Supplemental Information to:

Determining Sequential Micellization Steps of Bile Salts with Multi-CMC Modeling

David Rovnyak*, Jiayi He[1], Sophie Kong[2], Kyle W. Eckenroad[3], Gregory A. Manley[4], Raeanne M. Geffert[5], Michael R. Krout, Timothy G. Strein

1 Dent Drive, Department of Chemistry, Bucknell University, Lewisburg PA 17837

*communicating author

1. current address: University of Pennsylvania, Department of Chemistry, 231 S. 34 Street, Philadelphia, PA 19104-6323, hejiayi@sas.upenn.edu

2. current address: Department of Pharmaceutical Chemistry, School of Pharmacy, University of California San Francisco, 1700 4th St, San Francisco CA 94158, Sophie.Kong@ucsf.edu

3. current address: Bristol Myers Squibb, 1 Squibb Drive, 92-218, New Brunswick, NJ, 08901, USA, Kyle.Eckenroad@bms.com

4. current address: AB SCIEX LLC , 500 Old Connecticut Path, Framingham, MA 01701, gmanley@gmail.com

5. current address: The University of North Carolina at Chapel Hill, UNC Eshelman School of Pharmacy, Division of Pharmacotherapy and Experimental Therapeutics
Kerr Hall, Campus Box 7569, Chapel Hill, NC 27599-7569, geffert@unc.edu

Sections

S1: Phase transition and mass action models for one CMC

In the phase separation (PS) model, the micellar particles are assumed to be a distinct phase, with high aggregation numbers, and the CMC is strongly discretized (i.e., the value of the observable shows a sharp discontinuity at the CMC). The PS model does not consider the aggregation number of the micelle. When $[S_{tot}] \geq CMC$, the phase separation model views the free monomer concentration as a constant saturated value:

$$
\begin{aligned}
&\text{if } [S_{tot}] < CMC \\
&\qquad f_{\text{free}} = 1 \ ; \\
&\text{if } [S_{tot}] \geq CMC \\
&\qquad [S] = CMC \ , \\
&\qquad f_{\text{free}} = \text{CMC} / [[S_{tot}]] \ , \qquad \text{and} \qquad \text{(S.1)} \\
&\qquad f_{\text{micelle}} = 1 - f_{\text{free}} \ .
\end{aligned}
$$

The general appearance of the phase transition model (Eqn 2) is illustrated in **Figure S.1** below, where low concentration data will yield o_{free}. Applying the model generally involves adjusting only the CMC and o_{micelle} values. It is often convenient to use logarithmic or square root scaling of the concentration, where the latter is used in this work.

In principle there are up to three adjustable parameters (o_{free}, o_{micelle}, *CMC*) but a model to Eqn S.1 is normally reduced to a two-parameter fit since o_{free} is effectively fixed by the data below the CMC (**Figure S.1a**). Moreover, with finely grained concentration steps, the CMC is often well-defined by the position of the discontinuity (e.g. 7 mM in **Figure S.1bc**), such that the CMC may also be constrained to a narrow range. For these reasons, the PS model can reduce to a pseudo-single-parameter fit of just o_{micelle}, since it may not be feasible to obtain data for a range sufficient to approach o_{micelle} asymptotically (**Figure S.1d**).

Computational fitting of the PS model to experimental data is possible in principle but it is necessary to deal first with the piecewise nature of the model, where one problem is that computational fitting tends to overemphasize agreement of the model with the second piece of Eqn 1 (i.e., the long tail for concentrations above the CMC) and de-emphasize adherence to the CMC. A rater could constrain the CMC to augment a computational approach, but the fit would still drive the CMC to the edge of the constrained range requiring ever narrower CMC ranges specified by the user. Such a process would then be difficult to distinguish from supervised (manual) modeling. Additionally, the application of the PS model to real data is nontrivial since real data will exhibit varying degrees of equilibrium (mass action) behavior, and an experienced rater can ensure the PS model is applied correctly to the first CMC in particular by matching the inflection point of the PS model to the data as closely as possible (**Figure 2** of the

main text). Finally, it is usually not possible to sample the region after the CMC completely, meaning that the situation of saturation shown in **Figure S.1e** is unusual. The curvature can be truncated by a sequential CMC that follows shortly (e.g. consider **Figure 4** in the main text and others as well), again requiring manual intervention.

