Thermodynamics and structure of inclusion compounds of tauro- and glyco-conjugated bile salts and β-cyclodextrin

René Holm[*a], Wei Shi[b], Rune A. Hartvig[§b], Sune Askjær [c], Jens Christian Madsen[d], and Peter Westh[b]

Supporting information

Introduction

This appendix is divided into two sections, the first providing a detailed description of the molecular modeling used in this work and the second additional ^{1}H-ROSEY NMR spectra's for the interested reader

1. Molecular modeling

This appendix describes the theoretical foundation behind the in-house molecular modeling software used for the statistical thermodynamics calculations described in the paper. According to the general naming from molecular modeling of ligand-receptor complexes, the small bile salt molecules are termed ligands and the cyclodextrin host molecule is termed the receptor throughout the following text.

The reversible formation of non-covalent bonding between a ligand *L* and a receptor *R* into a ligand-receptor complex *RL* can be described by the equilibrium $R + L \leftrightarrows RL$. The position of the equilibrium is determined by the binding energy also called the change in Gibbs free energy of the process, which determines the relative concentrations of the three different species *L*, *R* and *RL*.

With these species (neglecting entropy of the water molecules in the binding pocket) the binding process can energetically be described by the overall change in free energy, which at standard concentration equals:

$$\Delta G^{\circ}_{RL} = \mu^{\circ}_{RL} - \mu^{\circ}_{R} - \mu^{\circ}_{L} \qquad \text{Eq. 1}$$

where μ_X is the chemical potential of species *X* that in statistical thermodynamics is expressed by:

$$\mu_X^\circ = -RT\ln\left(\frac{1}{VC^\circ}Z_X\right) \qquad \text{Eq. 2}$$

Here V is the volume of the system, C^o is the standard concentration and Z is the overall partition function or configurational integral defined by:

$$Z = \int \exp\left(-\frac{U(r)+W(r)}{RT}\right)dr \qquad \text{Eq. 3}$$

The integral over r of the potential energy U and the solvation energy W relates to the full potential energy surface (PES) of the species and thereby cover translational and rotational phase space as well as all internal degrees of freedom of the species. Getting a correct estimate of Z requires both a correct representation of the PES via U and W and an adequate sampling/coverage of all relevant areas on the PES. By splitting the partition function Z into translational, rotational and internal degrees of freedom the chemical potential of the species becomes:

$$\mu_X^\circ = -RT\ln\left(\frac{1}{VC^\circ}\, z_{Trans}\cdot z_{Rot}\cdot z_{Internal}\right) \qquad \text{Eq. 4}$$

For a free species that does not interact with other molecules (neglecting solvent) there is no resistance to either rotation or translation. In addition the external degrees of freedom are independent of the internal degrees of freedom. This allows for analytical integration over translational (x, y, z)-space and Euler-angle rotational space (α, β, γ) yielding a factor of $8\pi^2 V$ for $z_{Trans}z_{Rot}$.[1] The volume terms cancel the chemical potential for a free species in equation 4, and are therefore given by:

$$\mu_{X\,free}^\circ = -RT\ln\left(\frac{8\pi^2}{C^\circ}\, z_{Internal}\right) \qquad \text{Eq. 5}$$

Calculation and integration of the potential energy U and the solvation energy W over the remaining 3N – 6 internal degrees of freedom where N is the number of atoms comprising the species X, are hence left to be determined. This holds true for both the

free ligand and the free receptor, but the theoretical considerations for the bound ligand is more complicated. In the receptor-ligand complex the ligand is restricted in its translational and rotational freedom by an external force effected by the receptor. Therefore the integral over the 6 external degrees of freedom for the ligand in the ligand-receptor complex must be calculated numerically as they are dependent on the total complex. Under these circumstances, the standard chemical potential for the receptor bound ligand becomes:

$$\mu^{\circ}_{RL-complex} = -RT\ln\left(\frac{1}{VC^{\circ}} \underset{Trans}{\overset{R}{z}} \underset{Rot}{\overset{R}{z}} \underset{Internals}{\overset{R}{z}} \underset{Trans}{\overset{L}{z}} \underset{Rot}{\overset{L}{z}} \underset{Internals}{\overset{L}{z}}\right) \qquad \text{Eq. 6}$$

