protons on the side chains of Tyr, Phe, Leu, and Lys upon Q8 addition. In general, this result demonstrates the inclusion of these residues along with the neighboring Met within the Q8 cavity—that is, pair inclusion. For peptides containing MY and MF sequences, it is interesting to note that the signals corresponding to the beta protons on Tyr and Phe perturb downfield, suggesting their positioning near the portal carbonyl oxygens or near a peptide carbonyl in the complex. For peptides containing ML, the signals corresponding to the geminal beta protons perturb in opposite directions upon binding Q8, leading to unusually large separation and suggesting that the geminal protons project in different directions in the complex. 2D ROESY data show NOE crosspeaks between protons on the neighboring side chains of Met and Leu in the Q8•MLA complex, and of the sidechains of Met and Tyr in the Q8•MYA and Q8•YMA complexes (Figures 3b and S60, S68, S74). These NOEs are not present in the absence of Q8, as observed in the NOESY spectra (Figures S58, S66, S72).

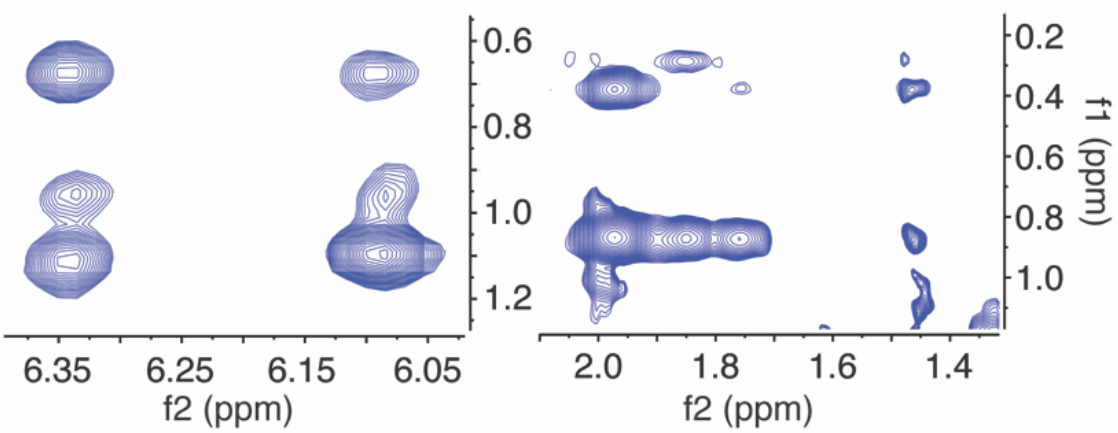


Figure 3b. 500 MHz 2D ROESY spectra of samples containing (a) a 1:1 mixture of Q8:MYA and (b) a 1:1 mixture of Q8:MLA. The regions of the spectra displayed here reveal NOE crosspeaks between protons on neighboring side chains. Spectra were acquired at 25 ºC in D_2O solution. Reprinted with permission from JACS.[25]

The thermodynamic data presented in Table 3 led us to conclude that the residue adjacent to the binding pair can impact the binding affinity, likely due to steric effects. When that residue is Ala, the NMR signal corresponding to the Ala methyl protons perturbs downfield upon the addition of Q8. This result was also observed previously for the Q8•YLA complex and indicates the methyl group is positioned near the Q8 portal. In MKVGY, the signal corresponding to one of the two Val methyl groups perturbs downfield upon addition of Q8. These data support our hypothesis that there may be steric interactions between the Q8 portal and the side chain of the third residue.

In the pentapeptide series, we included a Tyr residue at the C-terminus in order to aid in quantifying the peptides by UV spectroscopy. Prior work led us to predict that we would observe negligible competition for Q8 binding by this residue. To support this prediction, we studied the effects of Q8 addition on the NMR signals corresponding to the side chain of the C-terminal Tyr residues. For the pentapeptides containing an N-terminal inclusion pair (e.g., MLGGY), we observe minimal changes to the spectra of the C-terminal Tyr upon Q8 addition. In MYGGY, there are two Tyr residues, and chemical exchange is slow on the NMR time scale. We observe upfield perturbation of one of these sets of signals at a 1:1 Q8:peptide ratio, and we believe these signals correspond to the Tyr in the MY inclusion pair. For pentapeptides that do not contain a strong-binding inclusion pair, we observe significant broadening and upfield chemical shift perturbation for the C-terminal Tyr residue. This result is consistent with the absence of a competitive binding site.

Conclusions. This study establishes a novel approach to the molecular recognition of Met-terminated peptides and non-aromatic peptides with submicromolar affinity in aqueous solution. It also shows that cucurbit[n]uril-mediated peptide recognition does not require an aromatic residue for high affinity. In the absence of an aromatic residue, additional binding energy can be generated by the inclusion of the side chains of two immediately neighboring residues within the cavity of Q8. The detailed sequence-activity relationships presented here suggest several rules for targeting Met-containing peptides with Q8 (Figure 4). The residue paired with Met should be large and either hydrophobic or cationic (i.e., Leu, Lys, Arg, Tyr, or Phe), but it cannot have branching at the beta carbon (i.e., Val, Ile). The sequence order of the target pair is not as important is its amino acid composition. Targeting the pair at the N-terminus enhances binding affinity in all cases and is necessary for non-aromatic target pairs. The residue adjacent to the target pair on the C-terminal side should be as small as possible to minimize steric interactions.

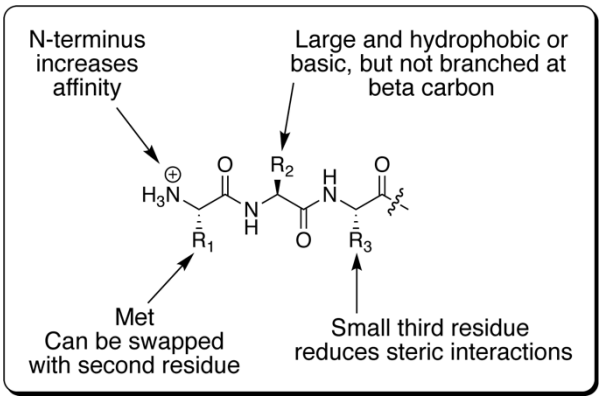


Figure 4. Summary schematic of the sequence determinants that lead to high-affinity binding to Q8. Reprinted with permission from JACS.[25]

Methionyl aminopeptidase (MetAP) is known to remove N-terminal Met from newly translated proteins when the second residue is small and neutral.[33] In the pair inclusion motif, Q8 binds tightly to N-terminal Met only when the second residue is large and hydrophobic or cationic, which should withstand degradation by MetAP. This characteristic combined with ability of cucurbit[n]urils to target N-termini[22] and disordered loops[35] in folded proteins suggests that Q8 could be useful for targeting newly translated

and unmodified proteins. More generally, the high affinity binding of recombinant proteins via a two-residue target site would constitute a truly minimal protein affinity tag that should have negligible impact on protein structure and function.

Experimental

Materials. The following compounds were of analytical purity grade and used without purification: (L)-Fmoc-Ala-OH, (L)-Fmoc-Arg-(Pbf)-OH, (L)-Fmoc-Asn(Trt)-OH, (L)-Fmoc-Asp(OtBu)-OH, (L)-Fmoc-Gly(OtBu)-OH, (L)-Fmoc-Gln(Trt)-OH, (L)-Fmoc-Gly-OH, (L)-Fmoc-His(Trt)-OH, (L)-Fmoc-Ile-OH, (L)-Fmoc-Leu-OH, (L)-Fmoc-Lys(Boc)-OH, (L)-Fmoc-Met-OH, (L)-Fmoc-Phe-OH, (L)-Fmoc-Pro-OH, (L)-Fmoc-Ser(tBu)-OH, (L)-Fmoc-Thr(tBu)-OH, (L)-Fmoc-Tyr(tBu)-OH, (L)-Fmoc-Val-OH, and O-(benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU) (Peptides International), biotech-grade dimethyl formamide (DMF), diisopropylethylamine (DIEA), triisopropylsilane (TIS), trifluoroacetic acid (TFA), piperidine, anhydrous dichloromethane (DCM), and HPLC-grade acetonitrile (Sigma-Aldrich); and monobasic and dibasic sodium phosphate (VWR). SynPhase Rink amide Lantern resins (8 μmol scale) were from Mimotopes. Bulk rink amide MBHA resin was from Peptides International. Deuterium oxide, "100%" (D, 99.96%), was from Cambridge Isotopes. The peptides Met-Ala-Ala, Met-Leu-Ala, Leu-Met-Ala, Ala-Met-Leu, Met-Ala-Leu, Met-Tyr-Ala, Tyr-Met-Ala, Ala-Met-Tyr, Met-Ala-Tyr, Met-Lys-Ala, Lys-Met-Ala, Ala-Met-Lys, Met-Ala-Lys, Met-Tyr-Gly-Gly-Tyr, Met-Leu-Gly-Gly-Tyr, Leu-Met-Gly-Gly-Tyr, Met-Ile-Gly-Gly-Tyr, Met-Val-Gly-Gly-Tyr, Met-Arg-Gly-Gly-Tyr, Met-Ser-Gly-Gly-Tyr, and Met-Gly-Gly-Gly-Tyr were from GenScript Biotech (purity > 95%) with amidation at C-termini and HPLC purification. The peptides Met-Lys-Val-Gly-Tyr, Met-Lys-Ala-Gly-Tyr, and Met-Lys-Gly-Gly-Tyr were also from GenScript Biotech (purity > 95%) with HPLC purification, but with carboxylic acids at C-termini. The peptide Met-Phe-Ala was individually synthesized and purified. Water was obtained from a Barnstead

Nanopure Infinity water system (18 MΩ cm). Cucurbit[8]uril (Q8) was synthesized according to a published procedure.[9] Tetramethylbenzobis-(imidazolium) diiodide (MBBI) was generously provided by Prof. Christopher Bielawski.

A stock solution of 10 mM sodium phosphate buffer was adjusted to pH 7.0 and sterile filtered. The concentration of MBBI was determined by UV spectroscopy (ε_{287} = 20,200 cm^{-1} M^{-1}). The concentration of Tyr-containing peptides was determined by UV spectroscopy (ε_{276} = 1280 cm^{-1} M^{-1}). The concentration of Q8 was determined by calorimetric titration with a standard solution of methyl viologen (ε_{257} = 20,400 cm^{-1} M^{-1}).

Peptide Synthesis. Due to the use of Rink amide resin, all peptides have a primary amide at the C-terminus. The library of 144 peptides used in the fluorescence screen for binding (Figures 2 of the main manuscript and Table S1 of the Supporting Information) were synthesized using parallel fmoc solid-phase synthesis with SynPhase Rink amide Lantern resins and characterized as described previously.[1,8] No purification was performed, and the peptides had an average purity of 79%, as determined by analytical HPLC. The average synthetic recovery was determined to be 67% by UV spectroscopy of peptides containing Tyr.

With the exception of Met-Phe-Ala, all individual peptides for binding characterization were ordered from GenScript. The peptide Met-Phe-Ala was synthesized individually on Rink amide MBHA resin using an automated peptide synthesizer (CS Bio CS 336X) according to the following procedure: The synthesis was carried out on the scale of 0.10 g resin (resin loading was 0.37 mmol/g). Fmoc deprotection was accomplished by shaking the resin in a solution of 20% piperidine in DMF at room temperature for 20

minutes, followed by extensive rinsing with DMF. For each peptide coupling, a mixture of Fmoc-protected amino acid (4 equivalents versus the resin), HBTU (3.9 equivalents) and DIEA (6 equivalents) was dissolved in 10 mL of DMF, mixed for 15 minutes, and then added to the resin. Couplings proceeded at room temperature for 60 minutes, followed by extensive rinsing with DMF. After the final Fmoc deprotection, the resin was transferred to a flask equipped with a course frit and stopcock and rinsed with dichloromethane (i.e., a manual solid phase peptide synthesis vessel), the solvent was drained, and the peptide was cleaved from the resin by treatment with two 10 mL portions of a mixture of 95% TFA, 2.5% TIS, 2.5% water for 10 minutes for the first portion and for one hour for the second portion at room temperature with shaking. The filtrate containing the crude peptide was collected and concentrated by rotary evaporation under high vacuum at 60 ºC for 20 minutes. The residue was dissolved in 30 mL of 0.1% aqueous TFA and purified in three 10 mL batches by reversed phase preparatory HPLC on a Gemini-NX 5 µm C18 column. Fractions were checked by analytical HPLC (Agilent 1100 system with a Phenomenex Gemini-NX C18 column, data not shown), and pure fractions were combined and lyophilized to dryness. Peptide purity and identity were verified by ^{1}H NMR spectroscopy in deuterium oxide on a Varian 500 MHz spectrometer at 25 ºC (Figure S51).

Library Screening by Fluorescence Spectroscopy. A solution containing a 1:1 stoichiometric ratio of Q8 and MBBI was prepared by first dissolving a quantity of MBBI in 10 mM sodium phosphate buffer, pH 7.0, and determining the concentration of the solution by UV spectroscopy (ε_{287} = 20,200 $cm^{-1}M^{-1}$). This standard MBBI solution was adjusted to 80 µM in the same buffer and added to an equimolar quantity of dry Q8 (whose

purity had already been determined by titration). The resulting mixture was solubilized by mixing and brief ultrasonication followed by heating at 50-60 °C for 15-20 minutes. The resulting colorless solution of 1:1 Q8 MBBI complex was allowed to cool to room temperature followed by sterile-filtration (0.4 μm, Teflon).

Fluorescence experiments were carried out in Corning 96-well, black, flat-bottomed plates. Q8•MBBI reference sample solutions were prepared by mixing 50 μL of 80 μM Q8•MBBI stock solution with an equal volume of 10 mM phosphate buffer. Peptide + Q8•MBBI solutions were prepared by mixing 50 μL of 80 μM Q8•MBBI stock solution with an equal volume of 400 μM peptide stock solution. Therefore, the final concentrations were 40 μM Q8•MBBI and 200 μM peptide. The solutions were mixed by pipetting before reading the fluorescence intensity on a Tecan M200 Pro fluorescence plate reader (top reading mode; 297 nm excitation, 5 nm bandwidth; 340 nm emission, 20 nm bandwidth).

Electrospray Ionization Mass Spectrometry (ESI-MS). Mass spectra were acquired by infusing into using an Agilent 6230 TOF LC/MS mass spectrometer with an electrospray ion source in the positive ion mode. Samples were dissolved in pure water at concentrations of 55 μM purified peptide in the absence and presence of one molar equivalent of Q8. The sample containing Q8 and MKA was analyzed at 5.5 mM.

Isothermal Titration Calorimetry (ITC). Titrations were carried out in 10 mM sodium phosphate buffer (pH 7.0) at 300 K on a VP-ITC calorimeter (Microcal, Inc). In a typical experiment, Q8 was in the sample cell at a concentration in the range 0.03 – 0.1 mM, and the peptide was in the injection syringe at a concentration in the range 0.5 – 1.2 mM. The

titration schedule consisted of 28 consecutive injections of 2-10 μL with at least a 200 s interval between injections. Heats of dilution, measured by titrating beyond saturation, were subtracted from each data set. All solutions were degassed prior to titration. The data were analyzed using Origin software and fit by nonlinear regression to the binary equilibrium model (non-interacting sites) supplied with the software in order to determine the molar enthalpy, equilibrium association constant, and stoichiometry of binding. These values were used to calculate the free energies of binding and the entropic contributions to the free energies of binding.

Nuclear Magnetic Resonance (NMR) Spectroscopy. ^{1}H NMR spectra were acquired in deuterium oxide at 25 ºC on a Varian 500 MHz spectrometer. Presaturation of the residual protiated solvent signal at 4.77 ppm was used as necessary in the acquisition of one-dimensional ^{1}H NMR spectra as well as two-dimensional ^{1}H-^{1}H COSY spectra, ^{1}H-^{1}H NOESY spectra, and ^{1}H-^{1}H ROESY spectra. 2-D COSY spectra of all 1:1 Q8:peptide mixture were acquired with a spectral width of 8012.8 Hz. 2-D NOESY and ROESY spectra were acquired using a 1 s relaxation delay and 200 ms mixing time.

References

(1) Smith, B.: *Synthetic Receptors for Biomolecules: Design Principles and Applications*; Royal Society of Chemistry: London, UK, **2015**.

(2) Kubota, R.; Hamachi, I. *Chem. Soc. Rev.* **2015**, *44*, 4454-4471.

(3) Hugget, J.; O'Grady, J.: *Molecular Diagnostics: Current Research and Applications*; Horizon Press: Norfolk, UK, **2014**.

(4) Patrick, G. L.: *An Introduction to Medininal Chemistry*; 5th Edition ed.; Oxford University Press: New York, NY, **2005**.

(5) Bertozzi, C. R. *Acc. Chem. Res.* **2011**, *44*, 651-653.

(6) Yin, H.; Hamilton, A. D. *Angew. Chem. Int. Ed.* **2005**, *44*, 4130-4163.

(7) Peczuh, M. W.; Hamilton, A. D. *Chem. Rev.* **2000**, *100*, 2479-2494.

(8) Bunka, D. H. J.; Stockley, P. G. *Nature Rev. Microbiol.* **2006**, *4*, 588-596.

(9) Moriera, I. S.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2007**, *68*, 803-812.

(10) Day, E. S.; Cote, S. M.; Whitty, A. *Biochem.* **2012**, *51*, 9124-9136.

(11) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42* , 1210-1250.

(12) Salonen, L. M.; Ellermann, M.; Diederich, F. *Angew. Chem., Int. Ed*. **2011**, *50*, 4808-4842.

(13) Martinez, C. R.; B. L. Iverson, B. L. *Chem. Sci.* **2012**, *3*, 2191-2201.

(14) Mahadevi, A. S.; Sastry, G. N. *Chem. Rev*. **2013**, *113*, 2100-2138.

(15) Ma,, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. *Proc. Natl. Acad. Sci. U. S. A*. **2003**, *100*, 5772-5777.

(16) Espinoza-Fonseca, L. M. *Mol. Biosyst.* **2012**, *8*, 237-246.

(17) Rahman, M. M.; Muhseen, Z. T.; Junaid, M.; Zhang, H. *Curr. Protein Pept. Sci.* **2015**, *16*, 502-512.

(18) By contrast, submicromolar binding of His-containing peptides has been observed using coordinate-covalent interactions. See Hortalá, M. A.; Fabbrizzi, L.; Marcotte, N.; Stomeo, F.; Taglietti, A. *J. Am. Chem. Soc.* **2003**, *125*, 20-21.

(19) Tashiro, S.; Tominaga, M.; Kawano, M.; Therrien, B.; Ozeki, T.; Fujita, M. *J. Am. Chem. Soc.* **2005**, *127*, 4546-4547.

(20) Urbach, A. R.; Ramalingam, V. *Israel J. Chem.* **2011**, *51*, 664.

(21) Bockus, A. T.; Adam R. Urbach: in Aromatic Interactions: Frontiers in Knowledge and Application, Eds. Darren W. Johnson and Fraser Hof, Royal Society of Chemistry: Cambridge, UK, **2017**.

(22) Chinai, J. M.; Taylor, A. B.; Ryno, L. M.; Hargreaves, N. D.; Morris, C. A.; Hart, P. J.; Urbach, A. R. *J. Am. Chem. Soc.* **2011**, *133*, 8810-8813.

(23) Li, W.; Bockus, A. T.; Vinciguerra, B.; Isaacs, L.; Urbach, A. R. *Chem. Commun.* **2016**, *52*, 8537-8540.

(24) Smith, L. C. Smith, Leach, D. G.; Blaylock, B. E.; Ali, O. A.; Urbach, A. R. *J. Am. Chem. Soc.* **2015**, *137*, 3663-3669.

(25) Hirani, Z.; Taylor, H. F.; Babcock, E. F.; Bockus, A. T.; Varnado, C. D. Jr.; Bielawski, C. W.; Urbach, A. R. *J. Am. Chem. Soc.*, **2018**, *140,* 12263-9.

(26) Mock, W. L. *Top. Curr. Chem.* **1995**, *175*, 1-24.

(27) Lagona, K.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. *Angew. Chem. Int. Ed.* **2005**, *44*, 4844-4870.

(28) Barrow, S. J.; Kasera, S.; Rowland, M. J.; del Barrio, J.; Scherman, O. A. *Chem. Rev.* **2015**, *115*, 12320-12406.

(29) Biedermann, F.; Nau, W. M.; Schneider, H.-J. *Angew. Chem. Intl. Ed.* **2014**, *53*, 11158-11171.

(30) Ali, O. A.; Olson, E. M.; Urbach, A. R. *Supramol. Chem.,* **2013**, *25*, 863-368.

(31) Biedermann, F.; Rauwald, U.; Cziferszky, M.; Williams, K. A.; Gann, L. D.; Guo, B. Y.; Urbach, A. R.; Bielawski, C. W.; Scherman; O. A. *Chem. Eur. J.* **2010**, *16*, 13716-13722.

(32) Heitmann, L. M.; Taylor, A. D.; Hart, P. J.; Urbach, A. R. *J. Am. Chem. Soc*, **2006**, *128*, 12574-12581.

(33) Moon, K.; Kaifer, A. E. *Org. Lett.* **2004**, *6*, 185-188.

(34) Hirel, P.-H.; Schmitter, J.-M.; Dessen, P.; Fayat, G.; Blanquet, S. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8247-8251.

(35) Sonzini, S.; Marcozzi, A.; Gubeli, R. J.; van der Walle, C. F.; Ravn, P.; Herrmann, A.; Scherman, O. A. *Angew. Chem. Int. Ed.* **2016**, *55*, 14000-14004.

Chapter Three

Creating Receptor Binding Sites by Selective Enzyme Processing

Abstract. The ability to recognize and label peptides and proteins site-selectively has broad applications in basic and applied science. To target or label proteins selectively, however, there needs to be an accessible binding site for the receptor. Cucurbit[7]uril (Q7) targets aromatic residues selectively at the N-terminal position. Most proteins, however, do not contain a N-terminal aromatic residue. The Urbach lab has been able to generate an N-terminal aromatic residue using a non-selective protease for Q7 to bind. This chapter describes efforts to extend this method to processing folded proteins with Q7 and to processing peptides with Q8.

Introduction. Proteases, also known as proteolytic enzymes, are enzymes that catalyze the breakdown of proteins and peptides by hydrolysis of peptide bonds[1]. They cleave protein substrates either from the N or C termini (exopeptidase) or in the middle of a molecule (endopeptidase). The macrocyclic host, cucurbit[7]uril (Q7), can exhibit specific inhibitory effects on the activity of proteases by binding to the substrate peptides. However, this approach requires a high concentration (mM) of Q7 in order to be effective (Figure 1)[2]. Q7 can recognize peptides and proteins that contain aromatic amino acid residues with equilibrium dissociation constant (K_d) values ranging from 0.01-100 μM[3]. The recognition of N-terminal aromatic residues by Q7 is mediated by the simultaneous inclusion of the aromatic side chain within the hydrophobic cavity of Q7 and ion-dipole interactions between the positively charged N-terminal ammonium group and the carbonyl oxygens lining the entrances to the cavity of Q7[3].

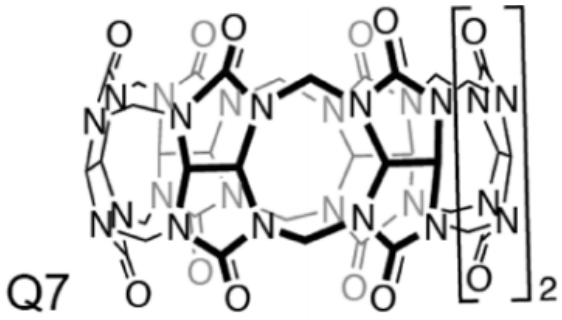


Figure 1. Model of Q7.

In 2013, the Urbach lab performed a study that demonstrated Q7 inhibiting a nonspecific exopeptidase, aminopeptidase N (APN), sequence-specifically. Upon enzymatic digestion of the first three residues of the pentapeptide, TGAFM, an N-terminal phenylalanine residue is generated (Figure 2). Q7 binds selectively to this N-terminal residue with a K_d value of approximately 1 μM[4]. The resulting Q7•FM complex was stable to proteolytic digestion for at least 24 hours. Digestion was monitored using analytical HPLC and mass spectrometry.

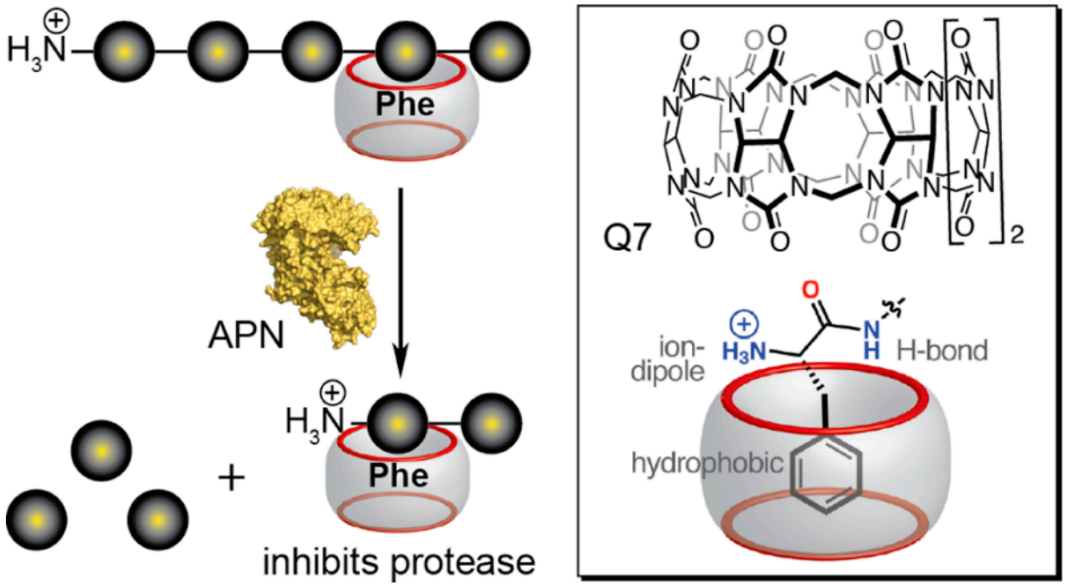


Figure 2. (Left) Schematic of the inhibition of APN-mediated peptide digestion at a Phe residue in the presence of Q7. (Right) Chemical formula of Q7 and a schematic of the molecular recognition of a N-terminal Phe by Q7. Reprinted with permission from JACS[3].

In order to reproduce the results from the 2013 paper and become familiar with the methodology, an experiment was carried out in which an aminomethyl group was added to at the 4-position of the phenylalaine side chain to make the residue, AmF (Figure 3). This residue was placed in the middle of the pentapeptide, TGGAmFM, and at the N-terminus of a dipeptide, AmFM. Q7 binds to a N-terminal AmF with a low nanomolar binding affinity (K_d = 0.95 nM) compared to the native peptide, FGG (K_d = 0.31 μM).[4] Due to AmF being effective at nanomolar concentrations, the generated N-terminal AmF group makes a sufficient model to reproduce the published results[4]. The degradation of the pentapeptide, TGGAmFM, by APN in the presence of Q7 yielded the dipeptide AmFM in a similar fashion as in the 2013 paper.

Here, we test conditions for monitoring digestion with APN of the protein, Hen Egg White Lysozyme (HEWL). HEWL was chose as our model protein because it contains a single polypeptide chain made up of 129 residues, it is one of the most characterized proteins, and it contains a Phe at the third position (near the N-terminus). By digesting HEWL, it could be possible to generate an N-terminal Phe residue to allow for the site-selective binding of Q7. This processing technique on proteins offers the opportunity for binding sites to be created quickly, conveniently, and selectively, which could be useful in the processing of other proteins to generate affinity tags without chemical or genetic modification.

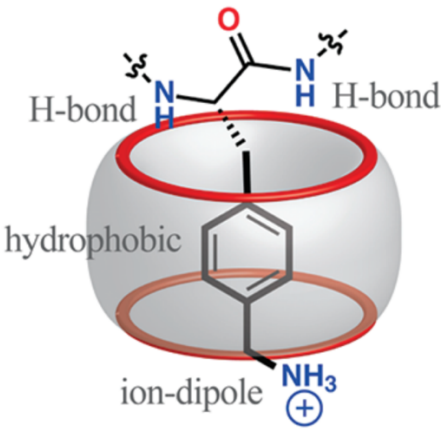


Figure 3. Schematic illustration of Q7 binding to an N-terminal amF residue from cooperative hydrophobic and electrostatic interactions. Reprinted with permission from JACS[3].

Results and Discussion

In addition to discussing our efforts to establish and optimize the conditions to achieve HEWL digestion, this chapter also discusses a scheme to use this processing technique on peptides that contain dipeptide sequences for Q8 recognition, as discussed in Chapter 2.

Reproducing the APN Protocol. Before we could use the APN processing technique on HEWL and peptides using Q7 and Q8, respectively, we needed to reproduce the conditions of the published experiments[4]. This reproduction of the results would enable us to determine the concentration ranges within which to work. The LC-MS is useful because it allows us to monitor quantitatively the conversion of the starting material to the product. On the LC, using a reversed phase C18 column, the starting material and product have different retention times. The starting material is the pentapeptide, while the product is the dipeptide•Q7 complex. The MS confirms the mass of the starting materials and the products, which allows us to track the digestion process unambiguously.

Confirmation of Enzymatic Activity. The enzymatic activity was determined using a Tecan Infinite 200 Pro fluorescent plate reader. A fluorogenic substrate, L-Met-7-amido-4-methylcoumarin (AMC), was used for the real-time monitoring of enzymatic activity. A series of different AMC concentrations (0 µM, 15.6 µM, 31.3 µM, 62.5 µM, 125 µM, 250 µM, 500 µM, and 1000 µM) was used to determine which concentrations would be in the dynamic range of the plate reader, and thus allow for the monitoring of the enzymatic activity by fluorescence changes. After the addition of APN, the hydrolysis of the substrate, which is indicated by an increase in fluorescence intensity, was observed using an optimal concentration of 15.6 µM AMC (Figure 4).

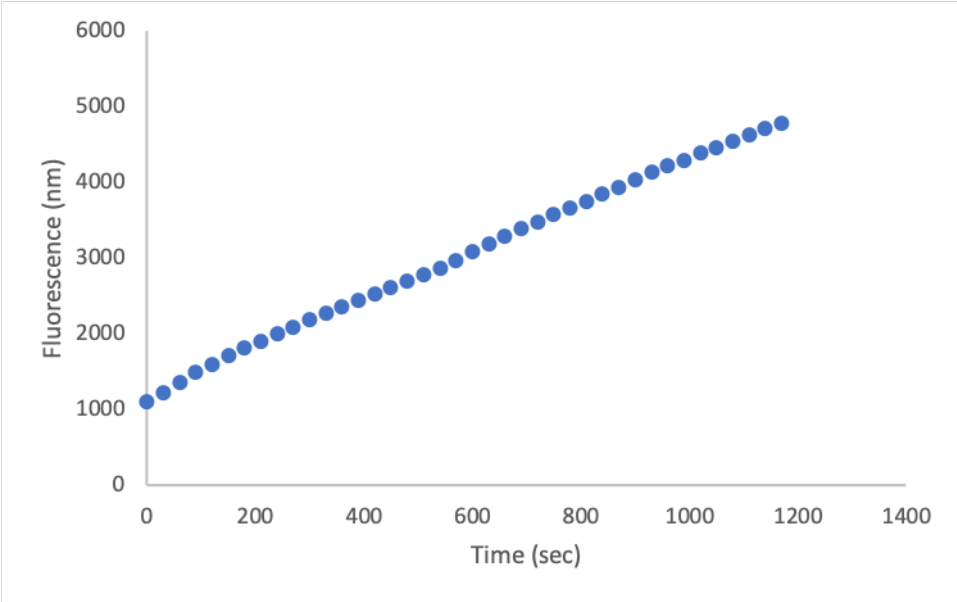


Figure 4. Continuous fluorescence plot (λ_{exc} = 380, λ_{obs} = 460) with AMC (15.6 µM) upon the addition of APN (7.14 nM), in 30 mM HEPES (pH 7.37), 250 mM NaCl, 100 µM $CoCl_2$, and 4% DMSO at 37 ℃.

Enzyme Kinetics. From the fluorescence assay of the different substrate concentrations, the kinetics for APN were characterized using Michaelis-Menten analysis (Figure 5). The enzyme kinetic values for APN are shown in Table 1.

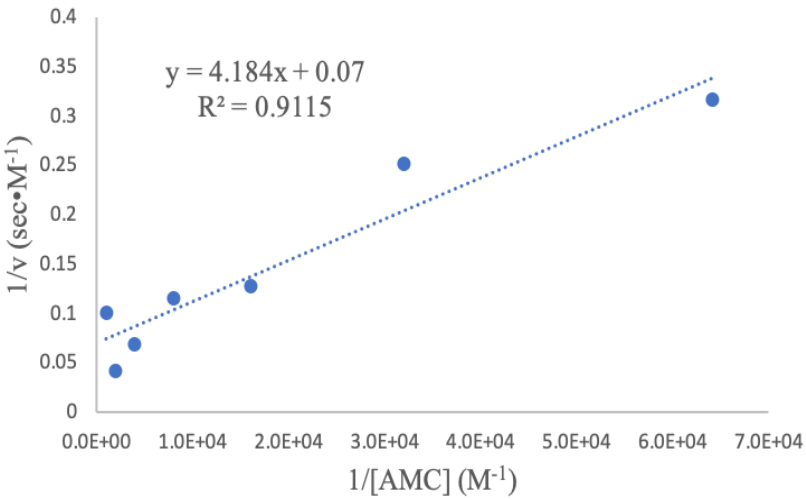


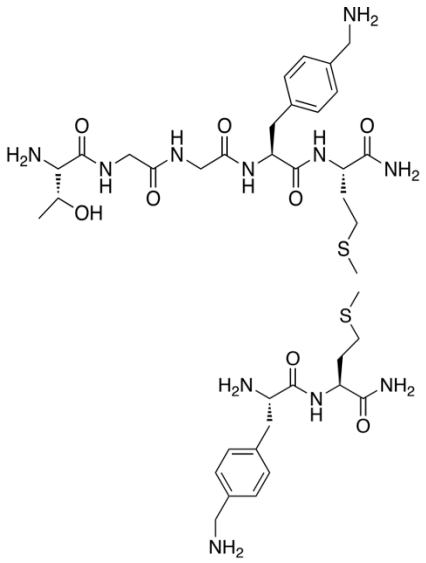
Figure 5. Representative Lineweaver-Burk Plot for APN.

Table 1. Enzyme Kinetics for APN.

V_{max} (mol•s^{-1})	15.9 (±1.9)
K_M (M)	99.9 (±39.)
k_{cat} (s^{-1})	2.3 (±0.3) x 10^9
k_{cat}/K_M ($s^{-1}M^{-1}$)	2.5 (±0.8) x 10^7

Mean values were measured from at least three fluorescence assays in 30 mM HEPES (pH 7.37), 250 mM NaCl, 100 µM $CoCl_2$, and 4% DMSO at 37 ℃.

Peptide Analytical HPLC Monitoring of APN Inhibition. To monitor the inhibition of APN digestion, we followed the degradation of the peptides, as shown in Figure 6, in the presence of Q7, using reversed-phase analytical HPLC. Peptide **1** has a low binding affinity to Q7, due to the internal AmF residue. The corresponding proteolytic product, peptide **2,** has an N-terminal AmF residue and therefore binds to Q7 more tightly due to the additional electrostatic interaction between the N-terminal ammonium group of the Phe with the carbonyl lined portal of Q7.



Peptide **1**:
Thr-Gly-Gly-**AmF**-Met

Peptide **2**:
AmF-Met

Figure 6. Chemical structures (left) and amino acid sequences (right) of the peptides used in this study; the N-termini are unprotected primary amines, and the C-termini are primary amides.

A baseline HPLC trace with Q7 was obtained for peptide **1** as a reference for the starting material (Figure 7). A baseline for peptide **2** was obtained in the presence of Q7 as a reference for the expected Q7•dipeptide product complex (Figure 8). In addition to the analytical HPLC traces, electrospray ionization time-of-flight mass spectra were used to confirm the starting material and product (Figures 9 and 10).

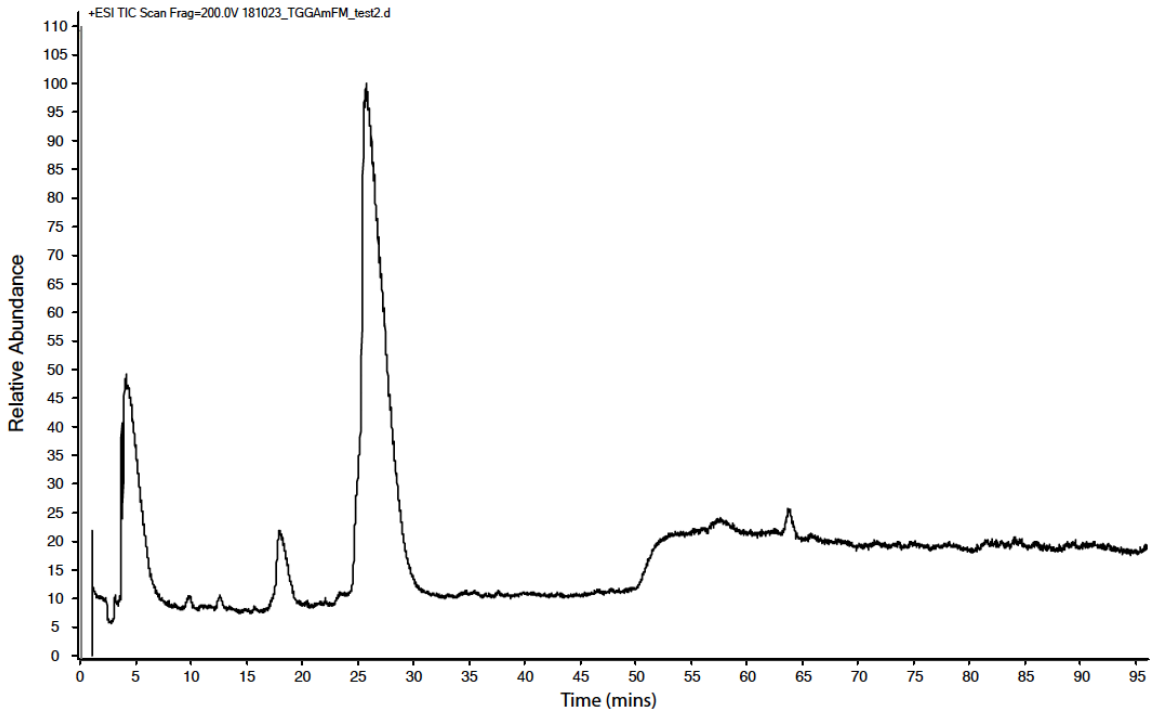


Figure 7. Analytical HPLC trace of 0.2 mM peptide **1** in the presence of 1 mM Q7 in 10 mM ammonium formate buffer (pH = 7.18). The method used a 0.33% acetonitrile per minute gradient in 0.1% aqueous TFA and monitored for absorbance at 220 nm.

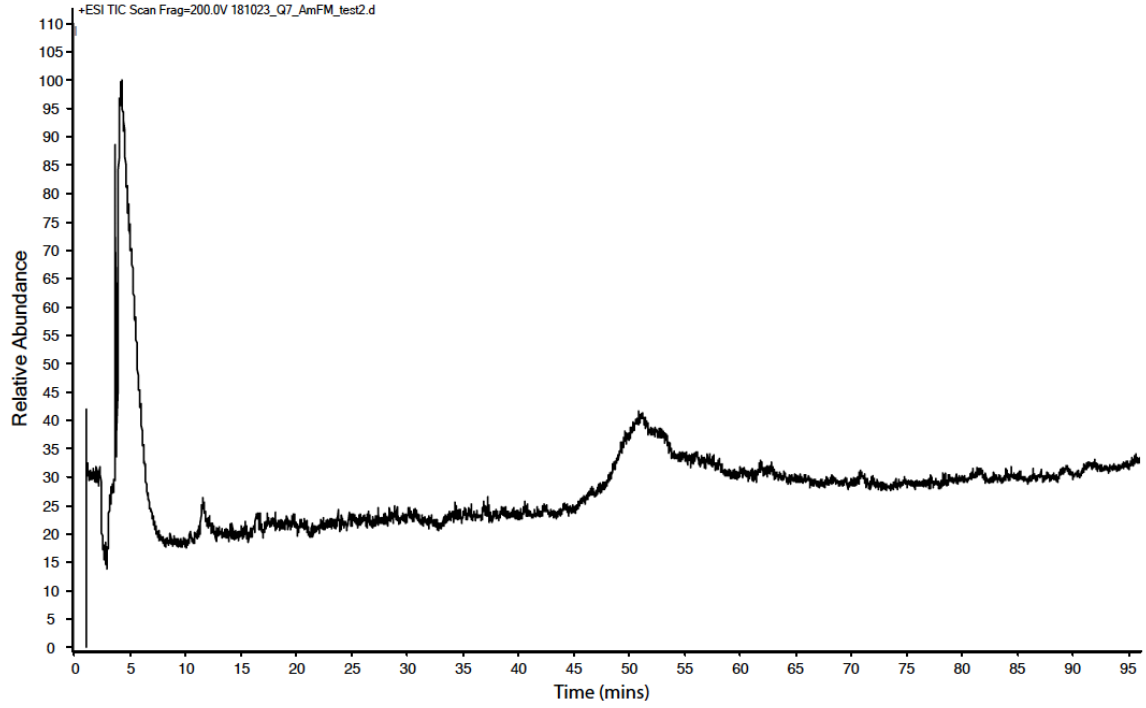


Figure 8. Analytical HPLC trace of 0.2 mM peptide **2** in the presence of 1 mM Q7 in 10 mM ammonium formate buffer (pH = 7.18). The method used 0.33% acetonitrile per minute gradient in 0.1% aqueous TFA and monitored for absorbance at 220 nm.

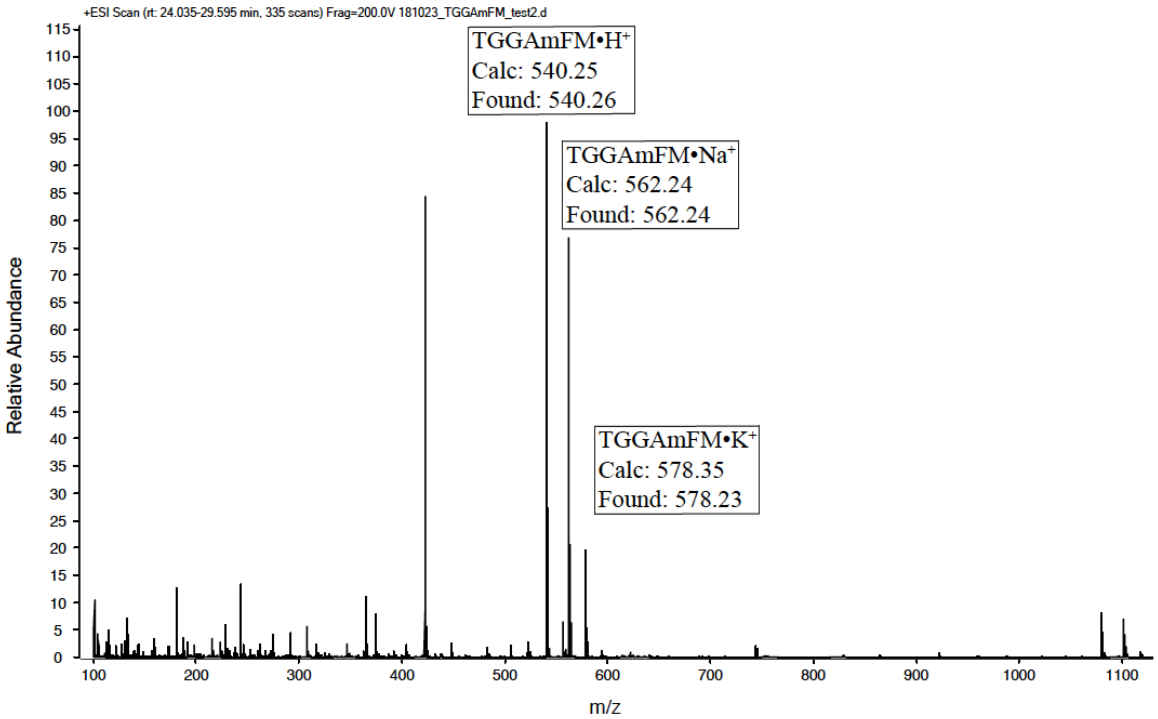


Figure 9. ESI-MS positive ion mode spectrum of Peptide **1**.

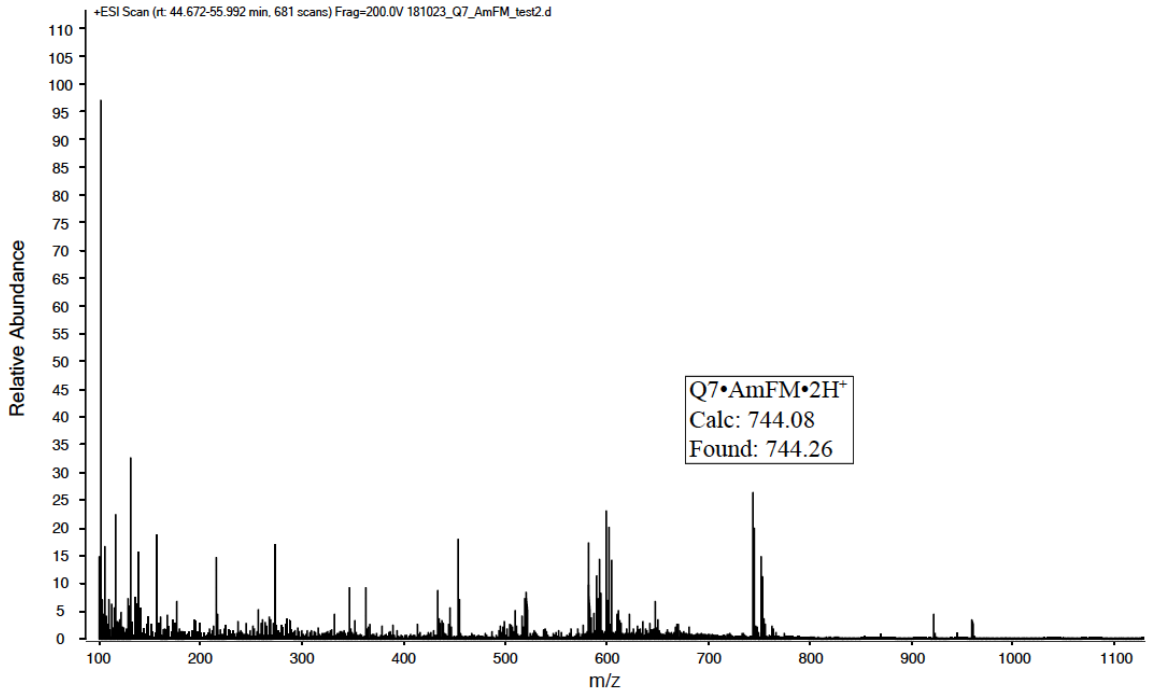


Figure 10. ESI-MS positive ion mode spectrum of Peptide **1**.

To follow the enzymatic digestion of peptide **1**, APN was added to peptide **1** in the presence of Q7, and the enzymatic reaction was followed over a period of 3 hours. The HPLC trace on the C18 column was unresolved (Figure 11) and so digestion was confirmed based on the MS spectra. The spectra for the beginning and end of the APN digestion were acquired by infusing the samples directly into the ESI-MS/TOF (Figures 12 and 13). The mass spectrum of the solution at the end of 3 hours exhibited a peak with m/z value of 744.26, which corresponds to the exact mass of the dipeptide product bound to Q7 (Figure 13).

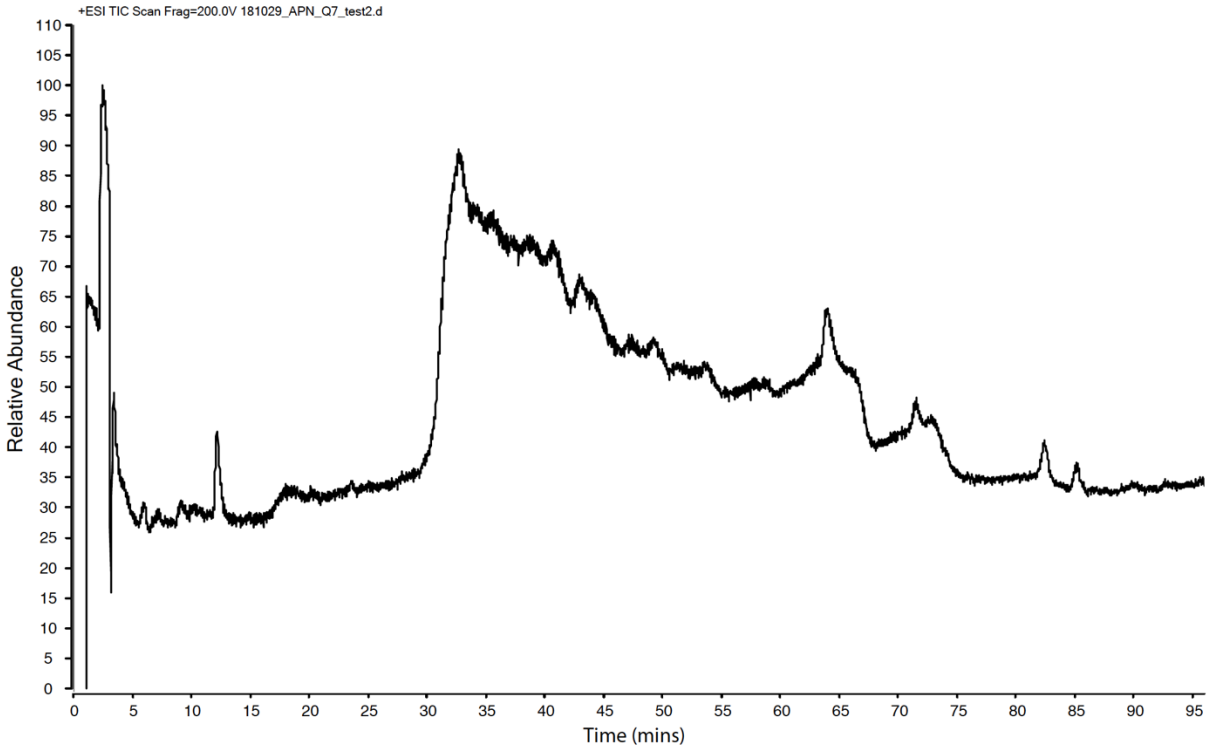


Figure 11. Analytical HPLC trace of 0.2 mM peptide **2** in the presence of 1 mM Q7 and 0.179 µM APN in 10 mM ammonium formate buffer (pH = 7.18). The trace was taken after the enzyme had been allowed to react for 3 hours. The method used 0.33% acetonitrile per minute gradient in 0.1% aqueous TFA and monitored for absorbance at 220 nm.

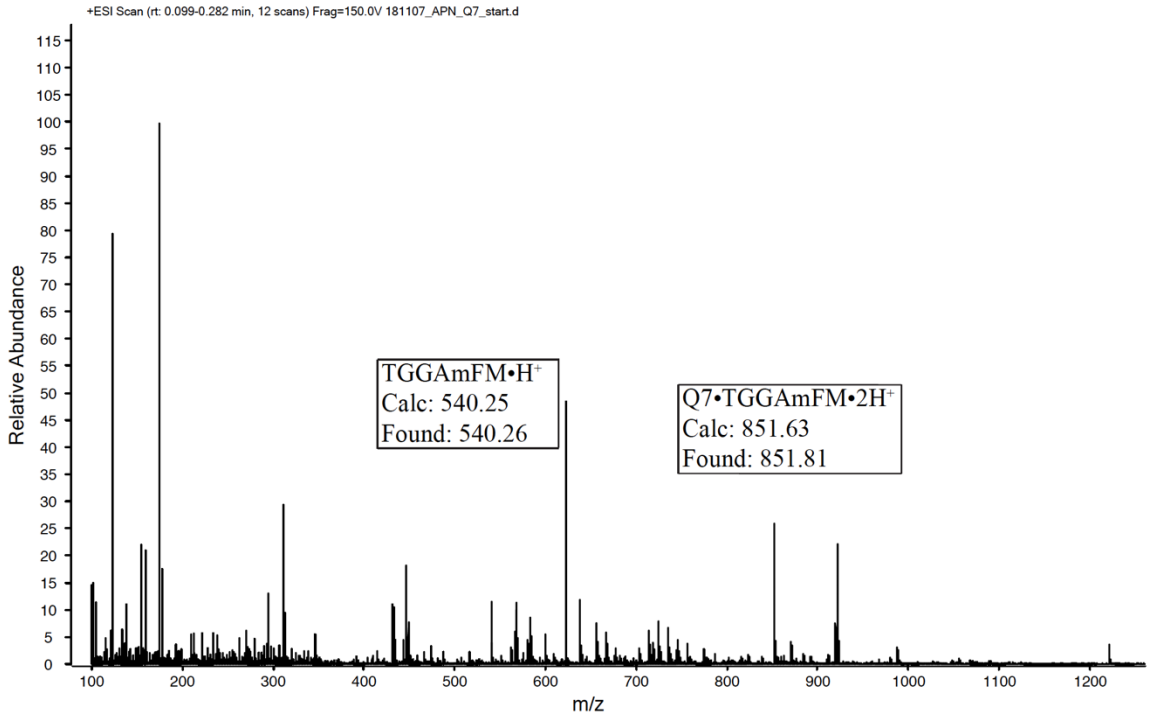


Figure 12. ESI-MS positive ion mode spectrum of Peptide **1** at the beginning of the digestion.

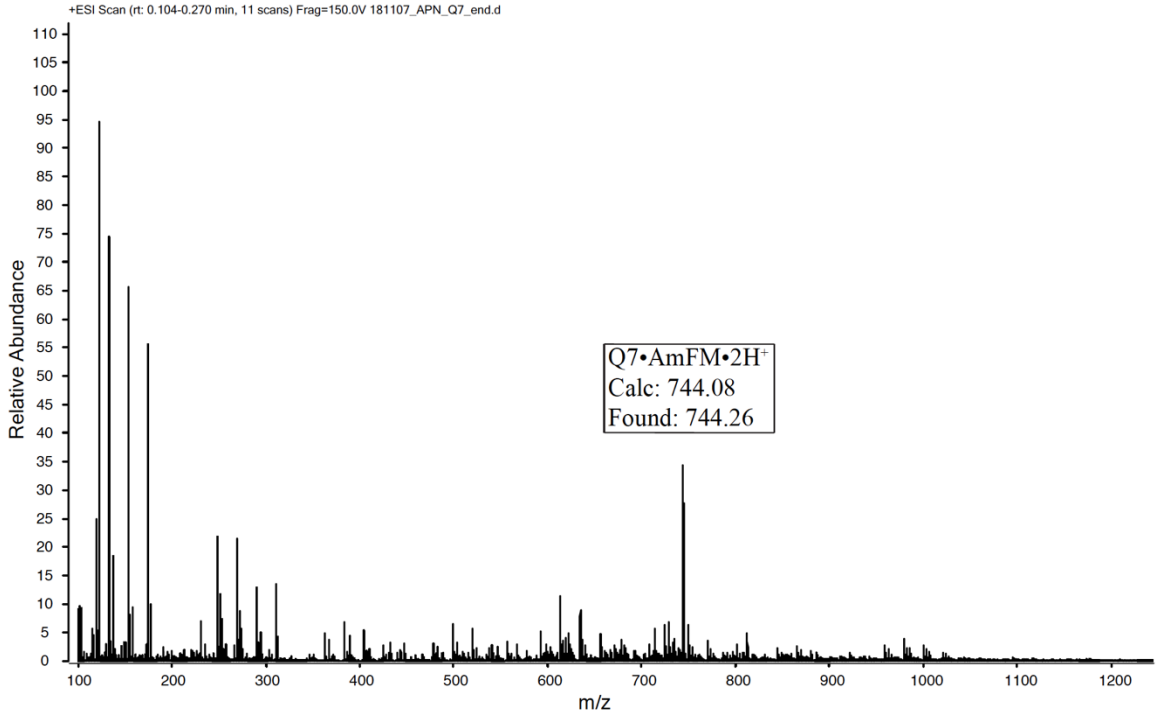


Figure 13. ESI-MS positive ion mode spectrum of Peptide **2** at the end of the digestion.

APN Inhibition using HEWL. A baseline HPLC trace of HEWL was obtained as a reference for the starting material (Figure 14). In addition, an electrospray ionization time-of-flight mass spectrometer was used to acquire reference data for the starting material (Figure 15).