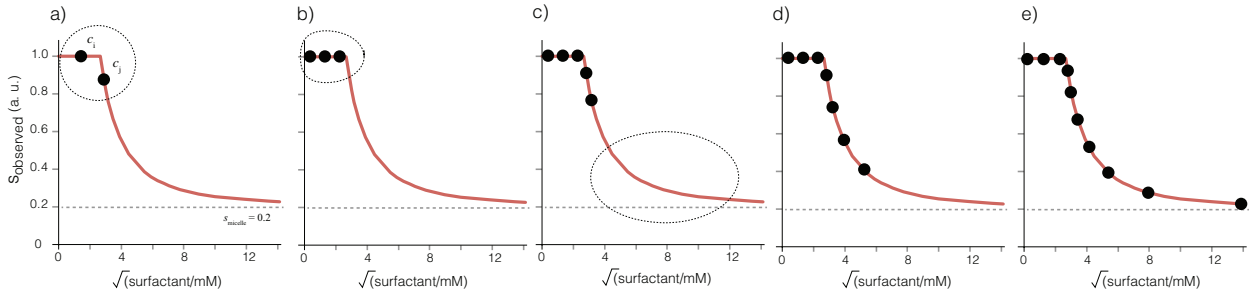

Figure S.1 Sampling considerations for use of the phase transition model (red line) are illustrated. See Section S.1 for a further discussion of the panels, where panels (d) and (e) show sampling that is amenable to accurate modeling, recognizing particularly in (d) that it may be difficult to sample the region after the CMC as fully as desired.

In the mass-action (MA) model, the aggregation is described as an equilibrium, where the aggregation number (n) and sometimes the counterion occupancy (β) are included in the model. [1] The equilibrium of interest for a nonionic surfactant S is

$$nS \leftrightarrow M \quad , \tag{S.2}$$

where S denotes the surfactant monomer, M the micelle, and n the aggregation number. The equilibrium constant and the mass action statement are, respectively,

$$K = \frac{[M]}{[S]^n} \ , \text{ and} \tag{S.3}$$

$$[S_{\text{tot}}] = [S] + n[M] \ , \tag{S.4}$$

where S_{tot} denotes the sum of concentrations of surfactant molecules present as free monomers and in micelles. Combining Eqns S.3 and S.4 yields

$$[S_{\text{tot}}] = [S] + nK[S]^n = [S](1 + nK[S]^{n-1}) \ . \tag{S.5}$$

It is common to recast Eqn 6 by considering the concentration $[S^*_{tot}]$ where half the surfactant is present in micelles. The total bile salt still follows Eqn S.4,

$$[S^*_{tot}] = [S^*] + n[M^*] \ , \tag{S.6}$$

where [S*] and [M*] are the monomer and micelle concentrations for the condition in which half the molecules are in micelles. Then, if half of the surfactant molecules are bound, then $[S^*] = n[M^*]$ and $S^*_{tot} = 2[S^*]$. Now K can be rewritten as

$$K = \frac{[M^*]}{[S^*]^n} = \frac{[S^*]/n}{[S^*]^n} \quad , \text{ and} \tag{S.7}$$

$$nK = \left(\frac{1}{[S^*]}\right)^{n-1} = \left(\frac{2}{[S^*_{tot}]}\right)^{n-1} \quad . \tag{S.8}$$

Substituting Eqn 8 into Eqn S.5 yields

$$[S_{\mathrm{tot}}] = [S]\left(1 + \left(\frac{2[S]}{[S^*_{tot}]}\right)^{n-1}\right) . \tag{S.9}$$

Equation S.9 is also Eqn 3 of the main text. Treating n and S^*_{tot} as fittable parameters, Eqn 9 predicts [S] implicitly for a given [S_{tot}]. The concentration $[S^*_{tot}]$ at which half the surfactant is bound is sometimes referred to as the critical concentration of the mass action model, which should not be confused with the CMC. When n is large, CMC ~ $[S^*_{tot}]/2$. When n is small, a CMC may still be defined and obtained from mass action models.[1] For ionic surfactants in the presence of counterions, the mass action model incorporates an additional parameter β for the counterion occupancy of a micelle, and has been presented in the context of bile salt ITC measurements. [1]