where R and L denotes the partition functions of the receptor and ligand parts in the bound complex. Presuming that the external degrees of freedom for the receptor are not affected by the ligand, the integration over translation and rotation can for this part of the partition function be done analytically, again yielding a factor of $8\pi^2 V$, this leaves equation 6 to:

$$\mu^{\circ}_{RL-complex} = -RT\ln\left(\frac{8\pi^2}{C^{\circ}} \underset{Internals}{\overset{R}{z}} \underset{Trans}{\overset{L}{z}} \underset{Rot}{\overset{L}{z}} \underset{Internals}{\overset{L}{z}}\right) \qquad \text{Eq. 7}$$

Given these equations the overall change in free energy upon ligand binding then sums to Equation 8:

$$\Delta G^{\circ}_{RL} = \mu^{\circ}_{RL} - \mu^{\circ}_{R} - \mu^{\circ}_{L}$$

$$\Delta G^{\circ}_{RL} = -RT\ln\left(\frac{8\pi^2}{C^{\circ}} \underset{Internals}{\overset{R,complex}{z}} \underset{Trans}{\overset{L,complex}{z}} \underset{Rot}{\overset{L,complex}{z}} \underset{Internals}{\overset{L,complex}{z}}\right) + RT\ln\left(\frac{8\pi^2}{C^{\circ}} \underset{Internals}{\overset{R,free}{z}}\right) + RT\ln\left(\frac{8\pi^2}{C^{\circ}} \underset{Internals}{\overset{L,free}{z}}\right)$$

Provided a correct representation and sampling of PES, the above equation allows for a correct calculation of ΔG for the ligand binding under the approximation of the implicit treatment of the solvent.

Approximating the PES by Harmonic Functions

Using the harmonic oscillator - rigid rotor approximation (HORRA), motion around energy minima can be approximated by a harmonic energy profile centered at the bottom of the energy well. A harmonic function gives a good local approximation and has the advantage that all thermodynamic functions (e.g. entropy, free energy etc.) can be calculated analytically by classical mechanics of a harmonic oscillator.[1] The disadvantage is that the approximation is only valid locally around the base of the energy well and that it is expensive in time to calculate for very large systems. Furthermore, at higher temperatures the thermal energy of the system increases to a level from the base of the minimum, making the approximation less precise. For drug size molecules at relevant temperatures however, the harmonic approximation has been shown to provide a fair sampling of the soft degrees of freedom on the potential energy surface of the molecule.[2] Based upon this empiric knowledge, a formula for $\mu^\circ = -RT \ln Z$, where Z is the configurational integral (equation 3) and the combined energy provided by U and W (equal to $E(\boldsymbol{r})$), are the energy profile of the PES, which needs to be estimated, can be derived. Dependent on the coordinate basis, a Jacobian determinant can be necessary in order to account for the connection between the external and internal coordinates.[2,3] However, as only the torsions angles and not the hard bond lengths and angles are considered, the Jacobian determinant is independent of the internal coordinates[3] and it was be omitted in the calculations.

If derived from a Taylor series expansion, the approximated harmonic energy profile $E(\boldsymbol{r})$, of the potential at the base of the energy well, r_0 takes the form:

$$E(r) = E(r_0) + \left(\frac{\partial E}{\partial r}\right)_{r=r_0} (r - r_0) + \frac{1}{2}\left(\frac{\partial^2 E}{\partial r^2}\right)_{r=r_0} (r - r_0)^2 + \ldots$$

As the gradient, $\left(\frac{\partial E}{\partial r}\right)_{r=r_0}(r-r_0)$ at this point was equal to zero the PES harmonic approximation was limited to the second order, hence the equation becomes:

$$E(r) = E(r_0) + \frac{1}{2}\left(\frac{\partial^2 E}{\partial r^2}\right)_{r=r_0}(r-r_0)^2 \qquad \text{Eq. 9}$$

where $\left(\frac{\partial^2 E}{\partial r^2}\right)_{r=r_0}$ is the curvature of the PES at the bottom of the well, which can be viewed as a force constant of a harmonic oscillator anchored at the base of the well. The whole concept is illustrated in Fig. 1S.

