Conformational Investigation of the Structure – Activity Relationship of GdFFD and Its Analogues on an Achatin-like Neuropeptide Receptor of Aplysia californica Involved in the Feeding Circuit

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-06-2018-003661
Article Type:	Paper
Date Submitted by the Author:	10-Jun-2018

Complete List of Authors:	Do, Thanh; University of Illinois, Urbana-Champaign, Chemistry Checco, James; University of Illinois, Urbana-Champaign, Chemistry Tro, Michael; University of California Santa Barbara Shea, Joan-Emma; UCSB, Chemistry Bowers, M.; University of California Santa Barbara, Sweedler, Jonathan; University of Illinois, Department of Chemistry

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

PDB files.zip

Journal Name

ARTICLE

Conformational Investigation of the Structure – Activity Relationship of GdFFD and Its Analogues on an Achatin-like Neuropeptide Receptor of *Aplysia californica* Involved in the Feeding Circuit

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Thanh D. Do,[a,*] James W. Checco,[a] Michael Tro,[b] Joan-Emma Shea,[b,c] Michael T. Bowers[b] and Jonathan V. Sweedler[a,*]

Proteins and peptides in nature are almost exclusively made from L-amino acids, and this is even more absolute in the metazoan. With the advent of modern bioanalytical techniques, however, previously unappreciated roles for D-amino acids in biological processes have been revealed. Over 30 D-amino acid containing peptides (DAACPs) have been discovered in animals where at least one L-residue has been isomerized to the D-form via an enzyme-catalyzed process. In *Aplysia californica*, GdFFD and GdYFD (the lower-case letter "d" indicates the amino acid is a D-amino acid) modulate the feeding behavior by activating the *Aplysia* achatin-like neuropeptide receptor *(*apALNR). However, little is known about how the three-dimensional conformation of DAACPs influences activity at the receptor, and the role that D-residues play in these peptide conformations. Here, we use a combination of computational modeling, drift-tube ion-mobility mass spectrometry, and receptor activation assays to create a simple model that predicts bioactivities for a series of GdFFD analogs. Our results suggest that the active conformations of GdFFD and GdYFD are similar to their lowest energy conformations in solution. Our model helps connect the predicted structures of GdFFD analogs to their activities, and highlights a steric effect on peptide activity at position 1 on the GdFFD receptor apALNR. Overall, these methods allow us to understand ligand-receptor interactions in the absence of high-resolution structural data.

Introduction

Molecular recognition, which includes protein-protein and protein-ligand interactions with high specificity and affinity, constitutes the basis of many fundamental processes that are essential to life.[5,6] High-resolution protein-ligand structures obtained from X-ray crystallography and nuclear magnetic resonance (NMR) experiments dramatically enhance our understanding of how biology is transacted in three dimensions. However, to fully illuminate the key elements in protein function, ideally one needs to determine high-resolution structures for both the apo (ligand-free) and holo (ligand-bound) proteins, and to dissect the thermodynamic energy terms that govern the conformational differences. These processes remain tedious and challenging for structure-based efforts, evidenced by the small number of protein structures with congeneric ligands deposited in the RCSB Protein Data Bank.[8] In addition, many endogenous ligands are not rigid molecules, but short peptides that may undergo fast molecular motions in their free forms. Many of these peptides are signaling molecules that bind to specific cell surface receptors and trigger intracellular effects.[9,10] A large number of bioinformatic tools[11-17] have been developed to predict the binding sites of flexible ligands, and to improve our ability to reliably estimate the affinity of a given protein-ligand pairing in the absence of high-quality apo- and holo-structures.

Several D-amino acid-containing peptides (DAACPs) are endogenously produced and act as neuropeptides in the central nervous system (CNS) of the model organism *Aplysia californica*. For example, GdFFD and GdYFD (where each D-residue is denoted using a lower case "d" followed by the one-letter amino acid code) were shown to act as extrinsic modulators of the feeding circuit[4,7] and intrinsic neuromodulators in the locomotor network.[7,18] In 2015, Bauknecht and Jekely[19] screened 126 neuropeptides against 87 orphan G protein-coupled receptors from the annelid *Platynereis* and identified ligands for 19 receptors. Through protein homology, this study identified an *Aplysia* receptor referred to as the *Aplysia* achatin-like neuropeptide receptor (apALNR), which was activated by GdFFD, but not by GFFD. Checco et al.[3] further investigated the substrate specificity of

[a] Department of Chemistry and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, United States
Email: jsweedle@illinois.edu
Email: thanhddo@illinois.edu
[b] Department of Chemistry and Biochemistry, [c] Department of Physics, University of California at Santa Barbara, Santa Barbara, Californica 93106, United States.
Electronic Supplementary Information (ESI) available: Additional Figures showing the computational modelling workflow, representative mass spectra and ATDs of GFFD, GdFFD and its analogues. Coordinates of GFFD, GdFFD and its analogues in PDB format. Peptide characterization data for previously unreported peptides.
See DOI: 10.1039/x0xx00000x

this receptor and showed that *apALNR* is expressed throughout the *Aplysia* CNS, suggesting that GdFFD (and the related ligand GdYFD) likely play many different physiological roles throughout the animal's nervous system.

Since GdFFD, but not GFFD, is capable of activating apALNR and mediating physiological functions,[4] it is of great interest to identify the factors underpinning the structure-bioactivity relationship. The presence of D-residues in dominantly L-residue peptides can induce conformational preferences through local constraints that may not be adopted by homochiral peptides. Notable examples are the D-residue-substituted analogues of the opioid peptide Leu-Enkephalin (YGGFL), which were tested for inhibitory activities toward electrically invoked contractions of mouse *vasa deferentia*. The [D-Ala][2] analogue (YdAFGL) was about ten times more active than the wild-type YGGFL, whereas other analogues had no more than 10% activity of the wild-type.[20] The D-residue substitution was shown to drastically alter the structure and intermolecular interactions of the peptide, and presumably accounts for the dramatic difference in biological activity.[21]

Interactions present in a protein-ligand complex often indicate an enthalpy/entropy compromise. Consequently, determining the energy difference (i.e., strain energy) between the free ligand state and the conformationally restricted bound state has posed a major challenge in predicting ligand activity, especially without *a priori* knowledge on the binding site.[22,23] Perola and Charifson[24] surveyed 150 crystal structures of pharmaceutically relevant protein-ligand complexes and showed that only about 10% of the ligands have calculated strain energies greater than 10 kcal/mol. In some cases, the structure of a free peptide ligand in solution provides valuable information for understanding ligand-receptor interactions and designing analogues with improved potency.[25,26] Our goal here is to make progress in this area by developing a relatively simple method to predict ligand activity based on lowest energy conformations.

Accordingly, we investigated conformational differences between the DAACPs that activate apALNR, those that do not, and their all-L-residue counterparts in their unbound states. Our first objective was to understand the structural differences, and then use them to devise a model with the power to reliably predict the activities of peptide analogues, and finally, rationally design new bioactive peptidic ligands for the same receptor. We utilized a computational modeling workflow that combines replica-exchange molecular dynamics (REMD)[27] simulations with density functional theory (DFT) calculations to determine the global energy minimum structure of each peptide of interest. The model structures were cross-validated with ion-mobility mass spectrometry (IM-MS), which can differentiate peptide epimers that adopt different conformations. A series of bioactive peptides were chosen as a training set to formulate a model that correlates structural data to experimental receptor activation (EC_{50}) values. The model can predict the activity of a series of analogues and provides new insights into the specificity of apALNR toward the endogenous DAACPs, and may be useful for the future design of chemical probes to modulate physiological responses mediated by these ligands.

Materials and Methods

Peptide synthesis and purification

We used the same procedure as previously reported.[3] Briefly, peptides were synthesized by solid-phase peptide synthesis based on Fmoc-protection of the main chain amine and purified by reversed-phase high performance liquid chromatography (HPLC). Final peptide purity was assessed by reversed-phase HPLC and identity was confirmed by matrix-assisted laser desorption/ionization time-of-flight MS.

Computational modeling workflow

Initial peptide conformations were built using the *tleap* module available in the Amber 12 package.[28] The peptide coordinates were then read into GROMACS v4.6.5[29,30] and the topology files were generated using the Amber FF99SB force field.[31] Oda et al.[32] recently showed that Amber FF99SB can be used for accurate calculations of proteins and peptides, including D-amino acids. Each peptide system was then solvated in a cubic water box containing approximately 1,400 TIP3P water molecules[33] under a periodic boundary condition. Positively and negatively charged ions (Na^+ and Cl^-) were added to neutralize the system, which was minimized using the steepest descent algorithm for 3 ns and then subjected to another NVT equilibration for another 3 ns. Initial guesses for temperature values in the T-REMD simulations with 32 replicas were taken from Patriksson and Spoel's temperature predictor[34] and then adjusted to obtain an exchange rate of approximately 25–30%. The temperature values ranged from 268 to 476 K. Each replica was equilibrated at the desired temperature for 6 ns before the production run for T-REMD was begun. Exchanges between replicas were attempted at every 3 ps. The LINCS algorithm[35] was employed to constrain bonds between heavy atoms and hydrogens, and the SETTLE algorithm[36] was used for water molecules. These constraints allow an integration time step of 2.0 fs. Electrostatic and dispersion forces were computed with a real space cut-off of 1.2 nm and the particle mesh Ewald method[37] was used to treat long-range electrostatics. Simulations were performed at neutral pH in which the temperature was maintained by the Nose-Hoover thermostat. The temperature and pressure coupling constants were 0.1 ps and 1.0 ps, respectively. The equations of motion were integrated according to the leap-frog algorithm. The production run was 200-ns long per replica, but only the last 100-ns data were subjected to analysis. The trajectory at 300 K of each peptide was clustered based on the end-to-end distance (d_{ee}) and the distance between the N-terminus (NH_3^+) and carboxylic sidechain of Asp ($d_{N\text{-}Asp}$) into families of folded, partially folded, and unfolded structures. Each family was further clustered using the Daura algorithm[38] available in the *g_cluster* program. A representative structure of each cluster was further subjected to quantum mechanics (QM) treatment using the Gaussian 09

program.[39] Specifically, the optimized geometries and relative energies of each structure in water were calculated at the B3LYP level of theory with the cc-pVDZ basis set, Grimme's dispersion correction GD3,[40] and polarized continuum model (PCM) for implicit water. From there, the global energy minimum structure was determined for each peptide.

IM-MS

Peptide powder was dissolved in water and diluted to a final concentration of 50 μM. Mass spectra and ion-mobility data were collected on a lab-built instrument consisting of an nano-electrospray source, a source funnel, a 2-m long drift cell, an exit funnel, and a quadrupole mass analyzer.[41] In the experiments, ions were generated through the means of nano electrospray ionization, stored in a source funnel, and subsequently pulsed into a drift cell filled with He gas at 10 torr. The ions drift through the cell with a constant velocity as the forces created by a weak electrical field on the ions and the drag force due to collisions with buffer gas molecules cancel each other. Drift velocity can be related to the reduced ion mobility K_0, and used to calculate the experimental collision cross sections σ given in Eq. 1

$$\sigma \approx \frac{(18\pi)^{\frac{1}{2}}}{16}\left[\frac{1}{m}+\frac{1}{m_b}\right]^{\frac{1}{2}}\frac{ze}{(k_BT)^{\frac{1}{2}}}\frac{1}{K_0}\frac{1}{N} \quad (Eq.1)$$

where m and m_b are the molecular weights of the ions and buffer gas molecules, respectively, ze is the charge of the ion, N is the buffer gas density.[42]

apALNR activation assays

We used the same assay as previously described in Checco et al.[3] The specific peptides tested in this study include GdFdFD, GdFVD, dPdFFD, PdFFD, PdFAD, dPdFFDGG and Aib-dFFD. In each test, GdFFD was used as a control. Other EC_{50} data were obtained from Checco et al.[3] CHO-K1 cells (ATCC, CLL-61) were transiently transfected with plasmids encoding for *apALNR* (in pcDNA3.1 (+)) and *Gα-16* (in pcDNA3.1 (+)) using the transfection reagent Turbofect (Thermo Fisher Scientific). After exposure to potential agonist peptides for 1 h, activation of apALNR was detected by monitoring IP1 accumulation using an IPOne Detection Kit (Cisbio, 62IPAPEB).

Results and Discussion

The lowest energy conformations of GdFFD and GFFD are structurally distinct, as supported by IM-MS cross section measurements

Ideally, one might be able to "fold" any peptide correctly using molecular dynamics (MD) if the system was simulated with a perfect force field and for an infinite amount of time. However, there is not yet a perfect force field and even the millisecond time scale is still not routinely accessed.[43-45] While proper sampling of the conformational landscape will undoubtedly benefit from high-level MD techniques such as REMD, QM calculations on the resulting structures are often necessary. It has also been shown on several occasions that QM refinement of ligand structures can substantially reduce conformational strain.[8,46,47] In all cases, experimental validation is invaluable. Traditional approaches to structure determination include NMR and X-ray crystallography, which are capable of providing atomistic models. However, data collection and structure refinement remain a bottleneck for studies that require a large number of analyses of similar peptides. IM-MS provides an alternative as the data can be collected in minutes or hours, and the collisional cross section (CCS) provides a coarse measurement of a molecule's size and shape.[48-50]

Since our modeling approach (see Supporting Information Figure S1) can sample multiple conformations for each peptide, for simplicity, we only refer to the peptide in its lowest energy conformation obtained from our modeling workflow (see Materials and Methods and Figure S1), unless explicitly stated otherwise. Figure 1A shows the predicted lowest energy structures of GdFFD and GFFD. The Ramachandran map of GdFFD (Figure 1B, bottom panel) is indicative for D-amino acid occurrences (positive φ and negative ψ). Both GFFD and GdFFD adopt compact conformations in which the N-terminus forms salt-bridges with both the C-terminus and the Asp sidechain. This type of interaction is consistent with the fact that removing the charge from either the N- or C-terminus dramatically decreases activity (e.g., the EC_{50} values of Ac-GdFFD and GdFFD-NH_2 are both 60-fold higher than GdFFD).[3] However, the relative positions of the two Phe residues ([D/L-Phe]^2 and [L-[Phe]^3) in these structures are distinct. In GdFFD, the two Phe sidechains are on the opposite sides of the plane created by the backbone atoms (Figure 1A). In GFFD, the two Phe sidechains are on the same side of the plane (Figure 1A). Furthermore, the [L-Phe]^3 sidechain in GdFFD projects toward the termini whereas that same residue in GFFD points to the opposite direction, away from the termini. Although the differences in conformation and shape of the overall molecules may account for the receptor specificity toward GdFFD but not GFFD, it is necessary to experimentally validate that the structures are reasonable.