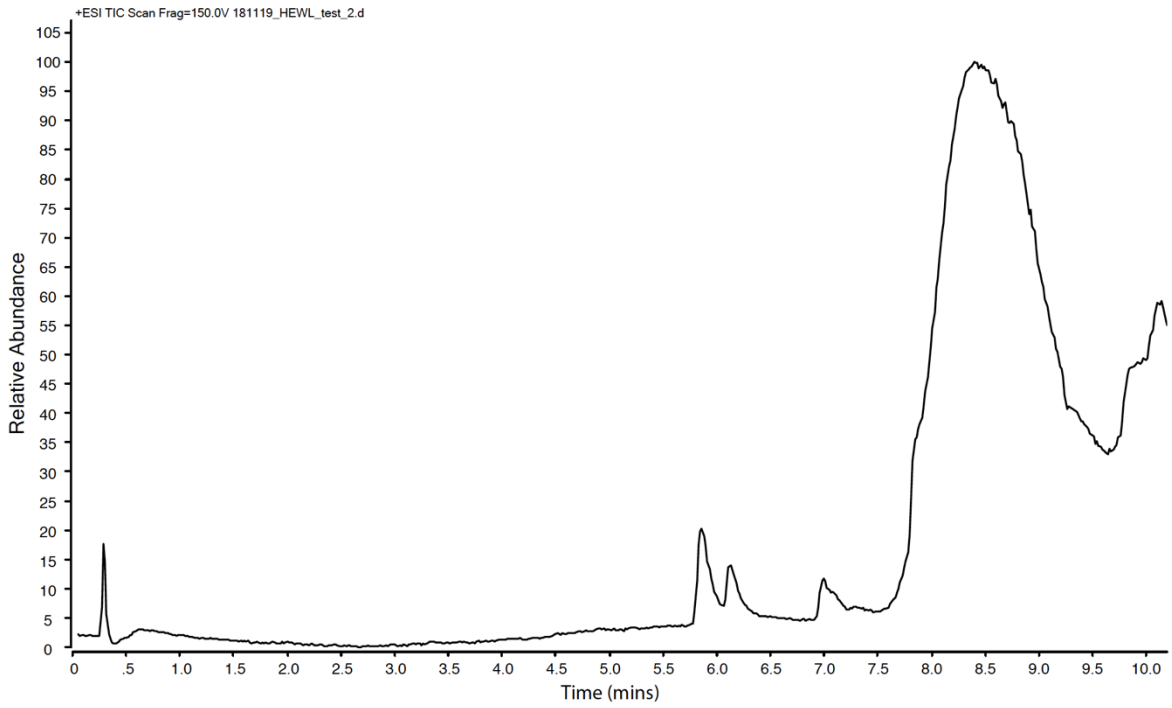


Figure 14. Analytical HPLC trace of 0.5 µM HEWL in 0.1% formic acid in water. The method used a 50% acetonitrile per minute gradient in 0.1% formic acid and monitored for absorbance at 220 nm.

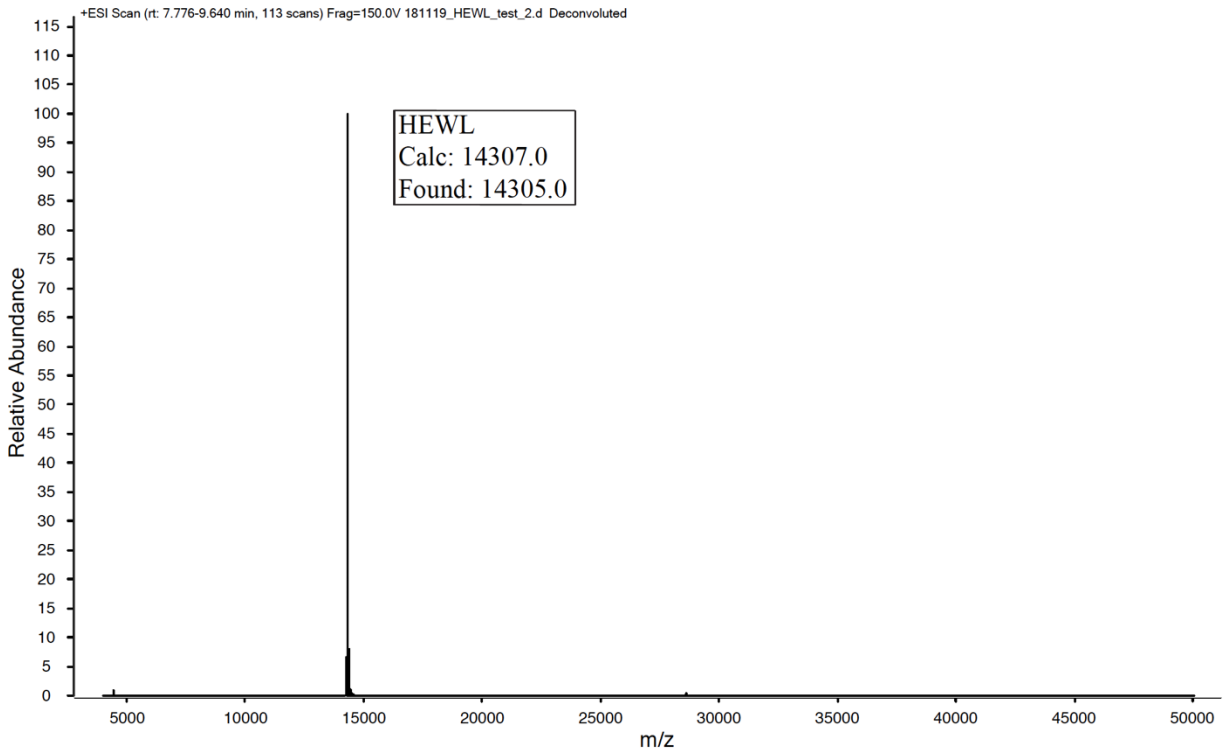


Figure 15. Deconvoluted ESI-MS positive ion mode spectrum of HEWL.

APN was added to HEWL in the presence of Q7, and the enzymatic reaction was followed over a period of 3 hours. An analytical HPLC trace was obtained after the three hours (Figure 16). Over this time, the main peak present in the MS corresponded to that of the starting material[5], i.e., no HEWL•Q7 complex was formed (Figure 17).

According to the crystal structure of HEWL (PDB code 1DPX), the N-terminus is not solvent accessible and is packed against the outside of the protein. Based on prior work[6,7], we hypothesize that the protein should be disordered at the N-terminus in order for APN to begin digestion at the N-terminus and for complex formation with Q7.

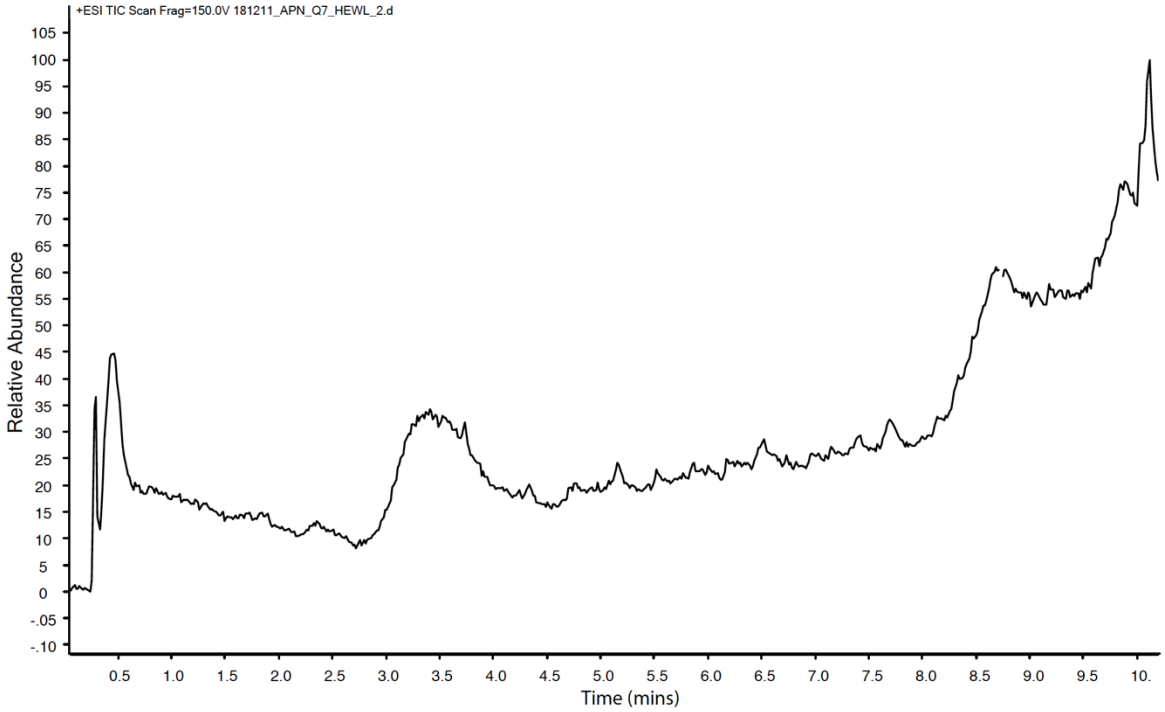


Figure 16. Analytical HPLC trace of 100 µM HEWL in the presence of 1 mM Q7 and 0.179 µM APN in 10 mM ammonium formate buffer (pH = 7.18). The trace was taken after the enzyme had been allowed to react for 3 hours. The method used a 50% acetonitrile per minute gradient in 0.1% formic acid and monitored for absorbance at 220 nm.

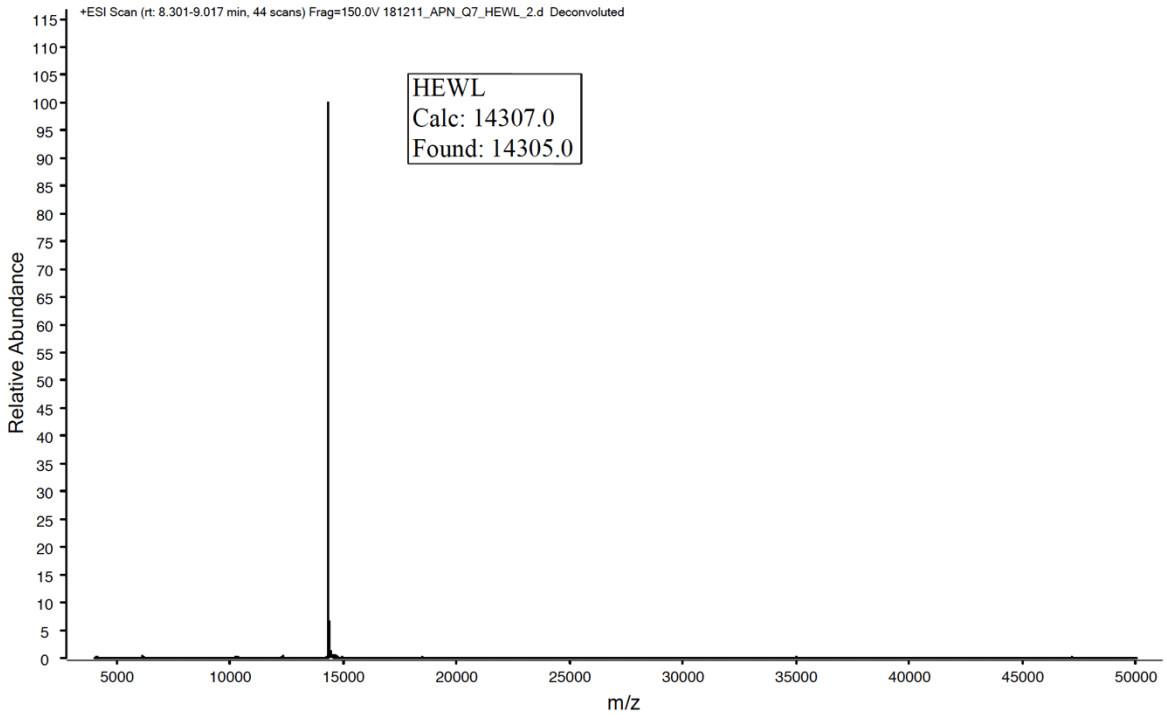
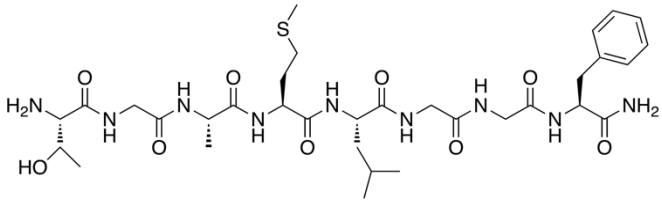
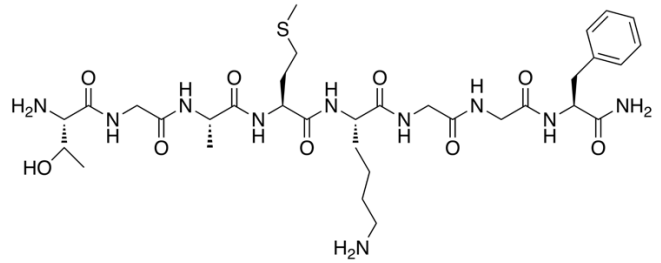


Figure 17. Deconvoluted ESI-MS positive ion mode spectrum of HEWL at the end of the digestion.

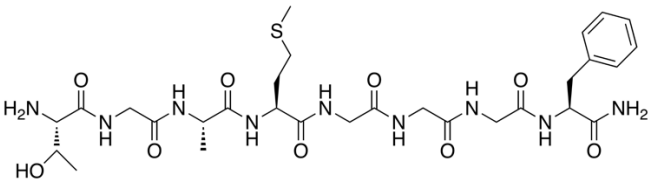
APN Inhibition using Q8. Now that we confirmed the conditions for APN digestion using the pentapeptides and Q7, we hypothesized that Q8 might also inhibit APN through the dipeptide sequences that were studied in **Chapter 2** thus allowing a broader ability for this technique. A series of three octapeptides were designed with the target dipeptide sequences inserted within the middle of the chain (Figure 18). Peptides **3** and **4** contain the internal dipeptide sequence ML and MK, which should bind to Q8 with submicromolar affinity once they are N-terminally exposed upon digestion. Control peptide **5** contains the dipeptide sequence MG, which is expected to bind to Q8 with a weak affinity. Further studies include taking pure HPLC traces of the three octapeptides to obtain the retention time as well as optimizing conditions to observe complex formation by analytical HPLC and LC-MS.



Peptide **3**:
Thr-Gly-Ala-**Met**-**Leu**-Gly-Gly-Tyr



Peptide **4**:
Thr-Gly-Ala-**Met**-**Lys**-Gly-Gly-Tyr



Peptide **5**:
Thr-Gly-Ala-**Met**-**Gly**-Gly-Gly-Tyr

Figure 18. Chemical structures (left) and amino acid sequences (right) of the peptides used in this study; the N-termini are unprotected primary amines, the C-termini are -NH_2 for a primary amide. Target sites are bolded.

Conclusion

In this project, we have aimed to generate a N-terminal aromatic binding site for Q7 on HEWL and proposed a scheme to generate N-terminal dipeptide sequences for Q8 on octapeptides. Reproducing the APN protocol corroborated that Q7 can inhibit APN. Q7 inhibits APN by binding to modified or unmodified N-terminal Phe peptides under the conditions used for the HPLC and MS experiments. In reproducing the published results[4], we have established conditions for monitoring peptide digestion using the HPLC in tandem with mass spectrometry. Currently, we are attempting to find conditions to achieve protein digestion by looking for ways in which we can disorder the N-terminus of HEWL without fully deactivating the protease. We are also looking ahead to finding conditions for complex formation using Q8 for APN mediated digestion on the octapeptides.

Experimental

Materials. The peptides Thr-Gly-Ala-Met-Leu-Gly-Gly-Tyr, Thr-Gly-Ala-Met-Lys-Gly-Gly-Tyr, and Thr-Gly-Ala-Met-Gly-Gly-Gly-Tyr, were purchased from GenScript Biotech (purity > 95%) with amidation at C-termini and HPLC purification.

Reagents. Water was obtained from a Barnstead Nanopure Infinity water system (18 M Ω cm). The following compounds were of analytical purity grade, unless otherwise noted, and were used without purification: Fmoc-Phe(4-CH_2NHBoc)-OH (Chem Impex International); Rink amide MBHA resin, HBTU, Fmoc-Met-OH, Fmoc-Gly-OH (Peptides International), Fmoc-Thr(tBu)-OH, biotech-grade dimethyl formamide (DMF), diisopropylethylamine (DIEA), triisopropyl silane (TIS), trifluoroacetic acid (TFA), piperidine, anhydrous dichloromethane, HPLC-grade acetonitrile, H-Met-AMC (Enzo Life Sciences), and aminopeptidase N (from porcine kidney, 50 U/mg) (EMD Millipore). All amino acids were the L enantiomer. Cucurbit[7]uril was synthesized according to the literature and quantified by isothermal titration calorimetry. [8-10]

Synthesis. Peptides **1-5** were synthesized by standard Fmoc solid-phase synthesis protocols on a Rink amide MBHA resin. Each synthesis was carried out on the scale of 0.2703 g resin (0.37 meq/g). Fmoc deprotection was accomplished by shaking out the resin in a solution of 20 % piperidine in DMF at room temperature for 20 minutes, followed by extensive rinsing with DMF. For each peptide coupling, a mixture of Fmoc-protected amino acid, HBTU, and DIEA (4 eq. each) was dissolved in 10 mL of DMF, mixed for 15 minutes, and added to the resin. Couplings were carried out by shaking at room temperature

for 60 minutes followed by extensive rinsing with DMF. After the final Fmoc deprotection, the resin was transferred to a glass-fritted flask and rinsed with dichloromethane, and then the peptide was cleaved from the resin by treatment with two portions of a 7 mL mixture of 95% TFA, 2.5% TIS, and 2.5% water for 10 minutes for the first part and for one hour for the second part at room temperature with shaking. The mixture was filtered, and the crude peptide was collected and concentrated by rotarty evaporation at 60 °C for 20 minutes. The crude peptide was then dissolved in 30 mL of 0.1% TFA and purified in three batches by reverse phase HPLC (Waters Delta 600 Semi-Prep system with a Phenomenex Gemini-NX C-18 column) using a 1% per minute gradient of acetonitrile in 0.1% aqueous TFA and monitored for absorbance at 220 nm (Agilent 1100 system with a Phenomenex Gemini-NX C-18 column). Pure fractions were combined and lyophilized to dryness.

Lyophilized peptides were dissolved in a Barnstead Nanopure Infinity water system (18 M cm) and aliquots were determined assuming the extinction coefficient of the 4-(CH_2NH_2)-Phe residue (ε_{260} = 208.5 M^{-1} cm^{-1}). Absorption measurements were performed using a UV-visible spectrophotometer.

Electrospray Ionization Mass Spectrometry (ESI-MS). Mass spectra were acquired using an Agilent 6230 TOF LC/MS mass spectrometer with an electrospray ion source in the positive ion mode.

APN and AMC Assay. Assays for APN and L-Met-7-amido-4-methylcoumarin (AMC) were performed as published[4] with a range of AMC concentrations of 0 μM, 15.6 μM, 31.3 μM, 62.5 μM, 125 μM, 250 μM, 500 μM, and 1000 μM in 30 mM HEPES (pH 7.37), 250

mM NaCl, 100 μM $CoCl_2$, and 4% DMSO at 37 ℃. An addition 1 μL of APN was used to initiate the reaction. The final concentration of APN was 7.14 nM in a total assay volume of 301 μL. The enzymatic activity was monitored for a total of 20 minutes. Fluorescence emission intensities were measured every 30 seconds. Fluorescence emission intensities were measured using a Tecan Infinite 200 Pro fluorescence plate reader with an excitation wavelength of 380 nm (bandwidth 9 nm) and emission wavelength of 460 nm (bandwidth 20 nm) with a constant amount of gain, signal averaging, integration time, and focal depth.

Monitoring Enzymatic Activity by Analytical HPLC. Baseline analytical HPLC and MS traces were obtained for both peptides at 0.2 mM. The HPLC and MS traces for peptide **1** was taken without Q7 whereas the traces for peptide **2** was taken with Q7. The method used a 0.33% acetonitrile per minute gradient in 0.1% aqueous TFA and monitored for absorbance at 220 nm on a Phenomenex Gemini-NS C-18 column. Stock concentrations for each peptide were determined by UV-vis spectroscopy in a 10 mM ammonium formate buffer (pH 7.37) with the extinction coefficient of AmF being ε_{260} = 208.5 M^{-1} cm^{-1}. Absorption measurements were performed on a UV-visible spectrophotometer. Q7 solutions were prepared from a Q7 stock that was standardized by ITC. All Q7 solutions were sonicated and heated at 60 °C for five minutes, then cooled, and syringe filtered.

APN (2.5 units, 50 μL from EMD Millipore) was activated for 2 hours in a non-shaking Fisher Scientific Isotemp Incubater at 37 °C. An addition of 18 μL APN was added to a solution of 0.2 mM peptide **1** with 1 mM Q7 in a 10 mM ammonium formate (pH 7.37) buffer. The solutions were placed in the incubator at 37 °C and allowed to react for 3 hours. An aliquot was taken from the solution after 3 hours and analyzed by analytical HPLC and

MS (Agilent 6230 TOF LC/MS with a coupler) at a 0.33% acetonitrile per minute gradient in 0.1% aqueous formic acid and monitored for absorbance at 220 nm.

The same was done with HEWL. An addition of 18 μL APN, after it had been activated, was added to a solution of 0.1 mM HEWL in the presence of 1 mM Q7 in a 10 mM ammonium formate (pH 7.37) buffer. The solutions were placed in the incubator at 37 °C and allowed to react for 3 hours. An aliquot was taken from the solution after 3 hours and analyzed by analytical HPLC and MS (Agilent 6230 TOF LC/MS with an Agilent C-18 column) at a 50% acetonitrile per minute gradient in 0.1% formic acid and monitored for absorbance at 220 nm.

References

(1) Ghale, G.; Ramalingam, V.; Urbach, A. R.; Nau, W. M. *J. Am. Chem. Soc.* **2011**, *133*, 7528-7535.

(2) Urbach, A. R.; Ramalingam V. *Israel J. Chem*. **2011**, *51*, 664-678.

(3) Logsdon, L. A.; Urbach, A. R *J. Am. Chem. Soc.*, **2013**, *135*, 11414-11416.

(4) Logsdon, L. A.; Schardon, C. L.; Ramalingam V.; Kwee, S. K.; Urbach, A. R. *J. Am. Chem. Soc*. **2011**, *133*, 17087-17092.

(5) Canfield, R. E. *J. Biol. Chem*. **1963**, *238*, 2698-2707.

(6) Chinai, J. M.; Taylor, A. B.; Ryno, L. M.; Hargreaves, N. D.; Morris, C. A.; Hart, P. J.; Urbach, A. R. *J. Amer. Chem. Soc. Comm.* **2011,** *133*, 8810-8813.

(7) Li, W.; Bockus, A. T.; Vinviguerra, B.; Isaacs, L.; Urbach, A. R. *Chem. Commun*. **2016**, *52*, 8537-8540.

(8) Day, A.; Arnold, A. P.; Blanch, R. J.; Snushall, B. *J. Org. Chem*. **2001**, *66*, 8094-8100.

(9) Kim, J.; Jung, I. S.; Kim, S. Y.; Lee, E.; Kang, J. K.; Sakamoto, S.; Yamaguchi, K.; Kim, K. *J. Am. Chem. Soc*. **2000**, *22*, 540-541.

(10) Marquez, C.; Huang F.; Nau, W. M. *IEEE Trans. Nanosci*. **2004**, *3*, 39-45.

Supporting Information

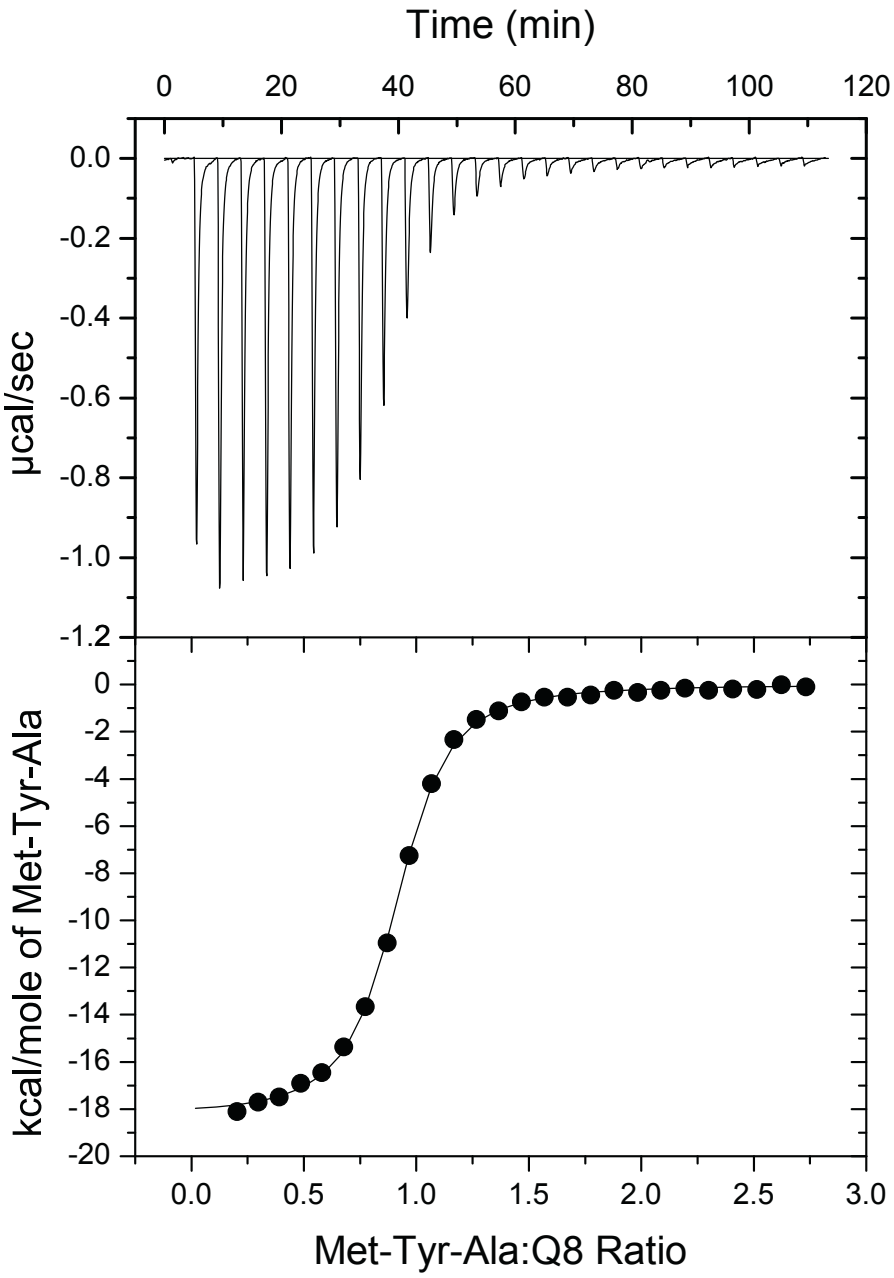


Figure S1. Representative isothermal titration calorigram of Met-Tyr-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

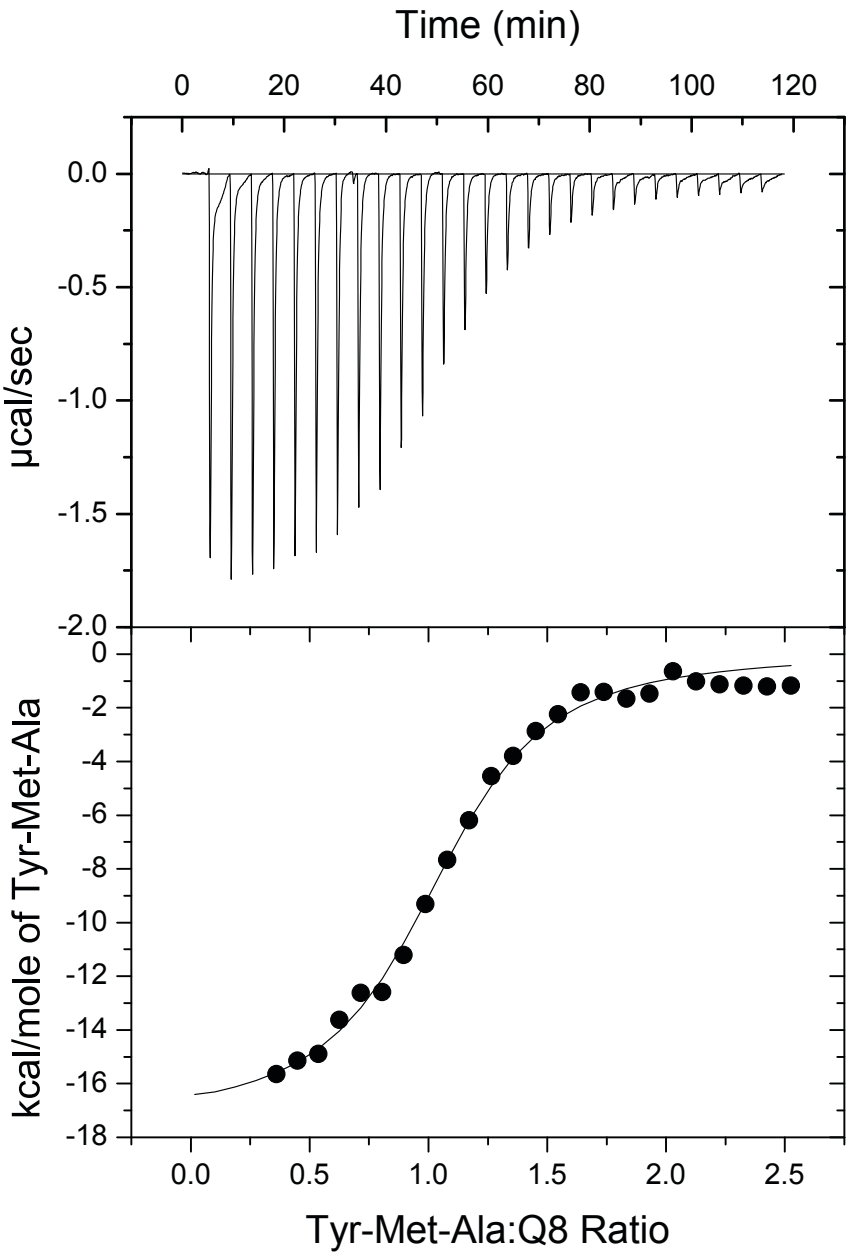


Figure S2. Representative isothermal titration calorigram of Tyr-Met-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

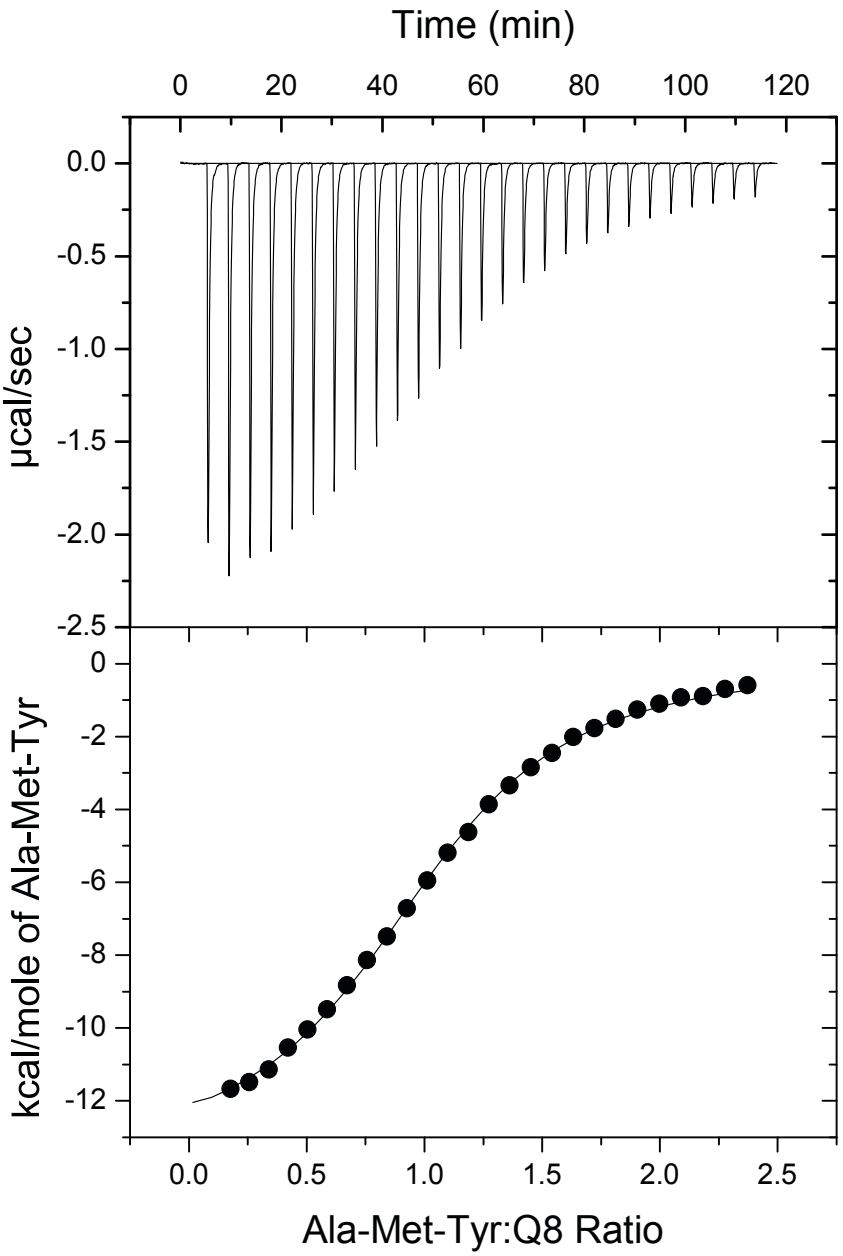


Figure S3. Representative isothermal titration calorigram of Ala-Met-Tyr titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

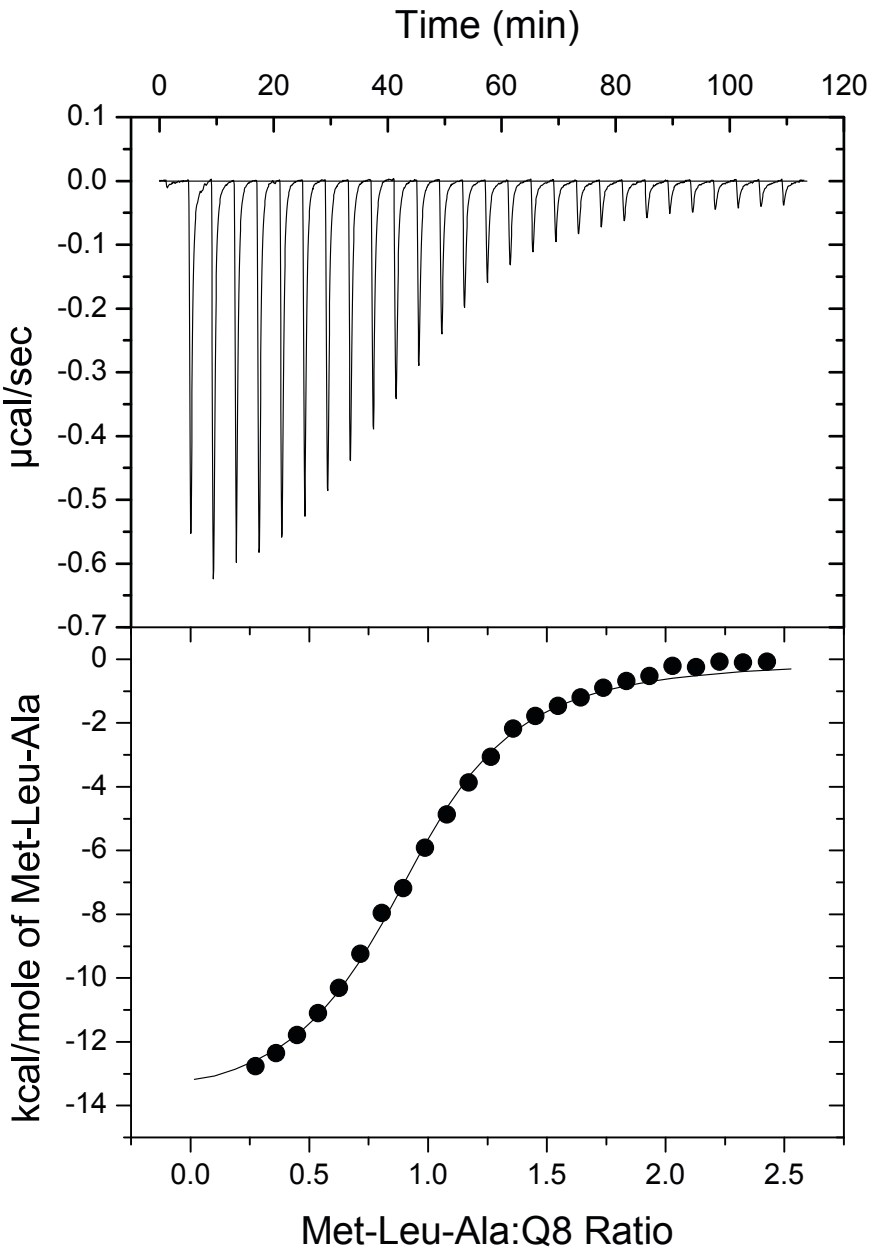


Figure S4. Representative isothermal titration calorigram of Met-Leu-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

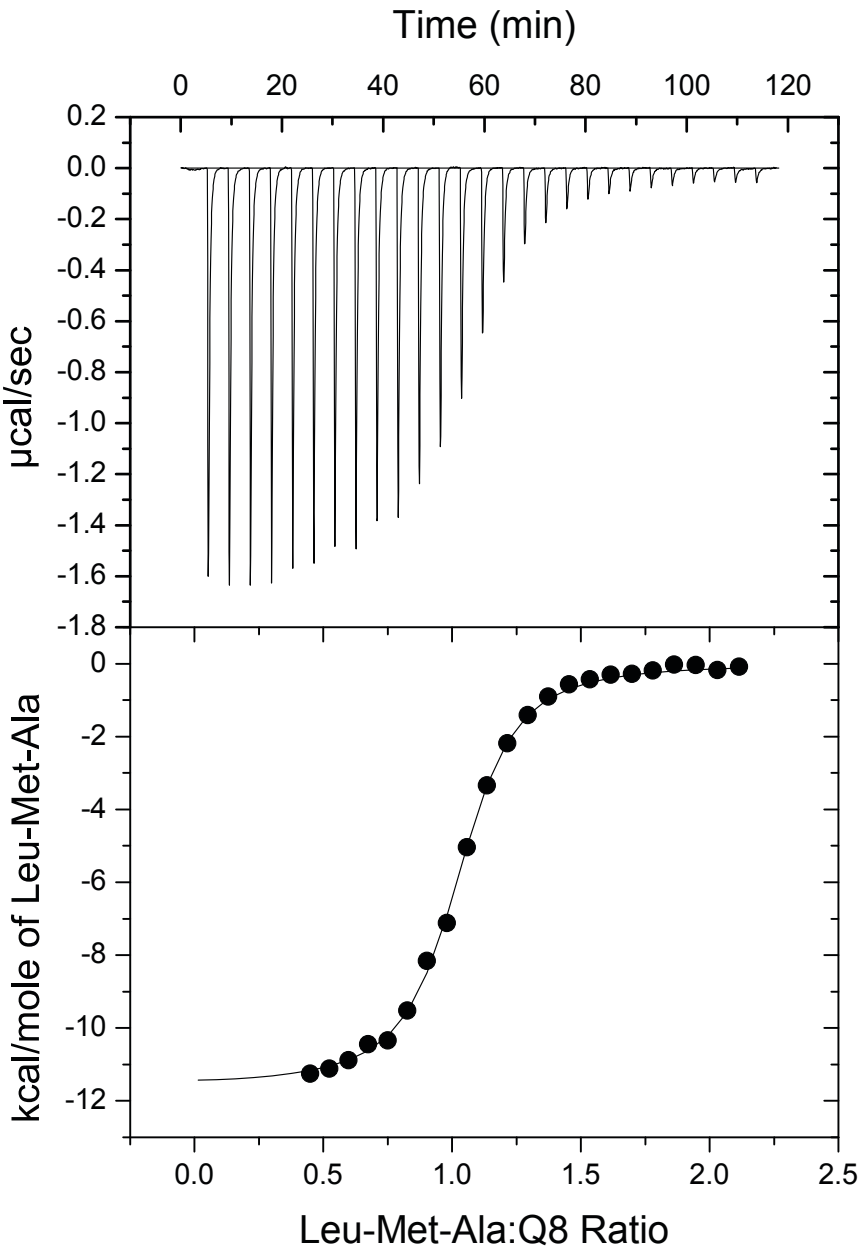


Figure S5. Representative isothermal titration calorigram of Leu-Met-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

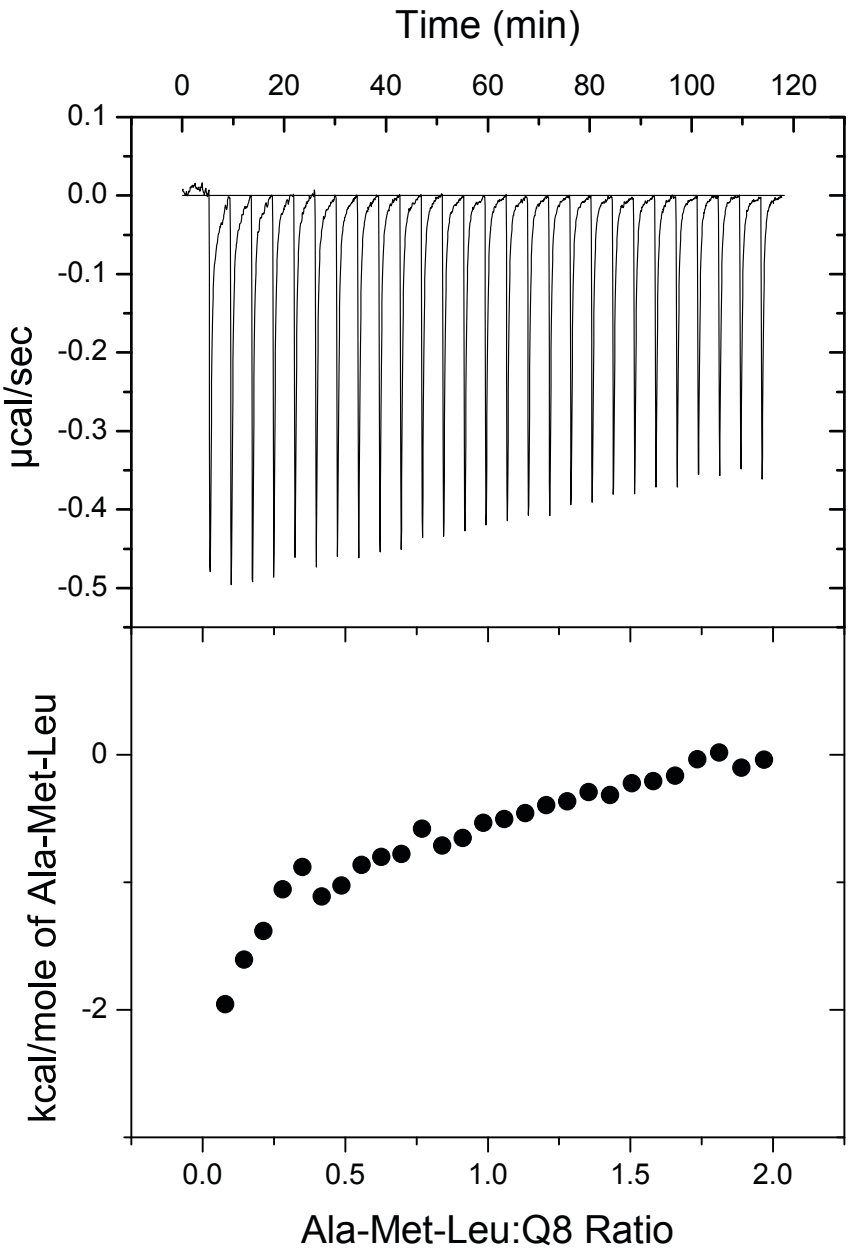


Figure S6. Representative isothermal titration calorigram of Ala-Met-Leu titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

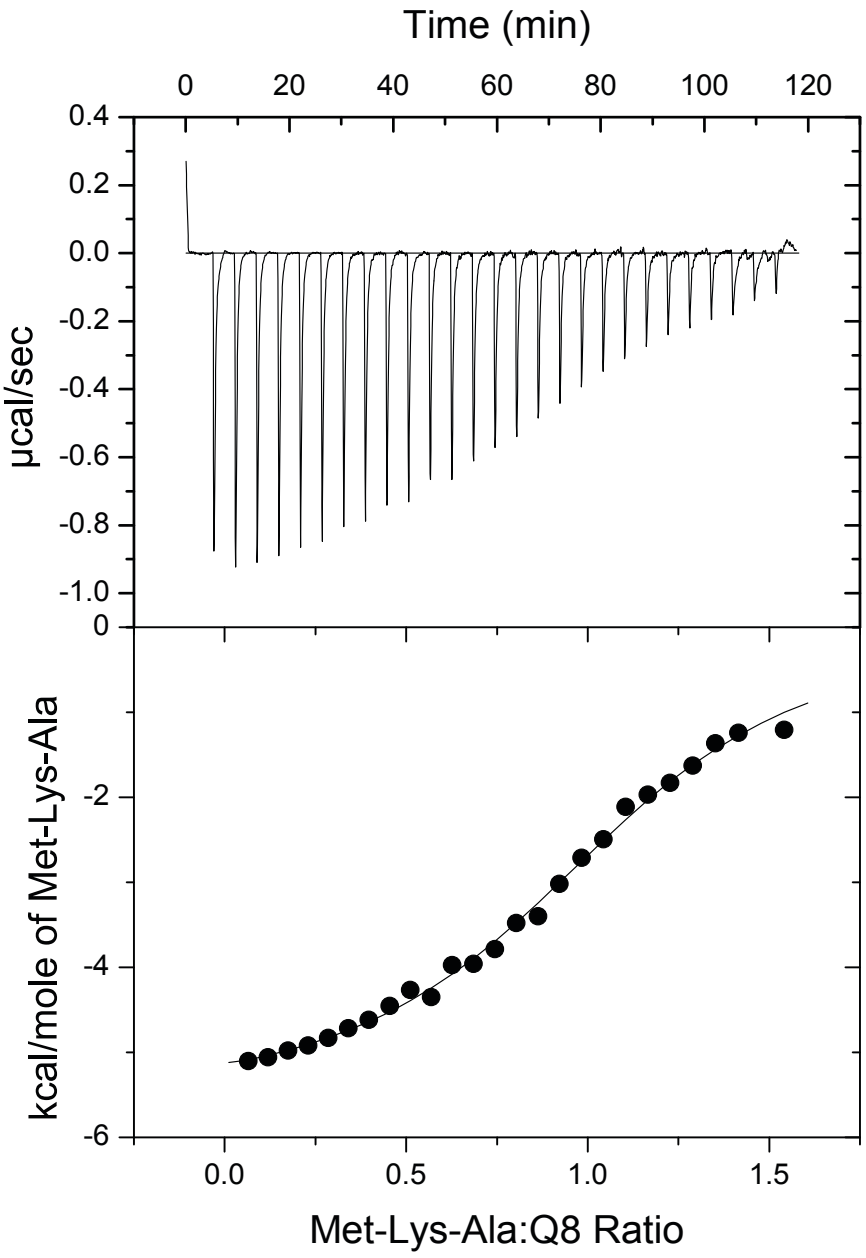


Figure S7. Representative isothermal titration calorigram of Met-Lys-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

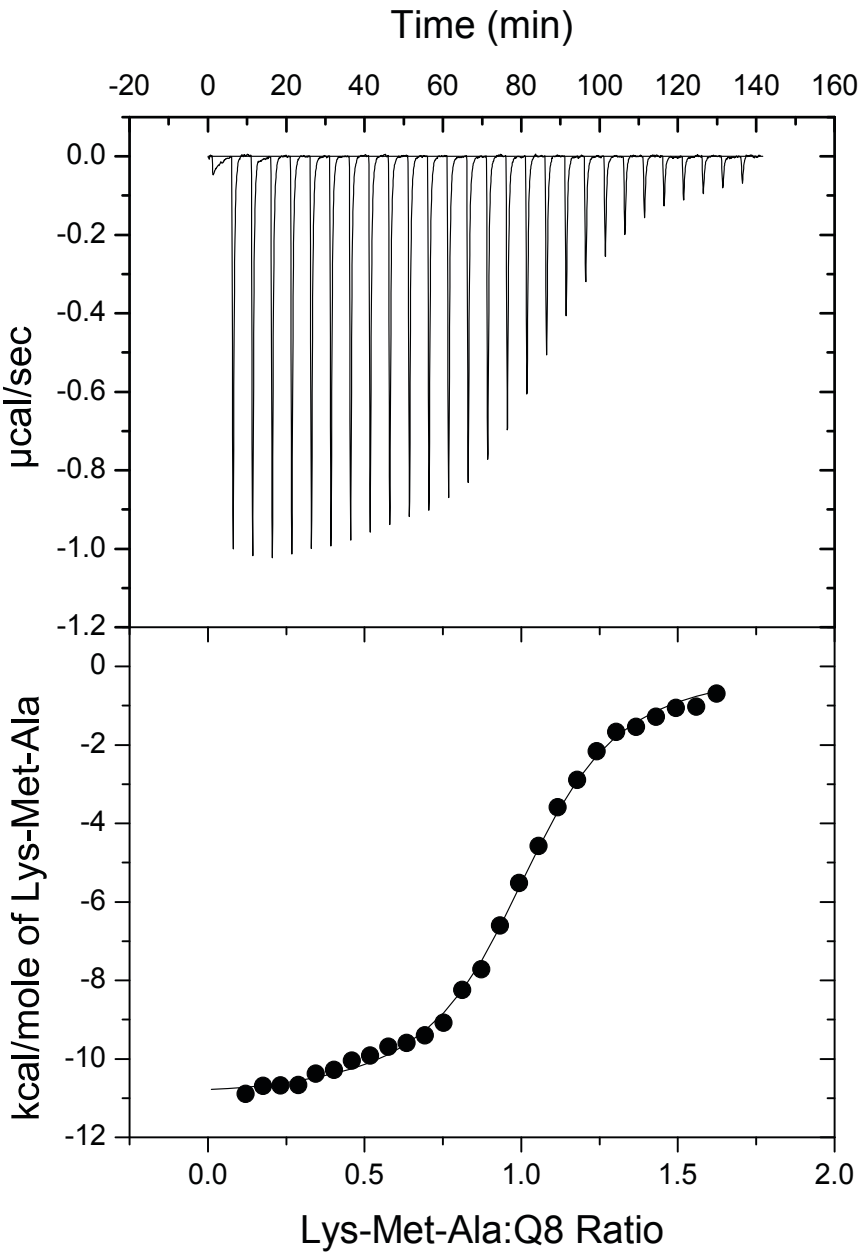


Figure S8. Representative isothermal titration calorigram of Lys-Met-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

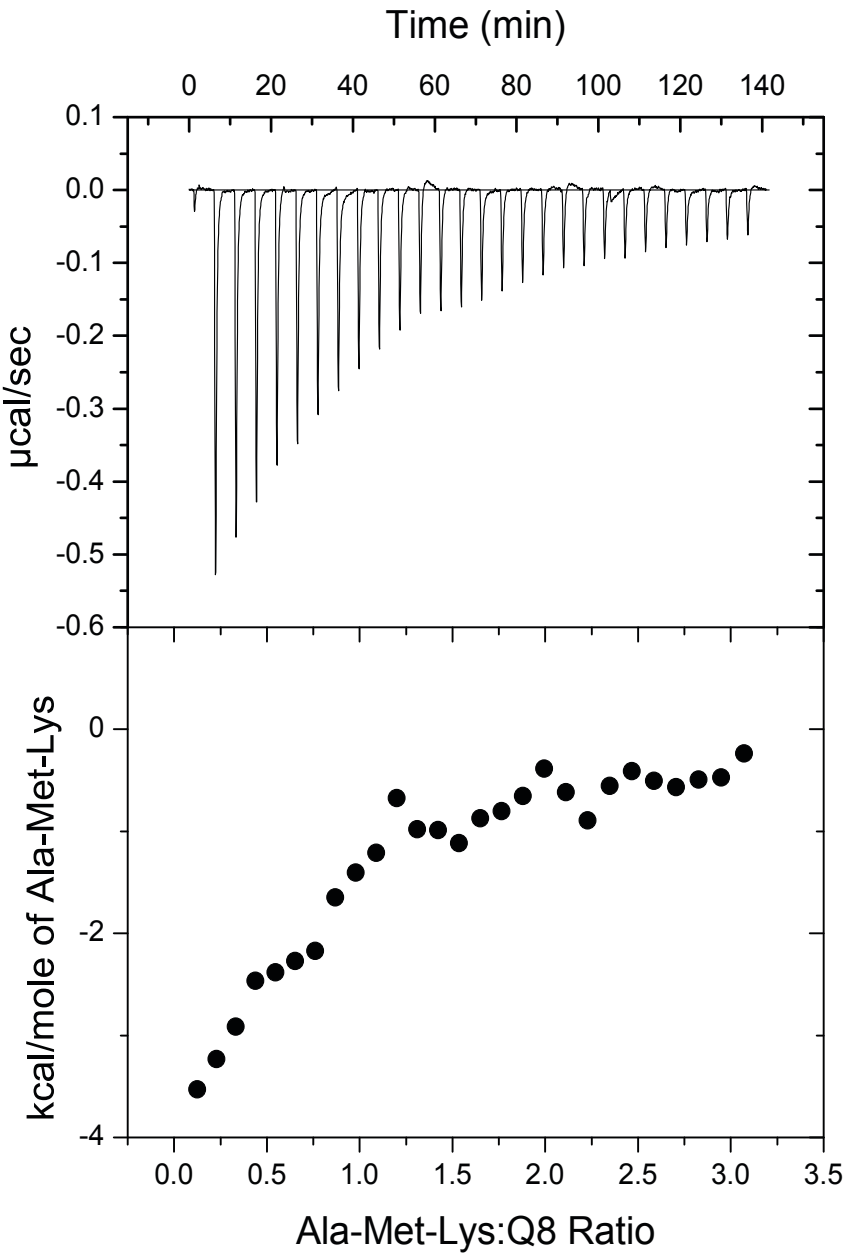


Figure S9. Representative isothermal titration calorigram of Ala-Met-Lys titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

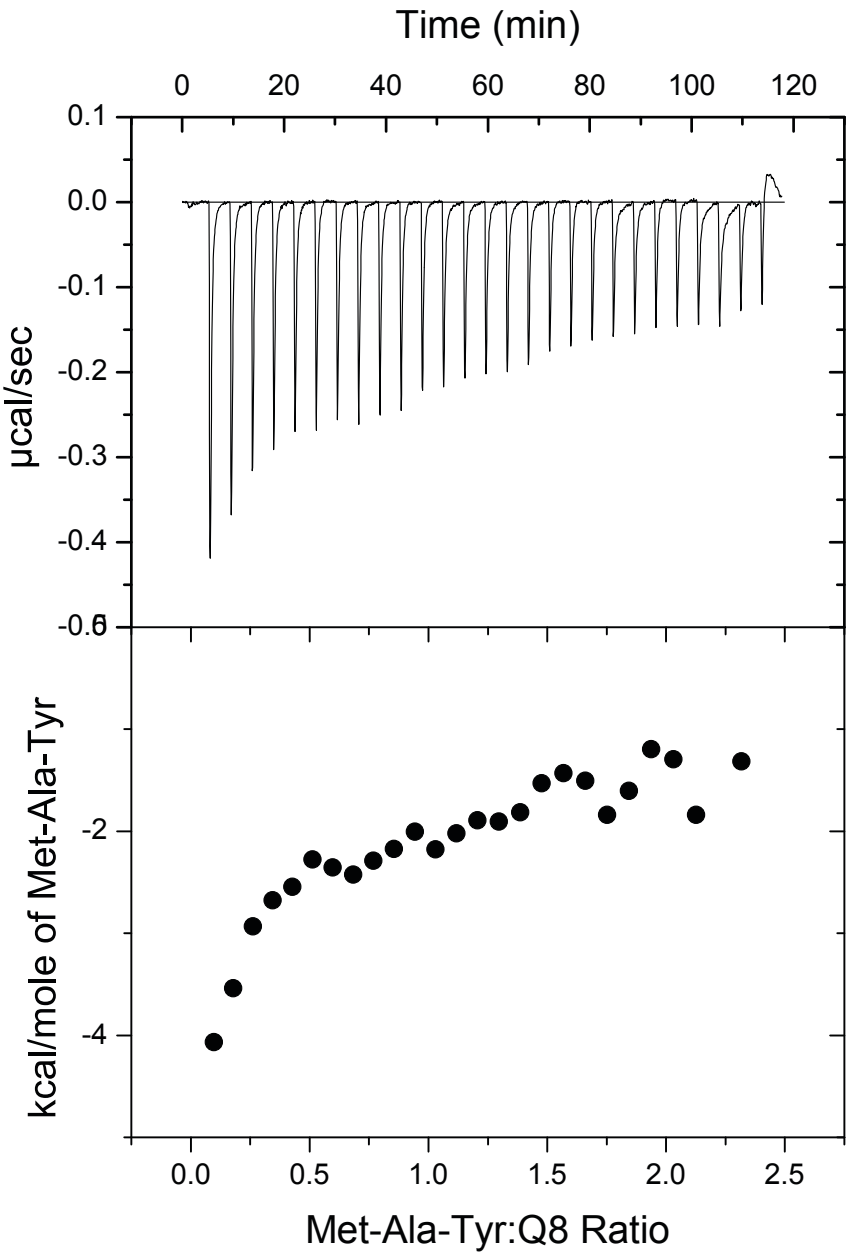


Figure S10. Representative isothermal titration calorigram of Met-Ala-Tyr titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

