Intramolecular bonding in a statistical associating fluid theory of ring aggregates

S. A. Febra, A. Aasen, C. S. Adjiman, G. Jackson and A. Galindo

Department of Chemical Engineering, Centre for Process Systems Engineering, Institute for Molecular Science and Engineering, Imperial College London, London, UK; Department of Energy and Process Engineering, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT

The first-order thermodynamic perturbation theory of Wertheim (TPT1) is extended to treat ring aggregates, formed by inter- and intramolecular association. The expression for the residual association contribution to the Helmholtz free energy for ring aggregates, incorporating the appropriate terms in Wertheim’s fundamental graph sum of the TPT1 density expansion, is derived to calculate the distribution of the molecular bonding states. This requires the introduction of two new parameters to characterise each possible ring type: the ring size $\tau$, which is equal to one in the case of intramolecular association, and a parameter $W$ that captures the likelihood of two ring-forming sites bonding. The resulting framework can be incorporated in equations of state that account for the residual association contribution to the free energy, such as the statistical associating fluid theory (SAFT) family, or the cubic plus association (CPA) equation of state. This extends the applicability of these equations of state to mixtures with an arbitrary number of association sites capable of hydrogen bonding to form intramolecular and intermolecular rings. The formalism is implemented within SAFT-VR Mie to calculate the fluid-phase equilibria of model chain-like molecules containing two associating sites A and B, allowing for the formation of open-chain aggregates and intramolecular bonds. The effect of adding a second component that competes for the association sites that mediate intramolecular association in the chain is also examined. Accounting for intramolecular bonding is shown to have a significant impact on the phase equilibria of such systems.

ARTICLE HISTORY

Received 12 July 2019
Accepted 12 September 2019

KEYWORDS

Association; ring formation; intramolecular association; hydrogen bonding; SAFT

1. Introduction

The hydrogen bond (HB) is defined [1] by the International Union of Pure and Applied Chemistry (IUPAC) as an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation.

Hydrogen bonds are characterised by their orientation, usually requiring the participating atoms to be collinear, and by their strength, with energies ranging from 8kJ mol$^{-1}$ in the case of weak hydrogen bonds involving $\pi$ orbitals to more than 40 kJ mol$^{-1}$ when charged acceptor groups are present [2]. The formation of hydrogen bonds leads to long-lived molecular aggregates that markedly influence the thermodynamic and phase behaviour of the system.

Most commonly, hydrogen bonds form between independent molecules (intermolecular HB), leading to the formation of linear or branched chain-like aggregates or networks that can extend in open form as well as include ring-like structures. In addition, hydrogen bonds may involve atoms in different parts of the same molecule (intramolecular HB), on occasion leading to bent X–H...X conformations in smaller molecules.
Hydrogen bonding can also be found within large macromolecules (polymers and proteins) with little constraint from the covalent bonds binding the atoms. The macroscopic manifestations of these two types of HB bond can be rather different and striking. Intermolecular HB are responsible for the high boiling temperatures and heats of vaporisation of small molecules such as water and ammonia, and the re-entrant vapour-phase behaviour reported in patchy colloids is also due to the formation of intermolecular ring-like aggregates [3]. Marked differences in solubility are seen between polar molecules in which intramolecular HBs form, where an enhancement in lipophilicity and cell-membrane permeability is observed [4], and the ‘open’ unbonded forms, which are more water-soluble [5,6]. Furthermore, the folding and aggregation of proteins are directly related to the formation of inter- and intramolecular HB (aggregation of mis-folded proteins is a common feature of diseases such as Alzheimer’s, Parkinson’s and type II diabetes [7]), as are the binding and specificity of ligands [8,9], the stabilisation of many polymer structures [10–12], and the changes reported in cloud-point pressures of polymer solutions [13,14].

Accounting for the effect of hydrogen bonding and aggregate formation has been of interest for some time in the development of theories for the fluid state. In 1908, Dolezalek [15] proposed a chemical approach to account for hydrogen bonding by considering that molecules aggregate into a new species through a chemical reaction, allowing for an explanation of many of the observed deviations from Raoult’s law seen in solutions of associating fluids. The main drawback of Dolezalek’s approach is that each of the association aggregates needs to be specified a priori. In the 1970s and 80s the advent of molecular-based association theories [16–25] provided a path towards understanding the distribution of association aggregates. Wertheim [26–29] proposed an approach to account for the residual contribution to the Helmholtz free energy arising from short-ranged and highly directional interactions in a fluid. This thermodynamic perturbation theory (TPT) became a landmark in the modelling of associating fluids. A number of constraints were originally imposed in the model to mimic common steric effects, such as one bond being restricted to involve only two molecules, and double-bond formation between two molecules being prohibited. Additionally, approximations were applied in the development of the theory neglecting cluster–cluster interactions, bond cooperation effects, and ring-like aggregate formation. Applying these constraints and approximations results in simple expressions for the distribution of molecular bonding states, directly obtained from the minimisation of the Helmholtz free energy.

At first order, Wertheim’s TPT (TPT1) requires knowledge of the pair (two-body) correlation function and the free energy of a reference non-bonded fluid. The approach has been shown to provide an excellent description of the thermodynamic properties of model associating fluids [30–34] and of self-assembling patchy particles [35–39]. Moreover, in the limit of complete association, covalent chain formation can be treated successfully [34,40], as shown early on by comparison with molecular simulation [41]. The need for a rigorous and reliable equation of state (EOS) to model the thermophysical properties of associating mixtures inspired the recasting of Wertheim’s TPT1 into the statistical associating fluid theory (SAFT) [42,43]. Following the original publications there has been a continuous effort to extend and improve equations of state that incorporate the TPT1 association term of Wertheim, giving rise to numerous versions of the SAFT approach, which have been demonstrated to deliver accurate thermodynamic properties for a large array of complex fluids and mixtures [44–51]. In most of these, the original assumptions of TPT1 are maintained, including the underlying approximation that limits the type of association aggregates considered to open clusters and neglects the formation of ring-like aggregates, whether through intermolecular or intramolecular hydrogen bonds.

The approximations of the TPT1 approach that lead to neglecting ring formation can be relaxed. Sear and Jackson [52,53] and Chapman and co-workers [54,55] have considered a pure fluid of fully flexible hard-chain molecules of $m$ spherical segments with two association sites (A and B) located in the terminal segments, incorporating inter- and intramolecular hydrogen bonding between A–B sites (with A–A and B–B site–site association prohibited). The authors noted that the correct free-energy term to account for intramolecular bonding must contain an intermolecular site–site correlation function for the ends of the chain. In the work of Sear and Jackson [52] this function is approximated as a product of the intermolecular correlation function with a parameter $W$ that accounts for the likelihood of the two terminal sites meeting, whereas in the work of Chapman and co-workers [54] the intermolecular site–site correlation function is retrieved from simulation.

Sear and Jackson adopted a formal approach through the modification of the fundamental graph sum of Wertheim by adding the ring integral and then proceeding with a minimisation of the residual Helmholtz
free energy by functional derivation, while Chapman and co-workers opted for a mass-balance approach. The two resulting approaches were later shown to be equivalent [56–58]. In their paper, Sear and Jackson also studied the formation of intermolecular rings from a specified number of associating spherical molecules and the competition of the formation of ring-like and (open) chain-like aggregates [52]. The approach was shown to be useful in modeling the phase behavior of hydrogen fluoride (HF), where accounting for ring and chain-like clusters is key to capture the maximum observed in the enthalpy of vaporisation [59].

García-Cuéllar et al. [56,60] followed on from the work of Ghonasgi and Chapman [54,55], extending their approach to mixtures, and modeling a telechelic polymer solution, while Avlund et al. [58] extended the theory to model intramolecular association in chain molecules with multiple association sites. The competition between intermolecular and intramolecular bonding was accounted for by designating a site to promote intramolecular association as well as intermolecular association (all other sites participating in either only intermolecular association or intramolecular association with site A). The theory was applied to model the phase behavior of binary mixtures of glycol ethers in water and alkanes [61]. Unfortunately, in comparisons with experimental data, no significant improvement over calculations with the original TPT1 approach was found.

Intramolecular hydrogen bonding was also studied in interfacial systems by Marshall et al. [62] by considering the inhomogeneous form of the Helmholtz free energy using classical density functional theory. The calculated surface tension and solid/liquid partitioning of four-segment chain molecules exhibiting intra- and intermolecular hydrogen bonding were found to be in excellent agreement with simulation. The method is computationally very demanding for molecules with more than five segments, and in later work the authors proposed an approach to speed up the calculation through the use of Monte Carlo ensemble averages to compute a number of the integrals required [63]. An extension for the case of mixtures of associating chains with multiple sites in a water-like solvent was also developed [64]. In the absence of ring formation, all molecules should be found to be non-associated in the low-density limit. In the case of two-site models the correct limit is recovered, but unfortunately, for larger number of sites, the wrong limit is obtained.

Wertheim’s TPT1 treatment has also proven to be valuable for the study of the interplay of chain and ring-like aggregation in colloidal patchy particles. Tavares and co-workers extended the approach to account for the competition of chain and ring formation in two-patch models (both type a) [65] and later in models with two patches of type a, situated in the poles of a hard-sphere, and a number of patches of type b placed along the equator [66]. In their initial methodology, ring formation was incorporated through aa bonding only, with the density distribution of rings obtained from independent computer simulations which were then incorporated in the extended free energy expression to obtain the required thermodynamic properties. It is important to acknowledge the earlier work of Almarza [3] in which it was shown that locating the a patches at the poles suppresses the formation of rings, while off-setting them from the poles alters the persistence of the chains and favors ring formation, which leads to re-entrant phase behavior in a pure fluid. A systematic investigation of the phase behavior of patchy fluids has also been carried out, considering varying geometries for the placement of patches and comparing computer simulation results with the extended theory [67]. The agreement between simulation and theory was found to be very good in general, except in models with small angles between the ring-forming sites, in which competition with branching may be important; this was addressed in a later generalisation of the model to account for ring formation [68].

Despite the progress in extensions of Wertheim’s TPT1 formalism, a general algebraic description to incorporate ring-like clusters as the result of intra- and intermolecular association, for arbitrary number of components and association sites, is still lacking. Here, we provide such a framework together with the corresponding contribution of association to the Helmholtz free energy, applicable to any EOS that incorporates Wertheim’s association term. We illustrate the application of the theory using two model systems to study the impact of intramolecular bonding using the statistical associating fluid theory for Mie potentials of variable range (SAFT-VR Mie) [69] within our generalised framework.

The model potentials considered are presented in Section 2, and the theory is developed in Section 3 following the approach of Sear and Jackson [52], in which ring graphs are added to the fundamental graph sum characteristic of Wertheim’s TPT to derive the relevant mass-action equations (Section 4). In order to facilitate the use of the new theory within an EOS, the residual free energy of a mixture of associating molecules that can form inter- or intramolecular bonds that lead to ring-like aggregates, is presented in Section 5. The specific expression for model systems using SAFT-VR Mie are given in Section 6 together with representative calculations, and final remarks are provided in Section 7.
2. Molecular model

We consider model associating molecules comprising \( m \geq 1 \) fused spherical segments of diameter \( \sigma \) (noting that \( m = 1 \) corresponds to a spherical model and \( m > 1 \) to a chain model), with short-range association sites embedded in the segments to mediate directional interactions of the hydrogen-bonding type (Figure 1). Further details are provided in the following subsections. We consider first the form of the pair potential needed to incorporate intra- as well as intermolecular interactions, impose a set of steric restrictions (following Wertheim’s original approach [26,27]), and discuss the types of aggregates resulting.

2.1. Molecular pair potentials

In order to treat both inter- and intramolecular interactions, two pair potentials are considered: a pair interaction potential \( \phi^{\text{inter}} \) between two molecules 1 and 2, and a pair interaction potential \( \phi^{\text{intra}} \) between two segments in the same molecule. The intermolecular pair potential is given as the sum of a reference \( \phi^{\text{ref}} \) and a perturbative association HB potential \( \phi^{\text{HB,inter}} [26,54] \):

\[
\phi^{\text{inter}}(r_{12}, \Omega_1, \Omega_2) = \phi^{\text{ref}}(r_{12}, \Omega_1, \Omega_2) + \sum_{\alpha} \sum_{\beta} \phi^{\text{HB,inter}}_{ss}(r_{1\alpha 2\beta}),
\]

where \( r_{12} \) represents the vector between the centres of two molecules with centres of mass in positions \( r_1 \) and \( r_2 \), i.e., \( r_{12} = r_2 - r_1 \), and \( \Omega_1 \) and \( \Omega_2 \) are vectors containing the respective angles characterising the molecular orientation and conformation, including all angles subtended by the constituting segments and the association sites. The reference intermolecular potential is assumed to be given by a sum of intersegment potentials between the segments that constitute the molecules [70], thus

\[
\phi^{\text{ref}}(r_{12}, \Omega_1, \Omega_2) = \sum_{\alpha} \sum_{\beta} \phi^{\text{ref}}_{ab}(r_{1\alpha 2\beta}),
\]

where \( r_{1\alpha 2\beta} \) is the distance between segment \( \alpha \) present in molecule 1 and segment \( \beta \) in molecule 2 given by \(|r_2 + d_\beta(\Omega_2) - (r_1 + d_\alpha(\Omega_1))|\), \( d_\beta \) is the displacement vector of the core of segment \( \beta \) from the centre of the molecule in position \( r_2 \), and \( d_\alpha \) is the displacement vector of the core of segment \( \alpha \) from the centre of the molecule in position \( r_1 \) (cf. Section 2.1). Any spherically symmetric potential (hard-sphere, square-well, Mie, Lennard-Jones, . . . ) can be used as \( \phi^{\text{ref}}_{ab}(r_{1\alpha 2\beta}) \).