We used IM-MS to complement our simulation results. IM-MS structurally characterizes biological molecules via measurements of CCS, σ, a quantity that is dependent on the conformation of the molecule in the gas phase.[50-52] Under some conditions, the solution-phase structures can be kinetically trapped after dehydration, allowing direct comparison to structures in solution. However, great care must be taken to treat the ions gently and the native charge states present in solution must be utilized. Our drift-tube IM-MS measurements using a lab-built instrument, with high mobility resolution and gentle conditions at the source,[41] allow baseline separation of conformers with cross section differences greater than 1%. This instrument offers a resolving power comparable to trapped ion mobility spectrometry (TIMS)[53] and higher than traveling-wave IM-MS;[54,55] these two IM-MS technologies have been recently utilized to study similar systems. The drift-tube IM-MS instrument used here

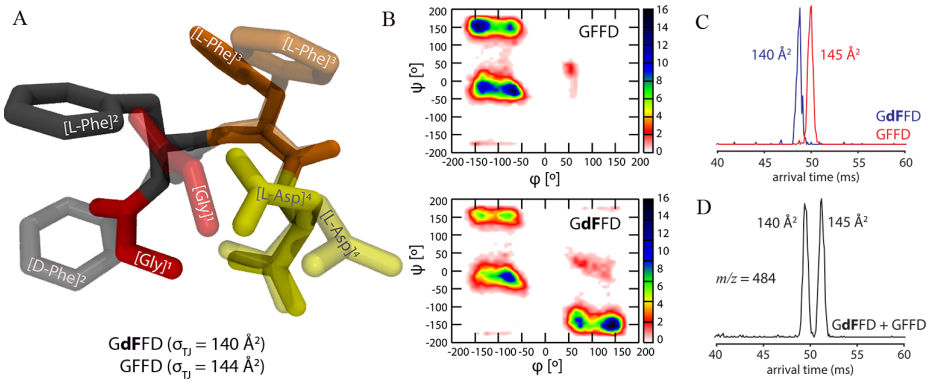


Figure 1. (A) Overlaid structures of GdFFD (semi-transparent) and GFFD (solid). Hydrogens are omitted for clarity. (B) Ramachandran plots of GFFD and GdFFD structures obtained from T-REMD. (C) ATDs of singly charged species (z = -1, m/z = 484) of separate samples of the peptides GdFFD (blue) and GFFD (red). (D) An ATD of an equimolar mixture of GdFFD and GFFD. The peptide concentration is 50 μM in water. σ_{TJ} is the theoretical CCS obtained from the trajectory method.[1,2]

also uses a gentle source condition and has reliable performance in negative polarity, and so may be particularly well-suited for the analysis of short anionic peptides such as GdFFD.

The mass spectra of GdFFD, GFFD, and their mixture in negative polarity show two peaks corresponding to singly (n/z = 1/1; where n denotes the number of peptide molecules) and doubly charged species (n/z = 1/2) (see Supporting Information Figure S2). The experimental CCS σ is an intrinsic value that can be directly compared with the theoretical CCSs of the model structures.[49,56] The theoretical CCSs of our predicted GdFFD and GFFD structures, computed using the trajectory method (TJ),[1,2] was 140 Å^2 for GdFFD (Figure 1) and 144 Å^2 for GFFD. Consistent with these theoretical CCSs, the arrival time distribution (ATD) of the singly charged (z = -1, the natural charge state in solution) GdFFD shows a single species with a CCS of 140 Å^2, while the ATD of the singly charged GFFD shows a single species with a CCS of 145 Å^2 (Figure 1B). Furthermore, the ATD of a 1:1 mixture of GdFFD and GFFD shows base-line separation of the two features (Figure 1C). For GdFFD, only the ground state theoretical structure matched the experiment at a CCS of 140 ± 1 Å^2, strongly supporting the lowest energy conformation (shown in Figure 1A). For GFFD (σ_{exp} = 145 Å^2), there were two higher-energy structures (ΔE = 1.3 and 2.4 kcal/mol; Figure S3A-B) with theoretical CCSs within 145 ± 1 Å^2. However, when we overlaid those structures with the lowest energy structure of GFFD, shown in Figure 1B, these somewhat higher-energy conformations were essentially identical to the lowest energy conformation, with a minor difference in the location of the first residue (Figure S3C, D).

The structure of GdFAD (also known as achatin I), a homologue of GdFFD, has been solved using X-ray crystallography by Kim et al.[57] The X-ray structure shows a bent conformation resembling a cyclic conformation, similar to the structure predicted by our modeling approach. While the sidechain positions of [D-Phe]^2 and [L-Asp]^4 are slightly different, both agree on the proximity between the amino group at the N-terminus and the β-carboxyl group at the C-terminus. Interestingly, the crystal structure for GFAD shows that this peptide adopts an anti-parallel β-pleated sheet structure in the crystal.[58] As such, GFAD may prefer an aggregation state in solution and not exist primarily as a monomer. Our modeling workflow focused on the monomer state of the peptides; hence, it was unable to capture the structures of oligomers. Chiral substitutions have been shown to affect the ability of peptides to interact. Bleiholder et al.[21] showed that oligomer formation is abundant for Enkephalin's YAGFL but greatly diminished for the heterochiral YdAGFdL. Therefore, the incorporation of a D-residue may enhance the peptide potency by maintaining its monomer state in solution, although in a few other cases, it may cause misfolding and aggregation. Overall, the excellent agreement between experimental and theoretical CCSs, along with the similar structure obtained from crystallographic methods for a similar peptide, especially for GdFFD, suggest that the model structures from our modeling workflow are reasonable.

Assessment of structure-activity relationship for peptides in the training set: a model to predict EC_{50} values of GdFFD analogues

Bioisosterism is a commonly employed strategy in medicinal chemistry for the rational design of new drugs through molecular modification of the lead compound.[59,60] To an extent, one can assume molecules of similar size and shape are likely to show similar activity towards the same target macromolecule. We aimed to develop a model to predict peptide activity at apALNR based on the simulated conformation of the ligand. To accomplish this goal, we chose a small set of peptides with known potencies at apALNR (Table 1) as a training set to develop a predictive model. This minimal set of five peptides spans a range of activity from the strongly active peptides GdFFD, GdYFD, and dAdFFD (EC50 < 100 nM) to the moderately active peptides GdFAD and dKdFFD (EC50 < 1000 nM).[3] Although the activity of these peptides at apALNR were recently evaluated,[3] their molecular conformations in solution have yet to be characterized.

Table 1. Peptides in the training set with experimental EC_{50} obtained from the cell-based assays and feeding circuit activity.

	Peptides	ΔE_{scaled} (kcal/mol)	EC_{50} (nM)[a]	*Max response[a] (%)*	*Feeding circuit activity[b]*
Training Set	GdFFD	0	30	100	Active
	GdYFD	0	30	100	Active
	dAdFFD	3.50	80	73	Active
	GdFAD	10.35	400	70	Active
	dKdFFD	11.82	600	70	*not tested*

[a] Activation of apALNR from Checco et al.[3]
[b] Feeding circuit activity data from Bai et al.,[4] Livnat et al.,[7] and Checco et al.[3]

Our approach to evaluating the conformation-dependent activity of the five peptides in the training set is shown in Figure 2A, B. First, we used the lowest energy structures of the two most active peptides GdFFD and GdYFD (which are nearly identical except for the presence of the hydroxyl group of Tyr) as the benchmark structures for comparison. For this analysis, we assumed that the energetic difference between the lowest energy structures of GdFFD/GdYFD and the conformation required to activate apALNR is small. Furthermore, we assumed that analogues would have to adopt similar conformations as GdFFD/GdYFD to activate apALNR. To appropriately compare the energy difference between the predicted lowest energy conformation for a given analogue

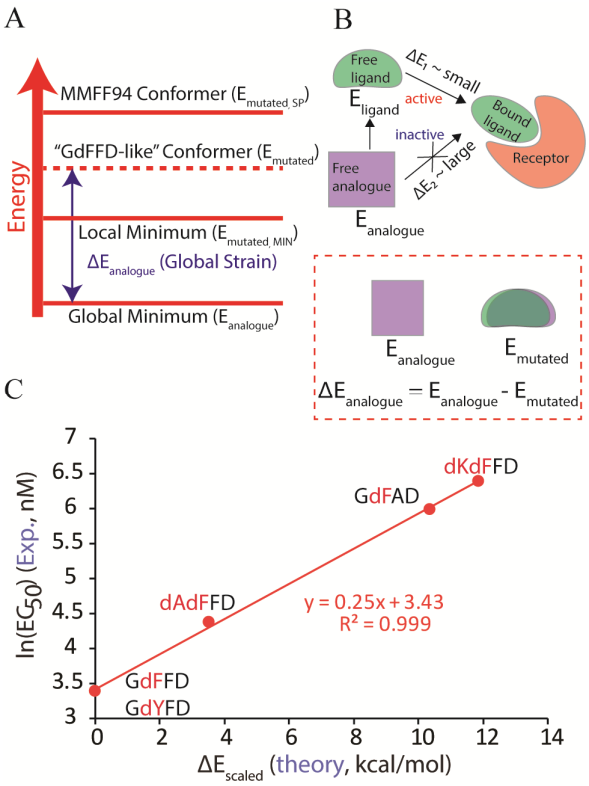


Figure 2. (A) A schematic illustration of the energy differences between structures used to obtain ΔE_{analog}, as discussed in the main text. (B) Cartoon description of ΔE_{analog}. (C) The linear fit model correlates theoretical ΔE_{scaled} to experimental $\ln(EC_{50})$ for the five peptides in the training set (from Table 1).

with the active conformation of GdFFD/GdYFD, we compared the lowest energy conformation of a given analogue (with energy $E_{analogue}$) with that of the analogue adopting a fictitious, "GdFFD-like" conformation with energy $E_{mutated}$ (Figure 2A).

To illustrate the calculation of $E_{mutated}$, we discuss dKdFFD as an example. The ideal "GdFFD-like" conformer of dKdFFD would have a minimal deviation in atom coordinates from GdFFD, with the lysine sidechain in an optimized position. Thus, starting from GdFFD, we first generated a dKdFFD conformer with a D-Lys sidechain at position 1 in a random conformation, then optimized the dKdFFD structure using the MMFF94 force field. We performed a single-point energy calculation on the resulting structure to obtain $E_{mutated,SP}$. This energy is expected to be higher than $E_{mutated}$ because the structure was not geometrically optimized to a local energy minimum. We then performed QM geometry optimization of the resulting structure to obtain its local minimum energy $E_{mutated,MIN}$. This energy is expected to be lower than $E_{mutated}$ since the entire structure was fully optimized, including the backbone atoms. As a result, we estimated $E_{mutated}$ by taking an average of $E_{mutated,MIN}$ and $E_{mutated,SP}$, as shown schematically in Figure 2A. This approach was also applied for GdFAD and dAdFFD in the training set and all analogues in the test set discussed below. While this approach appears to be *ad hoc*, it allows us to quickly and systematically approximate the energies of GdFFD-like conformations for all analogues.

After calculating $E_{mutated}$ for a given analogue of interest, the $\Delta E_{analogue}$ value is then defined as the energy difference between the global energy minimum structure of an analogue $E_{analogue}$ and $E_{mutated}$ (see Figure 2A):

$$\Delta E_{analogue} = E_{analogue} - E_{mutated} \quad \text{(Eq. 2)}$$

Furthermore, because different peptides would have different numbers of atoms, bonds, angles, etc., it is necessary to normalize $\Delta E_{analogue}$ to the same scale, as in Eq. 3.

$$\Delta E_{scaled\ (to\ GdFFD)} = \Delta E_{analogue} \times {E_{GdFFD}}/{E_{analogue}} \quad \text{(Eq. 3)}$$

The model that relates ΔE_{scaled} to $\ln(EC_{50})$ is shown in Figure 2C. Interestingly, a simple linear fit provides an excellent correlation between these two quantities ($R^2 \approx 1$). Note that the fit only includes peptides that are active in cell-based assays. Peptides that showed no activity in our cell-based receptor activation assays (such as GFFD), and thus have no EC_{50} value, cannot be plotted in this correlation.

The potential mean force (PMF) of the GdFFD structures obtained from REMD simulation is shown in Figure 3A. Aside from a dominant population of compact structures, a considerable number of structures are not folded. It is possible that the active conformation of GdFFD may be different than its lowest energy structure. To test the hypothesis that the active conformation is indeed a folded, cyclic-like structure, we

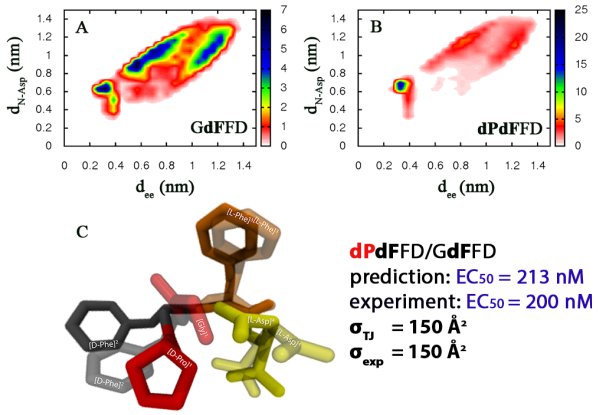


Figure 3. Potential mean force of (A) GdFFD and (B) dPdFFD obtained from REMD simulations. d_{ee} denotes the end-to-end distance. $d_{N\text{-}Asp}$ denotes the distance between the N-terminus (NH_3^+) and the carboxylic sidechain of [L-Asp][4]. Compact structures have small d_{ee} and $d_{N\text{-}Asp.}$ (C) An image of aligned structures of dPdFFD and GdFFD. GdFFD is shown in lighter colors. The theoretical and experimental CCSs of dPdFFD obtained from negative mode IM-MS are also shown.

evaluated dPdFFD. Proline residues contribute positively to the stability of turn-like structures,[61] although in most cases the proline residue is located in the second position. The PMF of dPdFFD (Figure 3B) indicates that a considerably higher percentage of structures adopt a folded, cyclic-like conformation than GdFFD (see also Figure S4 for IM-MS data that show CCS agreement between theory and experiment). Figure 3C shows an overlaid image of dPdFFD onto GdFFD. Because dPdFFD adopts a similar shape to GdFFD (backbone root-mean squared deviation (RMSD) = 0.69 Å), we predict that dPdFFD should be active. Our model calculates a ΔE_{scaled} value of 7.7 kcal/mol for dPdFFD and predicts an EC_{50} of 213 nM, which is in good agreement with the experimental EC_{50} of 200 nM. The experimental data indicate that dPdFFD has an activity slightly weaker than GdFFD, GdYFD, and dAdFFD, and the maximum response is lower (see Supporting Information Figure S5). The high propensity for dPdFFD to stabilize cyclic-like structures (Figure 3B) may compensate for the small mismatch in backbone atoms of dPdFFD and GdFFD (Figure 3C) (which will be discussed in the next section), resulting in a relatively low EC_{50} value.

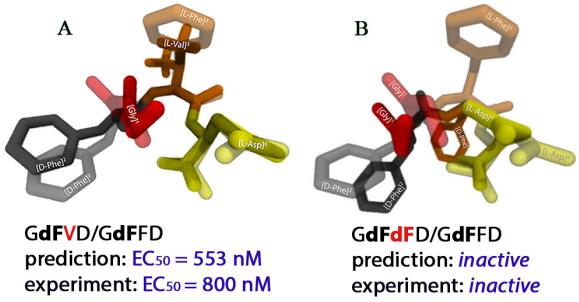
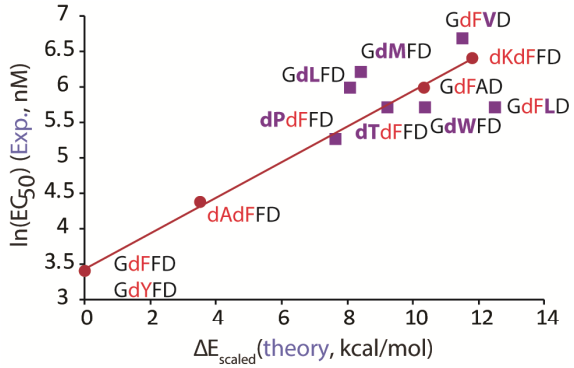


Figure 4. Images of backbone-aligned structures of (A) GdFVD and GdFFD, or (B) GdFdFD and GdFFD. GdFFD is shown in lighter colors.

We note that the ability to adopt the cyclic-like structures may be necessary but not sufficient to activate the receptor. Supporting Information Figure S6 shows the percentages of cyclic-like structures for GdFFD analogues that are found to be active at the receptor. This low-resolution parameter weakly correlates to experimental EC_{50}, indicating that other factors, such as the positions of sidechain atoms, are also critical to activity.

