Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering. This journal is © The Royal Society of Chemistry 2021

Supporting Info: Coarse-Grained Molecular Dynamics Study of the Self-Assembly of Polyphilic Bolaamphiphiles using SAFT- γ Mie Force Field

April 20, 2021

1 Interaction Parameters

1.1 SAFT- γ Mie Force Field

SAFT- γ Mie force field relies on a coarse-graining methodology where the nonbonded interaction parameters are calculated using a molecular based equation of state, SAFT- γ Mie EoS, and the bonded interactions are calculated using Boltzmann inversion from atomistic simulations. The method is summarised in Figure 1.

1.2 Non-Bonded Interactions

Knowing the molecules representing each moiety, and the corresponding CG representation, it is now possible to obtain non-bonded interaction parameters using the SAFT- γ Mie EoS. The equation of state defines the total Helmholtz free energy, A^{SAFT} , of a system of N particles at temperature T as a sum of four contributions[15]:



Figure 1: Schematic of the coarse-graining methodology used in this work.

$$\frac{\mathbf{A}^{SAFT}}{\mathbf{N}\mathbf{k}_B\mathbf{T}} = \frac{\mathbf{A}^{Ideal}}{\mathbf{N}\mathbf{k}_B\mathbf{T}} + \frac{\mathbf{A}^{Mono}}{\mathbf{N}\mathbf{k}_B\mathbf{T}} + \frac{\mathbf{A}^{Chain}}{\mathbf{N}\mathbf{k}_B\mathbf{T}} + \frac{\mathbf{A}^{Assoc}}{\mathbf{N}\mathbf{k}_B\mathbf{T}} \tag{1}$$

The first contribution is that of an ideal gas. The next perturbation, A^{Mono} , is a free energy contribution of all the CG segments having non-bonded interactions. The third contribution is due to the contribution of m monomers forming a chain and the final contribution is due to short-range directional forces, such as hydrogen bonding, by assuming associating "sticky" sites on the monomers. The full mathematical description of the EoS could be found in literature [15]. In this work, only the first three contributions are used to model polymers, as the non-continuous functions used to model association sites cannot currently be implemented in molecular dynamics simulations.

In the flavour of SAFT presented in this study, molecules are represented as chains of tangentially bonded spherical monomers (pearl necklace), that interact with a Mie potential[11] of two segments, i and j, at a distant r:

$$u_{ij}^{Mie}(r) = \mathcal{C}\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r}\right)^{\lambda_{r,\ ij}} - \left(\frac{\sigma_{ij}}{r}\right)^6 \right]$$
(2)

where ε_{ij} is the depth of the potential well, σ_{ij} is the average distance between the centres of mass of the two segments at contact, and $\lambda_{r,ij}$ is the repulsive exponent, controlling the softness and the range of the interaction between two segments. C is a pre-factor:

$$C = \left(\frac{\lambda_{r,ij}}{\lambda_{r,ij} - 6}\right) \left(\frac{\lambda_{r,ij}}{6}\right)^{6/(\lambda_{r,ij} - 6)} \tag{3}$$

Having defined the equation of state, a key question is how this is used as a method to find CG parameters. Parameters of the intermolecular potential defined in Equation 2 are used directly as input parameters of the EoS, i.e. $A^{SAFT} = f(\boldsymbol{m}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{\lambda}_r)$, where \boldsymbol{m} is a matrix describing the number of segments of each component i in each chain, and the other bold symbols being the matrix of parameters between species i and j.

This renders the EoS a useful parameter estimation tool to obtain "topdown" interaction parameters using experimental macroscopic bulk properties of fluids. The EoS could be used to find optimal parameters, $\alpha = [\boldsymbol{m}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{\lambda}_r]$, that minimise the error between experimental data and the calculations of the EoS. In this study, saturated liquid densities and vapour pressures were chosen as target experimental data in parameter estimation. This is in accordance with previous studies [10, 19, 12, 1, 5], where it has been shown that these two properties give parameters that could be used to predict a wide range of other bulk macroscopic properties. The objective function for the minimisation is thus:

$$\min_{\alpha} \left[F_{obj}(\alpha) = w_P \sum_{i=1}^{N_P} \left(\frac{\Delta P_v(T_i, \alpha)^2}{P_v^{Exp}(T_i)} \right) + w_\rho \sum_{i=1}^{N_\rho} \left(\frac{\Delta \rho_l(T_i, \alpha)^2}{\rho_l^{Exp}(T_i)} \right) \right]$$
(4)

where N_P and N_{ρ} are the number of data points used for pressure and saturated liquid density respectively and Δ denotes the difference between experimental data and the calculations of the EoS . w_P and w_{ρ} are weights given to each type of experimental data. Experimental saturation properties from literature were used in the python implementation of the gSAFT package [7] to find optimal parameters, and multiple (200) initial guesses sampled from a Sobol' sequence [22] were used to find optimal solutions. A non-linear least squares python package, Lmfit [13] was used in the optimisation procedure.

