Supplementary Information (SI): Differential Self-Assembly of Sequence-Isomeric Phosphoestamers

James Williamson, Tomasz Piskorz, Bini Claringbold, Alexandra R. Paul, Nikita Harvey, Fernanda Duarte, Christopher J. Serpell*

Contents

Materials and Instrumentation

All chemicals used were reagent grade and were purchased from Sigma Aldrich, Fischer Scientific, Alfa Aesar, Apollo Scientific, Link Technologies, ChemGenes, and Fluorochem. When specified as anhydrous acetonitrile, this was bought as such that it is free of water and the anhydrous solvent was transferred by nitrogen purged syringes with needles and septa.

DNA Synthesis: All phosphoestamers were synthesised on an Expedite™ 8909 Nucleic Acid Synthesiser system provided by Biolytic.

Mass Spectrometry: Mass spectrometry data was collected on a Shimadzu 2020 LCMS and an Agilent Q TOF6545 LC-MS/MS.

NMR: ^{31}P NMR spectra were recorded at a ^{31}P NMR frequency of 161 MHz using a Bruker Avance Neo 400 spectrometer equipped with a Bruker BBO iProbe and 9.4 T Bruker Ultrashield magnet. ^{1}H DOSY NMR spectra were recorded at a ^{1}H NMR frequency of 600 MHz using a Bruker Avance Neo spectrometer equipped with a QCI-F cryogenically cooled probe and 14.1 T Bruker Ascend magnet. Chemical shifts for all NMR data are in ppm (parts per million).

Synthesis of Tetramers

C12 phosphoramidite (0.25 g) and spacer 18 (**HEG**) phosphoramidite (1 g) were dissolved separately in anhydrous acetonitrile (5 mL/0.25 g of phosphoramidite) and affixed to the DNA synthesiser under a nitrogen atmosphere. Other reagents on the DNA synthesiser were as follows: oxidiser (0.02 M iodine, 20% pyridine), Cap A Mix (tetrahydrofuran/pyridine/acetic anhydride 8:1:1), Cap B Mix (10% methylimidazole in tetrahydrofuran), deblock solution (3% trichloroacetic acid in dichloromethane), and ETT activator solution (0.25 M 5-ethylthio-1H-tetrazole, acetonitrile). A leak test was undertaken to ensure nitrogen was not leaking from the lines, then the lines were primed with new reagent ready for synthesis. Next, UnyLinker™ universal support (0.021 g) was added to a synthesiser column, attached to the synthesiser, and this was flushed several times with acetonitrile. The Expedite™ 8909 DNA Synthesiser was connected to a computer with Validate XP software and, using this, a custom sequence (e.g. 7878 for **(C12-HEG)$_2$**) and a customised protocol of DNA synthesis at 1 µmol were selected.

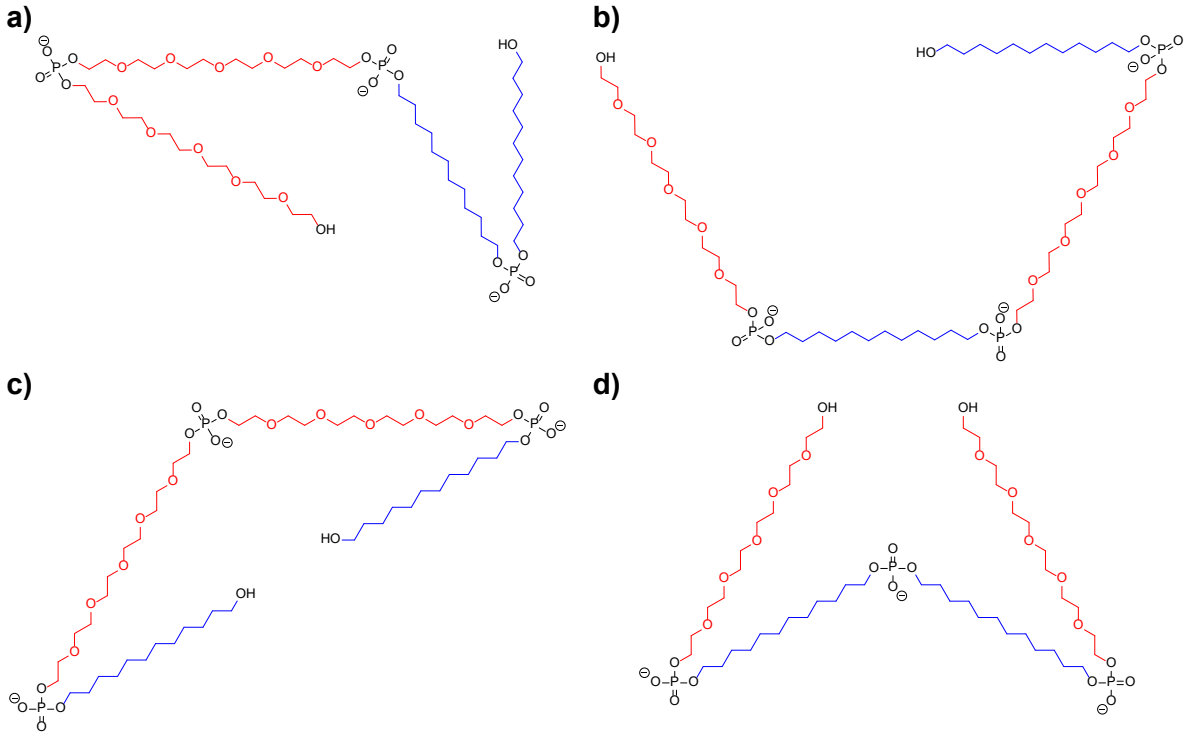

Figure S1: Structures of the four synthesised tetramers. (a) ***C12$_2$-HEG$_2$****, (b)* ***(C12-HEG)$_2$****, (c)* ***C12-HEG-HEG-C12****, (d)* ***HEG-C12-C12-HEG****.*

^{31}P NMR of Tetramers

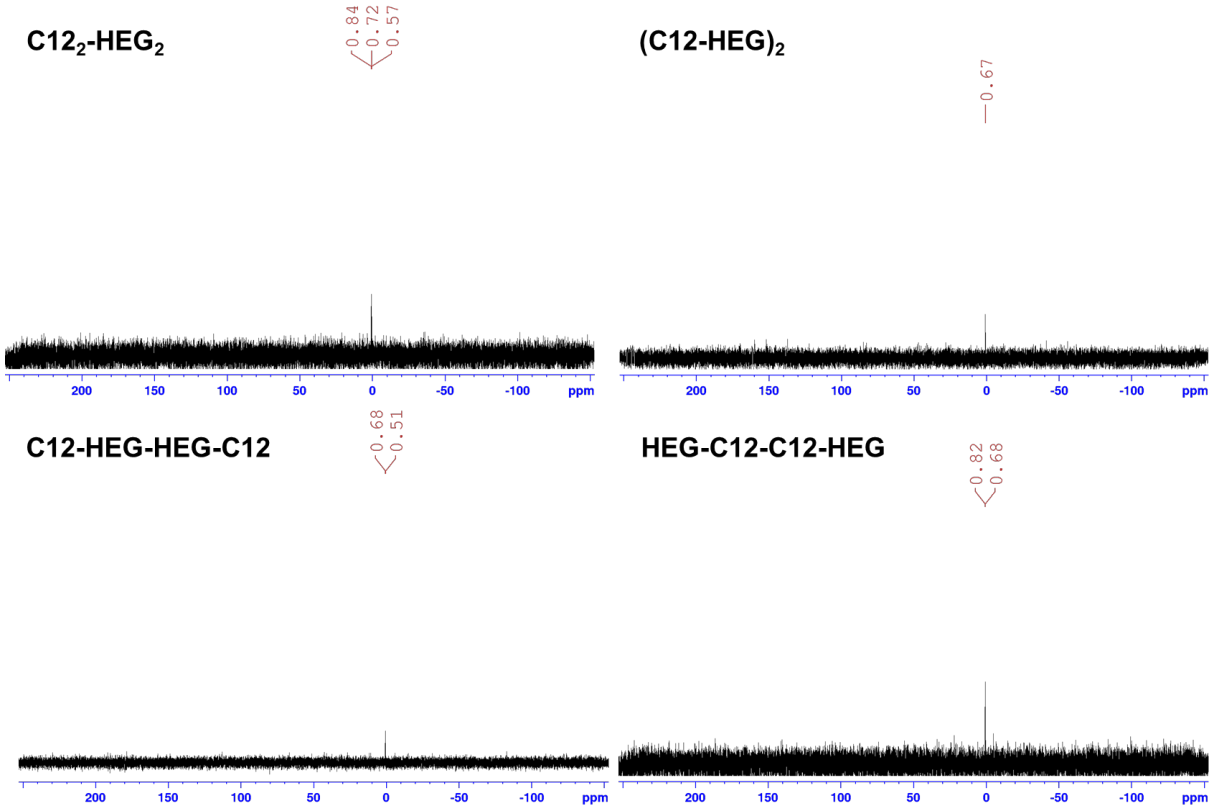

Figure S2: Entire ^{31}P NMR spectra of the ***C12*** *and* ***HEG*** *tetramers.*

Mass Spectrometry

The tetramers were prepared at a concentration of 1 mg/mL in MilliQ water and run for mass spectrometry.

$C12_2$-HEG_2

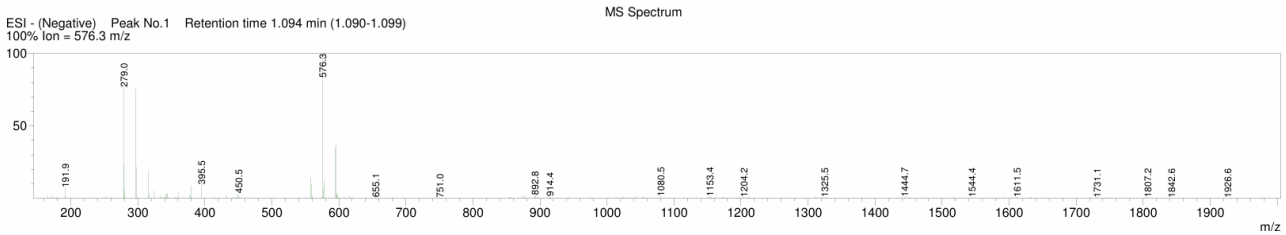

*Figure S3: Mass spectrum of the **$C12_2$-HEG_2** tetramer obtained on a Shimadzu 2020 LCMS.*

$(C12\text{-}HEG)_2$

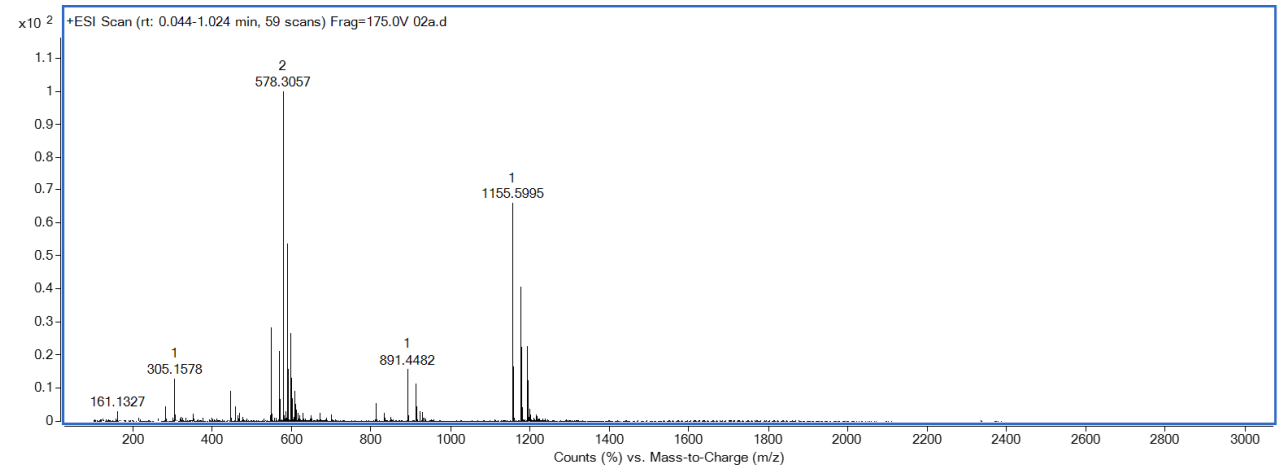

*Figure S4: Mass spectrum of the **$(C12\text{-}HEG)_2$** tetramer obtained on an Agilent Q TOF6545 LC-MS/MS.*