GdFVD was not tested in the previous report[3] and is another suitable candidate for assessing our model (Figure 4A). The observation that GdFAD is active at the receptor[3] indicates that the identity of the sidechain at position 3 is not critical for activity. Thus, one might predict that GdFVD may also be active. However, unlike [L-Phe][3] in GdFFD or [L-Ala][3] in GdFAD, [L-Val][3] in GdFVD is a β-branched residue, which can dramatically affect the conformational preferences of the resulting peptide.[62] We determined ΔE_{scaled} of GdFVD, using the protocol described above, to be 11.51 kcal/mol. The predicted EC_{50} based on the model is 553 nM, which qualitatively agrees with the experimental value of 800 nM. The same approach was utilized to obtain ΔE_{scaled} to "predict" EC_{50} for several other peptides that were previously evaluated,[3] including dTdFFD, GdMFD, GdLFD, GdWFD, and GdFLD (see Figure 5). Overall, we obtained a positive correlation between $\ln(EC_{50})$ and ΔE_{scaled} from our model. In particular, the ability of the model to accurately predict the difference in potency between dPdFFD and dTdFFD, despite both peptides featuring a D-residue at position 1, suggests the model can discriminate peptides based on their predicted conformations.

The active peptides plotted in Figure 5 are all predicted to adopt cyclic-like conformations similar to GdFFD. However, major changes in the primary sequence can dramatically alter the overall conformation of the peptide to the point where it is unable to adopt the predicted "active-like" conformation of GdFFD. Based on this assumption, we predicted that GdFdFD would be inactive (Figure 4B), and in fact, GdFdFD showed no activity in our cell-based assays (see Supporting Information Figure S5A).



Peptides	$\Delta E^{(kcal/mol)}_{scaled}$	ln (EC_{50}), exp.	EC_{50}, theo. (nM)	EC_{50}, exp. (nM)
dPdFFD	7.70	4.09	213	200
GdLFD	8.09	5.99	234	400
GdMFD	8.42	6.21	254	500
dTdFFD	9.23	5.70	310	300
GdWFD	10.39	5.70	417	300
GdFVD	11.51	6.68	553	800
GdFLD	12.50	5.70	708	300

Figure 5. Linear correlation between theoretical ΔE_{scaled} and experimental $\ln(EC_{50})$ for some GdFFD analogues. The circles and squares represent the peptides in the training set and test set, respectively. The purple letters highlight the single-residue substitution in each GdFFD analogue. See Figure S5 for all dose-response curves used to calculate EC_{50}, exp.

[L-Ala]^1 vs. [D-Ala]^1: stereochemistry at position 1 affects peptide activity

We next examined the peptide AdFFD. Figure 6 shows overlaid images of dAdFFD (panel A) and AdFFD (panel B) onto GdFFD. Previous experiments showed that only dAdFFD activates apALNR, while AdFFD does not.[3] Interestingly, the same study has also shown that AdFFD is physiologically weakly active in the feeding circuit of *Aplysia*,[3] although whether this activity stems from activation of apALNR or some other mechanism is not clear. As an experimental test of our predicted structures, we analyzed dAdFFD and AdFFD by IM-MS. In contrast to GdFFD and GFFD discussed above, we were unable to obtain adequate signals for dAdFFD or AdFFD in negative polarity. However, we were able to collect mass spectra and ATDs for dAdFFD and AdFFD in positive polarity, and the experimental singly charged CCSs agreed with the theoretical CCSs (Figure 6C, D), suggesting our modeled structures are reasonable.

Interestingly, we also observed the formation of oligomers for both peptides (see Supporting Information Figure S7 for representative mass spectra and other ATDs that show large oligomers). With IM-MS, oligomers having the same *m/z* ratio can be unambiguously identified due to their difference in mobility (for example, see Bernstein et al.[63,64]). Figure S7, panels E and F, compare the relative abundance of the peptide oligomers of dAdFFD and AdFFD. For AdFFD, the distribution is shifted toward larger oligomers (*n* = 6 and 8) whereas for dAdFFD, the lower-order oligomers (*n* = 2 and 4) are dominant.

Interestingly, Li and co-workers[55] also detected oligomer formation for DAACPs, including GFAD and GdFAD in positive polarity.

The discrepancy in the predicted activity for AdFFD (EC_{50}, theo. = 520 nM) and its experimental activity (EC_{50}, exp. > 500,000 nM) highlights the important limitations of our approach. Our model assumes that the lowest energy solution-state conformation of GdFFD is similar to the active conformation but does not consider how major modifications may negatively affect interactions within the receptor's ligand-binding site or other intermolecular interactions that may take place. For example, a ligand may be predicted to adopt an overall "active-like" backbone conformation in solution, but the position of sidechains may lead to steric interactions in the ligand binding site that disrupt productive activation of the receptor. The lowest energy structure of each peptide obtained from our modeling workflow is predicted to be the most probable structure existing in solution and initially interacts with the receptor. In the case of dAdFFD versus AdFFD (see Figure 6A, B), our modeling predicts that both peptides are able to adopt "active-like" conformations (i.e., the same conformation as GdFFD), but the inability of AdFFD to activate apALNR suggests two possibilities: (a) the projection of the [L-Ala]^1 sidechain may be involved in disruptive steric interactions in the ligand-binding site; or (b) there is a mismatch in sidechain orientation for one or more of residues 2–4. By comparing the structures of active peptides

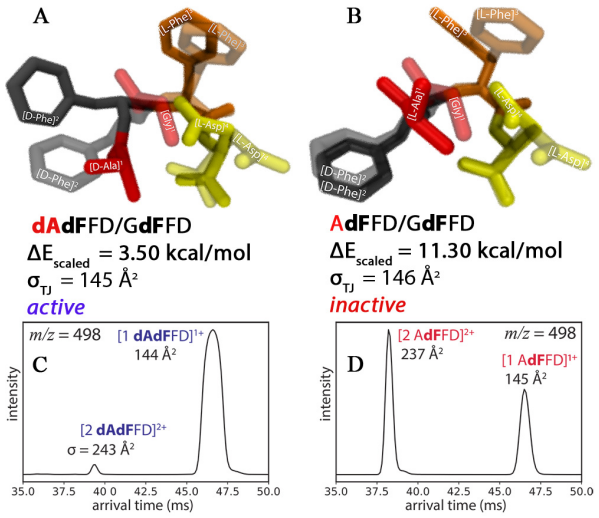


Figure 6. Aligned structures of (A) dAdFFD and GdFFD, or (B) AdFFD and GdFFD, (shown in lighter colors). The ΔE_{scaled} values are also listed together with the theoretical CCSs. (C-D) Representative ATDs collected at *m/z* 498 showing singly charged monomers and doubly charged dimers of dAdFFD and AdFFD. The data were collected in positive polarity at peptide concentration of 50 μM in water.

such as dPdFFD and dAdFFD to GdFFD (Figures 3C and 6A), we observed that minor differences in sidechain orientations of the second (D-Phe) and fourth (L-Asp) residues did not reduce peptide activity at the receptor. Since substitutions at position 3 of the peptides decreased activity at the receptor (e.g., GdFLD, GdFVD), we suggest that the mismatch in sidechain

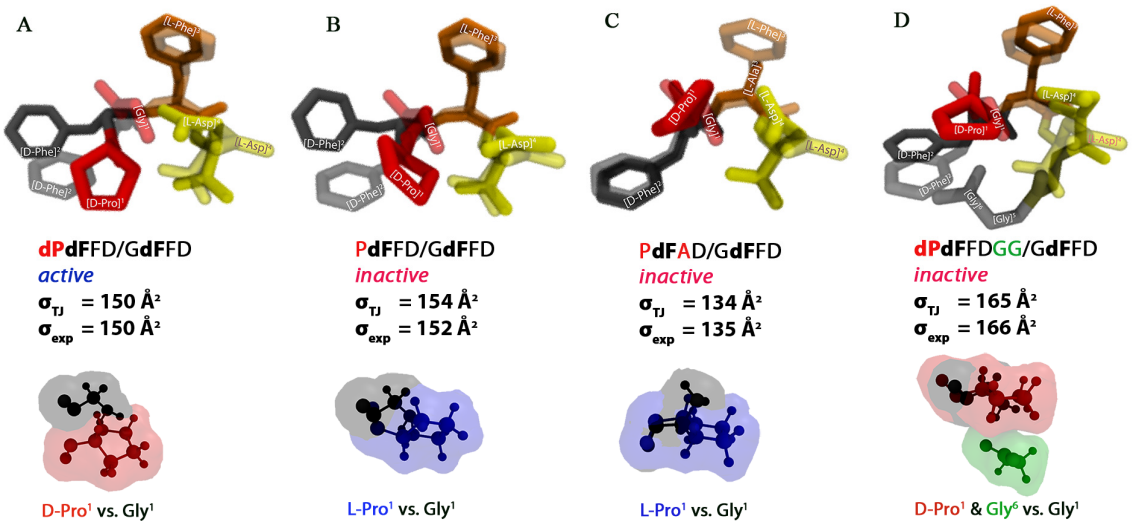


Figure 7. Images of backbone aligned (A) dPdFFD, (B) PdFFD, (C) PdFAD and (D) dPdFFDGG onto GdFFD (shown in lighter colors). The theoretical and experimental CCSs of each peptide are also shown. IM-MS experiments were performed in negative polarity at 50 μM peptide in water. The bottom panels show the position of the first residue in each peptide relative to $[Gly]^1$ in GdFFD. For dPdFFDGG, the C-terminal glycine is also shown.

orientation at position 3 is not the major cause of AdFFD inactivity. To illustrate this point, we tested PdFAD. This peptide does not have a bulky sidechain at position 3. Our cell-based assay showed that PdFAD is inactive (Figure S5C), whereas GdFAD and dPdFFD are both active, as mentioned above. Collectively, the data suggest that the presence of an L- or D-residue other than glycine at position 1 leads to subtle differences in peptide conformation that dramatically affect peptide activity at the receptor.

To investigate the effect of stereochemistry at position 1, we compared the lowest energy conformations of GdFFD with analogues bearing D- or L-residues at position 1 and two additional peptides, PdFFD and Aib-dFFD (where Aib is 2-methylalanine). When aligned by their backbone with GdFFD, we see that D-residues placed in position 1 (e.g., dPdFFD, Figure 7A; and dTdFFD, dKdFFD, Figure S8) occupy the space below that of the glycine in native GdFFD while L-residues at position 1 overlap well with $[Gly]^1$ of GdFFD when the peptide structures are aligned via backbone RMSD. PdFFD, PdFAD, and Aib-dFFD, which project sidechains in an "L-residue-like" manner, are unable to activate apALNR despite adopting conformations similar to GdFFD (Figure 7B, C and Supporting Information Figures S9–11 and S5C), suggesting that the presence of sidechain atoms of the L-residue can cause steric effects, preventing the peptide from making contact with the receptor's residues in the ligand-binding site.

The predicted structure of dPdFFDGG aligns well with the "active-like" conformation of GdFFD (Figures 7D and S12) in good agreement with the experimental CCS (Figure S5C) but was found to be inactive at apALNR. Apparently, the additional Gly residues (relative to dPdFFD) force D-Pro at position 1 to relocate, creating a disruptive steric effect similar to the case when the first residue is a L-residue. Another plausible reason for the inactivity is that the position of the C-terminal charge for dPdFFDGG, which is altered relative to GdFFD, negatively affects receptor activation. This hypothesis is supported by the fact that GdFFD-NH_2, in which the C-terminal charge is removed, is a significantly weaker agonist (by about two orders of magnitude) than GdFFD for activating apALNR.[3] Such molecular interactions could not have been predicted based on our shape-based model.

Summary and Conclusions

It has been shown previously that peptide analogues containing D-residues can display increased potency[20,65] and stability to proteases[3,66] relative to their all-L-residue counterparts. GdFFD and GdYFD are cell-to-cell signaling peptides that act in the CNS of *Aplysia* and activate apALNR. We report the first investigation of the relationships that link the conformations of GdFFD, GdYFD, and several analogues to their abilities to activate apALNR using a combination of computational modeling, IM-MS, and cell-based receptor activation assays. We constructed a simple model to predict the potency of peptide analogues for apALNR by examining the overall backbone conformation (RMSD relative to GdFFD) and correlating the global strain energy ΔE_{scaled} approximated by theory and the experimental $\ln(EC_{50})$ obtained from our cell-based assays. The global strain energy was calculated based on the lowest energy structure obtained from REMD followed by DFT optimization, and the energy of a hypothetical structure with similar conformation as the endogenous ligand GdFFD/GdYFD. We show that although the structure of the ligand bound to the receptor is not available, we can still approximate the energies of active conformers and use them to predict activities of several peptide analogues.

Our modeling approach allowed us to gain insight into the conformational space likely adopted by GdFFD and to correctly

predict the activity of several GdFFD analogues. Limitations in our predictive power were highlighted by analogues that dramatically deviated from the core backbone of GdFFD, such as those that altered stereochemistry or added additional residues. Thus, while our models account for intramolecular interactions that influence a peptide's conformation in solution, they do not account for intermolecular interactions with binding partners like the receptor. In addition, although examining the lowest energy conformation can provide valuable insight into activity,[25,26] we expect that a more sophisticated model, accounting for not only the lowest energy structure but also the relative distributions among stable conformations, will make the predictions more accurate. Such information can be obtained from conformational-sensitive techniques such as NMR or gas-phase infrared spectroscopy coupled with high-resolution IM-MS.[67-70] Nevertheless, approaches similar to ours may be effective for quickly exploring a variety of simple substitutions in peptide ligands for a given receptor, even in the absence of high-resolution structural information of the receptor or the ligand-receptor complex. Furthermore, information detailing the active conformation of specific ligands like GdFFD aids in the design of small molecule or peptide agonists or antagonists as chemical probes for receptor signaling. Such probes would be useful as physiological tools to interrogate peptide activity in *in vitro* and *in vivo* experiments in cases where little is known about the receptor.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We gratefully acknowledge support from the National Institutes of Health, Award No. P30 DA018310 from the National Institute on Drug Abuse and Award No. R01 NS031609 from the National Institute of Neurological Disorders and Stroke (both to J.V.S.), and the National Science Foundation, Award Nos. CHE-1565941 (M.T.B.) and MCB-1716956 (J.-E.S.). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources through the XSEDE grant number TG-CHE150069 (J.V.S.) and TG-MCA05S027 (J.-E.S.). We acknowledge support from the Center for Scientific Computing at the CNSI and MRL at the University of California at Santa Barbara: an NSF MRSEC (DMR-1121053) and NSF CNS-0960316. J.W.C. was supported in part by a Beckman Institute Postdoctoral Fellowship, funded by a Beckman Foundation gift to the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