Short-ranged directional site–site interactions between molecules are promoted through association sites. Segments may contain any number of sites. We use uppercase letters A,B, . . . , to refer to labelled sites, lowercase letters \( a, b, . . . \), to refer to association site types, and \( s, s', . . . \), to refer to sites in general. The interaction between sites \( s \) and \( s' \) in molecules 1 and 2 are modelled using square-well (SW) potentials, so that the intermolecular site–site interaction is given by

\[
\phi^{\text{HB,inter}}_{ss}(r_{12}, \Omega_1, \Omega_2) = \begin{cases} -\epsilon_{\alpha\beta}^{\text{HB}}, & \text{if } |r_{12} + d_s(\Omega_2) - d_s(\Omega_1)| < r_c, \\ 0, & \text{else}, \end{cases}
\]

where \( \epsilon_{\alpha\beta}^{\text{HB}} \) is the well depth of the square-well interaction between sites of types \( \alpha \) and \( \beta \), \( d_s(\Omega_1) \) is the displacement vector of site \( s \) from the centre of the molecule in position \( r_1 \) (cf. Figure 1), \( d_s(\Omega_2) \) is the displacement vector of site \( s' \) from the centre of the molecule in position \( r_2 \), \( |r_{12} + d_s(\Omega_2) - d_s(\Omega_1)| = |(r_2 + d_s(\Omega_2)) - (r_1 + d_s(\Omega_1))| \) denotes the centre-centre distance between association sites \( s \) and \( s' \) in molecules 1 and 2 at coordinates \((r_1, \Omega_1)\) and \((r_2, \Omega_2)\), respectively, and \( r_c \) is the range of the association interaction between sites of types \( \alpha \) and \( \beta \).

The intramolecular potential \( \phi^{\text{intra}} \) is analogously defined as

\[
\phi^{\text{intra}}(\Omega_1) = \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma} \phi^{\text{HB, intra}}_{ss'}(\Omega_1),
\]

where the distance between segments and sites is given by the orientational/conformational vector \( \Omega_1 \), and the
factor 1/2 is included to prevent double counting, since the associating sites of the pair $ss'$ are located in the same molecule. The reference intramolecular potential is assumed to be given by a sum of intersegmental potentials between the segments that constitute the molecule [70]:

$$
\phi^\text{ref}(\Omega_1) = \frac{1}{2} \sum_\alpha \sum_\beta \phi^\text{ref}_{\alpha\beta}(r_{1\alpha1\beta}).
$$

(5)

**Figure 2.** Constraints of TPT1 [28] restrict the following bonding schemes: (a) the repulsive cores of two associating molecules prevent a third core participating in bonding; (b) a site may participate in one bond only; (c) double-bonding between any two segments is not permitted.

**Figure 3.** Examples of association aggregates formed in a spherical model fluid of molecules shown in (a) with three association sites (labelled A, B, C). The check marks and crosses indicate whether the aggregate type is accounted for by the theory presented in our current work or not: (a) monomer species; (b) open branched chain; (c) intermolecular ring bonded through A–B bonds only; (d) intermolecular ring bonded through A–B bonds only, with further branching; (e) A–B intermolecular ring associated to a B–C intermolecular ring; and (f) intermolecular ring involving more than one type of site–site interaction.
The site–site intramolecular interaction is modelled using a SW potential so that,

\[ \phi_{ss'}^{\text{HB}, \text{intra}}(\Omega_1) = \begin{cases} -\epsilon_{ab}^{\text{HB}} & |d_{s'}(\Omega_1) - d_s(\Omega_1)| < r_{ab}^c, \\ 0 & \text{else} \end{cases} \]

and the potential is a function of the site–site distance and orientation of sites \(s\) and \(s'\) of a molecule in coordinates \(r_1\) given by the vector \(|d_{s'}(\Omega_1) - d_s(\Omega_1)|\). The intramolecular potential is non-zero only for the pair of sites in the molecule that are sterically (site–site distance < \(r_{ab}^c\)) and energetically (\(\epsilon_{ab}^{\text{HB}} > 0\)) favourable to association.

### 2.2. Steric restrictions

Bonding constraints restricting the type of possible site–site interactions, are imposed following those of the original TPT1 approach [26,27] (Figure 2). In the first restriction, when a bond between two sites located in different segments is formed, the repulsive cores of the two

![Figure 4. Examples of the association aggregates that can be found in a fluid of chain molecules shown in (a) with multiple association sites, A, B, C, ... The check marks and crosses indicate whether the aggregate type is accounted for by the theory presented in our current work or not: (a) monomer species; (b) intramolecular ring; (c) open chain aggregate; (d) intermolecular ring containing only one type of site–site bond (B–C); (e) intramolecular ring with a C–E bond associated to an A–E intramolecular ring; (f) branched A–E intramolecular ring; (g) molecule involved in more than one ring; and (h) ring involving more than one type of site–site interaction.](image-url)
segments prevent a third from participating in bonding without the cores overlapping (Figure 2(a)). In the second restriction the angle between any two given sites in a molecule is assumed to be large enough to prevent double (or multiple) bonding between two segments, such that a site may participate in one bond only (Figure 2(b)). Finally, double bonding between any two segments is not permitted (Figure 2(c)).

In the molecular model considered no limit is imposed on the number of association sites that are located in a given segment or molecule and the precise positions of the sites in a given segment are not specified. It is however assumed that their configuration is such to conform to the steric incompatibilities described earlier. Accordingly, a one-site molecule leads only to the formation of dimers, while in a two-site molecule linear chain aggregates (dimers, trimers, tetramers, . . .) as well as rings (of inter- or intramolecularly bonded molecules) of any size may form. Open, linear and branched chains as well as ring aggregates may form in the case of models with three or more association sites (see Figure 3 and 4).

In the case of the formation of ring-like clusters and intramolecular association, the steric restrictions do not limit the formation of such aggregates. In our framework we account for ring-like structures that are formed by molecules of the same species only and that have a single type of site–site interaction within the ring aggregate; different rings, each including a different type of site–site interaction are, nevertheless allowed. Moreover, ring-ring aggregate formation is also considered. In multi-component systems ring-like aggregates of any of the components of the mixture are taken into account, including larger clusters of rings of different components. Examples of the types of aggregates accounted for in our approach are presented in detail in the following sections.

2.3. Spherical molecules

In a fluid of spherical molecules with two or more association sites, following the previously discussed restrictions, association can result in the formation of open aggregates (linear or branched chains; cf. Figure 3(b)) and intramolecular rings containing three or more molecules; intramolecular bonding and a two-sphere ring, which amounts to double bonding, are not considered. Intramolecular ring-like aggregates may form by association of three or more molecules of a given component involving site–site interactions of only one type, see examples in Figures 3(c–e). Intermolecular rings of arbitrary size may form, and ring—ring association that leads to the formation of larger clusters is also taken into account (cf. Figure 3(e)). Intermolecular ring-like clusters involving more than one type of site–site interaction are however excluded in our framework (cf. Figure 3(f)).

2.4. Chain molecules

In a fluid of chain molecules with \( m \geq 2 \) segments (Figure 4), in addition to the types of intermolecular interactions discussed in the previous section, intramolecular site–site interactions are also possible. Examples of the types of possible aggregates in such a fluid are presented in Figure 4. The same restrictions for the formation of intermolecular open clusters and ring-like aggregates as described for the case of spherical molecules apply. Intramolecular association is also possible (i.e. aggregates involving only one molecule; cf. Figure 4(b)), as well as closed aggregates involving only two molecules (cf. Figure 4(d)). In the case of a ring-like aggregate with two molecules, only one type of site–site interaction is allowed within the ring.

3. The Helmholtz free energy

The Helmholtz free energy of a pure associating fluid can be obtained considering molecules in a non-associated monomer, dimer, chain, intramolecular ring of size \( \tau \), and intramolecular ring configurations, and any other cluster species that may be present in the fluid as distinct species, so that [55,71]

\[
A = N_0 \mu_0 + N_{\text{dimer}} \mu_{\text{dimer}} + N_{\text{trimer}} \mu_{\text{trimer}} + \cdots \\
+ \sum_{R=1}^{N_{\text{RS}}} N_{\tau R} \mu_{\tau R} + \cdots + N_{\text{intra}} \mu_{\text{intra}} + \cdots - PV,
\]

(7)

where \( P \) is the pressure, \( V \) the total volume, \( \mu \) the chemical potential, and the subscripts refer to the type of species. \( N_0 \) is the number molecules in a monomer configuration, \( N_{\text{dimer}} \) is the number of molecules participating in dimers, \( N_{\text{trimer}} \) is the number of molecules participating in aggregates of open chains of three molecules, and so on . . . , \( N_{\tau R} \) is the number of molecules which are intermolecularly bonded into rings of size \( \tau R \geq 2 \) for chain molecules or \( \tau R \geq 3 \) for spherical molecules (with \( \tau R \) corresponding to the number of molecules in the ring). The summation is over the number of different ring sizes, \( N_{\text{RS}} \). Double and larger ring clusters, as well as ring-chain clusters may also form, extending Equation (7). \( N_{\text{intra}} \) is the number of molecules with an intramolecular hydrogen bond only; i.e., not participating in any other type of aggregate. Furthermore, and \( \mu_0 \) is the chemical potential of the monomer species, \( \mu_{\text{dimer}} \) is the chemical potential of the dimers, \( \mu_{\text{trimer}} \) is the chemical potential of the trimers, etc.
of the trimers, \( \mu_{\text{tg}} \) is the chemical potential of intramolecular rings of size \( \tau_R \), and \( \mu_{\text{intra}} \) is the chemical potential of molecules with intramolecular association only. At equilibrium, the chemical potential of all the molecules in the system must be the same regardless of cluster arrangements, including monomers, which allows us to write Equation (7) as

\[
A = N\mu_0 - PV,
\]

where the total number of molecules \( N \) in the system, regardless of their bonded state, is given by

\[
N = N_0 + N_{\text{dimer}} + N_{\text{trimer}} + \cdots + \sum_{R=1}^{N_{\text{rg}}} N_{\tau_R} + \cdots + N_{\text{intra}}.
\]

Following an observation by Andersen [16,17] that the form of the combinatorial terms in the cluster expansion of associated fluids is independent of the density, the Helmholtz free energy can be determined by considering the limit of low density; the results obtained in the low-density limit are also valid at higher densities. It is thus possible to derive an expression for the association term of an ideal associating fluid following Dalton’s law, i.e., assuming a proportionality between the pressure and the total number of aggregates [55,59]:

\[
N_{\text{aggregates}} = N_0 + \frac{N_{\text{dimer}}}{2} + \frac{N_{\text{trimer}}}{3} + \cdots + \sum_{R=1}^{N_{\text{rg}}} \frac{N_{\tau_R}}{R} + \cdots + N_{\text{intra}},
\]

so that

\[
A_{\text{aggregates}} = N\mu_0 - N_{\text{aggregates}}kT,
\]

where \( k \) is the Boltzmann constant and \( T \) the absolute temperature. The ideal chemical potential of the monomer species is given by [72]

\[
\mu_0 = kT \ln (\rho_0 \Upsilon(T)),
\]

where \( \rho_0 = N_0/V \) is the number density of monomers and \( \Upsilon(T) \) is a function of temperature that includes the translational, rotational, and vibrational parts of the molecular partition function (the de Broglie volume) as well as other contributions from molecular configurations [70].

From this analysis it can be seen that the number of aggregates in the system and consequently the expression for the Helmholtz free energy directly depend on the type of species considered. We start by considering the simplest case of a non-associating fluid then proceed to incorporate association.

### 3.1. Non-associating monomers

In a non-associating system the only species present is the monomer, so that, \( N_{\text{aggregates}} = N_0 = N \) and the number density \( \rho \) is equal to the density of monomers; i.e., \( \rho_0 = \rho \). It therefore follows from Equations (11) and (12) that the free energy in a non-associating ideal system \( A_0 \) can be written in terms of the number density of the system as

\[
A_0 = NkT \left[ \ln (\rho \Upsilon(T)) - 1 \right],
\]

which corresponds to an ideal gas [72].