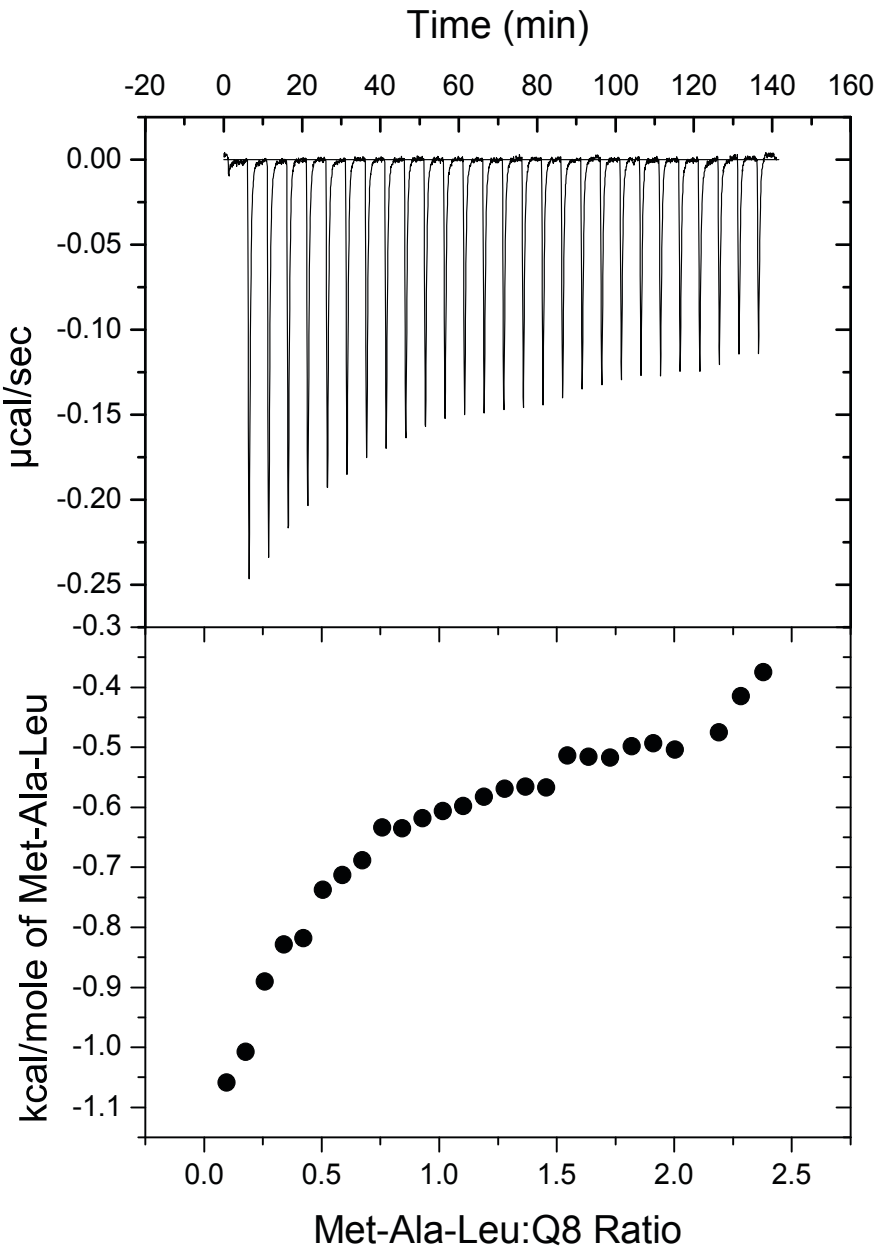


Figure S11. Representative isothermal titration calorigram of Met-Ala-Leu titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

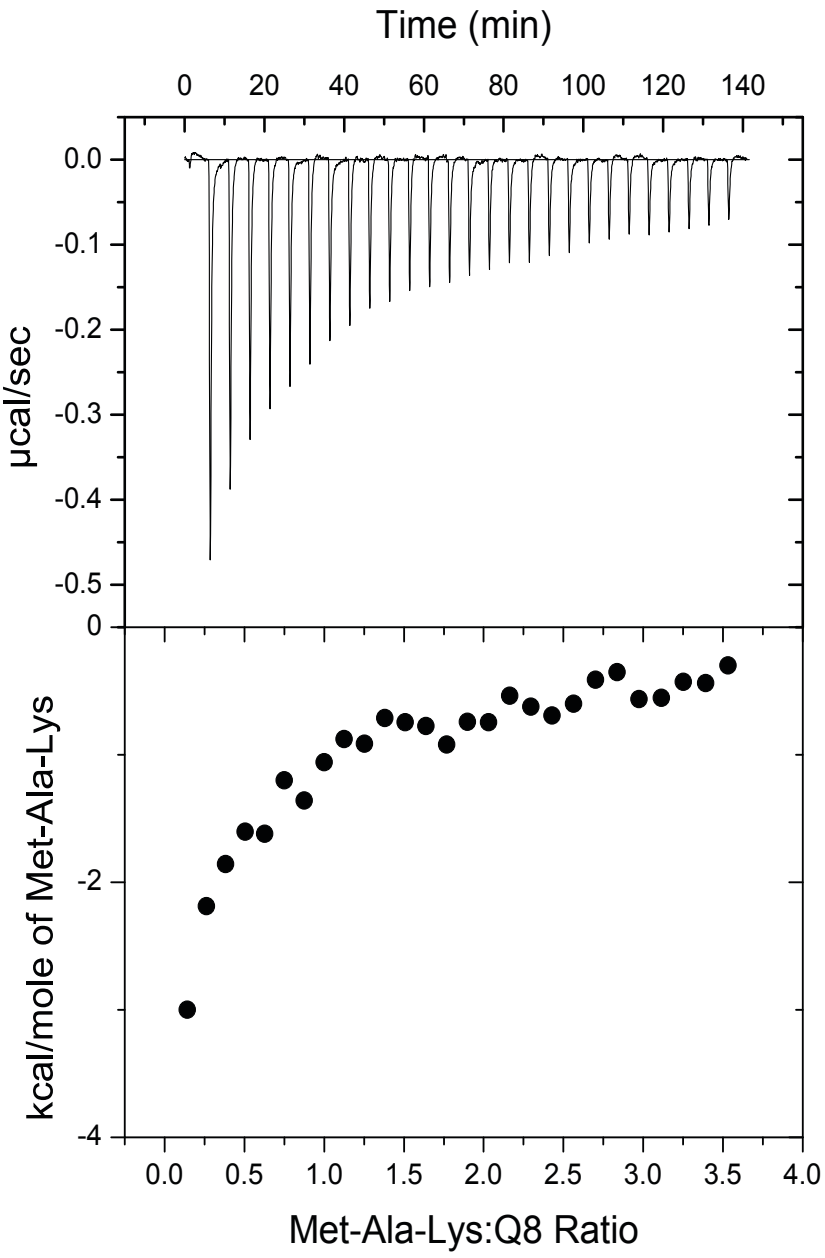


Figure S12. Representative isothermal titration calorigram of Met-Ala-Lys titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

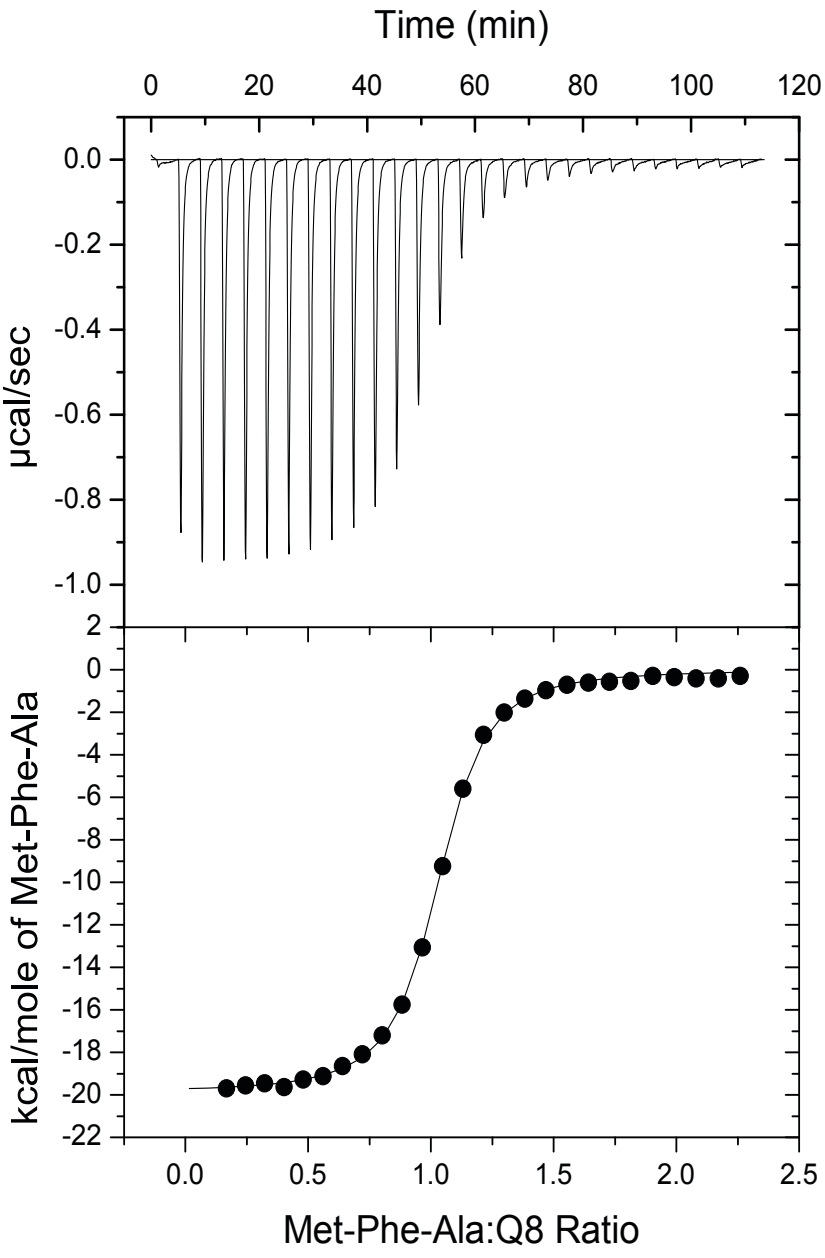


Figure S13. Representative isothermal titration calorigram of Met-Phe-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

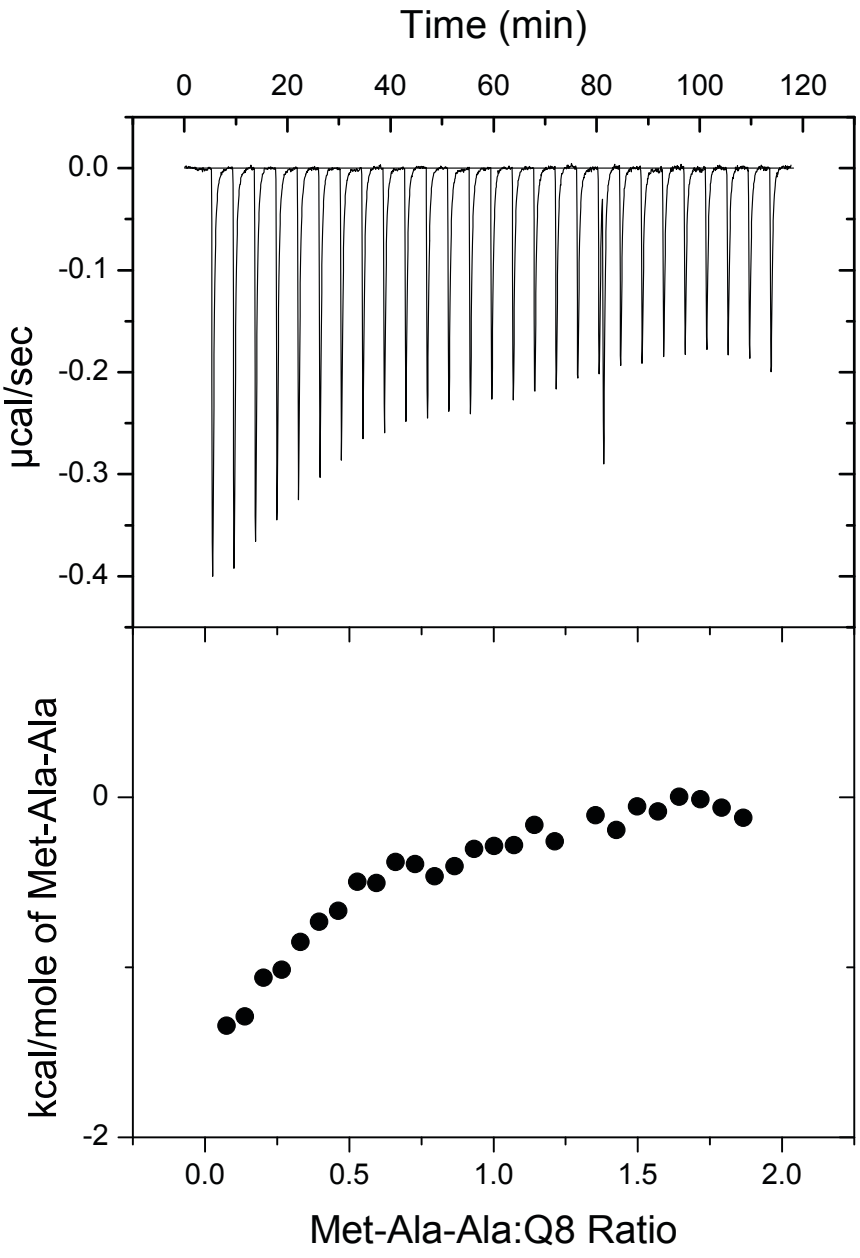


Figure S14. Representative isothermal titration calorigram of Met-Ala-Ala titrated into Q8 in 10 mM sodium phosphate, pH 7.0 at 27 ℃.

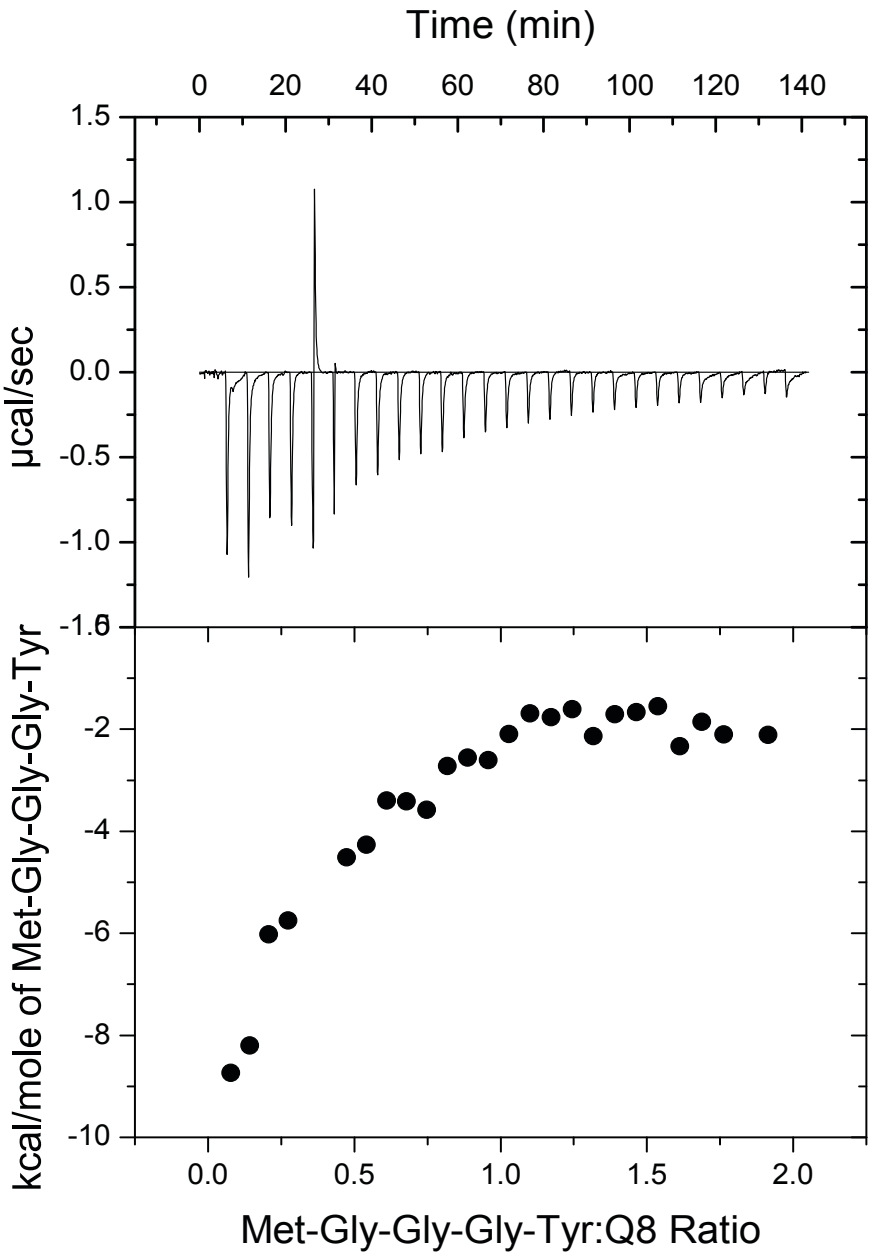


Figure S15. Representative isothermal titration calorigram of Met-Gly-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

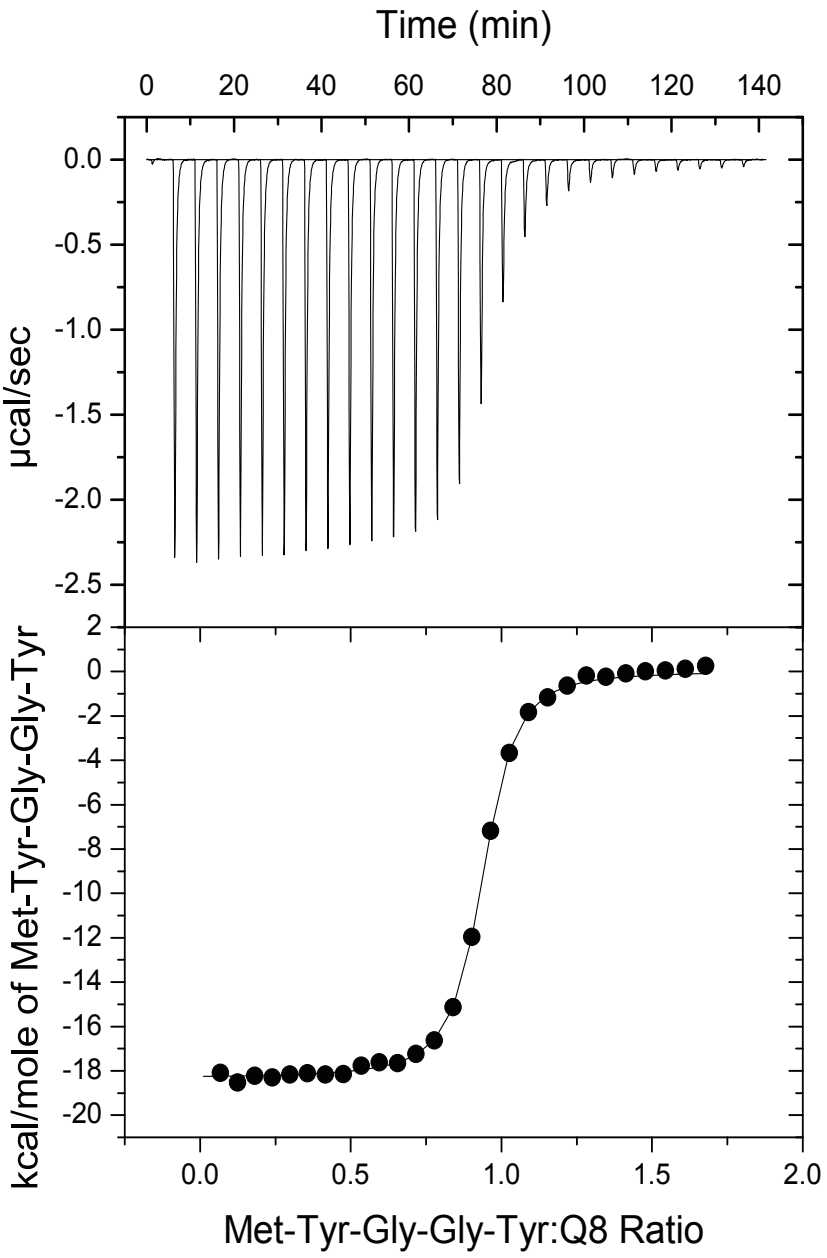


Figure S16. Representative isothermal titration calorigram of Met-Tyr-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

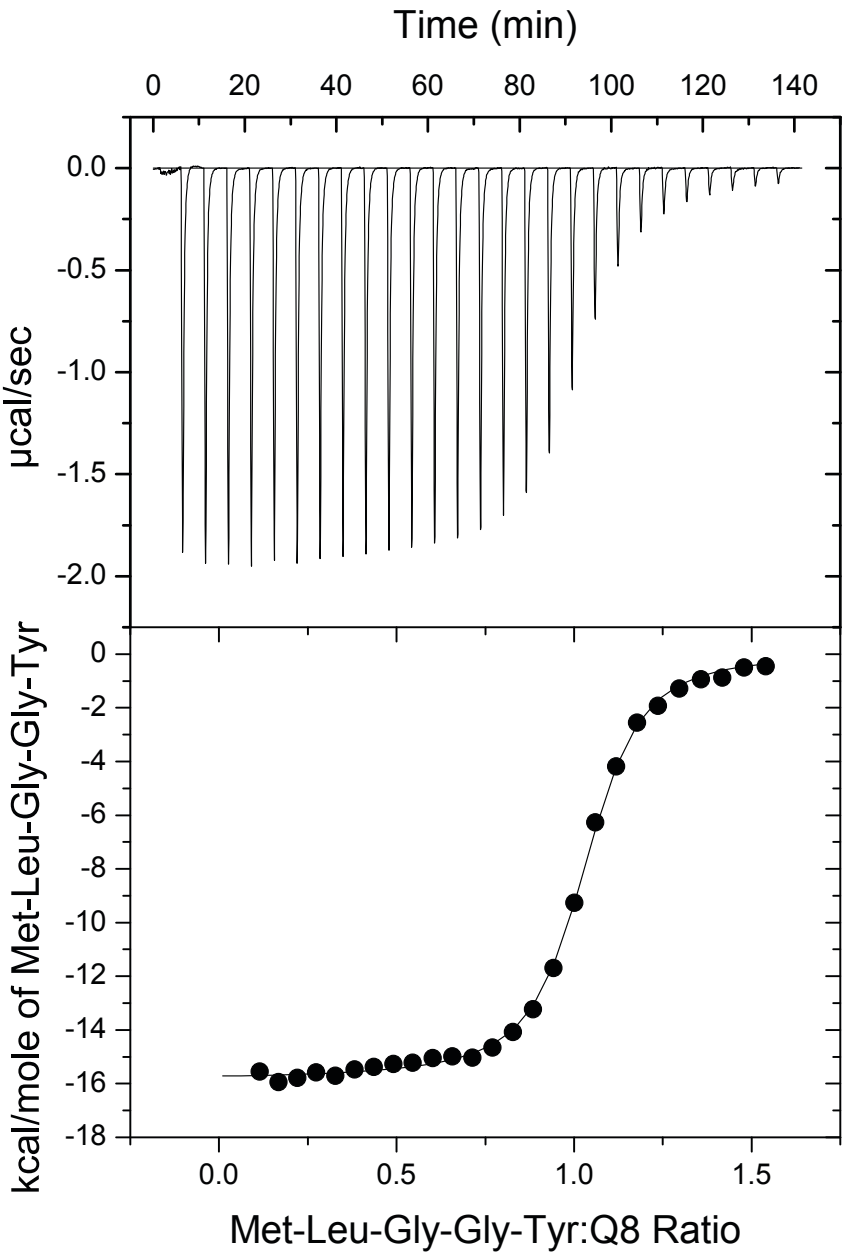


Figure S17. Representative isothermal titration calorigram of Met-Leu-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

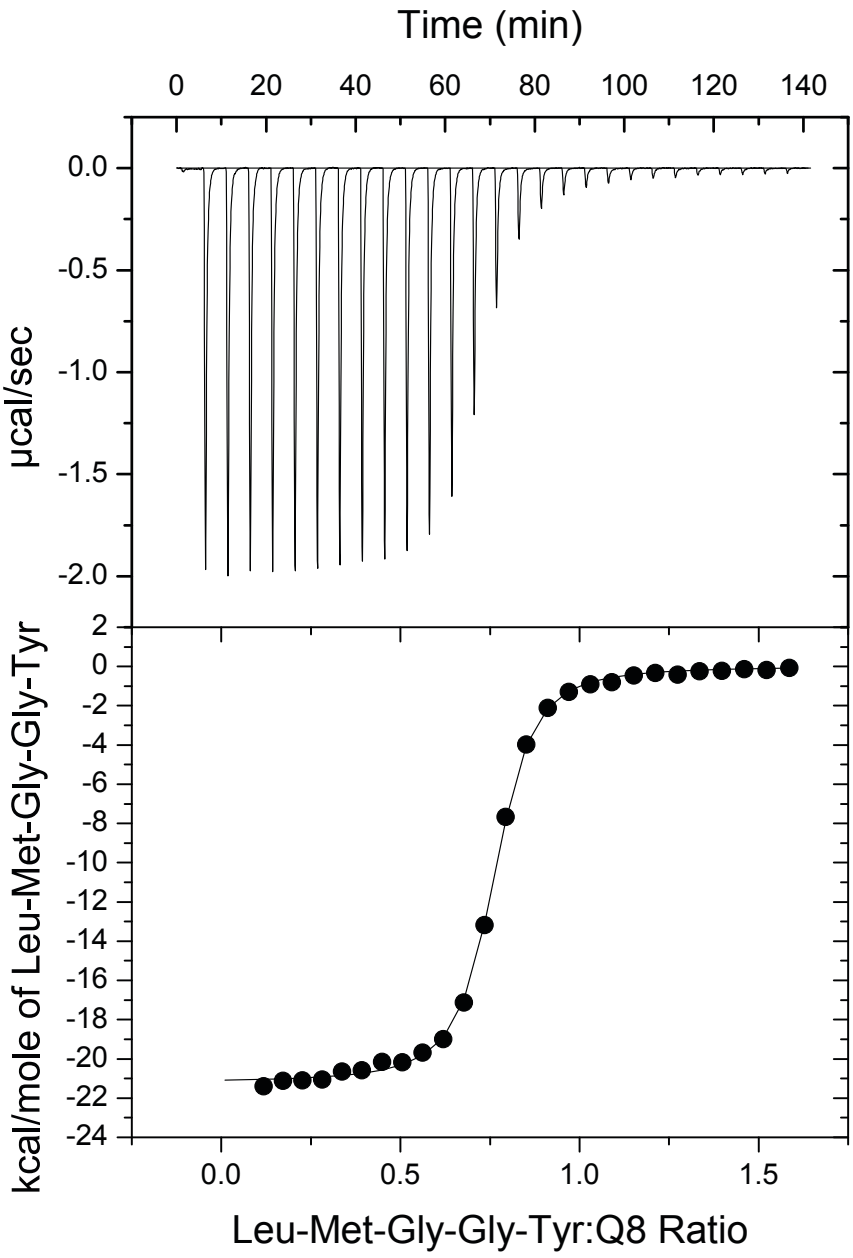


Figure S18. Representative isothermal titration calorigram of Leu-Met-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

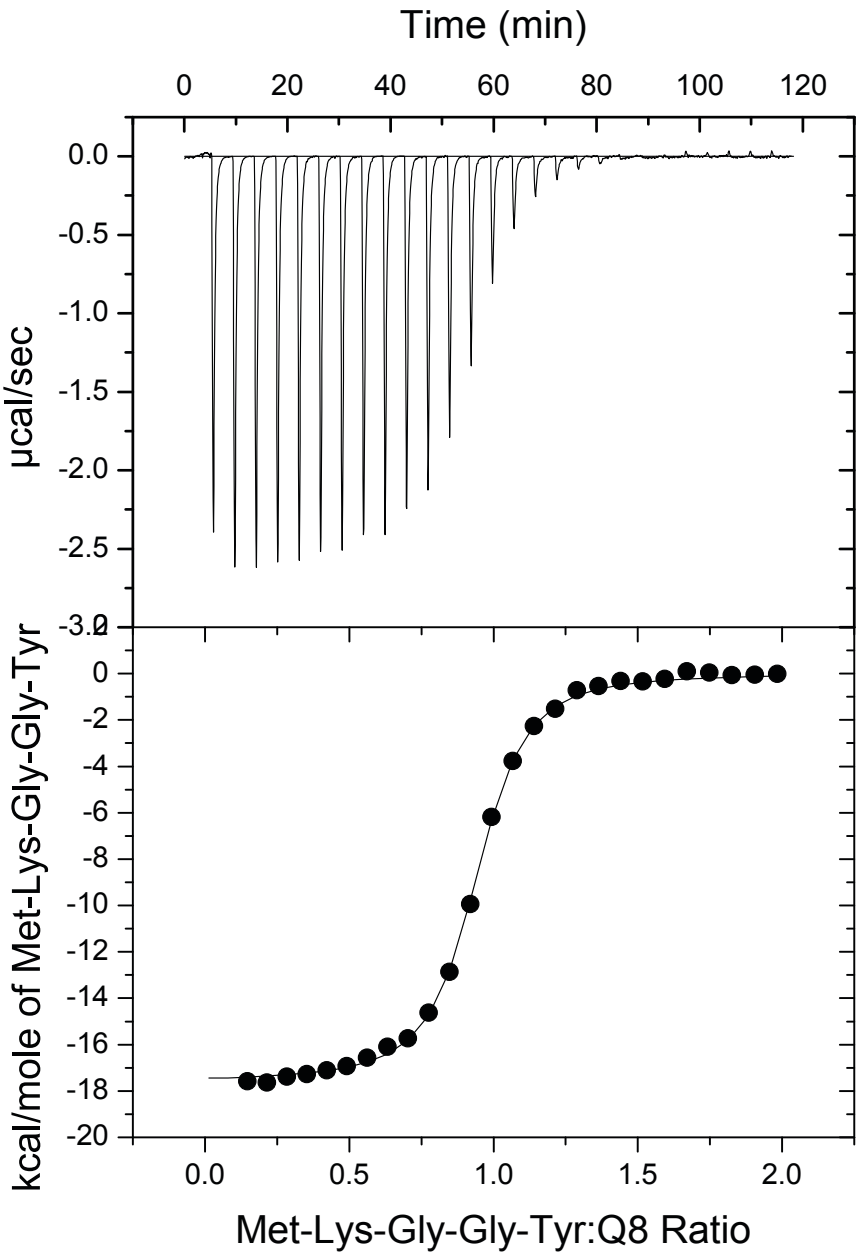


Figure S19. Representative isothermal titration calorigram of Met-Lys-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

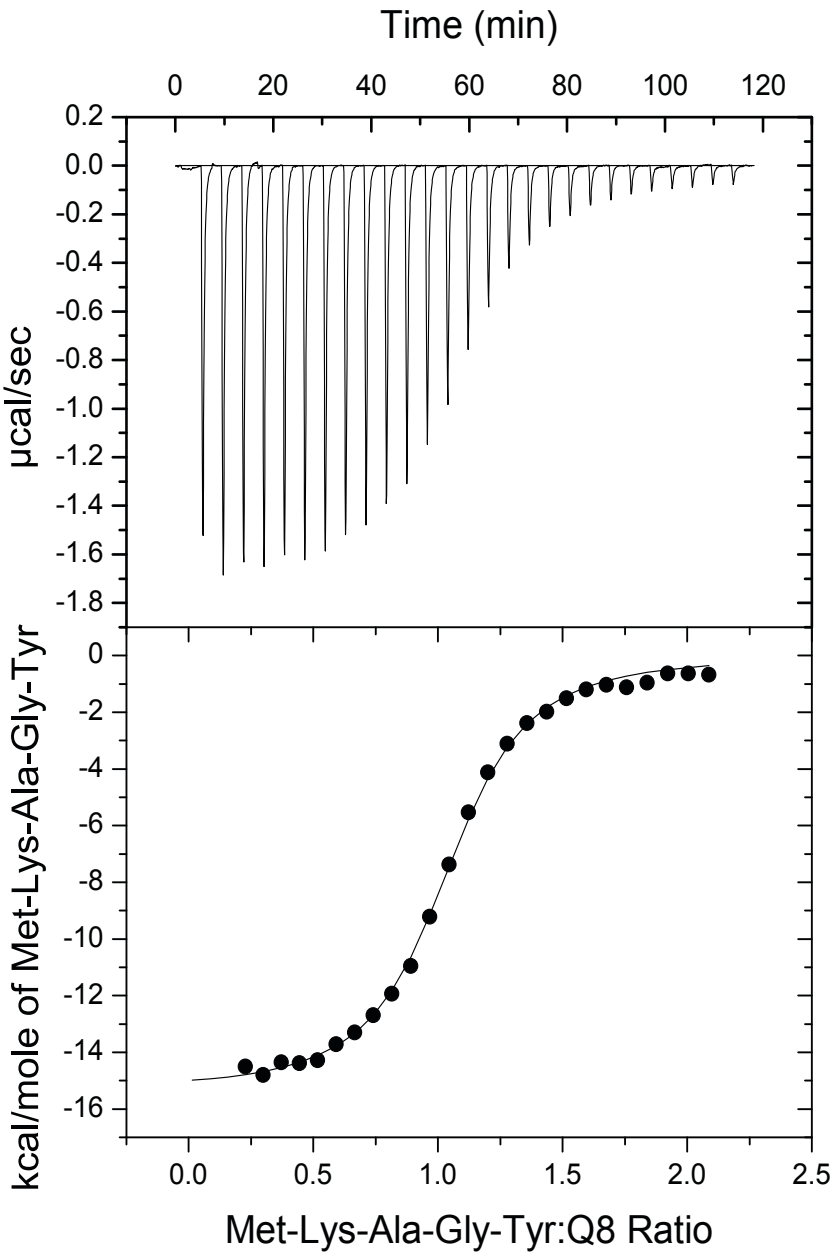


Figure S20. Representative isothermal titration calorigram of Met-Lys-Ala-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

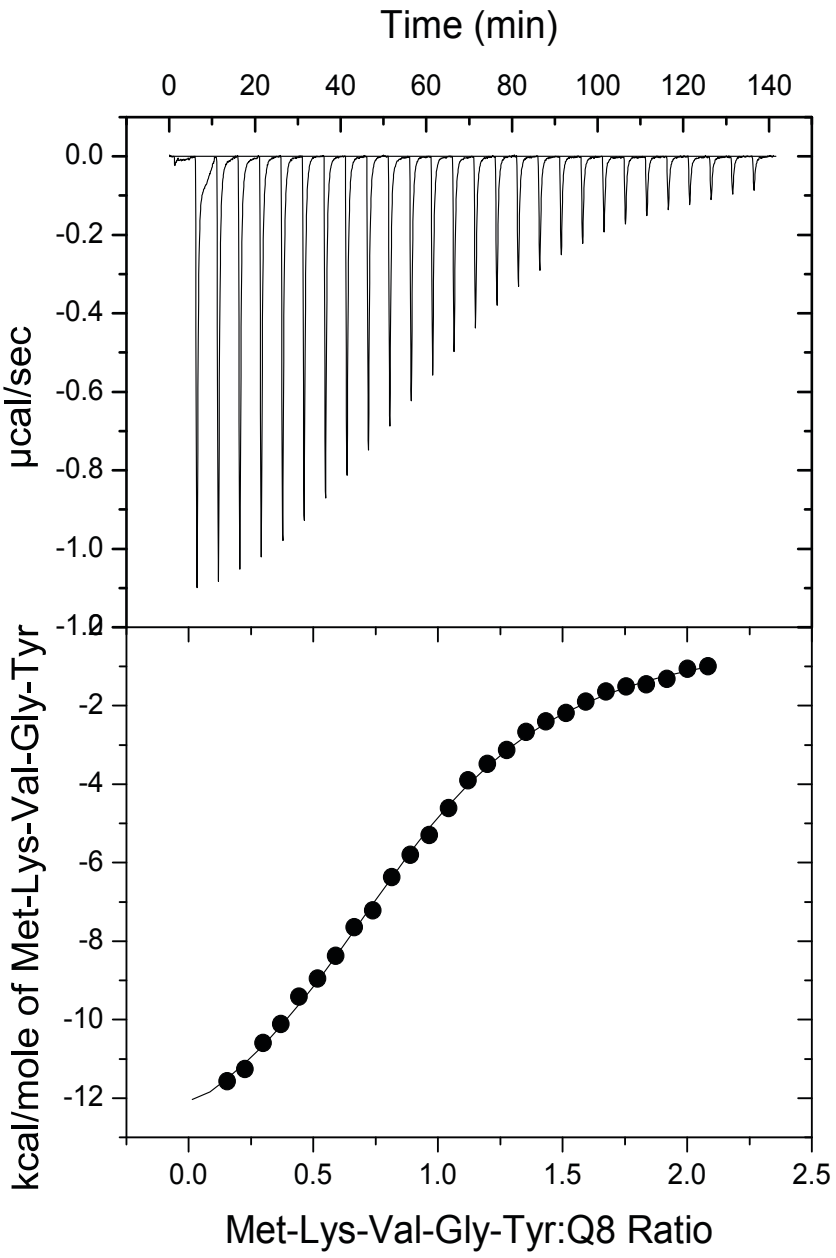


Figure S21. Representative isothermal titration calorigram of Met-Lys-Val-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

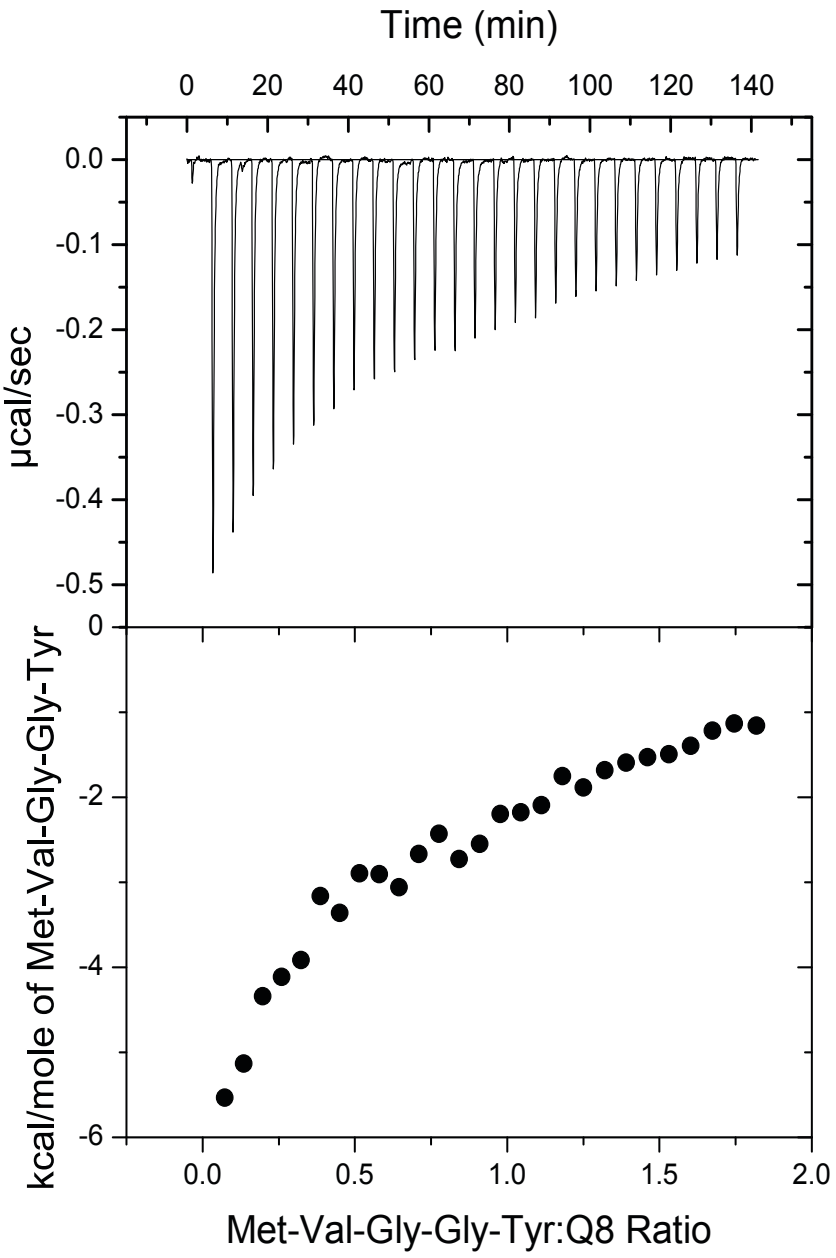


Figure S22. Representative isothermal titration calorigram of Met-Val-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

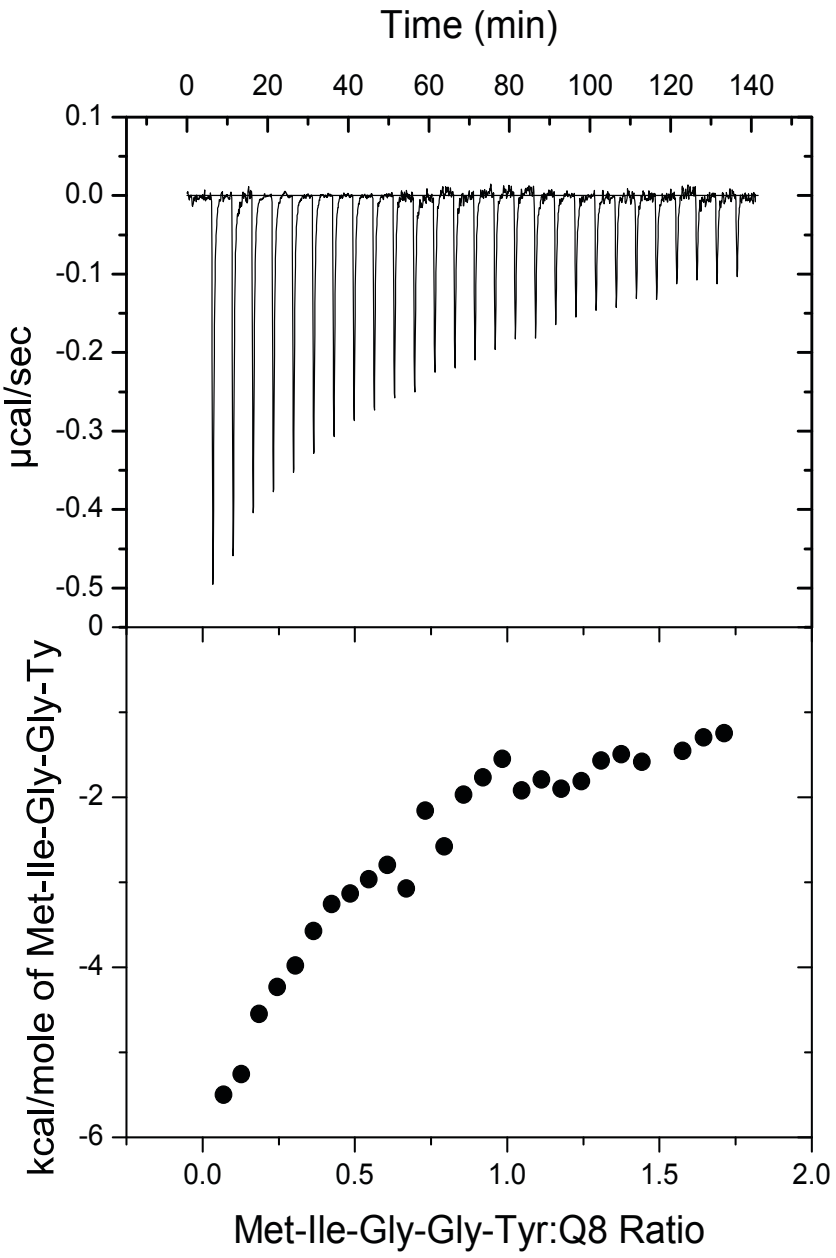


Figure S23. Representative isothermal titration calorigram of Met-Ile-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

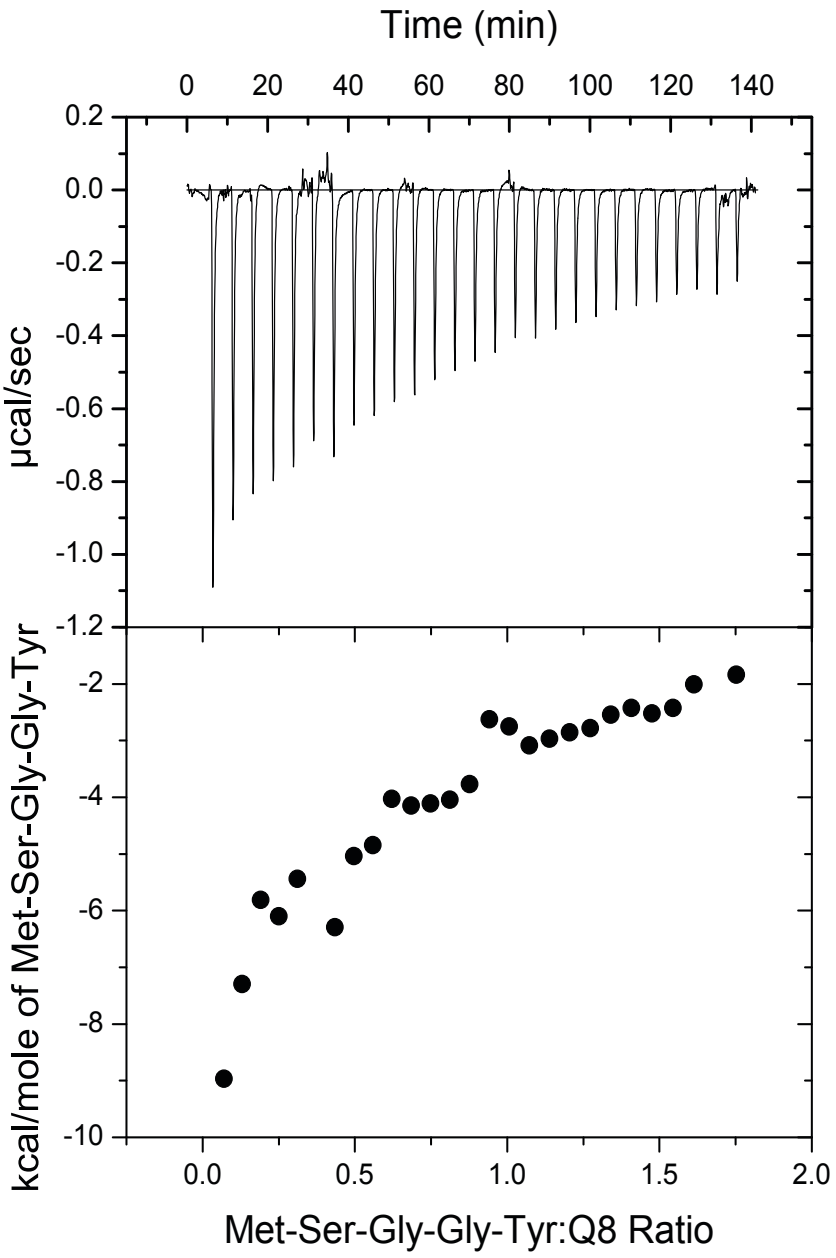


Figure S24. Representative isothermal titration calorigram of Met-Ser-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 °C.

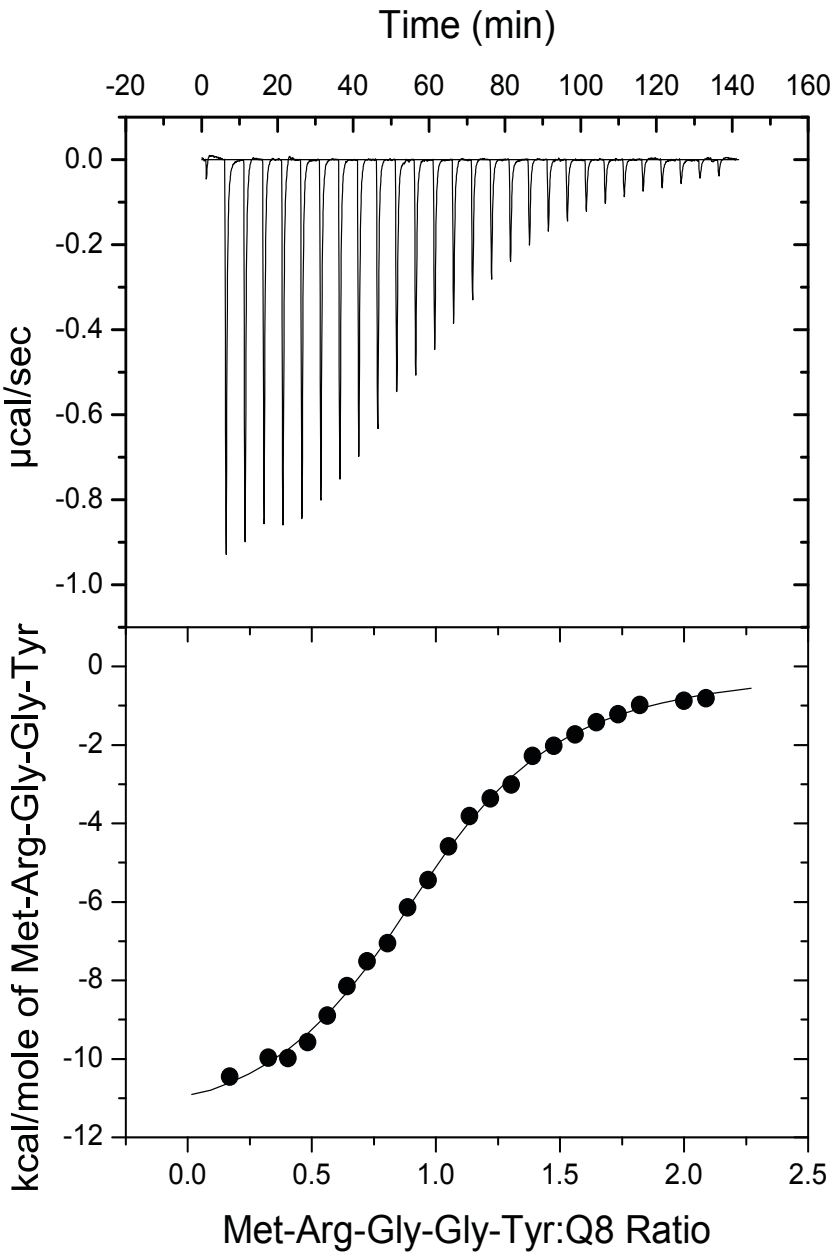


Figure S25. Representative isothermal titration calorigram of Met-Arg-Gly-Gly-Tyr titrated into Q8 in 10 mM sodium phosphate pH 7.0 at 27 ℃.

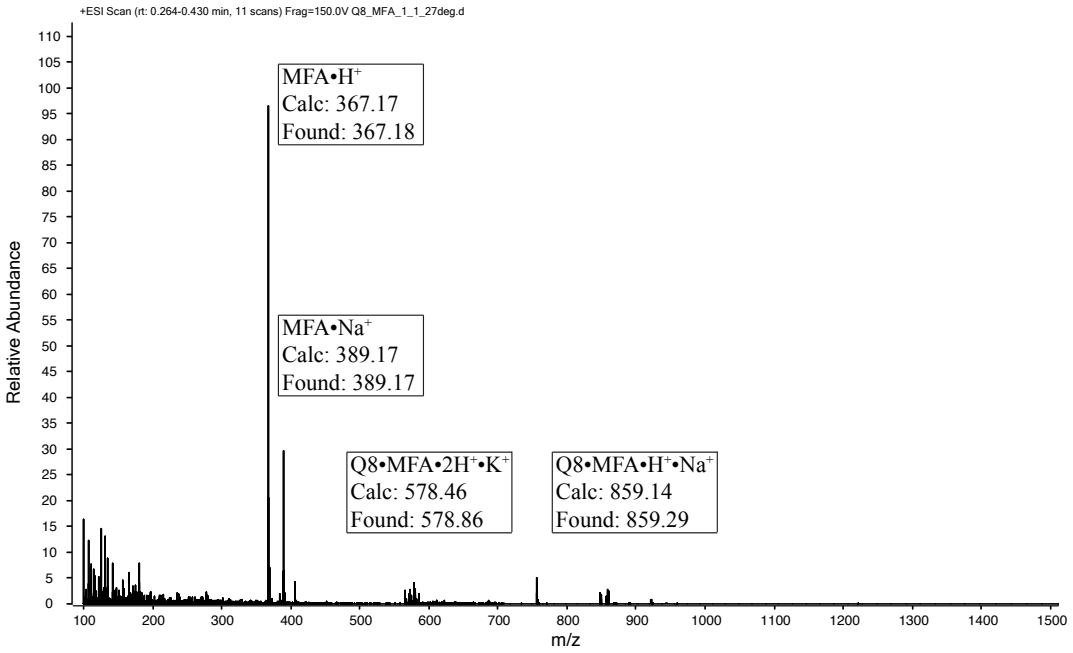


Figure S26. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Met-Phe-Ala.

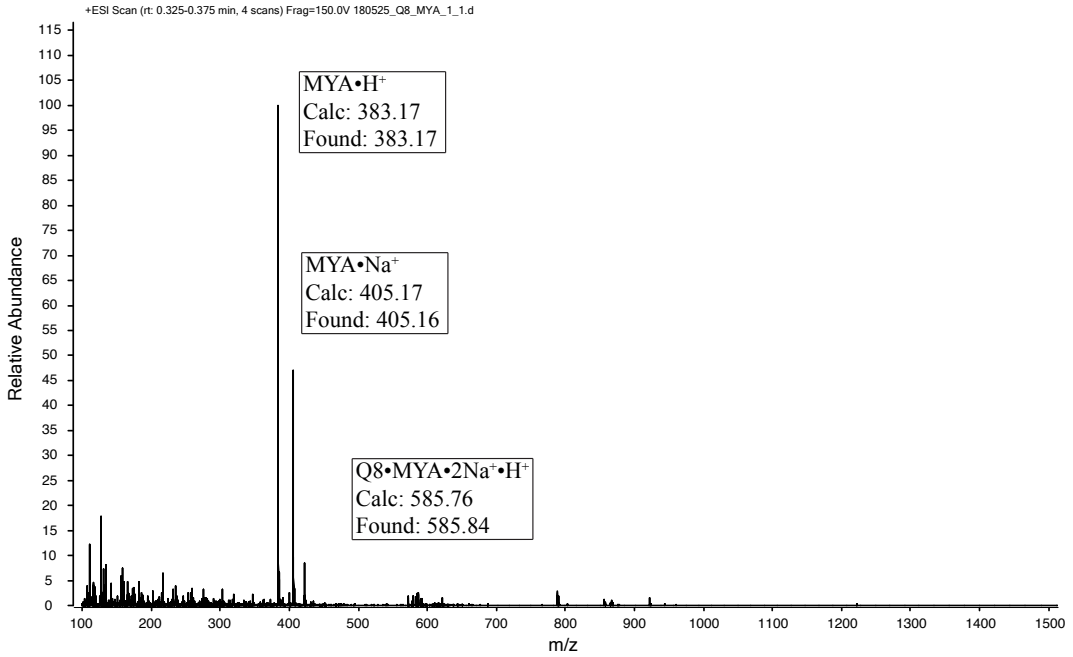


Figure S27. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 10 μM Met-Tyr-Ala.

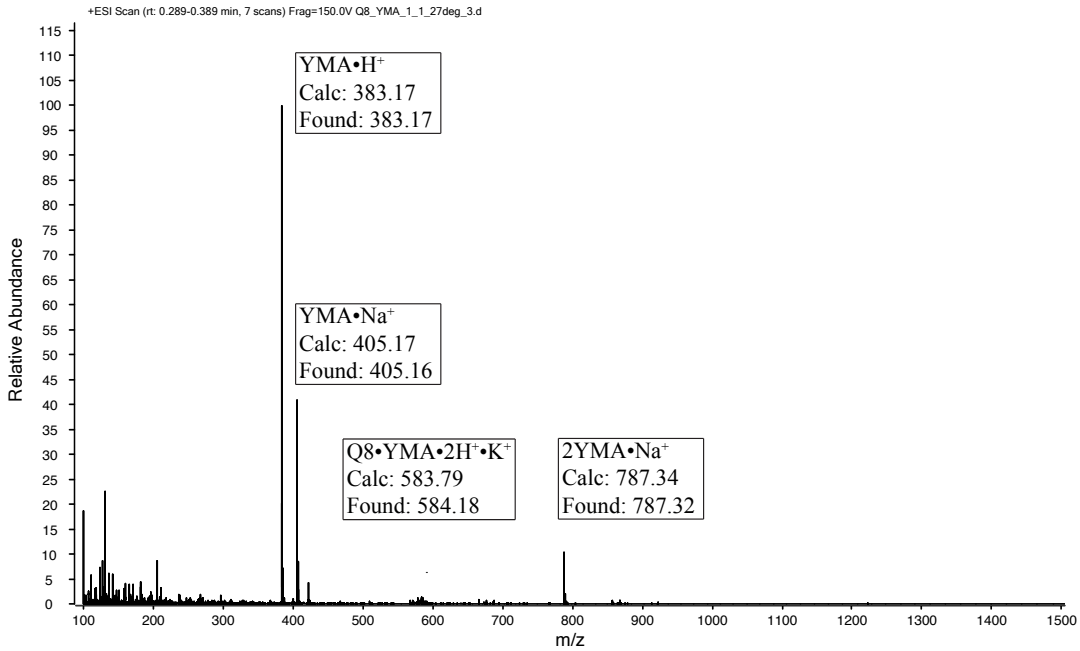


Figure S28. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Tyr-Met-Ala.