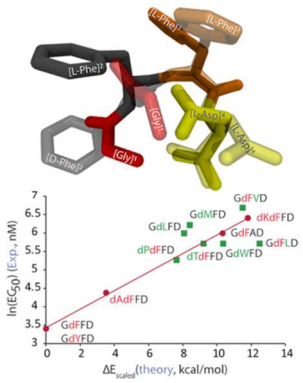
Author Information

***Corresponding authors:**
Jonathan V. Sweedler. Email: jsweedle@illinois.edu
Thanh D. Do. Email: thanhddo@illinois.edu

Notes and references

1. M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz and M. F. Jarrold, *J. Phys. Chem.*, 1996, 100, 16082-16086.
2. A. A. Shvartsburg and M. F. Jarrold, *Chem. Phys. Lett.*, 1996, 261, 86-91.
3. J. W. Checco, G. Zhang, W. Yuan, K. Yu, S. Yin, R. H. Roberts-Galbraith, P. M. Yau, E. V. Romanova, J. Jing and J. V. Sweedler, *ACS Chem. Biol.*, 2018, 13, 1343-1352.
4. L. Bai, I. Livnat, E. V. Romanova, V. Alexeeva, P. M. Yau, F. S. Vilim, K. R. Weiss, J. Jing and J. V. Sweedler, *J. Biol. Chem.*, 2013, 288, 32837-32851.
5. E. Persch, O. Dumele and F. Diederich, *Angew. Chem.*, 2015, 54, 3290-3327.
6. R. Baron and J. A. McCammon, *Annu. Rev. Phys. Chem.*, 2013, 64, 151-175.
7. I. Livnat, H. C. Tai, E. T. Jansson, L. Bai, E. V. Romanova, T. T. Chen, K. Yu, S. A. Chen, Y. Zhang, Z. Y. Wang, D. D. Liu, K. R. Weiss, J. Jing and J. V. Sweedler, *Anal. Chem.*, 2016, 88, 11868-11876.
8. G. L. Warren, T. D. Do, B. P. Kelley, A. Nicholls and S. D. Warren, *Drug Discov. Today*, 2012, 17, 1270-1281.
9. J. F. White, N. Noinaj, Y. Shibata, J. Love, B. Kloss, F. Xu, J. Gvozdenovic-Jeremic, P. Shah, J. Shiloach, C. G. Tate and R. Grisshammer, *Nature*, 2012, 490, 508-513.
10. B. E. Krumm and R. Grisshammer, *Front. Pharmacol.*, 2015, 6, 48.
11. R. Chakrabarti, A. M. Klibanov and R. A. Friesner, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, 102, 10153-10158.
12. M. Brylinski and J. Skolnick, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, 105, 129-134.
13. M. Lawrenz, D. Shukla and V. S. Pande, *Sci. Rep.*, 2015, 5, 7918.
14. R. Moretti, B. J. Bender, B. Allison and J. Meiler, *Methods Mol. Biol.*, 2016, 1414, 47-62.
15. M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri and D. R. Koes, *J. Chem. Inf. Model.*, 2017, 57, 942-957.
16. S. L. Swann, S. P. Brown, S. W. Muchmore, H. Patel, P. Merta, J. Locklear and P. J. Hajduk, *J. Med. Chem.*, 2011, 54, 1223-1232.
17. M. McGann, *J. Chem. Inf. Model.*, 2011, 51, 578-596.
18. C. Y. Yang, K. Yu, Y. Wang, S. A. Chen, D. D. Liu, Z. Y. Wang, Y. N. Su, S. Z. Yang, T. T. Chen, I. Livnat, F. S. Vilim, E. C. Cropper, K. R. Weiss, J. V. Sweedler and J. Jing, *PLoS One*, 2016, 11, e0147335.
19. P. Bauknecht and G. Jekely, *Cell Rep*, 2015, 12, 684-693.
20. D. H. Coy, A. J. Kastin, A. V. Schally, O. Morin, N. G. Caron, F. Labrie, J. M. Walker, R. Fertel, G. G. Berntson and C. A. Sandman, *Biochem Biophys Res Commun*, 1976, 73, 632-638.
21. C. Bleiholder, N. F. Dupuis, M. M. Gessel and M. T. Bowers, *Int. J. Mass Spectrom.*, 2017, 413, 52-60.
22. D. L. Mobley and K. A. Dill, *Structure*, 2009, 17, 489-498.
23. C. D. Blundell, M. J. Packer and A. Almond, *Biorg. Med. Chem.*, 2013, 21, 4976-4987.
24. E. Perola and P. S. Charifson, *J. Med. Chem.*, 2004, 47, 2499-2510.
25. A. S. Edison, E. Espinoza and C. Zachariah, *J. Neurosci.*, 1999, 19, 6318-6326.
26. M. Pinto, C. Rougeot, L. Gracia, M. Rosa, A. Garcia, G. Arsequell, G. Valencia and N. B. Centeno, *ACS Med. Chem. Lett.*, 2012, 3, 20-24.
27. Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, 314, 141-151.
28. D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER 12, University of California, San Francisco*, 2012.
29. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, 26, 1701-1718.
30. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, 4, 435-447.

31. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, 65, 712-725.
32. A. Oda, T. Nakayoshi, S. Fukuyoshi, E. Kurimoto, N. Yamaotsu, S. Hirono and O. Takahashi, *Chirality*, 2018, DOI: 10.1002/chir.22821.
33. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, 79, 926-935.
34. A. Patriksson and D. van der Spoel, *PCCP*, 2008, 10, 2073-2077.
35. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, 18, 1463-1472.
36. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, 13, 952-962.
37. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, 98, 10089-10093.
38. Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F. van Gunsteren and A. E. Mark, *Angew. Chem.*, 1999, 38, 236-240.
39. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT*, 2016.
40. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, 32, 1456-1465.
41. P. R. Kemper, N. F. Dupuis and M. T. Bowers, *Int. J. Mass Spectrom.*, 2009, 287, 46-57.
42. J. Gidden, A. Ferzoco, E. S. Baker and M. T. Bowers, *J. Am. Chem. Soc.*, 2004, 126, 15132-15140.
43. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, *Science*, 2010, 330, 341-346.
44. K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, *Science*, 2011, 334, 517-520.
45. T. J. Lane, D. Shukla, K. A. Beauchamp and V. S. Pande, *Curr. Opin. Struct. Biol.*, 2013, 23, 58-65.
46. Z. Fu, X. Li and K. M. M. Jr., *J. Comput. Chem.*, 2011, 32, 2587-2597.
47. C. H. Reynolds, *ACS Med. Chem. Lett.*, 2014, 5, 727-729.
48. C. Bleiholder and M. T. Bowers, *Annu. Rev. Anal. Chem.*, 2017, 10, 365-386.
49. G. Von Helden, M. T. Hsu, N. Gotts and M. T. Bowers, *J. Phys. Chem.*, 1993, 97, 8182-8192.
50. T. Wyttenbach and M. T. Bowers, in *Modern Mass Spectrometry*, ed. C. A. Schalley, Springer-Verlag, Berlin, Heidelberg, 2003, vol. 225, ch. 207-232.
51. T. Wyttenbach, G. von Helden and M. T. Bowers, *J. Am. Chem. Soc.*, 1996, 118, 8355-8364.
52. T. Wyttenbach, J. E. Bushnell and M. T. Bowers, *J. Am. Chem. Soc.*, 1998, 120, 5098-5103.
53. K. Jeanne Dit Fouque, A. Garabedian, J. Porter, M. Baird, X. Pang, T. D. Williams, L. Li, A. Shvartsburg and F. Fernandez-Lima, *Anal. Chem.*, 2017, 89, 11787-11794.
54. C. X. Jia, C. B. Lietz, Q. Yu and L. J. Li, *Anal. Chem.*, 2014, 86, 2972-2981.
55. X. Q. Pang, C. X. Jia, Z. W. Chen and L. J. Li, *J. Am. Soc. Mass Spectrom.*, 2017, 28, 110-118.
56. E. S. Baker, S. L. Bernstein and M. T. Bowers, *J. Am. Soc. Mass Spectrom.*, 2005, 16, 989-997.
57. K. H. Kim, H. Takeuchi, Y. Kamatani, H. Minakata and K. Nomoto, *Life Sci.*, 1991, 48, PL91-96.
58. T. Ishida, Y. In, M. Inoue, Y. Yasuda-Kamatani, H. Minakata, T. Iwashita and K. Nomoto, *FEBS Lett.*, 1992, 307, 253-256.
59. G. A. Patani and E. J. LaVoie, *Chem. Rev.*, 1996, 96, 3147-3176.
60. L. M. Lima and E. J. Barreiro, *Curr. Med. Chem.*, 2005, 12, 23-49.
61. H. Fu, G. R. Grimsley, A. Razvi, J. M. Scholtz and C. N. Pace, *Proteins*, 2009, 77, 491-498.
62. M. B. Swindells, M. W. MacArthur and J. M. Thornton, *Nat. Struct. Biol.*, 1995, 2, 596-603.
63. S. L. Bernstein, T. Wyttenbach, A. Baumketner, J. E. Shea, G. Bitan, D. B. Teplow and M. T. Bowers, *J. Am. Chem. Soc.*, 2005, 127, 2075-2084.
64. S. L. Bernstein, N. F. Dupuis, N. D. Lazo, T. Wyttenbach, M. M. Condron, G. Bitan, D. B. Teplow, J. E. Shea, B. T. Ruotolo, C. V. Robinson and M. T. Bowers, *Nat. Chem.*, 2009, 1, 326-331.
65. R. Maurer, B. H. Gaehwiler, H. H. Buescher, R. C. Hill and D. Roemer, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, 79, 4815-4817.
66. A. E. Rabideau and B. L. Pentelute, *Acs Central Science*, 2015, 1, 423-430.
67. F. Schinle, C. R. Jacob, A. B. Wolk, J. F. Greisch, M. Vonderach, P. Weis, O. Hampe, M. A. Johnson and M. M. Kappes, *J. Phys. Chem. A*, 2014, 118, 8453-8463.
68. A. Y. Pereverzev and O. V. Boyarkin, *PCCP*, 2017, 19, 3468-3472.
69. T. K. Roy, N. S. Nagornova, O. V. Boyarkin and R. B. Gerber, *J. Phys. Chem. A*, 2017, 121, 9401-9408.
70. D. E. Clemmer, D. H. Russell and E. R. Williams, *Acc. Chem. Res.*, 2017, 50, 556-560.



Computational modeling and ion-mobility mass spectrometry are used to understand and predict the activity of endogenous D-amino acid-containing neuropeptides at their cognate receptor.

Conformational Investigation of the Structure – Activity Relationship of GdFFD and Its Analogs on an Achatin-like Neuropeptide Receptor of *Aplysia californica* Involved in the Feeding Circuit

Thanh D. Do,[‡,] James W. Checco,[‡] Michael Tro,[±] Joan-Emma Shea,[¥,±] Michael T. Bowers[±] and Jonathan V. Sweedler[‡,*]*

[‡]Department of Chemistry and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States, and [±]Department of Chemistry and Biochemistry, [¥] Department of Physics, University of California at Santa Barbara, Santa Barbara, Californica 93106, United States.

***** Corresponding authors:
Jonathan V. Sweedler. Email: jsweedle@illinois.edu
Thanh D. Do. Email: thanhddo@illinois.edu

SUPPORTING INFORMATION

Table of Contents

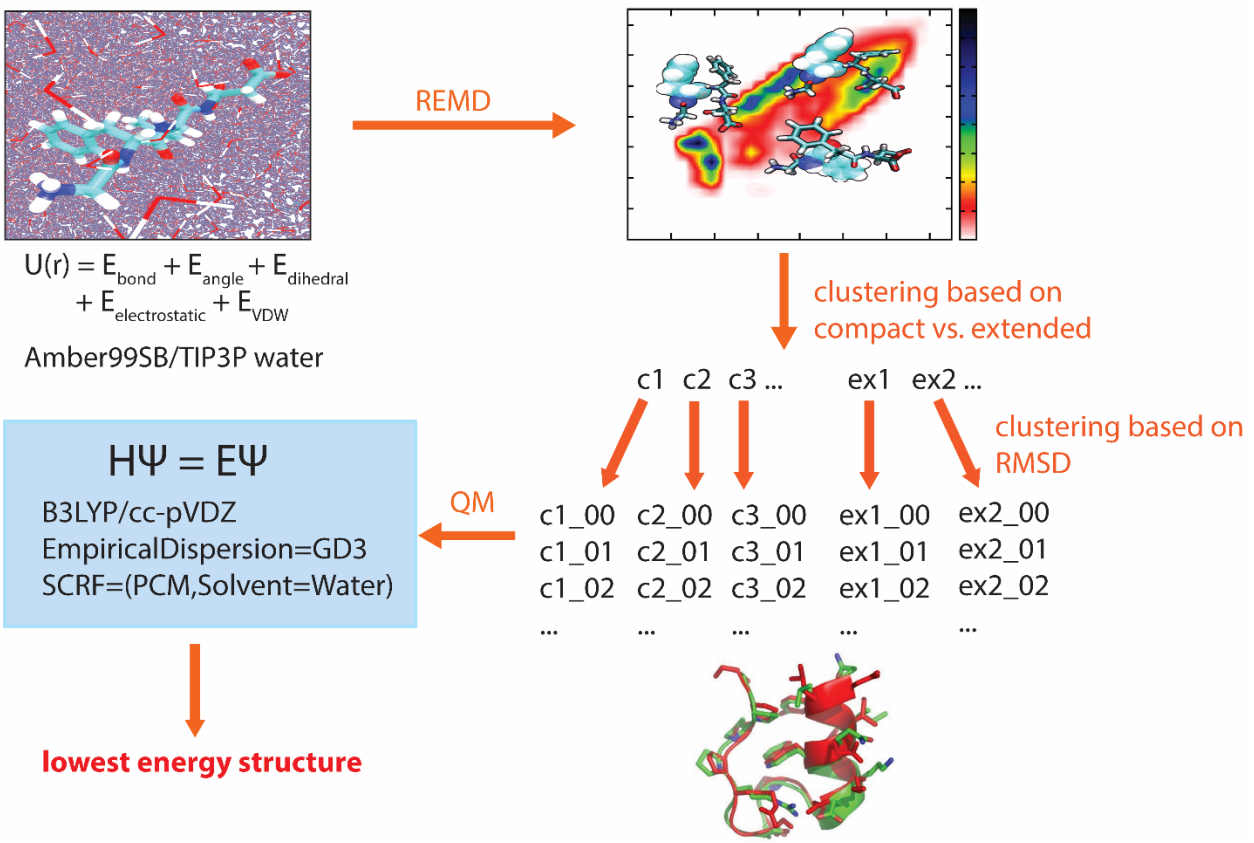


Figure S1. The computational modeling workflow to obtain global energy minimum structures of GdFFD and its analogs. For each peptide, an ensemble of structures was generated through the means of the enhanced sampling T-REMD method. The peptide was simulated using the combination of the Amber ff99SB force field and TIP3P explicit solvent water model. After the REMD simulation reached 200 ns, the trajectory at 300K was clustered into compact and extended families of structures. Each family was then clustered into smaller sub-populations. A representative structure of each subpopulation was subjected to geometry optimization and relative energy calculations using the B3LYP level of theory, the cc-pVDZ basis set, the Grimme's dispersion correction GD3, and the implicit water model PCM. The resulting structures were sorted based on energy to determine the global minimum structure.

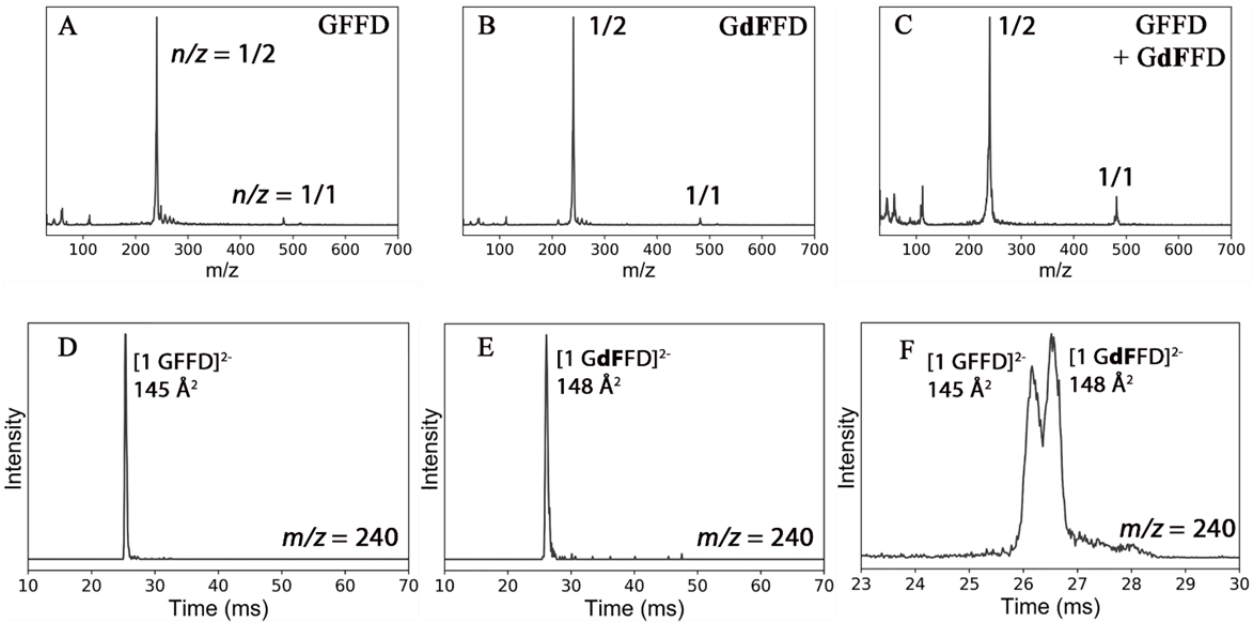


Figure S2. Representative mass spectra of (A) GFFD, (B) G**dF**FD, and (C) equal molar mixture of GFFD and G**dF**FD in water at the concentration of 50 µM. Mass spectral peaks are annotated with *n/z* where *n* is the number of peptide chain and *z* in the number of (negative) charge. (D–F) Representative ATDs of doubly charged species (*z* = -2, *m/z* 240) collected from pure standards of GFFD and G**dF**FD, and the mixture. All data were collected in negative polarity.