S2: Global fitting for deoxycholate with R-BNDHP probe

Global fitting of other protons (in addition to Figures 4-5 of the main text) confirms the sequential CMC values. Single CMC models are unable to explain the data in **Figure 3**, yet the introduction of a second CMC at 26 mM accounts for the flattening of the chemical shift trends after 26 mM for H5 and H7. An important contrast is drawn with H4 of R-BNDHP, which is insensitive to the 26 mM CMC, but is actually more sensitive to the high order CMC at about 60 mM (58 mM in the model shown below). This contrast helps to show the necessity of the 26 mM CMC for treating H5-H7. Moreover, it can be inferred that the secondary 26 mM CMC remodels the hydrophobic pocket sampled by H5-H7 but leaves the micelle surface sampled by H4 relatively conserved.

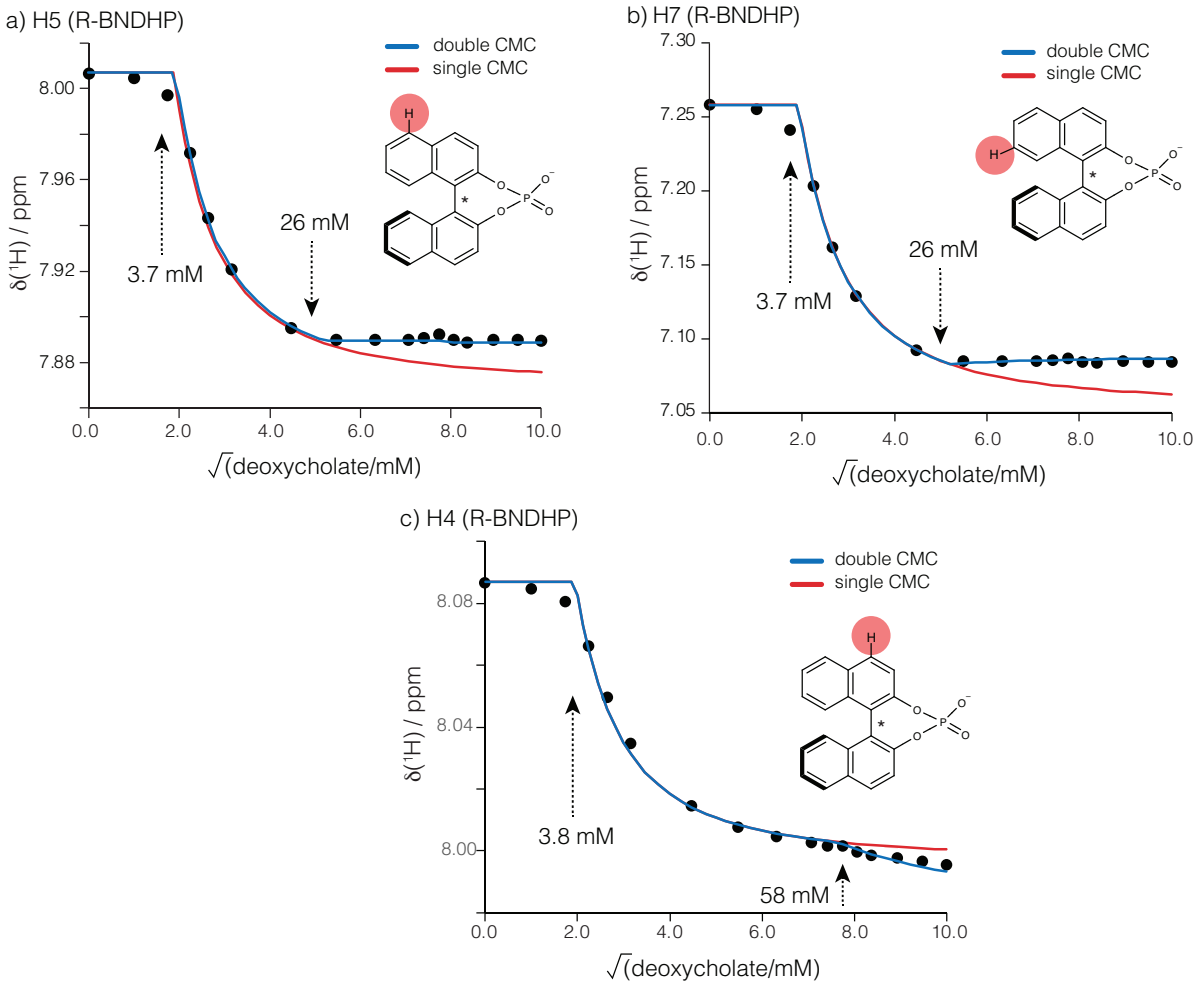

Figure S.2. (Updated to include H4) Global modeling builds confidence in the CMCs determined via chemical shift perturbations for different protons on the guest molecule with deoxycholate micelles. Here, the