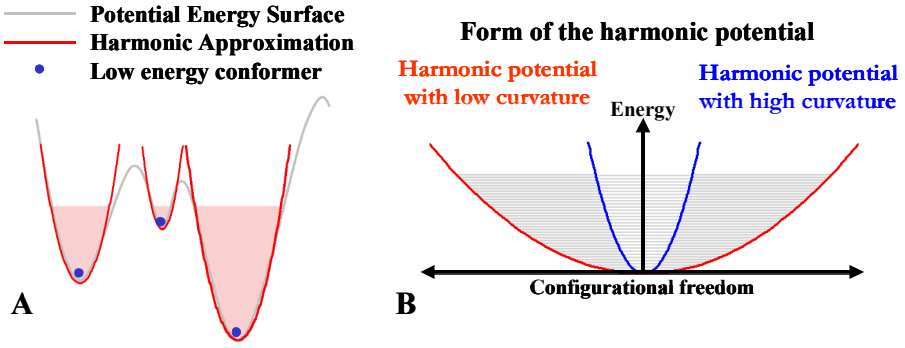

Fig. 1S*. A) The PES is approximated by harmonic potentials centered at the bottom of the energy wells. For each conformer harmonic potentials are fitted to approximate each degree of configurational freedom. B) Harmonic potentials with a high curvature are used for narrow energy wells where the thermal energy is more localized whereas softer curved harmonic potentials are used to model the broader energy wells where the energy is more dispersed.*

Combination of this PES approximation with equation 1, allows the formal requirements to calculate the chemical potential of the bound ligand to be developed.

Approximating ΔG by the Harmonic Oscillator Formalism

In cases were the internal degrees of freedom of the receptor are held fixed, like in most docking applications, and when the desolvation energy of the receptor is omitted the change in free energy of ligand binding can then be approximated by:

$$\Delta G^{\circ}_{RL} = \mu^{\circ}_{L_{bound}} - \mu^{\circ}_{L_{free}} = -RT \ln\left(\underset{Trans}{z^{L,bound}} \; \underset{Rot}{z^{L,bound}} \; \underset{Internals}{z^{L,bound}} \right) + RT \ln\left(\frac{8\pi^2}{C^{\circ}} \underset{Internals}{z^{L,free}} \right) \quad \text{Equation 10}$$

Ignoring the internal degrees of freedom for the receptor is a crude approximation that in effect sets this terms to zero.[4] In the case of the macro cyclic cyclodextrin molecule the internal degrees of freedom are highly correlated and they therefore cannot be easily treated, hence an approximation was applied.

The different factors in the partition function between the free and the bound ligand complex dependent functions, as they all relate to the reference frame of the receptor. The calculation of $z_{Trans}z_{Rot}z_{Internal}$ for the bound ligand must, therefore include considerations of the cross correlations between translation, rotation and the internal degrees of freedom. In practice, even the partition function for PES of the ligand alone needs a range of additional approximations in order to make the calculation feasible. For an extensive review of this subject see the review paper by Gilson et al.[5], but the most central is a way to approximate the configurational integral Z for both the free and the bound ligand. As mentioned previously this can be done with the HORRA across an ensemble of conformers that represent the relevant low energy minima on the PES.

Approximating the Chemical Potential of the Bound Ligand

The energy of the system is approximated as the potential energy and its energy minima plus the harmonic potential describing the energy change as the system moves away from each local energy minimum by the use of HORRA. The chemical potential of a bound ligand in translation, rotation and torsion angle space r around a given local energy minimum *i* on the potential energy surface is therefore given by Equation 11:

$$\begin{aligned}\mu_i &= -RT\ln(Z_i) \\ &= -RT\ln\int\exp(-G/RT)d\mathrm{r} \\ &= -RT\ln\int\exp\left(-\left(E_i+\frac{1}{2}\sum_j^N\sum_k^N E''_{jk}\Delta r_j\Delta r_k\right)\Big/RT\right)d\mathrm{r} \\ &= -RT\ln\int\exp\left(-\left(E_i+\frac{1}{2}\sum_{j=k}^N E''_{ij}\Delta r_j\Delta r_k+\sum_{j\neq k}^N E''_{jk}\Delta r_j\Delta r_k\right)\Big/RT\right)d\mathrm{r}\end{aligned} \qquad \text{(eq. 11)}$$

where *j* and *k* runs over all degrees of freedom considered.

