Supporting Material

Chain-length heterogeneity allows for the assembly of fatty acid vesicles in dilute solutions

Itay Budin[1,2], Noam Prwyes[1], Na Zhang[1,3] , and Jack W. Szostak[1]

[1] Howard Hughes Medical Institute, Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge St., Boston, MA 02114, USA

[2] Miller Institute for Basic Research in Science, University of California, Berkeley, 2536 Channing Way, Berkeley, CA 94720, USA

[3] High Magnetic Field Laboratory, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, P. R. China

This file includes Supplemental Text, Figures S1-S6, Table S1, and NMR spectra for Figure 7 and Table 1.

Derivation of the critical aggregation concentration (cac) of a multi-component mixture

We take the approach of Holland (1) to derive the cac for an ideal mixture of fatty acids. The chemical potential of species *i* as a monomer can be expressed as:

$$(1) \qquad \mu_i = \mu_i^{\circ} + RT \ln C_i^m$$

μ_i is the monomeric chemical potential of species *i*, μ_i° is its standard chemical potential, and C_i^m is the concentration of the monomer *i* in units of mole fraction in solvent (i.e. molar concentration divided by ~55). This assumes the activity coefficient of the free monomer is 1 (acts ideally), approximating free dispersal of monomers with no interactions. In a mixed aggregate,

$$(2) \qquad \mu_i^a = \mu_i^{a,\circ} + RT \ln f_i x_i$$

μ_i^a is the chemical potential of *i* in the mixed aggregate, $\mu_i^{a,\circ}$ is the chemical potential of *i* in a pure aggregate of species *i*, f_i is the activity coefficient for *i* in the aggregate (a measure of its interactions with other molecules), and x_i is the mole fraction of *i* in the mixed aggregate. For the pure aggregate of *i*, we can apply a pseudophase model, and thus:

$$(3) \qquad \mu_i^{a,\circ} = \mu_i^o + RT \ln cac_i$$

cac_i is the critical aggregation concentration (cac) for species *i* in a single-component system. At the concentration of the cac, monomers and aggregates are isoenergetic and the free energy of the monomer species is equal to that of the aggregate. Therefore,

$$(4) \qquad \mu_i^a = \mu_i$$

Substituting in Eq. 1 and 2:

$$(5) \quad \mu_i^{a,\circ} + RT \ln f_i x_i = \mu_i^{\circ} + RT \ln C_i^m$$

and Eq. 3:

$$(6) \quad \mu_i^{a} + RT \ln cac_i + RT \ln f_i x_i = \mu_i^{\circ} + RT \ln C_i^m$$

which simplifies to:

$$(7) \quad C_i^m = f_i x_i cac_i$$

If we take the case of a solution at the mixed cac (right at the formation of the first aggregates), the concentration of monomer i, C_i^m , is simply a product of its mole fraction of total lipids, X_i, and the total lipid concentration, which at that point is equal to the mixed cac (cac). Therefore,

$$(8) \quad C_i^m = X_i cac$$

Combining with Eq. 7 and rearranging:

$$(9) \quad \frac{X_i cac}{f_i cac_i} = x_i$$

The mole fractions of components i in the mixed aggregate of k components must sum to 1,

$$\sum_{i=1}^{k} x_i = 1$$

Therefore,

$$\sum_{i=1}^{k} \frac{X_i cac}{f_i cac_i} = 1 = cac \sum_{i=1}^{k} \frac{X_i}{f_i cac_i}$$

$$(10) \quad \frac{1}{cac} = \sum_{i=1}^{k} \frac{X_i}{f_i cac_i}$$

This is a general expression for the cac of a system (in total mole fraction units) with n components, each with mole fractions X_i, individual critical concentrations cac_i for their behavior

in single-component systems, and activity coefficient f_i. For micelles, fatty acids have shown ideal behavior (2), i.e. an activity coefficient of 1, thus potentially simplifying this further.

Estimation of vesicle partition of an individual species by NMR line width measurements

Because line width represents a weighted average between the species in a vesicle aggregate and as monomers (or small micelles), vesicle composition can be roughly estimated from the corresponding line widths of each component:

$$(11) \quad lw = X_m lw_m + X_a lw_a$$

Where lw is the measured line width of the sample, X_m the monomer fraction, lw_m the line width for a pure monomer sample, X_a is the aggregate (vesicle) fraction, and lw_a is the line width for a pure vesicle aggregate fraction. This is a simplification of actual fatty acid systems, which feature a ternary mixture of monomers, vesicles, and micelles (3). However, micelle samples prepared either at high pH (>10) or in the presence of detergent (Triton X-100) featured similar line widths (~2 Hz) to those of monomeric samples (i.e., those below the cac) and therefore these two states can be grouped together. A more significant assumption is that each of these states features a characteristic line width, regardless of concentration or their chemical environment. This is obviously a crude approximation, but we can test it against the line width vs. concentration data for 1-^{13}C decanoic acid from Fig. 7. We expect two regimes of line width: 1) below the cac, $lw = lw_m$, since $X_m = 1$ and 2) above the cac,

$$(12) \quad lw = X_m lw_m + X_a lw_a = \left(\frac{cac}{c}\right) lw_m + \left(1 - \frac{cac}{c}\right) lw_a$$

We set cac to 40 mM, as measured by turbidity and pinacyanol, lw_m to 2 Hz and performed a least squares fit of Eqn. 12 on the data for c > 40 mM to solve for lw_a (Fig. S6). This yielded a

reasonable fit (R^2 = 0.96) for the concentrations tested with a value for lw_a of ~130 Hz . In the case of pure oleic acid, which features a very low cac compared to the concentrations needed for NMR measurements, line width was independent of concentration under tested conditions (> 5 mM) at ~200 Hz, though fitting of Lorentzian functions to such broad peaks is a challenge. We, however, expect the value for lw_a to be intrinsically lower for decanoic acid, due to the abundance of coexisting micelles (which feature narrow peaks) characteristic of short-chain fatty acid vesicle solutions in these concentration ranges (3, 4).