chemical shift perturbations of H5- and H7-R-BNDHP are shown to be well modeled with the same CMCs determined in **Figures 3-4** for H6- and H3-R-BNDHP. The H4 proton is not perturbed at 26 mM in contrast.

S3: The 9.5 mM deoxycholate CMC by double-CMC models

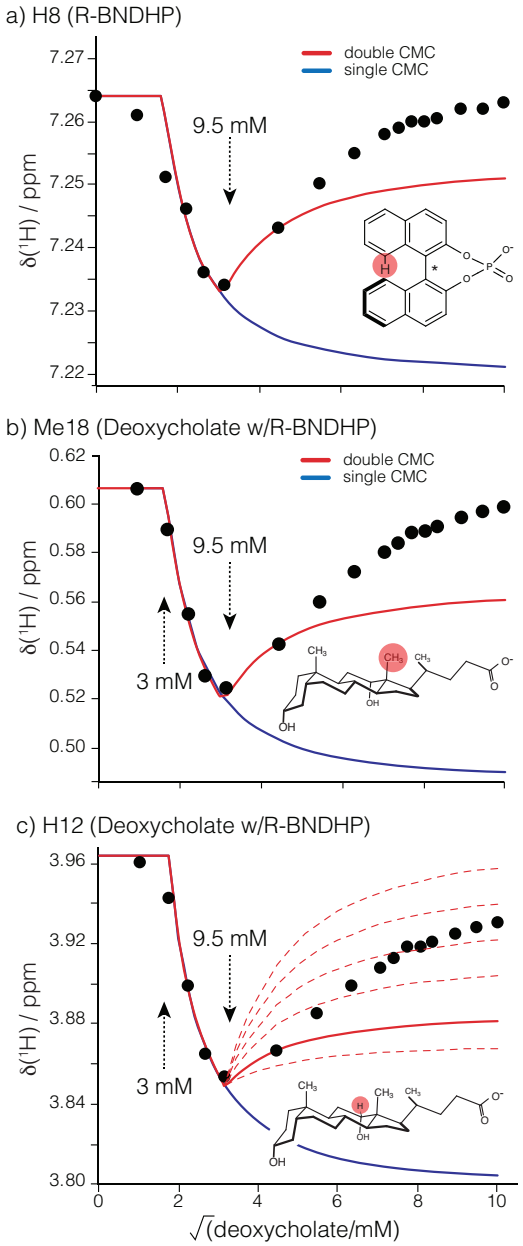

Figure S.3. The double-CMC model is applied to additional protons of the same data set, this time including Me19 of the bile salt to not only verify the CMC value(s) but also to verify that the aggregation witnessed by the bile salt corresponds to the events reported by the guest. In (a-c) the single and double CMC models are applied using 3 mM and 9.5 mM CMC values for DC; the small deviation of the first CMC from the 3.7 mM values utilized in **Figure 3** may owe in part to uncertainty in choosing PS parameters that approximate MA behavior. Part (c) explores the o_{mic2} parameter space to make it clear that no plausible choice of model can explain the trend in the data after 9.5 mM DC. The double-CMC model, while useful, is clearly inadequate.