For multicomponent mixtures, cross interactions between different species could be calculated using Lorentz-Berthelot combining rules, using only the pure component potential parameters[4]. In this study, σ_{ij} and $\lambda_{r,ij}$ were calculated using this approach, without any adjustable parameters:

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \tag{5}$$

$$\lambda_{r, ij} = 3 + \sqrt{(\lambda_{r, ii} - 3)(\lambda_{r, jj} - 3))}$$
(6)

For ε_{ij} , the cross interaction parameter could slightly deviate from the geometric mean of the energies of the pure components, with a state-independent adjustable parameter, k_{ij} :

$$\varepsilon_{ij} = (1 - k_{ij}) \frac{(\sigma_{ii}^3 \sigma_{jj}^3)^{1/2}}{\sigma_{ij}^3} \sqrt{\varepsilon_{ii} \varepsilon_{jj}}$$
(7)

When $k_{ij} = 1$ there is no interaction between the segments and a non-zero k_{ij} indicates non-ideal cross species interactions. This is particularly important in the simulation of interactions between glycerol end groups and other species, given the glycerol self-interactions are dominated by hydrogen bonding, whereas cross interactions with the other species does not involve short range hydrogen bonding, thus the cross interactions is not directly a function of the self-interactions.

A particular challenge in fitting cross-interaction parameters is a lack of experimental data. To obtain cross interaction parameters, two pathways were taken:

- 1. Finding experimental data for mixtures of components most similar to the molecules presented in Table 1 in the main article : For example instead of icosane using n-heptane, and instead of glycerol using ethanediol, where experimental solubility of p-terphenyl in those solvents are known, and using the k_{ij} parameters obtained there in for the moieties in the BA.
- 2. Using the current SAFT- γ library of united atom groups to build the molecules similar to molecules in Table 1 in the main article and generating pseudo-experimental data. Using the pseudo-experimental data, cross interaction parameters are then fitted to these data. The current SAFT- γ library has many small united atom groups, but these models cannot be used in simulation due to the presence of parameters that currently cannot be modelled in simulation. This methodology has not been implemented before and is a novel way of generating experimentally rare data. For the set of united atom models and their parameters one could refer to Hutacharoen et al.[5]

1.3 Bonded Interactions

In SAFT- γ Mie EoS, the derivations of the contributions to the free energy for the chain term assume a freely-jointed chain. This is not necessarily valid

at the levels of coarse-graining presented in this work. Thus the form of the intramolecular potential parameters used in this work is:

$$U_{Bonded} = \sum_{bonds} \frac{1}{2} k_b (b_i - b_{i,0})^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta_i - \theta_{i,0})^2$$
(8)

In the above equation, the bonding constraint between two particles is taken into effect by the *bonds* term, where b_0 is the equilibrium bond length and k_b is a harmonic bond constant. The angle θ formed between three bonded particles is constrained using the second term, with θ_0 being the equilibrium angle and k_{θ} being the bond constant. Both k_b and k_{θ} describe the stiffness of the potential. The lower their value, the more flexible the molecule.

Within the SAFT- γ framework, molecules are tangentially bonded spheres with intermolecular Mie potentials. They are tangentially bonded, so that the distance between the centre of mass of two tangentially bonded spheres is given by σ_{ij} from the theory. Thus in this work, $b_0 = \sigma$ and k_b , k_{θ} and θ_0 were fitted to results of atomistic simulations, as the effect of angles and chain stiffness cannot be ignored.