In matrix notation $\frac{1}{2}\sum_j^N\sum_k^N E''_{jk}\Delta r_j\Delta r_k = \frac{1}{2}\Delta\mathbf{r}\mathrm{H}(\mathbf{r})\Delta\mathbf{r}$, where **H(r)** is the second derivative or Hessian matrix over all degrees of freedom; translational, rotational and torsion angles. The Hessian is a symmetrical N_{tors} + 6 by N_{tors} + 6 matrix where N_{tors} is the number of relevant torsional degrees of freedom. The Hessian matrix can be diagonalized into a set of N_{tors} + 6 eigenvalues v_l and a set of N_{tors} + 6 eigenvectors e_l.[2,3,5] The off-diagonal elements are thereby eliminated and the remaining diagonal elements have the form of a Gaussian function. The transformation from r into eigenvector space, r → *e* therefore gives:

$$\begin{aligned}\mu_i &\approx -RT\ln\int\exp\left(-\left(E_i+\frac{1}{2}\sum_l^{N_{tors}+6}v_l e_l^2\right)\Big/RT\right)de \\ &\approx E_i - RT\ln\prod_l^{N_{tors}+6}\int\exp\left(-\frac{1}{2}v_l e_l^2/RT\right)de \\ &\approx E_i - RT\ln\prod_l^{N_{tors}+6}\sqrt{\frac{2\pi RT}{v_l}} \\ &\approx E_i - \frac{RT}{2}\ln\left(\frac{(2\pi RT)^{N_{tors}+6}}{\prod_l^{N_{tors}+6}v_l}\right)\end{aligned}$$

Equation 12

The product of the N_{tors} + 6 eigenvalues equals the determinant of the Hessian matrix with respect to the original torsional space. Back transformation from eigenvector to Cartesian space: $e \rightarrow$ r gives us:

$$\mu_{i,L_{bound}} \approx E_i - \frac{RT}{2}\ln\left(\frac{(2\pi RT)^{N_{tors}+6}}{\det(\mathbf{H}(\mathbf{r_i}))}\right), \qquad \text{Equation 13}$$

which is the approximation to the chemical potential of the bound ligand conformation i integrated over rotation, translation and torsion angle space. With the translation in the x, y and z directions, rotation around the axis r_x, r_y, and r_z and variation of N torsion angles ϕ the Hessian matrix becomes:

$$\mathbf{H}_{\mathbf{TRT}} = \begin{bmatrix} \frac{\partial^2 E}{\partial x^2} & \frac{\partial^2 E}{\partial x \partial y} & \frac{\partial^2 E}{\partial x \partial z} & \frac{\partial^2 E}{\partial x \partial r_x} & \frac{\partial^2 E}{\partial x \partial r_y} & \frac{\partial^2 E}{\partial x \partial r_z} & \frac{\partial^2 E}{\partial x \partial \phi_1} & \frac{\partial^2 E}{\partial x \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial x \partial \phi_N} \\ \frac{\partial^2 E}{\partial x \partial y} & \frac{\partial^2 E}{\partial y^2} & \frac{\partial^2 E}{\partial y \partial z} & \frac{\partial^2 E}{\partial y \partial r_x} & \frac{\partial^2 E}{\partial y \partial r_y} & \frac{\partial^2 E}{\partial y \partial r_z} & \frac{\partial^2 E}{\partial y \partial \phi_1} & \frac{\partial^2 E}{\partial y \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial y \partial \phi_N} \\ \frac{\partial^2 E}{\partial x \partial z} & \frac{\partial^2 E}{\partial y \partial z} & \frac{\partial^2 E}{\partial z^2} & \frac{\partial^2 E}{\partial z \partial r_x} & \frac{\partial^2 E}{\partial z \partial r_y} & \frac{\partial^2 E}{\partial z \partial r_z} & \frac{\partial^2 E}{\partial z \partial \phi_1} & \frac{\partial^2 E}{\partial z \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial z \partial \phi_N} \\ \frac{\partial^2 E}{\partial x \partial r_x} & \frac{\partial^2 E}{\partial y \partial r_x} & \frac{\partial^2 E}{\partial z \partial r_x} & \frac{\partial^2 E}{\partial r_x^2} & \frac{\partial^2 E}{\partial r_x \partial r_y} & \frac{\partial^2 E}{\partial r_x \partial r_z} & \frac{\partial^2 E}{\partial r_x \partial \phi_1} & \frac{\partial^2 E}{\partial r_x \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial r_x \partial \phi_N} \\ \frac{\partial^2 E}{\partial x \partial r_y} & \frac{\partial^2 E}{\partial y \partial r_y} & \frac{\partial^2 E}{\partial z \partial r_y} & \frac{\partial^2 E}{\partial r_x \partial r_y} & \frac{\partial^2 E}{\partial r_y^2} & \frac{\partial^2 E}{\partial r_y \partial r_z} & \frac{\partial^2 E}{\partial r_y \partial \phi_1} & \frac{\partial^2 E}{\partial r_y \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial r_y \partial \phi_N} \\ \frac{\partial^2 E}{\partial x \partial r_z} & \frac{\partial^2 E}{\partial y \partial r_z} & \frac{\partial^2 E}{\partial z \partial r_z} & \frac{\partial^2 E}{\partial r_x \partial r_z} & \frac{\partial^2 E}{\partial r_y \partial r_z} & \frac{\partial^2 E}{\partial r_z^2} & \frac{\partial^2 E}{\partial r_z \partial \phi_1} & \frac{\partial^2 E}{\partial r_z \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial r_z \partial \phi_N} \\ \frac{\partial^2 E}{\partial x \partial \phi_1} & \frac{\partial^2 E}{\partial y \partial \phi_1} & \frac{\partial^2 E}{\partial z \partial \phi_1} & \frac{\partial^2 E}{\partial r_x \partial \phi_1} & \frac{\partial^2 E}{\partial r_y \partial \phi_1} & \frac{\partial^2 E}{\partial r_z \partial \phi_1} & \frac{\partial^2 E}{\partial \phi_1^2} & \frac{\partial^2 E}{\partial \phi_1 \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial \phi_1 \partial \phi_N} \\ \frac{\partial^2 E}{\partial x \partial \phi_2} & \frac{\partial^2 E}{\partial y \partial \phi_2} & \frac{\partial^2 E}{\partial z \partial \phi_2} & \frac{\partial^2 E}{\partial r_x \partial \phi_2} & \frac{\partial^2 E}{\partial r_y \partial \phi_2} & \frac{\partial^2 E}{\partial r_z \partial \phi_2} & \frac{\partial^2 E}{\partial \phi_1 \partial \phi_2} & \frac{\partial^2 E}{\partial \phi_2^2} & \cdots & \frac{\partial^2 E}{\partial \phi_2 \partial \phi_N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial x \partial \phi_N} & \frac{\partial^2 E}{\partial y \partial \phi_N} & \frac{\partial^2 E}{\partial z \partial \phi_N} & \frac{\partial^2 E}{\partial r_x \partial \phi_N} & \frac{\partial^2 E}{\partial r_y \partial \phi_N} & \frac{\partial^2 E}{\partial r_z \partial \phi_N} & \frac{\partial^2 E}{\partial \phi_N \partial \phi_1} & \frac{\partial^2 E}{\partial \phi_N \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial \phi_N^2} \end{bmatrix}$$

This approximation is, however only valid if the Hessian matrix is positive, defined as positive eigenvalues for all the structures representing the true minimum. Effectively, this means that the energy gradient must be very close to zero so all force field energy minimizations should be performed to a gradient below 10^{-3} kcal/mol/Å.

With the chemical potential defined as shown above, Boltzmann weighting of all conformations of the ligand bound to the complex can be approximated with respect to the overall chemical potential of the bound ligand.