S4: Confirming DC CMCs in an independent series

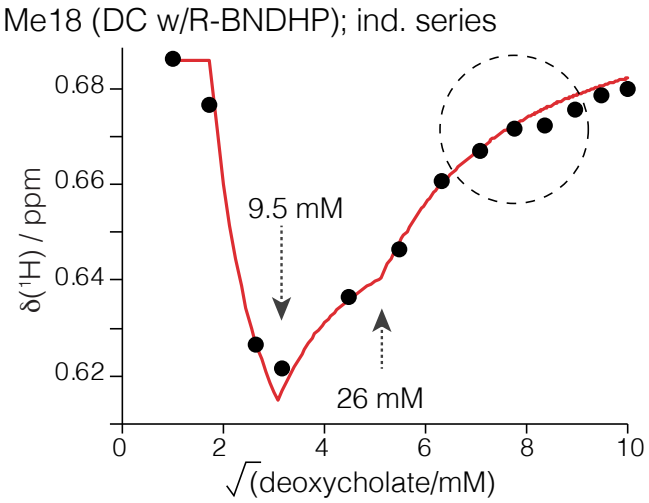

Figure S.4 A reanalysis of prior data with an independent concentration series of deoxycholate with R-BNDHP (2.5 mM) recorded at 900 MHz [2] confirms 3 mM, 9.5 mM, and 25 mM sequential CMCs with the multi-CMC PS model developed here. A small chemical shift offset from the series reported in the main text is noted, where this work does not use a chemical shift marker to avoid possible interactions with the bile acid micelles.

S5: The influence of the probe molecule on the CMC

Multi-CMC modeling can measure the influence of the probe molecule on the aggregation steps. Two independent concentration series of deoxycholate (pH 12, 298 K) were investigated with either 2.5 mM R-BNDHP or 2.5 mM S-BNDHP. Such concentrations are on a similar scale as the first two CMCs (about 3 mM and 9 mM DC) and could alter the bile aggregation. The S-BNDHP guest binds more strongly and could affect the bile micellization more strongly.

The H6-R-BNDHP chemical shift shows an initial 3.7 mM CMC, whereas a smaller 2.5 mM CMC is observed with 2.5 mM S-BNDHP. This change in the first CMC reflects the S-BNDHP guest molecule stabilizing the bile micelle and enabling aggregation at lower DC concentrations. Similarly a high order CMC is found at about 26 mM with R-BNDHP but 23 mM with S-BNDHP.

The R-BNDHP guest does not bind as strongly and is actually insensitive to the 9 mM CMC, requiring only a two-CMC fit, whereas the H6 chemical shift of the stronger binding S-BNDHP guest was sensitive enough to be treated by the first three CMCs.

The DC CMC values determined in the presence of 2.5 mM R-BNDHP are similar to the values obtained in probe-free measurements (**Table 2**), supporting that the weaker binding R-BNDHP probe is a less invasive reporter on DC aggregation, while relatively high concentrations of S-BNDHP can affect micellization.