In this work, harmonic potentials were used to model angle and bond stretching, using the Boltzmann inversion technique. First probability distributions of bonded and angular distributions were calculated from underlying atomistic simulations. The probability distributions, P(x), were fitted by weighted Gaussian distributions [21, 19]:

$$P(x) = \frac{w(x)}{\sqrt{2\pi s^2}} e^{-(x-\mu)^2/2s^2}$$
(9)

Here x is the intramolecular variable for which the probability distribution is calculated. In this work it is either the distance (x = b) between two neighbouring CG segments or the angle $(x = \theta)$ between three neighbouring CG segments. The mean distance or angle between the segments is given as μ ($\mu = b_0$ or θ_0) and s is the standard deviation of the Gaussian distribution. A weighting function w(x) is used to account for the asymmetry of the distributions. $w(r) = b^2$ and $w(\theta) = sin(\theta)$ were used as the weight functions for bond stretching and angular bending respectively. The harmonic potentials are then given as:

$$u_{Bond} = \frac{1}{2}k_b(b - b_0)^2 \approx k_B T \ln P(b)$$
(10)

$$u_{Angle} = \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 \approx k_B T \ln P(\theta)$$
(11)

Here, k_b and k_{θ} are measures of the stiffness of the harmonic potentials. In this work, given the constraints of the theory, $b_0 = \sigma$. However, k_b , k_{θ} and θ_0 were fitted using atomistic simulations, or if data were readily available from elsewhere.

For the flexible lateral chain, the k_b and k_θ parameters are the ones previously fitted using the Boltzmann inversion discussed for alkyl chains [19]. However, for the backbone, the value of k_θ was calculated using Boltzmann inversion from atomistic simulations using the OPLS-AA force field [6] as previously done for polymers. For glycerol, given that each segments represents one united atom carbon and one OH group, the angle potential is modelled with the same parameters as the angle potential of the three backbone carbons in the TraPPE UA model of 1,3-propanediol [24].

Table 1: Literature data used for obtaining cross interaction parameters of the three species using Method 1.

Mixture Needed	Mixture Found in Literature	Property	Ref
p-Terphenyl - Glycerol	p-Terphenyl - Ethanediol	Solubility (SLE)	[17]
p-Terphenyl - Icosane	p-Terphenyl - n-Heptane	Solubility (SLE)	[17]
Icosane - Glycerol	n-Heptane - Ethanediol	LLE	[23]

1.4 Cross Species Parameters: Using experimental data (Method 1)

As discussed in the methodology section, there is a lack of literature data for phase behaviours of mixtures containing p-terphenyl, icosane and glycerol. Thus two methods for obtaining cross interactions parameter were used. Table 2 is a summary of the literature data found for calculating the cross interaction parameters.

The first method is finding literature data of mixtures of molecules similar to those under investigation here. For the cross interactions between p-terphenyl glycerol, the closest experimental data found in literature was that of the solubility or solid liquid equilibrium (SLE) of p-terphenyl in ethanediol.

Solubility was calculated using following equation is used [18, 5]:

$$\ln(x_i(T,p)) = \frac{\Delta H_m^{fusion}}{R} \left(\frac{1}{T_m} - \frac{1}{T}\right) - \frac{\Delta c_p(T_m)}{R} \left[\ln\frac{T_m}{T} - \frac{T_m}{T} + 1\right] - \ln\gamma_i(T,p,\mathbf{x})$$
(12)

where x_i is the solubility of species i, ΔH_m^{fusion} is the molar enthalpy of fusion at melting point, T_m is the melting point and $\Delta c_p(T_m)$ is the difference in heat capacity of the solid and the liquid phases at the melting point. γ_i is the activity coefficient of species i in solution, which is calculated by the SAFT- γ Mie EoS and is therefore a function of the cross interaction parameter between the solvent and the solute i. Since the solid phase here is p-terphenyl, the values of the melting point properties for terphenyl are presented in Table 2.

To calculate activity coefficients using SAFT, the self-interaction parameters of p-terphenyl as given in Table 3 in the main article were used. For ethanediol, new parameters were fitted to NIST data, with the potential parameters being $N_{seg} = 3$, $\varepsilon = 503.67$ K, $\sigma = 3.04$ Å and $\lambda_{rep} = 22.26$. The cross interaction parameter was optimised to minimise the error for solubility of the solute. The optimal k_{ij} was found to be -2.8% which means a favourable interaction between the two species. The comparison between calculated and experimental solubilities is presented in Figure 2(a). In general there is good agreement between the fit and the experimental data. Given the strong self-interaction of ethanediol and glycerol due to hydrogen bonding, a favourable cross interaction is slightly surprising, however, without favourable cross-interaction, the solute was found to be insoluble in the ethanediol.