### 3.2. Free monomers and open-chain aggregates

We consider a system at fixed \( V, T \) containing ten molecules; if two molecules associate to form a dimer, nine aggregates remain, if further association occurs with another molecule to form a trimer, \( N_{\text{aggregates}} = 8 \) and so on. In the general case, the number of aggregates in a fluid where the molecules associate only in chains (branching is possible but not ring formation) is reduced by one per association bond formed. It is helpful to define at this point the fraction of molecules with sites in set \( \alpha \subseteq \Gamma \) free as \( X_\alpha \). Note that the molecules that contribute to \( X_\alpha \) have all sites in \( \alpha \) free at least, but not exclusively (other sites might be free). In this notation, \( X_{\Gamma} \) corresponds to the fraction of molecules with all sites free, i.e., the fraction of free monomer molecules. Traditionally this fraction is referred to as \( X_0 \) [30]; we follow this convention hereafter\(^1\). The fraction of molecules with (at least) site \( s \) free is given by \( X_s \), often referred to as \( X_s \), and the fraction of molecules with \( s \) bonded is then given by \( (1 - X_s) \). This means that for a model with the set of sites \( \Gamma \), the number of aggregates can be obtained as

\[
N_{\text{aggregates}} = N - \frac{1}{2}N \sum_{s \in \Gamma} (1 - X_s),
\]

where the factor 1/2 is included to avoid double counting the number of bonds, as each bond involves two sites. The corresponding form of the Helmholtz free energy is thus given by

\[
A = NkT \left[ \ln (\rho_0 \Upsilon(T)) - 1 + \frac{1}{2} \sum_{s \in \Gamma} (1 - X_s) \right].
\]

### 3.3. Free monomers, open chains, and intramolecular ring-like aggregates

A system in which intramolecular association is taken into account in addition to the formation of intermolecular chain-like aggregates is now considered. The formation of an intramolecular bond does not change the
number of aggregates in the system but, by competing with the formation of intermolecular bonds, it will affect the overall free energy of the system. To avoid confusion, we introduce \( X_s^{\text{open}} \) as the fraction of molecules with site \( s \) not bonded in an open chain, and re-express Equation (14) for the total number of aggregates as \[ N_{\text{aggregates}} = N - \frac{1}{2} N \sum_{s \in \Gamma} (1 - X_s^{\text{open}}), \] (16)

and the Helmholtz free energy as

\[
A = NkT \left[ \ln (\rho_0 \Upsilon(T)) - 1 + \frac{1}{2} N \sum_{s \in \Gamma} (1 - X_s) \right. \\
\left. - \frac{1}{2} N \sum_{s \in \Gamma} (1 - X_s^{\text{intra}}) \right].
\] (17)

One should also note that the sum of the fraction of molecules with site \( s \) bonded in open chains and the fraction of molecules with site \( s \) bonded in intramolecular rings \( (1 - X_s^{\text{intra}}) \) corresponds to the total fraction of molecules with site \( s \) bonded \( (1 - X_s) \). Therefore, for a given site \( s \) a relation

\[
(1 - X_s^{\text{open}}) + (1 - X_s^{\text{intra}}) = 1 - X_s,
\] (18)

can be written, which can be substituted in Equation (17) so that,

\[
A = NkT \left[ \ln (\rho_0 \Upsilon(T)) - 1 + \frac{1}{2} N \sum_{s \in \Gamma} (1 - X_s) \right. \\
\left. - \frac{1}{2} \sum_{s \in \Gamma} (1 - X_s^{\text{intra}}) \right].
\] (19)

The last term in this expression corresponds to the fraction of molecules forming intramolecular rings \( \xi_1 \):

\[
\xi_1 = \frac{1}{2} \sum_{s \in \Gamma} (1 - X_s^{\text{intra}}),
\] (20)

where we use the subscript \( 1 \) to indicate a ring-like aggregate involving one molecule only.

### 3.4. Free monomers, open chains, and intra- and intermolecular ring-like aggregates

The formation of each intermolecular ring cluster involving \( \tau \) molecules reduces the number of free aggregates in the system by \( \tau - 1 \). Accounting for this, the number of aggregates in the system is now given by

\[
N_{\text{aggregates}} = N - \frac{1}{2} N \sum_{s \in \Gamma} (1 - X_s^{\text{open}}) \\
- \sum_{R=1}^{N_{\text{RS}}} \frac{N_{\text{RS}}}{R_{\tau R}} (\tau - 1) \times N_{\tau R} / R_{\tau R},
\] (21)

where \( N_{\tau R} \) is the number of molecules in rings of size \( \tau_R \geq 1 \), which can be obtained in terms of the fraction of molecules involved in ring formation \( \xi_{\tau_R} \) as

\[
N_{\tau R} = \xi_{\tau_R} N.
\] (22)

It is also possible to write \( \xi_{\tau_R} \) in terms of the fraction of molecules with a ring-bonding site \( s \) free \( X_s^{\text{rings } \tau_R} \) as

\[
\xi_{\tau_R} = \frac{1}{2} \sum_{s \in \Gamma} (1 - X_s^{\text{rings } \tau_R}),
\] (23)

noting that in the case of \( \tau_R = 1 \), \( \xi_{\tau_R} = \xi_1 \) and \( X_s^{\text{rings } \tau_R} = X_s^{\text{intra}} \). Substituting Equations (22) and (23) in (21) allows one to rewrite the number of aggregates in the system as

\[
N_{\text{aggregates}} = N \left[ 1 - \frac{1}{2} \sum_{s \in \Gamma} (1 - X_s^{\text{open}}) \\
- \sum_{R=1}^{N_{\text{RS}}} \frac{N_{\text{RS}}}{R} (\tau - 1) \times \frac{1}{2} \sum_{s \in \Gamma} (1 - X_s^{\text{rings } \tau_R}) \right].
\] (24)

Analogously to Equation (18), the relation

\[
(1 - X_s^{\text{open}}) + \sum_{R=1}^{N_{\text{RS}}} (1 - X_s^{\text{rings } \tau_R}) = 1 - X_s,
\] (25)

holds and, upon insertion in Equation (24), the number of aggregates is simply given by

\[
N_{\text{aggregates}} = N \left[ 1 - \frac{1}{2} \sum_{s \in \Gamma} (1 - X_s) + \sum_{R=1}^{N_{\text{RS}}} \frac{1}{R_{\tau R}} \times \xi_{\tau_R} \right],
\] (26)

which can be used to write the total Helmholtz free energy of a fluid comprising open and ring-like aggregates (in the low-density limit) as

\[
A = NkT \left[ \ln (\rho_0 \Upsilon(T)) - 1 + \frac{1}{2} N \sum_{s \in \Gamma} (1 - X_s) \\
- \sum_{R=1}^{N_{\text{RS}}} \frac{1}{R_{\tau R}} \xi_{\tau_R} \right].
\] (27)

Equation (27) provides a way to calculate the total Helmholtz free energy with the proviso that the fractions of molecules in different bonded states \( X_0 \) (fraction of
monomers), \( X_{\alpha}, \alpha \subseteq \Gamma \), and the fraction \( \xi_{\alpha k} \) of molecules involved in clusters in rings of size \( r_k \) are known at a specified thermodynamic state \( N, V, T \). In order to derive expressions for these distributions of molecular bonded states we make use of Wertheim’s TPT1 [28,73], extending the framework to account for ring formation.

### 4. Mass action equations

Wertheim introduced the multi-density concept, which translates into treating molecules in different bonding states as different pseudo-species, mirroring the framework presented in Section 3. In his formalism, a density \( \rho_s \) is associated with molecules with site \( s \) bonded, a density \( \rho_{s'} \) to molecules with sites \( s \) and \( s' \) bonded, and so on. More generally, the number density of molecules with sites in the set \( \alpha \) bonded is \( \rho_{\alpha} \), where \( \alpha \) is a subset of the set of all sites \( \Gamma \), including the empty set \( \emptyset \) that conventionally corresponds to the density of molecules not bonded \( \rho_0 \), for \( \alpha = \emptyset \). Following Sear and Jackson [74], we present expressions considering an inhomogeneous system first, in which the densities are dependent on position coordinates \( r \) and the orientation and conformation vector \( \Omega \), and we make use of (1) as shorthand notation for \( (r_1, \Omega_1) \) which refers to all degrees of freedom of molecule 1.

The local number density \( \rho(1) \) of particles in coordinates (1) \( (\int \rho(1)d(1) = N) \) can be written as the sum of all bonding-state densities as [28]

\[
\rho(1) = \sum_{\alpha \subseteq \Gamma} \rho_\alpha(1). \tag{28}
\]

Furthermore, Wertheim [28] defined \( \sigma_\alpha(1) \) as the density of molecules that are not bonded plus the ones that are bonded exactly through one, some or all sites in the set \( \alpha \). Accordingly, \( \sigma_{\Gamma - \alpha}(1) \) is the density of molecules with (at least) the sites in the set \( \alpha \) free. As an example, in a two-site model with sites A and B, i.e., \( \Gamma = \{A, B\} \), \( \sigma_{\Gamma - A}(1) = \sigma_B(1) \) is the density of molecules with site B free, and is given by the sum \( \rho_0(1) + \rho_B(1) \). More generally,

\[
\sigma_\alpha(1) = \sum_{\delta \subseteq \alpha} \rho_\delta(1), \tag{29}
\]

with the special cases \( \sigma_\emptyset(1) = \rho_0(1) \) (the density of monomers) and \( \sigma_\Gamma(1) = \rho(1) \) the total number density of molecules in the system, regardless of their bonded state (given by \( N/V \), cf., Equation (9)). The density of molecules with sites in set \( \alpha \) free is related to the fraction of molecules with set of sites \( \alpha \) free by

\[
\sigma_{\Gamma - \alpha}(1) = \rho(1)X_\alpha(1), \quad \forall \alpha \subseteq \Gamma, \tag{30}
\]

and to the fraction of monomers

\[
\sigma_{\Gamma - \alpha}(1) = \rho_0(1) = \rho(1)X_\emptyset(1), \tag{31}
\]

for \( \alpha = \emptyset \).

In Wertheim’s work [28,29] the Helmholtz free energy expression is expressed as

\[
A_{\text{Werth}}[[\rho]] = kT \int \left[ \rho(1) \ln \left( \frac{\rho(1)}{\rho(1)} \right) \right] + \rho(1) + Q(1) \ d(1) - kT \Delta c(0)[[\rho]], \tag{32}
\]

where the notation \( \{\rho\} = \{\rho_0, \rho_A, \ldots\} \) highlights the dependence of the free energy on all bonding state densities. Following Wertheim’s formalism, each density \( \rho_\alpha \) is defined in terms of a sum of graphs [2]. \( A_{\text{Werth}} \) corresponds to the difference in free energy between the associating system and a reference (non-associating) system,

\[
A_{\text{Werth}} = A - A_{\text{ref}}, \tag{33}
\]

where all interactions other than association are accounted for in the reference fluid, including chain connectivity in the case of chain-like molecules as well any repulsive and dispersion interactions included in the model.

Equation (32) is a general expression that incorporates all possible aggregates (i.e., inter- and intramolecular rings as well as open chain aggregates); it provides a different route to express the association contribution to the free energy given by Equation (27). The difference between the two equations is that Equation (32) is a general expression, before equilibrium is imposed, that reduces to Equation (27) at equilibrium and after including the ideal contribution. As such, the mass action equations for the distribution of bonding states needed in Equation (27) can be obtained by minimisation of the free energy given by Equation (32) [26,28,52].

The function \( Q \) in Equation (32) results from the non-trivial derivation originally presented by Wertheim [28] and recently re-derived in an excellent review by Zimpias and Gross [73]. It is given by

\[
Q(1) = -\sum_{\alpha \subseteq \Gamma} \sigma_{\Gamma - \alpha}(1) + \rho_0(1) \times \sum_{\{\gamma_1, \ldots, \gamma_M\} \in \mathcal{P}(\Gamma), M \geq 2} (-1)^M(M - 2)! \prod_{i=1}^{M} \frac{\sigma_{\gamma_i}(1)}{\rho_0(1)}, \tag{34}
\]

where the first sum is over all sites and the second over all possible partitions of the set \( (\mathcal{P}(\Gamma)) \), i.e., the possible ways to divide \( \Gamma \) into pairwise disjoint subsets. The elements of the partition of \( \Gamma \) into \( M \) subsets are
\(\gamma_1, \gamma_2, \ldots, \gamma_M\) and the condition \(M \geq 2\) ensures that each partition considered must contain two or more elements [28].

The functional \(\Delta \epsilon^{(0)} = \epsilon^{(0)} - \epsilon^{(0)}_{\text{ref}}\) in Equation (32) is referred to by Wertheim [28] as the fundamental graph sum, where \(\epsilon^{(0)}\) is the correction quantity to the free energy due to cluster-forming interactions, while \(\epsilon^{(0)}_{\text{ref}}\) is the contribution from interactions existing in the reference fluid. While \(Q\) remains unchanged regardless of the types of association clusters considered, \(\Delta \epsilon^{(0)}\) is dependent on these. In the following section we study in detail the different contributions in this term.