While an approximation, a constant lw_a allows for the vesicle partition factor to be calculated from measured line widths. Setting the sum of X_m and X_a to 1, we can simplify Eq. 11:

$$X_a = \frac{lw - lw_m}{lw_a - lw_m} \quad (13)$$

Using the values above for decanoic acid, estimated partition coefficients were calculated for the systems and are shown in Table S1. While these partitions are dependent on a extrapolated value for lw_a (130 Hz, in this case), their relative values are not (e.g. at 50 mM vs. 15 mM). If we assume that the oleic acid features a very high aggregation partition (e.g. X_a of 1), we can further deduce vesicle compositions from these partition coefficients.

Supporting References

1. Holland, P.M. and D.N. Rubingh. 1983. Nonideal multicomponent mixed micelle model. *J. Phys. Chem.* 87:1984-1990.
2. Shinoda, K. 1954. The Critical Micelle Concentration of Soap Mixtures (Two-Component Mixture). *J. Phys. Chem.* 58:541-544.
3. Budin, I., A. Debnath, and J.W. Szostak. 2012. Concentration-driven growth of model protocell membranes. *J. Am. Chem. Soc.* 134:20812-20819.

4. Dejanović, B., K. Mirosavljević, …, P. Walde. 2008. An ESR characterization of micelles and vesicles formed in aqueous decanoic acid/sodium decanoate systems using different spin labels. *Chem. Phys. Lipids.* 156:17-25.

Supporting Figures

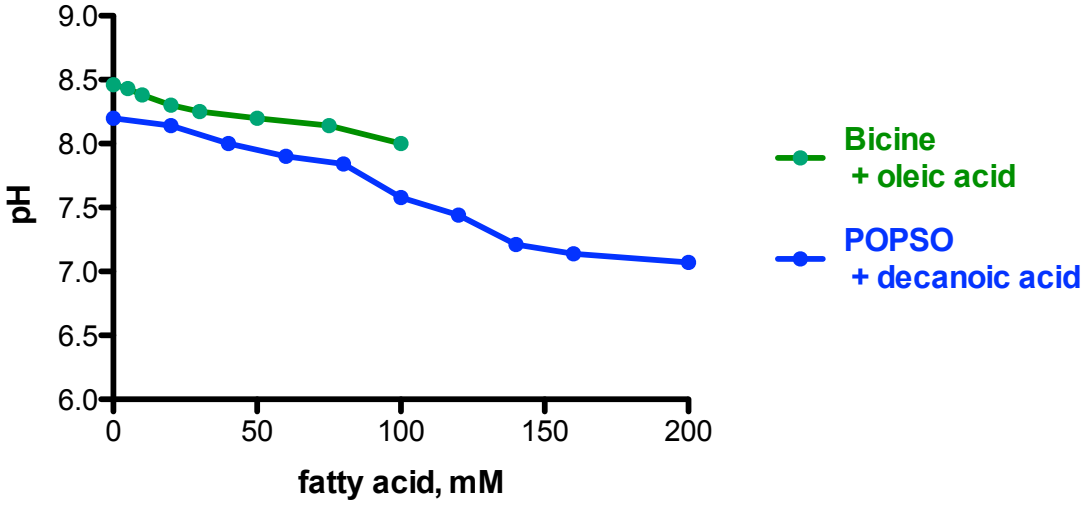


FIGURE S1 Effect of fatty acid concentration on the pH of buffered solutions used. The pH was measured of solutions of decanoic acid in 0.1 M POPSO (titrated to pH 8.2 with NaOH) and oleic acid in 0.2 M Bicine (titrated to pH 8.5 with NaOH). Connecting lines added for clarity. Samples were prepared at 30 °C and measurements taken at 23 °C.

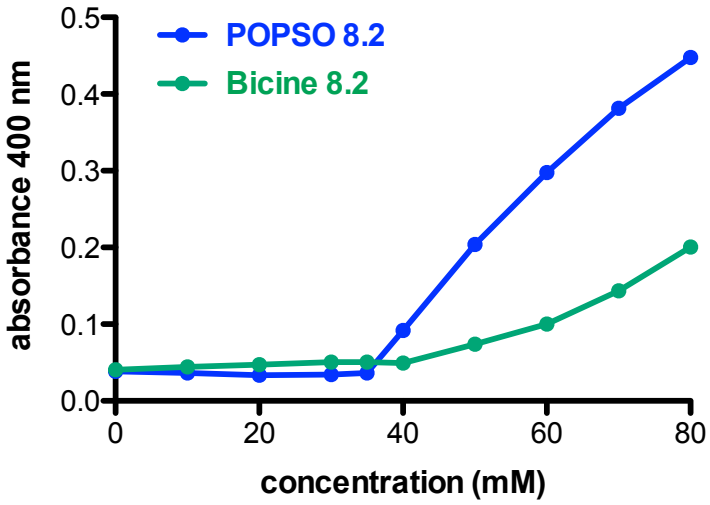


FIGURE S2 Effect of buffer used on decanoic acid vesicle assembly. POPSO and Bicine buffers, both titrated to pH 8.2 with NaOH, support decanoic acid vesicles in solutions above the cac (~40 mM in both cases). The POPSO buffers supports more abundant and/or larger vesicles, as indicated by the higher turbidity of the solution. Connecting lines added for clarity. Experiments were performed at 30 °C.