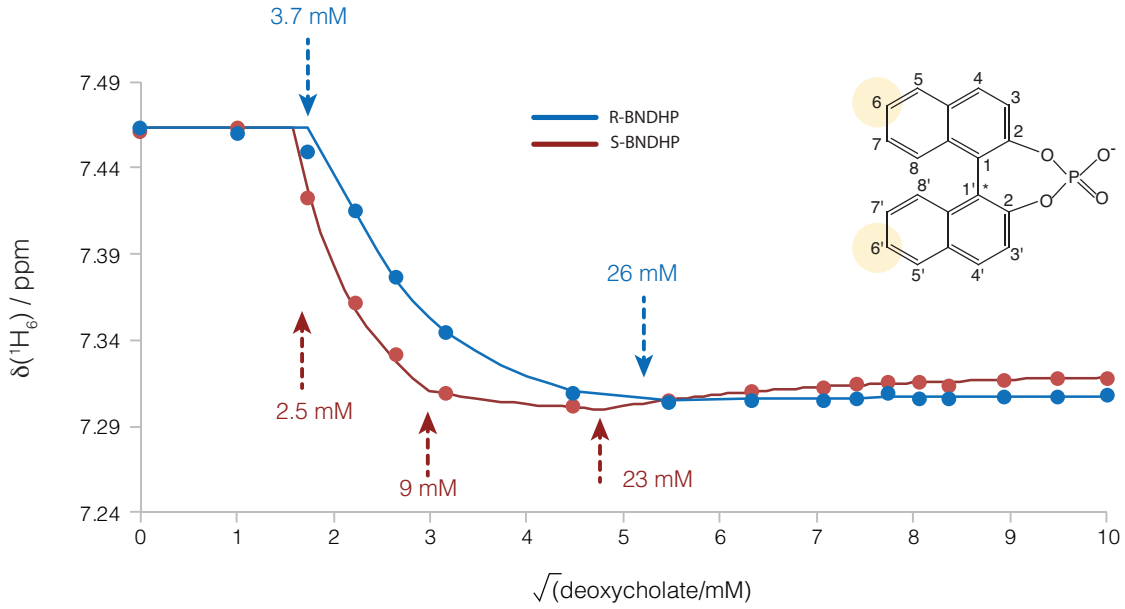

Figure S.5. The trends of δ(¹H6) of R- and S-BNDHP as a function of increasing deoxycholate concentration reflect different CMCs (pH 12). In each case the probe was present at 2.5 mM, a sufficient concentration to alter the micellization properties.

S6: Example Probe-Free Chemical Shift Cholate data with double-CMC models.

The preliminary CMC at about 7 mM cholate is not observed in the absence of the probe under these conditions (pH 12, 298 K). In considering additional fits, the consensus of the fits for the primary CMC is 16 mM, while a distribution of values for the secondary CMC place it at about 41 mM cholate (41 ± 4 mM).

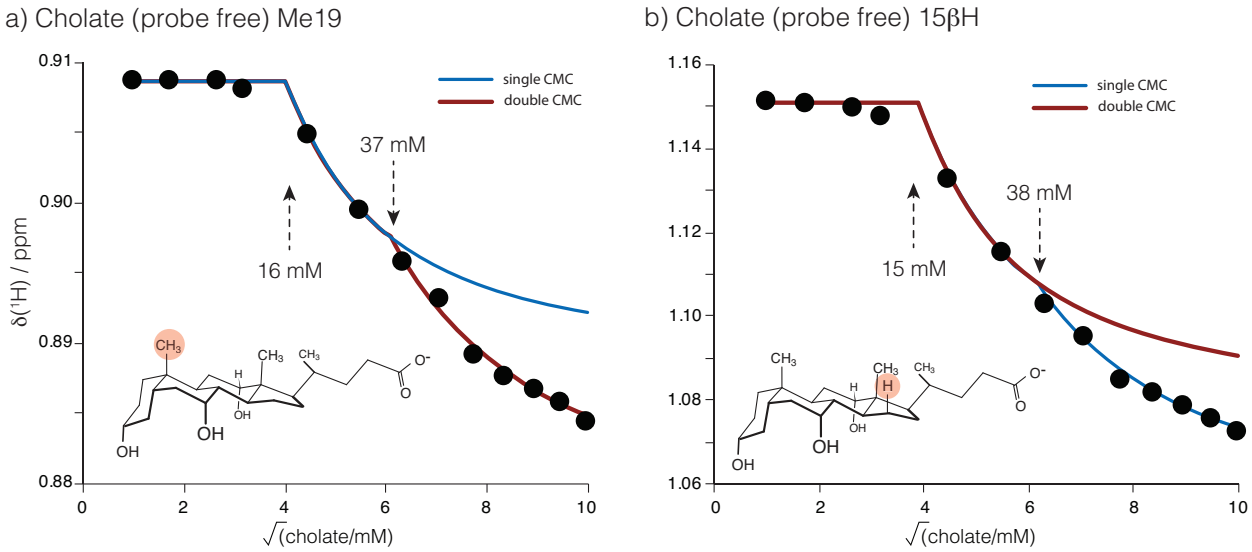

Figure S.6 Representative examples of modeling protons of basic cholate solutions that show a primary and a secondary micelle, as indicated on the figure.