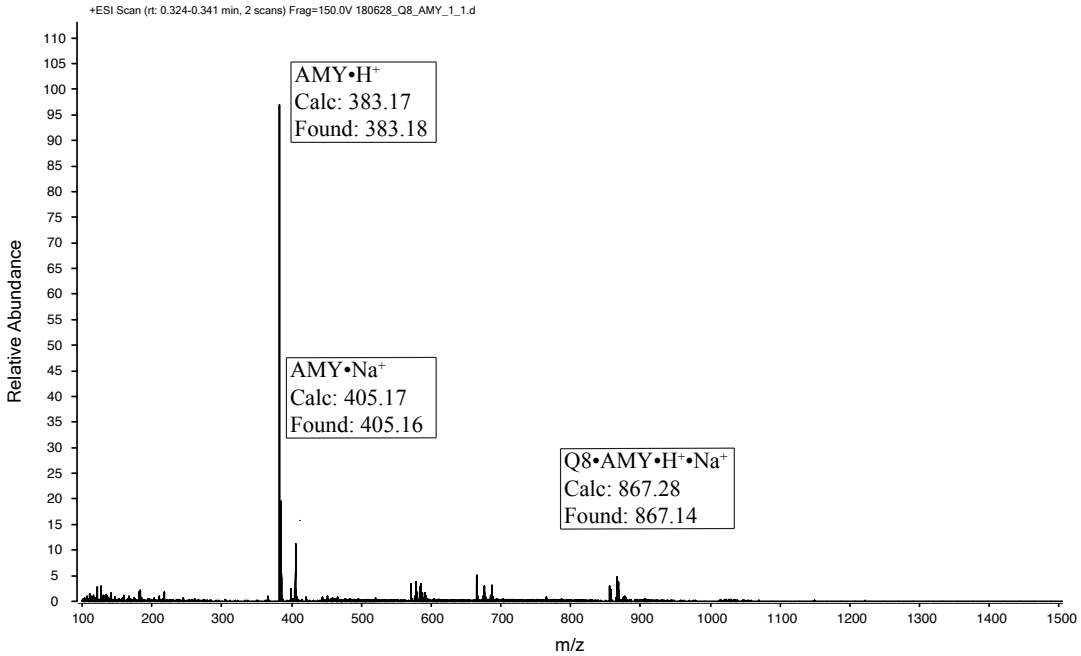


Figure S29. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 10 μM Ala-Met-Tyr.

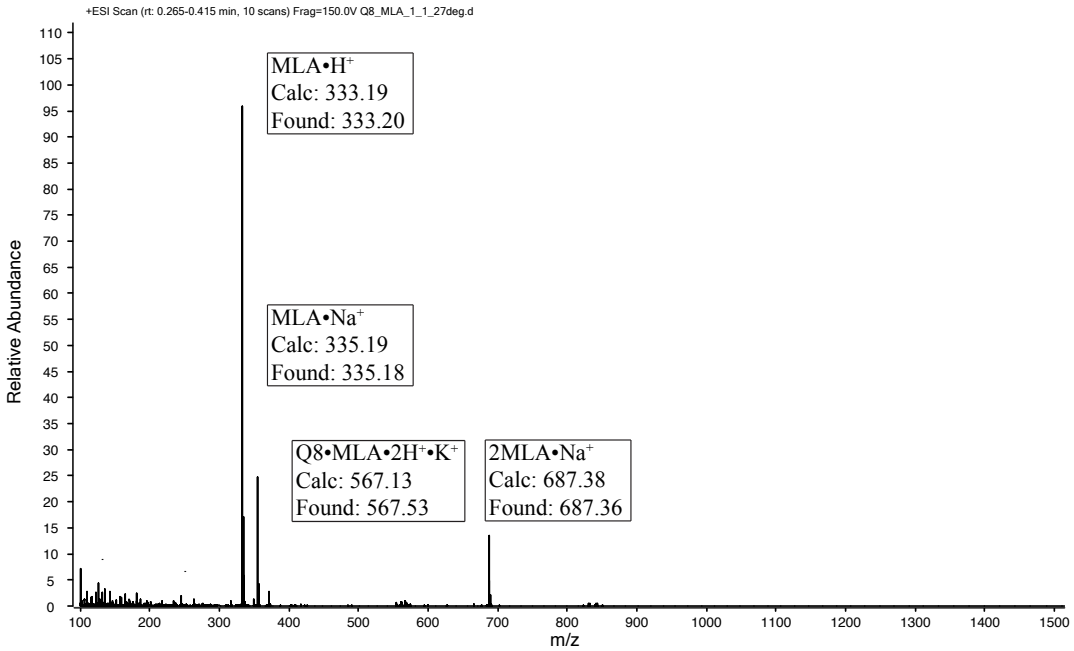


Figure S30. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Met-Leu-Ala.

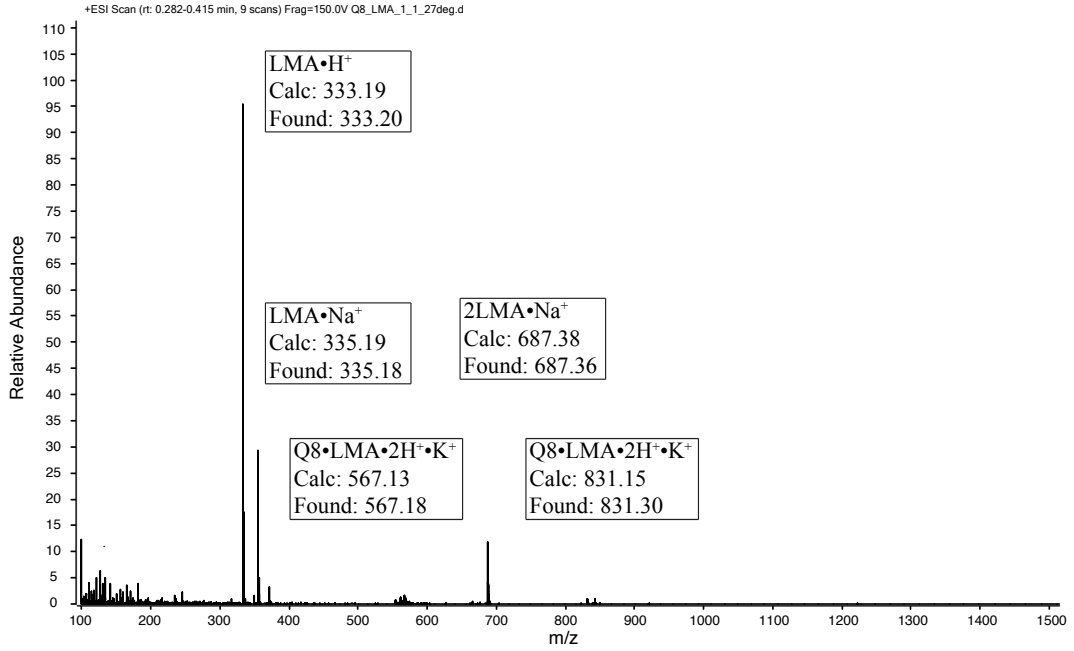


Figure S31. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Leu-Met-Ala.

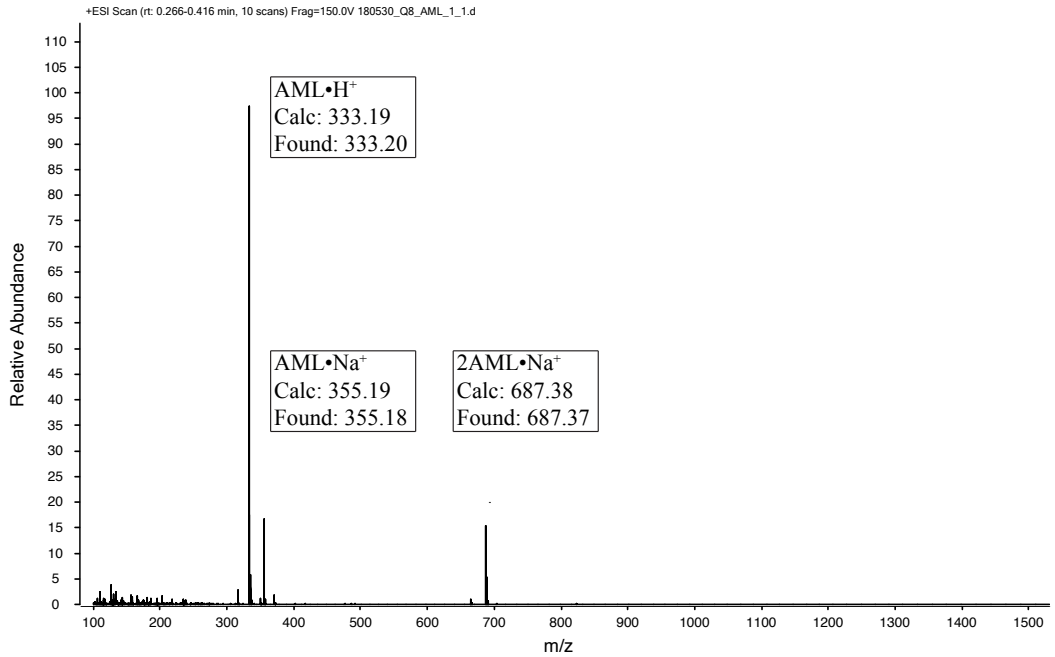


Figure S32. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Ala-Met-Leu.

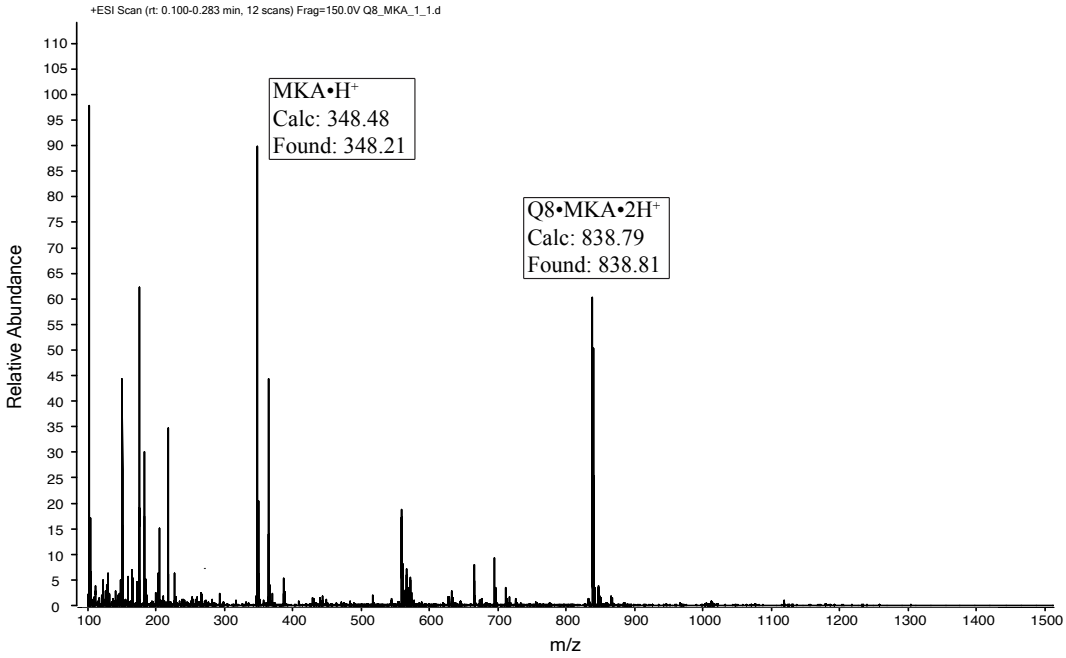


Figure S33. ESI-MS positive ion mode spectrum for a mixture containing 60 μM Q8 and 60 μM Met-Lys-Ala.

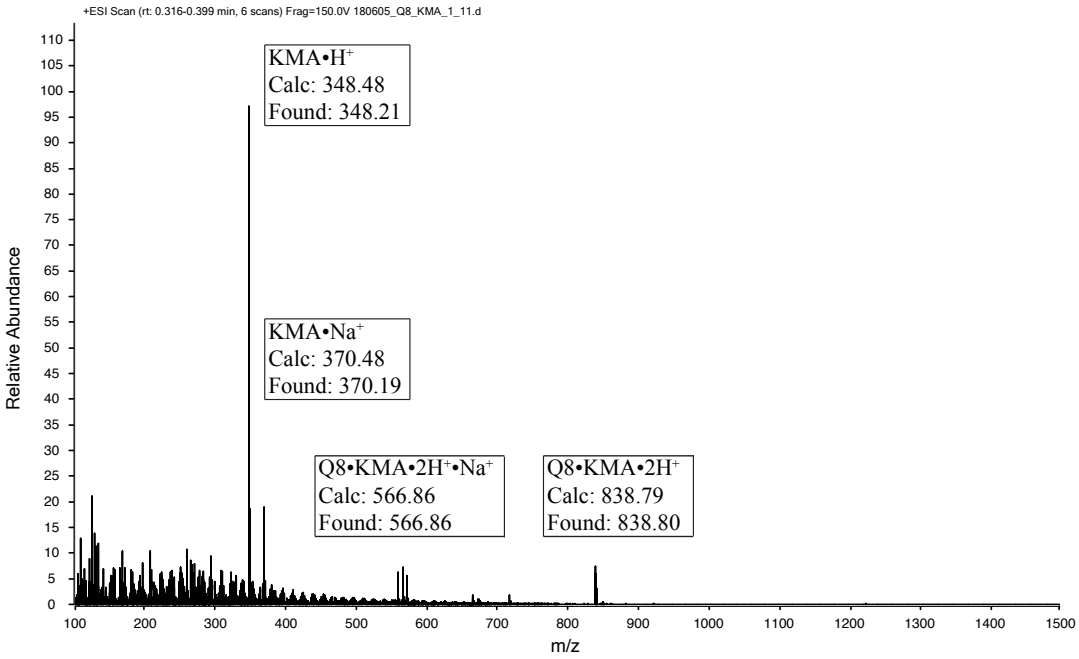


Figure S34. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 10 μM Lys-Met-Ala.

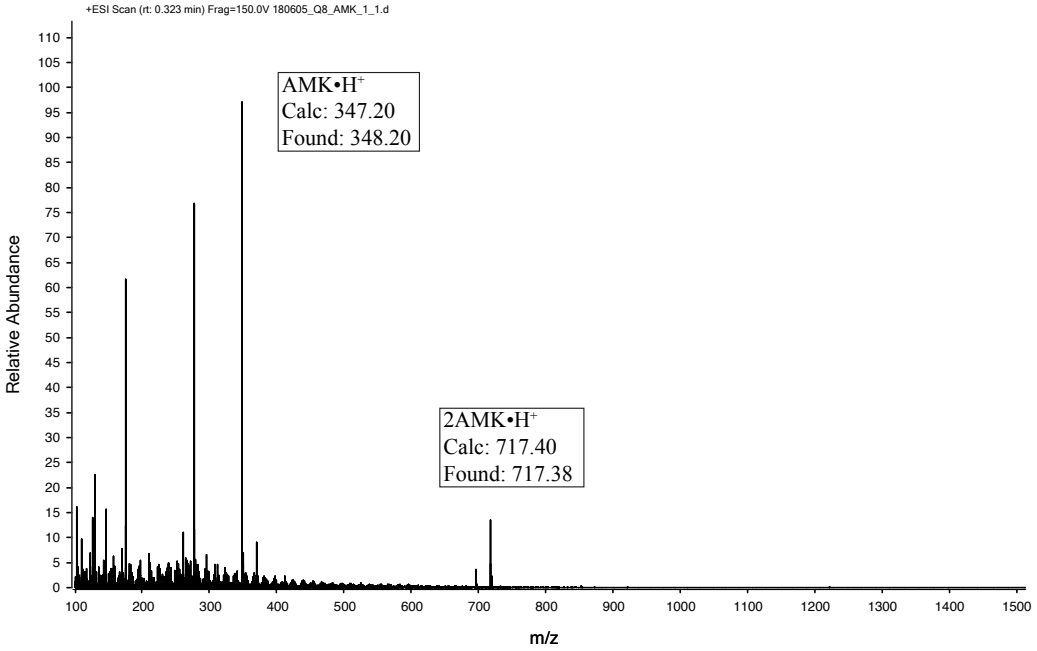


Figure S35. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 10 μM Ala-Met-Lys.

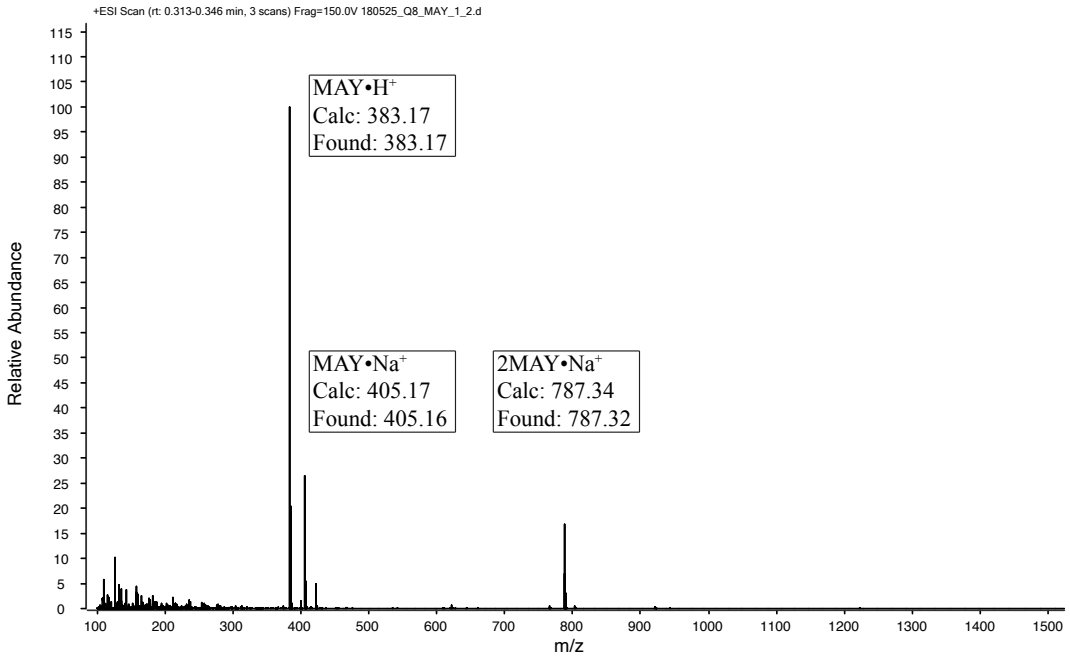


Figure S36. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 10 μM Met-Ala-Tyr.

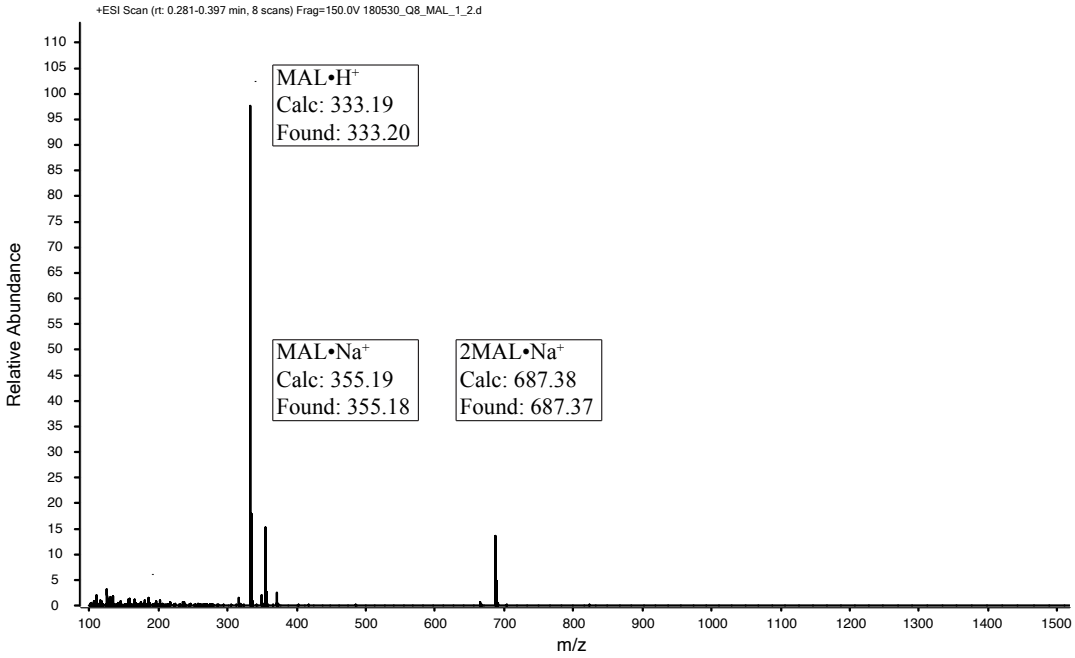


Figure S37. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Met-Ala-Leu.

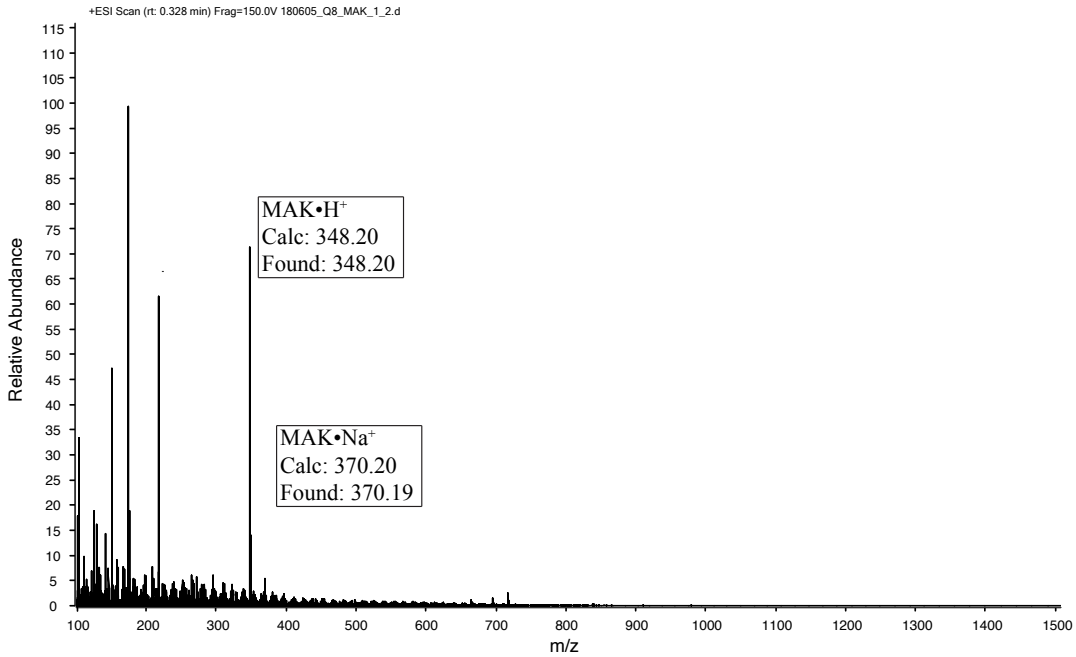


Figure S38. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 10 μM Met-Ala-Lys.

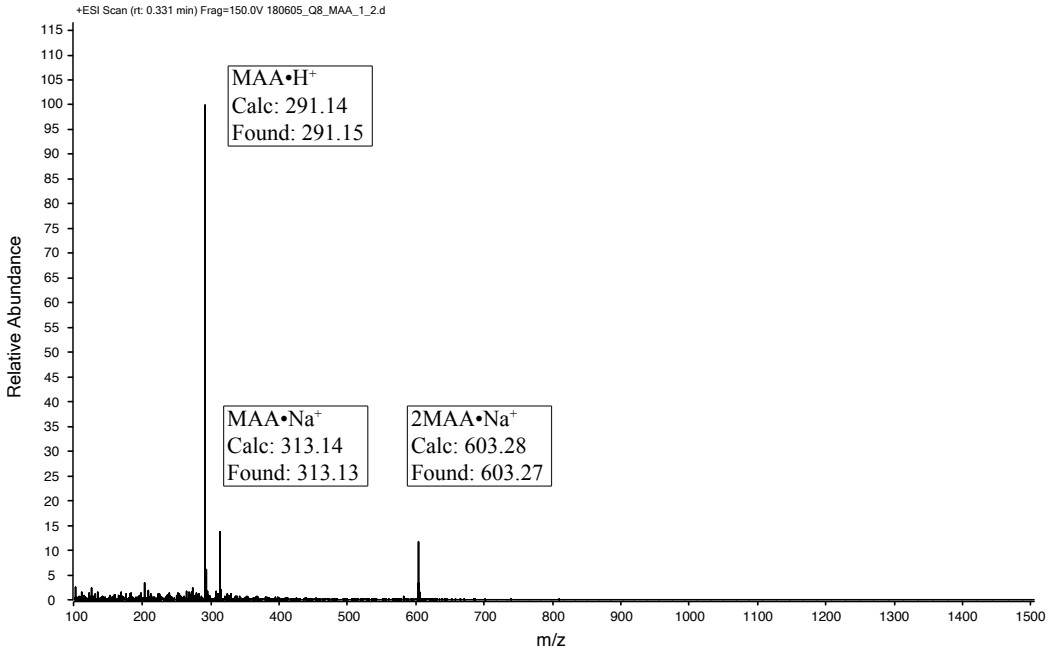


Figure S39. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 10 μM Met-Ala-Ala.

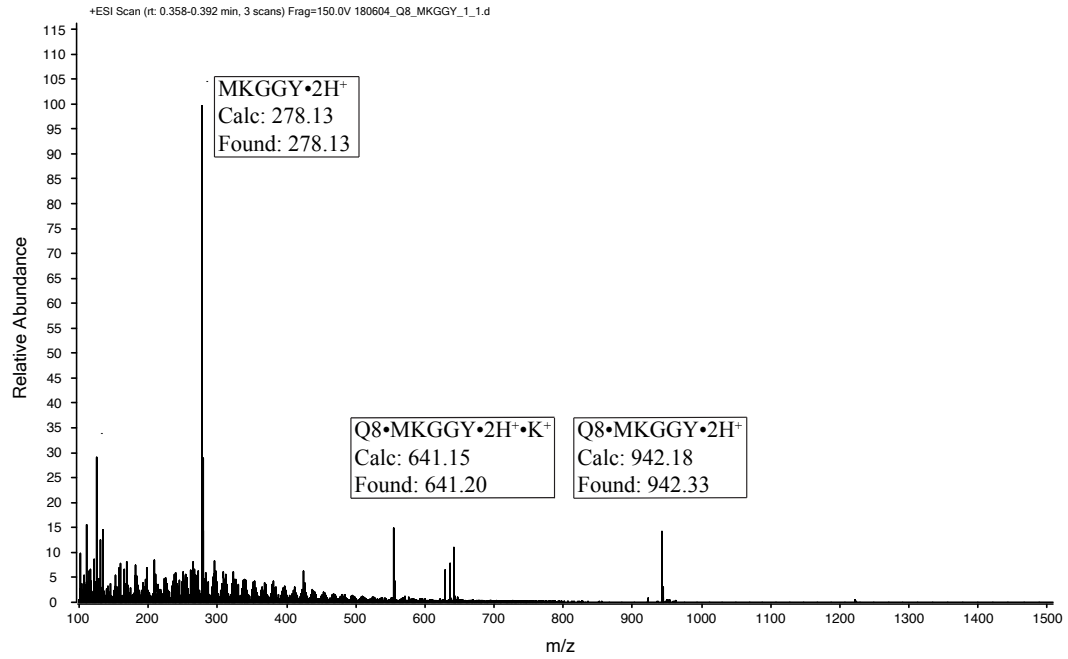


Figure S40. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Met-Lys-Gly-Gly-Tyr.

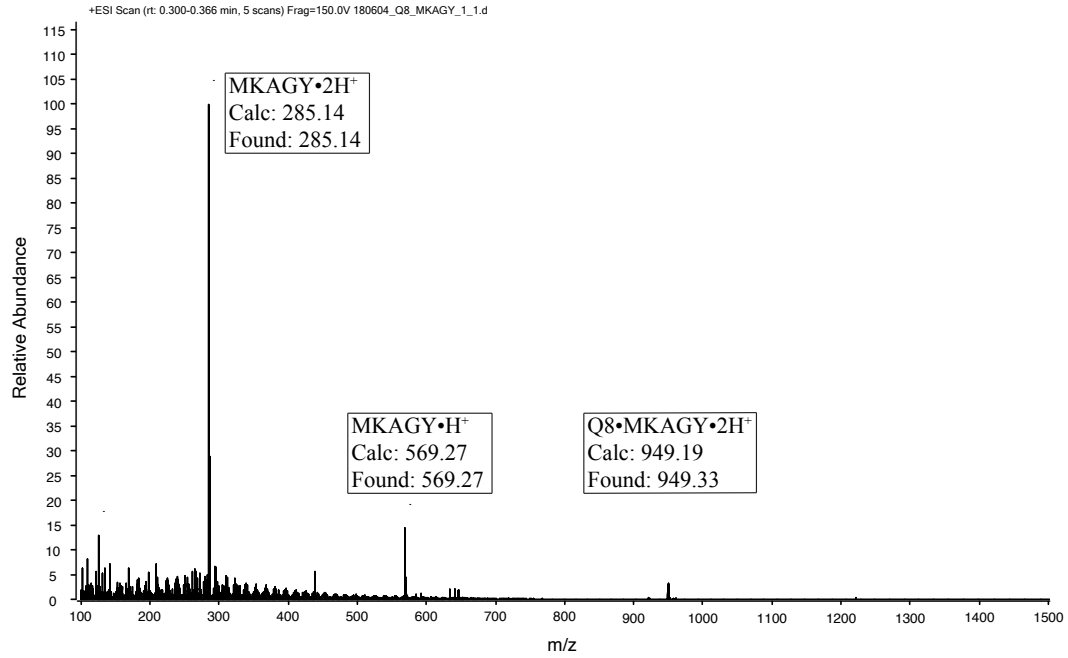


Figure S41. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Met-Lys-Ala-Gly-Tyr.

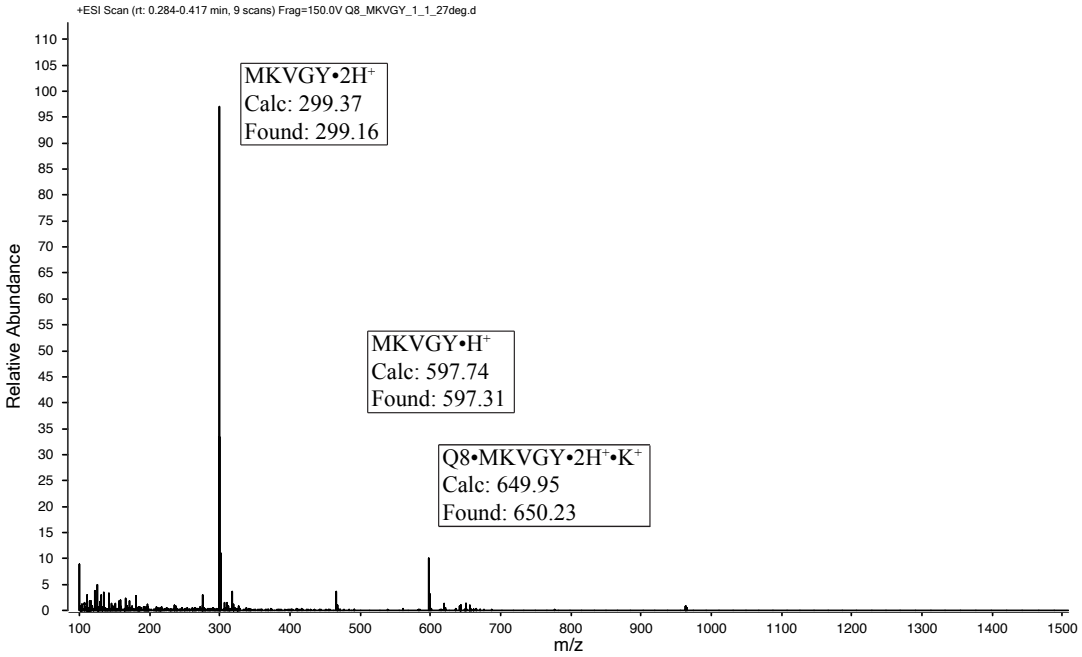


Figure S42. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Met-Lys-Val-Gly-Tyr.

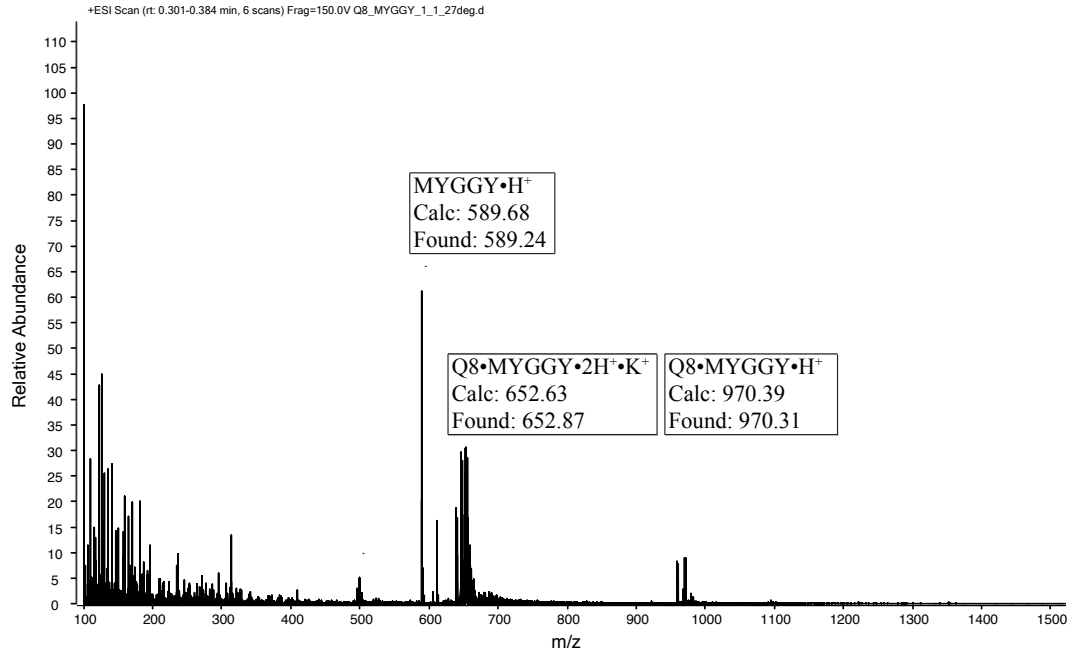


Figure S43. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Met-Tyr-Gly-Gly-Tyr.

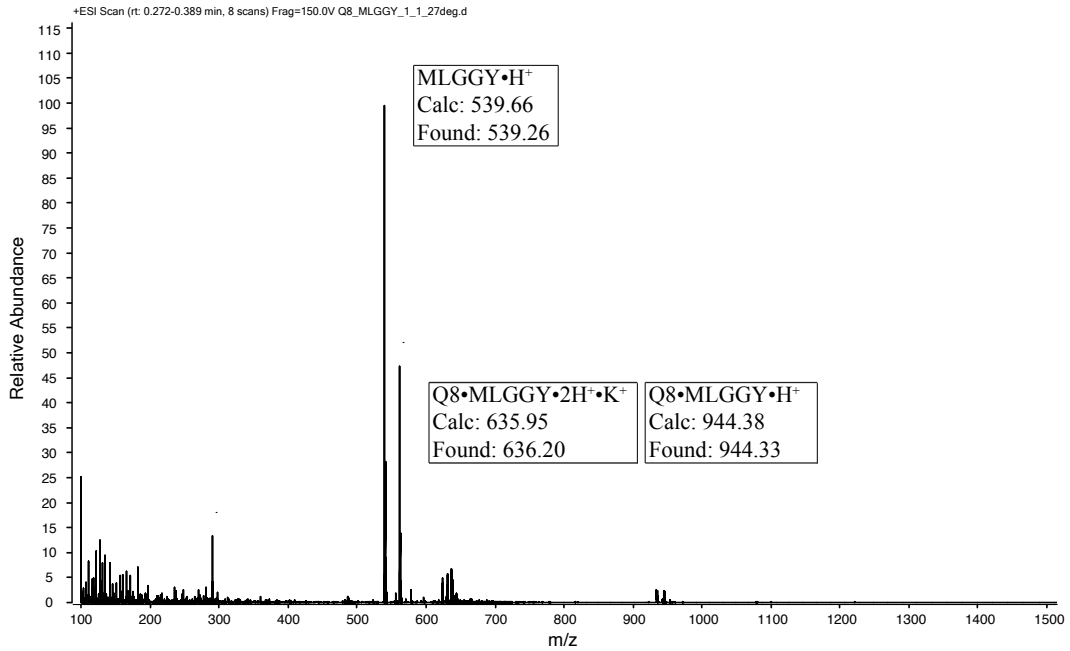


Figure S44. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Met-Leu-Gly-Gly-Tyr.

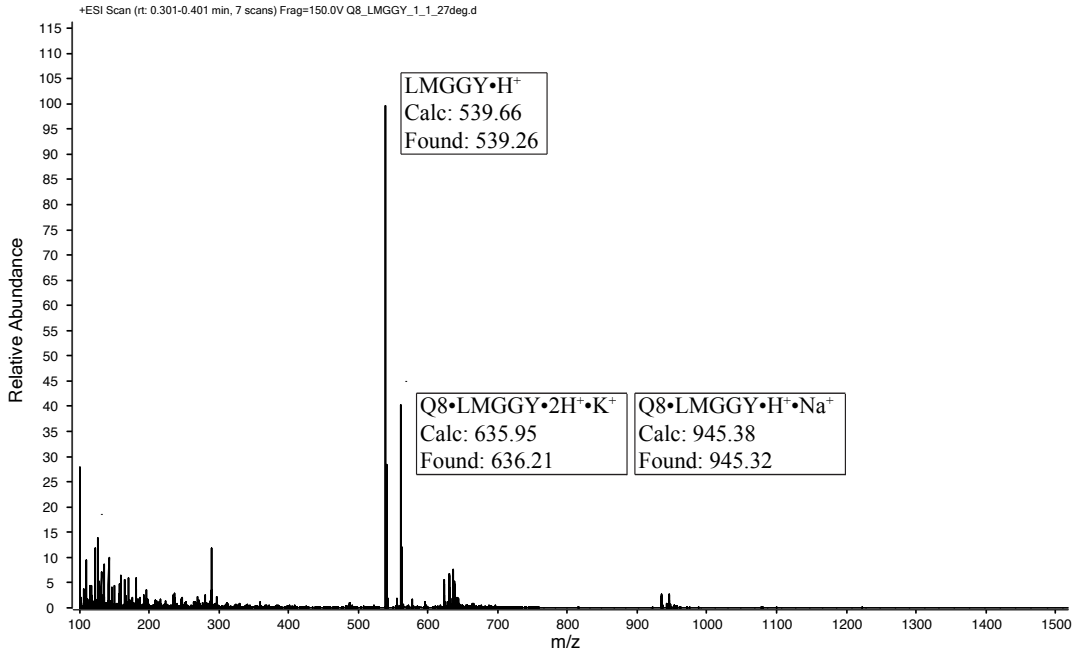


Figure S45. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Leu-Met-Gly-Gly-Tyr.

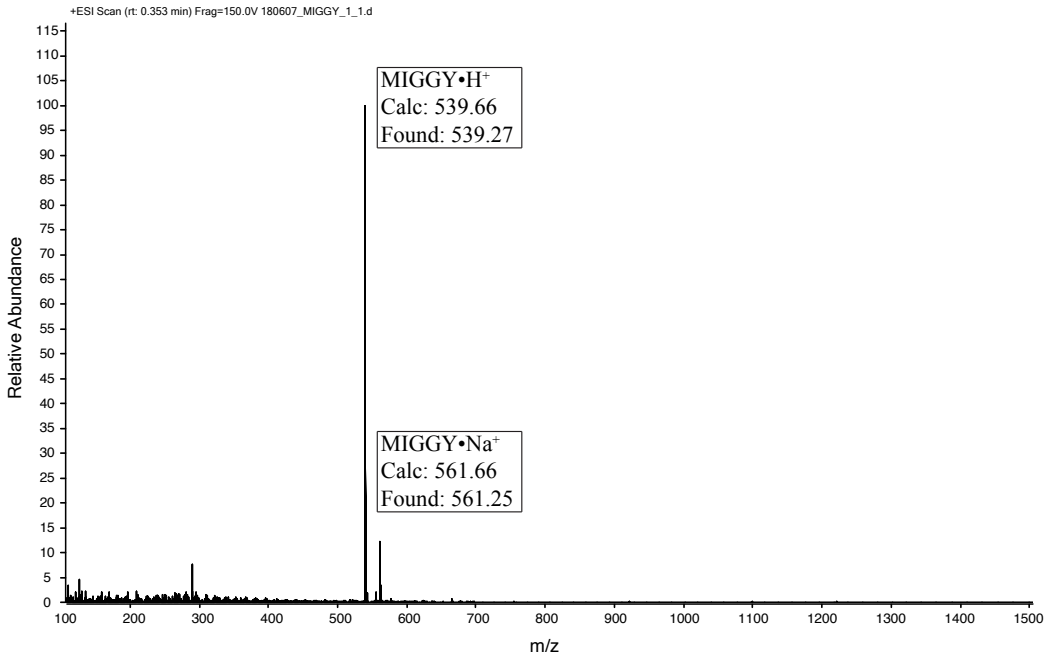


Figure S46. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Met-Ile-Gly-Gly-Tyr.

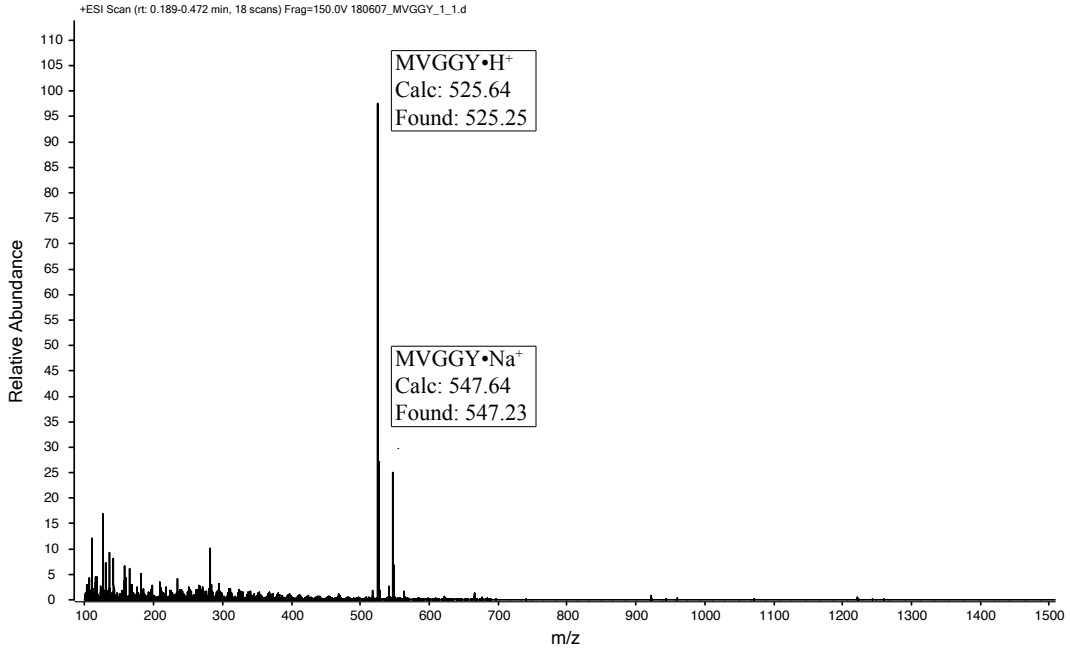


Figure S47. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Met-Val-Gly-Gly-Tyr.

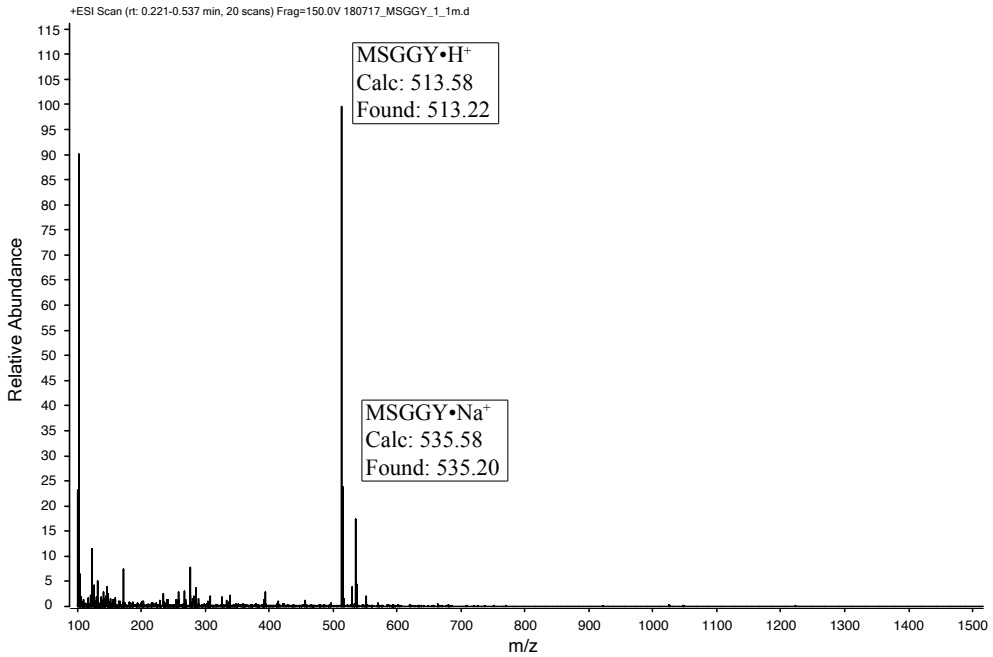


Figure S48. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Met-Ser-Gly-Gly-Tyr.

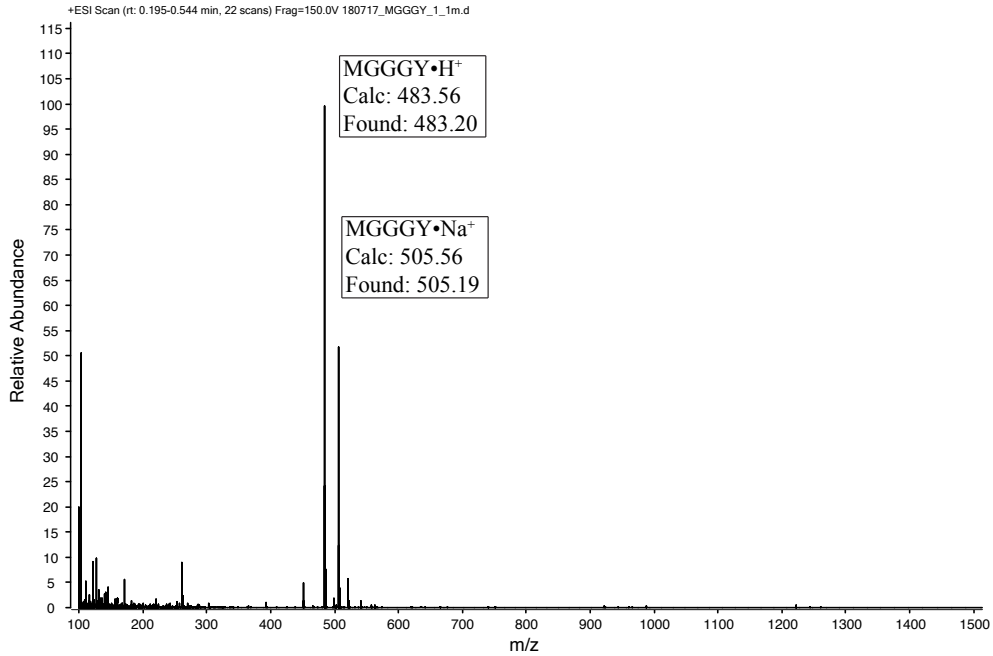


Figure S49. ESI-MS positive ion mode spectrum for a mixture containing 20 μM Q8 and 20 μM Met-Gly-Gly-Gly-Tyr.

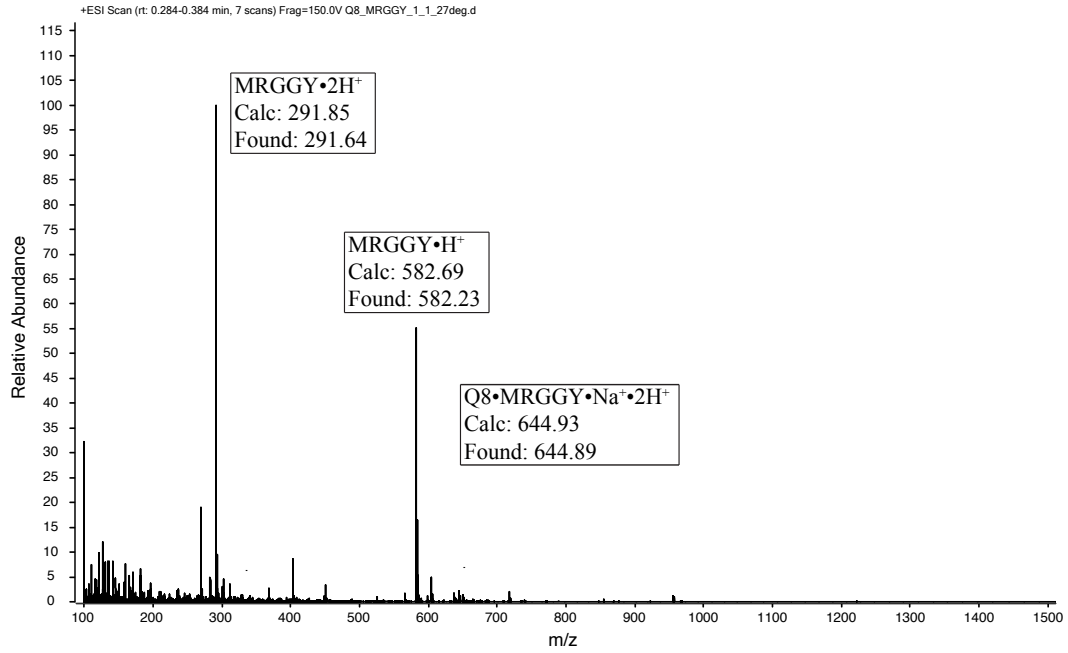


Figure S50. ESI-MS positive ion mode spectrum for a mixture containing 10 μM Q8 and 20 μM Met-Arg-Gly-Gly-Tyr.

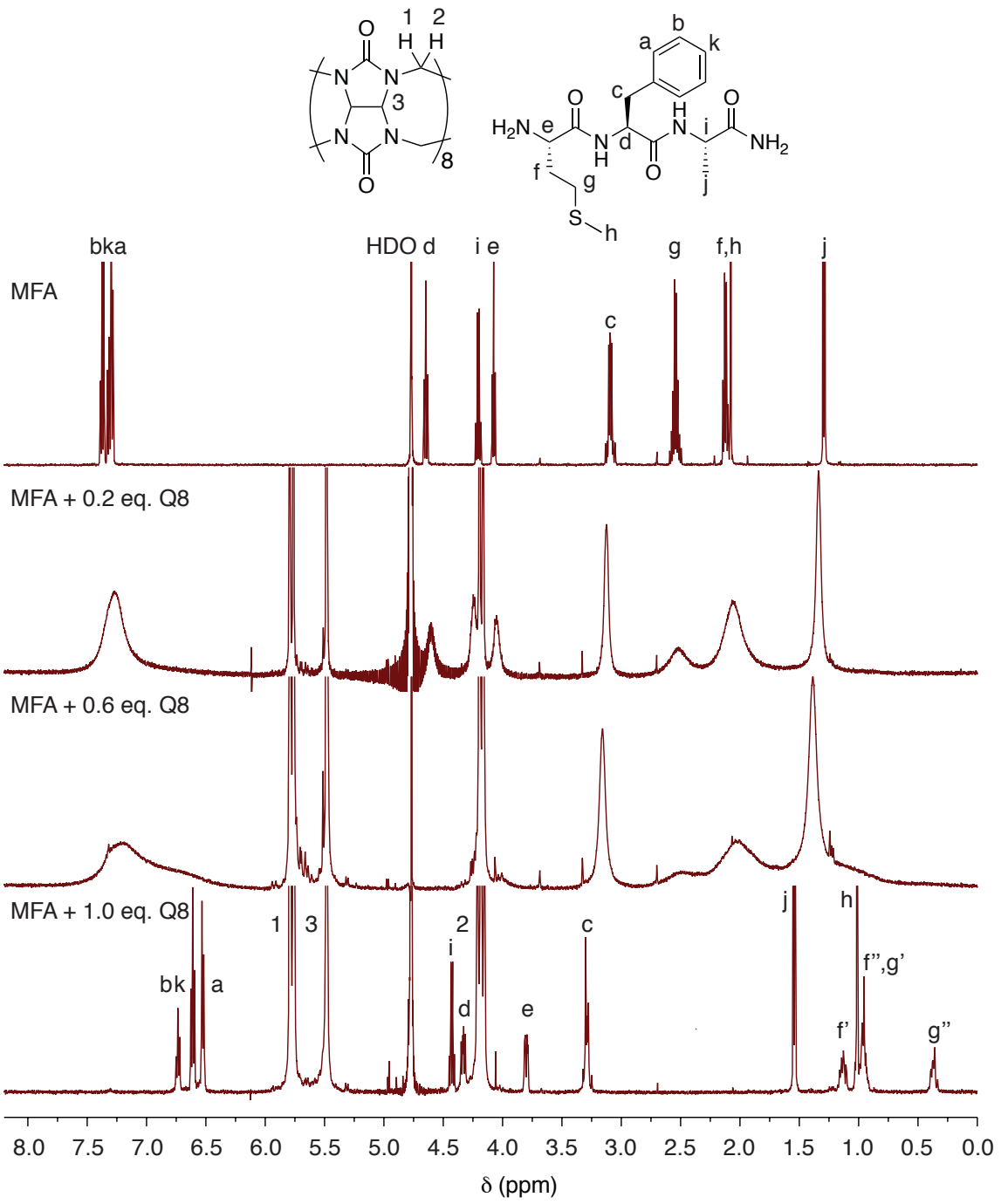


Figure S51. 500 MHz ^{1}H NMR speactra of Met-Phe-Ala with 0 eq. Q8, 0.2 eq. Q8, 0.6 eq. Q8, and 1.0 eq. Q8 at 25 ℃ in D_2O. Apostrophes indicate geminal separation.

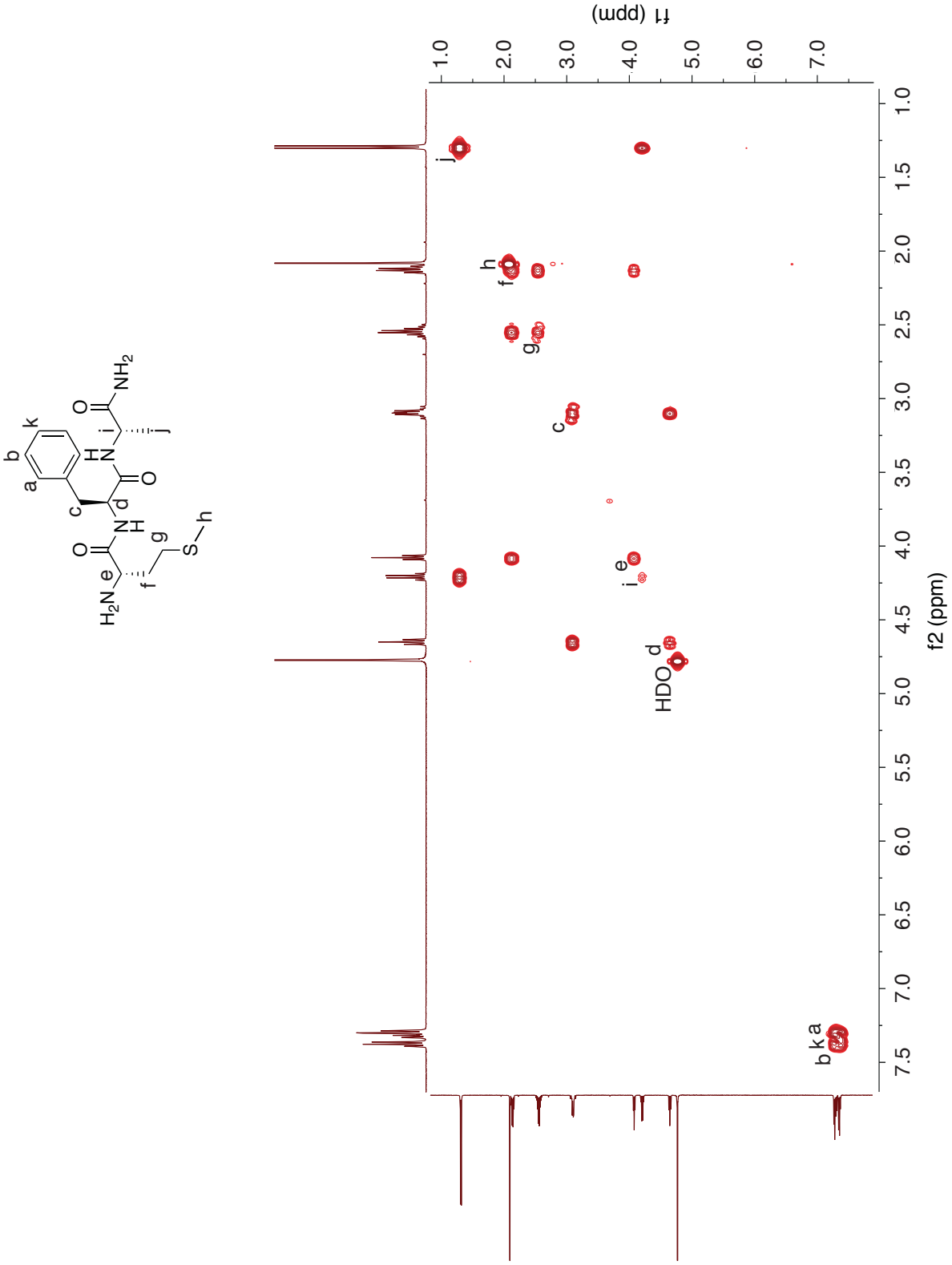


Figure S52. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Phe-Ala at 25 ℃ in D_2O.