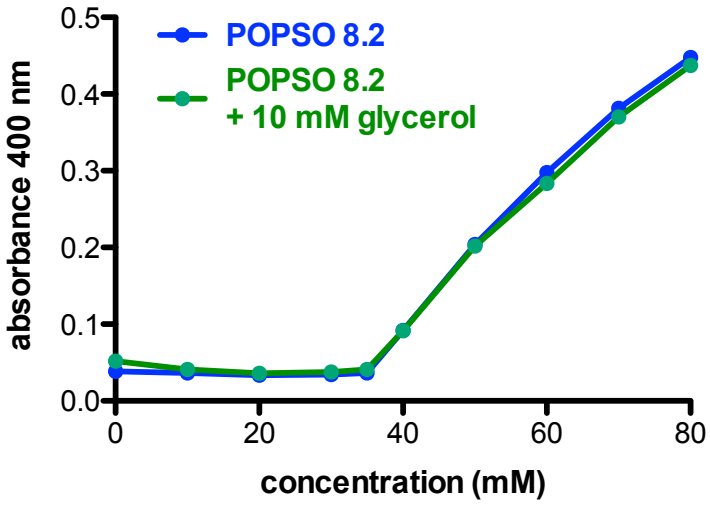


FIGURE S3 Effect of glycerol on decanoic acid vesicle assembly. The concentration of glycerol (10 mM) included in NMR experiments as an internal standard does not noticeable affect decanoic acid vesicle assembly as assayed by solution turbidity. Connecting lines added for clarity. Experiments were performed at 30 °C.

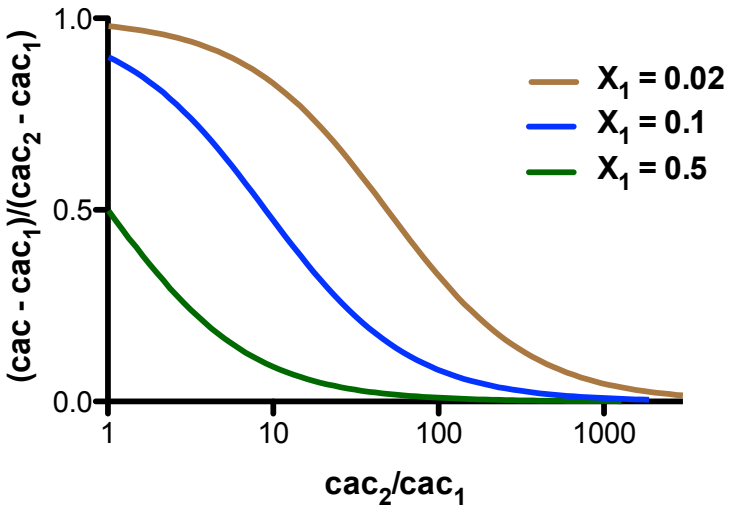


FIGURE S4 Predicted mixed cac values of a binary mixture (cac) converge to that of the longer chain species (cac_1) as the ratio of the two individual cacs (cac_2/cac_1) becomes large. Values were calculated from Eq. 4 for different mixture stoichiometries (X_1, the mole fraction of component 1). A cac_2/cac_1 ratio of 1000 corresponds to a difference of 8 carbons in the chain length of two fatty acid components.

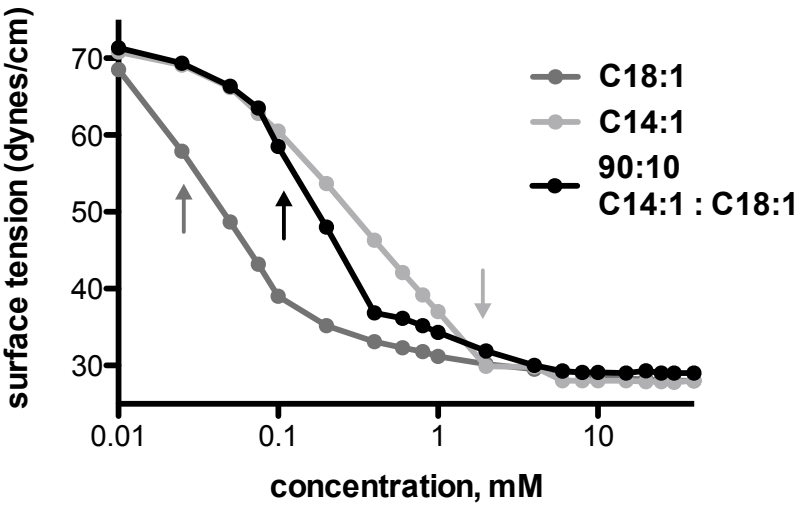


FIGURE S5 Surface tension of fatty acid solutions from Figure 2. Surface tension was measured from serial dilutions of myristoleic acid (C14:1), oleic acid (C18:1), and a 90:10 mixture of myristoleic acid : oleic acid. Surface tension decreases with increasing lipid concentration due to the ability of the amphiphilic monomers to lower the air-water interfacial energy. The decrease in surface tension plateaus due to the assembly of aggregates (vesicles or micelles), which do not interact with the interface. Arrows show the concentration at which vesicle assembly is detected using light scattering experiments shown in Figure 2. Because these concentrations are lower or equal to the concentration at which surface tension plateaus, we conclude that there is not a second, micelle critical concentration below the vesicle cac. In our model, fatty acid solutions feature only a single cac, with the structural composition of the resulting aggregate phase (i.e. vesicle vs. monomer) is dependent on solution pH. Experiments were performed at room temperature in Bicine buffer, pH 8.5.