C12-HEG-HEG-C12

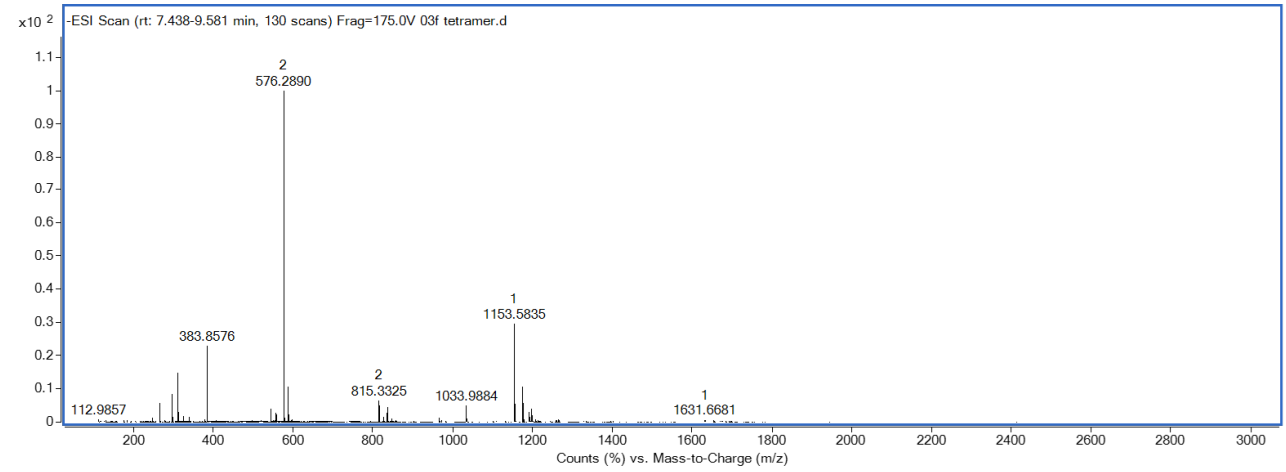

*Figure S5: Mass spectrum of the **C12-HEG-HEG-C12** tetramer obtained on Agilent Q TOF6545 LC-MS/MS.*

HEG-C12-C12-HEG

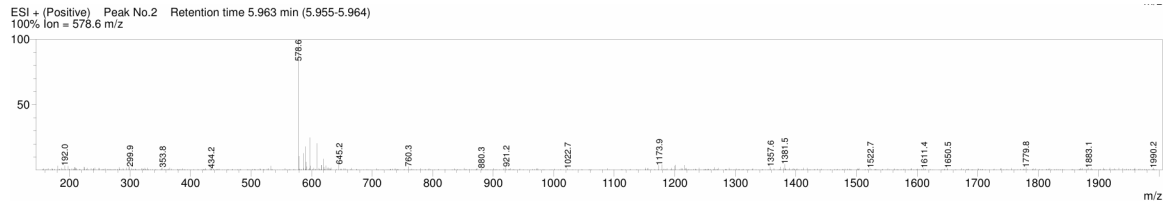

*Figure S6: Mass spectrum of the **HEG-C12-C12-HEG** tetramer obtained on a Shimadzu 2020 LCMS.*

^{1}H DOSY NMR Experimental

Samples were prepared in 3 mm NMR tubes using deuterium oxide/deuterium oxide with ammonium acetate as the solvent (170 µL). A 600 MHz Bruker Advance Neo spectrometer was set up for the DOSY NMR experiment in manual mode using a gradient range of 5%-95%, a diffusion delay of 100 ms, a relaxation delay of 2 s, 16 scans, 4 dummy scans, a diffusion gradient pulse length of 2.5 ms, 32 steps in the F1 dimension, a linear gradient ramp and an exponential fitting routine used for diffusion dimensions. A longitudinal eddy current delay pulse sequence was used with a gradient pulse, bipolar pulse pairs, presaturation, and 2 spoil gradients (ledbpgppr2s).

The results were processed initially using TopSpin (version 4.4.1) from Bruker and then diffusion coefficients were calculated using the main resonance of the alkoxy protons (3.60-3.70 ppm), and that of the alkyl protons (1.20-1.30 ppm) (Figure S7), using the General NMR Analysis Toolbox (version 1.3.2) software developed by the University of Manchester (GNAT).[1]

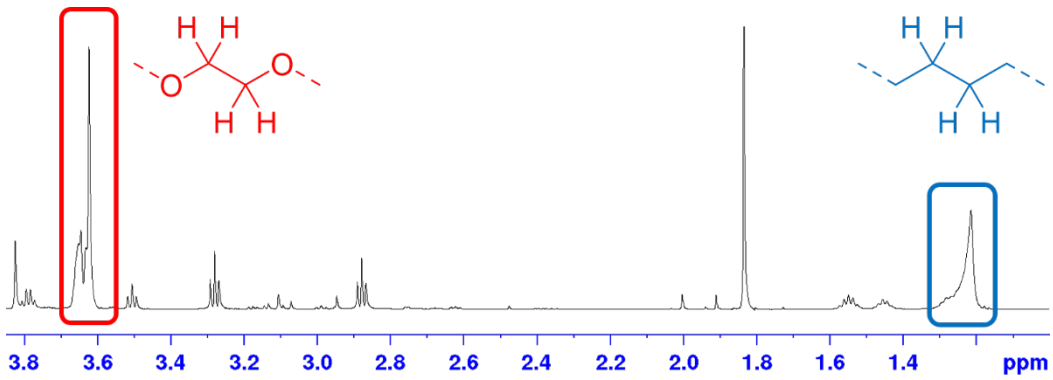

*Figure S7: 1D ^{1}H NMR spectrum of **C12$_2$-HEG$_2$**.*

DOSY NMR data (Tetramers 1 mM)