S7: Representative glycocholate models

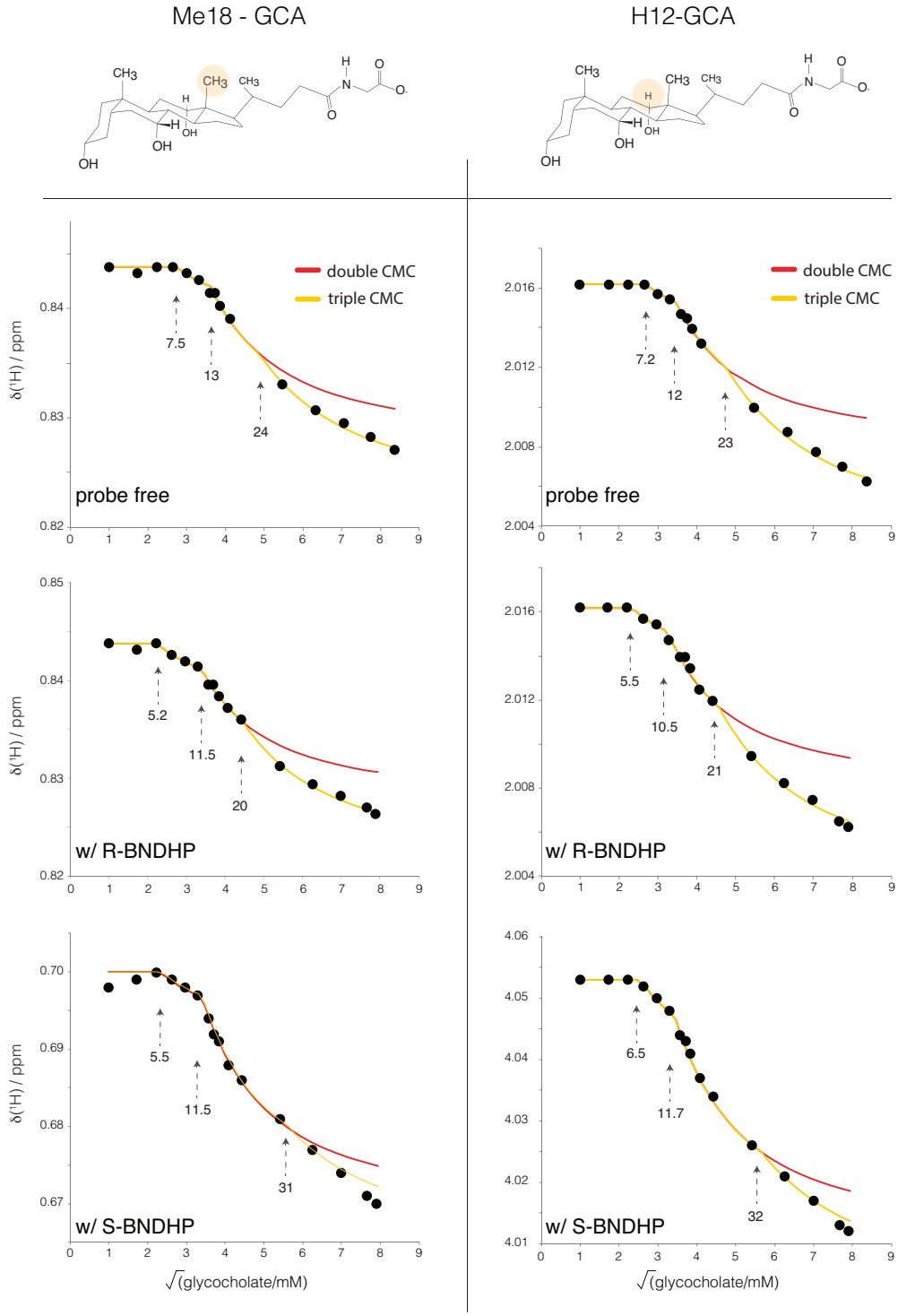

Figure S.7 Determining three sequential CMCs in basic (pH 12) solutions of glycocholate for the Me18 and H12 protons. The double-CMC model is included to illustrate the need for a third CMC.