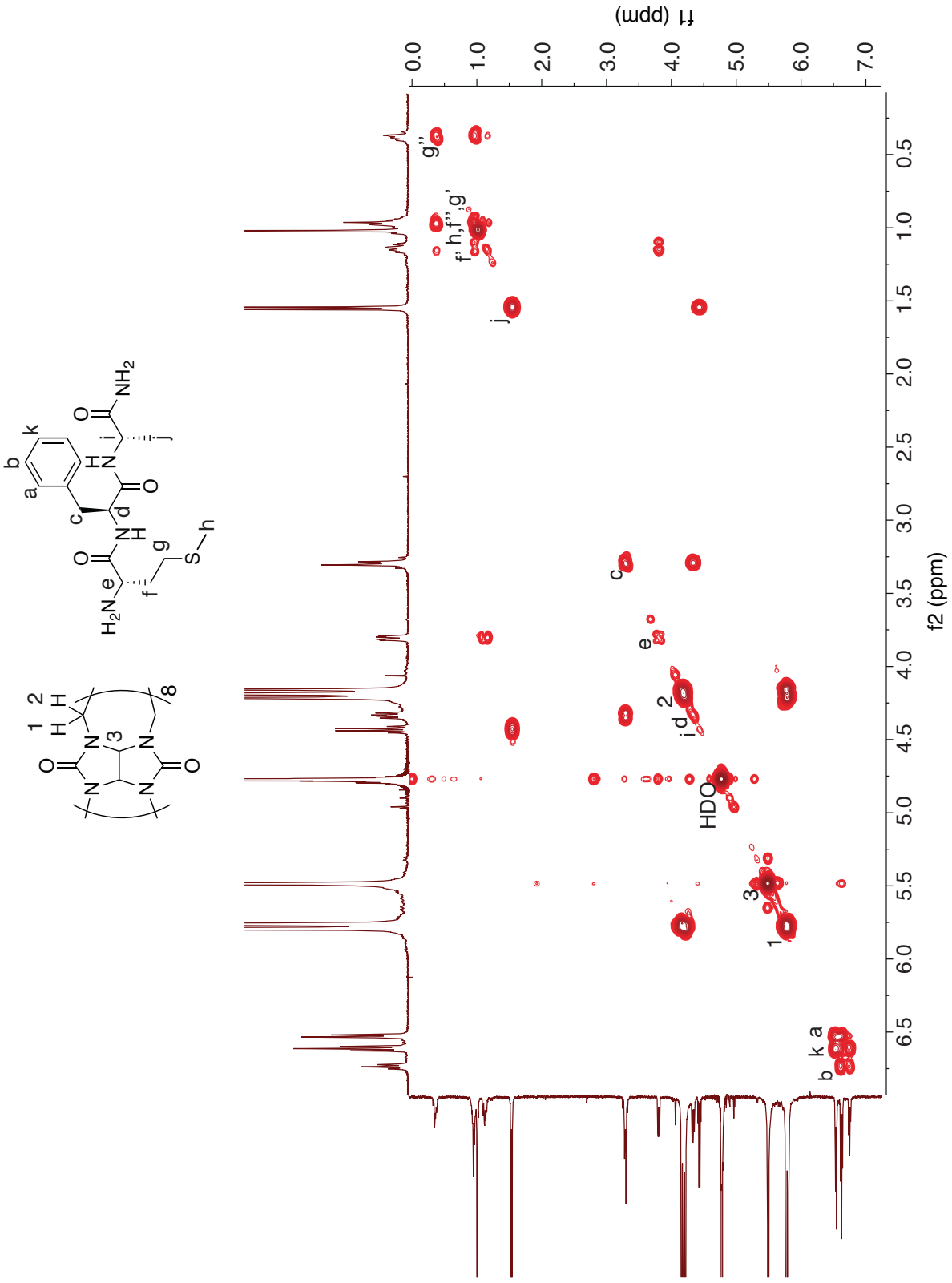


Figure S53. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Phe-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

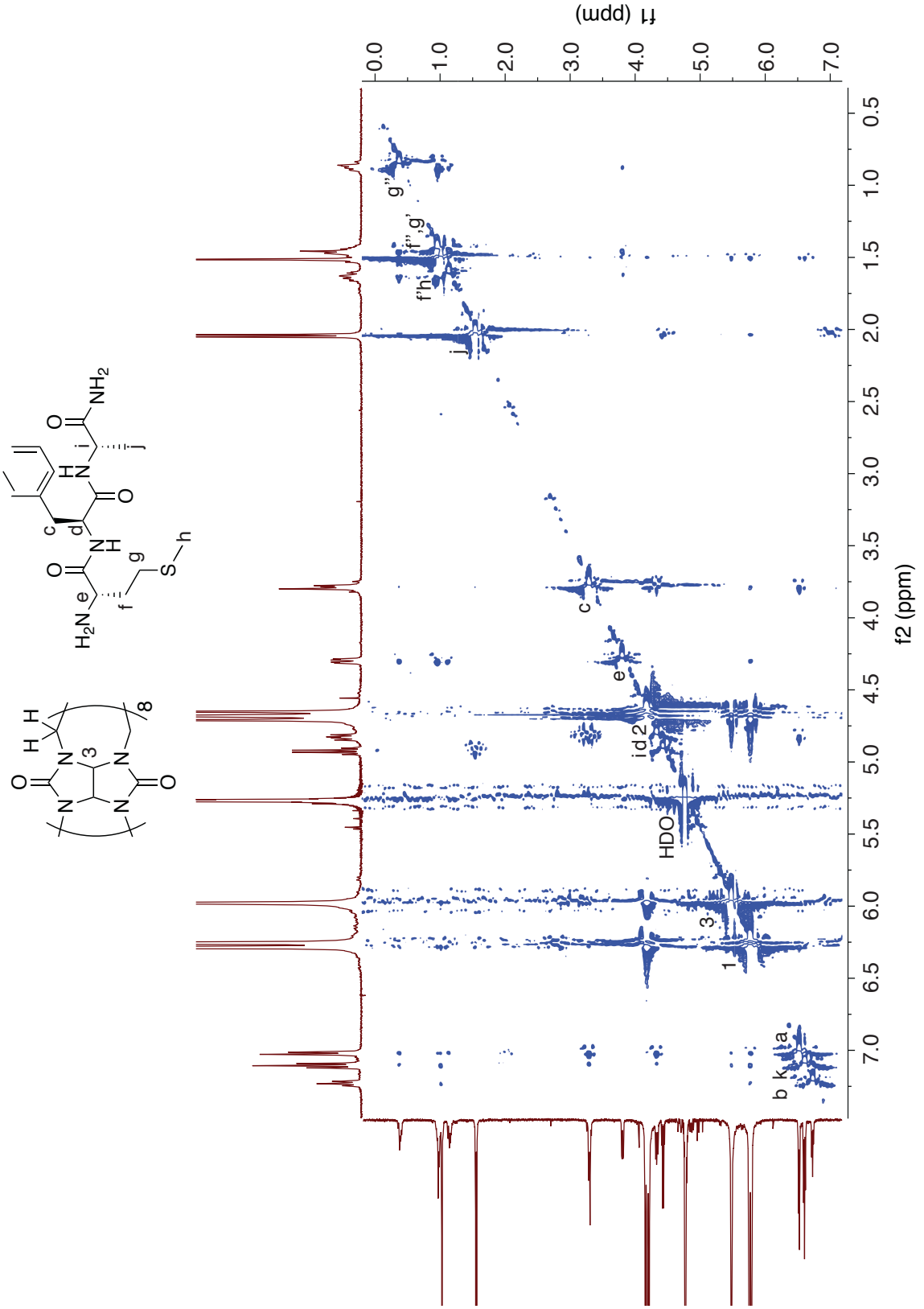


Figure S54. 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Met-Phe-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

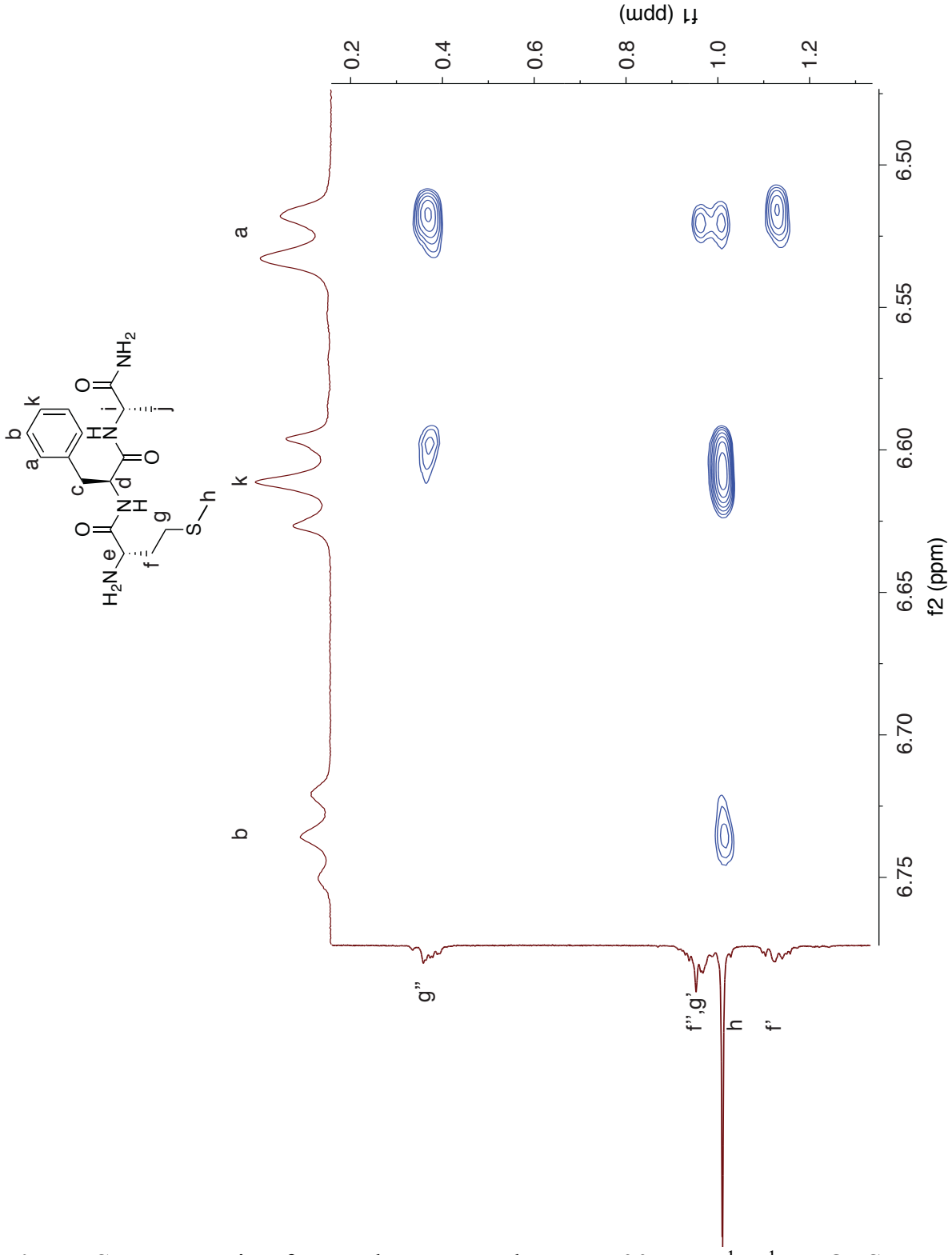


Figure S55. Zoom-in of Met-Phe crosspeaks on a 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Met-Phe-Ala at 25 ℃ in D_2O. Apostrophes indicate geminal separation.

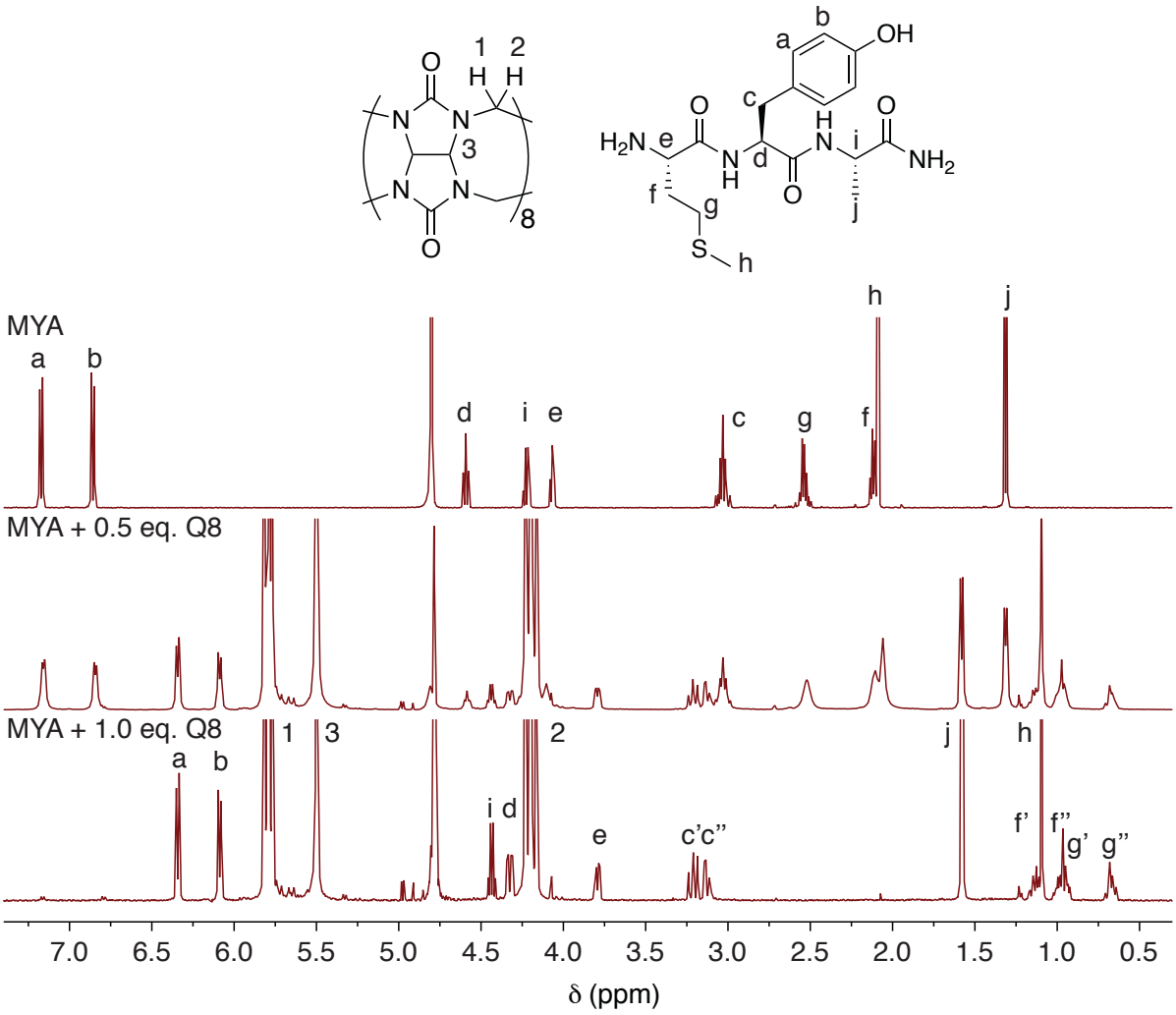


Figure S56. 500 MHz ^{1}H NMR speactra of Met-Tyr-Ala with 0 eq. Q8, 0.2 eq. Q8, 0.6 eq. Q8, and 1.0 eq. Q8 at 25 ℃ in D_2O. Apostrophes indicate geminal separation.

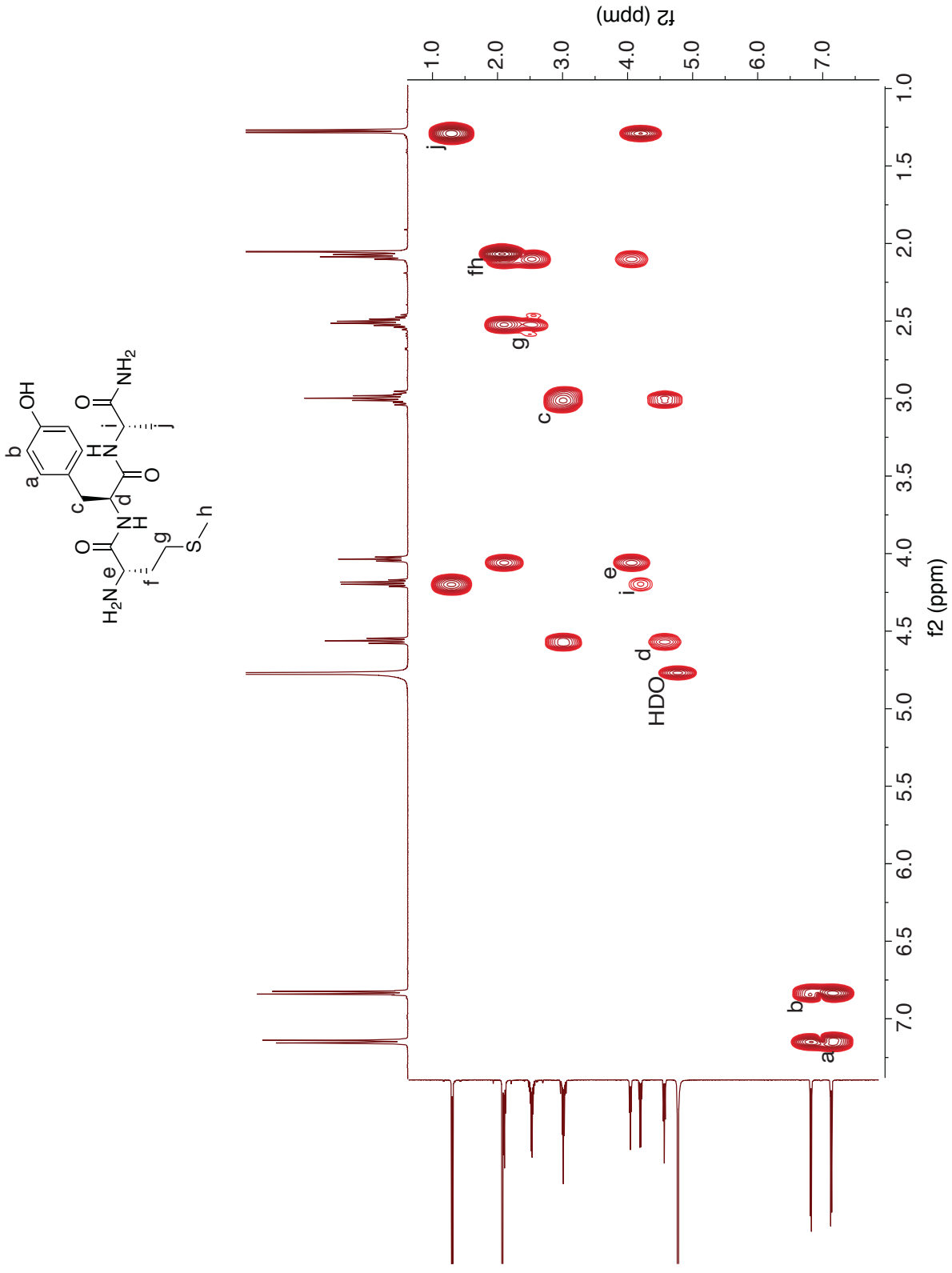


Figure S57. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Tyr-Ala at 25 ℃ in D_2O.

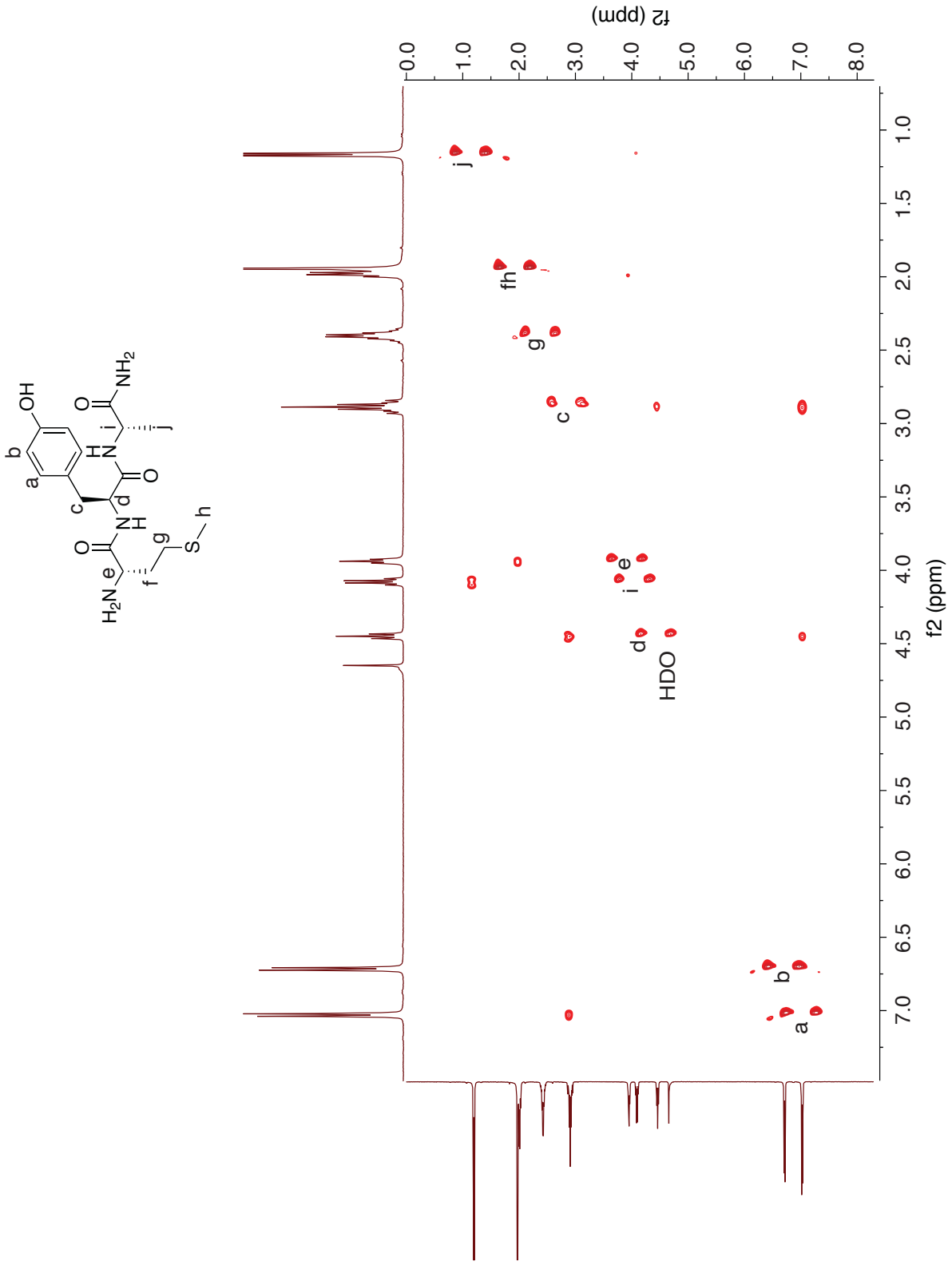


Figure S58. 500 MHz ^{1}H-^{1}H NOESY spectrum of Met-Tyr-Ala at 25 °C in D_2O.

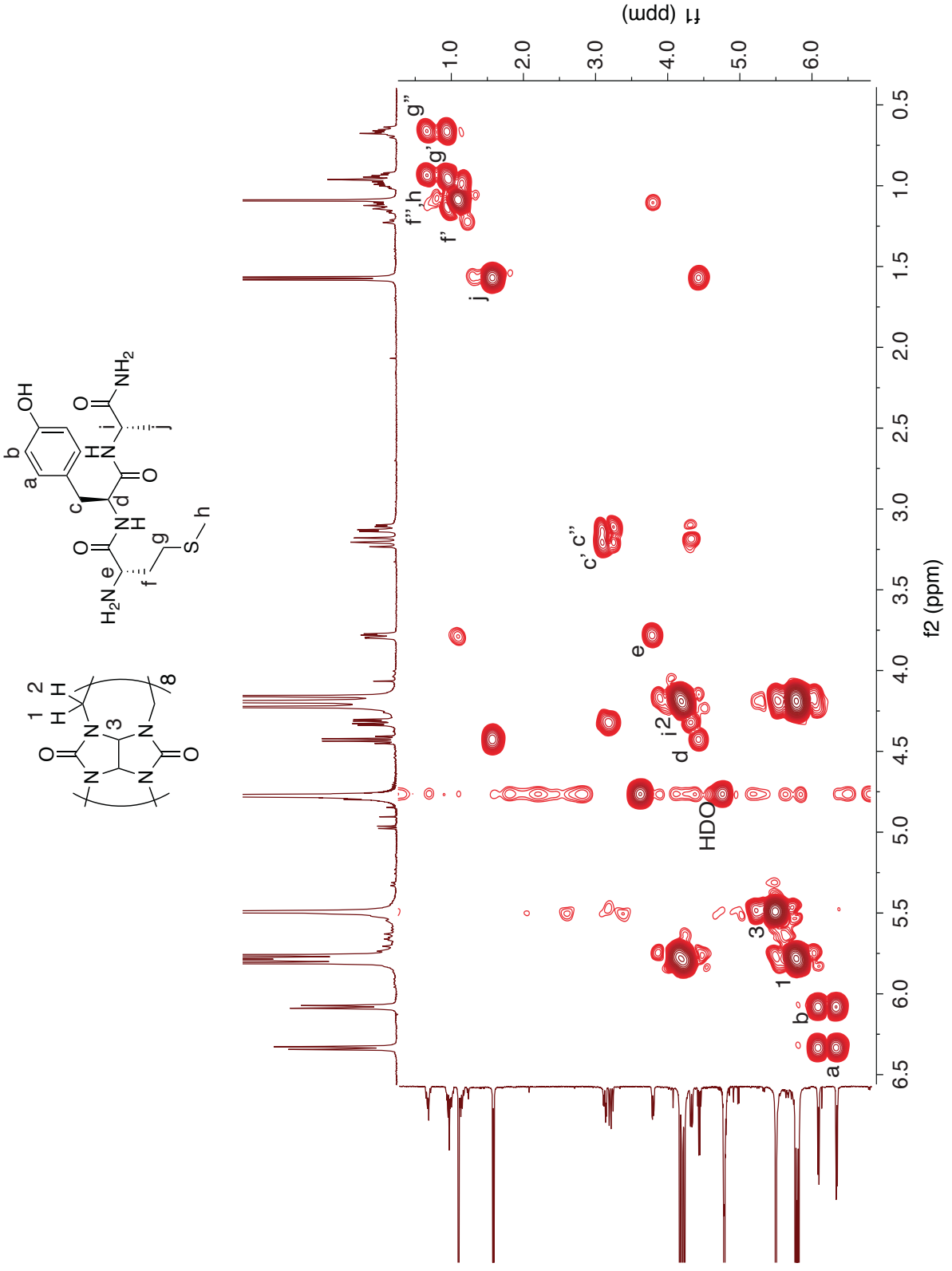


Figure S59. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Tyr-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

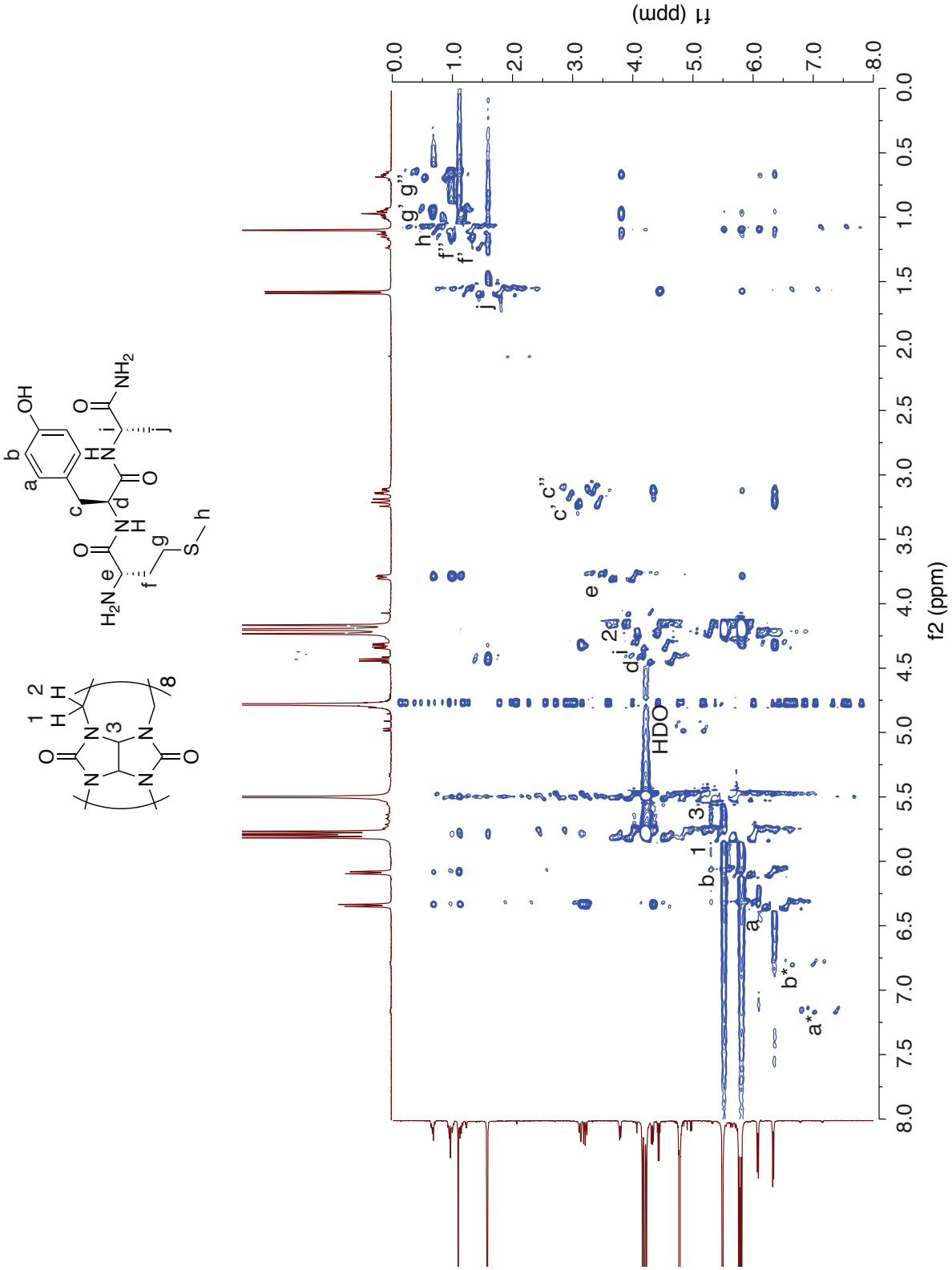


Figure S60. 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Met-Tyr-Ala at 25 °C in D_2O. Crosspeaks corresponding to unbound peptide are marked by an asterisk (*). Apostrophes indicate geminal separation.

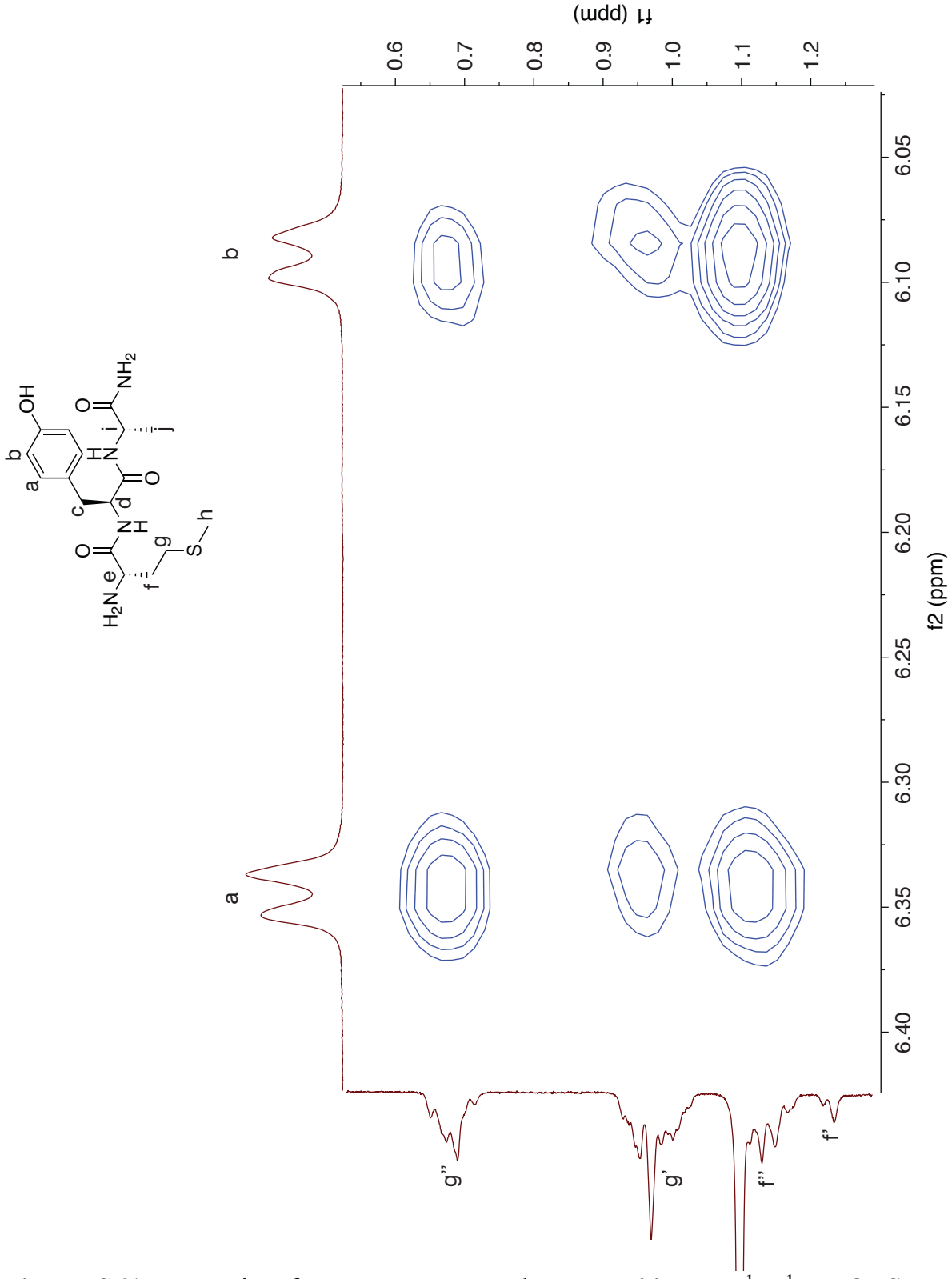


Figure S61. Zoom-in of Met-Tyr crosspeaks on a 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Met-Tyr-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

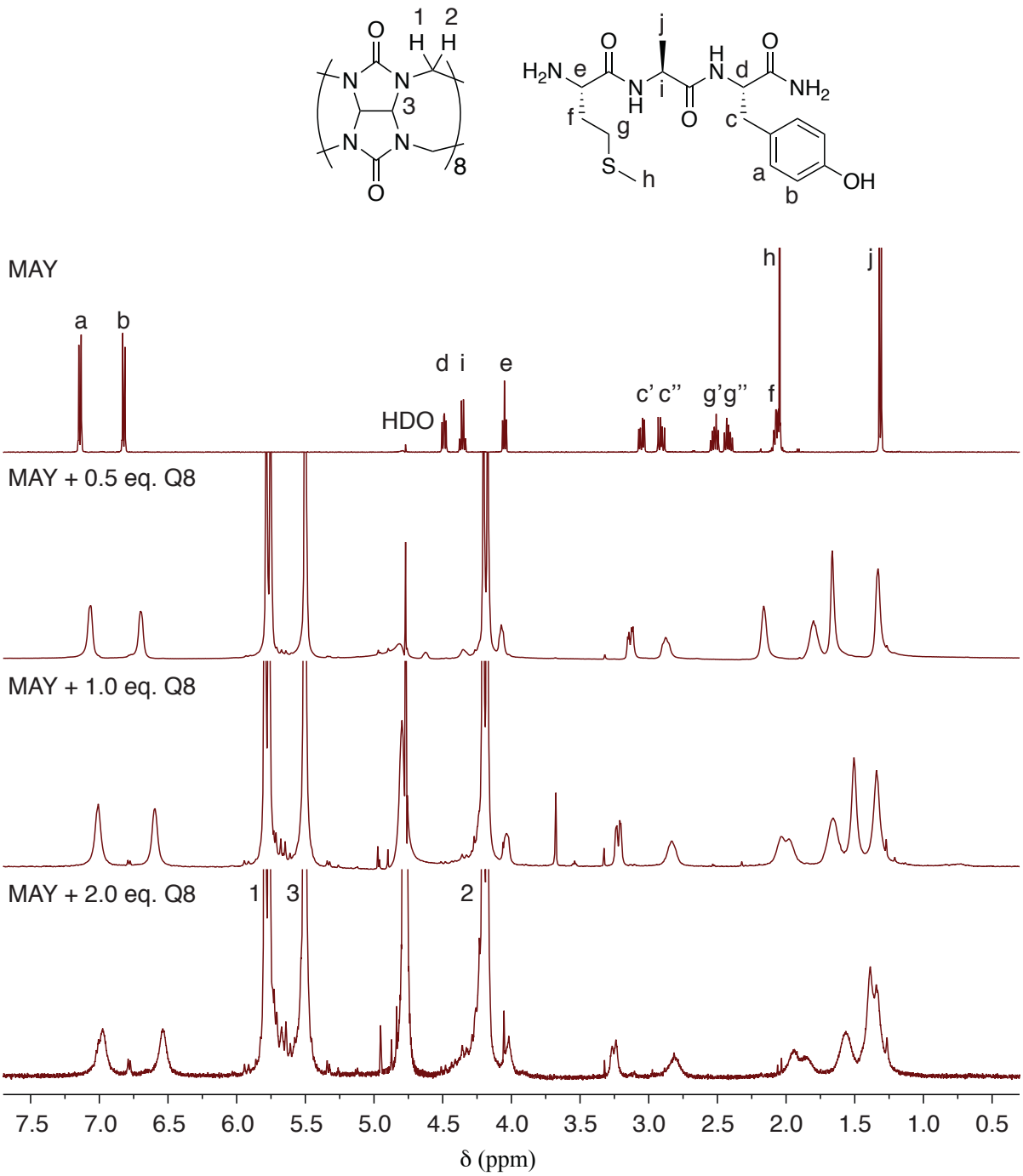


Figure S62. 500 MHz ^{1}H NMR spectra of Met-Ala-Tyr with 0 eq. Q8, 0.5 eq. Q8, 1.0 eq. Q8, and 2.0 eq. Q8 at 25 °C in D_2O.

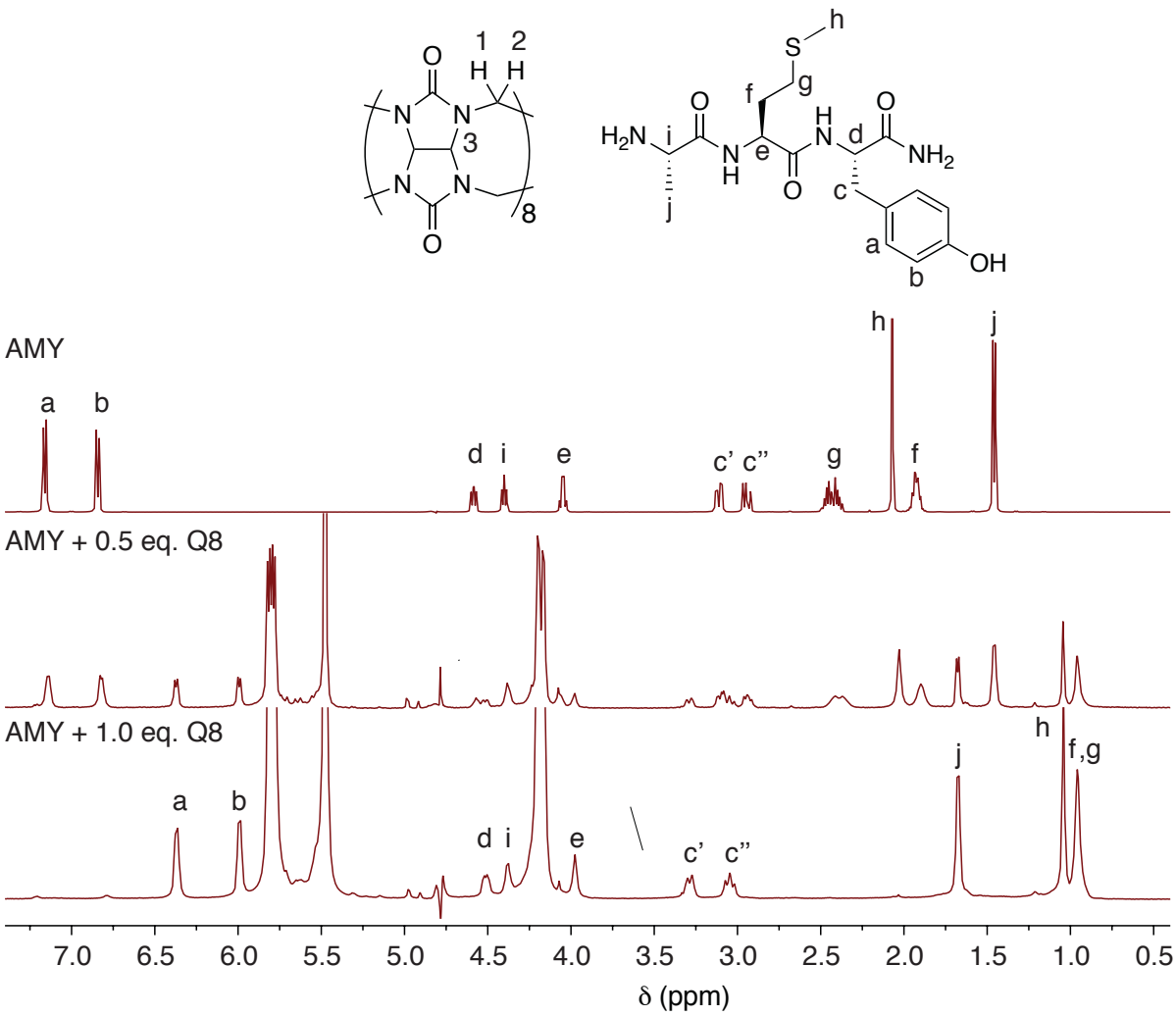


Figure S63. 500 MHz ^{1}H NMR spectra of Ala-Met-Tyr with 0 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

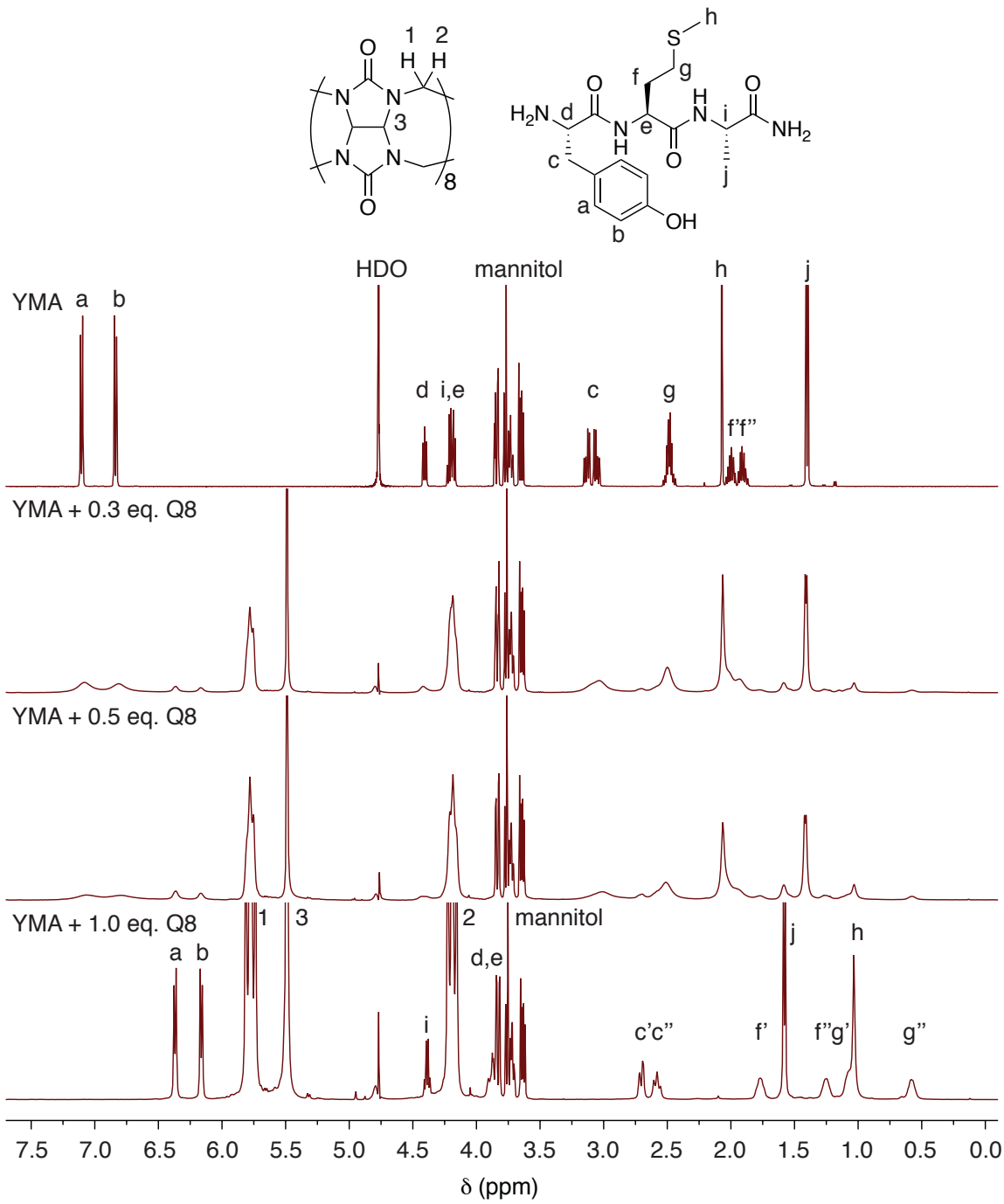


Figure S64. 500 MHz 1H NMR spectra of Tyr-Met-Ala with 0 eq. Q8, 0.3 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

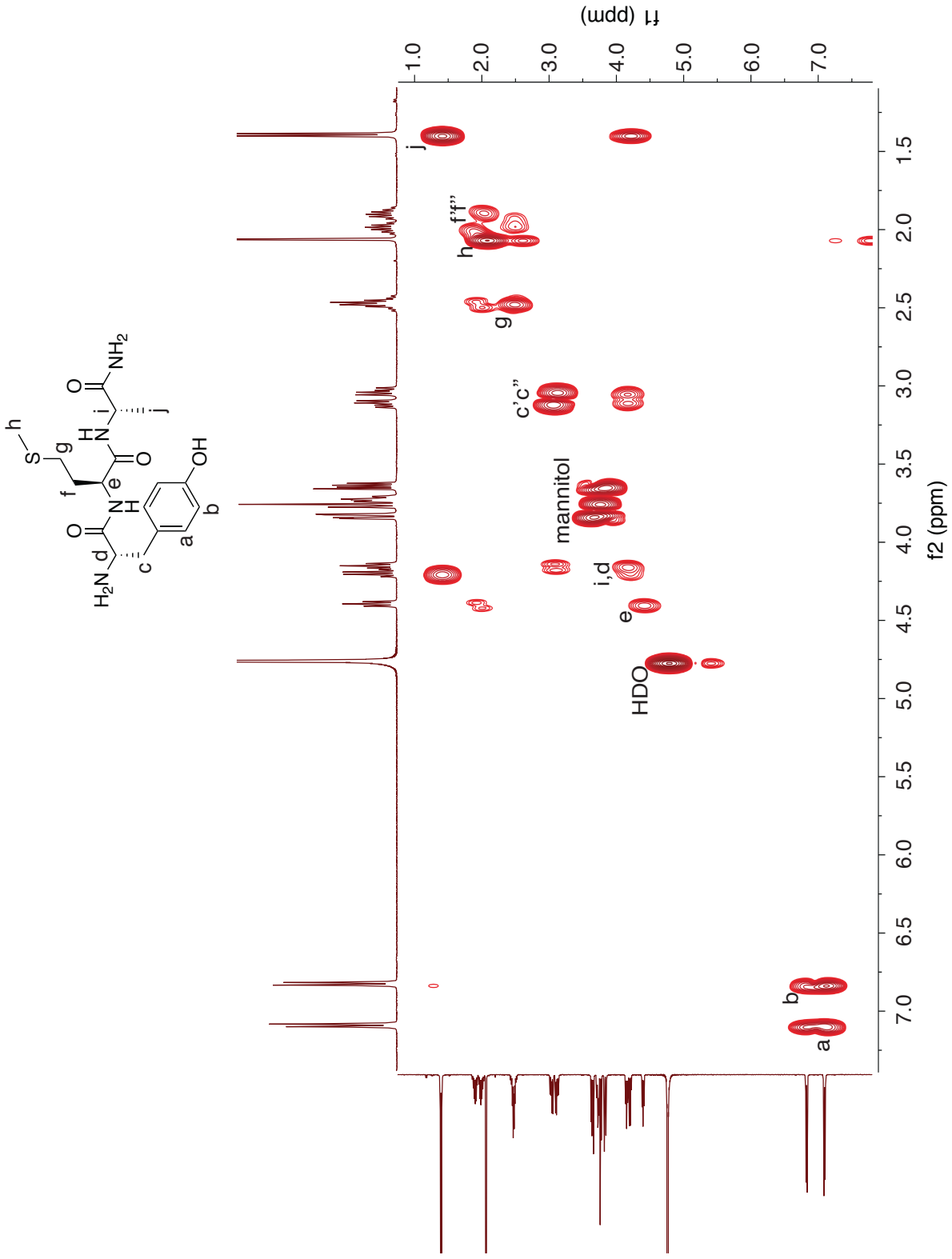


Figure S65. 500 MHz ^{1}H-^{1}H COSY spectrum of Tyr-Met-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

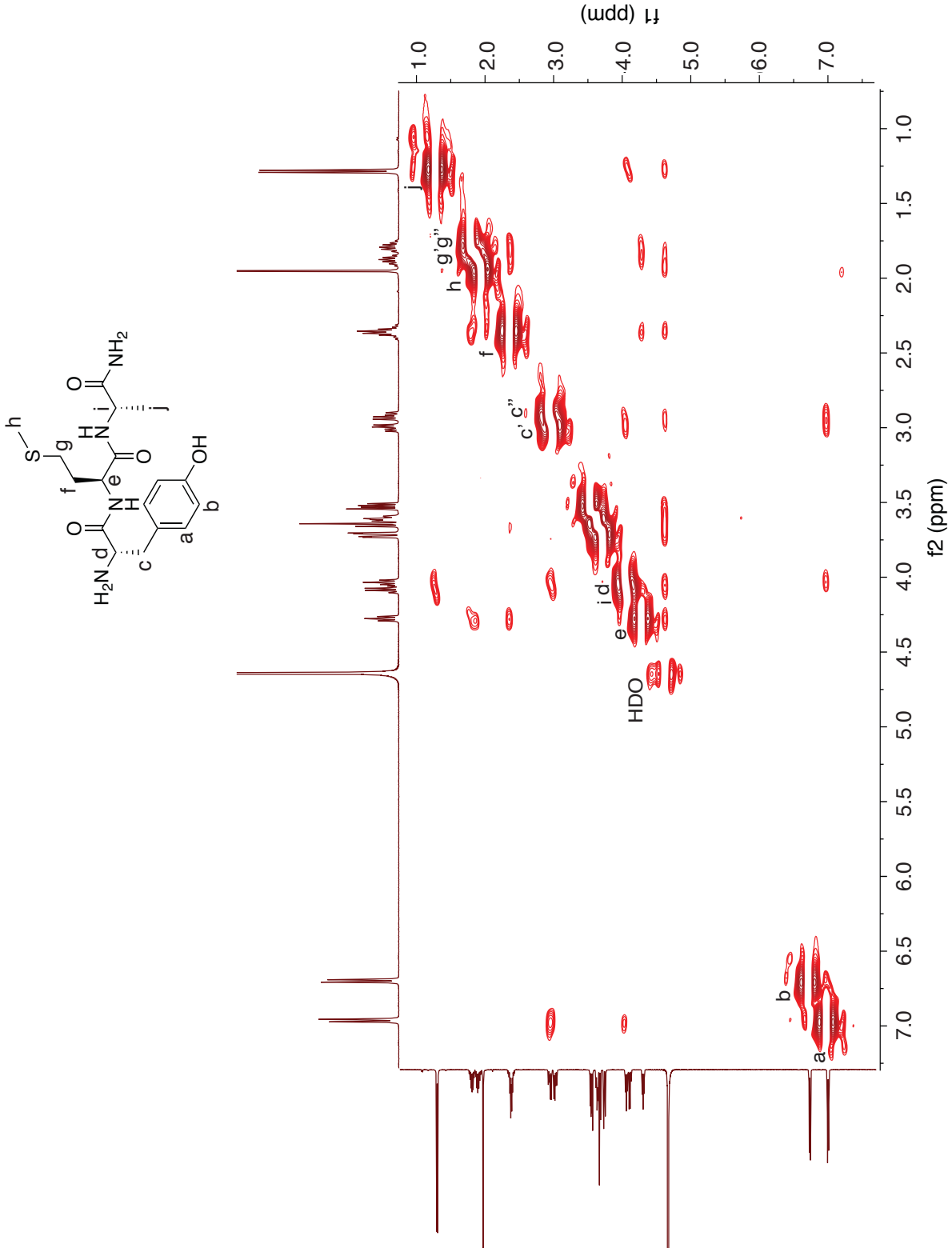


Figure S66. 500 MHz ^{1}H-^{1}H NOESY spectrum of Tyr-Met-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

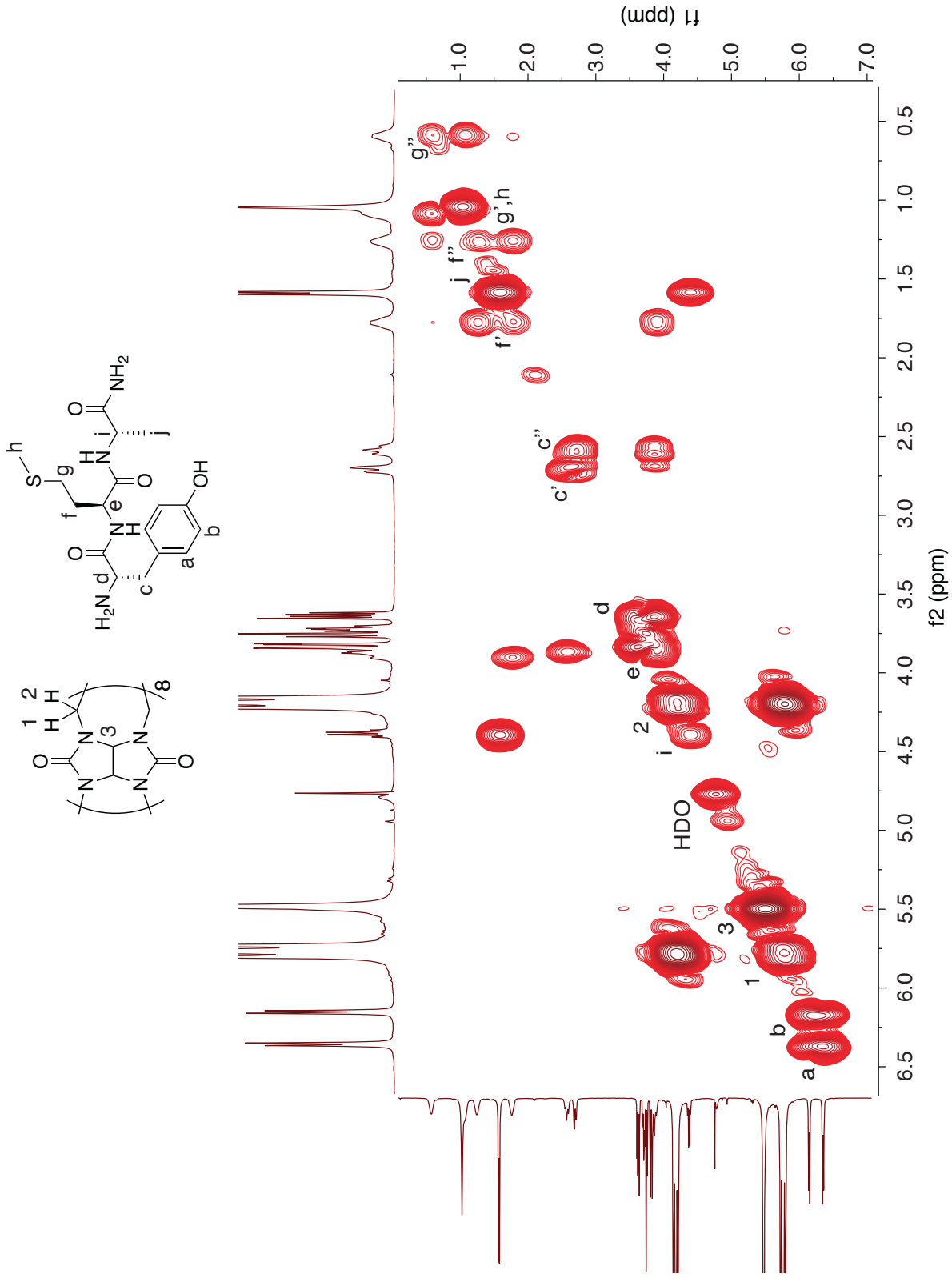


Figure S67. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Tyr-Met-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