$C12_2$-HEG_2

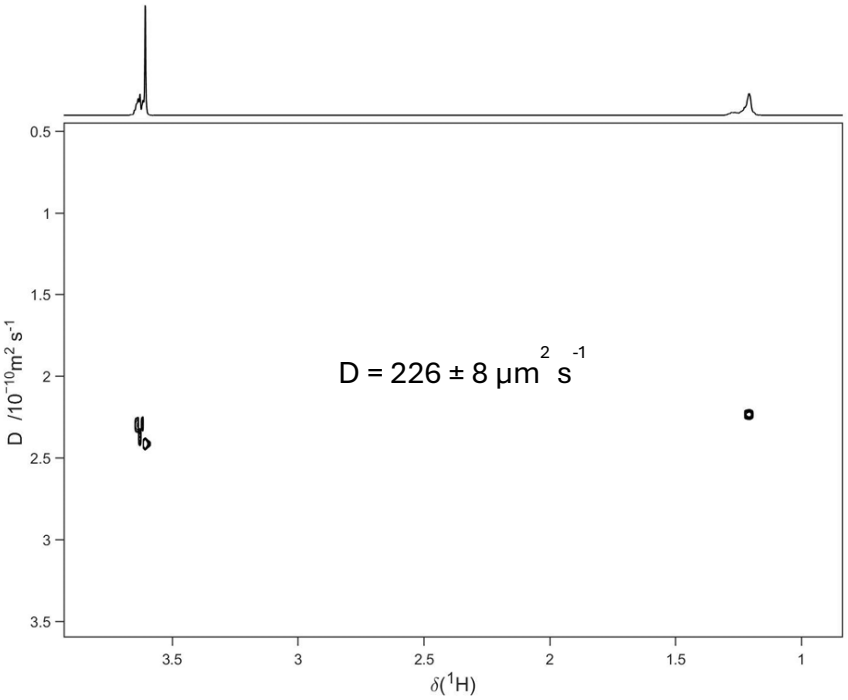

*Figure S8: Repeat 1 for **$C12_2$-HEG_2**.*

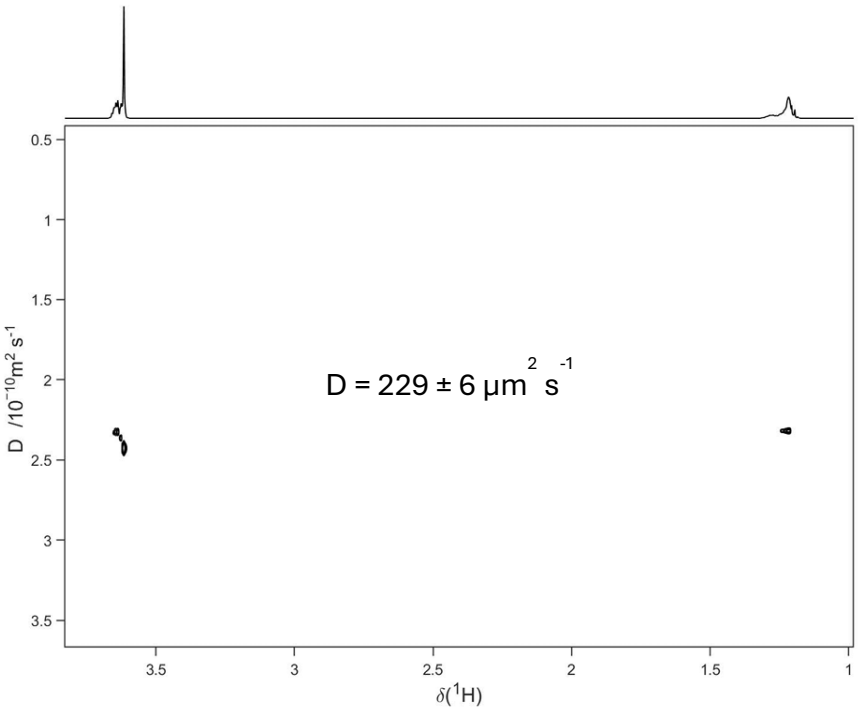

*Figure S9: Repeat 2 for **$C12_2$-HEG_2**.*

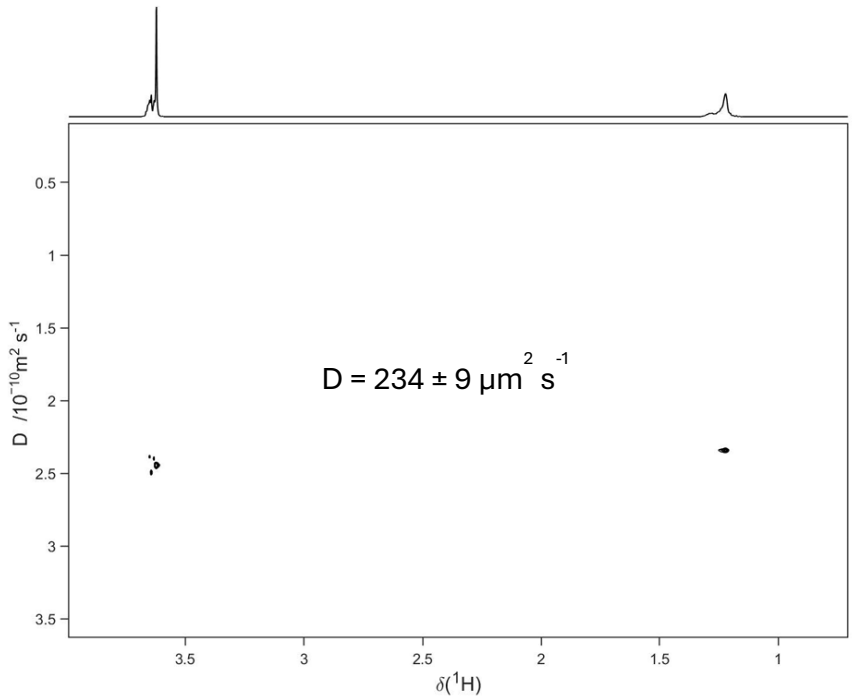

Figure S10: Repeat 3 for ***$C12_2$-HEG_2****.*

$(C12\text{-}HEG)_2$

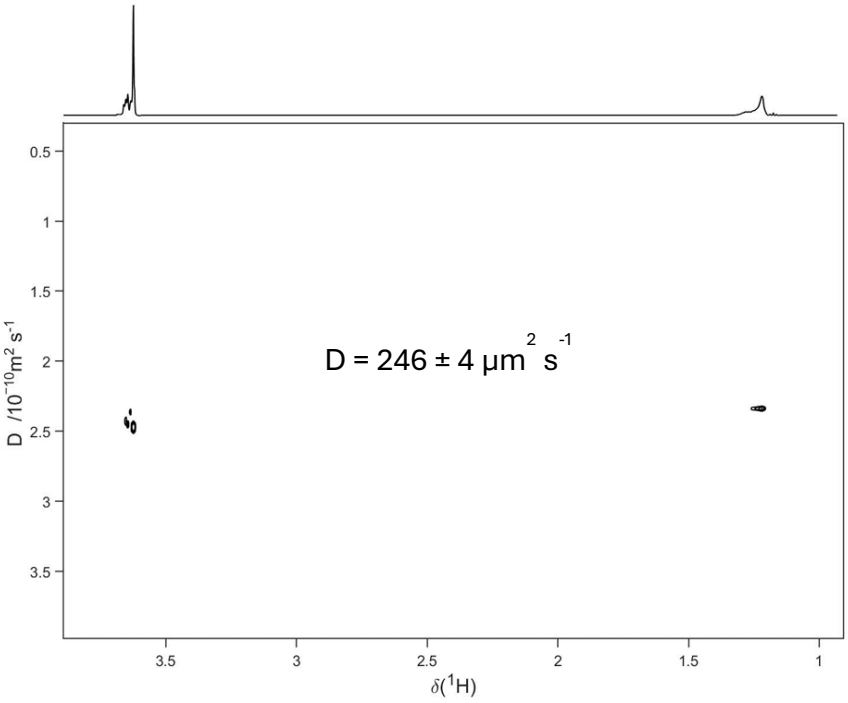

Figure S11: Repeat 1 for ***$(C12\text{-}HEG)_2$****.*

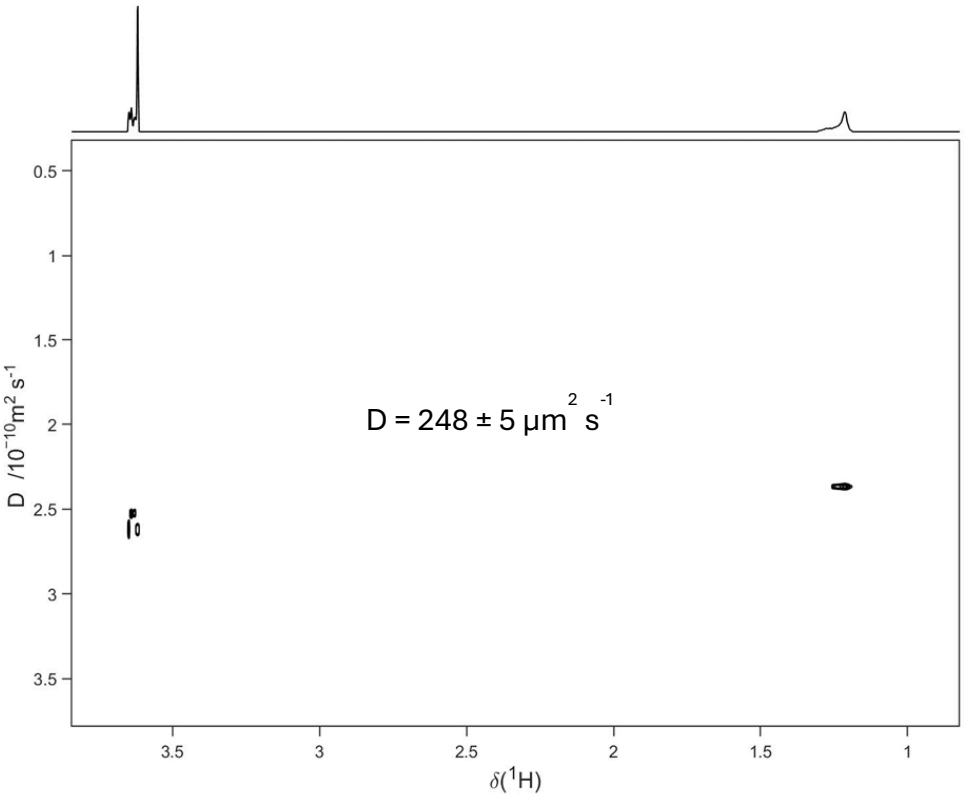

Figure S12: Repeat 2 for ***(C12-HEG)₂.***

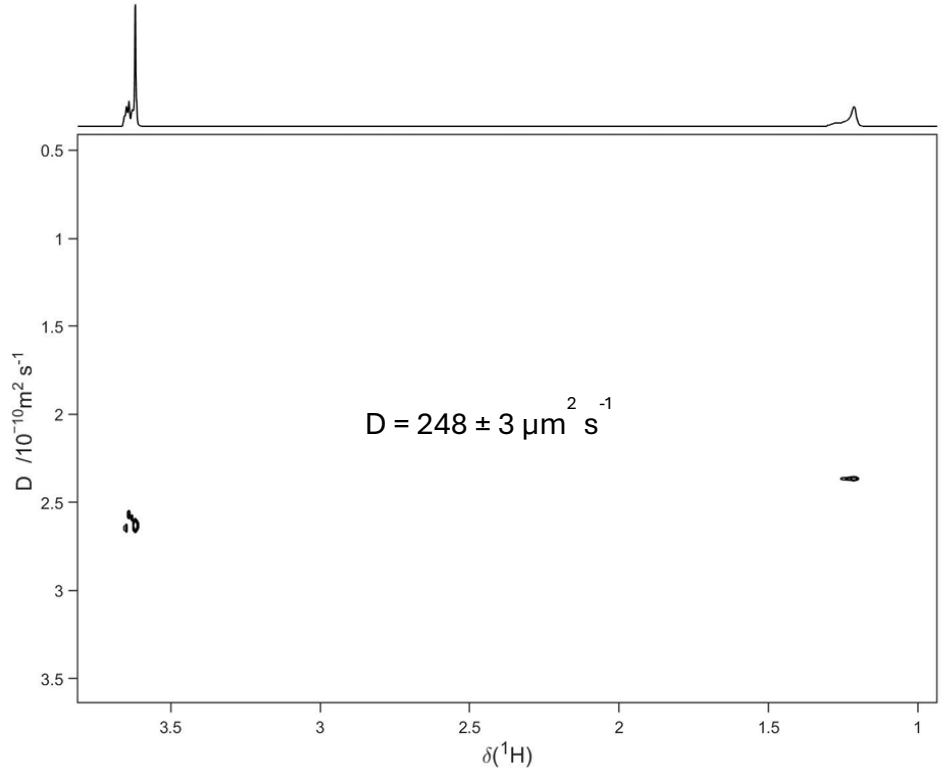

Figure S13: Repeat 3 for ***(C12-HEG)₂.***

C12-HEG-HEG-C12