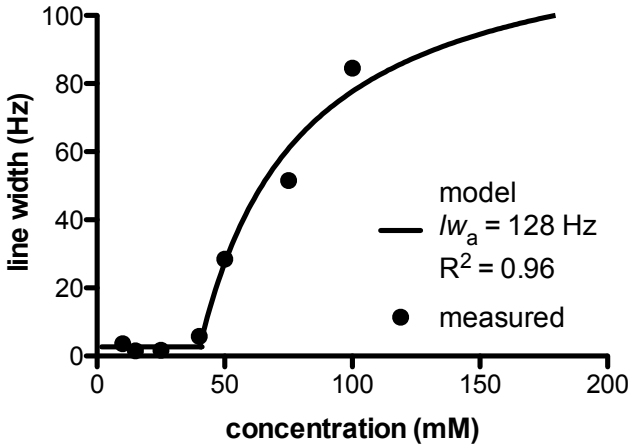


FIGURE S6 Weighted average model for 1-^{13}C decanoic acid line widths. Data from concentrations less than the cac (40 mM) were modeled as a constant line width (2 Hz); for concentrations greater than the cac, line widths were fit with a least squares regression of Eqn. 12.

Fatty acid mixture	15 mM decanoic acid	50 mM decanoic acid	15 mM 95:5 decanoic acid: oleic acid	50 mM 95:5 decanoic acid: oleic acid	15 mM 50:50 decanoic acid: oleic acid
^{13}C decanoic acid line width	1.6 Hz	28.6 Hz	14.2 Hz	49.1 Hz	36.3Hz
Estimated partition coefficient of decanoic acid into vesicles	0	0.21	0.10	0.38	0.28

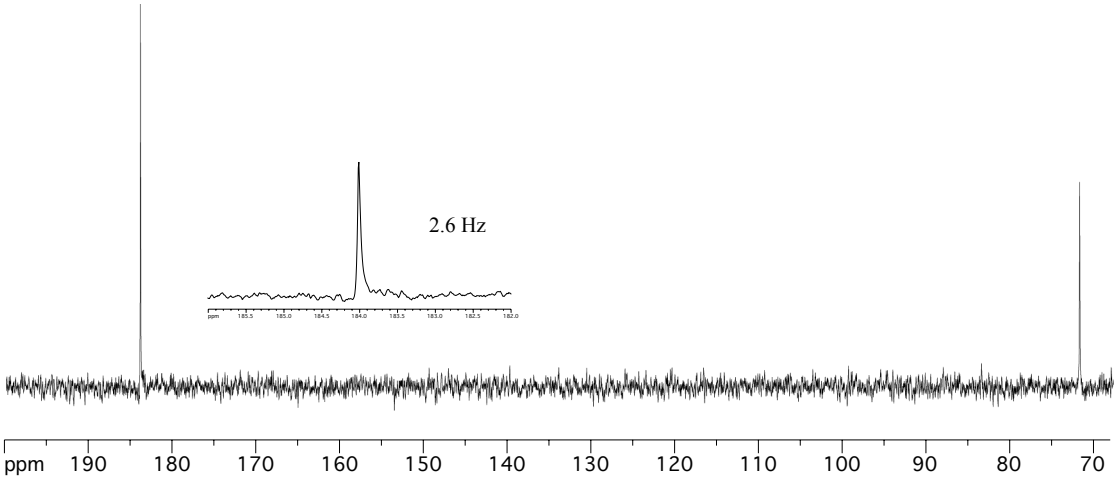
TABLE S1 Estimated partition coefficients of decanoic acid into pure or mixed vesicles as extrapolated from 1-^{13}C decanoic acid line widths. Experiments were performed in POPSO buffer at 30 °C.

1D ^{13}Carbon NMR spectra for Figure 7 and Table 1

Main spectra intensities are normalized to the 10 mM glycerol standard at 72 ppm.
Inserts show a 4 ppm region at the fatty acid peak (184 or 181 ppm) and the fitted line width.