### 4.1. The fundamental graph sum \(\Delta \epsilon^{(0)}\)

The fundamental graph sum \(\Delta \epsilon^{(0)}\) relates the possible bonded states of the molecular model to the number of species in the system \(N_{\text{species}}\) (cf. Equation (10)). It includes all irreducible graphs responsible for the different aggregates. Graphs corresponding to open-chain aggregates can be topologically reduced [29,72] to graphs containing an association bond between pairs of particles (dimers), while ring graphs are irreducible and cannot therefore be accounted for by the same graphs corresponding to open chains [52,62]. Here, we consider open-chain (\(\Delta \epsilon^{(0)}_{\text{open}}\)), intermolecular ring (\(\Delta \epsilon^{(0)}_{\text{ring}}\)), and intramolecular association (\(\Delta \epsilon^{(0)}_{\text{intra}}\)) contributions separately:

\[
\Delta \epsilon^{(0)}\{\rho\} = \Delta \epsilon^{(0)}_{\text{open}}\{\rho\}
+ \Delta \epsilon^{(0)}_{\text{ring}}\{\rho\}
+ \Delta \epsilon^{(0)}_{\text{intra}}\{\rho\}. \tag{35}
\]

#### 4.1.1. Open chain-like aggregates

The graph sum for open-chain aggregates \(\Delta \epsilon^{(0)}_{\text{open}}\{\rho\}\) can be expressed as [40]

\[
\Delta \epsilon^{(0)}_{\text{open}}\{\rho\} = \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma} \int \sigma_{r-s}(1) \sigma_{r-s'}(2) g_{\text{ref}}(r_{12}, \Omega_1, \Omega_2) \times f_{s's'}(r_{12}, \Omega_1, \Omega_2) d(1) d(2), \tag{36}
\]

and accounts for the formation of open chain aggregates of any size. This equation is equivalent to the standard \(\Delta \epsilon^{(0)}\{\rho\}\) of TPT1 [40] for the case when \(\sigma_a\) is defined in terms of segments, i.e., if \(\sigma_{r-s}(1)\) is defined as the singlet number density of segments (or spherical molecules) in coordinates (1) with site \(s\) free. Our model molecule may, more generally, be non-spherical, instead consisting of an arbitrary number of segments. The graph sum corresponding to Equation (36) thus differs from the one used by Wertheim [40] in the sense that it involves molecular graphs, simplifying the analysis; this concept was employed by Marshall and Chapman [75] who showed the relation between segment-based and molecular graphs.

It is possible to interpret Equation (36) by physically considering that the formation of an association bond between sites \(s\) in molecule 1 and \(s'\) in molecule 2 is a function of the number densities of molecules with sites \(s\) and \(s'\) free, \(\sigma_{r-s}\) and \(\sigma_{r-s'}\), respectively, and the likelihood of these sites forming a bond, given by the product of the radial distribution of the reference (unbonded) fluid \(g_{\text{ref}}\) and the Mayer \(f\)-function of the \(ss'\) interaction, in this case, \(f_{s's'}(r_{12}, \Omega_1, \Omega_2) = \exp\left(-\beta \phi_{s's'}^{HB}\right) - 1\). Molecules interact across all volume and configurations, and therefore their coordinates are integrated over all space and orientations. A sum over all sites \(s\) and \(s'\) is carried out and a factor of one half is included to avoid counting the same pair of sites twice. Any pair of sites is allowed to associate provided that the respective \(\epsilon^{\text{HB}}_{ab}\) is larger than zero.

#### 4.1.2. Intermolecular ring aggregates

The intermolecular ring contribution from clusters of the types represented in Figures 3(c–e) and 4(d), can be written as [76]

\[
\Delta \epsilon^{(0)}_{\text{ring}}\{\rho\} = \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma} \sum_{R=1}^{N_{\text{RS},s's'}} \frac{1}{\tau_R} \times \int \prod_{i=1}^{\tau_R} \left[g_{\text{ref}}(r_{i,i+1}, \Omega_i, \Omega_{i+1}) f_{s's'}(r_{i,i+1}, \Omega_i, \Omega_{i+1})\right] \times \prod_{j=1}^{\tau_R} \sigma_{r-s}(j) d(j), \tag{37}
\]

generalising the expression provided in reference [53] to an arbitrary number of ring-forming sites, and where the number of molecules per intermolecular ring \(\tau_R \geq 2\), \(\forall R\). The density of molecules with both sites \(s\) and \(s'\) free, \(\sigma_{r-s's'}\), is required to model the extent of rings formed, hence the summations in Equation (37) are over sites \(s\) and \(s'\), with \(s \neq s'\). This differs from Equation (36) where association requires only one site per molecule to be free (\(\sigma_{r-s}\)), where \(s\) and \(s'\) may be the same site (in a different molecule).

In a ring aggregate we consider each molecule in the ring to be close to contact with their immediately-neighbouring molecules only. Accordingly, in Equation (37) we have followed the approximation of Sear and Jackson [74] for the \(\tau_R\)-body distribution function, which is given in terms of pair distribution functions, one per
pair of sites, so that

\[ g^{(r)}(12 \ldots r_R) = \prod_{i=1}^{r_R} g(r_{i,i+1}, \Omega_i, \Omega_{i+1}), \]

(38)

where the convention \((i + 1) = 1\), for \( i = r_R \) is used.

The analysis of the expression for \( \Delta c_{\text{intra}}^{(0)} \) follows the same rationale as for \( \Delta c_{\text{open}}^{(0)} \): in order for an intermolecular ring of size \( \tau_R \) to form, the molecules must come close enough to form \( \tau_R \) links between them. It is thus required that sites \( a \) and \( b \) in each molecule involved are free, correctly oriented and that their interaction energy parameter \( \varepsilon_{ab}^{\text{HB}} \) is larger than zero. As rings of different sizes may form, a sum over the ring size \( R \) index is carried out and a factor \( 1/\tau_R \) is introduced to avoid multiple counting. The last bond, \( g_{\text{ref}}(r_{r_1,1}, \Omega_1, \Omega_1) f_{\text{intra}}^{(1)}(\Omega_1) \) is treated as of a different nature considering its formation to turn an open-chain cluster into a ring. The expression presented in Equation (37) ensures that a given intermolecular ring is composed by molecules bonded through site–site interactions of only one type. Ring aggregates and structures involving more than one type of site–site interaction (see Figure 3(f) and Figure 4(g,h)) can be accounted for in Equation (37) by incorporating the relevant Mayer \( f \)-functions and densities, but have not been considered in our current work.

### 4.1.3. Intramolecular association

The intramolecular association contribution from clusters of the types represented in Figures 4(b,e,f), is given by [76]

\[ \Delta c_{\text{intra}}^{(0)} \]

\[ \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma_{-s}} \int \left[ g_{\text{ref}}^{(1)}(\Omega_1) \right] f_{\text{intra}}^{(s,s')}(\Omega_1) \sigma_{\Omega_1}^{-1}(1) d(1), \]

(39)

where \( g_{\text{ref}}^{(1)}(\Omega_1) \) is the reference intramolecular distribution function, and \( f_{\text{intra}}^{(s,s')}(\Omega_1) = \exp \left( -\beta \phi_{\text{HB,intra}}^{(s,s')}(\Omega_1) \right) - 1 \). The intramolecular distribution function contains information on the likelihood of any two site-containing segments in a given molecule being at a close distance enough to bond. Additionally the number density of molecules of coordinates (1) with sites \( s \) and \( s' \) simultaneously free, \( \sigma_{\Omega_1}^{-1}(1) \), is required. This is different from the expression published by Sear and Jackson [52] as we are using molecular instead of segment graphs [75]. In our current work only one intramolecular association interaction is allowed for a given molecule, at a given time (involving any two sites). It is possible to include further intramolecular association interactions within a molecule by adding further contributions to Equation (39); for simplicity, we do not consider such cases here.

#### 4.1.4. Homogeneous formulation

In homogeneous systems the multiple densities can be treated as independent of position and orientation. Thus, following the definition of the fraction of molecules with a set of sites \( \alpha \) free (\( X_\alpha \)) in Equation (30), the free energy expression (Equation (32)) and function \( Q \) (Equation (34)), can be rewritten as

\[ A_{\text{Wert}} = NkT \left[ \ln X_0 + 1 + \frac{Q}{\rho} - \frac{\Delta c^{(0)}}{N} \right], \]

(40)

and

\[ \frac{Q}{\rho} = -\sum_{s \in \Gamma} X_s + X_0 \]

\[ \times \sum_{\gamma_1, \ldots, \gamma_M} (-1)^M (M - 2)! \prod_{j=1}^{M} \frac{X_{\gamma_j}}{X_0}, \]

(41)

with the contributions to the fundamental graph sum \( \Delta c^{(0)} \) in Equation (35) given by

\[ \frac{\Delta c_{\text{intra}}^{(0)} \left[ \{ \rho \} \right]}{N} = \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma} X_s X_{s'} \rho \Delta s_{s'}, \]

(42)

for open-chain aggregates from Equation (36), with

\[ \Delta s_{s'} = \int g_{\text{ref}}(r_{12}, \Omega_1, \Omega_2) f_{\text{intra}}^{(s,s')}(r_{12}, \Omega_1, \Omega_2) d(r_{12}) \]

\[ \times d(\Omega_1) d(\Omega_2), \]

(43)

and

\[ \frac{\Delta c_{\text{inter}}^{(0)} \left[ \{ \rho \} \right]}{N} = \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma} \sum_{\gamma_1, \ldots, \gamma_M} \frac{1}{\tau_R} (X_{ss'})^{\tau_R} \rho^{\tau_R} \Delta s_{s', \Gamma}, \]

(44)

for intermolecular ring aggregates from Equation (37), with

\[ \Delta s_{s', \Gamma} = \int g_{\text{ref}}(r_{i,i+1}, \Omega_i, \Omega_{i+1}) f_{\text{intra}}^{(s,s')} (r_{i,i+1}, \Omega_i, \Omega_{i+1}) \]

\[ \times \prod_{j=1}^{\tau_R - 1} \sum_{k=1}^{\tau_R} d(r_{ij,j+1}) \prod_{k=1}^{\tau_R} d(\Omega_k), \]

(45)

Here, ring sizes \( \tau_R \geq 2 \) for all \( R \).
The formation of an intermolecular ring aggregate can be rationalised as involving two steps: first, a chain cluster of \( \tau - 1 \) association links between sites \( s \) in position \( r_i \) and \( s' \) in \( r_{i+1} \) is formed; once a chain is formed, there is a probability of the two ends of the chain being at a close enough distance to form a final \( s - s' \) bond that completes the ring. Since in the TPT1 formalism the angles between segments are not accounted for explicitly and complete flexibility of the molecular bond angles is assumed, the pairs of sites that promote ring formation are an explicit input in the model. Following previous work \[52,74\], a parameter \( W_{ss',R} \) is introduced to capture the likelihood of ring-forming sites being in contact. An approximate expression,

\[
\Delta_{ss',R}^{\text{inter}} \approx (\Delta_{ss'})^{\tau_R} W_{ss',R} \tag{46}
\]

is proposed, which corresponds to having \( \tau_R - 1 \) site–site bonds, each characterised by \( \Delta_{ss'} \), and an additional bond characterised by \( \Delta_{ss'} W_{ss',R} \). The parameter \( W_{ss',R} \) has dimensions of inverse volume and, at a first level of approximation, it is here considered to be independent of density, temperature, and composition.

Lastly, for the case of an intramolecular bond, from Equation (39),

\[
\frac{\Delta_{\text{intra}}^{(0)}([\rho])}{N} = \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma - s} X_{ss'} X_{ss'}^{\text{intra}}, \tag{47}
\]

with

\[
X_{ss'}^{\text{intra}} = \int g_{\text{ref}}^{\text{intra}}(\Omega_1) g_{\text{intra}}^{\text{intra}}(\Omega_1) d(\Omega_1). \tag{48}
\]

The formation of an intramolecular site–site bond is considered to be of similar nature to the last bond leading to the formation of an intermolecular ring-like aggregate, thus,

\[
\Delta_{ss'}^{\text{intra}} \approx \Delta_{ss'} W_{ss'}. \tag{49}
\]

Comparing Equation (44) and (47), with the definitions given in Equation (46) and (49), it becomes apparent that the formation of an intramolecular bond corresponds to a special case of ring-aggregate formation with \( \tau = 1 \); i.e., a ring aggregate involving one molecule only. As a result, the contribution from the formation of intermolecular rings or intramolecular bonds to the fundamental graph sum can be collected in one term with \( \tau_R \geq 1 \):

\[
\frac{\Delta_{\text{ring}}^{(0)}([\rho])}{N} = \frac{\Delta_{\text{inter}}^{(0)}([\rho])}{N} + \frac{\Delta_{\text{intra}}^{(0)}([\rho])}{N}
\]

\[
\approx \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma - s} \frac{1}{\tau_R} \sum_{R=1} X_{ss'} \rho \Delta_{ss'}^{R} W_{ss',R} / \rho, \tag{50}
\]

which can be related to the fraction of molecules in rings of size \( \tau_R \) simply as

\[
\xi_{\tau_R} = \frac{1}{2} \sum_{s \in \Gamma} \sum_{s' \in \Gamma - s} (X_{ss'} \rho \Delta_{ss'})^{\tau_R} W_{ss',R} / \rho \tag{51}
\]

as required in Equation (27).