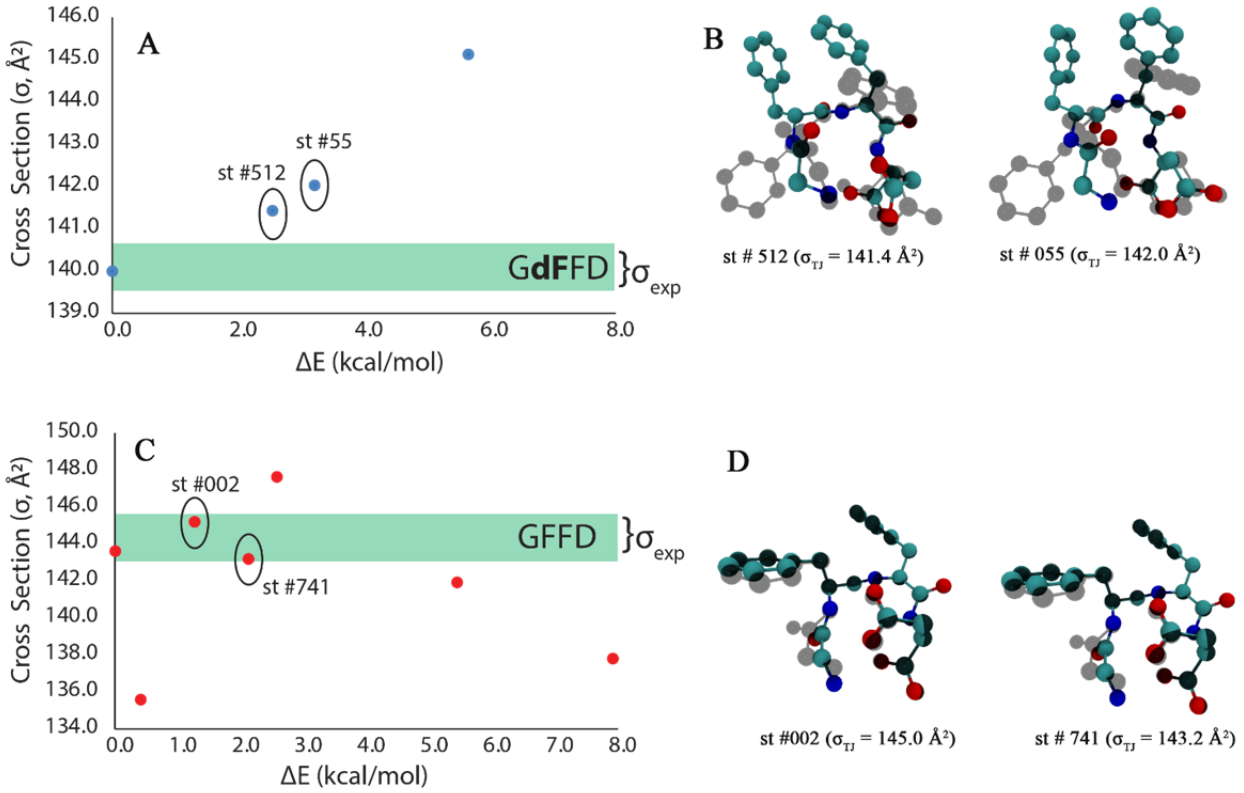


Figure S3. Theoretical CCS vs. ΔE for (A) G**dF**FD and (C) GFFD. (B, D) Overlaid images of structures (shown in colors) with similar CCSs onto the lowest energy structures of GdFFD/GFFD (shown as shadow).

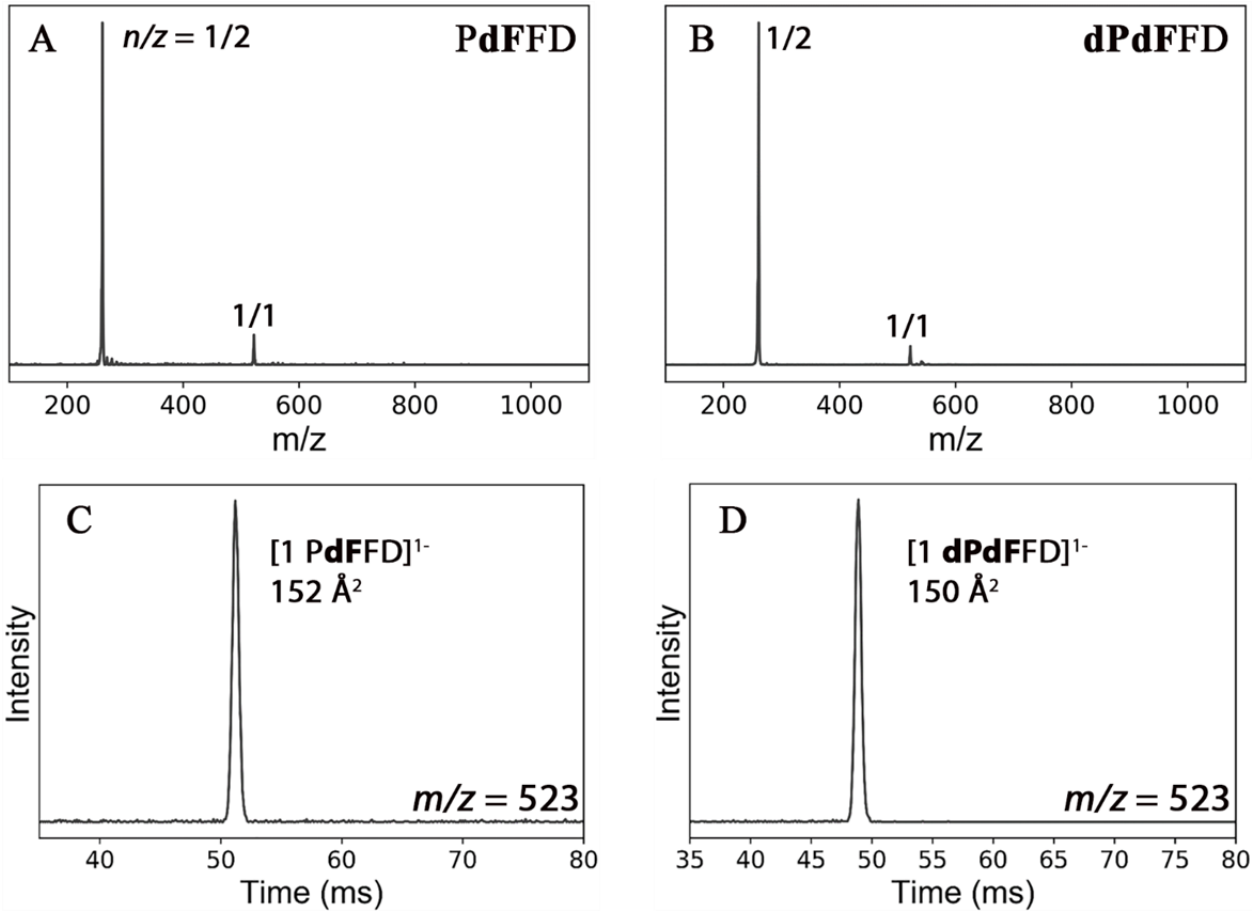


Figure S4. (A, B) Representative mass spectra of P**dF**FD and **dPdF**FD in water. Mass spectral peaks were annotated with *n/z* where *n* the number of peptide chain and *z* is the charge. (C, D) Representative ATDs of singly charged P**dF**FD and **dPdF**FD. The CCSs are also shown.

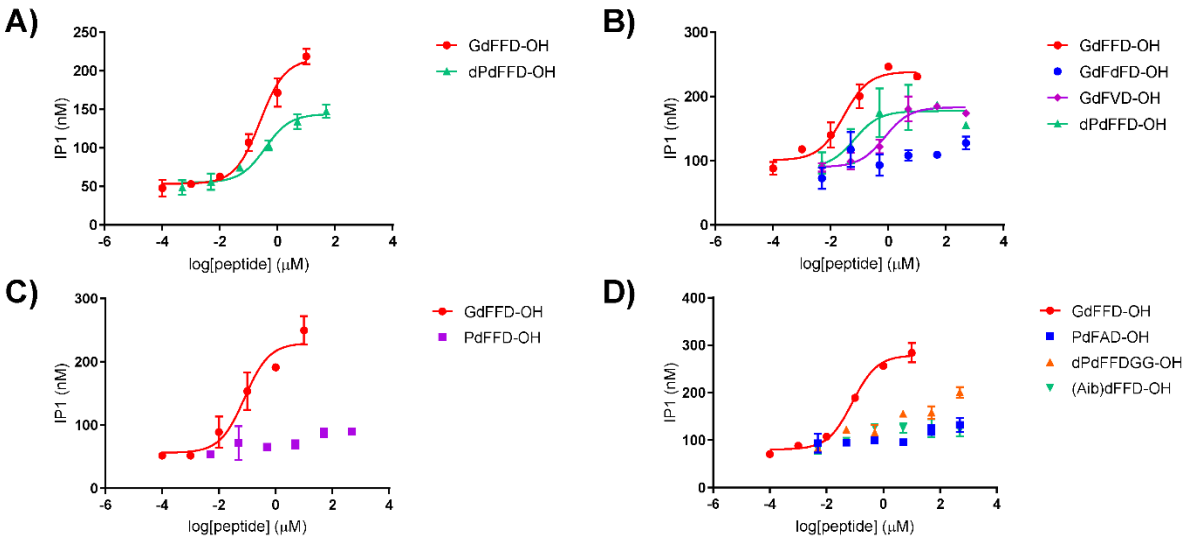


Figure S5. Dose-response curves of peptides tested for activation of apALNR, as determined by IPOne assay. CHO-K1 cells were co-transfected with plasmids for both *apALNR* and *Gα-16*. Each point represents the mean ± SEM from duplicate wells on the plate. Calculated EC_{50} values for each experiment are: (A) GdFFD-OH = 252 nM, dPdFFD-OH = 373 nM; (B) GdFFD-OH = 28 nM, dPdFFD-OH = 66 nM, GdFVD-OH = 748 nM; (C) GdFFD-OH = 81 nM; (D) GdFFD-OH = 82 nM. The EC_{50} value for dPdFFD-OH given in the main text (200 nM) is the average of the value obtained from panels (A) and (B), rounded to 1 significant figure.

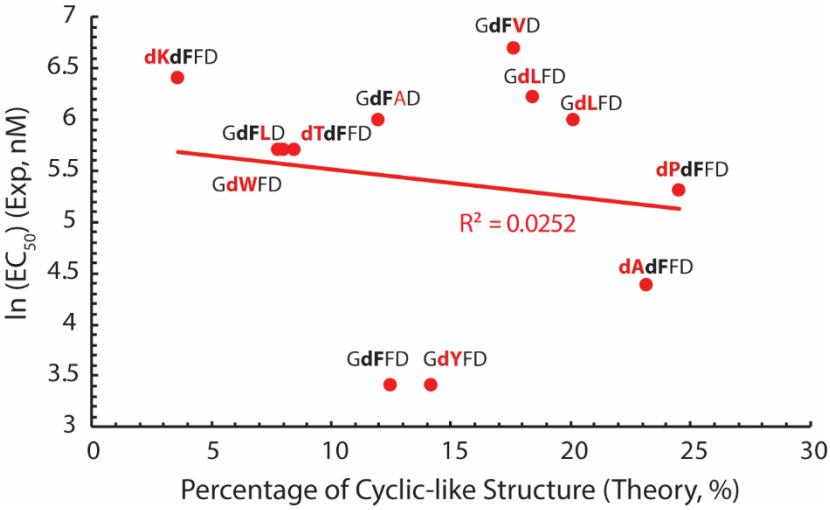


Figure S6. Correlation between the percentage of cyclic-like structures; i.e., end-to-end distance < 0.42 nm, to natural log of experimental EC_{50} for active peptides discussed in this work. In general, peptides with a low percentage of cyclic-like structures (< 10%) have weak activities. On the other hand, those with a high percentage of cyclic-like structures are not necessarily strongly active.

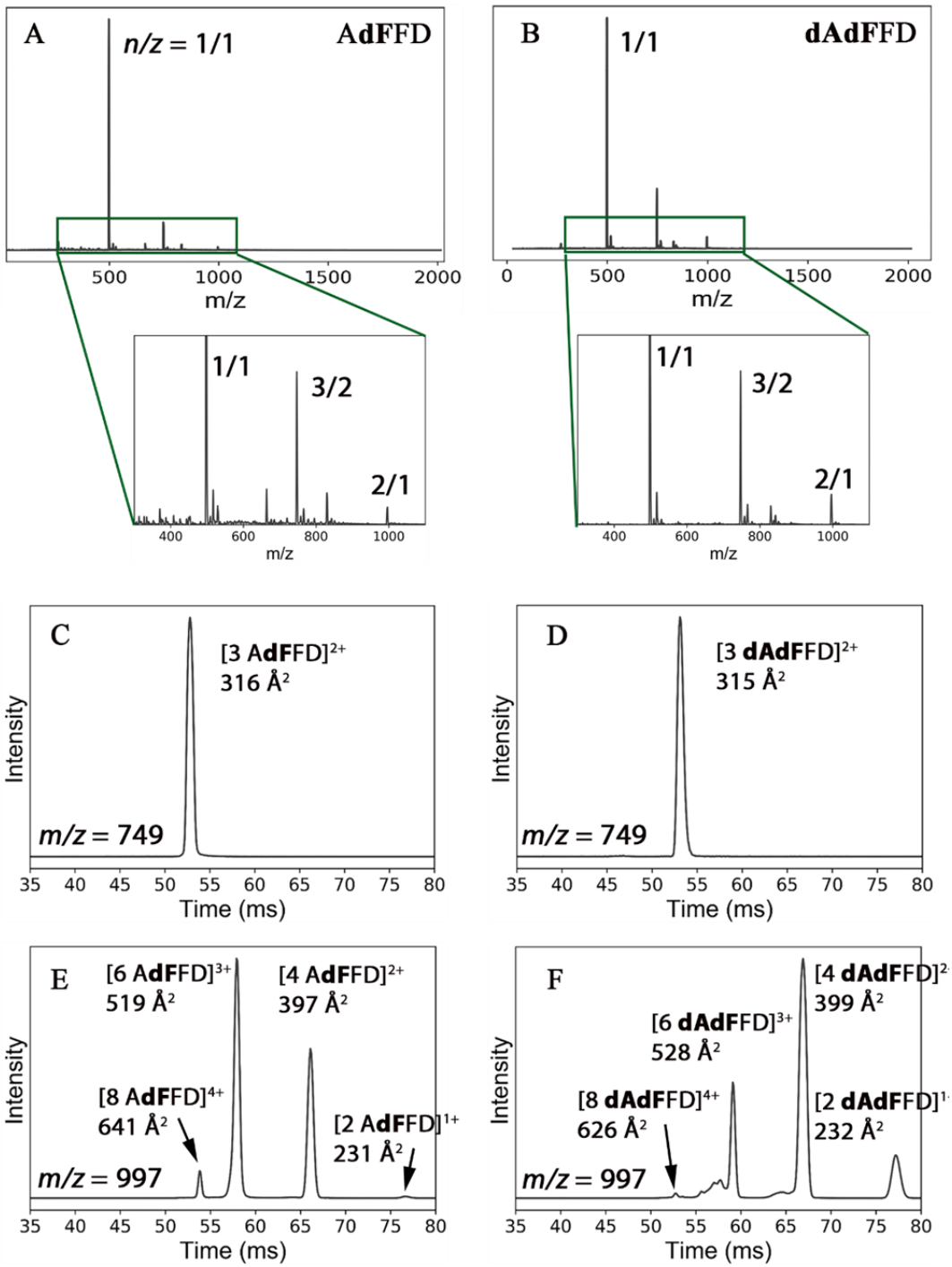


Figure S7. Representative mass spectra of (A) A**dF**FD and (B) **dAdF**FD in water. The *m/z* regions between 300 and 1100 are also shown. Mass spectral peaks are annotated with *n/z* where *n* is the oligomer size and *z* is the number of (positive) charge. (C, D) Representative ATDs of *n/z* = 3/2 at *m/z* = 749. (E, F) Representative ATDs of *n/z* = 2/1 at *m/z* = 997 show the formation of large oligomers up to octamer (*n* = 8). Each feature is labeled with the oligomer species and its corresponding CCS. For [8 dAdFFD]$^{4+}$, the intensity of the feature is low, thus the CCS may not be accurate. All data were collected in positive polarity.