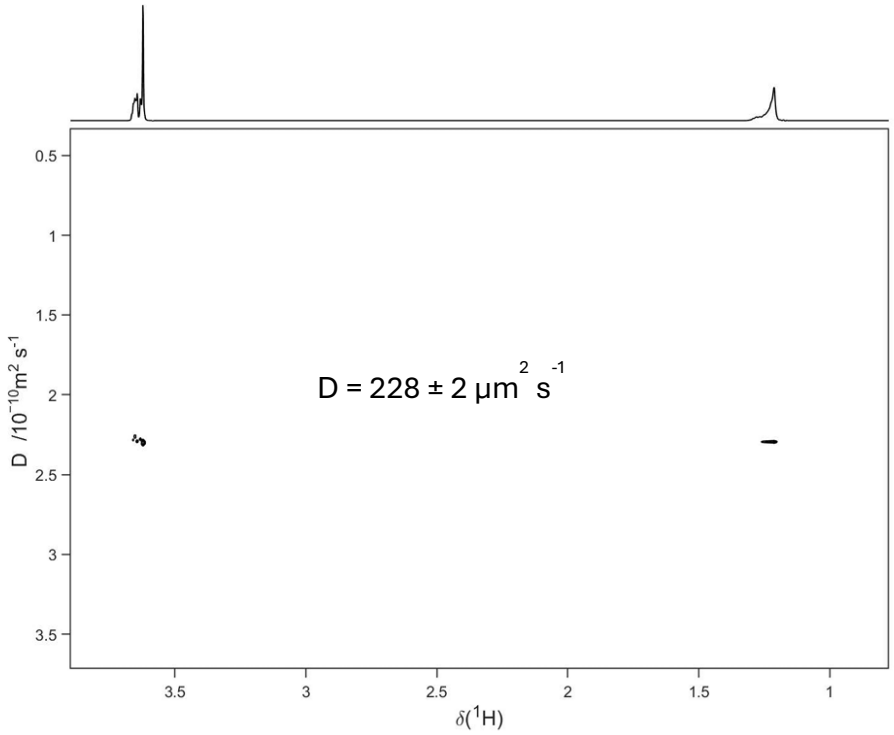

*Figure S14: Repeat 1 for **C12-HEG-HEG-C12**.*

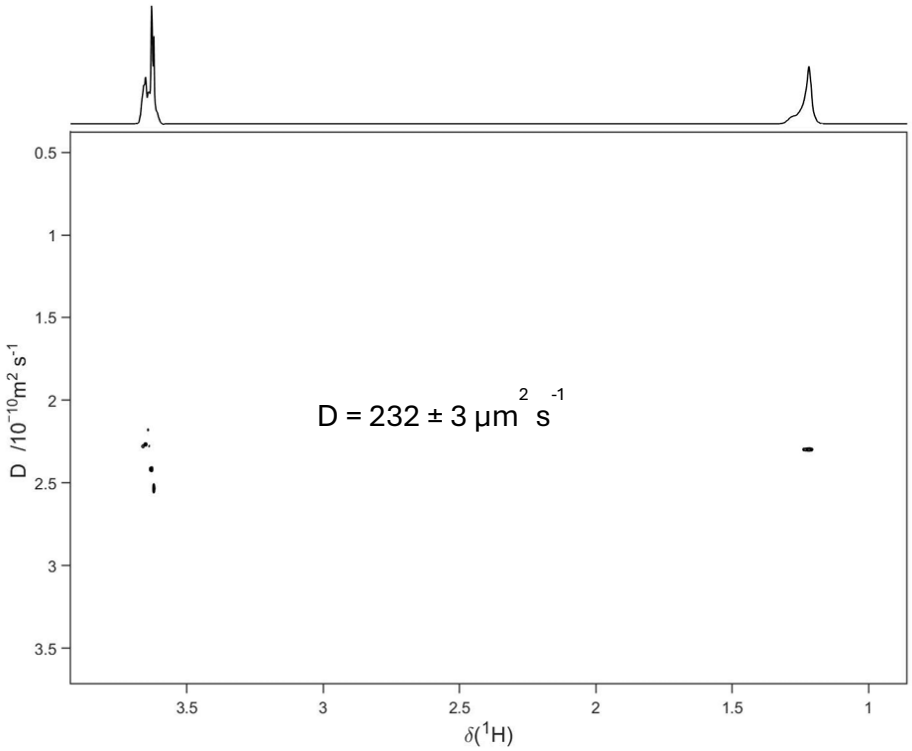

*Figure S15: Repeat 2 for **C12-HEG-HEG-C12**.*

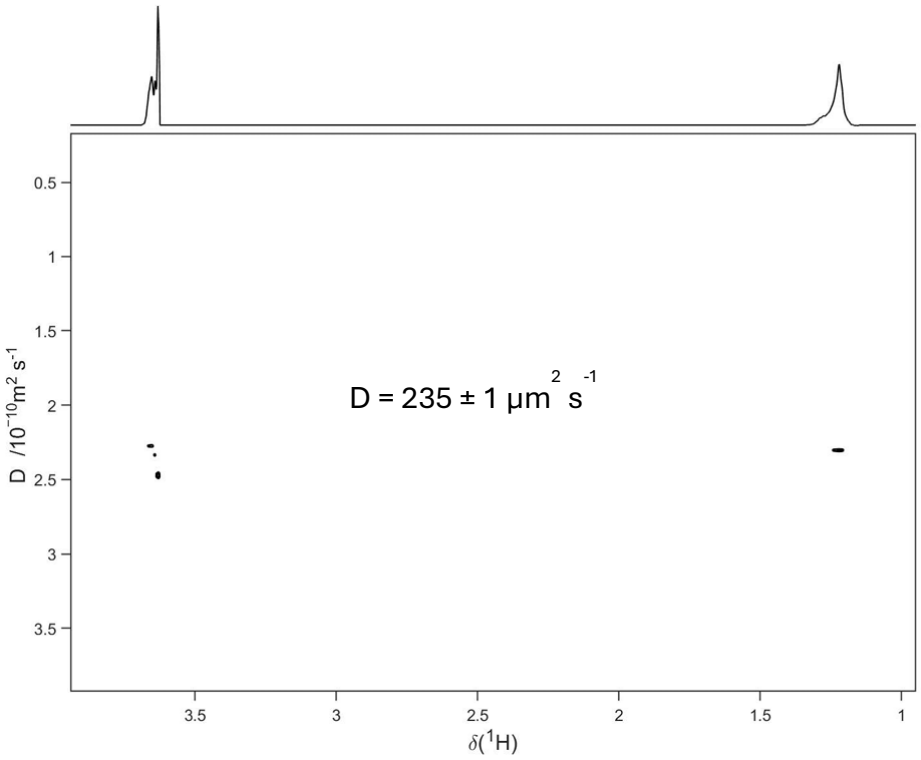

*Figure S16: Repeat 3 for **C12-HEG-HEG-C12**.*

HEG-C12-C12-HEG

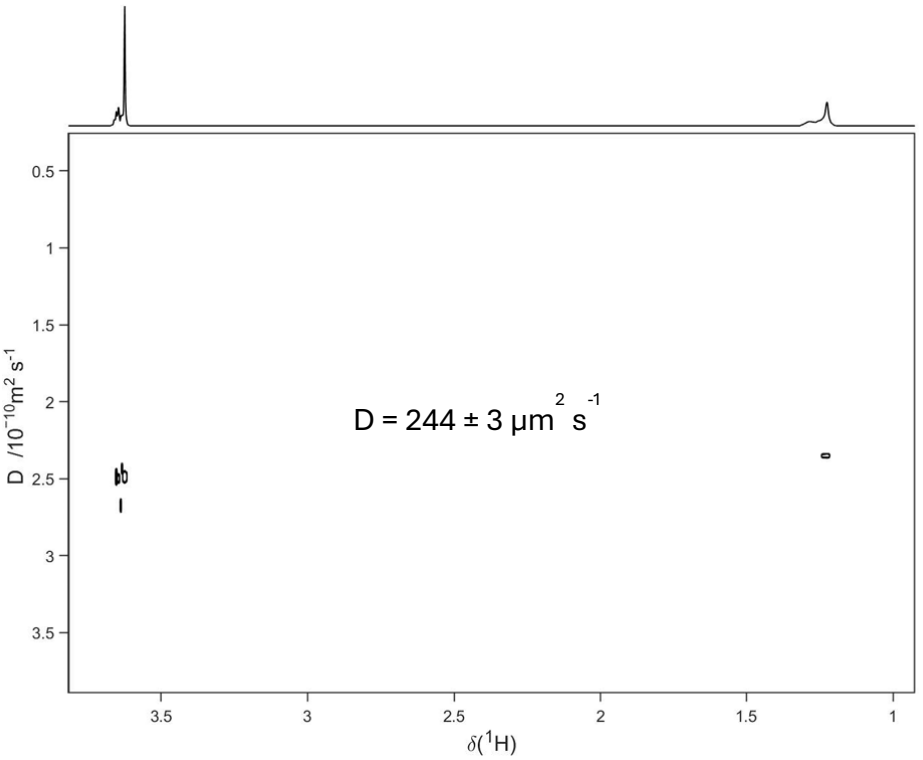

*Figure S17: Repeat 1 for **HEG-C12-C12-HEG**.*

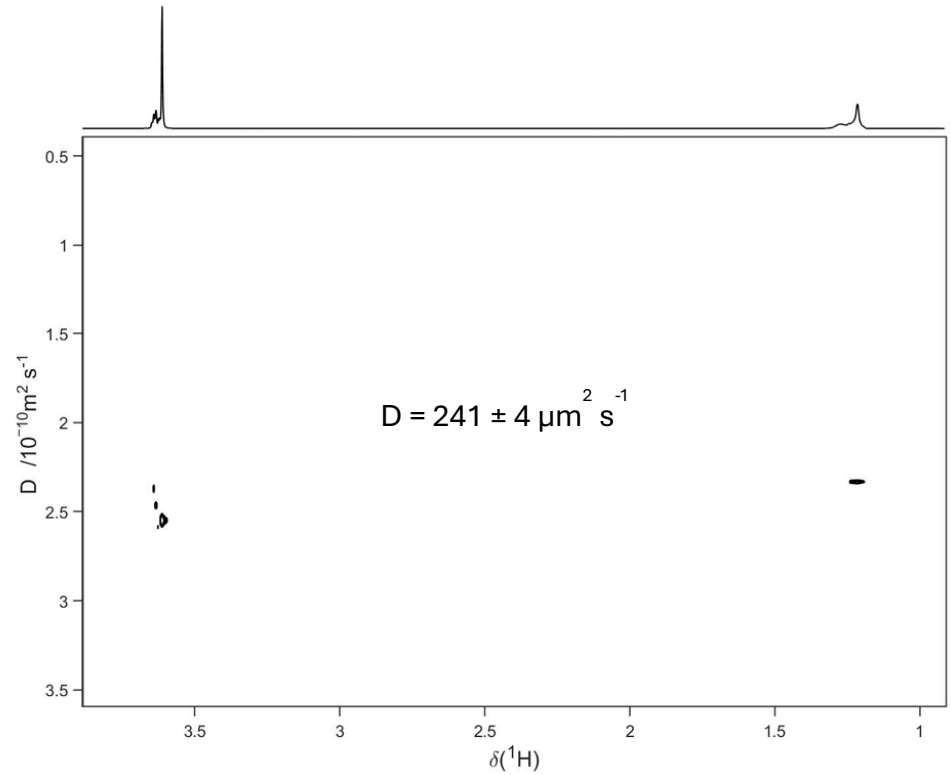

*Figure S18: Repeat 2 for **HEG-C12-C12-HEG**.*

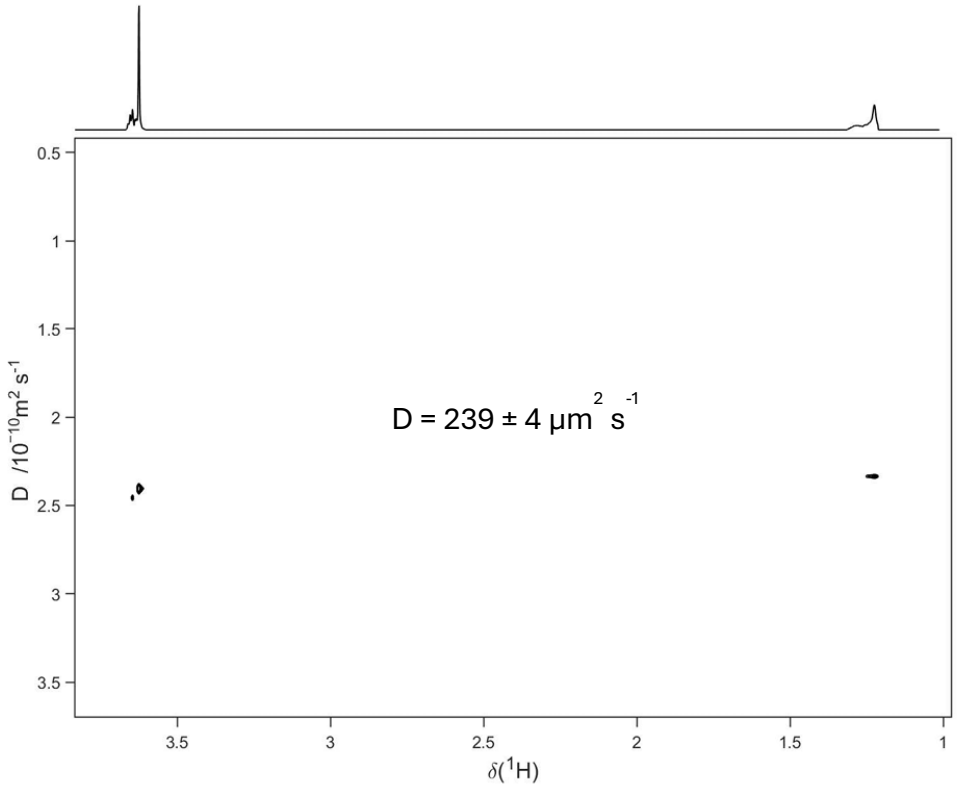

*Figure S19: Repeat 3 for **HEG-C12-C12-HEG**.*

DOSY NMR data (Tetramers 1 mM, 1 M NH_4OAc)