4.1.5. Homogeneous formulation considering site types

It is useful to define a more compact notation in which the number of sites of the same type \( a \) is labelled \( s_a \). Taking \( a \) to be the type of site \( s \) and \( b \) to be the type of site \( s' \), the \( \Delta \) and \( W \) are unchanged, i.e., \( \Delta_{ab} = \Delta_{ss'} \), \( \Delta_{ab,R} = \Delta_{ss',R} \), \( W_{ab} \), and \( W_{ab,R} = W_{ss',R} \). The fraction of molecules of a site of type \( a \) free is the same as the fraction of molecules with site \( s \) free, i.e., \( X_a = X_s \), so that

\[
\frac{\Delta_{\text{open}}^{(0)}([\rho])}{N} = \frac{1}{2} \sum_{a=1}^{N_a} \sum_{b=1}^{N_b} s_a X_a s_b X_b \rho \Delta_{ab}, \tag{52}
\]

where the summations are now over site types. Similarly, Equation (50) and (51) can be extended to multiple sites of each type so that\[76\]

\[
\frac{\Delta_{\text{ring}}^{(0)}([\rho])}{N} = \frac{1}{2} \sum_{a=1}^{N_a} \sum_{b=1}^{N_b} \sum_{R=1}^{N_R} \frac{1}{\tau_R} \sum_{s' \in \text{op}_{ab}} X_{ss'} \rho \Delta_{ab}^{R} W_{ab,R} / \rho, \tag{53}
\]

and

\[
\xi_{\tau_R} = \frac{1}{2} \sum_{a=1}^{N_a} \sum_{b=1}^{N_b} \left( \sum_{s' \in \text{op}_{ab}} X_{ss'} \rho \Delta_{ab} \right)^{\tau_R} W_{ab,R} / \rho, \tag{54}
\]

where \( \text{op}_{ab} \) is the set of ordered pairs of site-type \( ab \) starting with a site of type \( a \) followed by a site of type \( b \). The
number of pairs in the set $o p_{ab}$ is given by,

$$|o p_{ab}| = \begin{cases} s_a(s_a - 1), & \text{if } a = b \\ s_a b, & \text{otherwise} \end{cases}.$$  
(55)

Substituting Equations (41), (52), and (53) in equation (40), we write the general expression for $A_{\text{Wert}}$ as

$$A_{\text{Wert}} = N k T \left[ \ln X_0 + 1 - \sum_{a=1}^{N_{\text{ST}}} s_a X_a + X_0 \sum_{\{\gamma_1, \ldots, \gamma_M\} \in \mathcal{P}(\Gamma), M \geq 2} (-1)^{M} (M - 2)! \prod_{i=1}^{M} \frac{X_{\Gamma - \gamma_i}}{X_0} \right]$$

$$- \sum_{a=1}^{N_{\text{ST}}} \sum_{b=1}^{N_{\text{ST}}} s_a X_a s_b X_b \rho \Delta_{ab}$$

$$- \sum_{a=1}^{N_{\text{ST}}} \sum_{b=1}^{N_{\text{ST}}} \sum_{R=1}^{N_{\text{RS}, ab}} \frac{1}{\mathcal{P}}$$

$$\times \left[ \left( \sum_{s,s' \in o p_{ab}} X_{s,s'} \right) \rho \Delta_{ab} \right] \frac{\tau_{R}}{W_{ab,R}} \right],$$

(56)

from which the distribution of bonding states $X_0, X_a$, and $X_{ab}$ can be obtained.

### 4.2. Distribution of bonding states

The equilibrium conditions that determine the self-consistent values for $X_0, X_a, \ldots, X_{ab}, \ldots$ are those that make $A_{\text{Wert}}$ stationary with respect to these parameters [40]:

$$\frac{\partial A_{\text{Wert}}}{\partial X_\delta} \bigg|_{\{\gamma, \tau, \gamma, N\}} = 0, \quad \gamma \cap \delta = \emptyset,$$

(57)

where the number of variables $X_\delta$ for a component with a set of sites $\Gamma$ is $2^{\mid \Gamma \mid} - 1$, a number which grows exponentially with the number of sites $\mid \Gamma \mid$: $X_0$ for one-site, $X_0, X_A$, and $X_0$ for $\mid \Gamma \mid = 2$, $X_0, X_A, X_B, X_C, X_{AB}, X_{AC}$, and $X_{BC}$ for $\mid \Gamma \mid = 3$, etc. Minimising the free energy with respect to each of these variables will generate an increasingly large and rather complex system of equations. In this section, a compact formulation to determine the value of $X_\delta$ is presented, which results from the minimisation of the Helmholtz free energy for systems in which chain and ring-like aggregates are considered. We examine first the formulation in the absence of rings, recovering the well-known expressions of the statistical associating fluid theory (SAFT) [33,34,42] family of equations, then proceed to present the new expressions.

#### 4.2.1. Open chain clusters

In the original TPT1 formalism of Wertheim [26–29] $\Delta c^{(0)} = \Delta c_{\text{open}}^{(0)}$ since $\Delta c_{\text{ring}}^{(0)} = 0$. On minimisation of the Helmholtz free energy all of the site–site interactions are found to be independent of each other [33,37], so that

$$X_\alpha = \prod_{s \in \alpha} X_s = \prod_{i=1}^{M} X_{\beta_i},$$

$$\forall\{\beta_1, \ldots, \beta_M\} \in \mathcal{P}(\alpha), \ M \geq 1, \quad (58)$$

for any $\alpha \subseteq \Gamma$, which includes the special case $\alpha = \Gamma$ of molecules with no sites bonded,

$$X_0 = \prod_{s \in \Gamma} X_s = \prod_{i=1}^{M} X_{\beta_i},$$

$$\forall\{\beta_1, \ldots, \beta_M\} \in \mathcal{P}(\Gamma), \ M \geq 1. \quad (59)$$

This independence property can be used to define $X_0 = X_{\Gamma - \gamma} X_{\gamma}$ in general which, when used in Equation (41), leads to

$$X_0 = \sum_{\{\gamma_1, \ldots, \gamma_M\} \in \mathcal{P}(\Gamma), \ M \geq 2} (-1)^{M} (M - 2)! \prod_{i=1}^{M} \frac{X_{\Gamma - \gamma_i}}{X_0}$$

$$= \sum_{\{\gamma_1, \ldots, \gamma_M\} \in \mathcal{P}(\Gamma), \ M \geq 2} (-1)^{M} (M - 2)!,$$

(60)

Despite the rather complicated expression, the sum over the partition elements on the right hand side of Equation (60) is equal to $\mid \Gamma \mid - 1$ [57]. Substituting this result in Equation (56) leads to the the free energy of a system forming only chain-like aggregates given as

$$A_{\text{Wert}} = N k T \left[ \sum_{a=1}^{N_{\text{ST}}} s_a (\ln X_a - X_a + 1) \right]$$

$$- \frac{1}{2} \sum_{a=1}^{N_{\text{ST}}} \sum_{b=1}^{N_{\text{ST}}} s_a X_a s_b X_b \rho \Delta_{ab} \right],$$

(61)

By minimisation with respect to the density $\sigma_{\Gamma - a} = \rho X_a$, which is equivalent to minimisation with respect to fractions of molecules with a site of type $a$ free,

$$\frac{\partial A_{\text{Wert}}}{\partial X_a} \bigg|_{\{X_b \neq a, \tau, \gamma, N\}} = 0,$$

(62)

a system of equations is obtained determining the fractions of molecules not bonded at a site $a$,

$$X_a = \frac{1}{1 + \sum_{b=1}^{N_{\text{ST}}} s_b X_b \rho \Delta_{ab}},$$

(63)

collectively known as the mass action equations. In the ideal low-density limit, $X_a = 1$ for all $a \in \Gamma$, which
results in $\lim_{\rho \to 0} A_{\text{Wert}} = 0$. $A_{\text{Wert}}$ here is thus a residual contribution from association $A_{\text{assoc}}$ and the expected expression \cite{33,34,40} is recovered after substituting Equation (63) in Equation (61):

$$A_{\text{assoc}} = NK T \sum_{a=1}^{N_C} s_a \left( \ln X_a + \frac{1 - X_a}{2} \right),$$  
(64)

### 4.2.2. Open-chain and ring clusters

In the case of a system in which ring formation is accounted for, the site–site interactions are no longer independent and relations (58) and (59) are not valid. Instead, upon minimisation of the free energy in Equation (56) with respect to each $X_\delta$ for $\delta \subseteq \Gamma$, the fraction of molecules with the sites in set $\delta$ free is given by Wertheim \cite{28} as

$$X_\delta = X_0 \sum_{\psi \subseteq \Gamma - \delta} \frac{\sum_{(\gamma_1, \gamma_2, \ldots, \gamma_M) \in P(\psi), M \geq 1} \prod_{i=1}^{M} c_{\gamma_i}}{\Theta(\gamma)},$$  
(65)

where the first summation is over all subsets $\psi$ of the set $\Gamma - \delta$, including the empty set $\emptyset$ for which we follow the convention that $X_\emptyset = X_0$ for $\delta = \emptyset$. The second summation is over partitions of $\psi$ with elements $\gamma_i$. All the site fractions in Equation (65) are re-expressed in terms of the $c_{\gamma}$ functions, a new set of intensive variables, defined by Wertheim as

$$c_{\gamma} = \frac{\partial (\Delta c(0)/N)}{\partial X_\gamma}. \tag{66}$$

Accordingly, for $|\gamma| = 1$,

$$c_a = \sum_{s' \in \Gamma - s} X_{s'} \rho \Delta_{ss'}, \tag{67}$$

from Equation (42) in terms of sites or, equivalently,

$$c_a = \sum_{b=1}^{N_C} s_b X_b \rho \Delta_{ab}, \tag{68}$$

in terms of site types. For $|\gamma| = 2$, using Equation (50),

$$c_{ss'} = \sum_{R=1}^{N_{RS,ss'}} (X_{ss'})^{R-1} (\rho \Delta_{ss'})^{TR} \frac{W_{ss',R}}{\rho}, \tag{69}$$

and using Equation (53) and (66), which allows one to account for multiplicity of sites of the same type,

$$c_{ab} = \sum_{R=1}^{N_{RS,ab}} \left( \sum_{s' \in \Gamma - ab} X_{ss'} \right)^{R-1} (\rho \Delta_{ab})^{TR} \frac{W_{ab,R}}{\rho}, \tag{70}$$

in the case of $ab$ being a ring-forming pair of sites (i.e., $W_{ab,R} > 0$), which can now be rewritten in terms of site types as

$$c_{ab} = \sum_{R=1}^{N_{RS,ab}} (|o_{ab}| X_{ab})^{R-1} (\rho \Delta_{ab})^{TR} \frac{W_{ab,R}}{\rho}. \tag{71}$$

If $a$ is not involved in ring formation, then $W_{ab} = 0$ for all $b \in \Gamma$ and all ring sizes, and $c_{ab} = 0$. Furthermore, for any $|\gamma| \geq 3$, $c_{\gamma} = 0$, as neither the open-chain graphs (Equation (42)) nor the ring graphs (Equation (50)) depend on $X_\gamma$ for $|\gamma| \geq 3$. Applying these constraints allows one to rewrite Equation (65) in a simpler form cancelling all $c-$functions except $c_a$ and $c_{ab}$. Thus,

$$\frac{X_\delta}{X_0} = \sum_{(\gamma_1, \gamma_2, \ldots, \gamma_M) \in P(\Gamma - \delta), M \geq 1} \prod_{j=1}^{M} \Theta(\gamma), \tag{72}$$

where $M$ is the number of elements in a given partition of $\Gamma - \delta$ and

$$\Theta(\gamma) = \begin{cases} 1 + c_a & \text{if } \gamma = \{s\}, s \in \Gamma \\ c_{ab} & \text{if } \gamma = \{s, s'\}, (s, s') \in \Gamma, \end{cases} \tag{73}$$

where $a$ is the type of site $s$ and $b$ is the type of site $s'$. The closed system of Equations (67)–(73) provides a route to calculate the fractions of molecules with any subset (of $\Gamma$) of sites free. However, on inspection of Equation (27) we note that the evaluation of the residual Helmholtz free energy requires only $X_0, X_a(= X_s)$, and $X_{ab}(= X_{ss'})$, for $s, s' \in \Gamma$. Noting that $X_0 = 1$, the law of mass action to calculate $X_0$ is obtained for $\delta = \emptyset$ in Equation (72) as

$$(X_0)^{-1} = \sum_{(\gamma_1, \gamma_2, \ldots, \gamma_M) \in P(\Gamma - \emptyset), M \geq 1} \prod_{j=1}^{M} \Theta(\gamma). \tag{74}$$