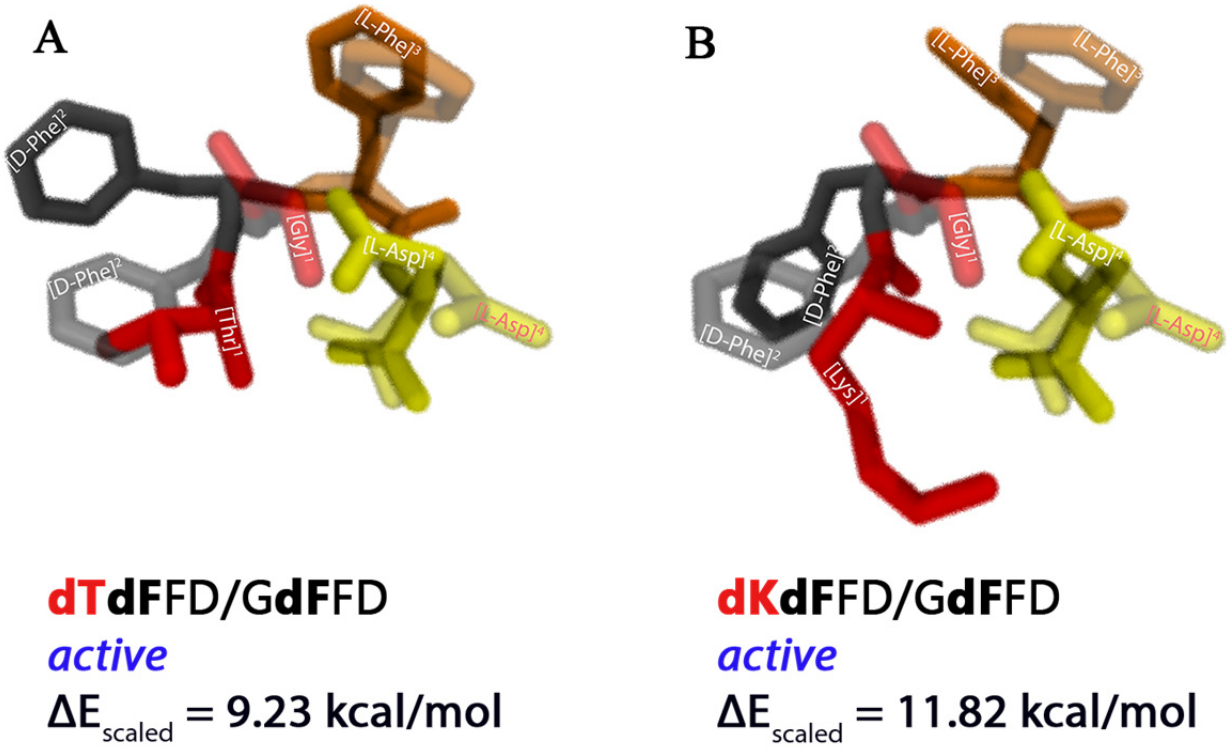


Figure S8. Overlaid structures of (A) **dTdF**FD and (B) **dKdF**FD onto G**dF**FD (shown in lighter colors). In both cases, dT and dK residues do not overlap in coordinate with $[Gly]^1$ of GdFFD.

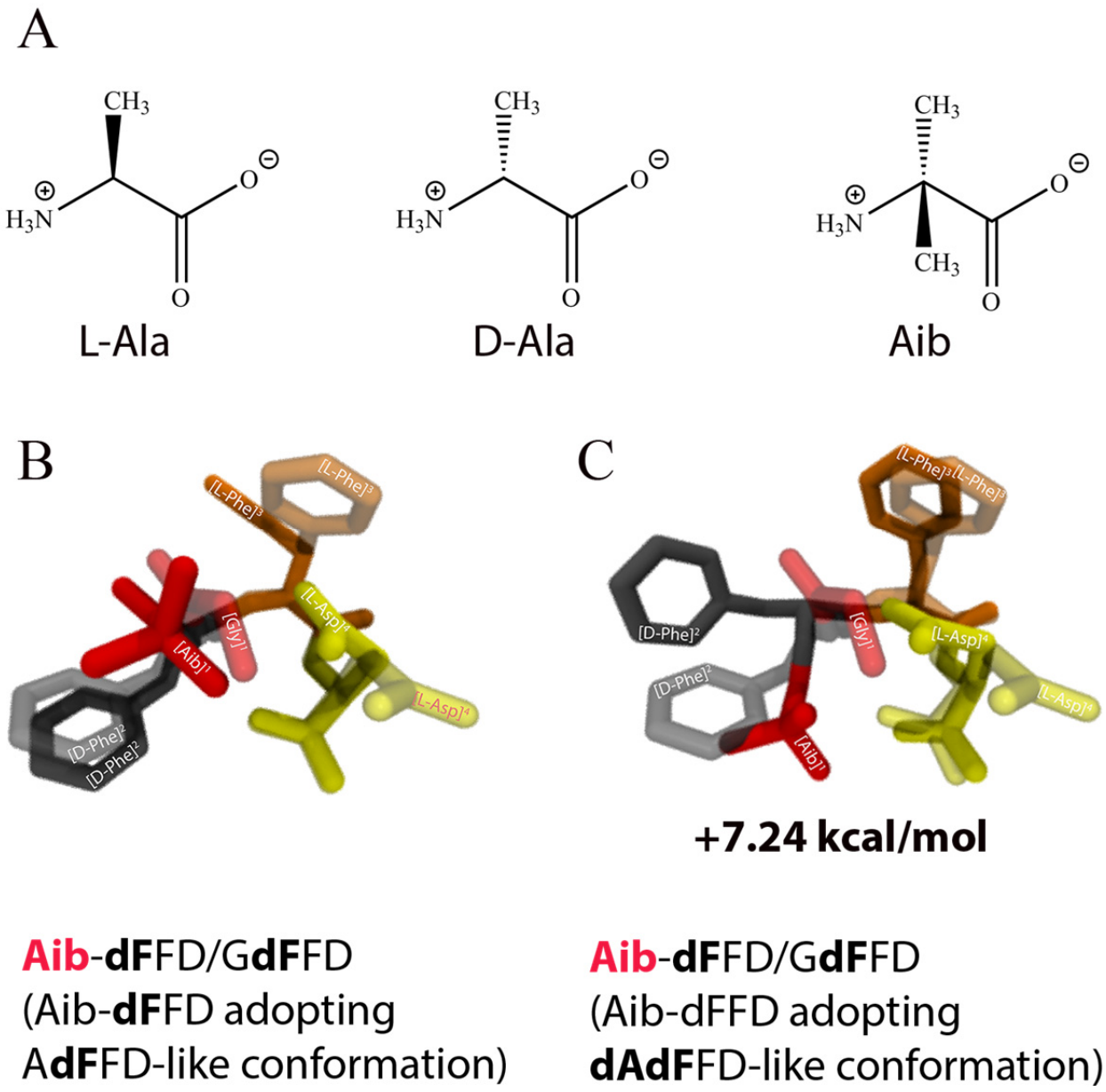


Figure S9. (A) Chemical structures of L-Ala, D-Ala and Aib amino acids. (B, C) Overlaid structures of two Aib-**dF**FD conformations onto G**dF**FD. These two structures were obtained by replacing Hα atoms in A**dF**FD and **dAdF**FD with a methyl (CH_3-group). The resulting structures were geometry optimized following the same protocol as discussed in the main text. The relative positions of Aib residues (in colors) to $[Gly]^1$ in G**dF**FD (in grey) are shown. G**dF**FD is shown in lighter colors in the overlaid images.

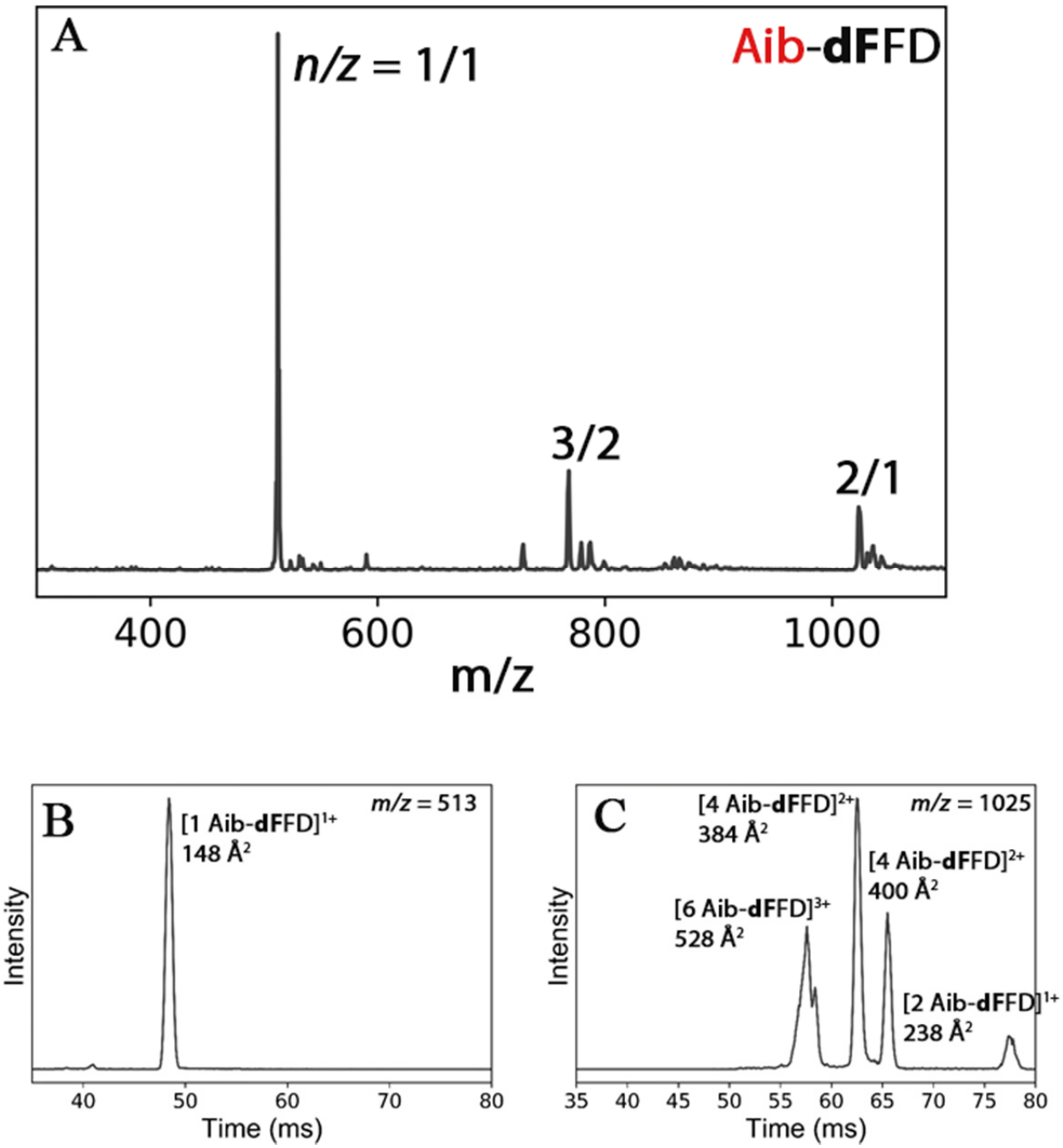


Figure S10. (A) Representative mass spectrum of Aib-**dF**FD in water. Mass spectral peaks were annotated with *n/z* where *n* the number of peptide chain and *z* is the charge. (B, C) Representative ATDs of *n/z* = 1/1 and 2/1. The CCSs are also shown. The data were obtained in positive polarity.

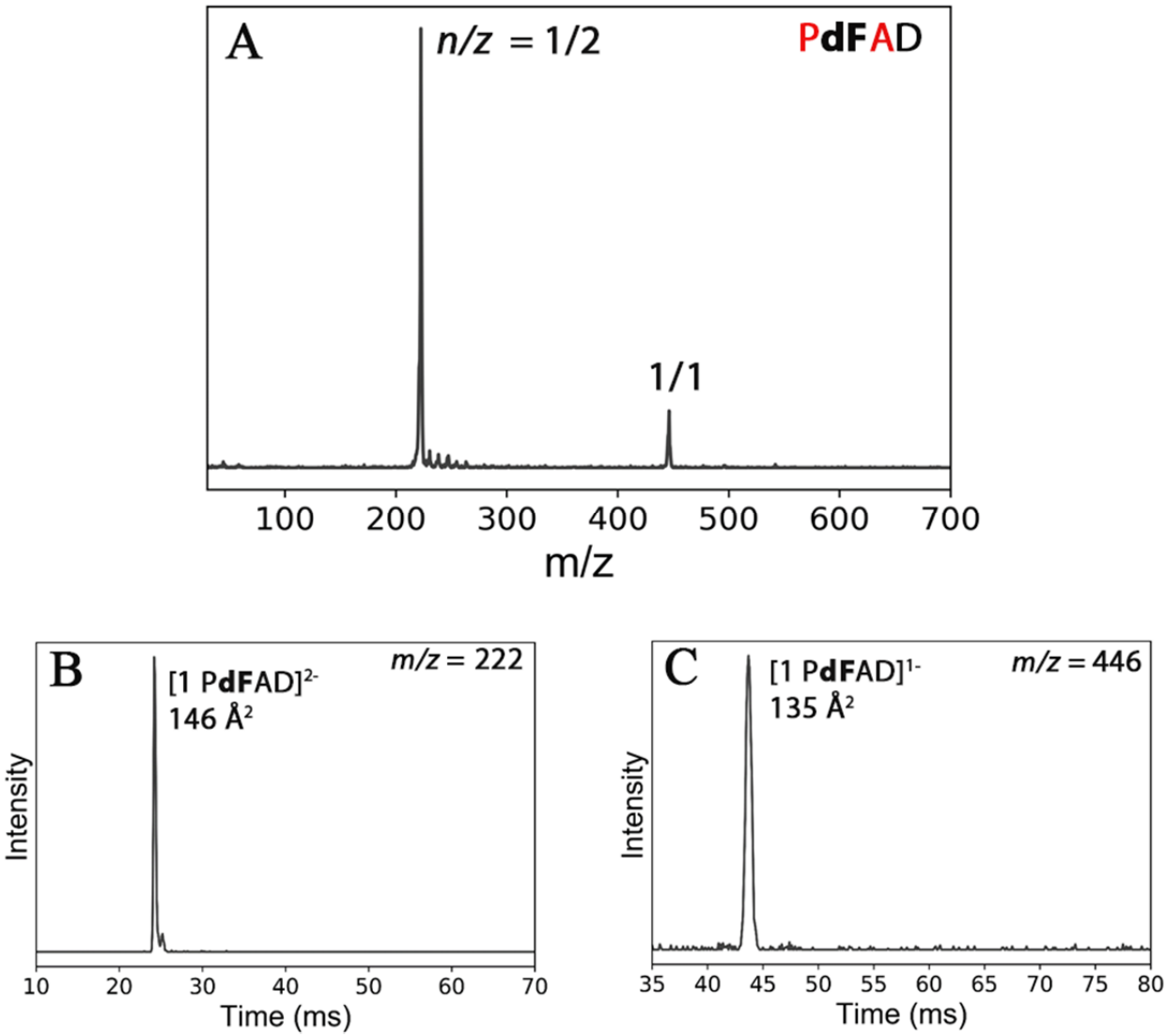


Figure S11. (A) Representative mass spectrum of P**dF**AD in water. Mass spectral peaks were annotated with *n/z* where *n* the number of peptide chain and *z* is the charge. (B, C) Representative ATDs of *n/z* = 1/1 and 1/2. The CCSs are also shown. The data were obtained in negative mode.