$C12_2$-HEG_2

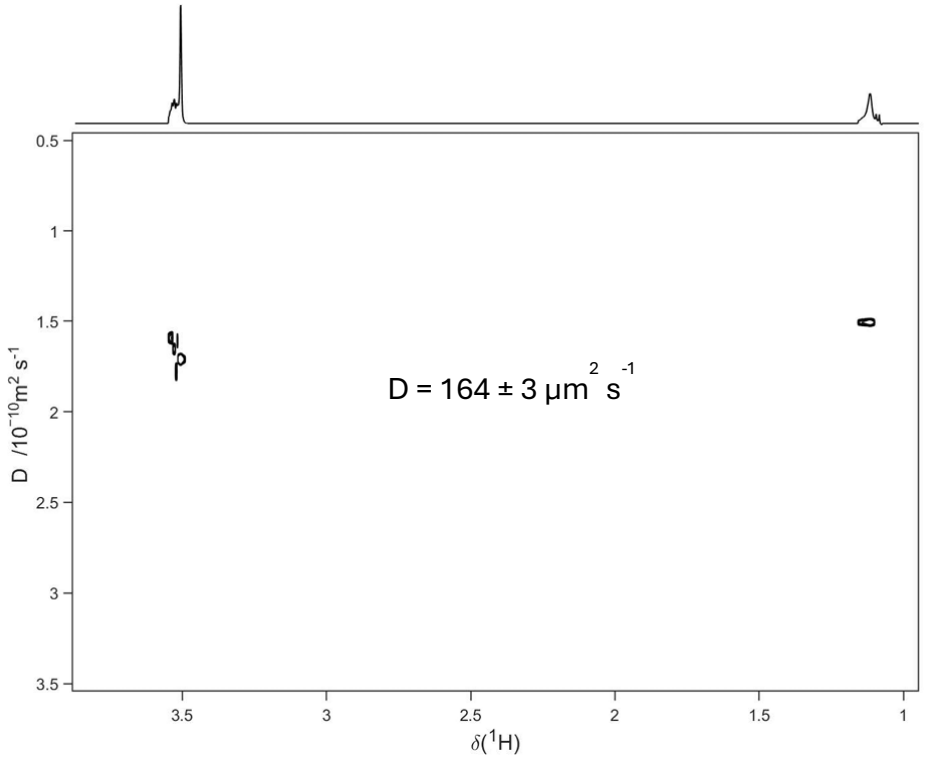

*Figure S20: Repeat 1 for **$C12_2$-HEG_2** with NH_4OAc.*

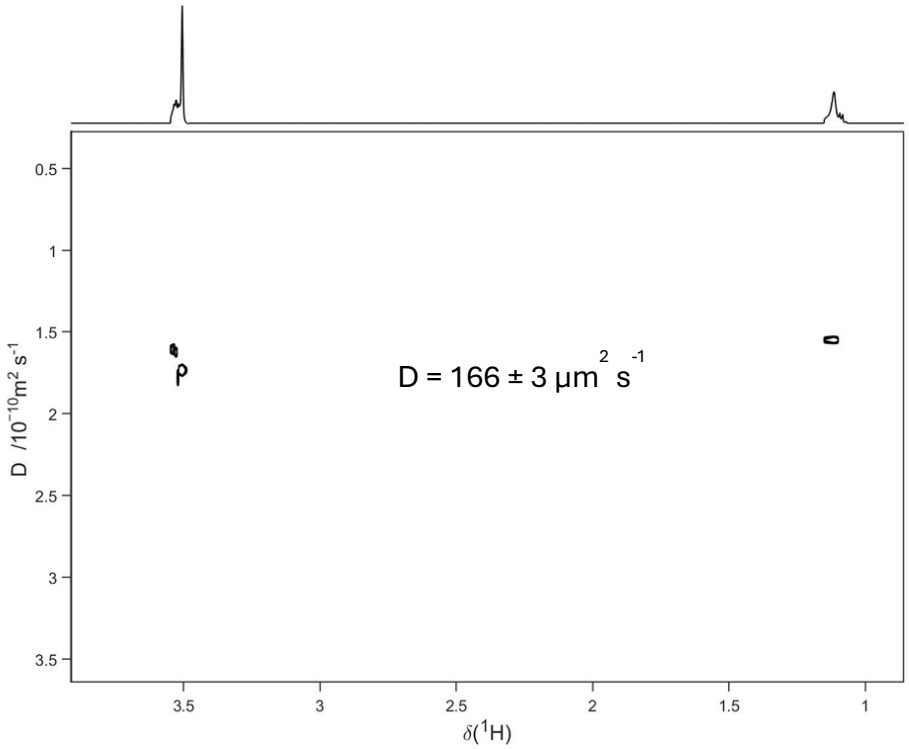

*Figure S21: Repeat 2 for **$C12_2$-HEG_2** with NH_4OAc.*

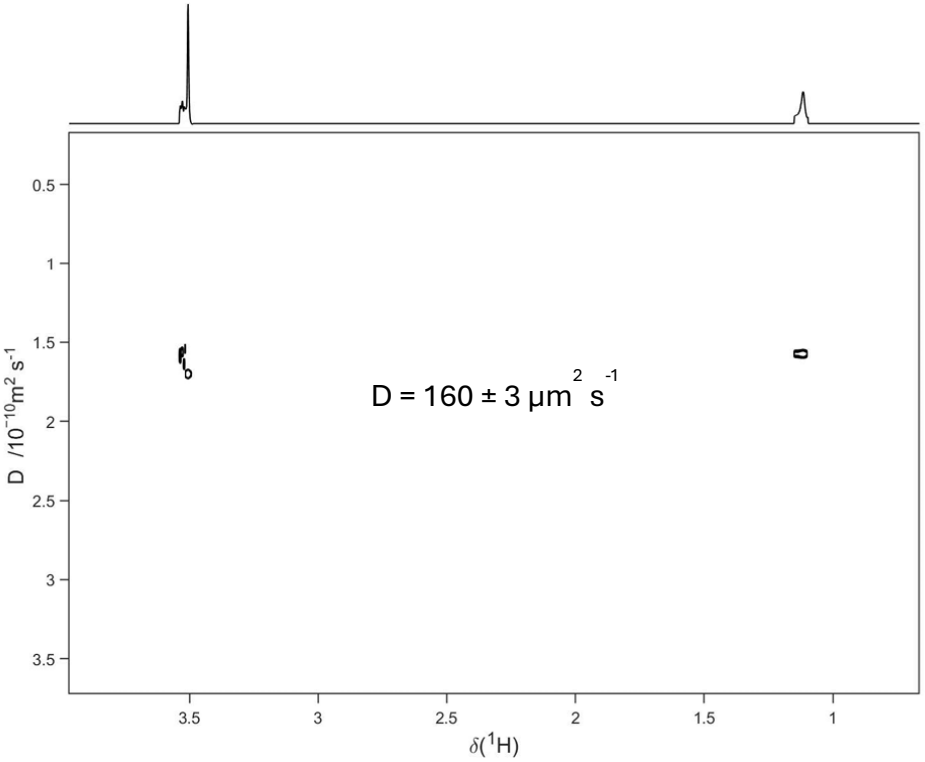

*Figure S22: Repeat 3 for **$C12_2$-HEG_2** with NH_4OAc.*

$(C12\text{-}HEG)_2$

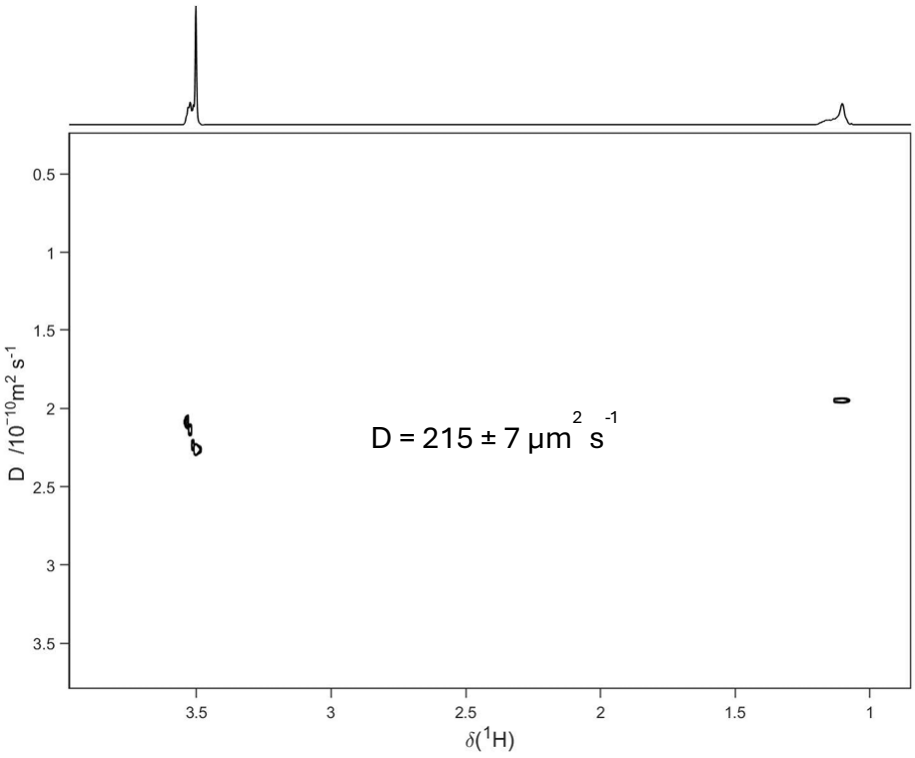

*Figure S23: Repeat 1 for **$(C12\text{-}HEG)_2$** with NH_4OAc.*

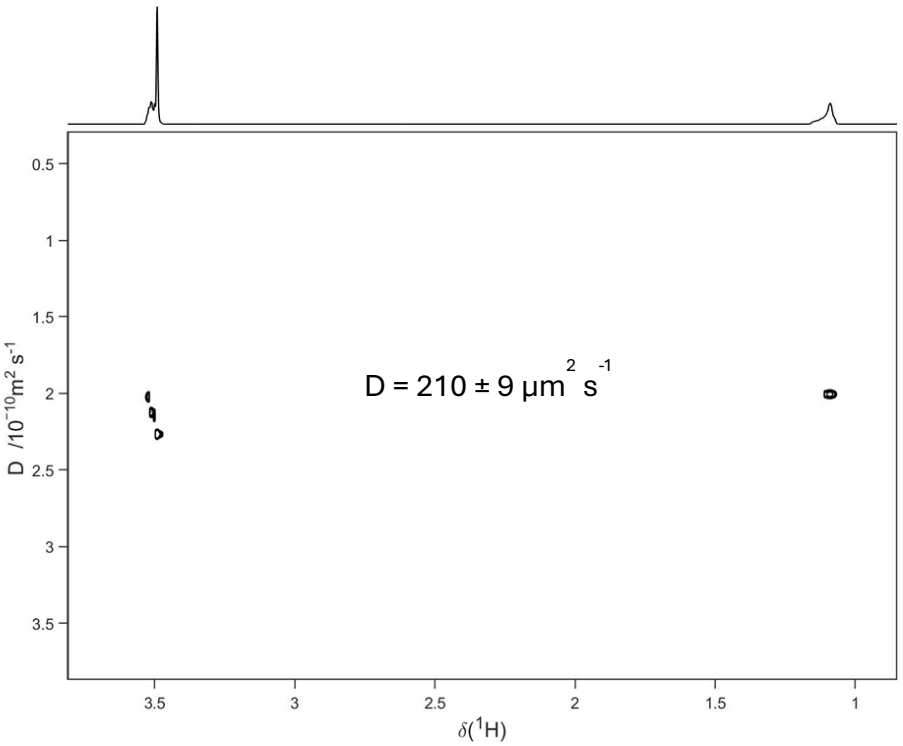

Figure S24: Repeat 2 for ***(C12-HEG)₂*** *with* NH_4OAc.