In a similar way, setting $\delta = ab$ in Equation (72), the fraction of molecules with the arbitrary pair $ab$ free is obtained so that

$$\frac{X_{ab}}{X_0} = \sum_{(\gamma_1, \gamma_2, \ldots, \gamma_M) \in P(\Gamma - ab), M \geq 1} \prod_{j=1}^{M} \Theta(\gamma). \tag{75}$$

An expression explicit in $X_a/X_0$ can be obtained following an analogous procedure but an alternative simpler expression can be derived by expanding the sum in Equation (74). In any particular partition of $\Gamma$ an arbitrary site $s$ appears either in an element $\gamma = \{s\}$ or in an element $\gamma = \{s, s'\}$. The sum over the partitions of the set...
of association sites \( \Gamma \) can thus be separated into two disjoint sums: the sums where site \( s \) appears in terms of the form \( (1 + c_s) \); and the sums where site \( s \) appears in terms of the form \( c_{ss'} \):

\[
(X_0)^{-1} = (1 + c_s) \times \sum_{s' \in \Gamma}^{M} \prod_{j=1}^{\Theta(y_j)} \left( \frac{y_j}{(\Gamma-s',M \geq 1)} \right) + \sum_{s' \in \Gamma}^{M} c_{ss'} \times \prod_{j=1}^{\Theta(y_j)},
\]

These two sums can be translated into quotients of the molecular fractions over the monomer fractions using Equation (72), so that Equation (76) is rewritten as

\[
(X_0)^{-1} = (1 + c_s) \frac{X_s}{X_0} + \sum_{s' \in \Gamma}^{M} c_{ss'} \frac{X_{ss'}}{X_0},
\]

and equivalently in terms of site types,

\[
(X_0)^{-1} = (1 + c_s) \frac{X_s}{X_0} + (s_a - 1) c_{aa} \frac{X_{aa}}{X_0} + \sum_{b=1,b \neq a}^{N_{ST}} s_b c_{ab} \frac{X_{ab}}{X_0},
\]

which can be multiplied by \( X_0 \) to obtain

\[
X_a = \frac{1 - (s_a - 1) c_{aa} X_{aa} - \sum_{b=1,b \neq a}^{N_{ST}} s_b c_{ab} X_{ab}}{1 + c_a}.
\]

The mass action equations (Equation (74)–(79)) can then be substituted in Equation (56) to obtain the association contribution to the Helmholtz free energy as

\[
A_{\text{Wert}} = NkT \left[ \ln X_0 + \frac{1}{2} \sum_{a=1}^{N_{ST}} s_a (1 - X_a) \right. \\
\left. - \frac{1}{2} \sum_{a=1}^{N_{ST}} \sum_{b=1}^{N_{ST}} s_{ab} c_{ab} X_{ab} \right] \frac{\Delta_{ii,ab}}{\rho}.
\]

or more compactly as

\[
A_{\text{Wert}} = NkT \left[ \ln X_0 + \frac{1}{2} \sum_{a=1}^{N_{ST}} s_a (1 - X_a) \right. \\
\left. - \sum_{R=1}^{N_{RS,ab}} \frac{1}{\tau_R} \times \xi_{\rho R} \right],
\]

which is reassuringly in agreement with the expression presented in Section 3 (once the contribution from the reference system of non-bonded molecules \( A_0 \) is discounted from Equation (27)).

5. Residual free energy of association for mixtures

Equation (80) can be straightforwardly written for a mixture of \( N_C \) associating components, as

\[
A_{\text{Wert}} = NkT \sum_{i=1}^{NC} x_i \ln X_{i,0} + \frac{1}{2} \sum_{a=1}^{N_{ST,i}} s_{ia} \left( 1 - X_{i,a} \right) \\
- \frac{1}{2} \sum_{a=1}^{N_{ST,i}} \sum_{b=1}^{N_{ST,i}} \frac{1}{\tau_R} \left( \rho \Delta_{i,ab} \right)_{\rho R} \times \frac{W_{i,ab, R}}{\rho},
\]

where the subscript \( i \) refers to each component, and \( x_i \) is the mole fraction.

In the ideal low-density limit this expression retains the chain connectivity contribution, so that in order to obtain the corresponding residual free energy of association \( A^{\text{assoc}} \), contributions remaining in the ideal limit (\( \rho \rightarrow 0 \)) need to be subtracted:

\[
A^{\text{assoc}} = A_{\text{Wert}} - \lim_{\rho \rightarrow 0} A_{\text{Wert}},
\]

with

\[
\lim_{\rho \rightarrow 0} A_{\text{Wert}} = NkT \left[ \sum_{i=1}^{NC} x_i \ln X_{i,0}^{\text{ideal}} + \sum_{a=1}^{N_{ST,i}} s_{ia} X_{i,a}^{\text{ideal}} \right. \\
\left. - \sum_{a=1}^{N_{ST,i}} \sum_{b=1}^{N_{ST,i}} \frac{1}{\tau_R} \left( \rho \Delta_{i,ab} \right)_{\rho R} W_{i,ab, \text{intra}} \right],
\]

where the superscript ‘ideal’ is shorthand for the low density limit (\( \lim_{\rho \rightarrow 0} \)) of the given variable. \( W_{i,ab, \text{intra}} \) refers to the likelihood of an intramolecular hydrogen bond forming between sites of type \( a \) and \( b \) in component \( i \). Intermolecular rings are not present in the ideal limit.
The fraction of molecules with sites in set $\delta$ free in the mixture is given as
\[
X_{i,\delta} = X_{i,0} \sum_{\gamma_1, \ldots, \gamma_M \in \mathcal{P}(\Gamma_i - \delta), M \geq 1} \prod_{j=1}^{M} \Theta_i(\gamma_j),
\]
(85)
with $\Theta_i(\gamma)$ defined by
\[
\Theta_i(\gamma) = \begin{cases} 
1 + c_{i,a} & \text{if } \gamma = \{s\}, s \in \Gamma_i \\
0 & \text{if } \gamma = \{s, s'\}, (s, s') \in \Gamma_i \end{cases}
\]
(86)
where $a$ is the type of site $s$, $b$ is the type of site $s'$,
\[
c_{i,a} = \rho \sum_{j=1}^{N_c} \sum_{b=1}^{N_S_{ST}} s_{i,b} X_{i,b} \Delta_{i,ab},
\]
(87)
and
\[
c_{i,ab} = \sum_{R=1}^{N_{RS,ab}} \left[ |o_p_{ab}| X_{i,ab} \right]^{R - 1} \left( \rho \Delta_{i,ab} \right)^R \frac{W_{i,ab,R}}{\rho}.
\]
(88)
Here, $c_{i,ab}$ is non-zero only if sites $a$ and $b$, both of the same species $i$, constitute a ring/intramolecular bonding pair, i.e., if $W_{i,ab,R} > 0$. The expressions for $X_{i,0}, X_{i,a}$, and $X_{i,ab}$ required in Equation (82) for the case of a mixture are given as
\[
(X_{i,0})^{-1} = \sum_{\gamma_1, \ldots, \gamma_M \in \mathcal{P}(\Gamma_i), M \geq 1} \prod_{j=1}^{M} \Theta_i(\gamma_j),
\]
(89)
\[
X_{i,ab} = X_{i,0} \sum_{\gamma_1, \ldots, \gamma_M \in \mathcal{P}(\Gamma_i - \delta), M \geq 1} \prod_{j=1}^{M} \Theta_i(\gamma_j),
\]
(90)
and
\[
X_{i,a} = \frac{1 - (s_{i,a} - 1)c_{i,a,a} X_{i,a} - \sum_{b=1, b \neq a}^{N_{ST,i}} s_{i,b} c_{i,ab} X_{i,ab}}{1 + c_{i,a}}.
\]
(91)
The ideal limit of Equation (85), from which the limits of Equation (89) and (90) follow naturally, is given as
\[
X_{i,\delta}^{\text{ideal}} = X_{i,0} \sum_{\gamma_1, \ldots, \gamma_M \in \mathcal{P}(\Gamma_i - \delta), M \geq 1} \prod_{j=1}^{M} \Theta_i^{\text{ideal}}(\gamma_j),
\]
(92)
with
\[
\Theta_i^{\text{ideal}}(\gamma) = \begin{cases} 
1 & \text{if } \gamma = \{s\}, s \in \Gamma_i \\
0 & \text{if } \gamma = \{s, s'\}, (s, s') \in \Gamma_i
\end{cases}
\]
(93)
where $a$ is the type of site $s$ and $b$ is the type of site $s'$,
\[
c_{i,a}^{\text{ideal}} = 0,
\]
(94)
and
\[
c_{i,ab}^{\text{ideal}} = \Delta_{i,ab}^{\text{ideal}} W_{i,ab,\text{intra}}.
\]
(95)
Lastly, the ideal limit of Equation (91) is given by
\[
X_{i,a}^{\text{ideal}} = 1 - (s_{i,a} - 1)c_{i,a,a}^{\text{ideal}} X_{i,a}^{\text{ideal}} - \sum_{b=1, b \neq a}^{N_{ST,i}} s_{i,b} c_{i,ab}^{\text{ideal}} X_{i,ab}^{\text{ideal}},
\]
(96)
which completes the set of expressions needed to calculate the residual association contribution for the Helmholtz free energy of the mixture, accounting for ring formation. Four specific cases of this general theory are presented in the Appendix. In these specific cases, where the types of rings allowed are restricted, we are able to avoid the use of partitions and the resulting equations are significantly simpler, easing the implementation of the theory.

In the following section we develop our new approach with the SAFT-VR Mie equation of state [69] and apply it to specific model systems.

### 6. Intramolecular association in SAFT-VR Mie

In order to study the impact of intramolecular association on the phase behaviour of fluids we examine chain systems incorporating intramolecular bonding, including the addition of a second associating component that competes for the sites that mediate the intramolecular association. Calculations are carried out using the SAFT-VR Mie equation of state [69], in which segment-segment interactions are treated via Mie potentials of general repulsive and attractive range (characterised by $\lambda_r$ and $\lambda_a$, respectively), and embed association sites to mediate hydrogen bonding, accounting for inter- and intramolecular association following the expressions presented in the previous sections.

#### 6.1. Models

Compound 1 is modelled as an associating chain fluid comprising $m$ tangentially bonded Lennard-Jones ($\lambda_r = 12, \lambda_a = 6$) spherical segments with two association sites of different types labelled $A$ (of type $a$) and $B$ (of type $b$, $b \neq a$), in which $ab$ association promotes the formation of open chain-like aggregates and intramolecular association (intermolecular ring formation is however not accounted for); no $aa$ or $bb$ bonding is allowed.
Table 1. Like and unlike interaction parameters for the model systems, where \( i, j \) are the component indexes (1 for the chain and 2 for the sphere), \( \sigma_{ij} \) is the segment diameter, \( \lambda_{ij}^r \) and \( \lambda_{ij}^a \) are the repulsive and attractive exponents of the Mie potential, respectively, and \( \varepsilon_{ij} \) is the depth of the potential well.

| \( \sigma_{ij}/\sigma_{11} \) | \( \lambda_{ij}^r \) | \( \lambda_{ij}^a \) | \( \varepsilon_{ij}/\varepsilon_{11} \) | \( \varepsilon_{ij}^{HB}/\varepsilon_{11} \) | \( \varepsilon_{ij}^{kin}/\varepsilon_{11} \) |
|---|---|---|---|---|---|
| 11 | 1.00 | 12.0 | 6.00 | 1.00 | 10.0 | 1.00 |
| 12 | 0.95 | 12.0 | 6.00 | 1.00 | 10.0 | 1.00 |
| 22 | 0.90 | 12.0 | 6.00 | 2.00 | -- | -- |

Notes: The association interaction is characterised by a square-well potential of depth \( \varepsilon_{ij}^{assoc} \) and the bonding volume \( K_{ij}^{HB}/\sigma_{ij}^3 \). Combining rules [69] are used for the unlike parameters.

for this compound. Compound 2 is modelled as a single Lennard-Jones sphere containing only one association site of type \( a \), which can only bond to sites of type \( b \), since \( aa \) association (dimerisation) is not permitted. In the binary mixture, the two components may associate through \( AB \) site–site bonding, component 2 hence competes with the formation of \( AB \) bonds in component 1.