The same procedure was carried out for the cross interaction of p-terphenyl in n-heptane. For n-heptane, the same parameters of icosane was used, with two icosane segments representing an n-heptane. Here, the rationale was to ensure a transferability of the cross interaction parameter to use when modelling icosane



Figure 2: Comparison of the calculations of SAFT- γ Mie with experimental data for binary mixtures with components similar to the moieties of the bolaamphiphile studied here: **a**) Solubility of p-terphenyl in ethanediol, **b**) Solubility of p-terphenyl in n-heptane, **c**) T-x diagram of ethanediol-n-heptane mixtures (LLE) at P = 1 bar. The experimental data for **a** and **b** are from ref [17], and for **c** are taken from ref [23]. A k_{ij} of -0.028, 0.057 and 0.132 was used for the mixtures in **a**, **b** and **c** respectively.

Table 2: Melting point properties of p-terphenyl used in solubility calculations in glycerol and n-heptane, taken from DIPPR 801 [20].

	· · · · [·]
Property	Value
T_m	$485 \mathrm{K}$
ΔH_m^{fusion}	$33.70 \text{ kJ mol}^{-1}$
$\Delta c_p(T_m)$	$34.82 \text{ J} \text{ mol}^{-1} \text{ K}^{-1}$



Figure 3: Comparison of T-x diagrams at P = 1 bar for binary mixtures of Left) benzene and glycerol, Middle) glycerol and icosane and Right) benzene and icosane as calculated using the parameters estimated in this work (SAFT- γ Mie CG) and the united atom models of Hutacharoen et al. [5] (SAFT- γ Mie UA). For the SAFT- γ Mie CG models, k_{ij} of 0.126, 0.145, 0.078 was used for the mixtures in a, b and c respectively.

and p-terphenyl, and therefore the self-interaction parameters were not altered. Solubility of p-terphenyl in n-heptane is presented in Figure 2(b). This set of experimental data has very few data points and it is not possible to assess the robustness of the cross interaction parameter over the whole temperature range, however, it is possible to find a cross interaction parameter that can give solubilities in quantitative agreement with the 3 data points that exists in literature.

Finally for n-heptane and ethanediol, the same parameters used in the solubility calculations were used, with the cross interaction fitted to experimental data of T-x diagram of ethanediol-alkane mixtures at P = 1 bar. The results can be seen in Figure 2(c). The results are only in qualitative agreement, with the slopes of the calculations of the EoS and the experimental data being different. It could thus be concluded that away from the temperature ranges plotted, the EoS and experiments would most likely not agree. However, temperatures of 320 to 380 K is within the temperature range where liquid crystalline phases of BAs are observed, and so the cross interaction parameters is valid for temperature ranges of 300 to 400 K.

The optimal k_{ij} for the interaction of alkanes with ethanediol was found to be 0.138, which is much larger than the k_{ij} for the other binary mixtures. This shows that unlike previous studies of BAs where all moieties are strongly immiscible, with very high repulsions, here only the interaction of alkanes and polar groups deviate considerably from values obtained by combining rules.

1.5 Cross Species Parameters: Using SAFT- γ Mie UA to generate surrogate models (Method 2)

It is possible to build SAFT- γ Mie UA models of icosane and glycerol using these UA groups. However for p-terphenyl this is not possible since there are

Table 3: Molecules used in Method 2 to find cross interaction parameters of the three moieties studied in this work and their SAFT- γ Mie UA groups used to build these molecules to generate pseudo-experimental data of mixtures. The Mie interaction parameters for these groups could be found in ref [5].

Molecule	UA Groups (N^{repeat})
Bezene	aCH(6)
Glycerol	OH (3); CH2 (2); CH (1)
Icosane	CH3 (2); CH2 (18)

no parameters for the carbons connecting the phenyl rings to each other (aC group). It is possible to build a benzene model using six aromatic aCH groups.

For p-terphenyl it could be assumed that each CG segment of the 3-segment model represents a benzene molecule, and therefore benzene-glycerol and benzeneicosane binary mixtures can be appropriate mixtures to find cross interaction parameters between p-terphenyl, icosane and glycerol. In this method, therefore, the fluid phase properties (namely T-x diagrams at P = 1 bar) for mixtures of icosane-glycerol, as well as benzene-glycerol and benzene-icosane were first calculated using the UA groups specified in Table 4. Once pseudo-experimental data for each of the mixtures is generated, the same T-x diagrams are generated using the SAFT- γ CG parameters in Table ??. Again, the self-interaction parameters of benzene was assumed to be the same as one segment of p-terphenyl. The cross interaction parameters were then adjusted to minimise the error between the T-x diagrams generated using the two different approaches.