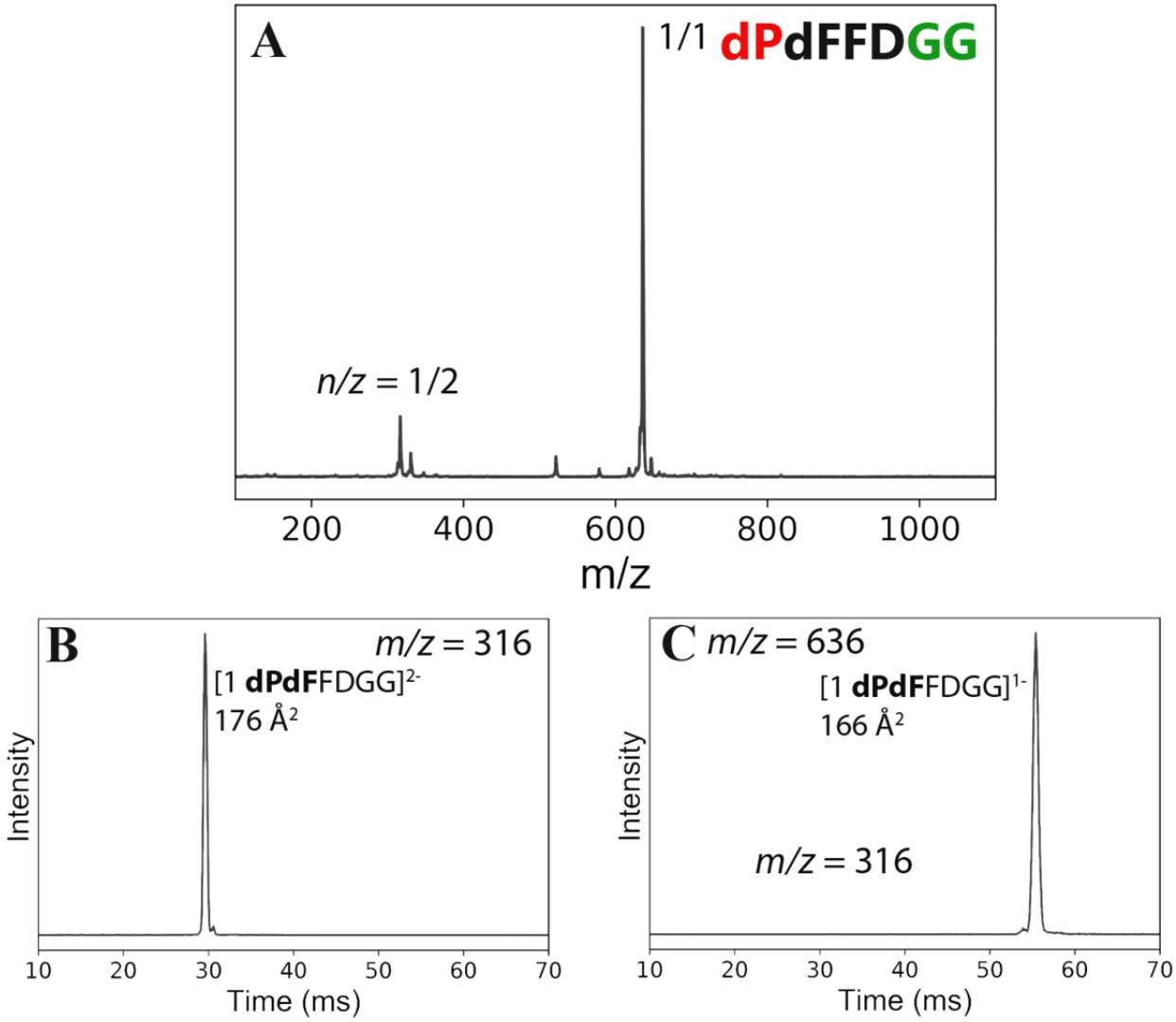
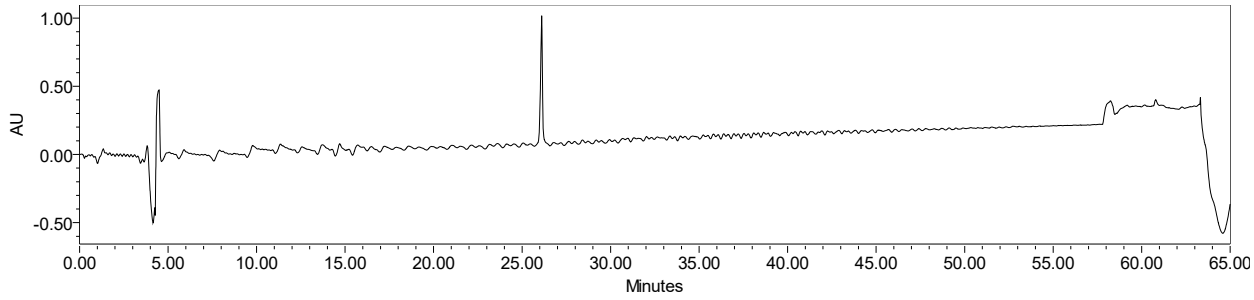
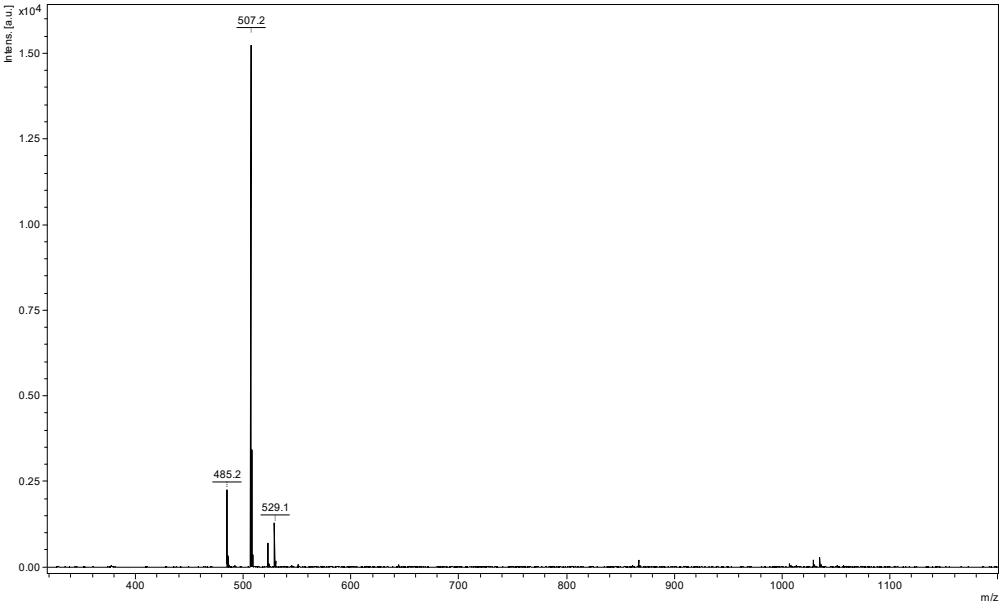


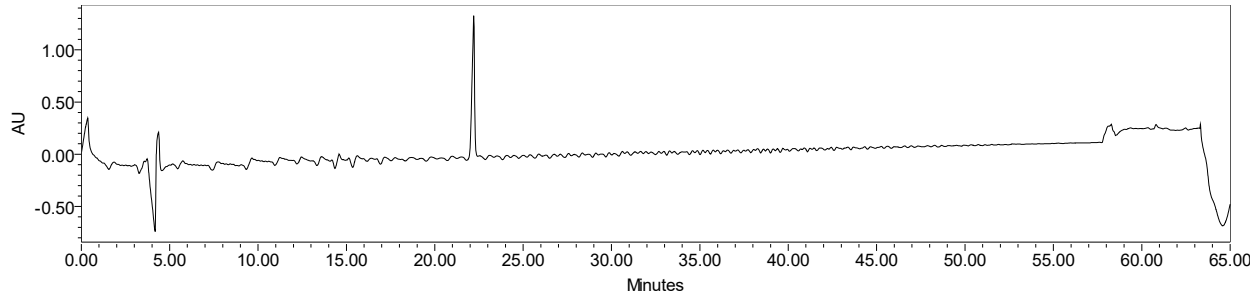
Figure S12. (A) Representative mass spectrum of **dPdFF**DGG in water. Mass spectral peaks were annotated with *n/z* where *n* the number of peptide chain and *z* is the charge. (B, C) Representative ATDs of *n/z* = 1/1 and 1/2. The CCSs are also shown. The data were obtained in negative mode.

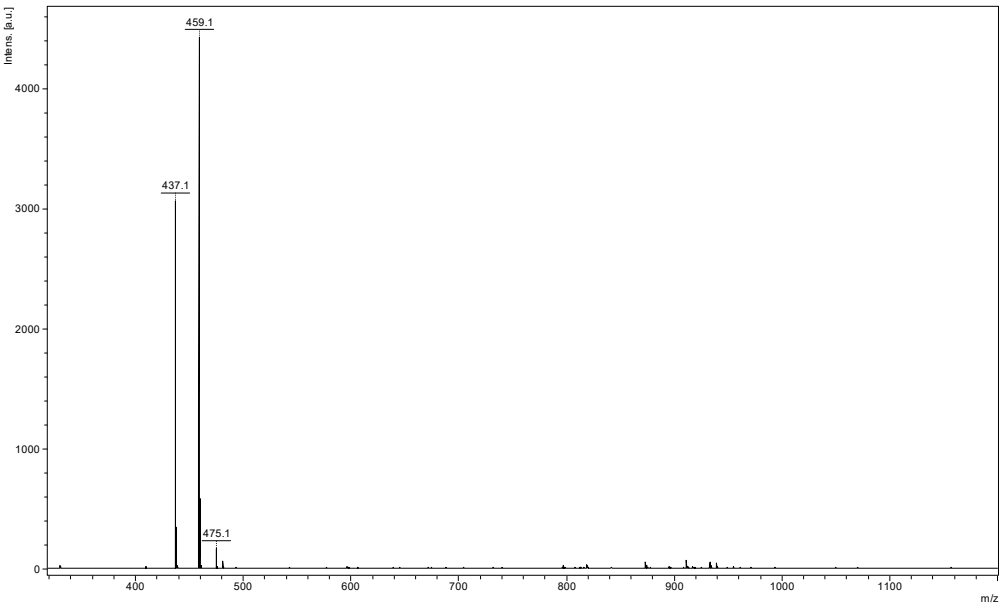
A. G**dFdF**D. Expected $[M+H]^+$ = 485.2, $[M+Na]^+$ = 507.2



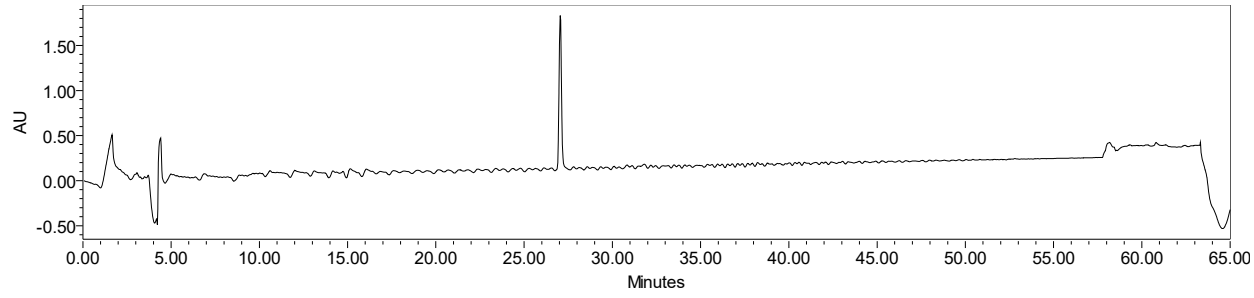


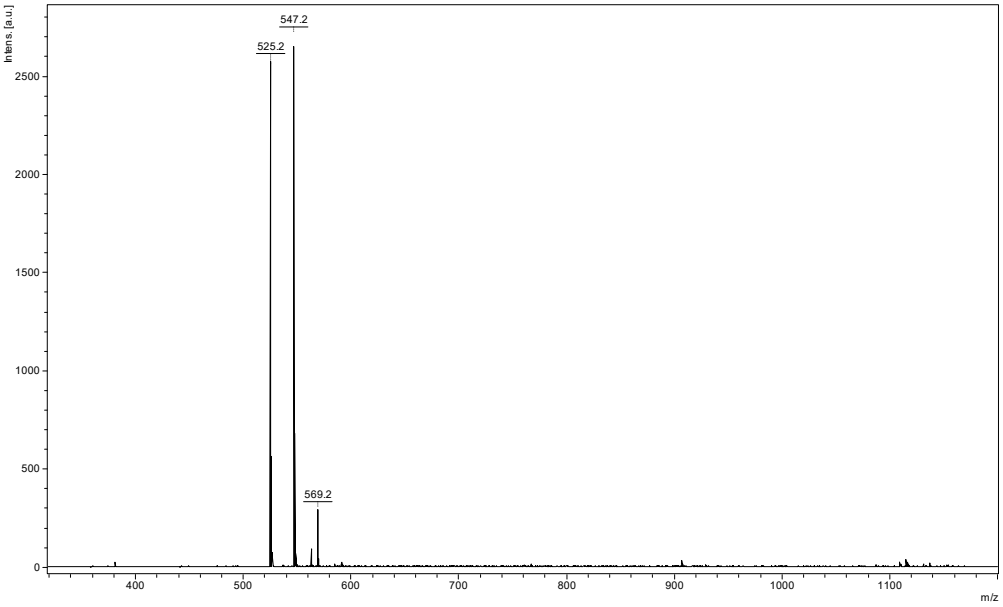
B. G**dF**VD. Expected $[M+H]^+$ = 437.2, $[M+Na]^+$ = 459.2