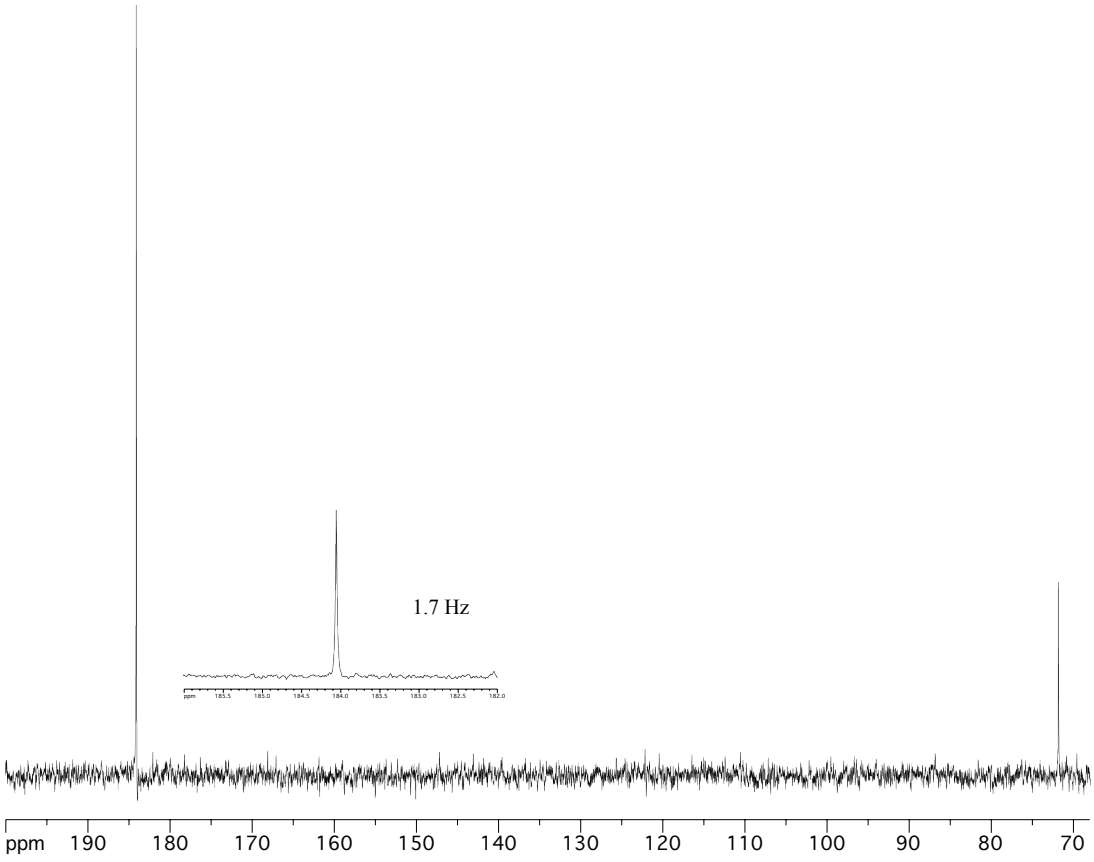
A

10 mM 1-^{13}C decanoic acid



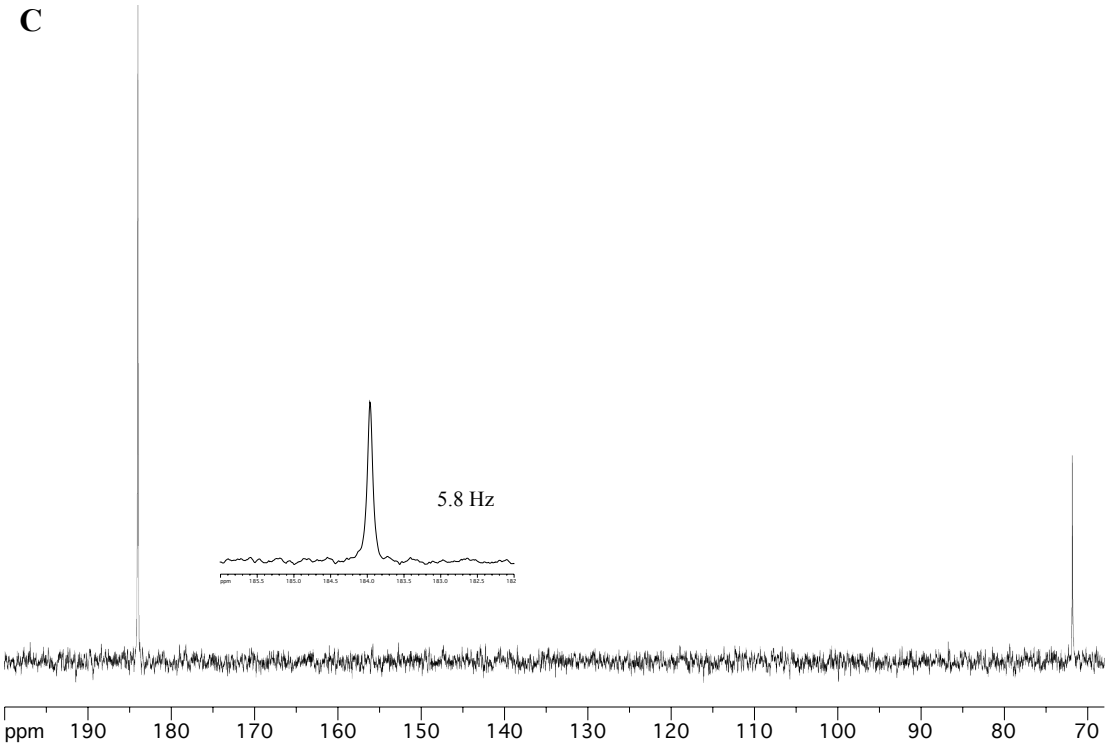
25 mM 1-^{13}C decanoic acid

B



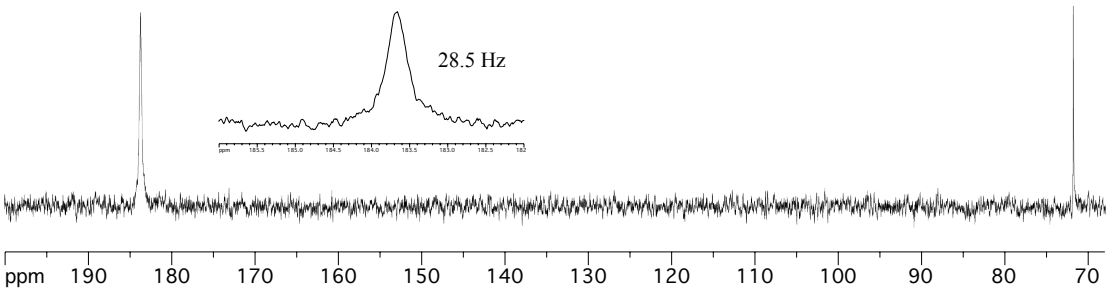
40 mM 1-^{13}C decanoic acid

C



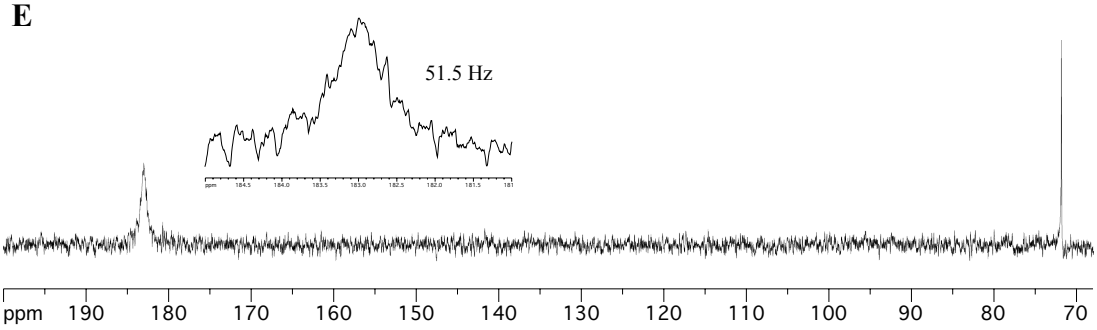