The extent of intramolecular association is characterised by the value of \( W_1,AB,\text{intra} \). The expression of Treloar for the end-to-end distribution function at contact for a freely-jointed chain [77,78],

\[
W_n = \frac{n(n-1)}{8\pi \sigma^3} \sum_{j=0}^{k} \frac{(-1)^j}{j!(n-j)!} \left[ \frac{n-1-2j}{2} \right]^{n-2}, \tag{97}
\]

where \( k \) is the smallest integer satisfying

\[
k \geq \frac{n-1}{2} - 1, \tag{98}
\]

and \( n = m-1 \) is the number of links in a chain, was first used by Sear and Jackson [74] and later shown to be useful in the prediction of phase behaviour of HF [59]. Due to the steric constraints present in a chain of interacting segments, however, we do not expect Equation (97) to be a good approximation of the end-to-end probability density in our model system. Instead, we treat \( W_1,AB,\text{intra} \) as an additional model parameter. A range for the reduced parameter \( W_n \), defined in terms of the corresponding \( W_n \), is used in the calculations.

The like and unlike model parameters for both components are presented in Table 1. The unlike values for the segment size \( c_{12} \) and dispersion energy \( \varepsilon_{12} \) are obtained from standard combining rules [69].

### 6.2. SAFT-VR Mie equation of state and association contribution

In the SAFT-VR Mie equation of state the Helmholtz free energy of a mixture of associating chain fluids is written as the sum of four separate contributions [69]

\[
A = A^{\text{ideal}} + A^{\text{mono}} + A^{\text{chain}} + A^{\text{assoc}}, \tag{99}
\]

where \( A^{\text{ideal}} [72] \) is the free energy of the ideal mixture and all other terms are residual contributions. \( A^{\text{mono}} \) accounts for the change in free energy due to monomer-monomer Mie repulsive and dispersion interactions, obtained following a third-order high-temperature expansion [69]; \( A^{\text{chain}} \) accounts for the residual free energy due to the formation of chains from the segments of the reference fluid [34,69]; and \( A^{\text{assoc}} \) is the residual contribution due to hydrogen bonding [79].

The association term in the original formulation of SAFT-VR Mie EOS is given by the original TPT1 theory of Wertheim, i.e., Equation (64) in our current work. Accounting now for the formation of intramolecular bonds, the residual association contribution \( A^{\text{assoc}} \) of the model system is obtained following from Equation (82):

\[
A_{\text{Wert}} = NkT \left[ X_{1,0} + \frac{1}{2} (1 - X_{1,A}) \right. \\
+ \left. \frac{1}{2} (1 - X_{1,B}) - X_{1,0} \Delta_{1,AB} W_{1,AB,\text{intra}} \right] \\
+ NkT \left[ \ln X_{2,0} + \frac{1}{2} (1 - X_{2,0}) \right], \tag{100}
\]

where the mass action equations are given as

\[
\frac{1}{X_{1,0}} = (1 + c_{1,A})(1 + c_{1,B}) + c_{1,AB}, \tag{101}
\]

\[
\frac{X_{1,A}}{X_{1,0}} = (1 + c_{1,B}), \tag{102}
\]

and

\[
\frac{X_{1,B}}{X_{1,0}} = (1 + c_{1,A}), \tag{103}
\]

for component 1, and

\[
\frac{1}{X_{2,0}} = (1 + c_{2,A}), \tag{104}
\]

for component 2, and where

\[
c_{1,A} = \rho X_{1,0} \Delta_{11,AB}, \tag{105}
\]

\[
c_{1,B} = \rho \left( X_{1,1,B} \Delta_{11,AB} + X_{2,0} \Delta_{12,AB} \right), \tag{106}
\]

\[
c_{2,A} = \rho X_{1,1,B} \Delta_{12,AB}, \tag{107}
\]

and

\[
c_{1,AB} = X_{1,AB,\text{intra}} \Delta_{11,AB} \frac{W_{1,AB,\text{intra}}}{\rho}. \tag{108}
\]
The ideal low-density contribution is obtained as

$$\lim_{\rho \to 0} A_{\text{Wert}} = N k T x_{1,0} \left[ \ln x_{1,0}^{\text{ideal}} + \frac{1}{2} (1 - x_{1,0}^{\text{ideal}}) \right] + \frac{1}{2} (1 - x_{1,0}^{\text{ideal}})$$

$$- x_{1,0}^{\text{ideal}} + 1 + \frac{\alpha_{1,AB} W_{1,AB,\text{intra}}}{X_{1,0}^{\text{ideal}}} \right].$$

In the low-density limit $x_{2,0}^{\text{ideal}} = 1$, while

$$x_{1,0}^{\text{ideal}} = x_{1,0} = 1 - c_{1,AB}^{\text{ideal}} x_{1,0},$$

and

$$\frac{1}{X_{1,0}^{\text{ideal}}} = 1 + \epsilon_{1,AB}^{\text{ideal}}.$$

where

$$c_{1,AB}^{\text{ideal}} = \frac{\Delta_{11,AB} W_{1,AB,\text{intra}}}{X_{1,0}^{\text{ideal}}}.$$

In the SAFT-VR Mie approach the integrated association parameter is expressed in a factored form as [69]

$$\Delta_{ij,AB} = F_{ij,AB} \epsilon_{ij,AB} K_{ij,AB} I_{ij},$$

where $F_{ij,AB} = \exp(\beta \epsilon_{ij,AB}^{\text{HB}}) - 1$, $\beta = 1/(kT)$, $K_{ij,AB}$ is a bonding volume parameter, and $I_{ij}$ is a temperature-density correlation of the association kernel given by [79,80]

$$I_{ij} = \sum_{p=0}^{10} \sum_{q=0}^{10-p} d_{pq} \left( \rho \sigma_x^3 \right)^{\rho} \left[ \frac{kT}{\epsilon_{ij}} \right]^q.$$

The coefficients $d_{pq}$ are given in reference [79], a van der Waals one-fluid mixing rule is used for the size parameter $\sigma_x^3$ (cf. equation (25) of reference [81]), while a two-fluid mixing rule is used for the energy parameter $\epsilon_{ij}$ (cf. equation (26) in reference [81]). The number density of segments is given by $\rho = \rho \sum_{i=1}^{NC} (x_i m_i)$. In the ideal limit

$$I_{ij}^{\text{ideal}} = \sum_{q=0}^{10} d_{pq} \left[ \frac{kT}{\epsilon_{ij}} \right]^q,$$

since the only terms of the polynomial that do not vanish at low density are the ones corresponding to $p = 0$.

### 6.3. Results

Thermodynamic properties and phase behaviour can be obtained from the Helmholtz free energy through standard relations and solving the phase equilibrium conditions of equality of temperature $T$, pressure $P$, and chemical potential $\mu_i$ of each component $i$ across phases. We use reduced variables throughout, with the temperature given as $T^* = kT/\epsilon_{11}$, the volume as $V^* = V/\sigma_{11}^3$, and the pressure as $P^* = P/\sigma_{11}^3$. We start by studying pure-compound model systems of chains with $m = 5$ segments and interaction model parameters corresponding to component 1 in Table 1 with varying degrees and types of association: a non-associating system (where $\epsilon_{11,ab}/\epsilon_{11} = 0$); an associating system in the absence of intramolecular interaction ($W^* = 0$); and associating systems with intramolecular association with $W^* = 10, 100, \text{and } 10^7$. The reduced parameter $W^*$ is defined in terms of the end-to-end function for a chain of five segments; i.e., $W^* = W_{1,AB,\text{intra}}/W_4$ (four links). The large value of $W^* = 10^7$ is chosen as representative of a model in the limit in which all molecules exhibit intramolecular association (equivalent to covalently bonded ring-like molecules).

In Figure 5, the vapour–liquid coexistence boundaries and enthalpy of vaporisation for the chain fluids comprising five segments are presented. When comparing the calculations for the non-associating and associating chain fluids, an increase in the critical temperature and pressure is clearly seen for all associating systems including the system without intramolecular association. As can be seen in the figure, accounting for intramolecular interactions in the fluids leads to a further increase of the critical temperature and pressure and to a clear decrease in the vapour pressure at high temperatures.

The low-temperature region of the vapour-pressure curve is better appreciated in a Clausius–Clapeyron representation (Figure 5(b)). Here, for temperatures $T^*/T^* > 0.5$, the vapour pressures of the systems with intramolecular association are found to be close to those of the non-associating system, and are visibly larger than those of the system in which only intermolecular association is taken into account. This behaviour suggests intramolecular association is favoured at low temperatures, which results in less intermolecular association as only one bond per pair of association sites is allowed. As the formation of an intramolecular bond does not affect the number of species in the fluid, the density of the vapour and liquid phase in these conditions can be expected to be close to those of the non-associating system, leading to vapour pressures close to those of the non-associating system. This effect is confirmed by examining the temperature–density phase diagram.

The coexistence densities (in terms of packing fractions) of the liquid and vapour phases can be seen in Figure 5(c). The increase of the critical temperature of the fluids with intramolecular association is clearly seen in this figure, together with slightly wider
Figure 5. Fluid-phase coexistence properties of two-site associating chain fluids (comprising $m = 5$ segments) in different associating schemes: (a) vapour pressure; (b) Clausius–Clapeyron representation; (c) saturated densities with inset of saturated volumes; and (d) vaporisation enthalpy. The reduced pressure is defined by $P^* = P \sigma_{11}^3/\varepsilon_{11}$, the reduced temperature by $T^* = kT/\varepsilon_{11}$, the packing fraction by $\eta = (\pi/6)\rho m \sigma_{11}^3$, the reduced volume by $V^* = V/\sigma_{11}^3$, and the reduced enthalpy by $\Delta H_{\text{vap}}^* = \Delta H_{\text{vap}}/(NkT)$. The dashed black curves correspond to calculations for a non-associating fluid, the continuous black curves to an associating fluid with $W^* = 0$, the continuous green curves to $W^* = 10$, the continuous blue curves to $W^* = 100$, and the continuous pink curves to $W^* = 10^7$.

Envelopes for increasing intramolecular association in the high-temperature region of the phase diagram. At low temperature, the densities of the saturated liquid are very close for all of the associating systems (and close to those of the non-associating fluid). The impact of incorporating intramolecular association is more clearly seen by examining the saturated volume (inset Figure 5(c)), where the larger volume (smaller packing fraction) of the saturated gas phase for the system without intramolecular association can be seen, while the saturated volumes of the fluids with intramolecular association are seen to be close to that of the non-associating system.

Consistent trends are observed (Figure 5(d)) in the calculated enthalpy of vaporisation. At low temperatures ($T^* < 2$), the systems with intramolecular association exhibit values of the enthalpy close to that of the non-associating system and markedly lower than those of the system with intermolecular association only. The trend is reversed in the high-temperature region partly driven by the higher critical temperature of the systems with intramolecular interactions.

In order to gain further understanding of the impact of intramolecular association on the fluid-phase behaviour of our model systems we analyse the effect of temperature and packing fraction on the distribution of aggregates. In Figure 6, we present calculations for three of the associating model fluids: associating chains in the absence of intramolecular bonding ($W^* = 0$), and associating chains with intramolecular association characterised by $W^* = 10$ and $W^* = 100$. Low-density states are represented in Figures 6(a–c), high-density states in Figures 6(d–f), and variable density at a constant temperature of $T^* = 2.0$ in Figures 6(g–i). Given the proposed association model (two sites labelled A and B, with only AB bonding allowed), molecules may be in one of three bonding states: free monomer, bonded in a linear chain aggregate, or intramolecularly bonded. In the figures, each of the fractions are represented by the width of the corresponding area. In order to appreciate better the change in the numbers of aggregate species for changing thermodynamic conditions, a broad temperature/density range is presented, including states within the two-phase envelope. On inspection it can be clearly seen that a decrease in temperature or an increase in packing fraction results in an increase of the fraction of associated molecules (given by the sum of inter- and intramolecular associated molecules), as can be expected.
Moreover, regardless of the value of $W^* > 0$ or packing fraction, at sufficiently low temperatures, the fractions of both monomers and molecules with intermolecular bonds are expected to approach zero and all molecules to exhibit an intramolecular bond, since such a state maximises the number of site–site interactions in the fluid. In any open-chain aggregate two sites remain unbonded, while in intramolecular hydrogen-bonded configurations, a limit with all of the sites bonded may be reached.

In the low-density states considered ($\eta = 0.007$) few molecules are found presenting intermolecular bonds; intermolecular bonding is appreciable only at low temperatures, and principally when intramolecular association is not considered $W^* = 0$ (Figures 6(a–c)). The extent of intramolecular association is dramatically reduced in the models with $W^* \neq 0$ due to the competition between inter- and intramolecular association. In the absence of intramolecular HB (Figure 6(a)), the percentage of molecules bonded is seen to reach close to 50% at the lowest temperature considered. By contrast, when intramolecular association is incorporated (Figures 6(b,c)) fewer than 0.01% of the molecules form part of open-chain aggregates, while almost all molecules present intramolecular association at the lowest temperatures. In the intermediate temperature region, a large proportion of the molecules are also found to present intramolecular HB when $W^* \neq 0$, and especially so in the case of $W^* = 100$.