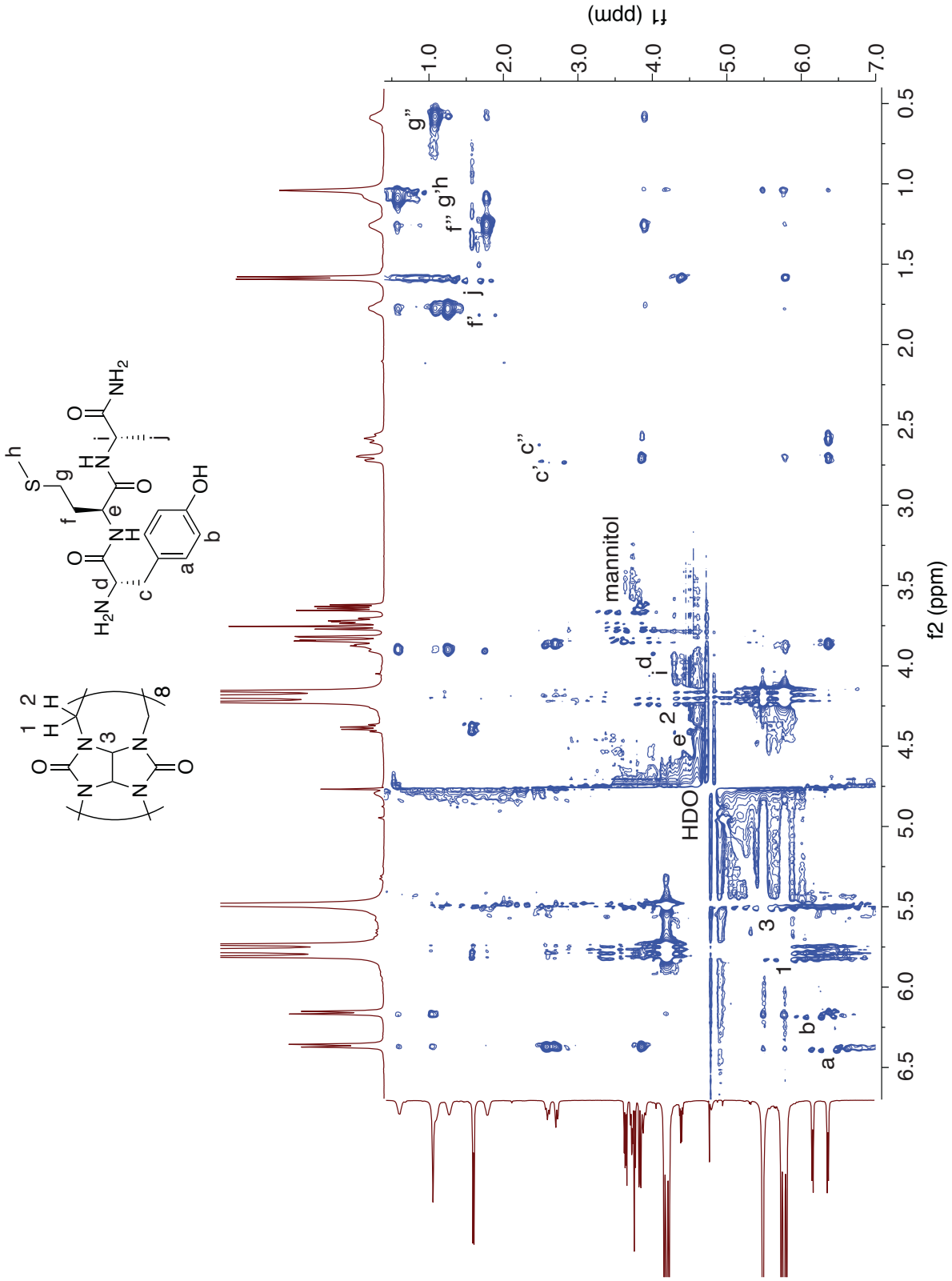


Figure S68. 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Tyr-Met-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

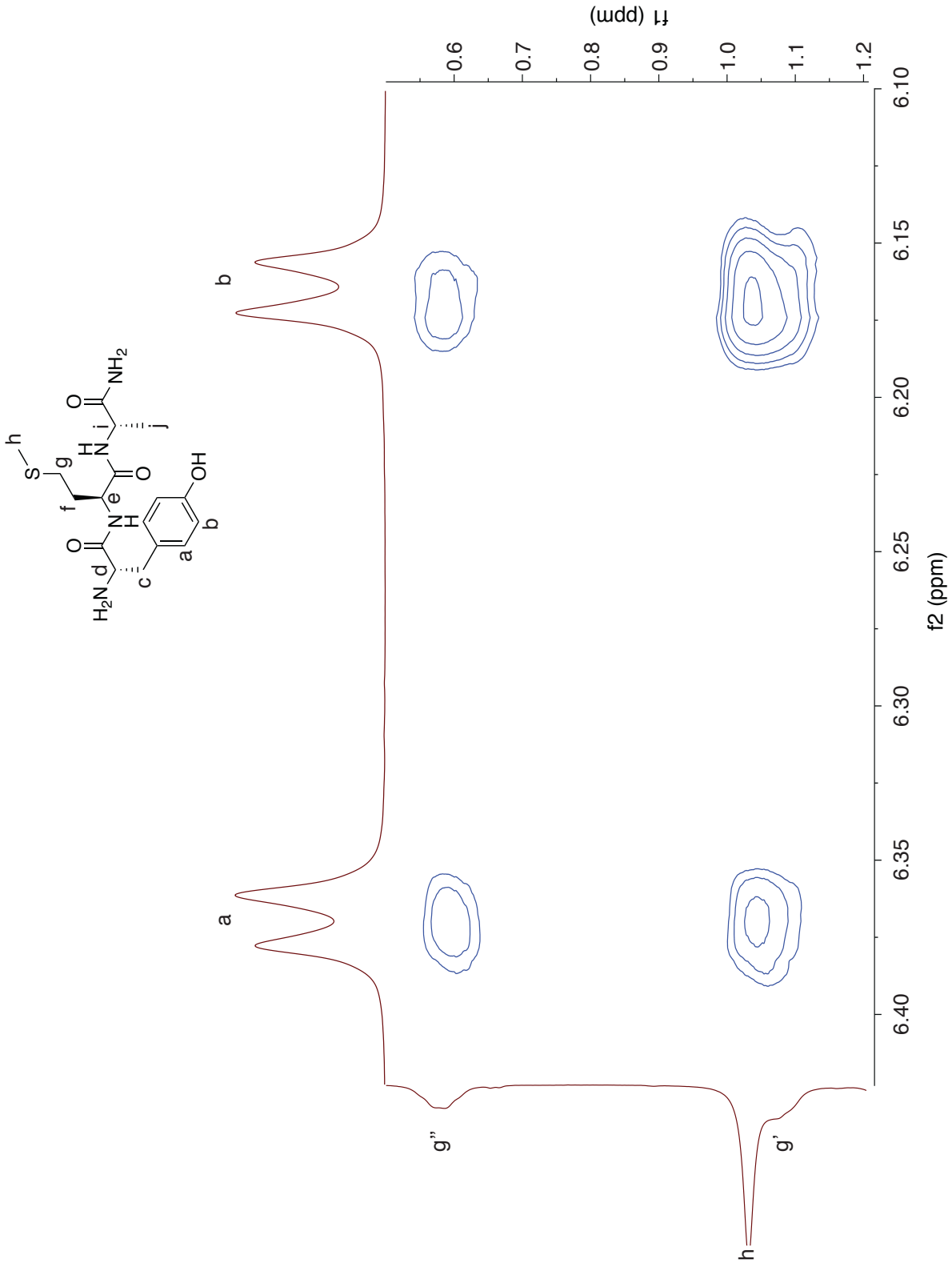


Figure S69. Zoom-in of Tyr-Met crosspeaks on a 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Tyr-Met-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

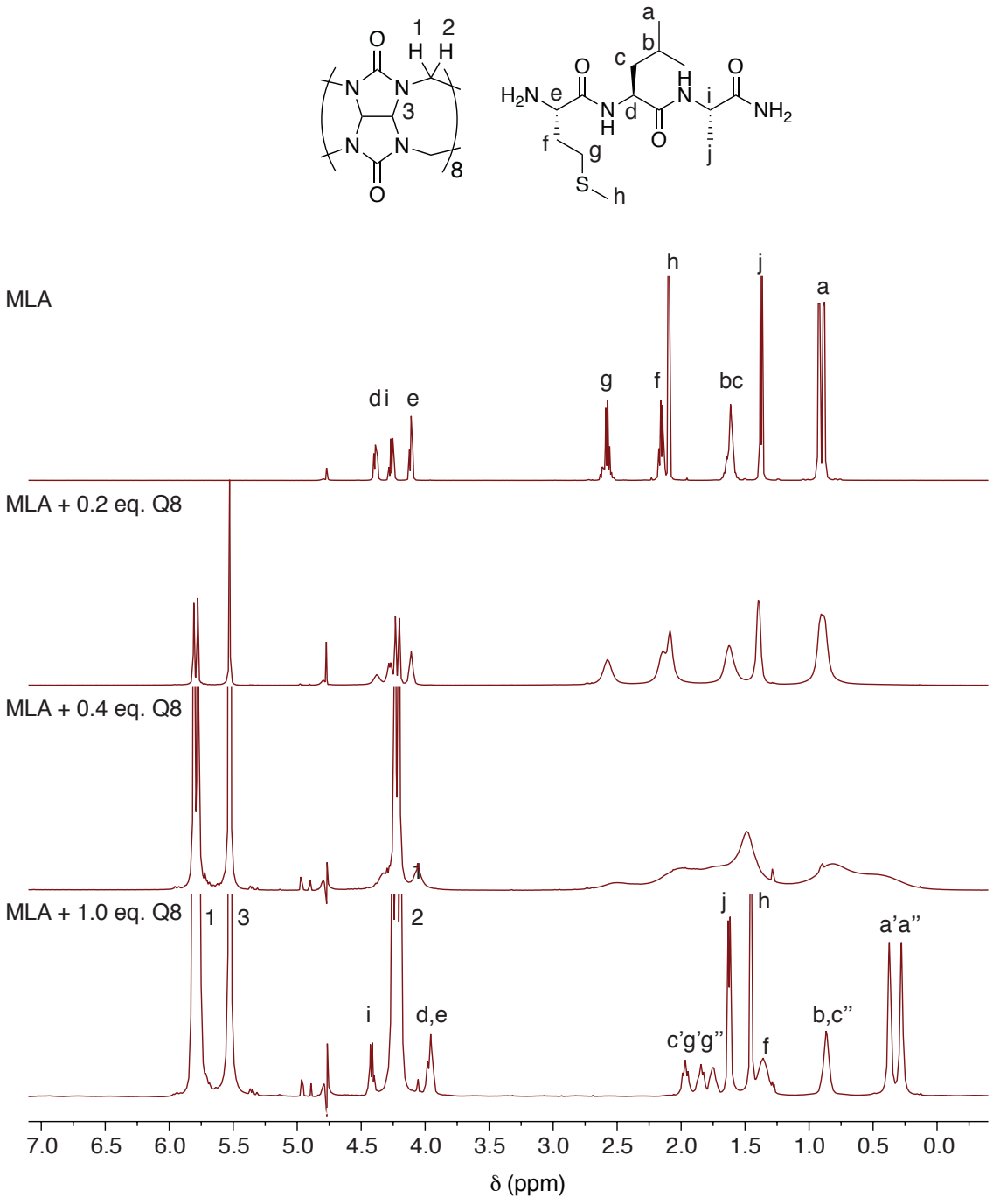


Figure S70. 500 MHz ^{1}H NMR spectra of Met-Leu-Ala with 0 eq. Q8, 0.2 eq. Q8, 0.4 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

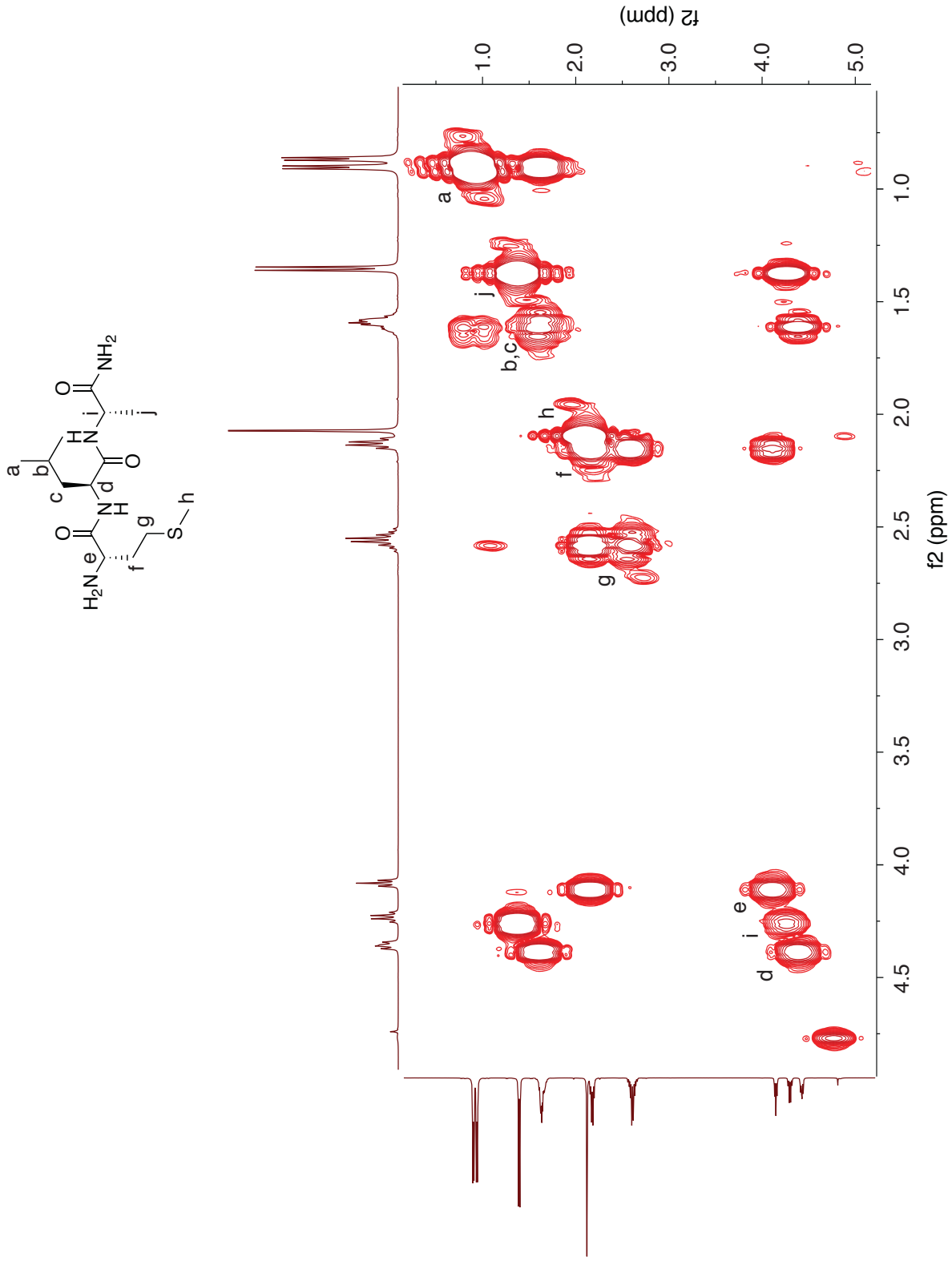


Figure S71. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Leu-Ala at 25 °C in D_2O.

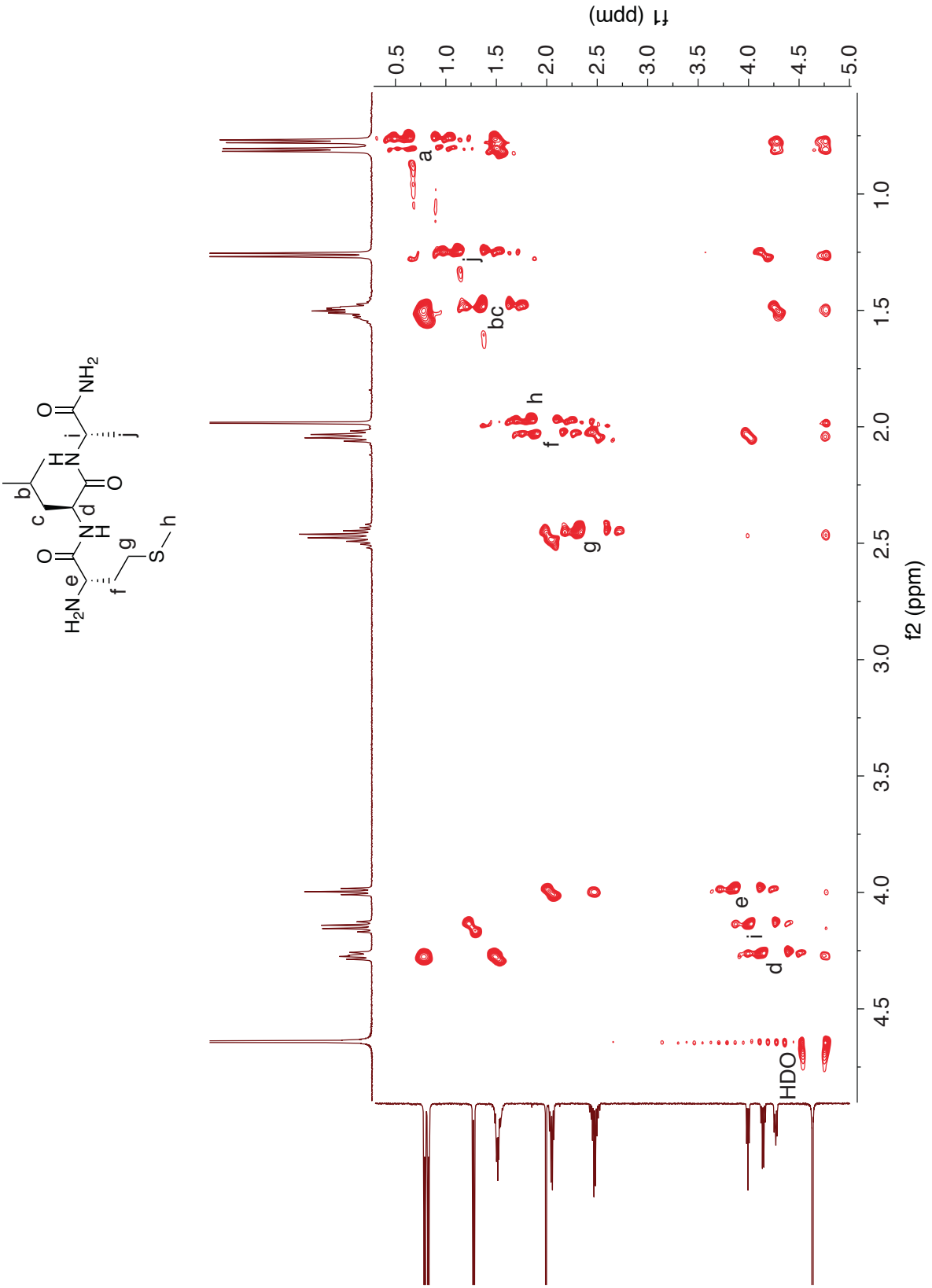


Figure S72. 500 MHz ^{1}H-^{1}H NOESY spectrum of Met-Leu-Ala at 25 °C in D_2O.

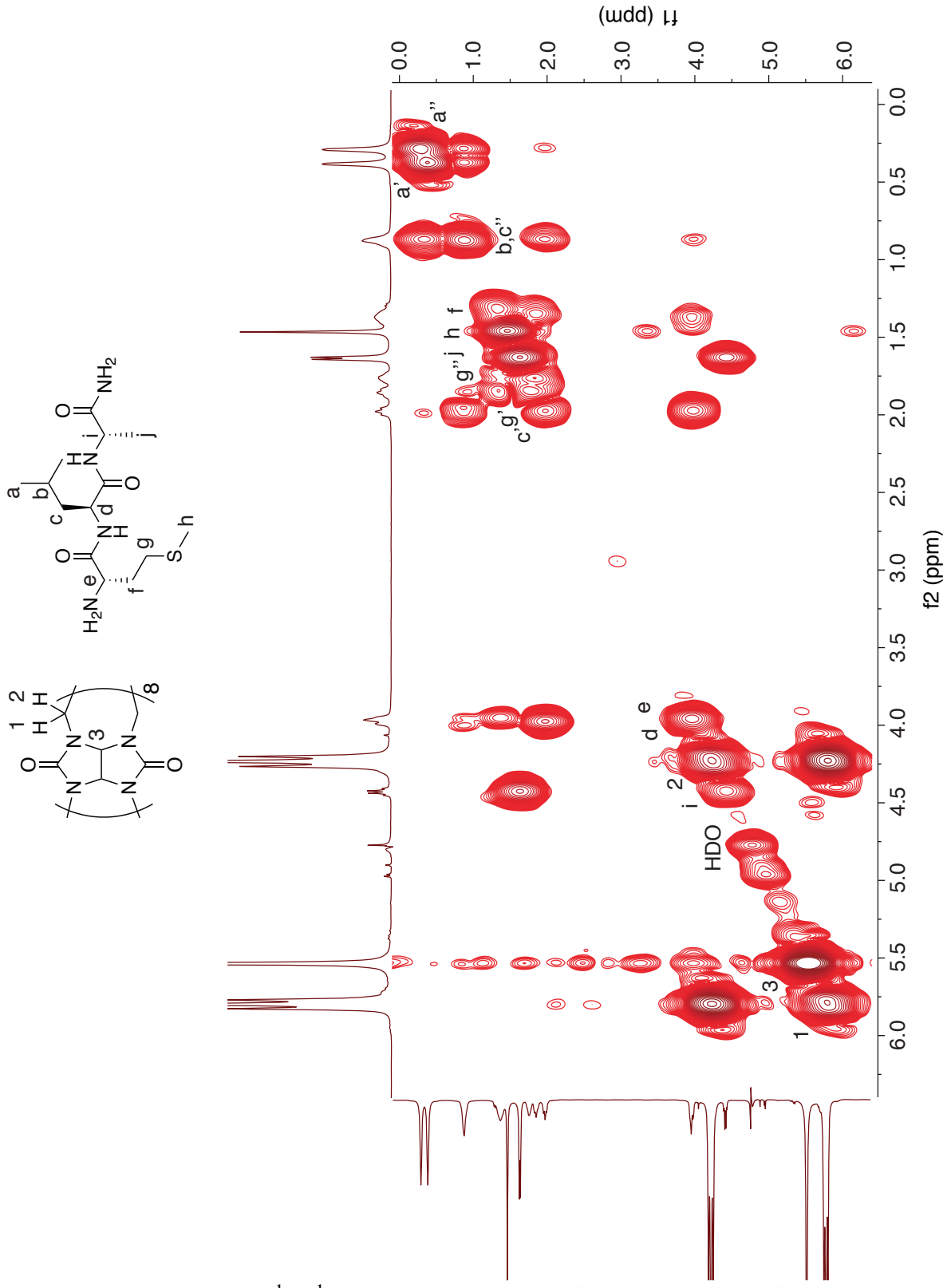


Figure S73. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Leu-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

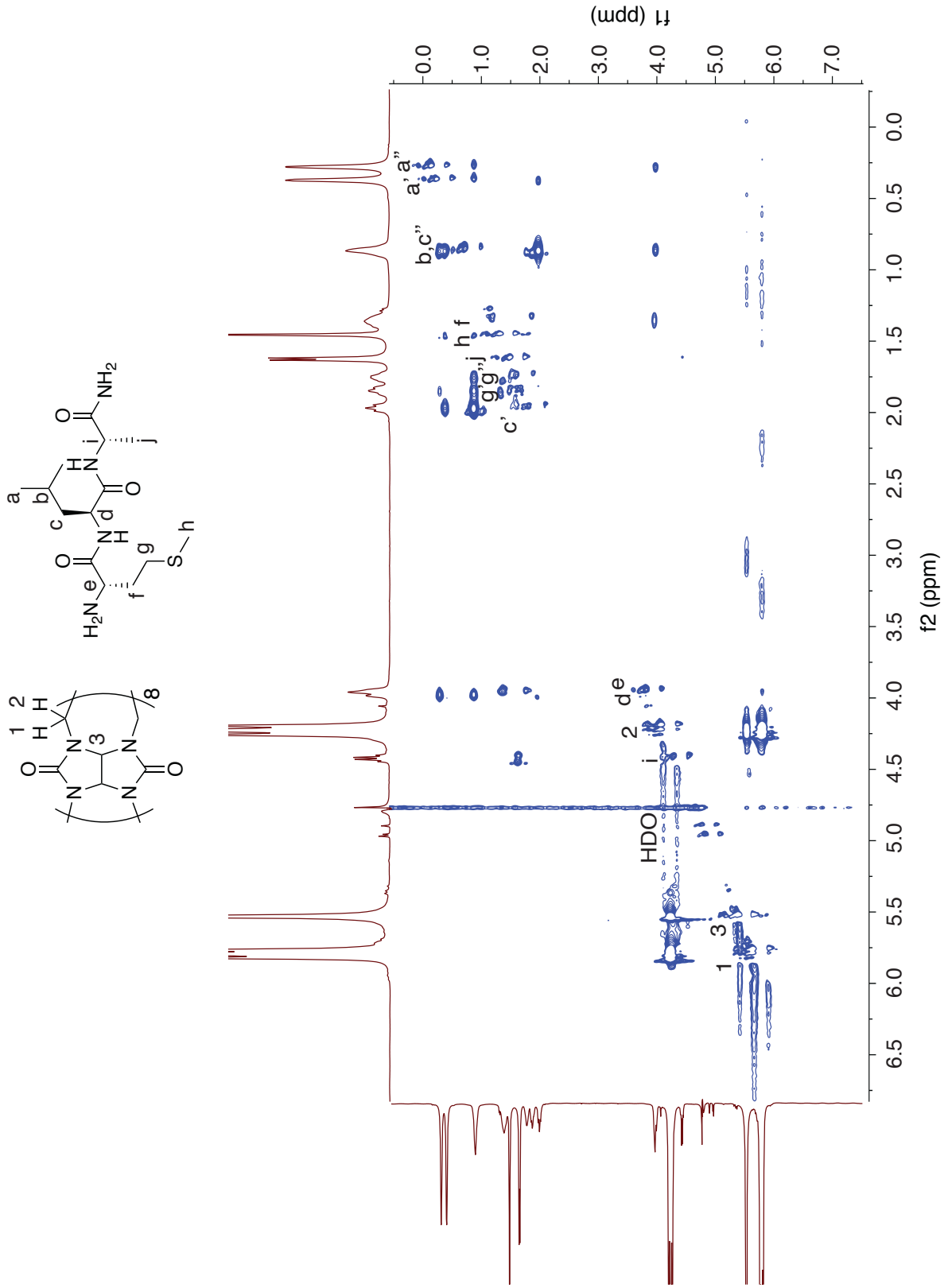


Figure S74. 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Met-Leu-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

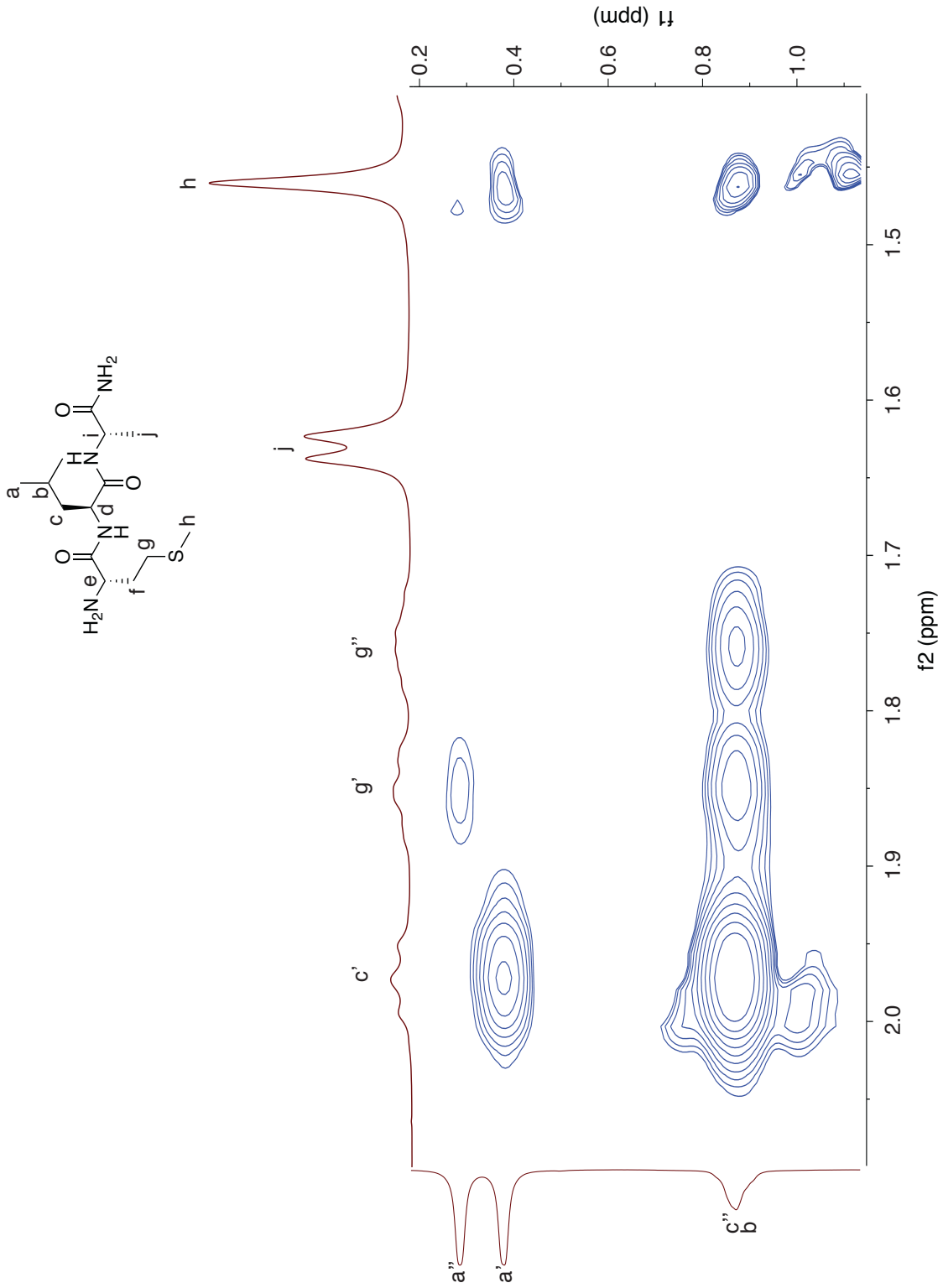


Figure S75. Zoom-in of Met-Leu crosspeaks on a 500 MHz ^{1}H-^{1}H ROESY spectrum of a 1:1 mixture of Q8 and Met-Leu-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

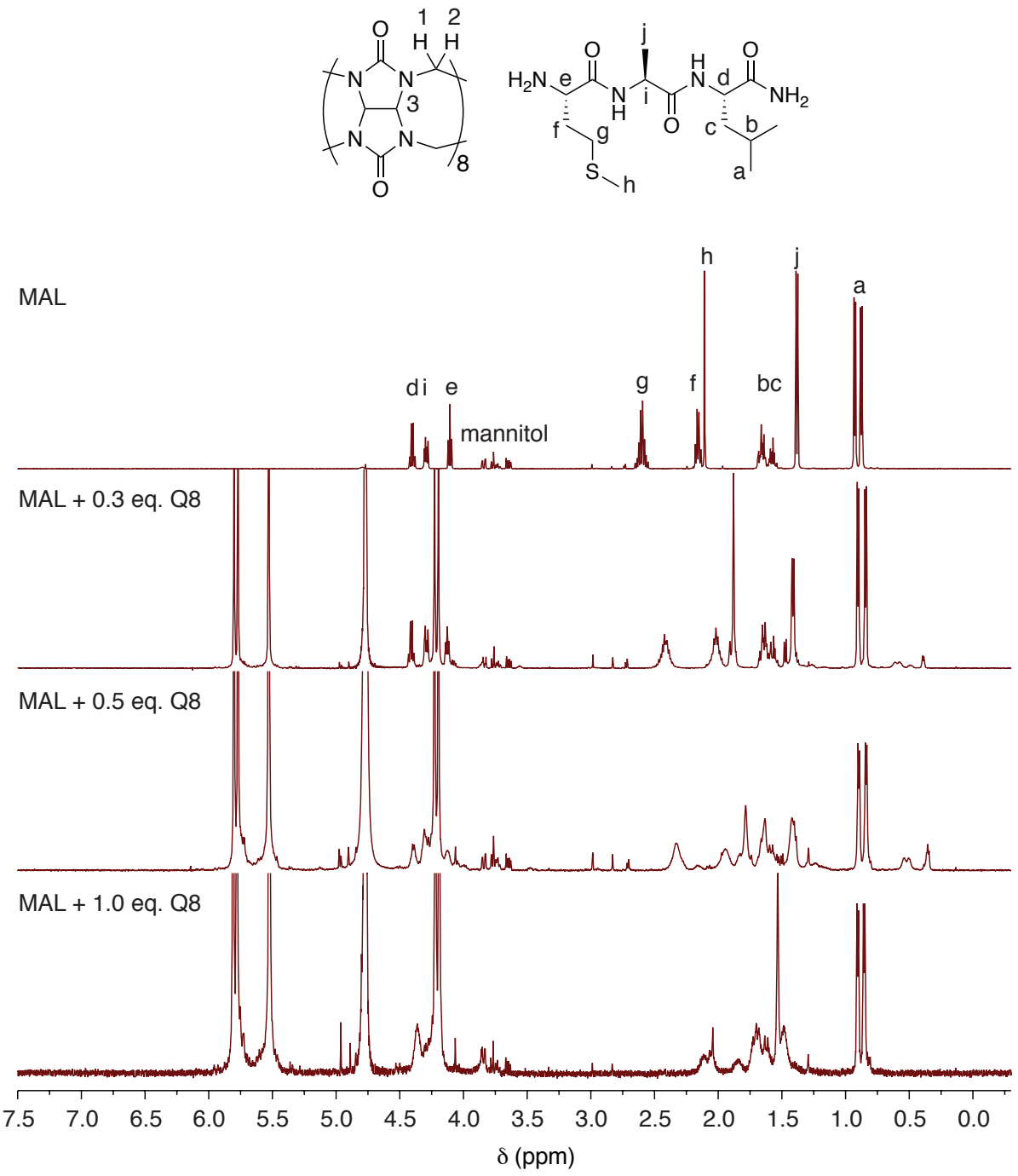


Figure S76. 500 MHz ^{1}H NMR spectra of Met-Ala-Leu with 0 eq. Q8, 0.3 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O.

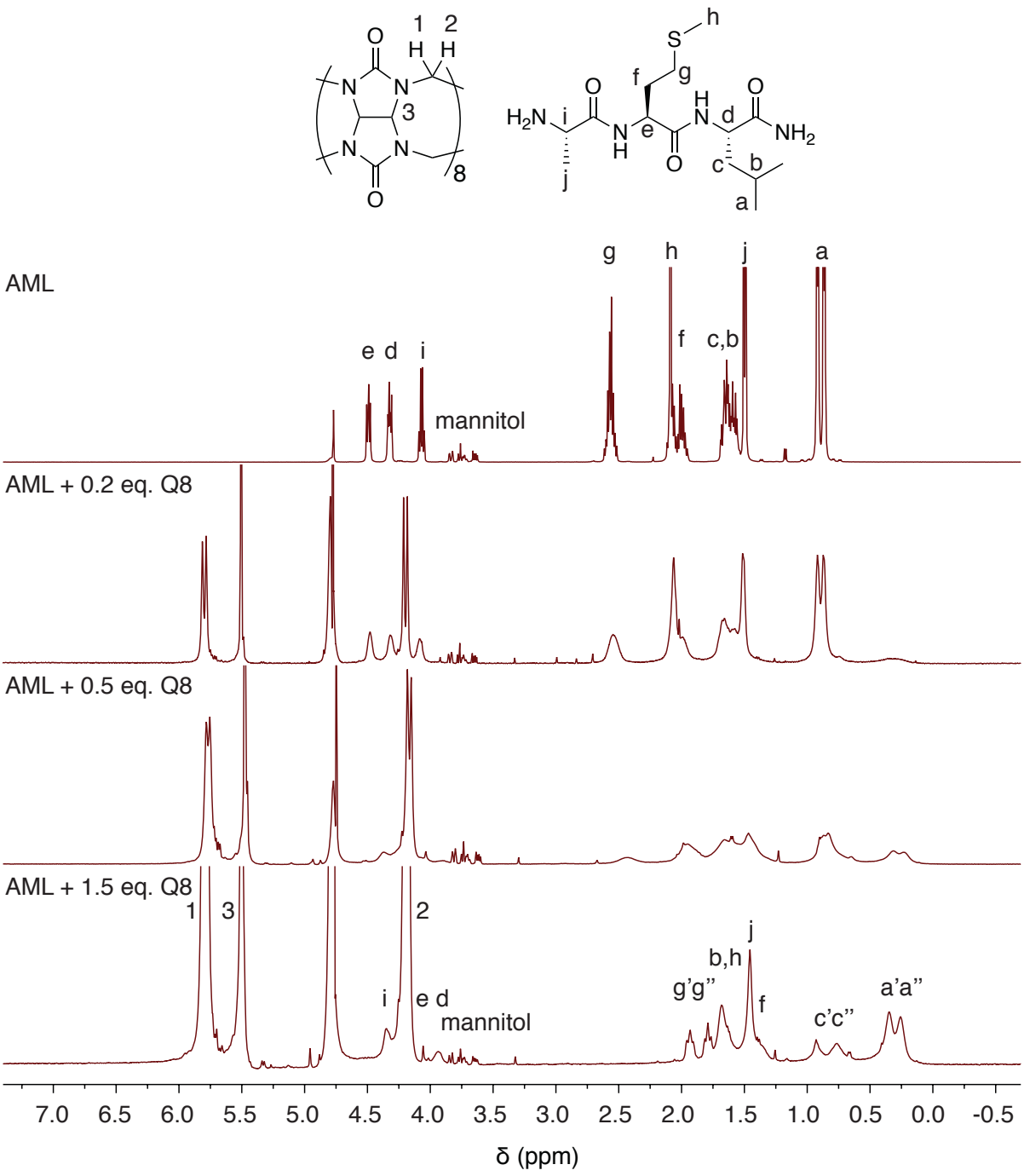


Figure S77. 500 MHz ^{1}H NMR spectra of Ala-Met-Leu with 0 eq. Q8, 0.2 eq. Q8, 0.5 eq. Q8, and 1.5 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation. Proton assignments at 1.5 eq. Q8 are based on a ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Leu-Ala.

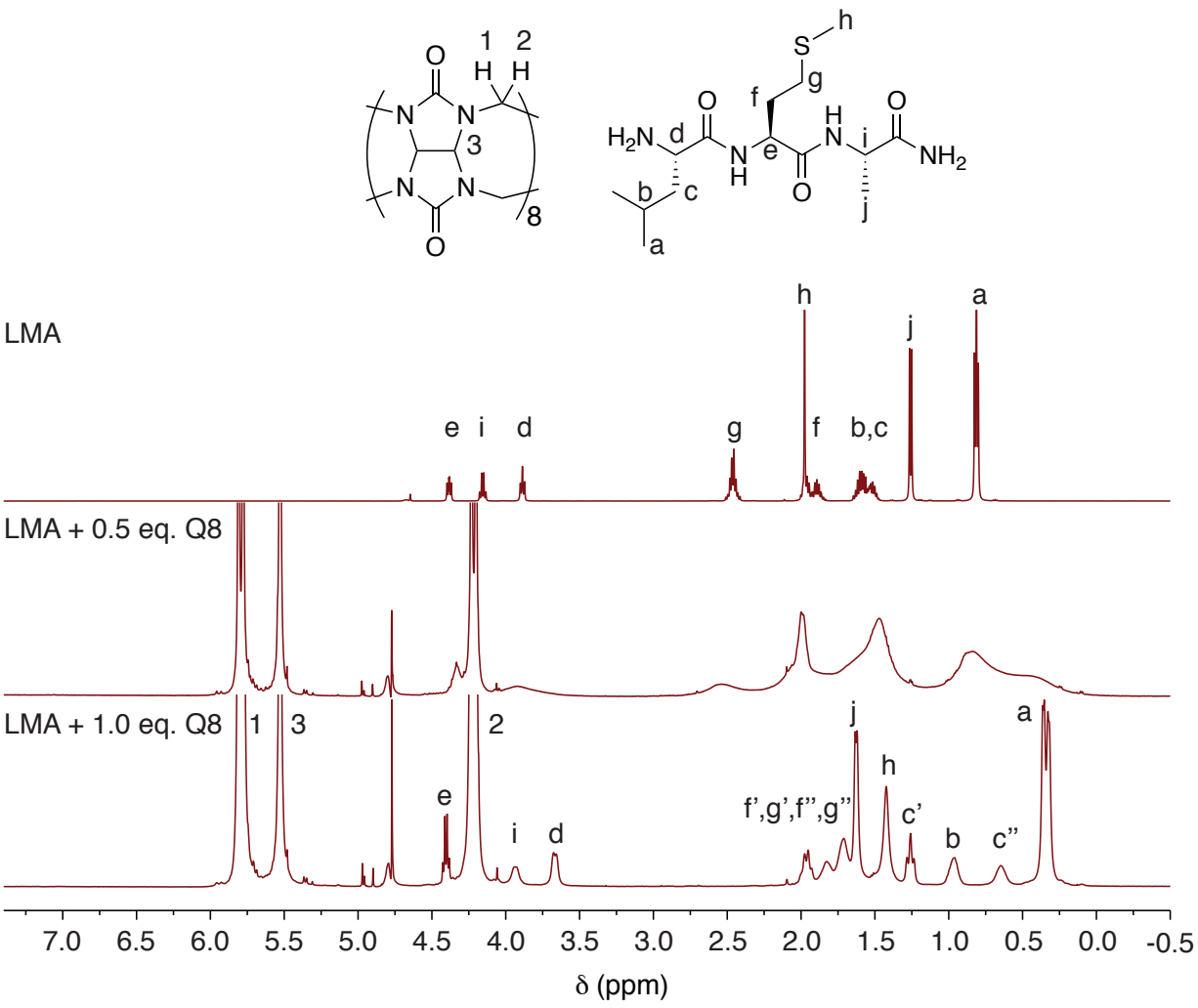


Figure S78. 500 MHz ^{1}H NMR spectra of Leu-Met-Ala with 0 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

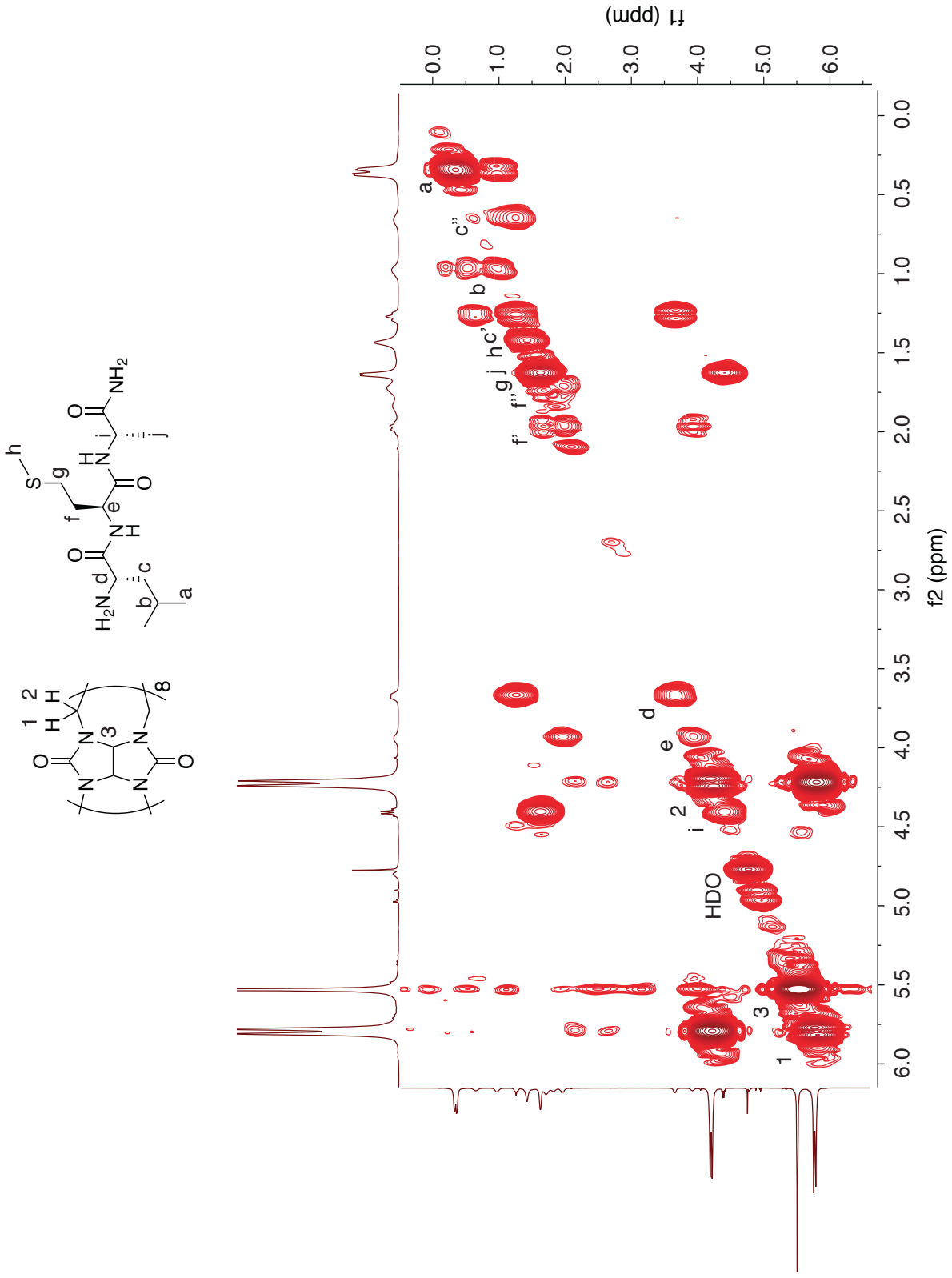


Figure S79. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Leu-Met-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

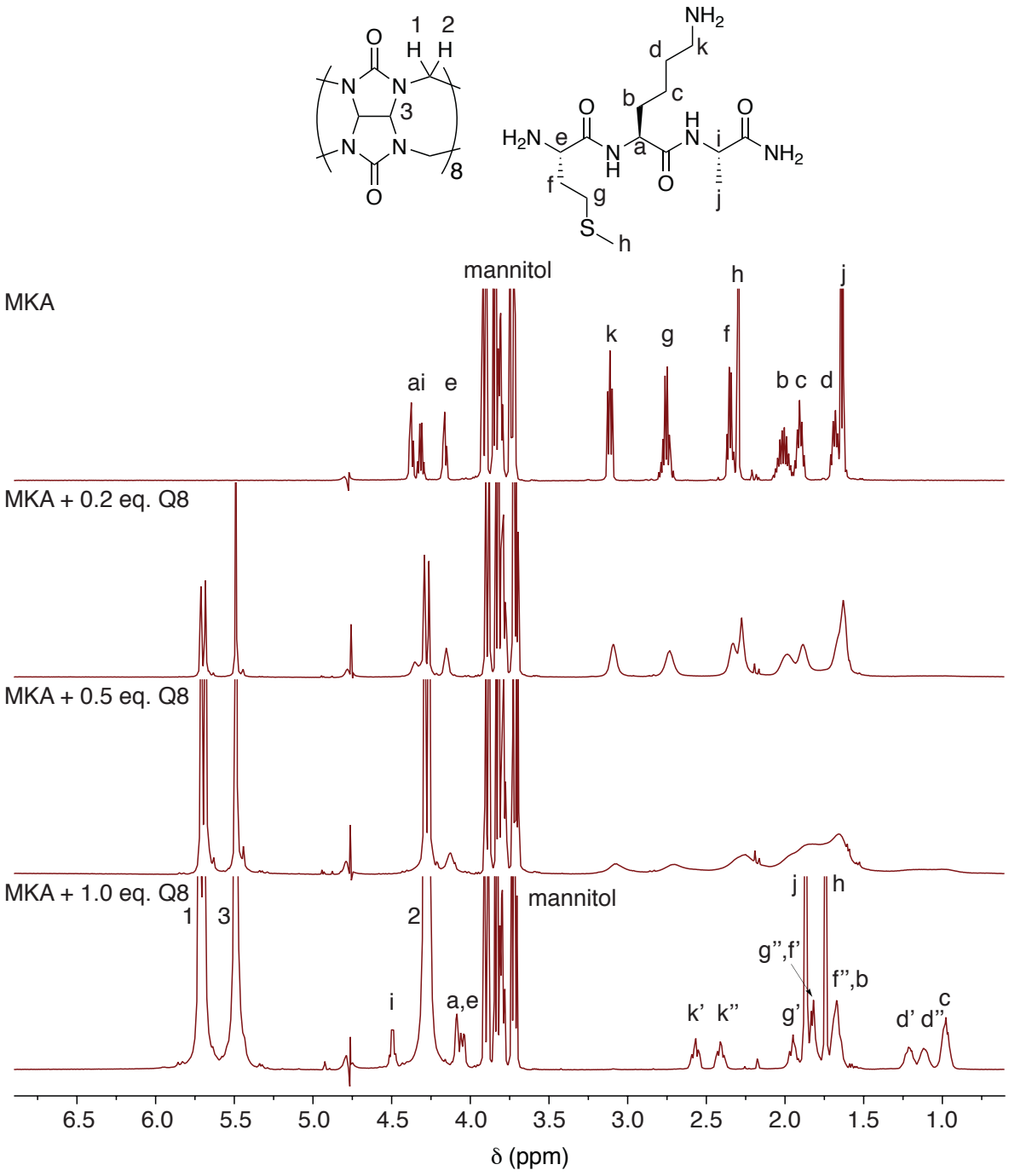


Figure S80. 500 MHz ^{1}H NMR spectra of Met-Lys-Ala with 0 eq. Q8, 0.2 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

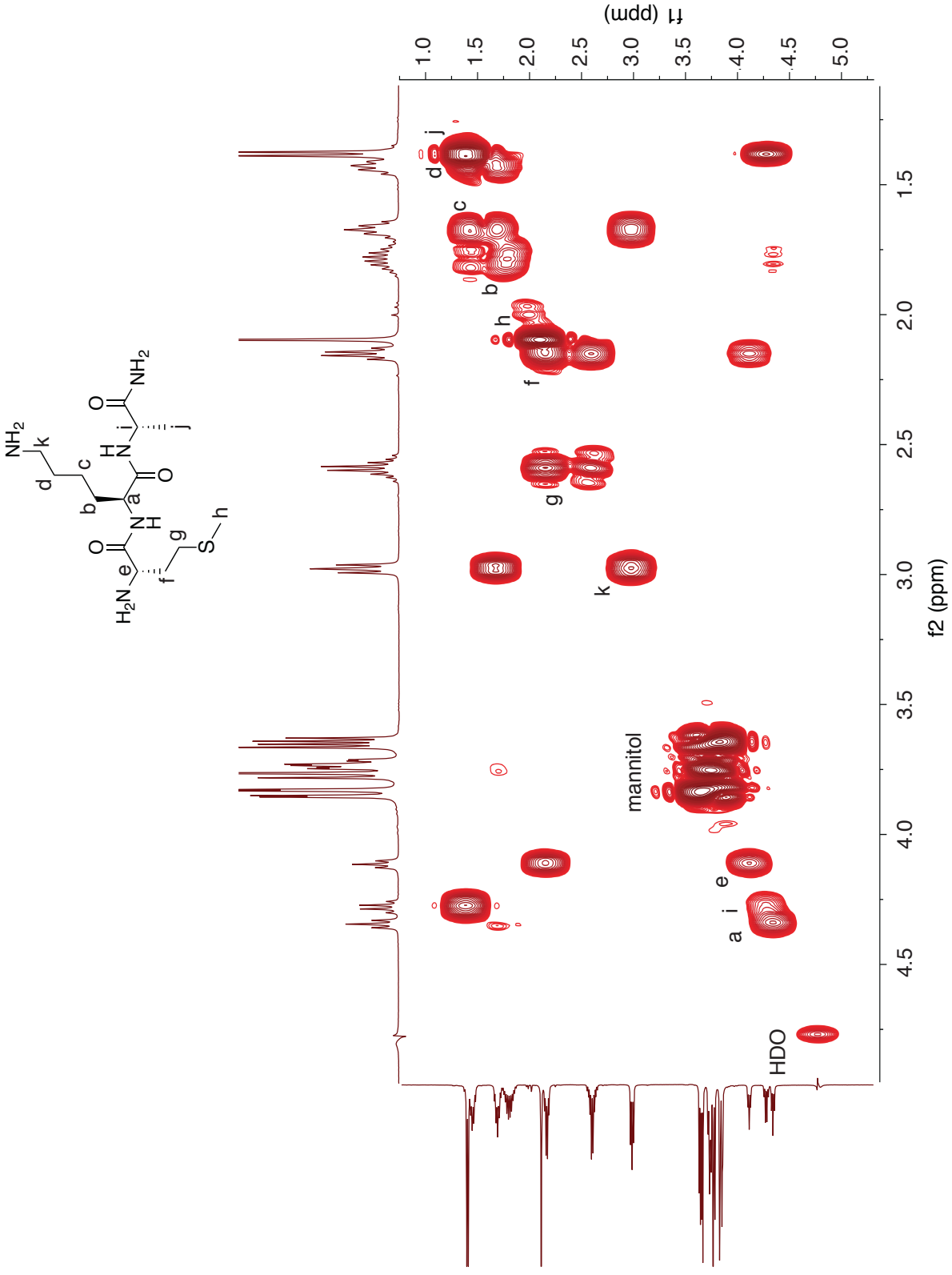


Figure S81. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Lys-Ala at 25 °C in D_2O.

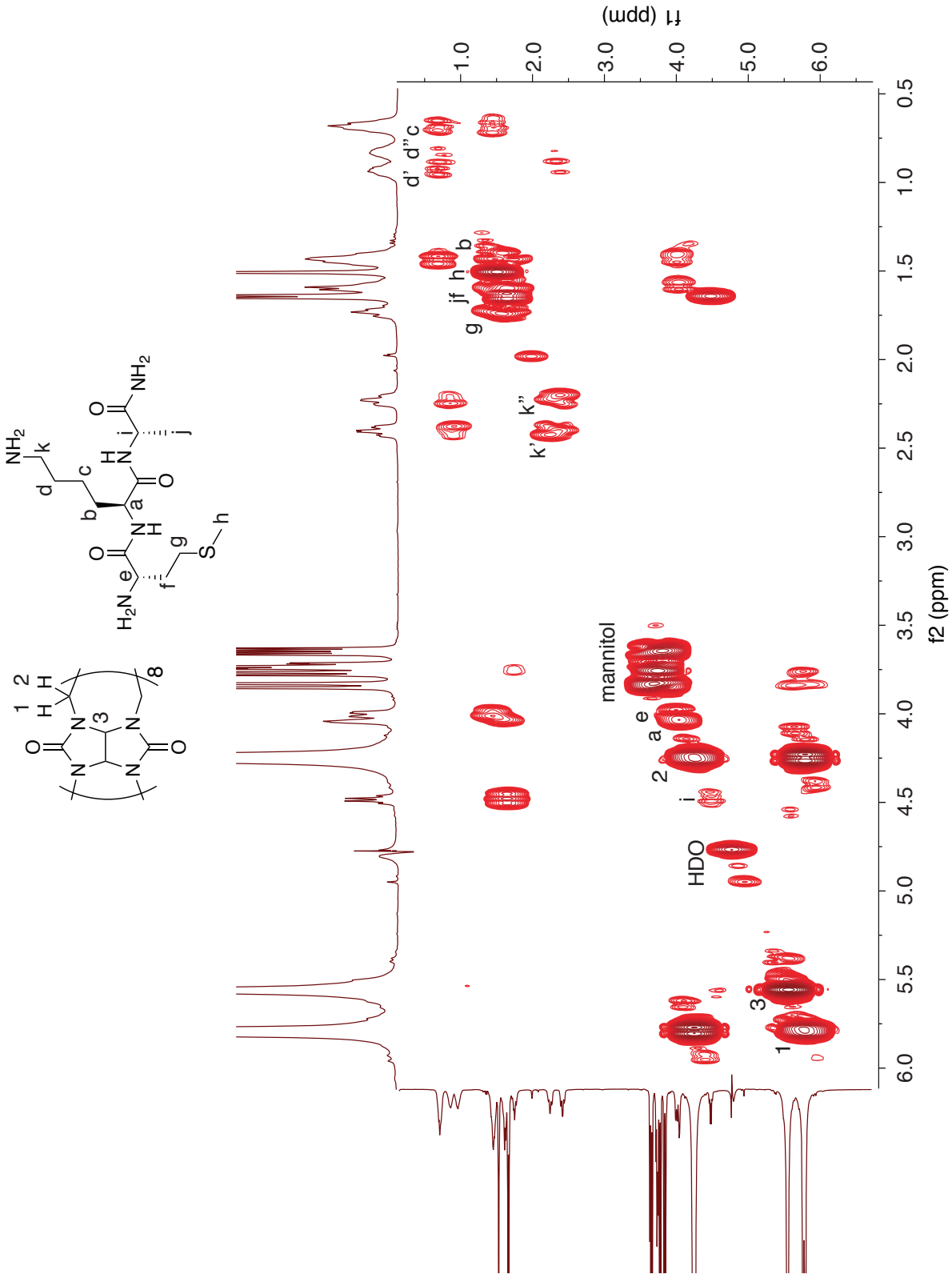


Figure S82. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Lys-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

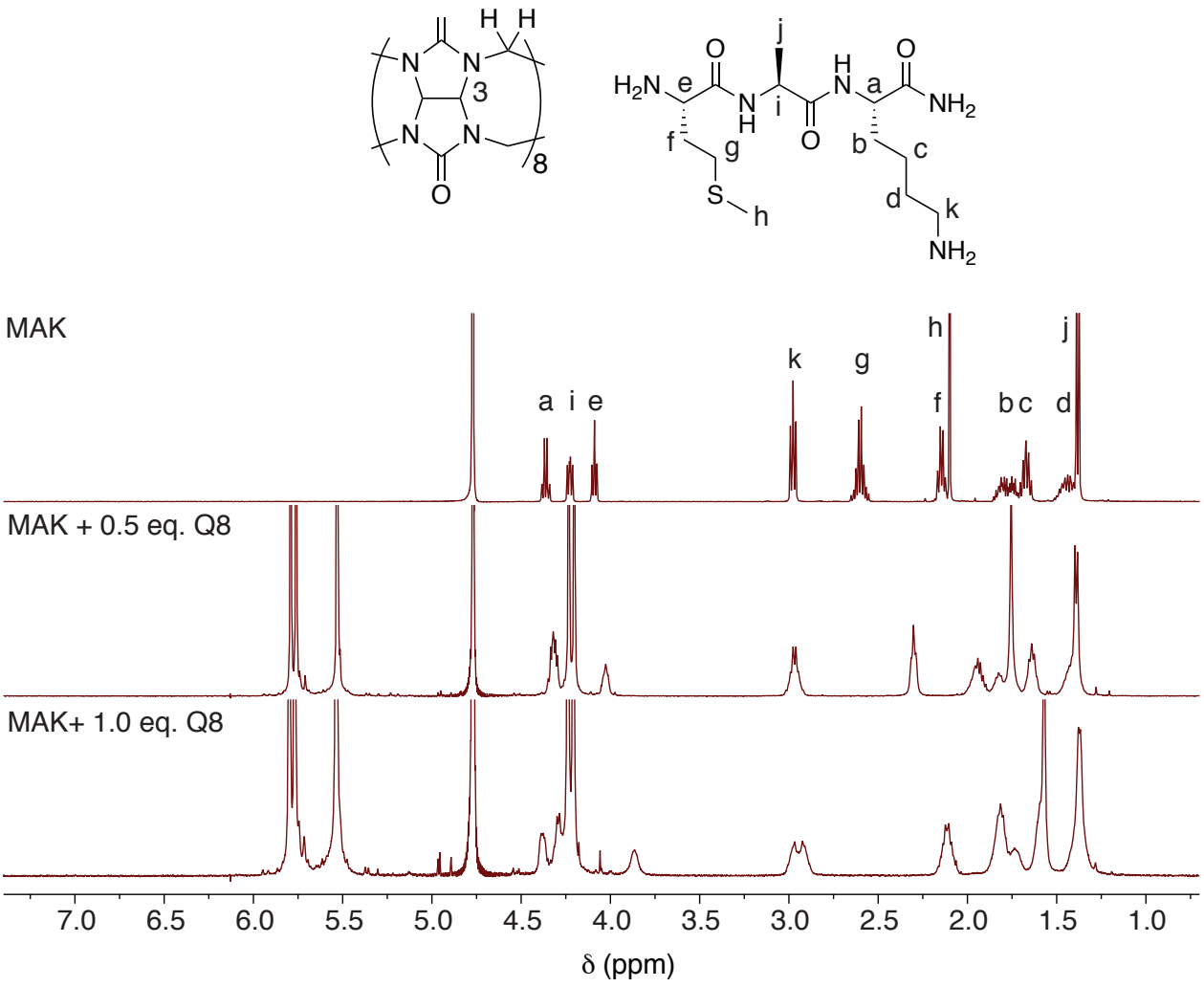


Figure S83. 500 MHz ^{1}H NMR spectra of Met-Ala-Lys with 0 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Solubility problems occurred after 0.5 eq. of Q8 were added.