50 mM 1-^{13}C decanoic acid

D



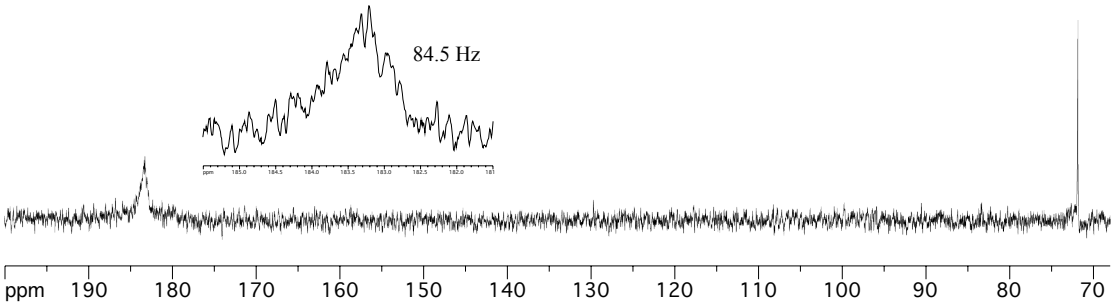
75 mM 1-^{13}C decanoic acid

E



100 mM 1-^{13}C decanoic acid

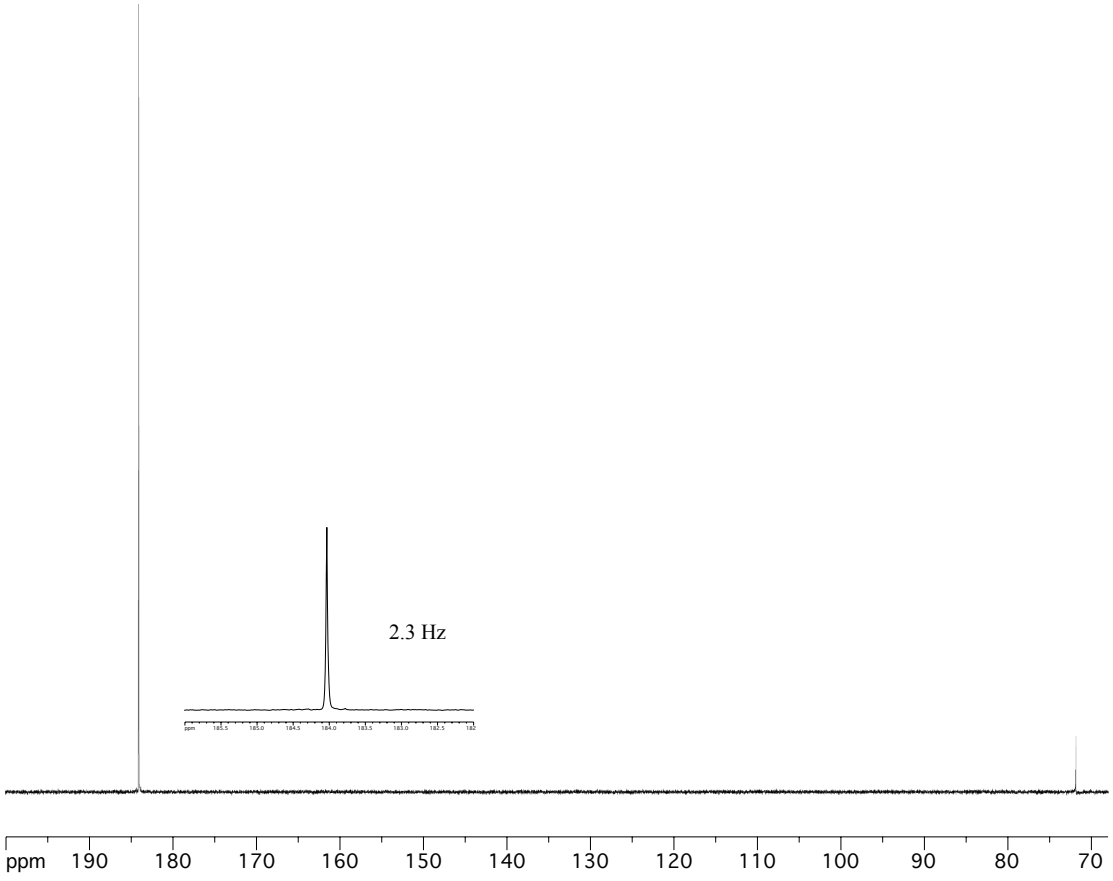
F



100 mM 1-^{13}C decanoic acid, pH 10

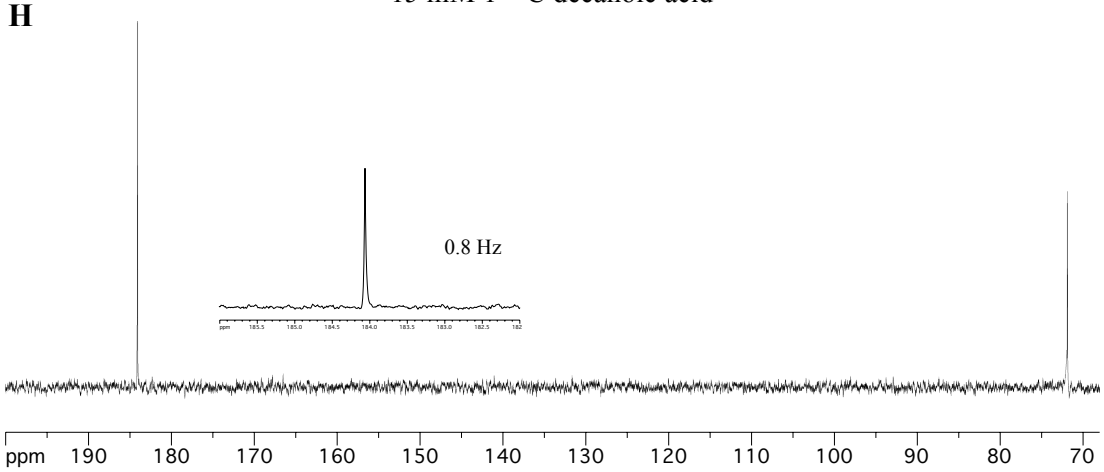
(vertical scale = 0.25x)