The results can be seen in Figure 3, with both sets of parameters generating very similar phase diagrams. For mixtures containing glycerol, the VLE and LLE regions merge with a VLLE line separating the two regions observed. It is known that for mixtures of bezene and glycerol, no azerotropes are observed experimentally at P = 1 bar [8]. For mixtures of heavier alkanes with alcohols, however, homogeneous azeotropes have been observed, such as mixtures of ethanol with n-heptane [14] and ethanediol with tetradecane [16]. This is predicted by the SAFT- γ UA models, for both binary mixtures of glycerol. It can be seen with the help of adjustable parameters that both approaches produce the same shape of phase envelope for all three mixtures. For the mixture of benzene-glycerol, the optimal value of k_{ij} was found to be 12.60%. This indicates a slightly unfavourable cross interaction between the species, which is different to the k_{ij} obtained in method 1, where the interactions were more favourable than predicted using combining rules. However, for the mixtures of benzene-icosane and glycerol-icosane, the k_{ij} values agree with those calculated in the previous section, being 7.80% and 14.5% respectively. This methodology indicates alcohol and alkane, as well as alcohol moieties interact extremely unfavourably with each other, whereas the cross interactions between the benzene groups and alkyl groups are only slightly different to the geometric mean of the self-interaction ε s.

Table 4: Values of k_{ij} obtained from both methods. The k_{ij} values obtained by method two were used in simulations of the liquid crystallines systems in this work. Parameters from method 2 were used in simulations.

cis nom memou z were used in simulations.				
Mixture	Method 1	Method 2		
p-Terphenyl - Glycerol	-0.028	0.126		
p-Terphenyl - Icosane	0.057	0.112		
Glycerol - Icosane	0.132	0.145		

1.6 Intramolecular Interactions

The first method was to fit the angle distribution to atomistic simulations using Boltzmann inversion. Atomistic simulations were run for a system containing 500 p-terphenyl molecules at P = 1 bar and T 500 K, with 500,000 steps for equilibration and 2,000,000 time steps ($\delta t = 1$ fs) for analysis. The centre of mass of each phenyl ring in the atomistic simulation was then taken as the centre of mass of each segment of the CG model and the angle distributions between the three CG segments were calculated. Using the probability distributions of the angles, θ and $k_{\theta, CG}$ were calculated using Boltzmann inversion. θ_{terph} was found to be 180° with the optimal k_{θ} being 225 kJ mol⁻¹ rad⁻². The results of the probability distributions can be seen in Figure 4(Left). Even though the angular distribution is correctly captured, it could be seen in Figure 4(Middle) that the p-terphenyl centre of mass-centre of mass (CoM-CoM) radial distribution function (rdf) of the CG segments with the aforementioned angle potential parameters is very different to the rdf of the same atomistic simulation used in the Boltzmann inversion. The peaks occcur at the same distances, but the magnitude of the peaks of the rdf is much higher for the CG model relative to the atomistic rdf. The logarithm of rdf is the potential of mean force between two p-terphenyl molecules, with a higher peak describing a stronger attraction between the molecules. This suggests that a more rigid CG model can capture the correct structure of the molecule, but at the same time attractions between two rigid molecules as modelled with SAFT are much stronger than the atomistic representation of p-terphenyl.

The second method therefore involved fitting k_{θ} to the CoM-CoM rdf of the atomsitic simulations, instead of the distribution of the angles of the three benzenes in the backbone. Using this approach, k_{θ} was found to be an order of magnitude smaller than previously calculated (225 kJ mol⁻¹ rad⁻²). The rdf of the CG model and the atomistic simulations remarkable agree (4(Middle)), however the distribution of the angle and therefore the structure of the molecule clearly deviates from the atomistic representations. It was decided to use intramolecular parameters obtained using the first method, as it accounted for the correct structure of the backbone, important in studying self-assembly.

In terms of bonded potentials, SAFT- γ Mie theory assumes a molecule to be composed of tangentially bonded spheres and therefore bond distances were not fitted to atomistic data, and σ s presented in Table ?? were used as bond distances for segments of the same moieties. For CG segments of different groups attached to each other, the arithmetic mean of the σ s was used as the bond distance. Assuming bond distances being the same size of σ is not strictly valid, however since the CG parameters are fitted to thermodynamic properties,



Figure 4: Comparison between the properties of course-grained models with atomistic simulations of p-terphenyl. **Left** Angle distributions of the three phenyl rings in p-terphenyl using two CG models that differ in the value of the angle potential only, with blue dash and dot line representing a model with $k_{\theta} = 225 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ and the red dash having $k_{\theta} = 25 \text{ kJ mol}^{-1} \text{ rad}^{-2}$. Both CG models have $\theta = 180^{\circ}$. **Middle** CoM-CoM RDF of p-terphenyl with the colors commensurate with the left figure. **Right** The bond distances between the two neighbouring phenyl rings in p-terphenyl as calculated by atomistic OPLS-AA simulations (black line) and the CG model (dashed blue lines).

instead of structural properties, it is expected that bond distances would be different to the underlying atomistic simulations. This could be seen in Figure 4(Right) where the bond distances between two neighbouring phenyl groups is around 10% higher in the CG models than in the atomistic simulations.