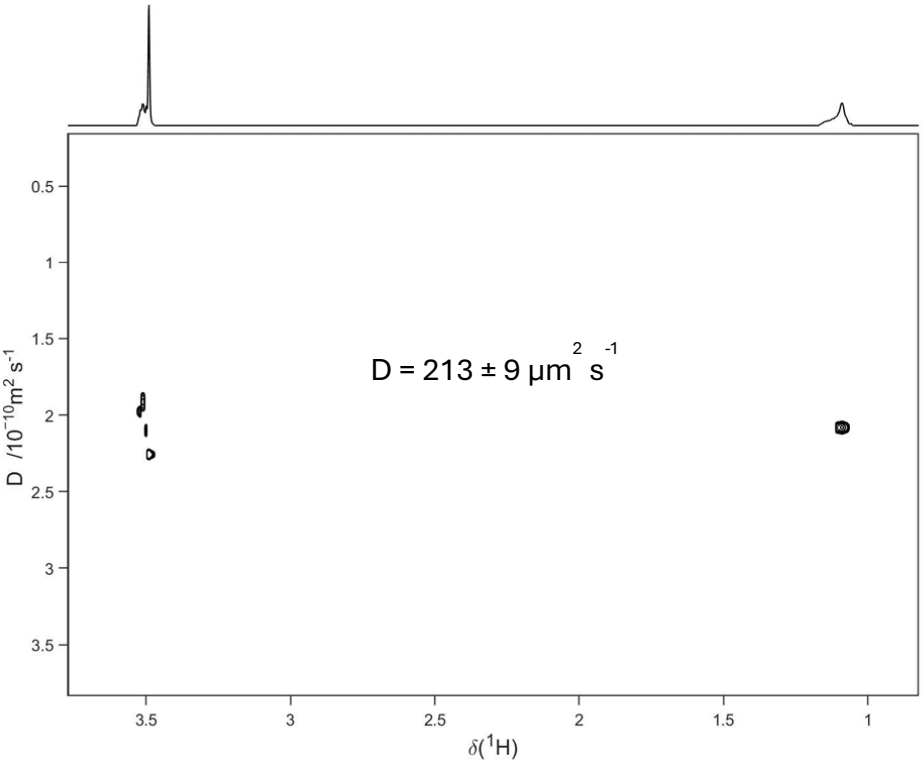

Figure S25: Repeat 3 for ***(C12-HEG)₂*** *with* NH_4OAc.