G



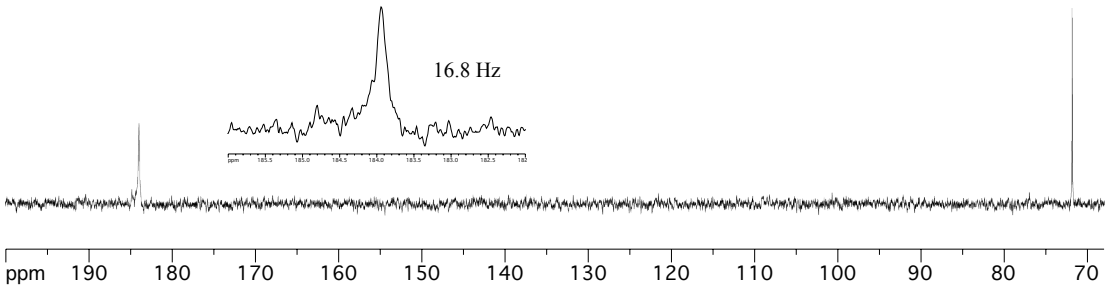
15 mM 1-^{13}C decanoic acid

H



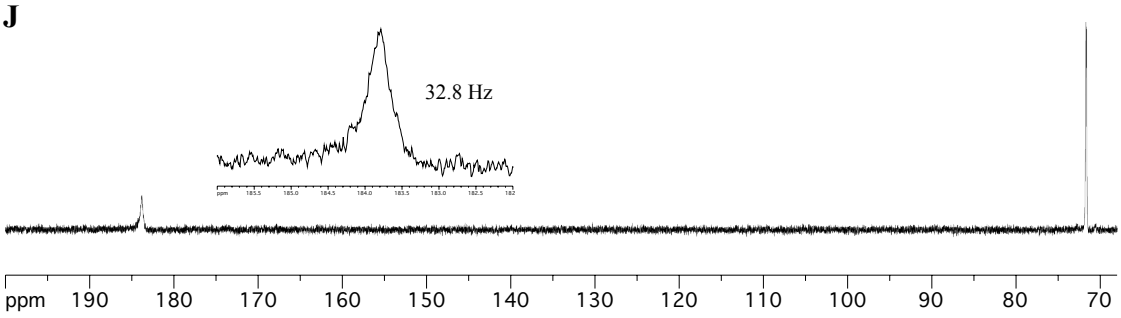
15 mM 95:5 1-^{13}C decanoic acid : oleic acid

I



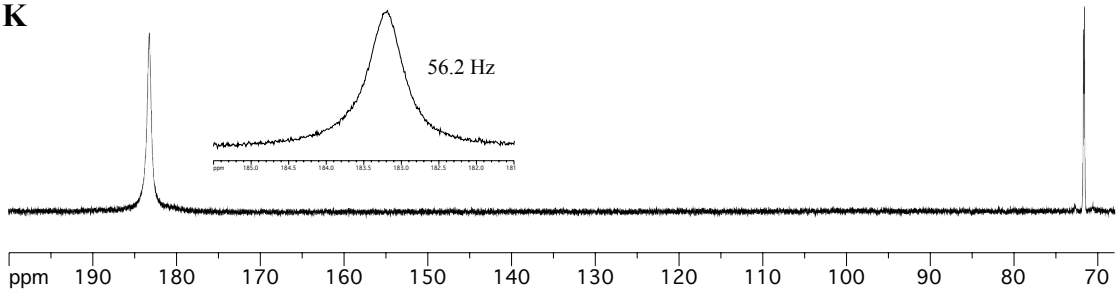
15 mM 50:50 1-^{13}C decanoic acid : oleic acid