S8: Global modeling of glycodeoxycholate (GDC) NMR titrations

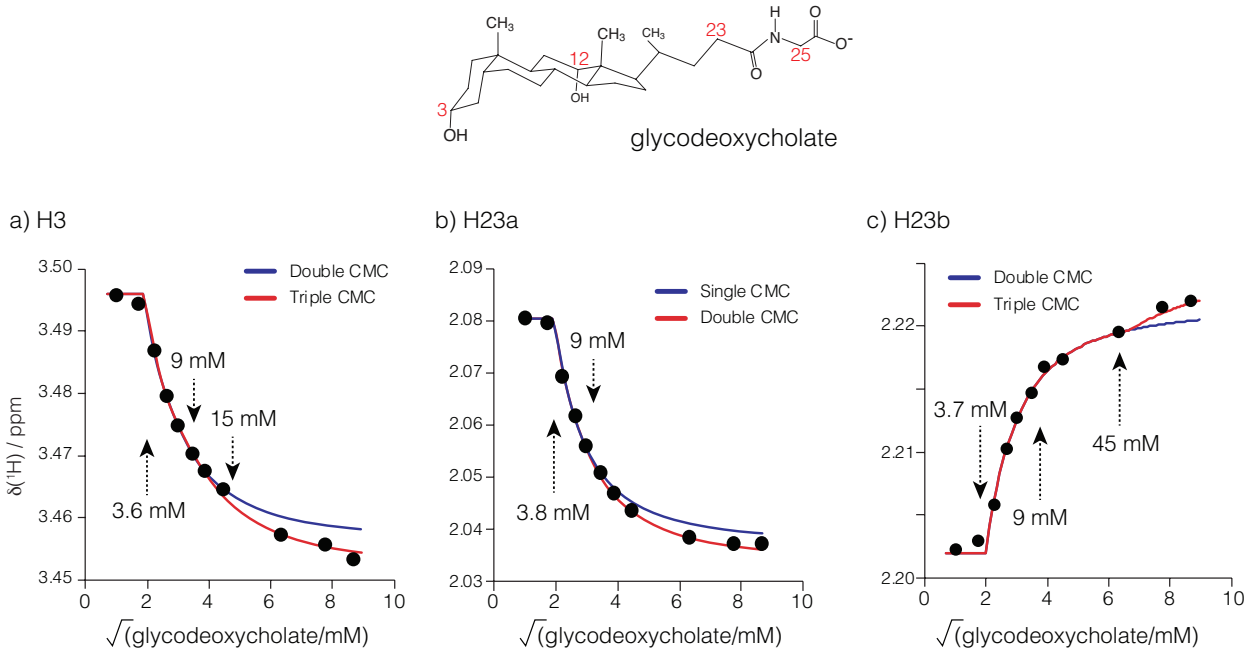

Figure S.8 Investigation of other protons in GDC builds consensus CMCs consistent with the strong trends seen in δ(¹H25) in the main text. Importantly, the simplest model should always be used since different protons do not all witness the same CMC events; for example, a double CMC model closely explains the behavior of one of the protons on C23 in panel (b) In contrast, examining the other proton in panel (c) shows more sensitivity to the higher order aggregate at about 45 mM GDC.

References

[1] N.E. Olesen, P. Westh, R. Holm, Determination of thermodynamic potentials and the aggregation number for micelles with the mass-action model by isothermal titration calorimetry: A case study on bile salts, J Colloid Interface Sci. 453 (2015) 79-89, https://doi.org/10.1016/j.jcis.2015.03.069.

[2] K.W. Eckenroad, G.A. Manley, J.B. Yehl, R.T. Pirnie, T.G. Strein, D. Rovnyak, An Edge Selection Mechanism for Chirally Selective Solubilization of Binaphthyl Atropisomeric Guests by Cholate and Deoxycholate Micelles, Chirality. 28 (2016) 525-533, https://doi.org/10.1002/chir.22609.