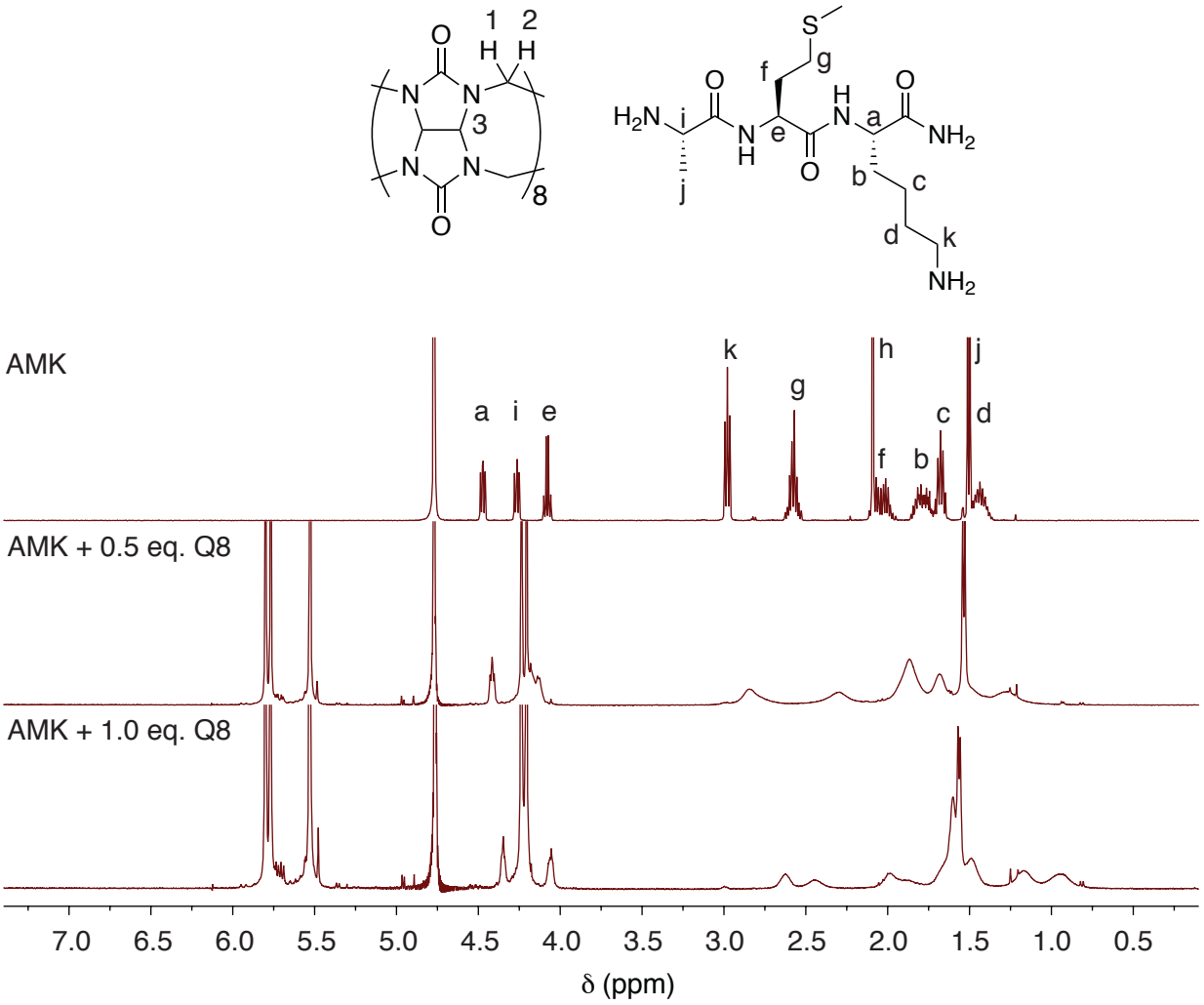


Figure S84. 500 MHz ^{1}H NMR spectra of Ala-Met-Lys with 0 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Solubility problems occurred after 0.5 eq. of Q8 were added.

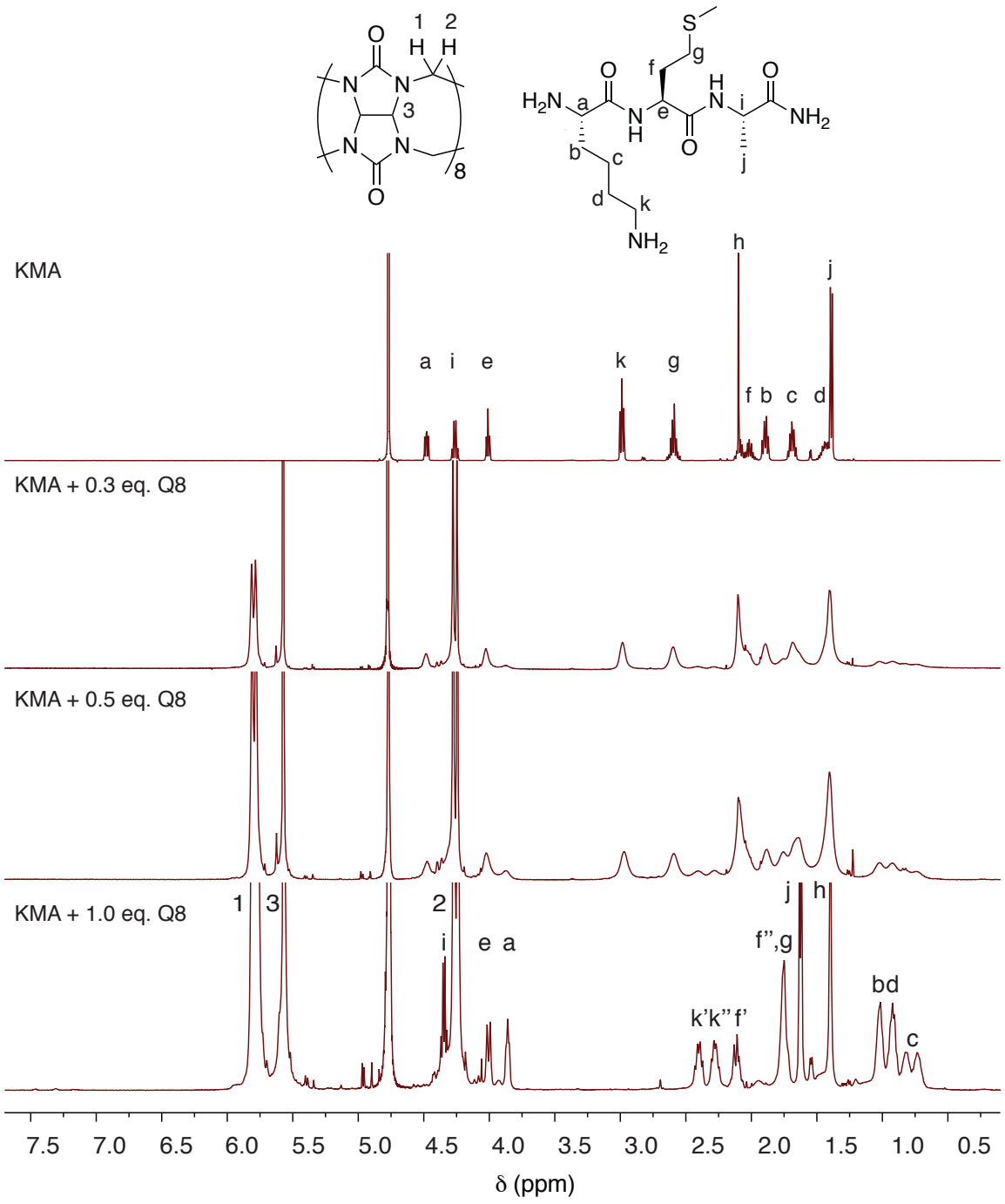


Figure S85. 500 MHz ^{1}H NMR spectra of Lys-Met-Ala with 0 eq. Q8, 0.3 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O.

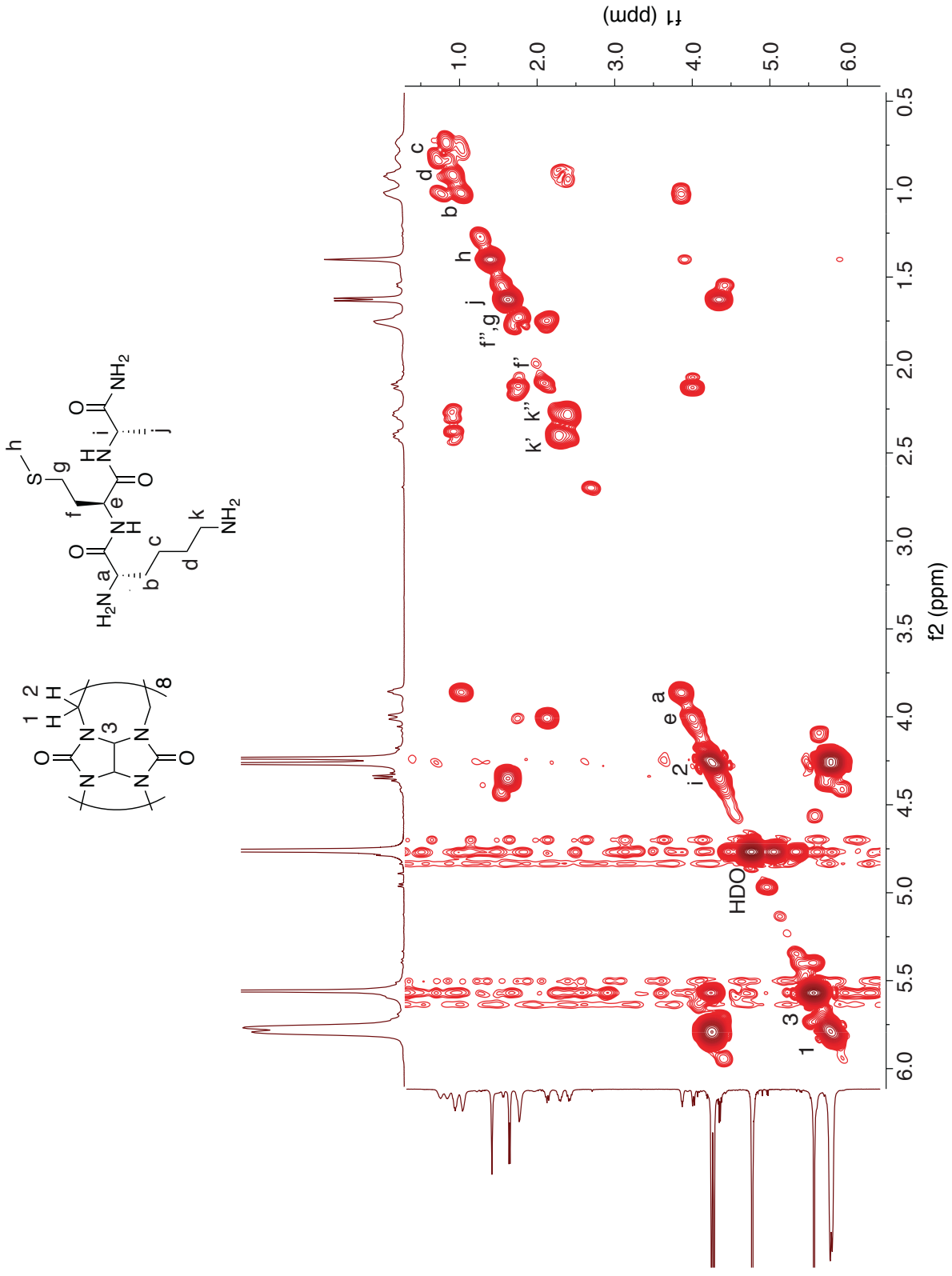


Figure S86. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1.1 mixture of Q8 and Lys-Met-Ala at 25 °C in D_2O. Apostrophes indicate geminal separation.

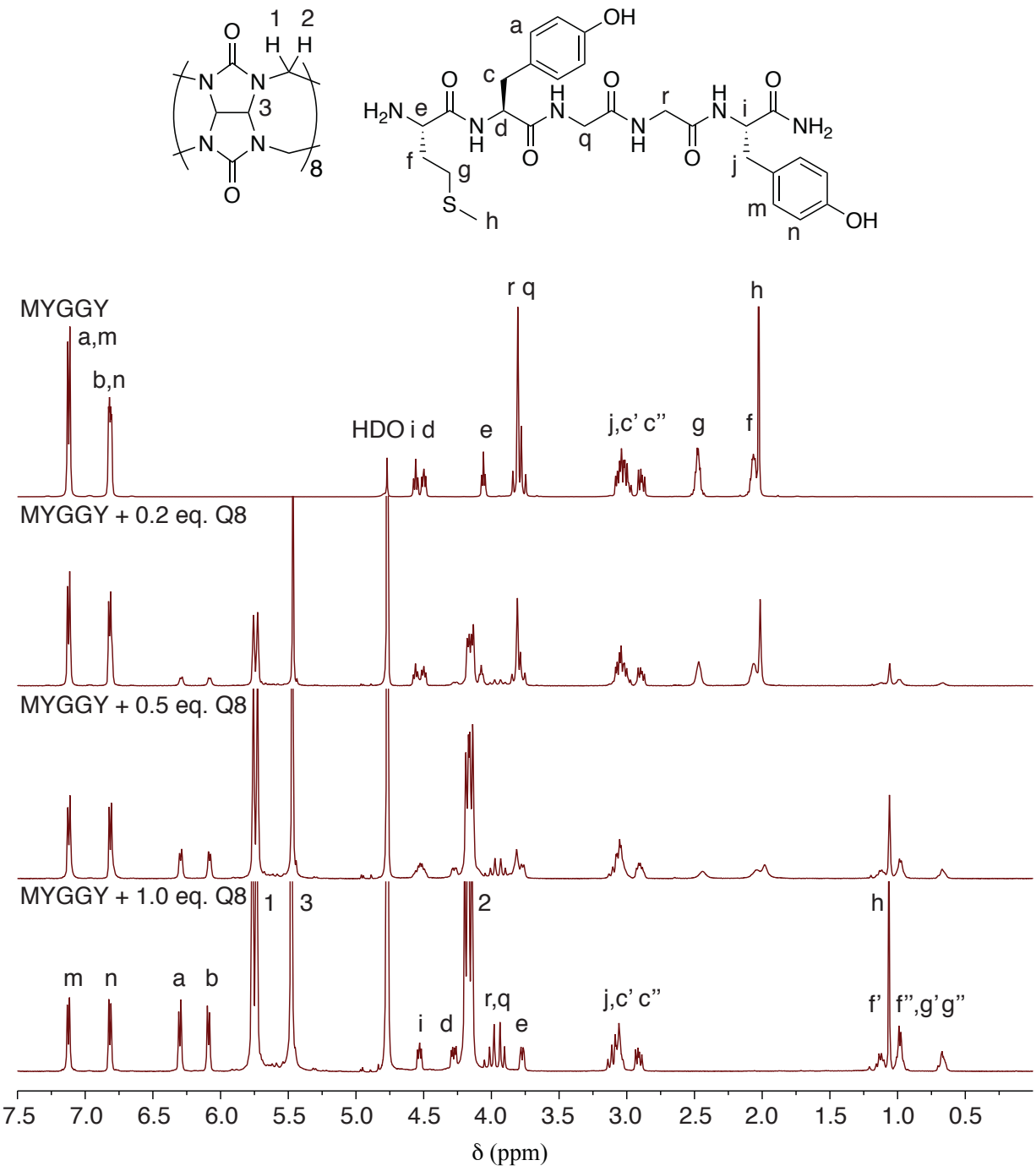


Figure S87. 500 MHz ^{1}H NMR spectra of Met-Tyr-Gly-Gly-Tyr with 0 eq. Q8, 0.2 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

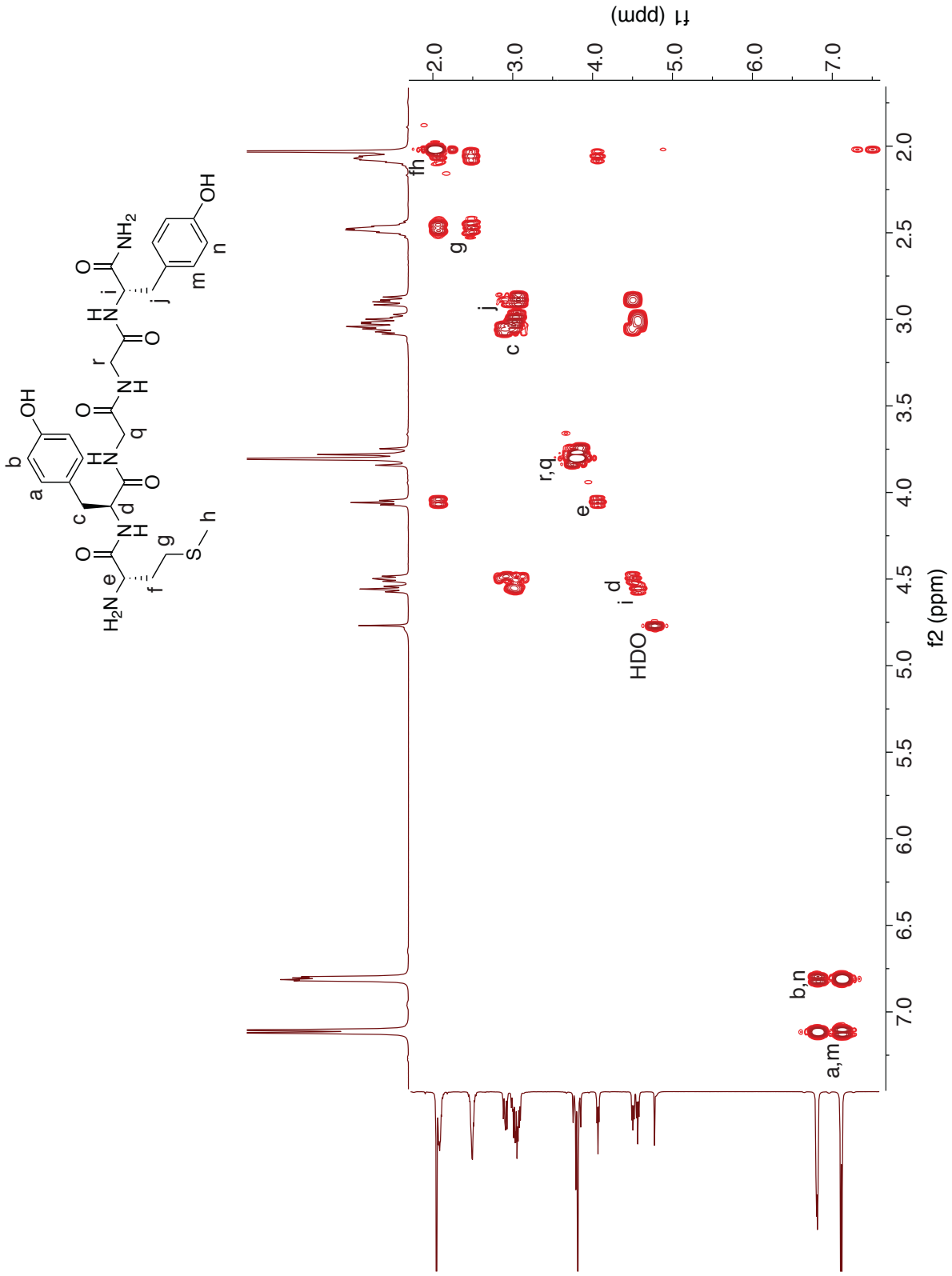


Figure S88. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Tyr-Gly-Gly-Tyr at 25 °C in D_2O.

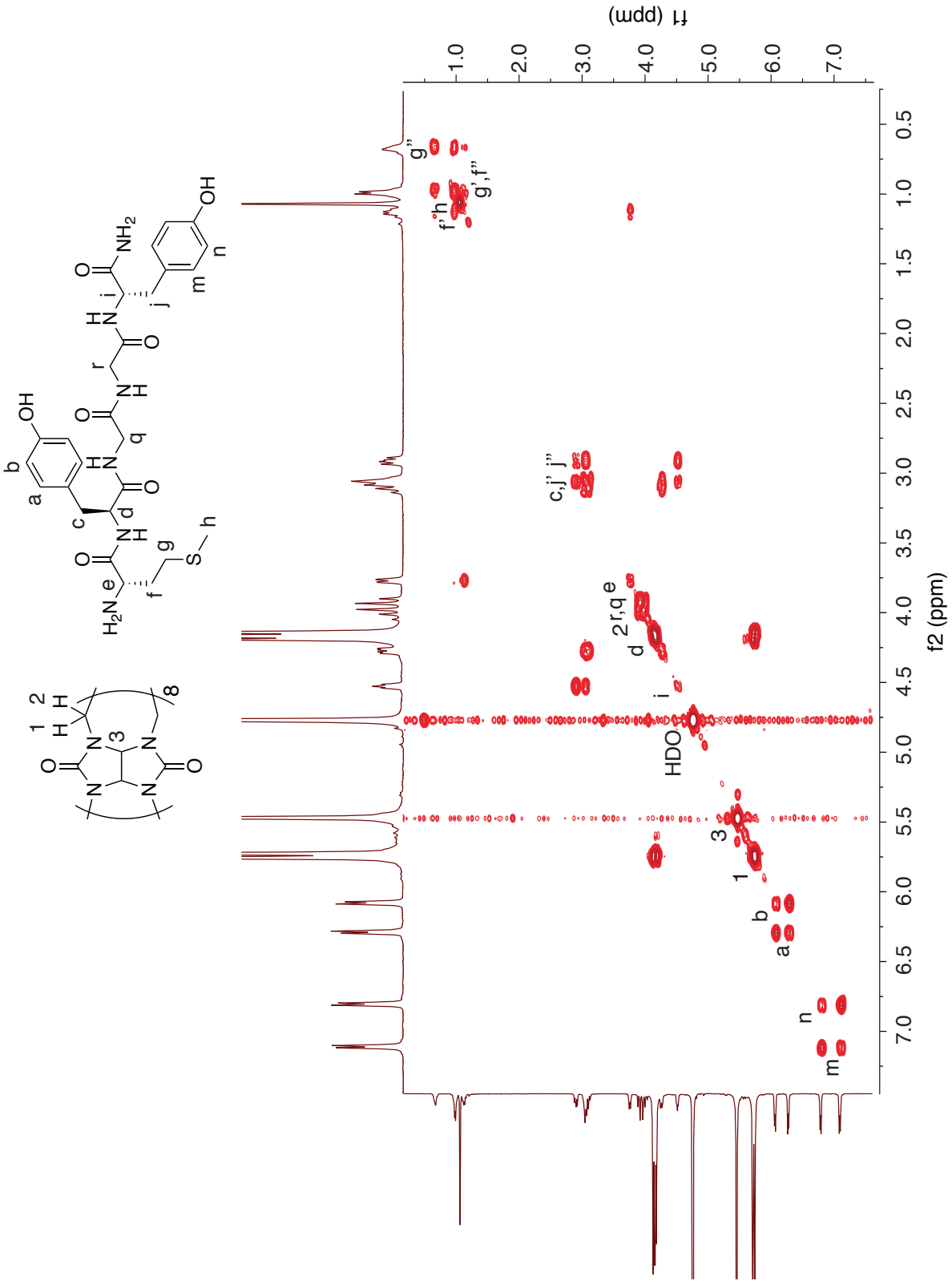


Figure S89. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Tyr-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

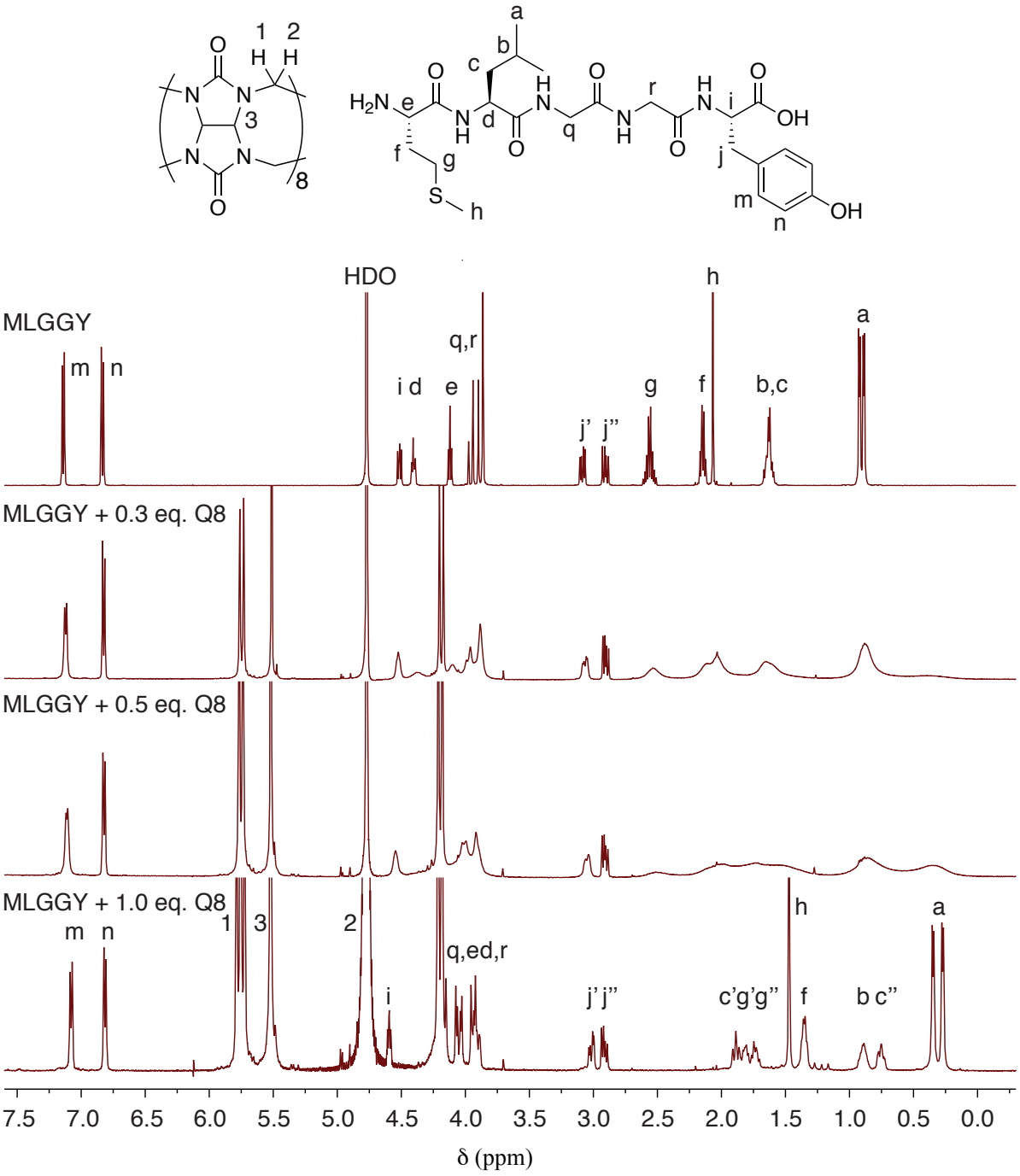


Figure S90. 500 MHz ^{1}H NMR spectra of Met-Leu-Gly-Gly-Tyr with 0 eq. Q8, 0.3 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

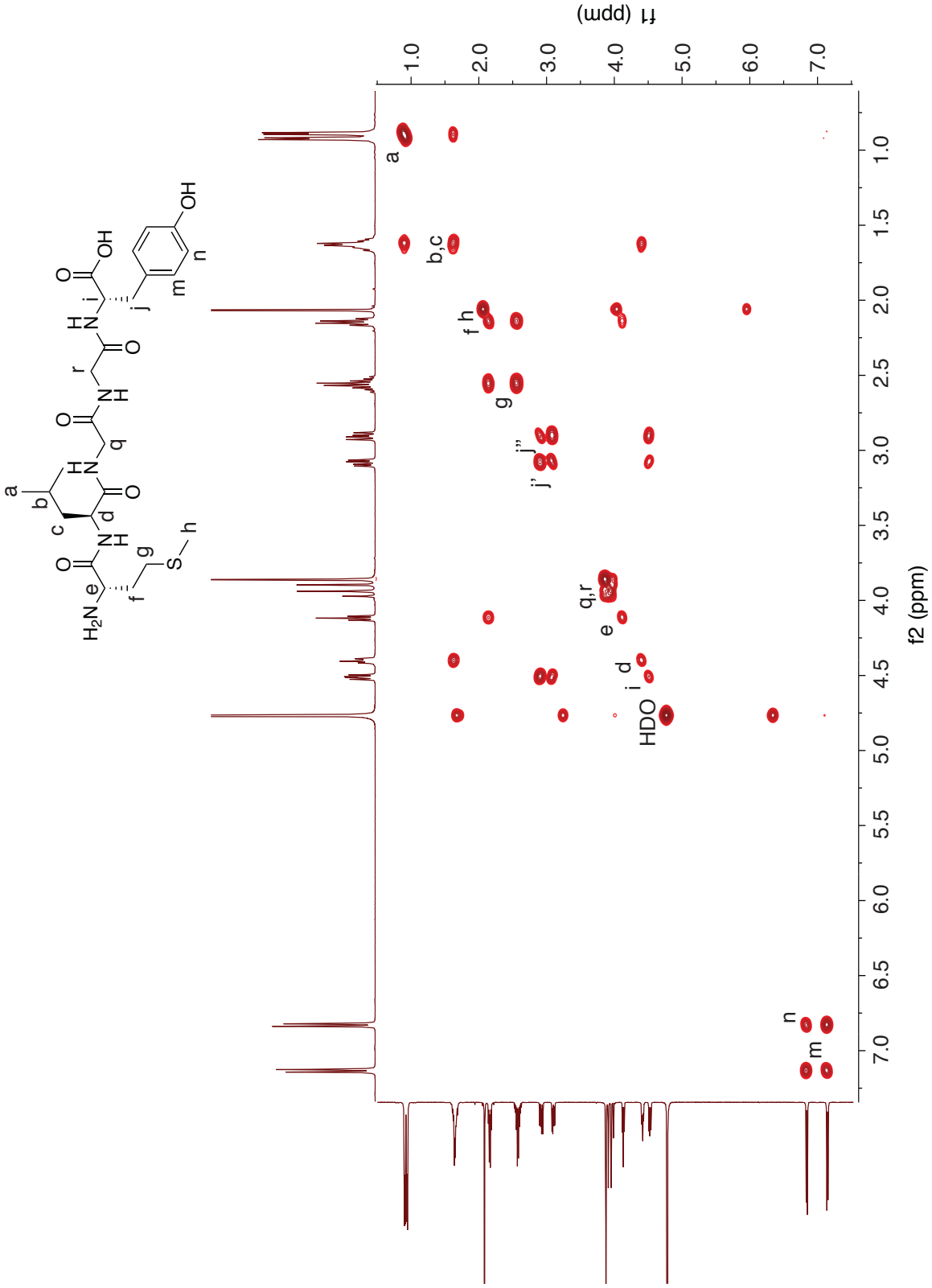


Figure S91. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Leu-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

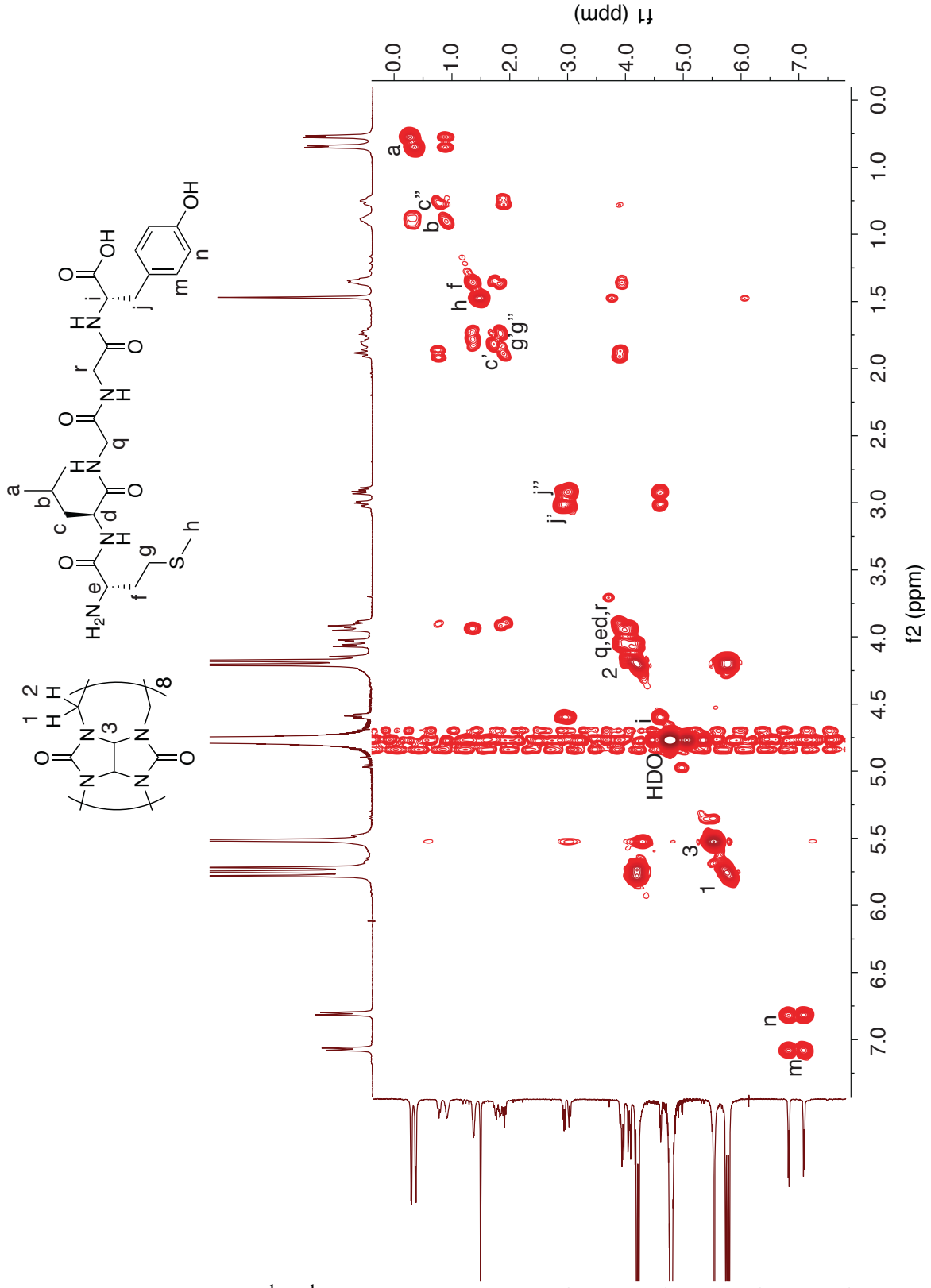


Figure S92. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Leu-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

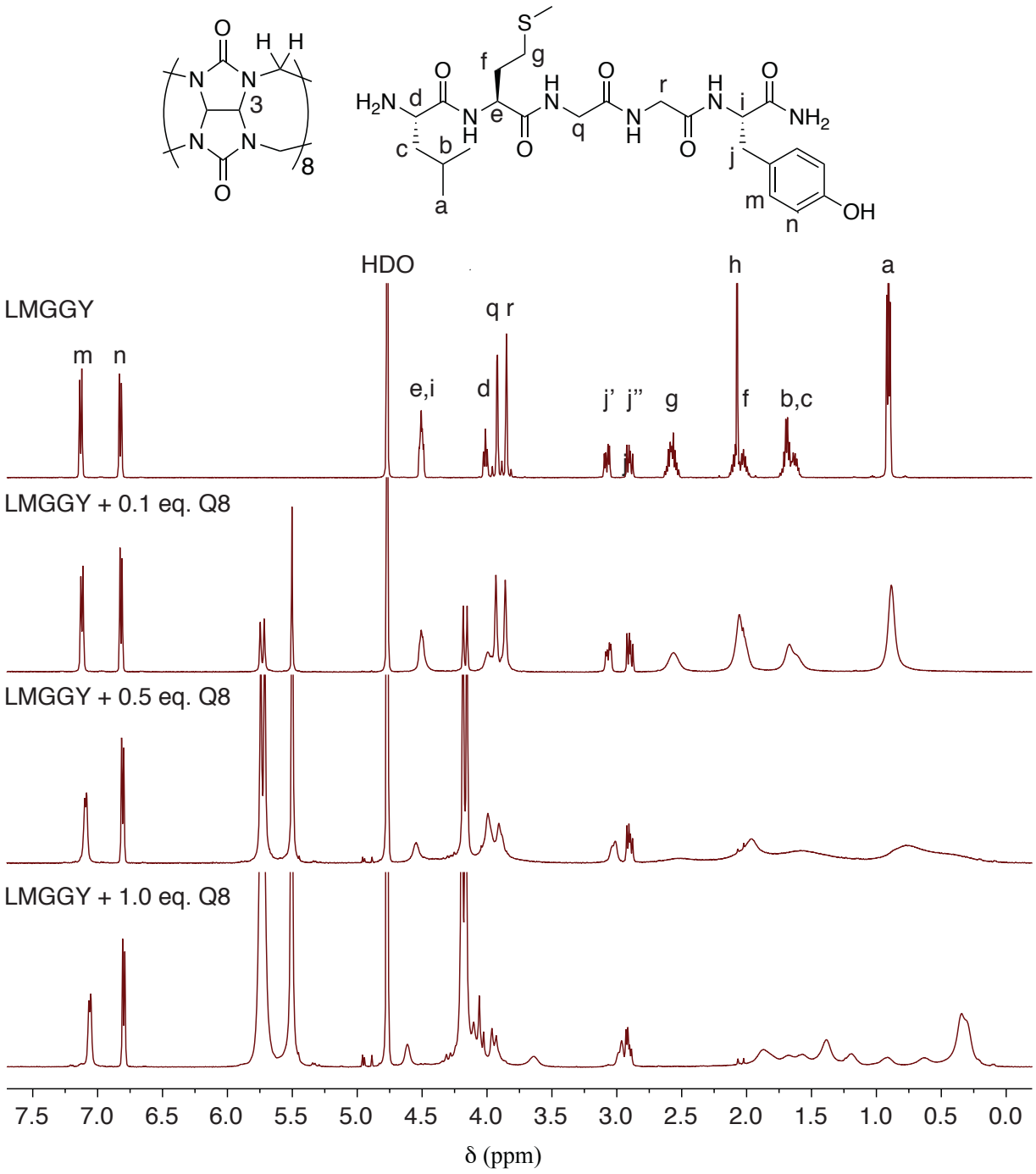


Figure S93. 500 MHz ^{1}H NMR spectra of Leu-Met-Gly-Gly-Tyr with 0 eq. Q8, 0.1 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

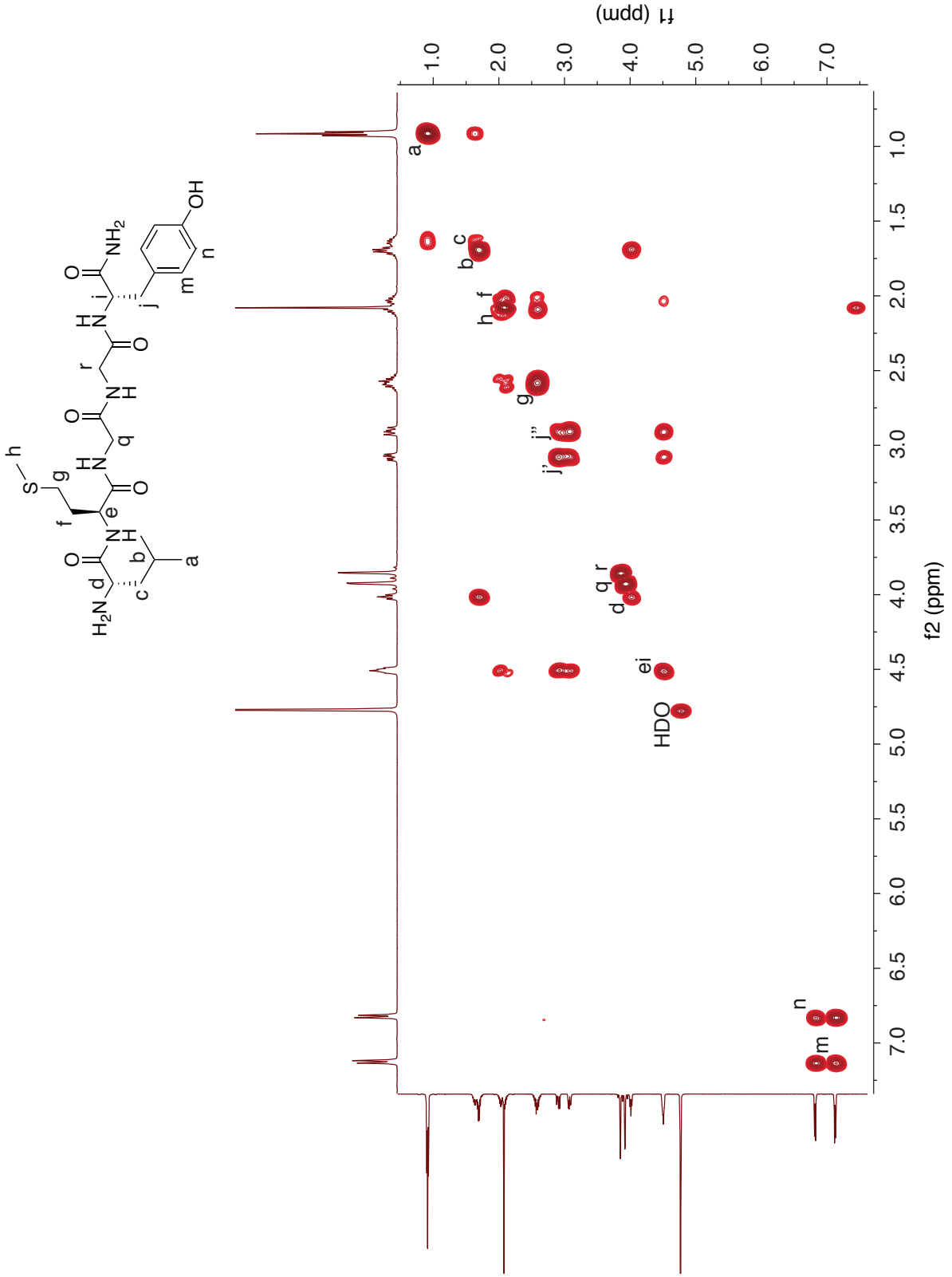


Figure S94. 500 MHz ^{1}H-^{1}H COSY spectrum of Leu-Met-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

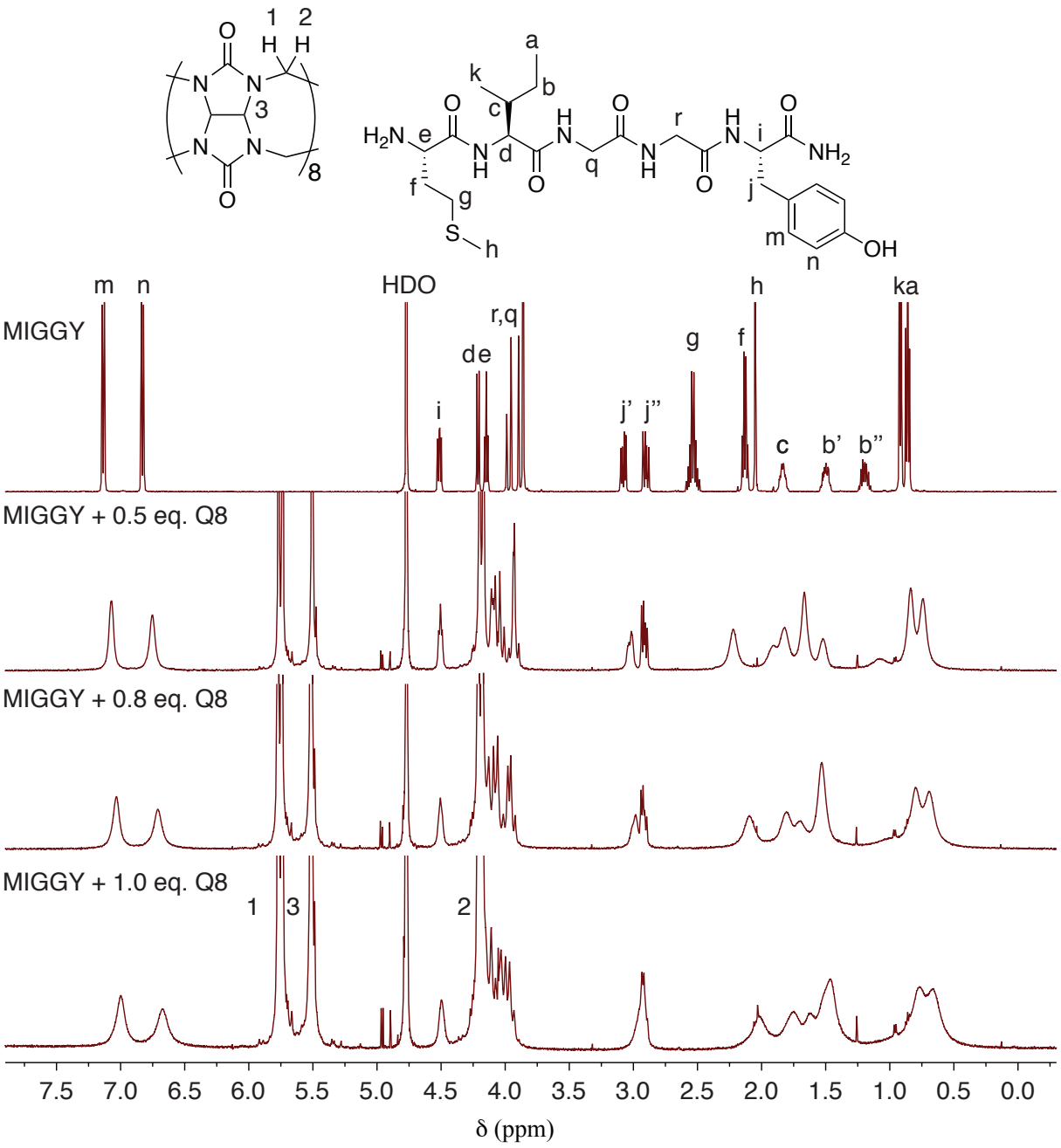


Figure S95. 500 MHz ^{1}NMR spectra of Met-Ile-Gly-Gly-Tyr with 0 eq. Q8, 0.5 eq. Q8, 0.8 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation. Solubility problems occurred after 0.5 eq. of Q8 were added.

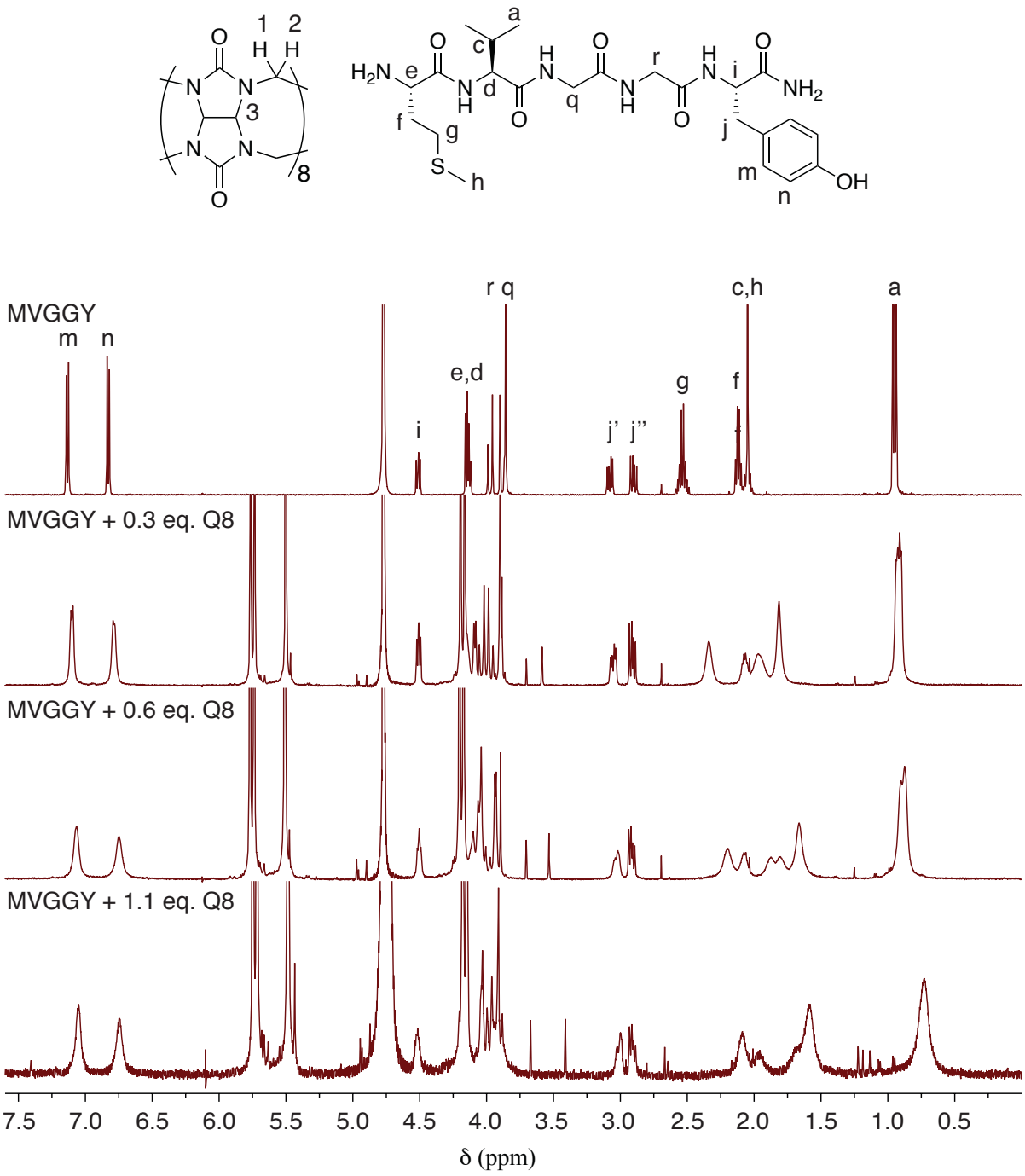


Figure S96. 500 MHz ^{1}H NMR spectra of Met-Val-Gly-Gly-Tyr with 0 eq. Q8, 0.3 eq. Q8, 0.6 eq. Q8, and 1.1 eq. Q8 at 25°C in D_2O. Apostrophes indicate geminal separation.

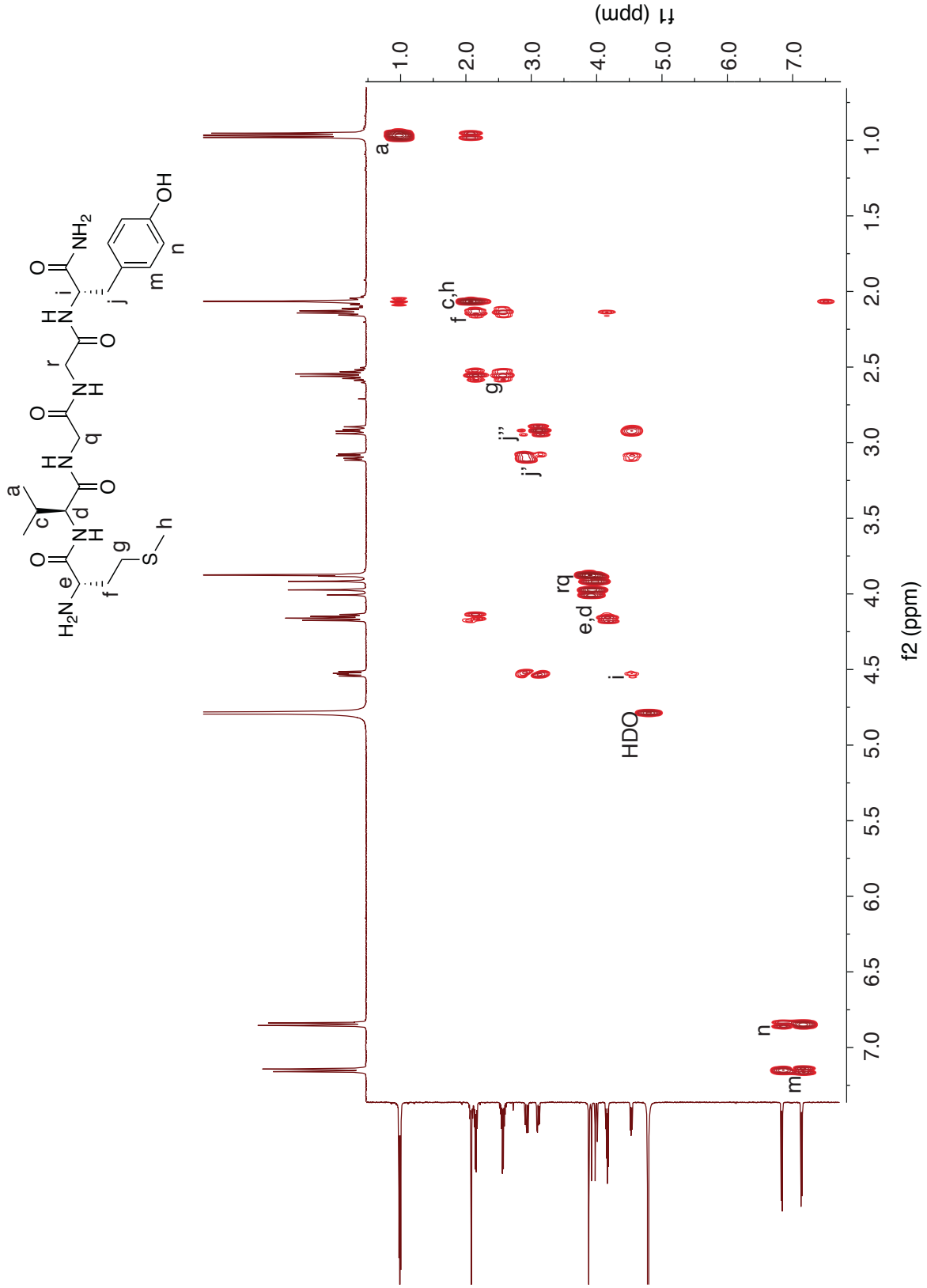


Figure S97. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Val-Gly-Gly-Tyr at 25 °C in D_2O.