J



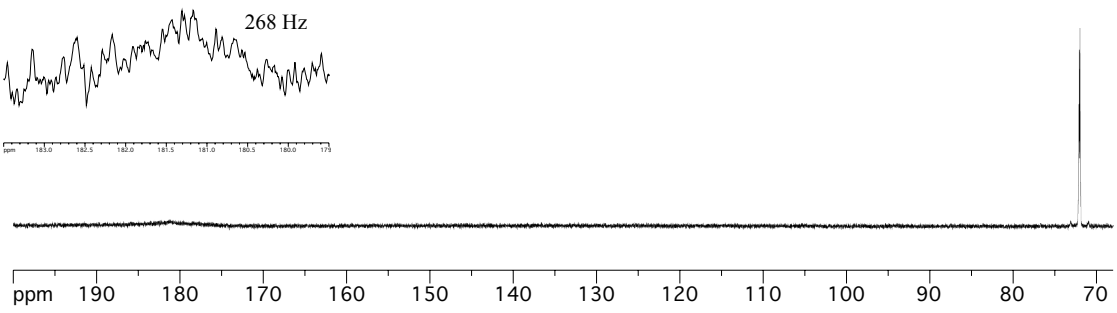
50 mM 95:5 1-^{13}C decanoic acid : oleic acid

K



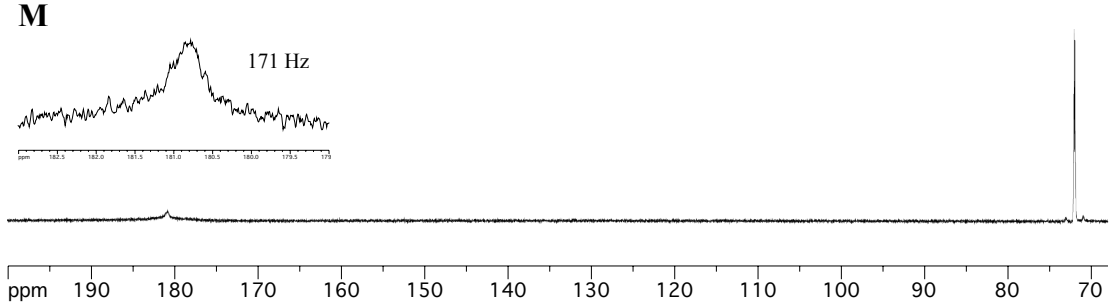
L

15 mM 1-^{13}C oleic acid



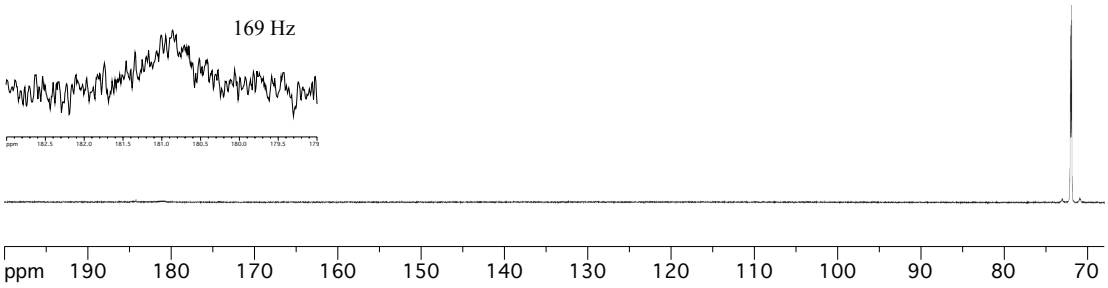
50 mM 1-^{13}C oleic acid

M



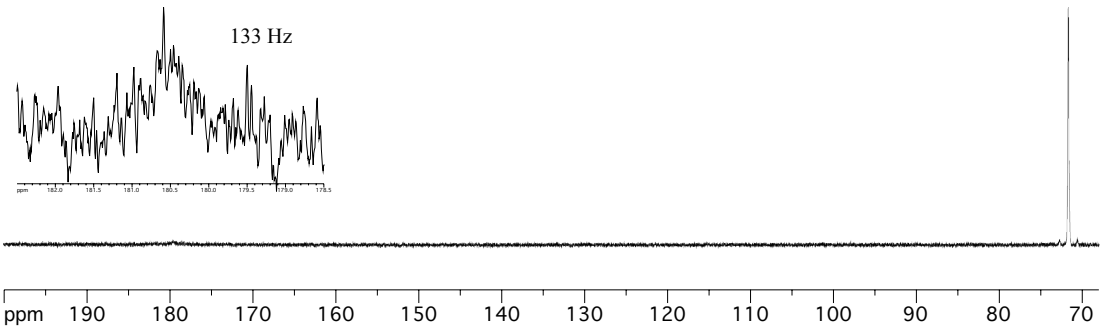
15 mM 95:5 decanoic acid : 1-^{13}C oleic acid

N



50 mM 95:5 decanoic acid : 1-^{13}C oleic acid

O



15 mM 50:50 decanoic acid : 1-^{13}C oleic acid