From previous works, it has been shown that the value of k_b does not significantly affect the phase behaviour of SAFT- γ Mie CG models, and therefore all values of k_b were set to 6130 kJ mol⁻¹ nm⁻² used in previous simulation studies [12, 19].

1.7 Comments on the Interaction Parameters

Pure component simulations were run using the inter and intramolecular parameters presented before. Each simulation involved a system containing 1000 molecules of each component, with an equilibration run of 200,000 steps and 500,000 steps of the main simulation run using an elongated box ($L_Z > 5 L_{X,Y}$). The cut-off for each simulation was 5 σ with no tail corrections.

As can be seen in Figure ??, for pure components the calculations of the simulation agree remarkably well with the results of the EoS, with less than 0.5% error between simulations and the calculations of the EoS.

Moreover, for the mixture of benzene (i.e. one p-terphenyl CG segment) and icosane, simulations were run with global composition of 50 mol% of icosane, with the calculations of the VLE presented in Figure 3(c). It could be seen that the parameters are robust for simulating both pure components and mixtures.

However, as previously mentioned, structural properties of CG models parametrised by top down approaches are not guaranteed to match the structural properties of real molecules. Moreover, coarse-graining in itself can affect the properties



Figure 5: Comparison of simulations of surface tension of **Top** p-Terphenyl and **Bottom** Glycerol with experimental data [20]. Developing non-associating models of molecules with hydrogen bonding using VLE data leads to CG models with very high surface tensions.

of the materials under study. For example, non-associating CG models are not realistic for hydrogen bonding molecules, as directional interactions cannot be averaged out into an isotropic spherical interaction [9]. A particular consequence of course-graining associating molecules as non-associating spheres can be seen in Figure 5, where the surface tension of glycerol is significantly overestimated. This is not observed for molecules that do not have hydrogen bonding such as p-terphenyl. Surface tension can have significant effects on the self-assembly structural properties of amphiphilic molecules, such as the curvature of the interfaces. An underlying problem is that the parameter estimation employed in this work is based on fitting the parameters to VLE data. Future work is currently being undertaken to improve parameter estimation of glycerol and other hydrogen bonding molecules.

Having highlighted these problems, the models presented in this work are still robust in simulating BAs as the parameters are thermodynamically consistent and are transferable over a wide range of state points. As will be shown in the following sections, these models can predict many properties of BAs. First, simulations of the bolaamphiphiles without any lateral side chains is presented. This is then followed by the simulation of a T-shaped bolaamphiphile (m = 6, n = 0) to assess the predictability of the model presented. This is followed by studying swallow-tails BAs.



Figure 6: 800 ns N σ T simulation of p-terphenyl at T = 390 K and σ = 1 bar starting with two different initial configurations: *Left* Random initial configuration. The nematic order parameter indicates no nematic order, *Right* Smectic layering initial configuration. The final configuration is not a smectic, but individual segments rearrange themselves to form FCC lattices, with the RDF of the segments provided.

2 Phase Behaviour of the Backbone without End Groups

Before the studies of swallow-tailed simulations can be carried out, it is imperative to investigate the phase behaviour of the rigid backbone and the backbone with glycerol end groups.

From previous studies, it is known that systems containing p-terphenyl exhibit nematic and smectic phase behaviour [2] at high temperatures and high densities. However, it is known from previous studies that nematic phases cannot be observed for molecules with chain lengths smaller than 5 segments [26]. For rigid molecules composed of less than 5 segments, the isotropic-nematic phase transition occurs at packing fractions higher than the isotropic-solid transition [25]. The OVL-LH theory of van Westen [25] predicts the isotropic-nematic coexistence packing fractions of hard trimers to be $\eta_{iso} = 0.541$ and $\eta_{nem} = 0.551$, which is higher than the isotropic-solid coexistence point calculated from molecular simulations of Vega et al [26], i.e. $\eta_{iso} = 0.430$ and $\eta_{iso} = 0.529$.

In this work, two sets of simulations were set up to study the phase behaviour of the rigid backbone. First set of simulations was set up by randomly inserting 500 molecules in the simulation box allowing the system to equilibrate at 390 K and 1 bar in an N σ T ensemble. The second set of simulations was run with a smectic initial configuration.