C12-HEG-HEG-C12

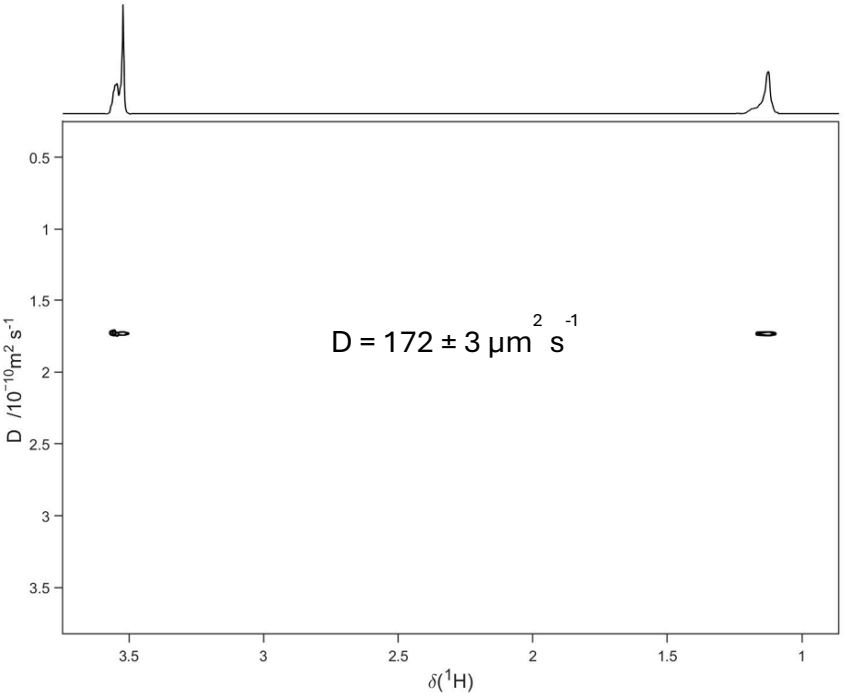

*Figure S26: Repeat 1 for **C12-HEG-HEG-C12** with NH_4OAc.*

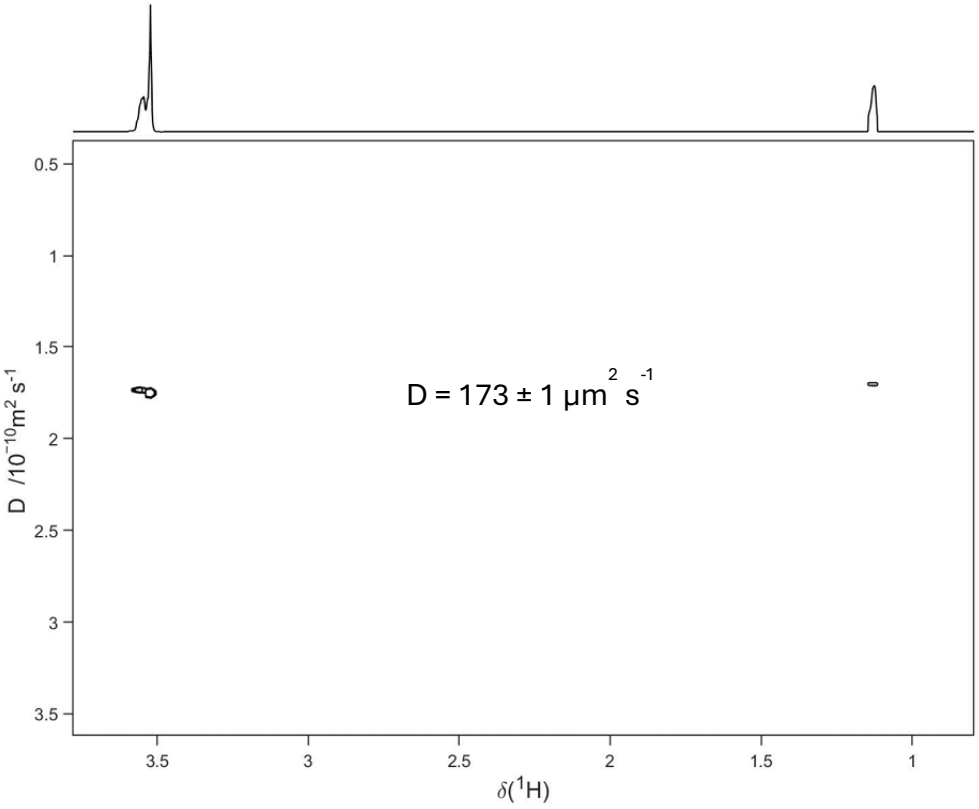

*Figure S27: Repeat 2 for **C12-HEG-HEG-C12** with NH_4OAc.*

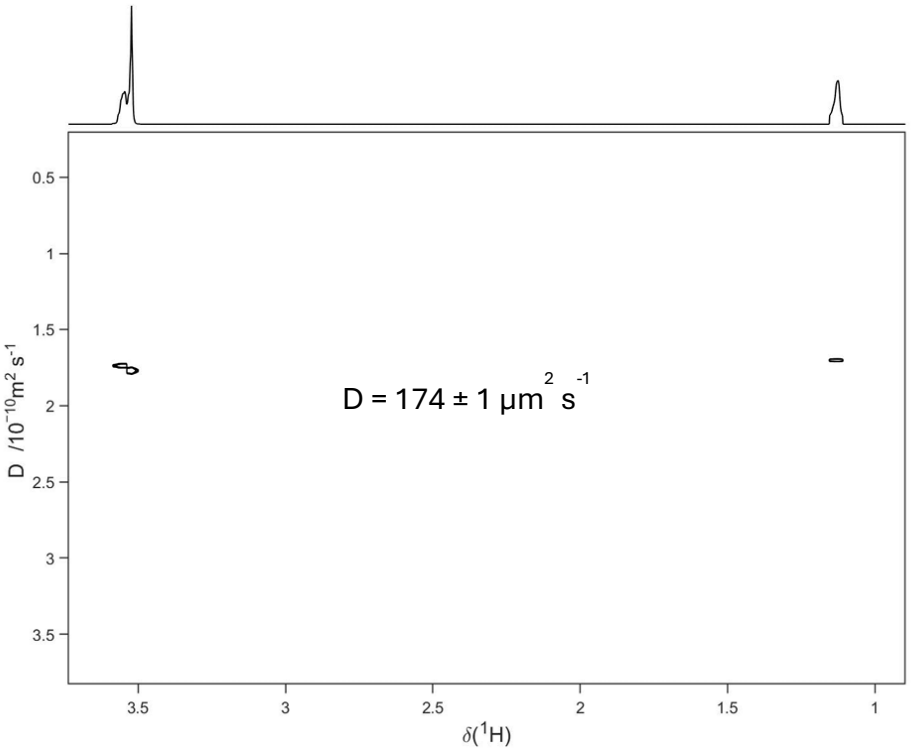

*Figure S28: Repeat 3 for **C12-HEG-HEG-C12** with NH_4OAc.*

HEG-C12-C12-HEG

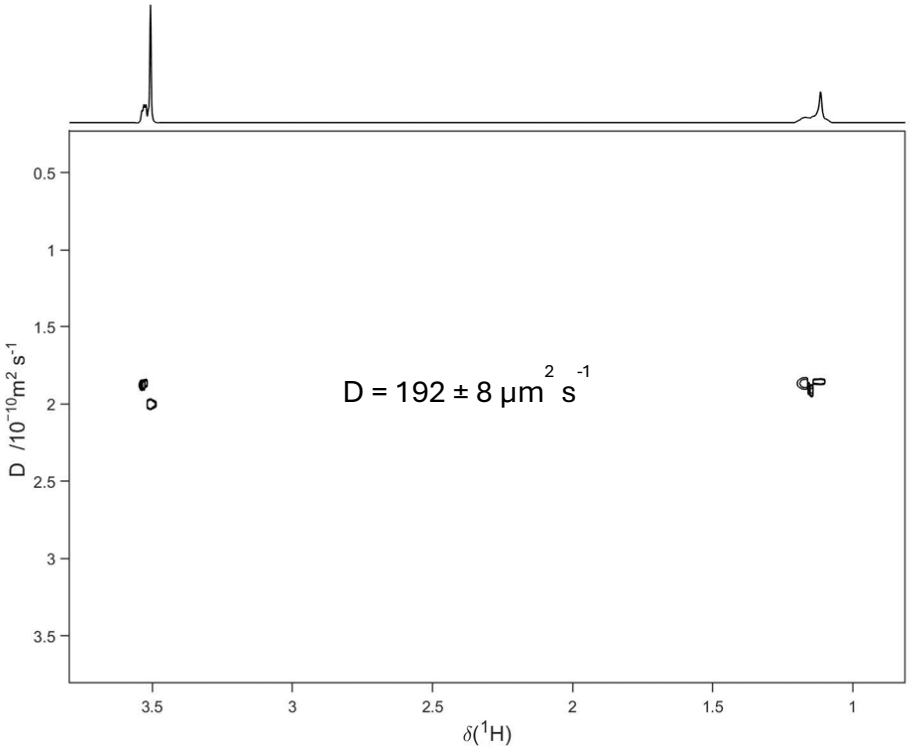

*Figure S29: Repeat 1 for **HEG-C12-C12-HEG** with NH_4OAc.*

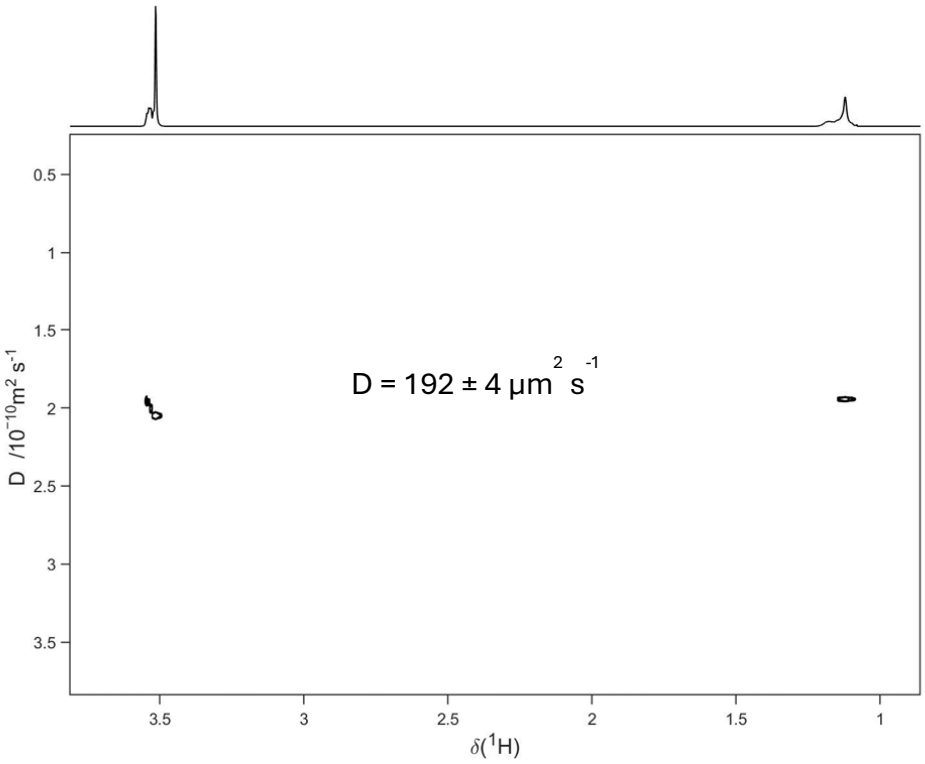

*Figure S30: Repeat 2 for **HEG-C12-C12-HEG** with NH_4OAc.*

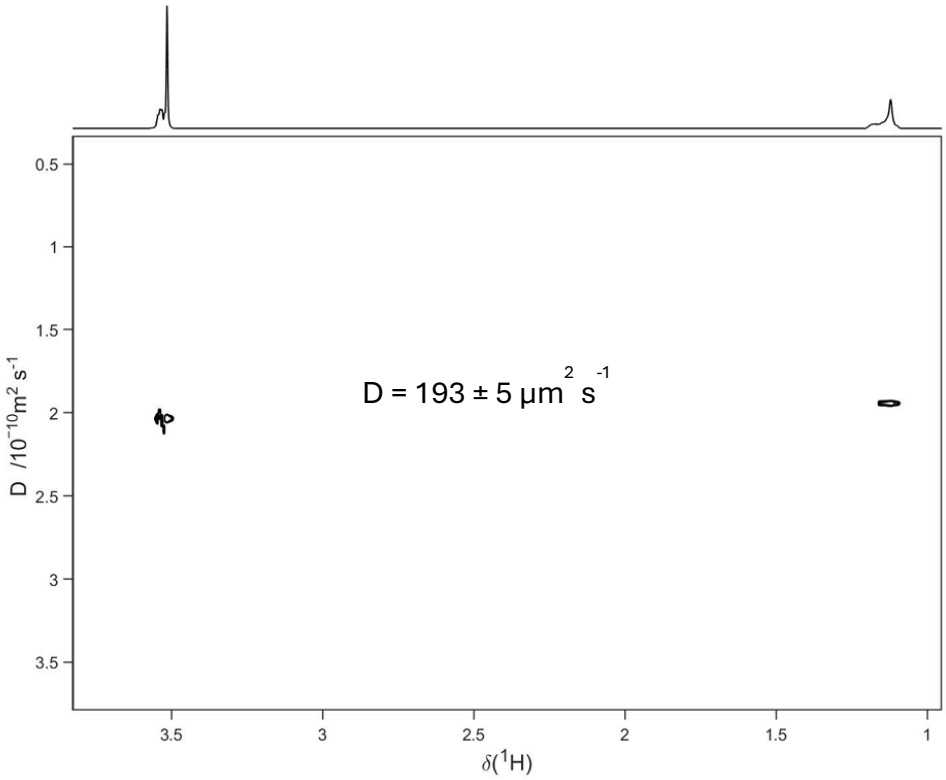

*Figure S31: Repeat 3 for **HEG-C12-C12-HEG** with NH_4OAc.*

Diffusion Coefficients plotted against Average Block Length

We plotted the experimental diffusion coefficients of the tetramers in 1 M NH_4OAc against the average block length in each tetramer and established a correlation between the two, i.e. the longer the average block length, the lower the diffusion coefficient.

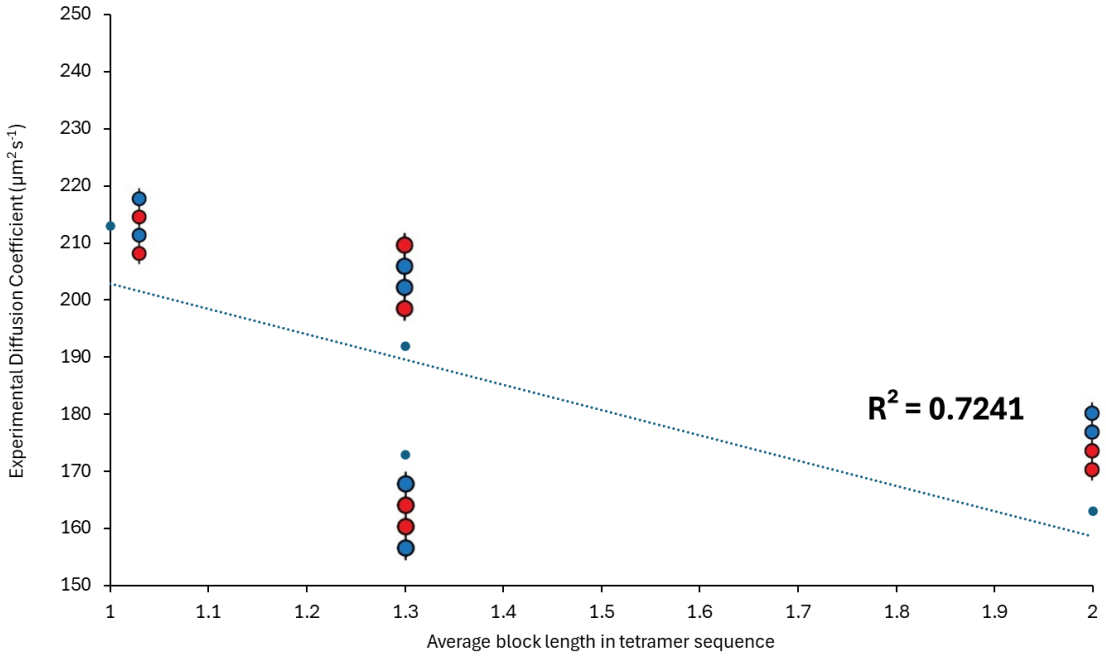

Figure S32: Plot of tetramer diffusion coefficients (1 mM in 1 M NH_4OAc) against average block length in each tetramer. Average block length in each tetramer was determined by dividing the number of monomers (4) by the number of blocks in the sequence (2, 3 or 4).

Supplementary Computational Data

Methods. A coarse-grained system consisting of the tetramers in water was simulated using the Martini force-field[2] within the GROMACS simulation package (version 2019.2).[3,4] In Martini, each coarse-grained (CG) bead represents ~4 heavy atoms, which significantly decreases computational cost. We used standard parameters for the alkane chain and phosphate group (i.e., C1 and Qa respectively, with bond lengths 0.47 nm). The initial system consisted of 200 randomly distributed molecules solvated in 30,000 CG water with 600 ammonium cations to neutralise the system. The ammonium cations were modelled using a +1-charged bead (Qd) designed to reproduce their hydrogen-bonding donor capabilities. Berendsen thermostat and pressure weak coupling[5] were applied with coupling constants of 3 ps^{-1} and 12 ps^{-1}, respectively. Two separate thermostats have been used for water, and the tetramers, both kept at 300 K.

Four 2 μs simulations were performed, each having 200 tetrameric phosphoestamers. The progress of the simulations was monitored by the solvent-accessible surface area (SASA), employing a 0.26 nm probe. Cluster analysis of the final structure was achieved with agglomerative clustering implemented in scikit-learn,[6] with a single linkage criterium with a 0.85 nm distance threshold. The tetramers have been clustered using the distance between aliphatic groups.

Clustering. The distance threshold for clustering was chosen using the silhouette score. The silhouette score also shows the quality of the clustering. The data is clustered using distances between hydrophobic **C12** groups since they are buried inside the structure and less likely to be in contact with molecules from another assembly. We observed that the clustering algorithm has the largest score for both systems above 0.85 nm (see Figure S33). Therefore, this value has been chosen for clustering. The largest clusters are present in the text.

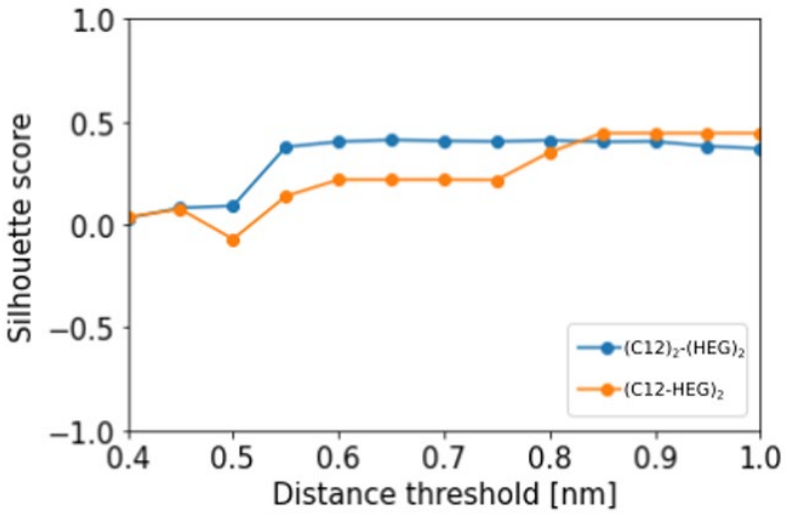

Figure S33: Silhouette score for agglomerative clustering of the simulated final assembly of the diblock and alternating tetramers.