Approximating the Chemical Potential of the Free Ligand

The derivation of the chemical potential for the free ligand, is similar to the approximation of the bound ligand discussed above, though with the exception that the external degrees of freedom are treated analytically. The chemical potential for the free ligand therefore becomes:

$$\mu_i^{\circ}{}_{L_{free}} \approx E_i - RT \ln Z' = E_i - RT \ln \left(\frac{8\pi^2}{C^{\circ}} \int \exp \left(- \frac{E_i + \frac{1}{2} (\phi - \phi_i)^T \, \mathrm{H}_{Torsion} (\phi - \phi_i)}{RT} \right) d(\phi - \phi_i) \right)$$

$$\approx E_i - RT \ln \left(\frac{8\pi^2}{C^{\circ}} \right) - \frac{RT}{2} \ln \left(\frac{(2\pi RT)^{N_{Torsion}}}{\det(H_{Torsion})} \right)$$

where $N_{Torsion}$ is the number of degrees of freedom (i.e. relevant torsion angles) and $H_{Torsion}$ is the Hessian matrix for the torsion angles only. The Hessian matrix has the form:

$$H_{Torsions} = \begin{bmatrix} \frac{\partial^2 E}{\partial \phi_1^{\,2}} & \frac{\partial^2 E}{\partial \phi_1 \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial \phi_1 \partial \phi_n} \\ \frac{\partial^2 E}{\partial \phi_2 \partial \phi_2} & \frac{\partial^2 E}{\partial \phi_2^{\,2}} & \cdots & \frac{\partial^2 E}{\partial \phi_2 \partial \phi_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial \phi_n \partial \phi_1} & \frac{\partial^2 E}{\partial \phi_n \partial \phi_2} & \cdots & \frac{\partial^2 E}{\partial \phi_n^{\,2}} \end{bmatrix}$$

The joint chemical potential over all conformers representing PES can easily be calculated with a Boltzmann weighting of the chemical potential of the individual conformers. Based on these approximations to the chemical potentials of the ligand in its bound and free form, $\Delta G_{Lig\text{-}binding}$ for the formation of the complex can be calculated as the difference between the two chemical potentials, and the likely binding positions can be approximated.

2. ^{1}H-ROSEY NMR

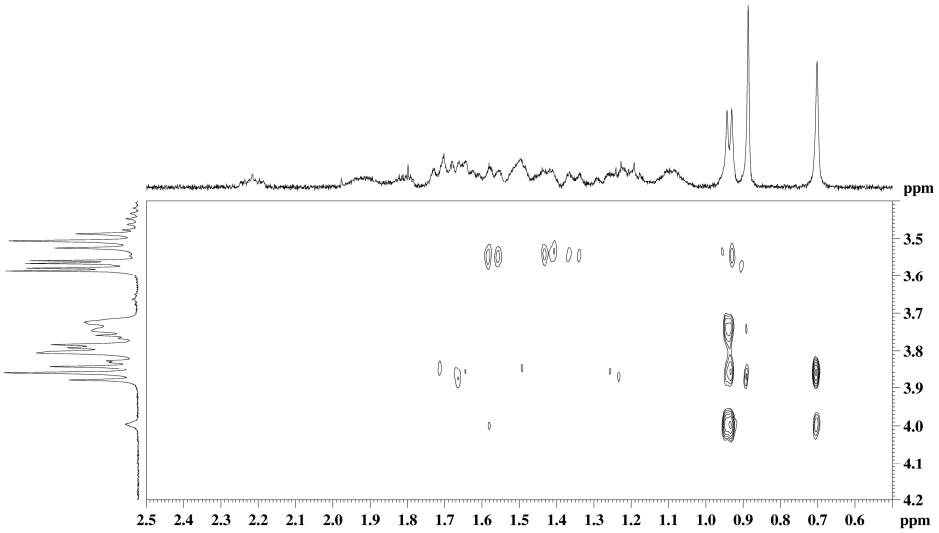

Fig. 2S. Partial ROESY spectrum of TDC and βCyD. The region of mostly bile salt chemical shift is depicted at the x-axis and the region of CD chemical shift at the y-axis (intermolecular region).