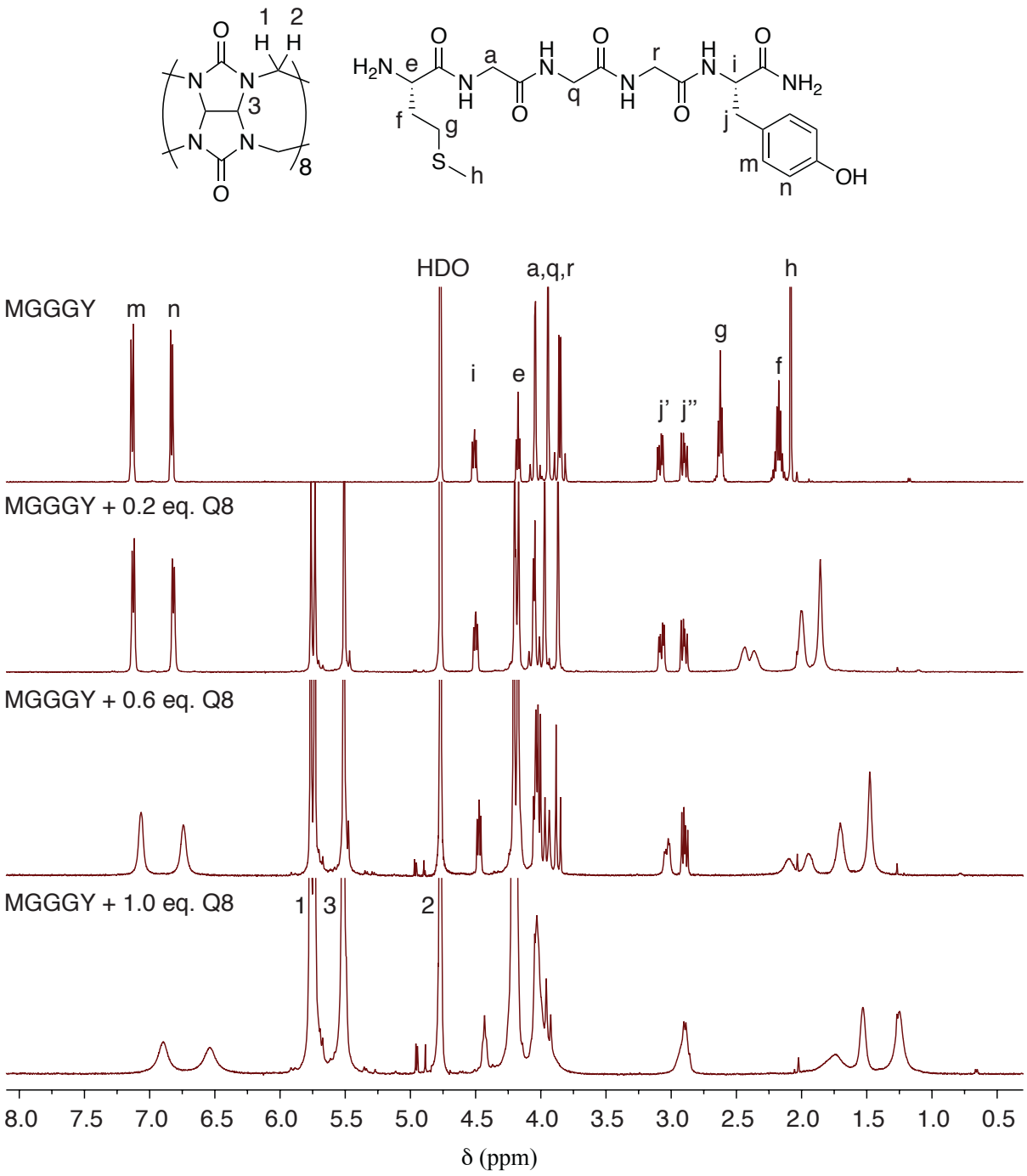


Figure S98. 500 MHz ^{1}H NMR spectra of Met-Gly-Gly-Gly-Tyr with 0 eq. Q8, 0.2 eq. Q8, 0.6 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

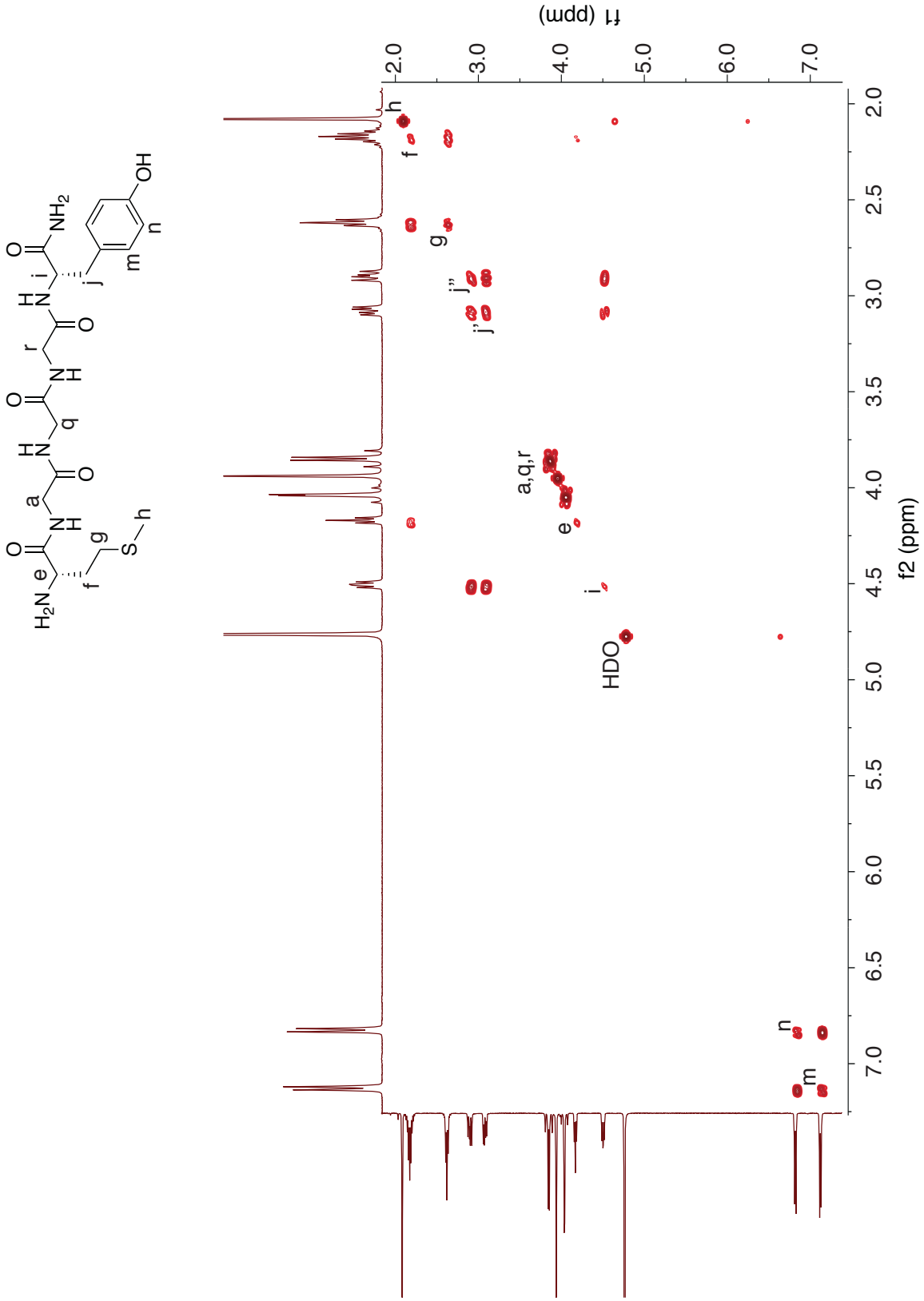


Figure S99. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Gly-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

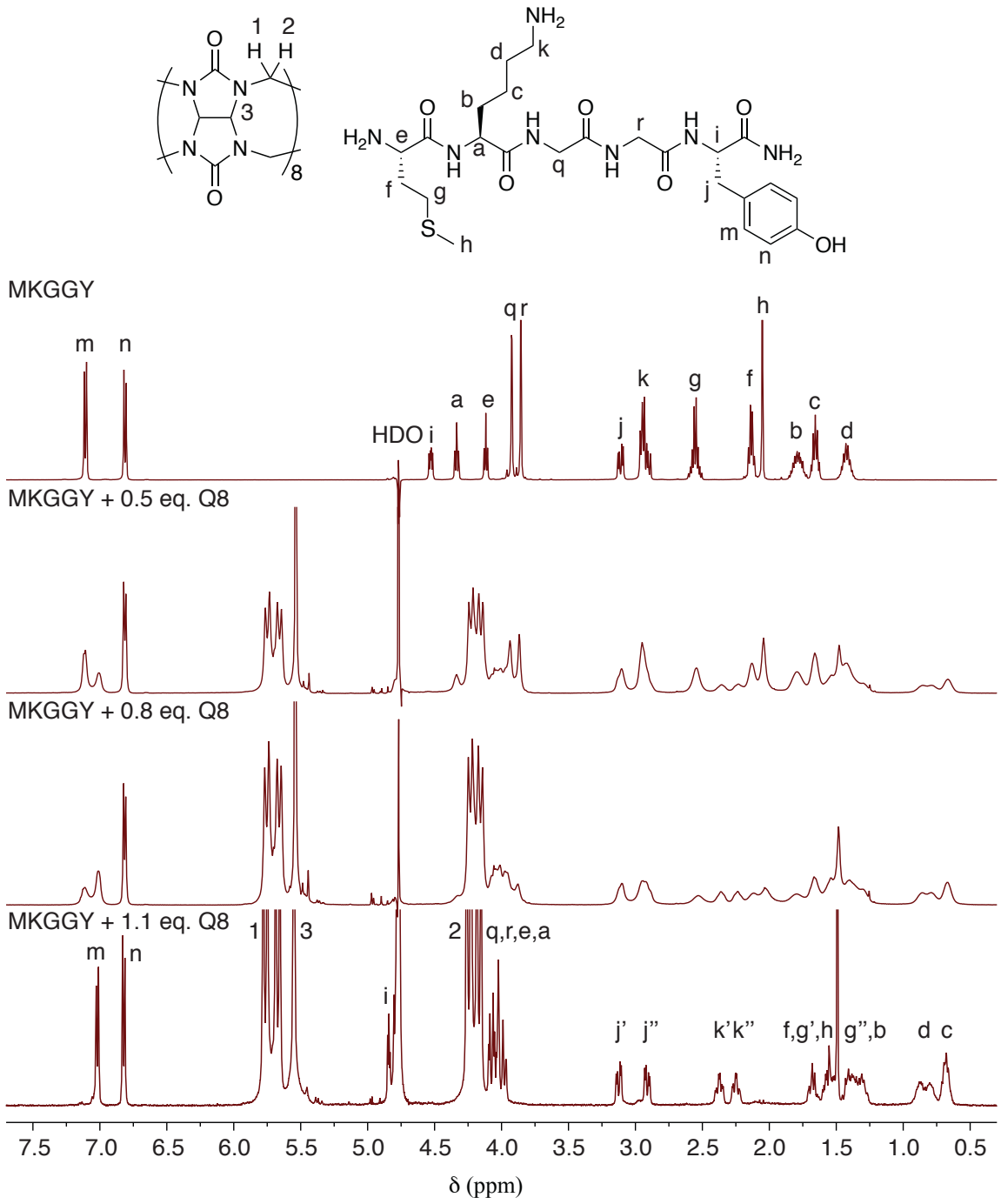


Figure S100. 500 MHz ^{1}H NMR spectra of Met-Lys-Gly-Gly-Tyr with 0 eq. Q8, 0.5 eq. Q8, 0.8 eq. Q8, and 1.1 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separations.

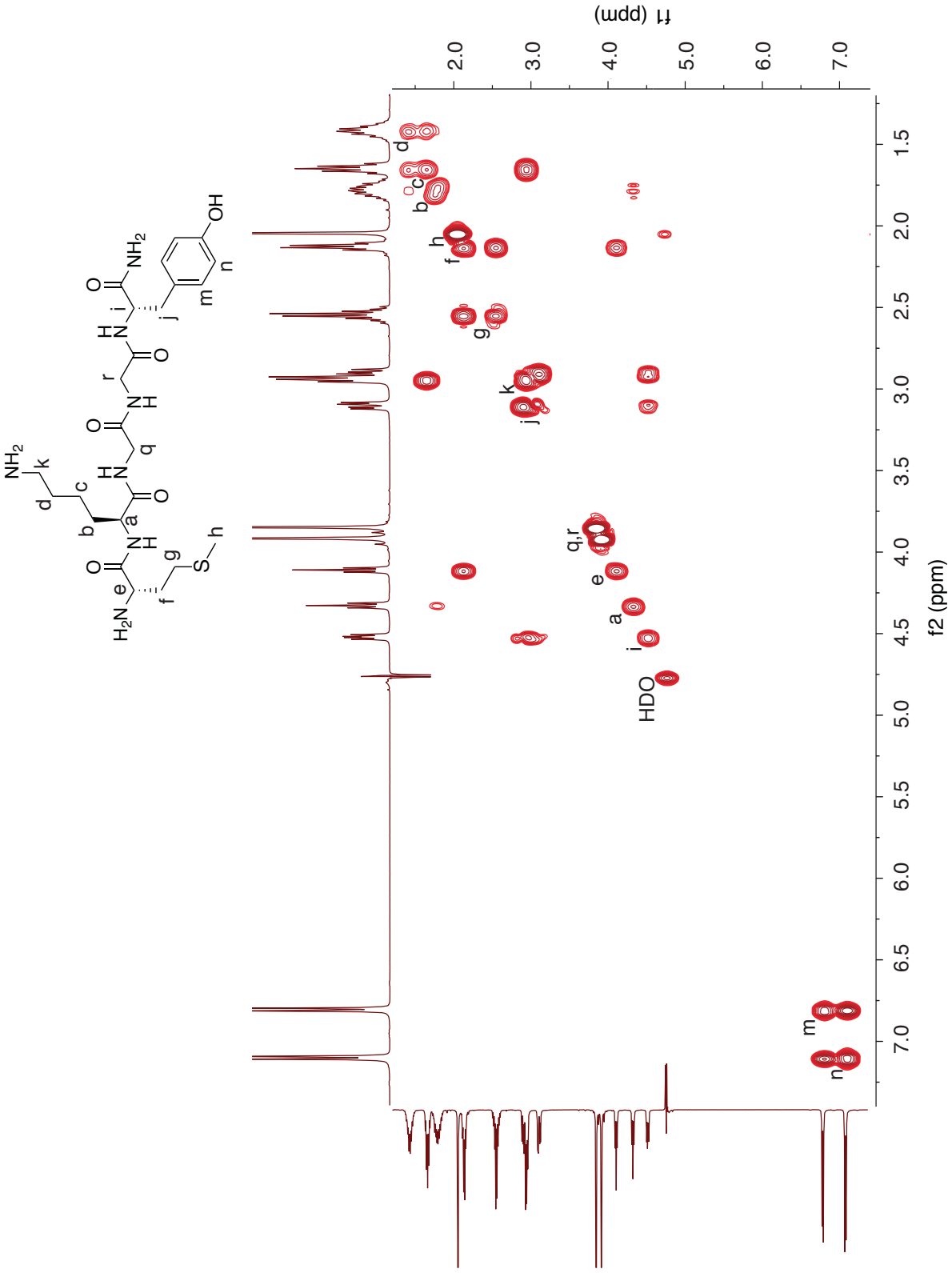


Figure S101. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Lys-Gly-Gly-Tyr at 25 °C in D_2O.

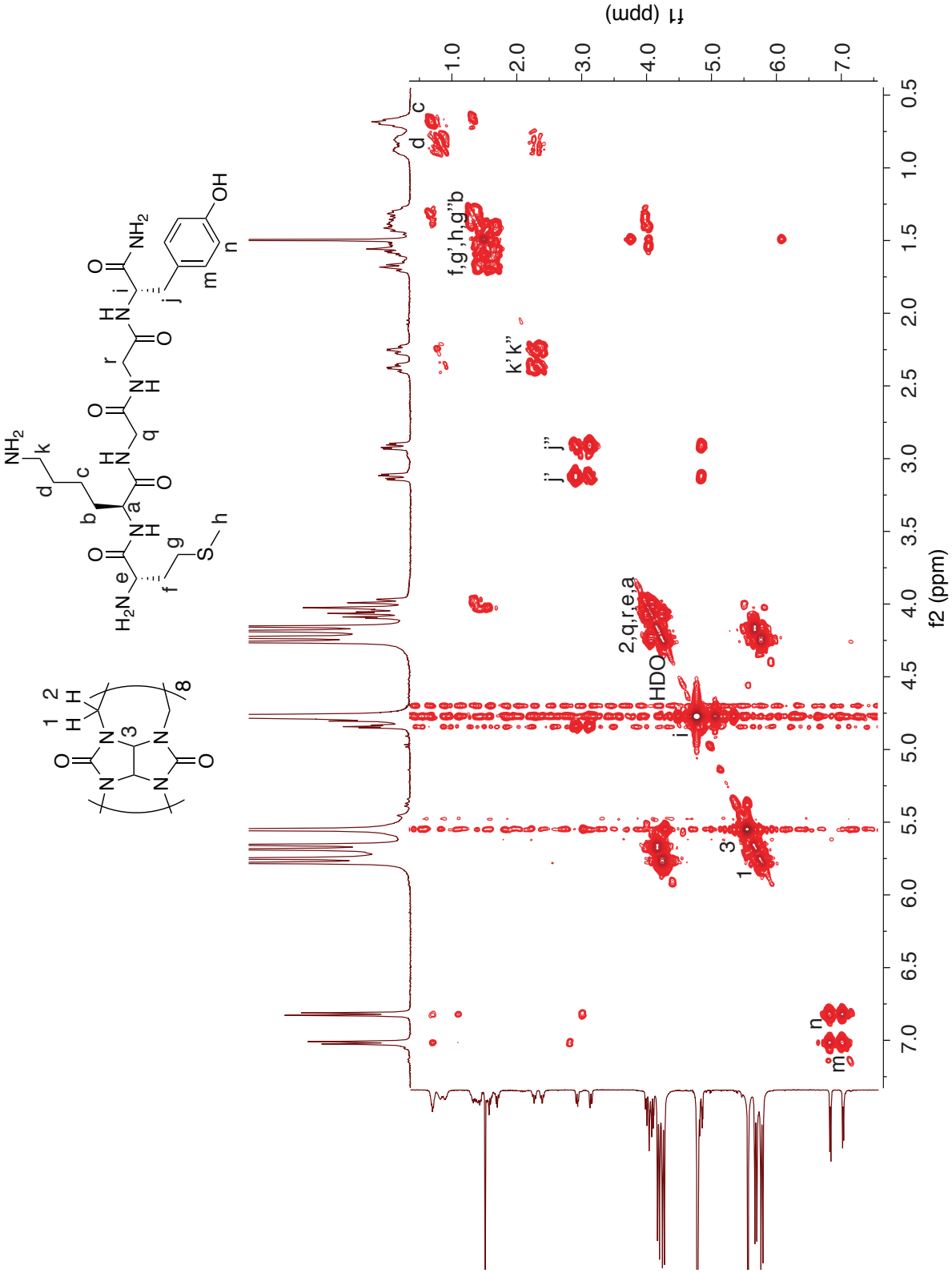


Figure S102. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Lys-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separations.

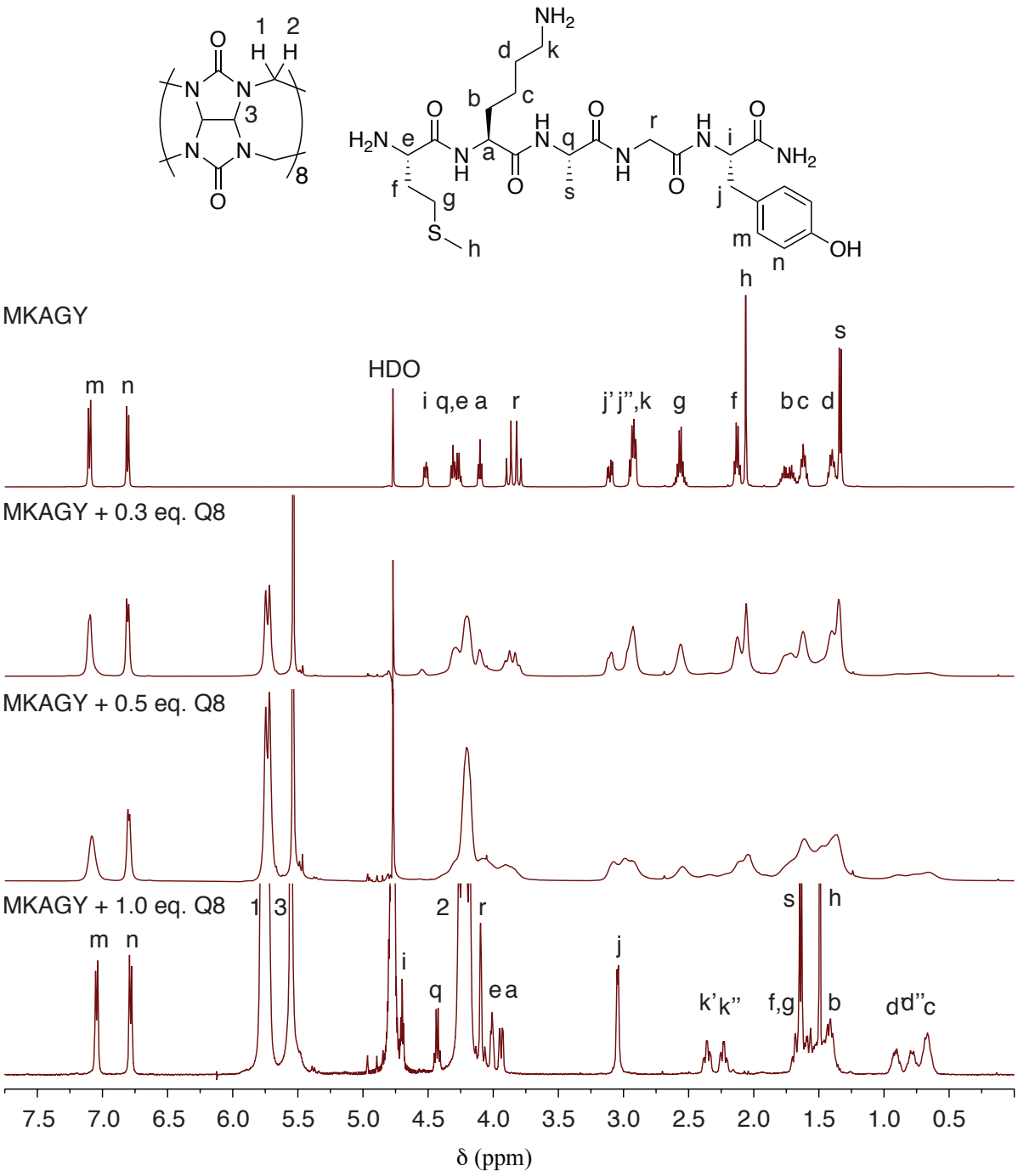


Figure S103. 500 MHz ^{1}H NMR spectra of Met-Lys-Ala-Gly-Tyr with 0 eq. Q8, 0.3 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

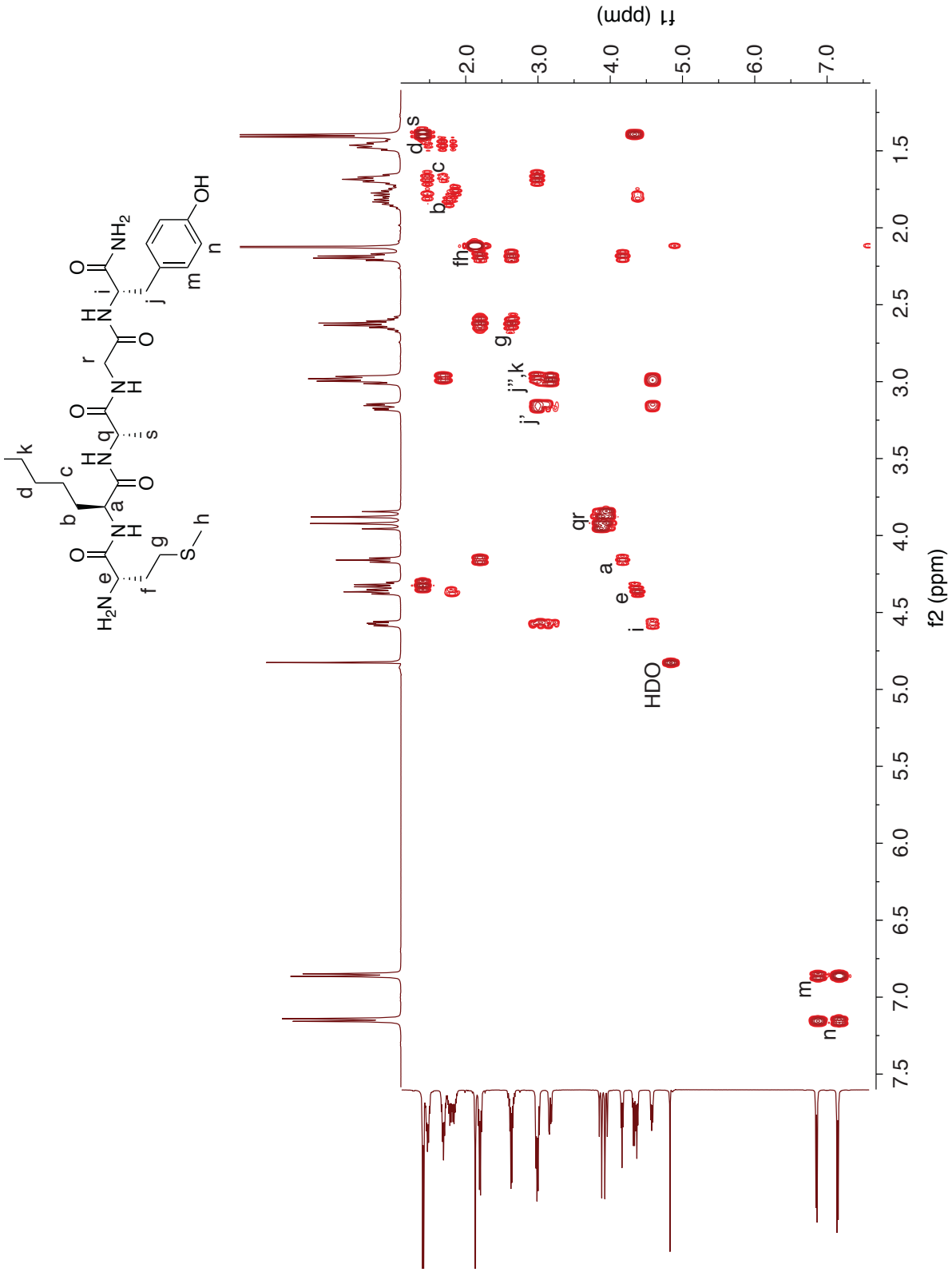


Figure S104. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Lys-Ala-Gly-Tyr at 25 °C in D_2O.

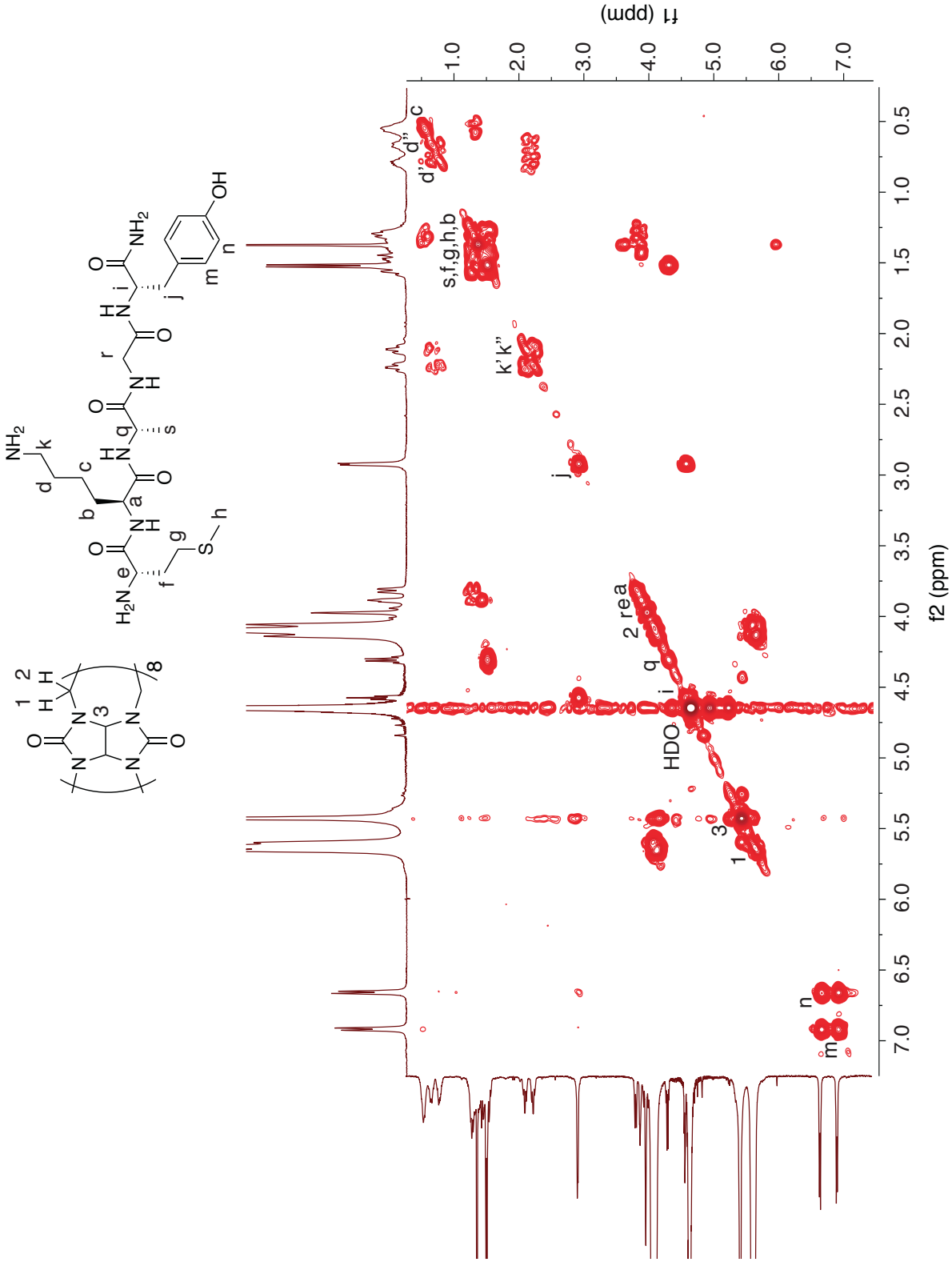


Figure S105. 500 MHz ^{1}H-^{1}H COSY spectrum of a 1:1 mixture of Q8 and Met-Lys-Ala-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

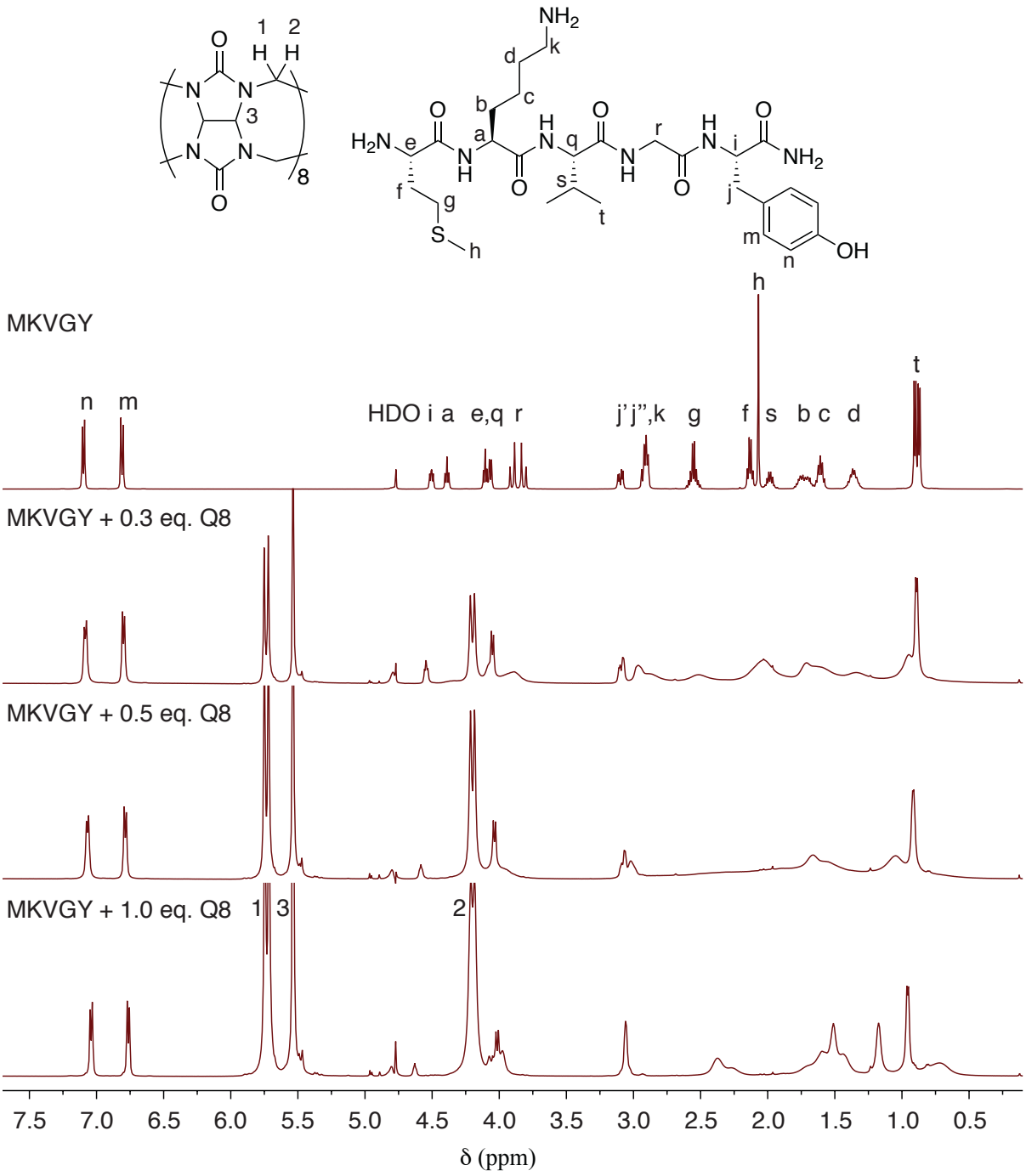


Figure S106. 500 MHz ^{1}H NMR spectra of Met-Lys-Val-Gly-Tyr with 0 eq. Q8, 0.3 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O.

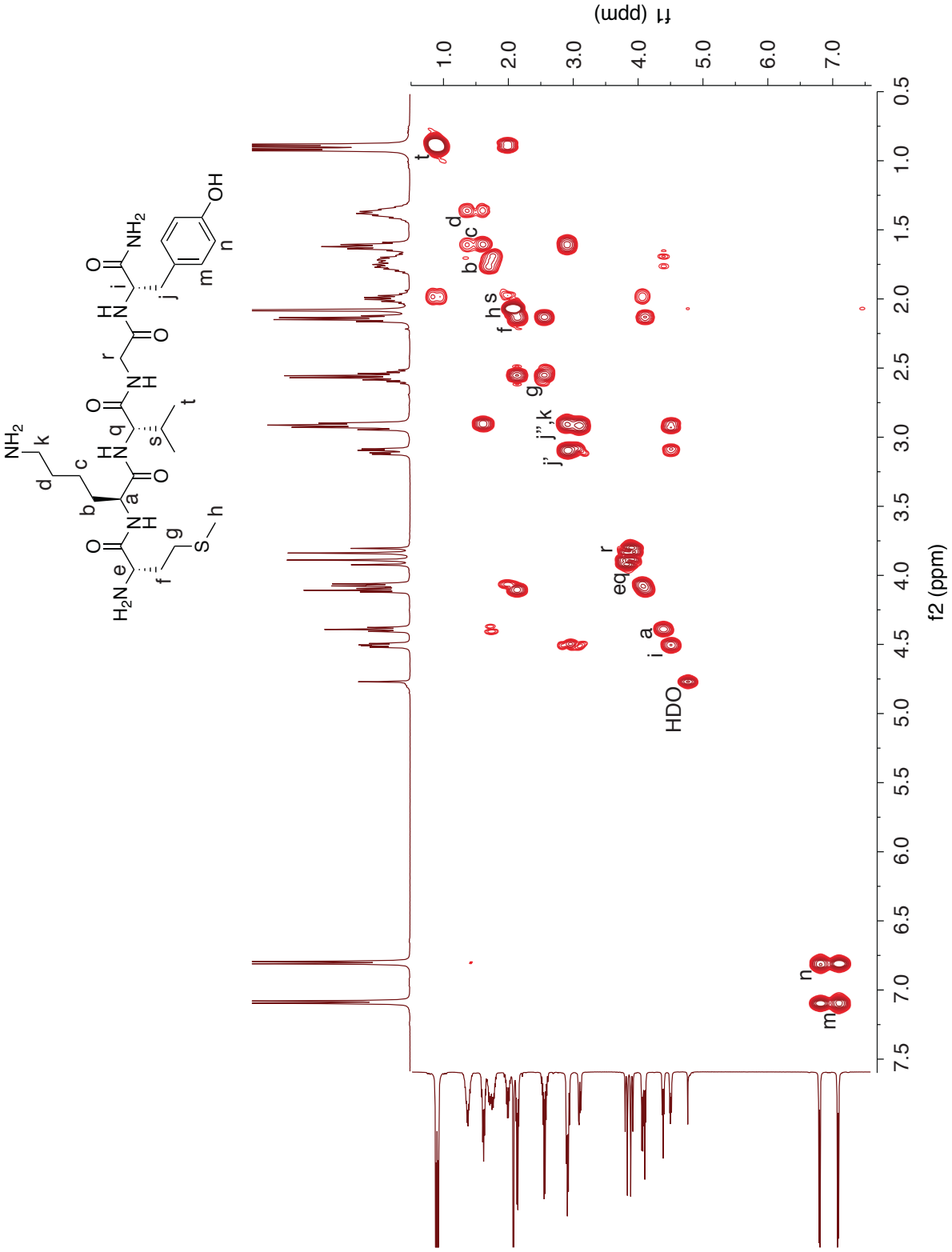


Figure S107. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Lys-Val-Gly-Tyr at 25 °C in D_2O.

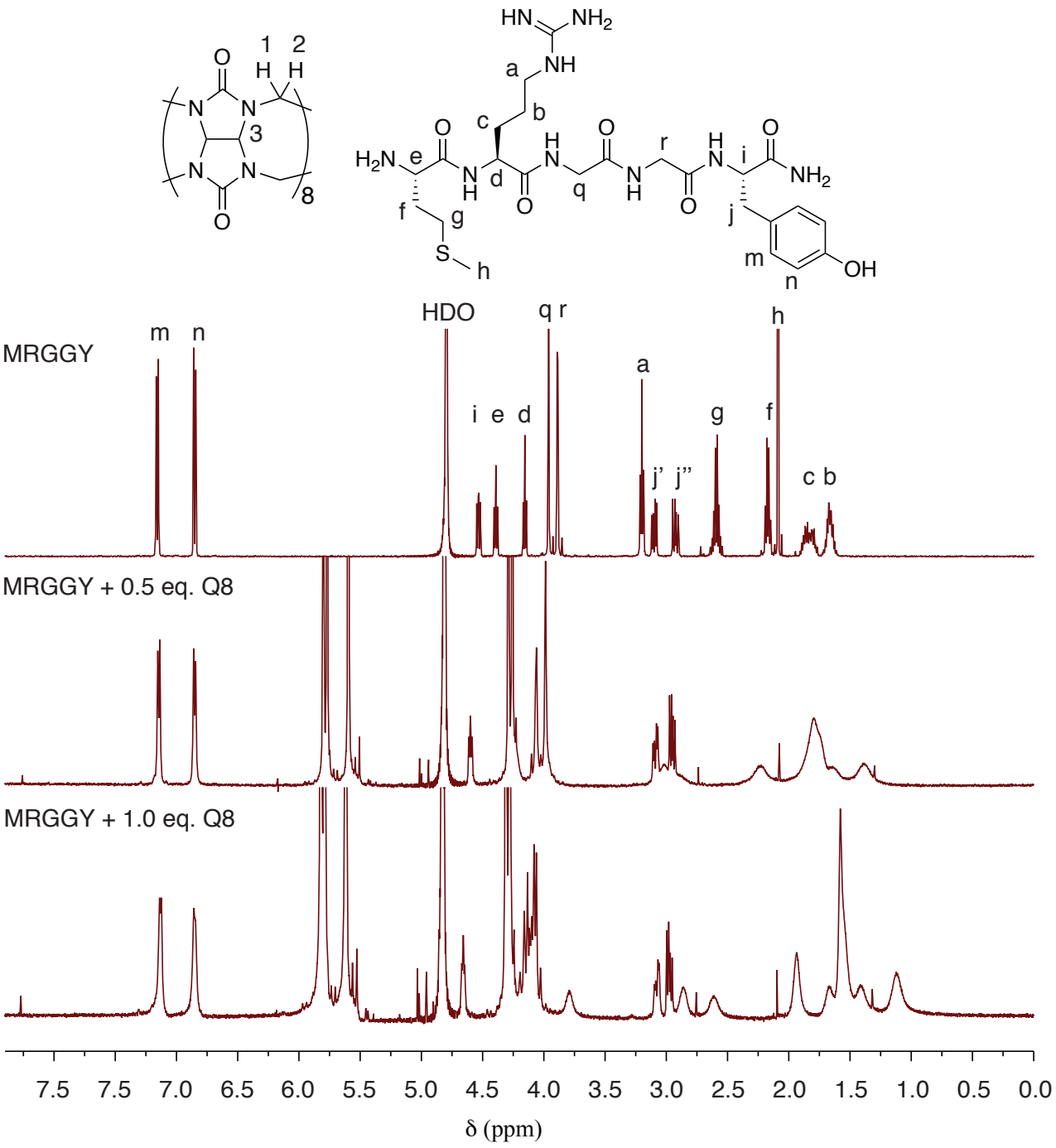


Figure S108. 500 MHz ^{1}H NMR spectra of Met-Arg-Gly-Gly-Tyr with 0 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Apostrophes indicate geminal separation.

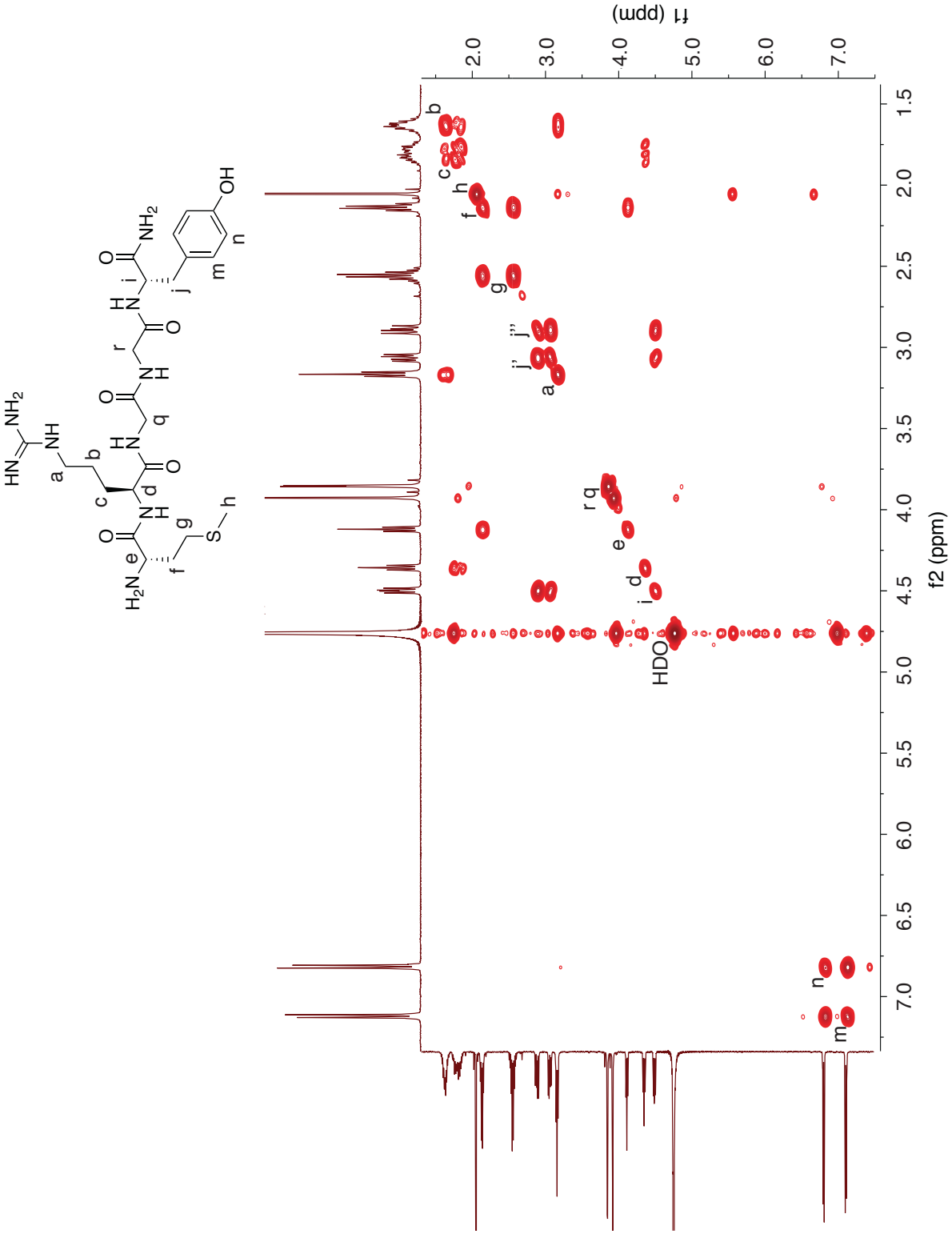


Figure S109. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Arg-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

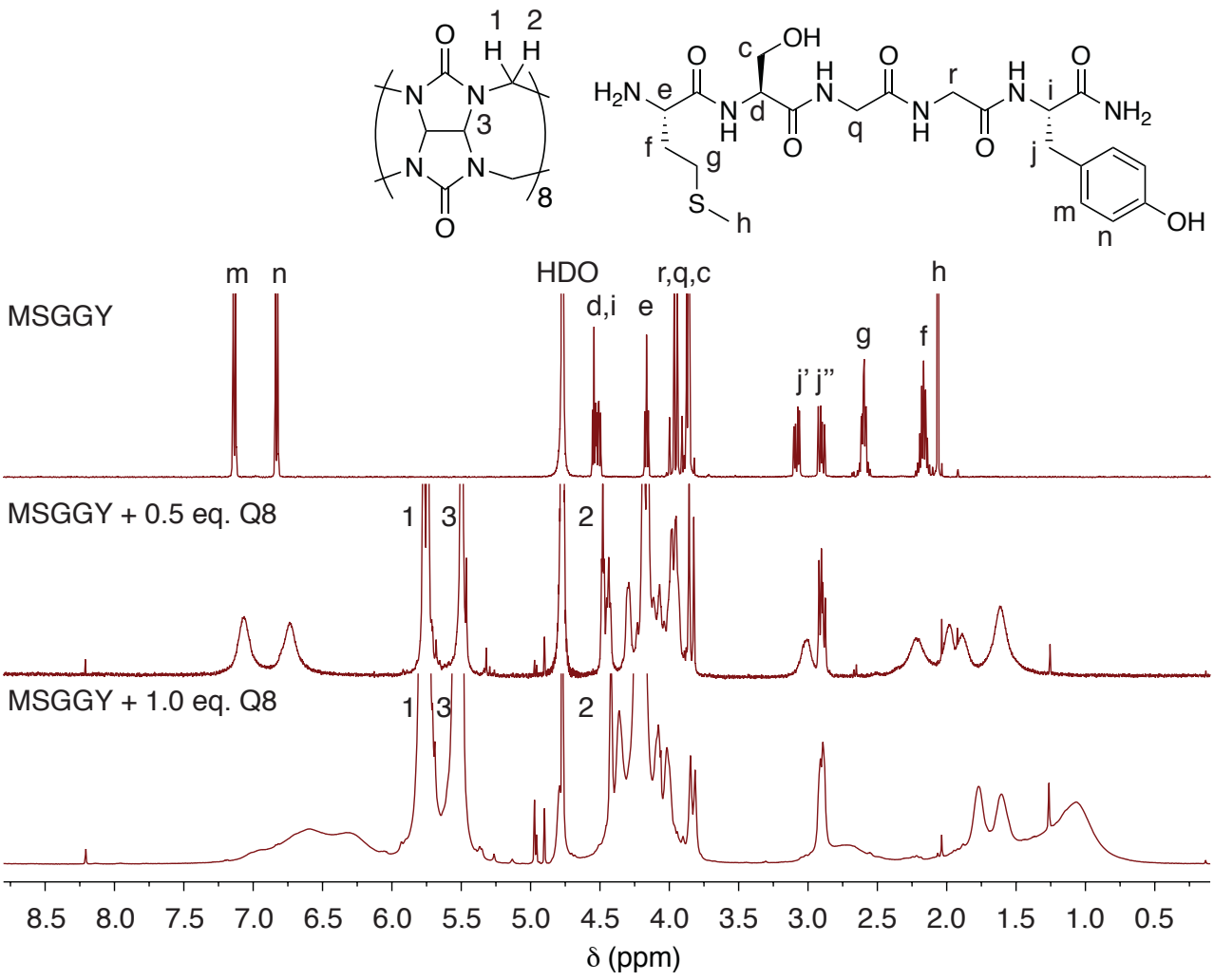


Figure S110. 500 MHz ^{1}H NMR spectra of Met-Ser-Gly-Gly-Tyr with 0 eq. Q8, 0.5 eq. Q8, and 1.0 eq. Q8 at 25 °C in D_2O. Solubility problems occurred after 0.5 eq. Q8 was added.

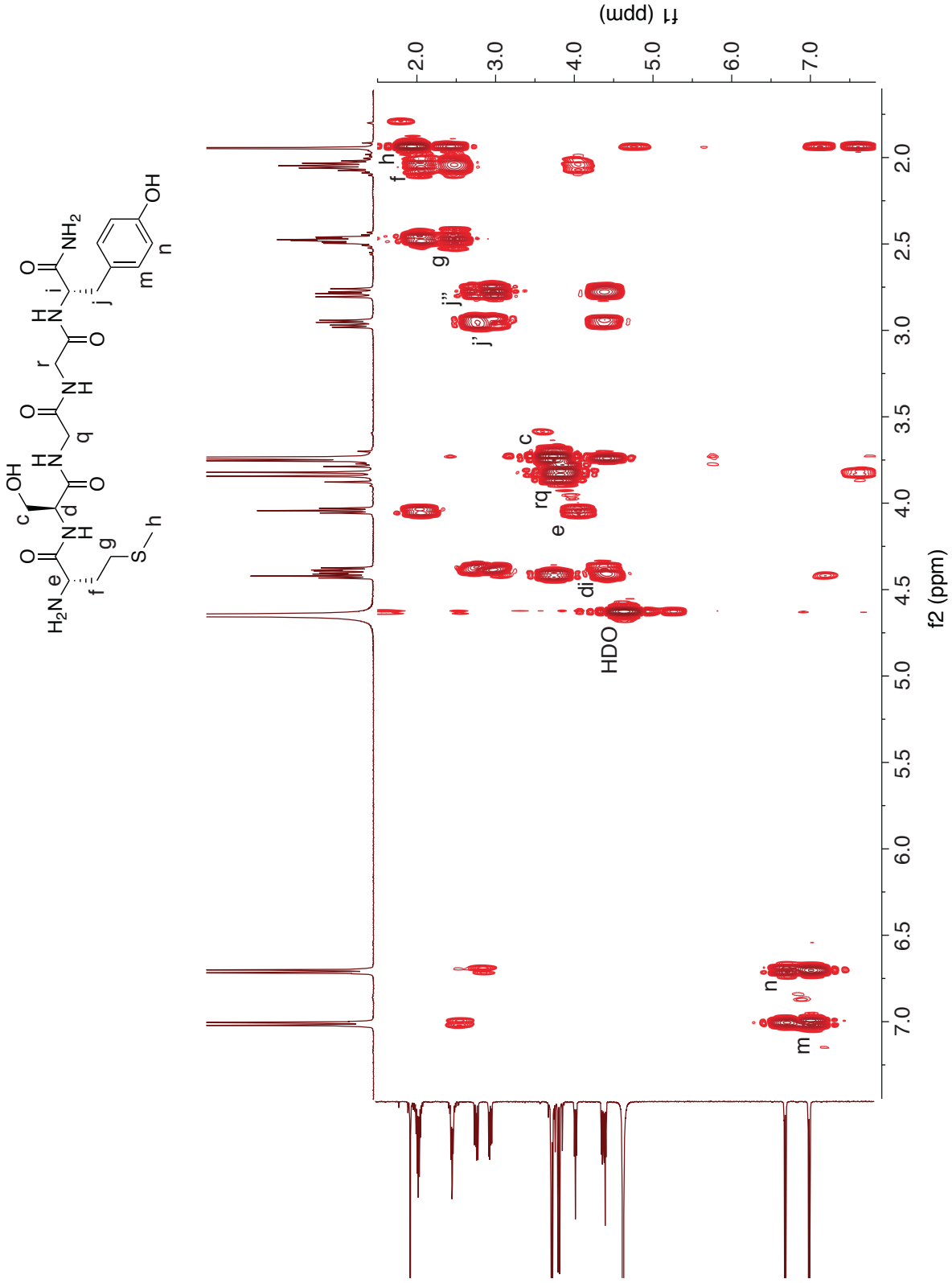


Figure S111. 500 MHz ^{1}H-^{1}H COSY spectrum of Met-Ser-Gly-Gly-Tyr at 25 °C in D_2O. Apostrophes indicate geminal separation.

Table S1. Fluorescence relative to the Q8•MBBI complex for mixtures of Q8•MBBI with each of 144 tripeptides of sequence Var1-Var2-Ala. Samples contained 200 μM peptide and 40 μM Q•MBBI at room temperature in 10 mM sodium phosphate buffer (pH 7.0). Values are averages of at least three experiments, and standard deviations are reported in parenthesis.

Var_1 \ Var_2	Y	F	I	L	M	P	V	A	G	Q	N	T	S	H	R	K	E	D
Y	28.5	38.4	23.6	41.5	30.6	25.1	33.3	30.8	29.8	23.6	24.9	28.7	25.7	19.5	26.6	40.8	27.0	29.2
	(3.0)	(2.7)	(3.5)	(3.3)	(2.3)	(3.8)	(3.1)	(3.0)	(4.3)	(4.9)	(2.6)	(3.8)	(3.4)	(8.6)	(3.5)	(2.6)	(4.2)	(0.8)
F	24.2	31.1	31.8	39.3	33.7	27.2	27.3	36.4	41.3	34.5	38.6	33.7	28.9	37.5	20.8	40.2	39.5	40.4
	(0.5)	(2.8)	(2.9)	(2.4)	(2.4)	(1.6)	(0.4)	(1.9)	(3.1)	(2.4)	(2.3)	(1.5)	(0.4)	(2.0)	(1.3)	(3.6)	(3.8)	(2.0)
I	40.7	36.3	30.4	35.1	27.3	27.2	30.9	21.6	30.6	26.9	25.8	25.8	23.0	32.0	23.4	36.7	32.6	23.2
	(1.1)	(2.8)	(2.5)	(2.0)	(1.7)	(2.3)	(1.5)	(1.9)	(2.1)	(1.4)	(2.1)	(1.8)	(2.5)	(2.6)	(2.7)	(2.1)	(2.7)	(2.7)
L	41.2	33.5	29.8	36.3	29.3	33.4	29.2	32.3	29.4	33.4	30.3	30.6	30.6	30.8	17.5	44.3	32.9	28.7
	(1.5)	(3.1)	(0.6)	(1.7)	(7.9)	(2.1)	(3.0)	(1.1)	(1.7)	(2.1)	(2.2)	(1.8)	(3.1)	(1.1)	(2.1)	(2.8)	(2.2)	(3.1)
M	40.4	39.4	18.2	35.2	22.4	20.8	15.6	19.1	19.2	18.5	20.0	20.3	18.2	22.8	19.7	32.9	17.2	17.6
	(0.6)	(1.1)	(0.8)	(1.3)	(1.0)	(0.7)	(1.5)	(0.6)	(1.1)	(1.6)	(3.1)	(0.8)	(3.8)	(0.7)	(1.6)	(0.2)	(1.3)	(1.3)
P	19.2	15.7	16.9	22.9	24.0	20.8	16.9	14.1	21.0	15.6	18.5	25.1	21.6	22.5	19.7	25.6	17.2	18.8
	(0.8)	(3.9)	(3.0)	(2.6)	(1.1)	(1.7)	(1.0)	(1.6)	(1.9)	(1.3)	(0.8)	(1.3)	(0.8)	(3.0)	(2.3)	(1.4)	(0.7)	(1.3)
R	36.0	36.4	11.6	34.3	21.3	17.0	19.4	17.0	12.1	20.1	14.6	15.0	15.7	17.5	18.7	20.5	14.4	14.1
	(1.0)	(2.6)	(3.2)	(3.2)	(2.8)	(1.2)	(1.1)	(1.2)	(2.0)	(1.8)	(1.1)	(1.3)	(2.3)	(2.7)	(0.3)	(1.7)	(1.7)	(0.2)
K	33.9	42.9	16.6	39.5	35.4	19.9	15.5	24.9	24.3	14.9	19.4	18.0	12.4	27.0	25.8	26.1	18.7	19.3
	(1.2)	(2.3)	(3.1)	(3.7)	(1.8)	(2.7)	(1.8)	(3.6)	(2.8)	(3.4)	(2.1)	(1.8)	(5.5)	(1.4)	(2.6)	(1.5)	(3.1)	(3.7)