Results

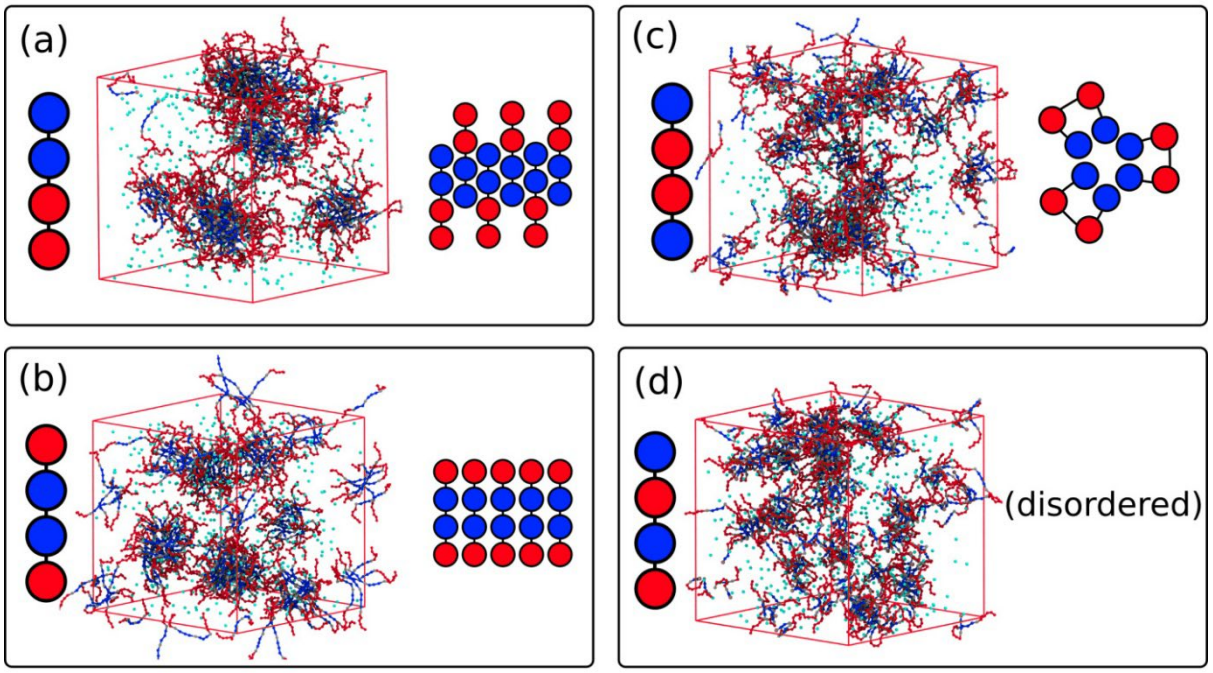

Figure S34: Snapshots of the final frame of the 2 µs simulations for (a) ***C12$_2$-HEG$_2$****, (b)* ***HEG-C12-C12-HEG****, (c)* ***C12-HEG-HEG-C12,*** *and (d)* ***(C12-HEG)$_2$****. Red and blue beads represent phosphoestamer and turquoise counterions. Water is not shown for clarity. The right side of the panels presents a schematic representation of the largest observed cluster.*

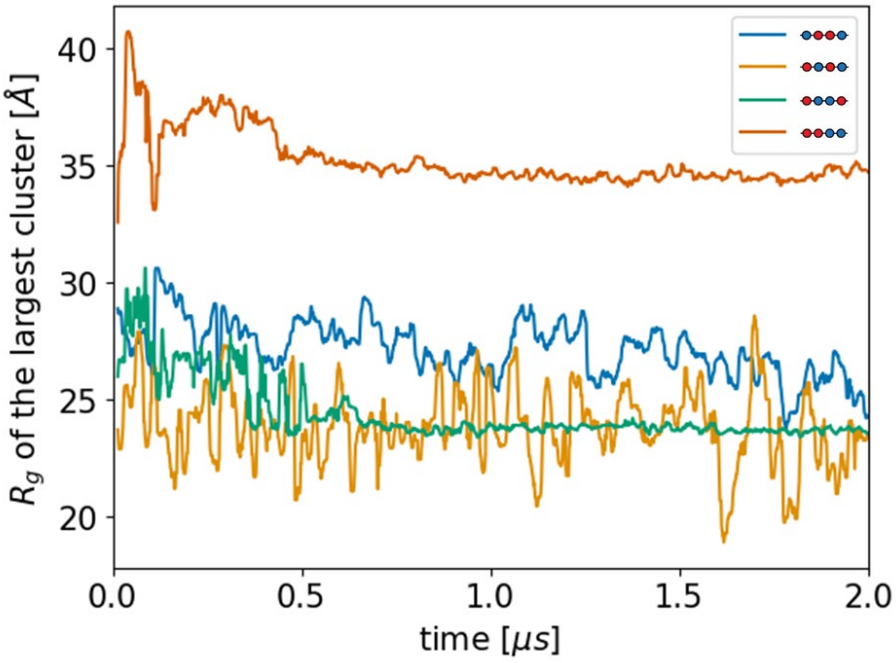

*Figure S35: Evolution of the radius of gyration (R_g) of the largest cluster simulated for each tetramer. This gave mean R_g values of 34.6 Å (****C12$_2$-HEG$_2$****), 23.6 Å (****(C12-HEG)$_2$****), 26.7 Å (****C12-HEG-HEG-C12****) and 23.8 Å (****HEG-C12-C12-HEG****).*

Additional Data

Tetramer Sequence	R_h in pure D_2O (nm)	R_h in 1 M NH_4OAc (nm)
$C12_2$-HEG_2	0.86	1.22
$(C12\text{-}HEG)_2$	0.80	0.93
C12-HEG-HEG-C12	0.85	1.15
HEG-C12-C12-HEG	0.82	1.03

Figure S36: Table of hydrodynamic radii (R_h) of the tetramer self-assemblies calculated using the Stokes-Einstein equation $R_h = k_BT/6\pi\eta D$. The viscosity (η) value for D_2O at 25 °C (1.1. x 10^{-3} Pa s) was used in all calculations.[7]

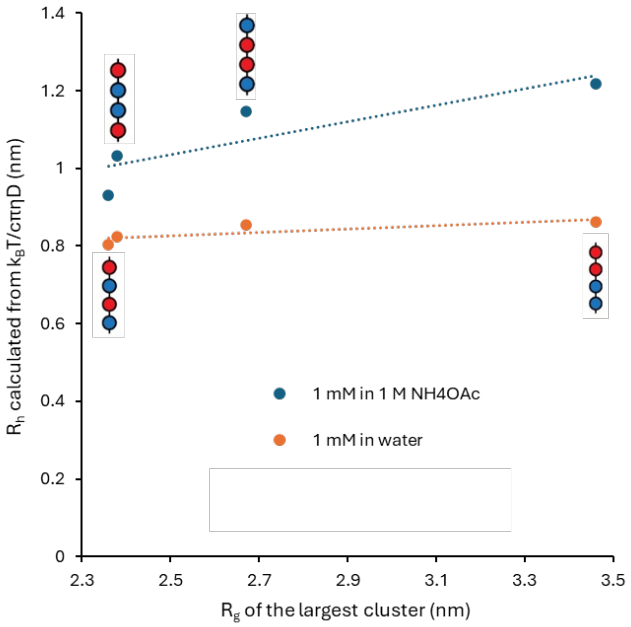

Figure S37: Plot of tetramer self-assembly hydrodynamic radius (R_h) against radius of gyration (R_g) of the largest cluster simulated for each tetramer. Since the cation was not varied in the simulation, the value on that axis is fixed for each sequence.

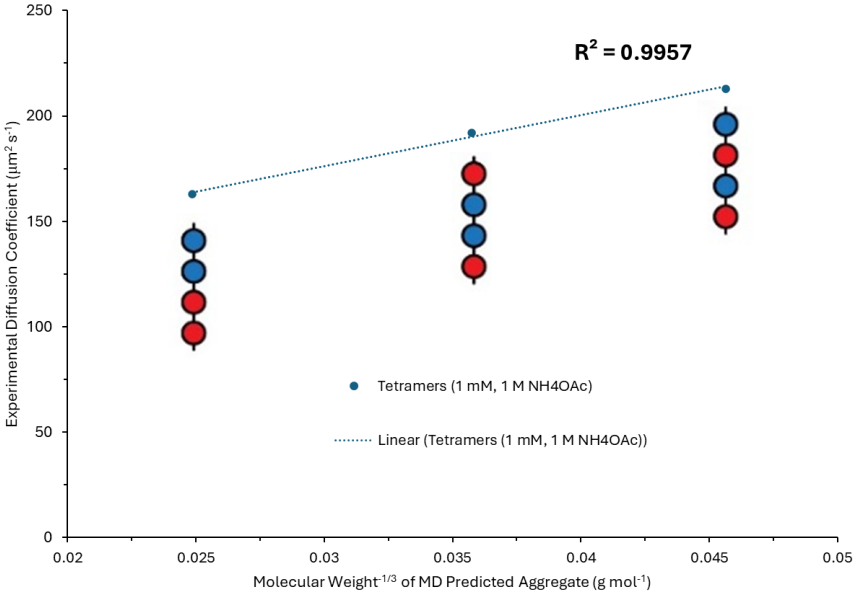

*Figure S38: Plot of tetramer diffusion coefficient against $M_w^{-1/3}$ of predicted aggregates without **C12-HEG-HEG-C12.***

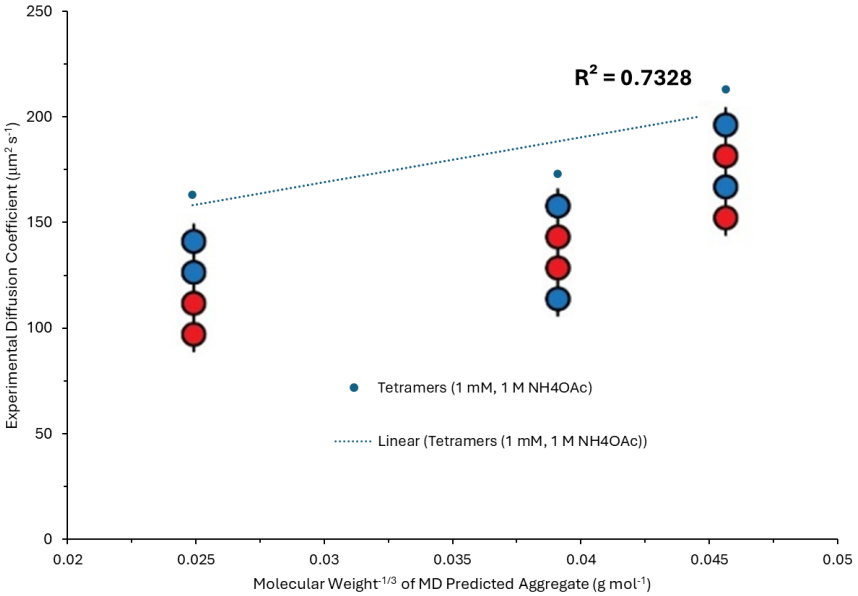

*Figure S39: Plot of tetramer diffusion coefficient against $M_w^{-1/3}$ of predicted aggregates without **HEG-C12-C12-HEG.***

References

(1) Castañar, L.; Poggetto, G. D.; Colbourne, A. A.; Morris, G. A.; Nilsson, M. The GNAT: A New Tool for Processing NMR Data. *Magn Reson Chem* **2018**, *56* (6), 546–558. https://doi.org/10.1002/mrc.4717.

(2) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824. https://doi.org/10.1021/jp071097f.

(3) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.
(4) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
(5) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.
(6) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (null), 2825–2830.
(7) Millero, F. J.; Dexter, Roger.; Hoff, Edward. Density and Viscosity of Deuterium Oxide Solutions from 5-70.Deg. *J. Chem. Eng. Data* **1971**, *16* (1), 85–87. https://doi.org/10.1021/je60048a006.