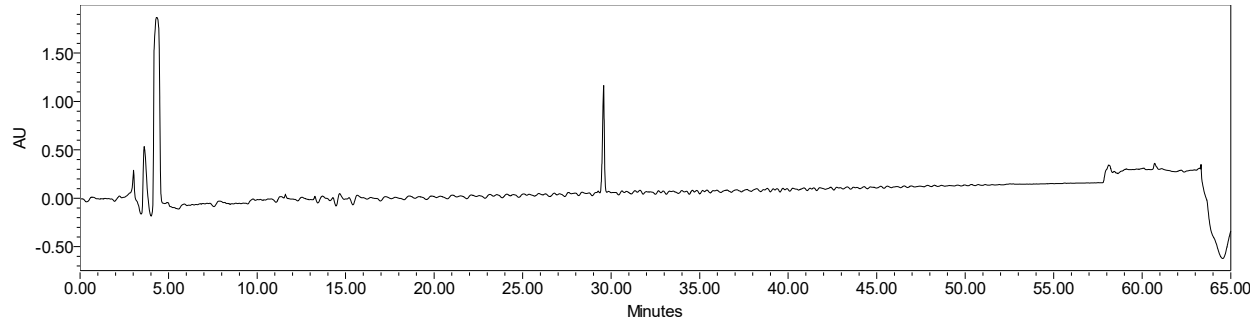


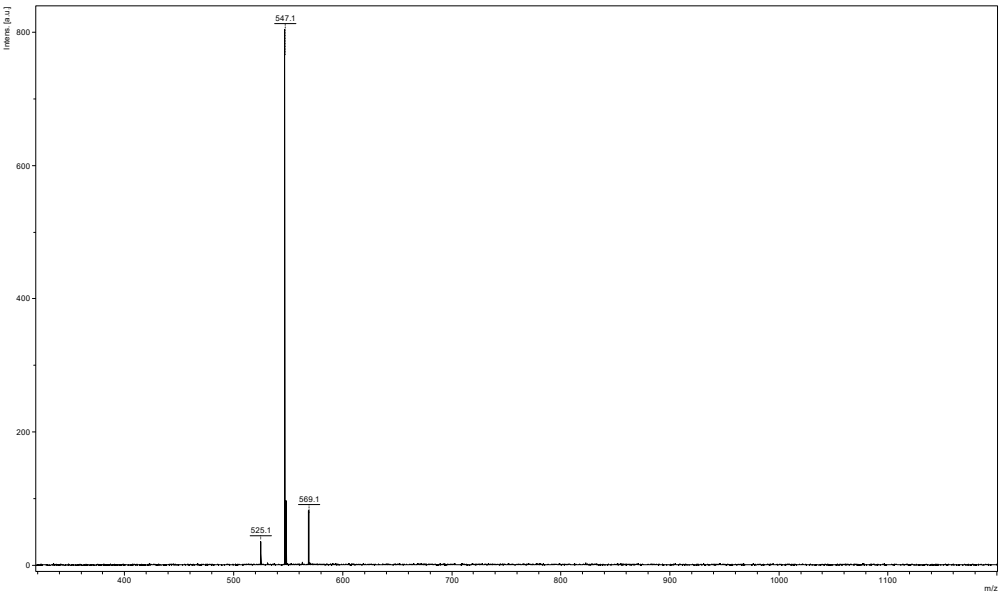
C. dPdFFD. Expected $[M+H]^+$ = 525.2, $[M+Na]^+$ = 547.2



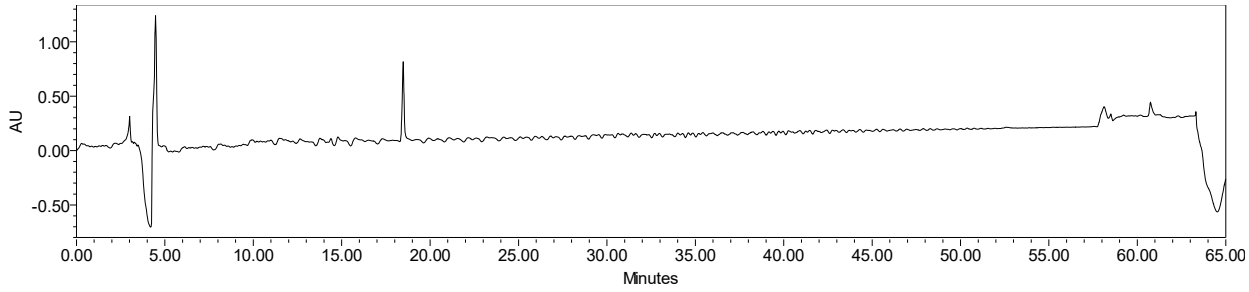


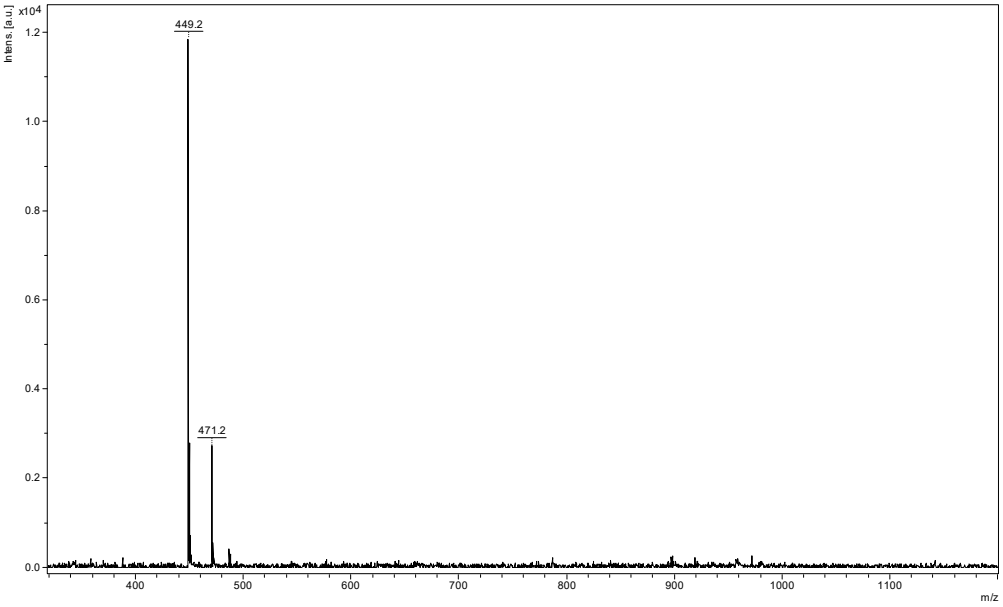
D. P**dF**FD. Expected $[M+H]^+$ = 525.2, $[M+Na]^+$ = 547.2



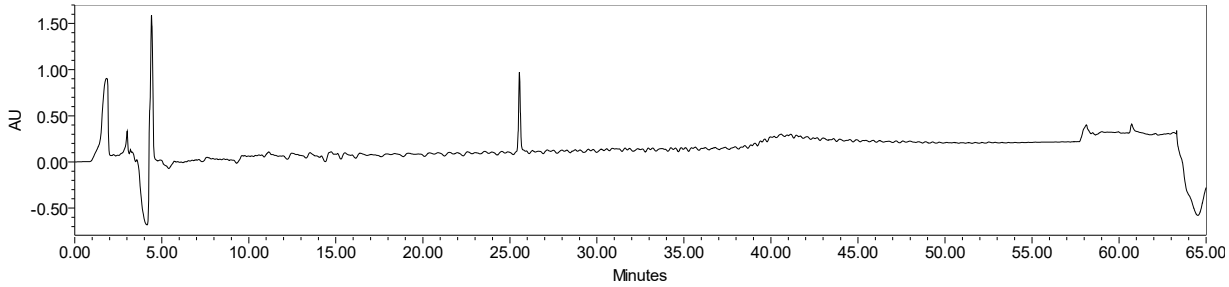


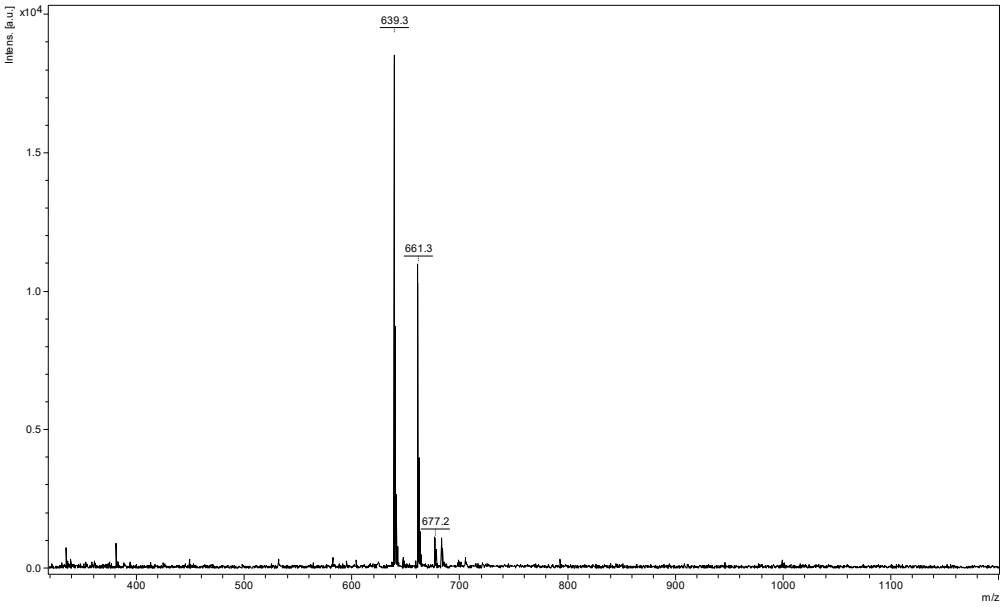
E. P**dF**AD. Expected $[M+H]^+$ = 449.2, $[M+Na]^+$ = 471.2





F. dPdFFDGG. Expected $[M+H]^+$ = 639.3, $[M+Na]^+$ = 661.3





G. Aib-**dF**FD. Expected $[M+H]^+$ = 513.2, $[M+Na]^+$ = 535.2

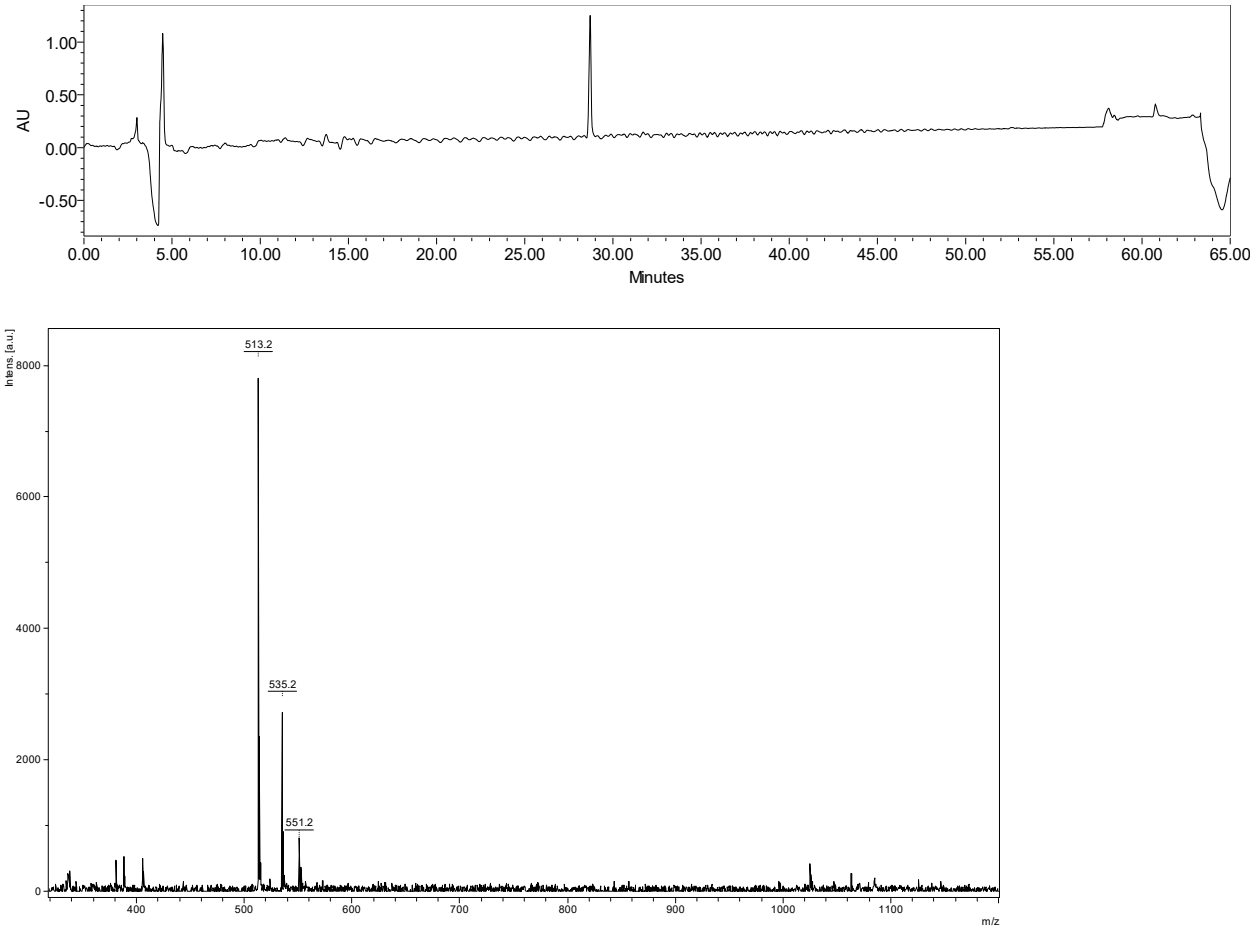


Figure S13. Peptide characterization. The MS-based characterization was performed for each new peptide synthesized and evaluated in this report. After purification by reversed-phase HPLC, peptide purity was assessed by analytical-scale reversed-phase HPLC (top panel) and peptide identity was confirmed by MALDI-TOF MS (bottom panel). Purity analysis by HPLC was monitored based on UV absorbance at 220 nm, using the following conditions: solvent A = 0.1% formic acid (FA) in H^2O, solvent B = 0.1% FA in acetonitrile, flow rate = 500 µL/min, gradient = 5–55% solvent B over 50 min, temperature = 35 °C. For MALDI-TOF MS analysis, the monoisotopic $[M+H]^+$ is labeled, along with the $[M+Na]^+$ (+23), $[M+K]^+$ (+39), or $[M+2Na-H]^+$ (+45) in some spectra.

Department of Chemistry
600 S. Mathews Ave. 63-5
Urbana IL 61801

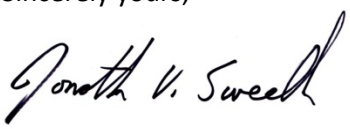
Jonathan V. Sweedler
June 10, 2018

Letter to Referees:

I am pleased to submit our manuscript titled ***Conformational Investigation of the Structure – Activity Relationship of GdFFD and Its Analogues on an Achatin-like Neuropeptide Receptor of Aplysia californica Involved in the Feeding Circuit*** for publication in **Physical Chemistry Chemical Physics (PCCP)**.

Although proteins and peptides are overwhelmingly made of L-amino acids, recent advances in analytical measurement technologies have revealed previously unappreciated roles of D-amino acids in biological processes. Over 30 D-amino acid containing peptides (DAACPs) have been discovered in animals where at least one L-residue has been isomerized to the D-form via an enzyme-catalyzed process. In *Aplysia californica*, an opisthobranch mollusk whose neural network has been extensively investigated by us and others, we previously showed that GdFFD and GdYFD (the lower-case letter "d" indicates the amino acid is a D-amino acid) modulate the feeding behavior by activating the *Aplysia* achatin-like neuropeptide receptor (apALNR). On the other hand, the all L-residue counterparts are completely inactive. Due to the lack of high-resolution structural data for apALNR, little is known about the roles of the D-residue in the peptide conformations that influence the peptide activities at the receptor and the ligand binding site. Here we performed replica exchange molecular dynamics and quantum mechanics calculations to elucidate the conformations of GdFFD and its analogs. The resulting structures were experimentally validated by gentle ion-mobility mass spectrometry measurements via cross section comparison. We then constructed a simple model that predicts bioactivities for a series of GdFFD analogs. Our model highlights a steric effect on peptide activity at position 1 on the GdFFD receptor apALNR. Our approach provides a better understanding of ligand-receptor interactions in the absence of high-resolution structural data. We feel this is an important result and one that will be well appreciated by the readership of **PCCP**.

Sincerely yours,



Jonathan V. Sweedler
James R. Eiszner Family Endowed Chair in Chemistry
Director, School of Chemical Sciences
Professor of Neuroscience and Physiology and the Beckman Institute

telephone 217-244-7359 • *fax* 217-265-6290 • *email* jsweedle@illinois.edu