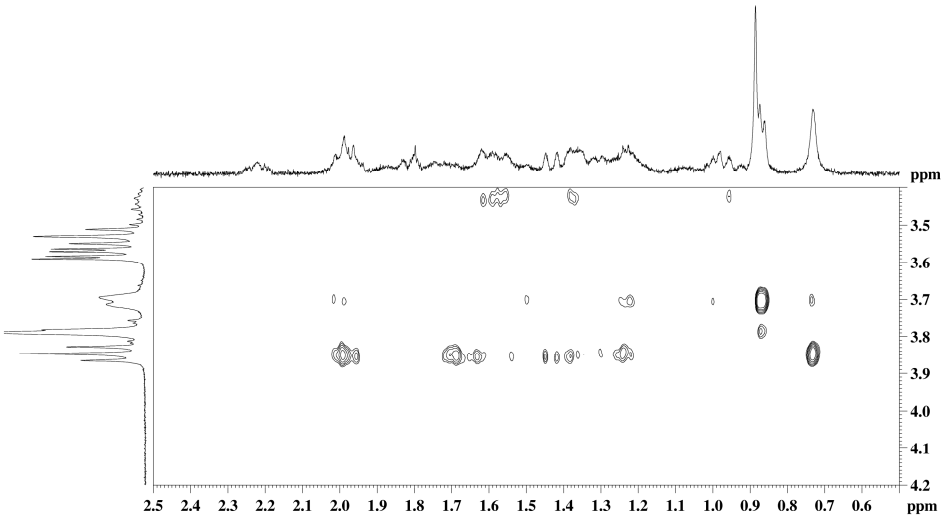

Fig.3S. Partial ROESY spectrum of TCDC and βCyD. The region of mostly bile salt chemical shift is depicted at the x-axis and the region of CD chemical shift at the y-axis (intermolecular region).

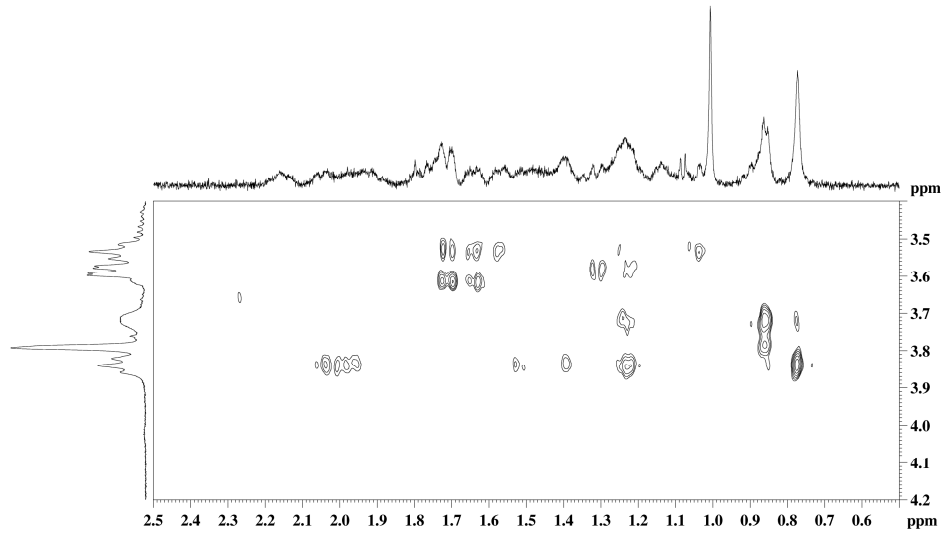

Fig.4S. Partial ROESY spectrum of TβMC and βCyD. The region of mostly bile salt chemical shift is depicted at the x-axis and the region of CD chemical shift at the y-axis (intermolecular region).

References

1 P. W. Atkins, *Physical Chemistry*. Oxford University Press, Oxford, 1994.

2 C. E. Chang, W. Chen and M. K. Gilson, *J. Chem. Theo. Comput.*, 2005, **1**, 1017-1028.

3 C. E. Chang, M. J. Potter and M. K. Gilson, *J. Phys. Chem. B*, 2003, **107**, 1048.1055

4 P. Cozzini, M. Fornabaio, A. Marabotti, D. J.Abraham, G. E. Kellogg and A. Mozzarelli, *Curr. Med. Chem.*, 2004, **11**, 3093-3118.

5 M. K. Gilson, J. A. Given, B. L. Bush and J. A. McCammon, *Biophys. J.*, 1997, **72**, 1047-1069.