The results for the the first set of simulations can be seen in Figure 6. After 80 million time steps $(0.8 \ \mu s)$, there was no drift in the energy with an average nematic order parameter is -0.0176 (Figure 6(Left). The snapshot of the final configuration can be seen in the aforementioned figure, with every molecule highlighted with a different colour. It can be visually observed that no nematic phase could be observed. After 80 million timesteps, the system pressure was changed only in the Z direction, at pressures of up to 10,000 bar at different temperatures. No nematic phase could be observed, agreeing with previously studies.

To further assess the phase behaviour of the backbone, the second set of simulations were run, where the initial configuration was a smectic phase. As can be seen in Figure 6(Right), after 800 ns, the smectic layer is disrupted as the individual segments form an FCC lattice. The RDF of individual CG segments could be seen in the figure, clearly highlighting a solid FCC lattice.

It can be concluded therefore that the backbone model used in this work is not representable of the real p-terphenyl molecule in terms of liquid crystalline phase behaviour, as it is not possible to observe smectic layering or nematic liquid phases. This can be improved in the future by implementing better models, such as using a five bead model for p-terphenyl, or anisotropic models based on the Gay-Berne potential.



Figure 7: Snapshots of the equilibration of the BA molecule at 400 K and $\sigma = 1$ bar at different times. After 1 μ s a smectic phase is observed.



Figure 8: The P_2 order parameter vs temperature for the bolaamphiphile without any lateral side chains.

3 Phase Behaviour of the Backbone with End Groups

Althought it has been shown that using the three bead model is not appropriate in studying the isotropic nematic phase behaviour of pure terphenyl, it can still be used in modelling the phase behaviour of bolaamphiphiles, as the unfavourable cross interaction between the moieties can induce smectic layering. The phase behaviour of the pure bolaamphiphile, without any lateral side chains, was investigated with the potentials developed in the previous section. For the first set of simulations, 500 molecules were randomly inserted in a simulation box, with the system equilibrated at 400 K and $\sigma = 1$ bar. The simulation was run in the N σ T ensemble. As can be seen in Figure 7, a lamellar phase is observed after 1 μ s.

To investigate this more accurately, the temperature of the same simulation box was raised to 600 K and slowly cooled at a rate of 20 K per 3 μ s. As can be seen in Figure 8, there is an isotropic-smectic phases transition at around 570 K. This value is comparable to that experimentally measured, which for a biphenyl based bolaamphiphile with no lateral side chain is 260 °C or 563 K [3].



Figure 9: Distribution of the different branches of the side chain in the columns, with green highlighting the shorter chain and orange highlighting the longer chain.

4 Distribution of Different Swallow-tail Branches in the Self-Assembly

References

- Carlos Avendaño, Thomas Lafitte, Amparo Galindo, Claire S Adjiman, George Jackson, and Erich A Muüller. Saft-γ force field for the simulation of molecular fluids. 1. a single-site coarse grained model of carbon dioxide. *The Journal of Physical Chemistry B*, 115(38):11154–11169, 2011.
- [2] MA Bates and GR Luckhurst. Computer simulation of liquid crystal phases formed by gay-berne mesogens. In *Liquid Crystals I*, pages 65–137. Springer, 1999.
- [3] Xiaohong Cheng, Marko Prehm, Malay Kumar Das, Jens Kain, Ute Baumeister, Siegmar Diele, Dag Leine, Alfred Blume, and Carsten Tschierske. Calamitic bolaamphiphiles with (semi) perfluorinated lateral chains: polyphilic block molecules with new liquid crystalline phase structures. Journal of the American Chemical Society, 125(36):10977–10996, 2003.
- [4] Andrew J Haslam, Amparo Galindo, and George Jackson. Prediction of binary intermolecular potential parameters for use in modelling fluid mixtures. *Fluid Phase Equilibria*, 266(1-2):105–128, 2008.
- [5] Panatpong Hutacharoen, Simon Dufal, Vasileios Papaioannou, Ravi M Shanker, Claire S Adjiman, George Jackson, and Amparo Galindo. Predicting the solvation of organic compounds in aqueous environments: from alkanes and alcohols to pharmaceuticals. *Industrial & Engineering Chemistry Research*, 56(38):10856–10876, 2017.
- [6] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.
- [7] Thomas Lafitte, Vasileios Papaioannou, Simon Dufal, and Constantinos C Pantelides. gsaft: Advanced physical property prediction for process modelling. In *Computer Aided Chemical Engineering*, volume 40, pages 1003– 1008. Elsevier, 2017.
- [8] Maurice Marie Albert Lecat. *Tables azéotropiques: Azéotropes binaires orthobares*, volume 1. chez l'auteur, 1949.
- [9] Olga Lobanova, Carlos Avendaño, Thomas Lafitte, Erich A Müller, and George Jackson. Saft-γ force field for the simulation of molecular fluids: 4. a single-site coarse-grained model of water applicable over a wide temperature range. *Molecular Physics*, 113(9-10):1228–1249, 2015.
- [10] Olga Lobanova, Andrés Mejía, George Jackson, and Erich A Müller. Saftγ force field for the simulation of molecular fluids 6: Binary and ternary mixtures comprising water, carbon dioxide, and n-alkanes. *The Journal of Chemical Thermodynamics*, 93:320–336, 2016.
- [11] Gustav Mie. Zur kinetischen theorie der einatomigen körper. Annalen der Physik, 316(8):657–697, 1903.

- [12] Pedro Morgado, Olga Lobanova, Erich A Müller, George Jackson, Miguel Almeida, and Eduardo JM Filipe. Saft-γ force field for the simulation of molecular fluids: 8. hetero-segmented coarse-grained models of perfluoroalkylalkanes assessed with new vapour–liquid interfacial tension data. *Molecular Physics*, 114(18):2597–2614, 2016.
- [13] Matthew Newville, Till Stensitzki, Daniel B Allen, Michal Rawlik, Antonino Ingargiola, and Andrew Nelson. Lmfit: Non-linear least-square minimization and curve-fitting for python. ascl, pages ascl-1606, 2016.
- [14] Dong NguyenHuynh. Correlation and prediction of liquid–liquid equilibria for alcohol/hydrocarbon mixtures using pc-saft equation of state at high pressure up to 150 mpa. *Fluid Phase Equilibria*, 425:206–214, 2016.
- [15] Vasileios Papaioannou, Thomas Lafitte, Carlos Avendaño, Claire S Adjiman, George Jackson, Erich A Müller, and Amparo Galindo. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from mie segments. *The Journal of Chemical Physics*, 140(5):054107, 2014.
- [16] Jensen Ernest Petersen. Azeotropic distillation, October 1944. US Patent 2,360,685.
- [17] GJ Pierotti, CH Deal, and EL Derr. Activity coefficients and molecular structure. Industrial & Engineering Chemistry, 51(1):95–102, 1959.
- [18] John M Prausnitz, Rudiger N Lichtenthaler, and Edmundo Gomes de Azevedo. *Molecular thermodynamics of fluid-phase equilibria*. Pearson Education, 1998.
- [19] Sadia Rahman, Olga Lobanova, Guadalupe Jiménez-Serratos, Carlos Braga, Vasilios Raptis, Erich A Müller, George Jackson, Carlos Avendaño, and Amparo Galindo. Saft-γ force field for the simulation of molecular fluids. 5. hetero-group coarse-grained models of linear alkanes and the importance of intramolecular interactions. *The Journal of Physical Chemistry B*, 122(39):9161–9177, 2018.
- [20] RL Rowley, WV Wilding, JL Oscarson, Y Yang, NA Zundel, TE Daubert, and RP Danner. Dippr data compilation of pure chemical properties. *De*sign Institute for Physical Properties, 2006.
- [21] Wataru Shinoda, Russell DeVane, and Michael L Klein. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation*, 33(1-2):27–36, 2007.
- [22] Il'ya Meerovich Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802, 1967.
- [23] LAK Staveley and GL Milward. 874. solutions of alcohols in non-polar solvents. part iv. some thermodynamic properties of glycols in benzene, heptane, and cyclohexane. *Journal of the Chemical Society (Resumed)*, pages 4369–4375, 1957.

- [24] John M Stubbs, Jeffrey J Potoff, and J Ilja Siepmann. Transferable potentials for phase equilibria. 6. united-atom description for ethers, glycols, ketones, and aldehydes. *The Journal of Physical Chemistry B*, 108(45):17596– 17605, 2004.
- [25] T Van Westen. Development of an equation of state for nematic liquid crystals. Thesis: Development of an Equation of State for Nematic Liquid Crystals, 2015.
- [26] Carlos Vega, Carl McBride, and Luis G MacDowell. Liquid crystal phase formation for the linear tangent hard sphere model from monte carlo simulations. *The Journal of Chemical Physics*, 115(9):4203–4211, 2001.