For the liquid-like states at $\eta = 0.3$ (Figures 6(d–f)), a larger extent of association is seen throughout in comparison to the gas-like states, as a consequence of the higher density. The effect of incorporating intramolecular association is clearly appreciated when comparing the low temperature states. At $T^* = 1.1$, 96% of the molecules are found to be bonded in open chains in the fluid with $W^* = 0$, with only 4% for the system with $W^* = 10$, and virtually no molecules forming part of open-chain aggregates for $W^* = 100$, as the fluid in this case is dominated by molecules presenting intramolecular hydrogen bonds.

In the intermediate temperature region for the system with $W^* = 10$, the competition of intramolecular and intermolecular association is clearly seen, with an especially interesting feature in the calculated fraction of molecules, intermolecular hydrogen bonding being the maximum that can be seen at $T^* = 1.85$. At low temperatures, intramolecular bonding is favoured, even at this liquid-like density, as fewer sites remain free (unbonded) through intramolecular association than through intermolecular association, which is energetically favourable. This is particularly clear in the case of $W^* = 100$ where the fluid is seen to be dominated by molecules with intramolecular bonds throughout.

Similarly, an inspection of the different bonding states as a function of packing fraction at constant temperature $T^* = 2.0$ (Figure 6(g–i)) highlights the overall increase in the fraction of molecules bonded as the density increases. In terms of the ratio of molecules presenting inter- or intramolecular association, it is interesting to compare the highest packing fraction considered ($\eta^* = 0.5$), for which roughly two thirds of the molecules are found in open chains for $W^* = 0$, while only 31% are in open chains for $W^* = 10$, and fewer than 0.5% for $W^* = 100$. As $W^*$ is increased it drives the formation of intramolecular interactions, and the fraction of molecules with intramolecular bonds increases over the entire density range.

Thus far we have considered systems with a chain length of $m = 5$. It is now interesting to consider the effect of molecular length on the phase equilibrium of the systems. Following from the models presented we have carried out fluid-phase equilibrium calculations for model chain fluids comprising $m = 5$, $m = 8$, and $m = 10$ segments, considering non-associating and associating models with $W^* = 0$ (no intramolecular HB), $W^* = 10$, and $W^* = 100$. For each chain length, the corresponding reference $W_n$ is used, so that $W^* = W/W_n$ as before, where $n = m - 1$ is the number of links in the chain.

As can be seen in Figure 7, the expected increase in the critical temperature is observed the chain length is increased [34]. The critical temperature and pressure are seen to increase with association and more so for an increase in the likelihood of the formation of intramolecular bonds (as given by larger values of $W^*$), for all chain lengths considered. Furthermore, we note that the larger the chain length, the less impact is seen in terms of the contribution of association to the fluid-phase behaviour. This can be understood in terms of the overall free energy of the systems. For the same association energy, bonding volume and $W^*$, the contribution to the total free energy due to covalent chain formation increases with increasing $m$, while the value of the residual contribution from association remains comparatively unchanged. Hence, the differences in the phase behaviour between the fluids with and without ring formation are less significant for larger chains.

The corresponding fractions of molecules in each of the possible bonding states are plotted in Figure 8 along the saturated liquid and vapour phase boundaries. The fraction of molecules non-bonded at any site $X_0$ markedly increases for increasing temperature and decreasing values of $W^*$ in both the liquid and vapour phases, mirroring the findings of Figure 6. The fraction
of molecules associated in linear clusters is higher when intramolecular association is not considered, as can be seen in Figure 8(b), even though the fraction of non-bonded molecules is higher in these systems. In each of the systems studied, larger fractions of non-bonded molecules are seen in the gas phases, where the density is low and intramolecular association is predominant, as seen in Figure 8(b,c).

Increasing the values of $W^*$ directly promotes intramolecular bonding as seen in Figure 8(c) (and previously in Figure 6), which competes with intermolecular bonding. In Figure 8(b) the fraction of molecules involved in open chain aggregates is seen to decrease with increasing chain length. This is partly due to the molecular density of the saturated-liquid phase being lower for longer chains (Figure 7(c)), and partly due to the value of $W$ decreasing with increasing chain length, according to Equation (97). However, this trend inverts with increasing values of $W^*$, where the fraction of molecules in open-chain aggregates is found to be smaller for the shorter chains as a result of more intramolecular association occurring for the systems with shorter chain length.

Interestingly, the saturated-vapour and liquid curves cross each other for the systems with $W^*=10$ but not for the systems with $W^*=100$ (Figure 8(c)). A higher fraction of intramolecular-bonded molecules is found in the vapour-phase for temperatures below the crossing point. At low temperatures there are more rings in the vapour than in the liquid phase because in the vapour-phase ring formation does not compete with intermolecular hydrogen bonding as in the case of the liquid phase. At $W^*=100$, the crossing of the vapour and liquid curves is no longer observed because the fraction of open-chain aggregates is insignificant in both phases.

Lastly, we consider a binary mixture of the associating chain molecules with $m=5$ (component 1) and a spherical component (component 2) which contains one association site of type $a$ which can bond to a site of type $b$ of component 1 only; no $aa$ association (dimerisation) is allowed in component 2. The parameters of the segment-segment interactions are given in Table 1.
In Figure 9 we present the phase diagrams, in which vapour–liquid (VLE), liquid–liquid (LLE) equilibria, and vapour–liquid–liquid equilibria (VLLE) can be seen, together with corresponding distributions of bonding states of the chain component at a fixed pressure of $P^* = 0.002$. In this model mixture, the chain component can be found in one of four bonding states: both sites non-bonded (we refer to this state as ‘free’); bonded to a spherical molecule through site B, with site A of the chain free or associated to other chains (referred to as ‘cross-associated’); bonded intermolecularly to at least one other chain molecule but not to component 2 (‘self-associated’); or intramolecularly bonded if $W^* > 0$ (‘intramolecular’).

Examining the phase behaviour in Figure 9(a) it is noticeable that including cross association leads to a narrowing of the LLE region and to higher saturation temperatures of the bubble and dew curves (following the higher saturation temperature of the associating chain fluid). The overall fraction of molecules of the chain component (component 1) that are found to be bonded increases with decreasing temperature, in both the vapour and the liquid phases. Virtually no cross-association between the two components is seen in the vapour phase or in L1, in which the mole fraction of chain molecules is very small. In liquid phase L2, which is rich in chain molecules, a large fraction of cross-associated molecules is seen, which increases with the enrichment of the solution in spherical molecules.

With regards to the formation of intramolecular bonds, in the mixture with associating chains with $W^* = 10$ (Figure 9(b)), a broader LLE region is seen compared to the case of the associating system with no intramolecular bonding. This is because fewer molecules of component 1 participate in cross-association with component 2 when intramolecular bonding occurs. This can be easily confirmed by studying the corresponding bonded states of the chain molecules. The overall fraction of bonded chain molecules in the L1 and L2 phase boundaries for the mixture with chains with $W^* = 10$ is only slightly higher than in the previous case of the chains with no intramolecular association, but with the difference that now a large proportion of the molecules are seen to present intramolecular bonding.

The most noticeable change is seen along the vapour phase boundary. In the case of $W^* = 0$ the fraction of molecules bonded is non-negligible only at lower temperatures, while in the case with $W^* = 10$ between 50%

---

**Figure 7.** Vapour–liquid equilibria of pure chain fluids with $m = 5$, $m = 8$, and $m = 10$ segments with interaction parameters given in Table 1: (a) vapour pressures; (b) Clausius–Clapeyron representation; (c) saturated densities; and (d) vapourisation enthalpies. The dashed curves correspond to non-associating systems, the continuous curves to systems with intermolecular association only, the long-dashed curves to those with intramolecular association with $W^* = 10$, and the dash-dotted curves to those with intramolecular association with $W^* = 100$. 
Figure 8. Temperature $T^*$ vs. (a) fraction of monomers $X_0$, (b) fraction $f_{\text{chains}}$ of molecules in open-chain aggregates, and (c) fraction $\xi_1$ of molecules in intramolecular rings, along the saturated-vapour and liquid states for chain fluids of $m = 5$, $m = 8$, and $m = 10$ segments. The continuous curves correspond to models with $W^* = 0$, the dash-dotted curves to $W^* = 10$, and the long-dashed curves to $W^* = 100$.

(at higher temperature) and 90% (at lower temperature) of the chain molecules are seen to be intramolecularly bonded. A further feature worth highlighting refers to the L2 branch, in which the fraction of self-associated chains can be seen to remain essentially constant as the temperature is decreased, while for the case of the mixture with $W^* = 0$ it can be seen to increase (Figure 9(a), right-hand panel).

Considering the mixture with chains characterised by $W^* = 100$ (Figure 9(c)), similar trends as for the system with $W^* = 10$ are seen, although these are of course more pronounced. The LLE boundaries and the bubble curve are found close to those of the mixture with non-associating chains, since with more intramolecular association in the system, less intermolecular and cross association occurs. The overall fraction of associated chains is found to be the largest of the three mixtures considered, in all phases, as this corresponds to the largest value of $W^*$.

7. Final remarks

The theory presented in our current work enables the modelling of chain molecules of $m \geq 1$ tangentially bonded spherical segments with embedded short-range association sites that mediate hydrogen-bond type interactions and result in the formation of closed aggregates in addition to open linear and branched aggregates. The intermolecular as well as intramolecular formation of ring-like aggregates are accounted for in the theory.

The formalism followed is that of Wertheim with the standard TPT1 steric restrictions of two sites per bond, one bond per site, and no double bonding between two segments. Wertheim’s approximation neglecting closed aggregates is relaxed to allow for the contribution to the Helmholtz free energy of the formation of intramolecular bonds and intermolecular ring aggregates to be accounted for.

After imposing steric restrictions, an expression for the Helmholtz free energy of the distribution of bonding states is derived based on the aggregate count following a pedagogical approach. The mass action equations determining the distribution of bonding states at equilibrium are found by minimisation of Wertheim’s expression for the Helmholtz free energy modified to include an extended fundamental graph sum to account for the closed aggregates. The theory presented is thus an extended TPT1 to treat ring-like aggregates. Two new model parameters are required per ring type: the ring size $\tau$ (i.e., number of molecules in a ring, with $\tau = 1$ for an intramolecular ring) and a parameter $W$ that captures the probability of two segments of the same molecule or of the same chain-like aggregate being within the appropriate bonding distance. As a first approximation, $W$ is considered to be independent of density and temperature and to be specific for a pair of sites types $ab$ in a given component $i$ (single-component rings) capable of
Figure 9. Fluid-phase equilibria and distribution of bonded states along the phase boundaries for a binary mixture of two-site (A and B) associating chains of \( m = 5 \) segments (component 1) and spheres with one association site A (component 2), which can bond to site B in component 1 only (no AA association is allowed) at a constant pressure \( P^* = 0.002 \). The phase boundaries are labelled as V for the saturated vapour, and L1 and L2 for the saturated liquid phases rich in component 1 and 2, respectively. The black dashed curves in (a), (b), and (c) correspond to calculations for non-associating chains, the black continuous curves to associating chains with \( W^* = 0 \), and the coloured curves in (b) and (c) to associating chains with \( W^* = 10 \) and \( W^* = 100 \), respectively. The fractions of chain molecules in the possible bonding states along the saturation curves are represented in the middle and right-hand side panels. For a given temperature, the fraction of cross-associated molecules of component 1 is represented by the width of the dark-colour solid area, the fraction of self-associated molecules by the width of the light-colour solid area, the fraction of chains with intramolecular HB by the width of the square-patterned area, and the fraction of non-bonded (free) chains by the width of the white area. (a) Associating fluid in the absence of ring formation (\( W^* = 0 \)); (b) associating fluid with \( W^* = 10 \); and (c) associating fluid with \( W^* = 100 \).

Associating to form ring-like clusters of a given size \( \tau_R \):

\[ W_{i,ab,\tau_R} \]

The free energy is presented as a residual contribution of the total Helmholtz free energy so it can be implemented in equations of state of the SAFT (and CPA) family. In particular, we have presented calculations using the SAFT-VR Mie equation of state for chains with two (AB) sites that promote both intramolecular hydrogen bonding and intermolecular association. Different chain lengths and different degrees of intramolecular association as determined by the value of \( W \) are studied and the competition between different bonding states upon the introduction of a one-site associating spherical component is also examined.

Larger values of \( W \) promote intramolecular bonding, lead to higher critical temperatures and pressures in the pure chain fluids and to higher vapour pressures at low temperatures as compared to the case of associating chains with no intramolecular bonding. Association is promoted by the decrease of temperature and at sufficiently low temperatures intramolecular bonded molecules are seen to dominate in the fluids since the number of bonds in the fluid is maximised in this configuration and
hence these states are the most energetically favourable.

At intermediate temperatures the competition between inter- and intramolecular hydrogen bonding is discussed through the analysis of the distribution of bonding states. The extent of both inter- and intramolecular association decreases with the decrease of packing fraction. This effect is naturally more pronounced in terms of the intermolecular aggregates than in terms of the intramolecular association.

The extension of the association contribution presented requires the introduction of new physically-meaningful parameters since the types of rings to be considered in the system need to be specified, with a ring type characterised by its size $\tau \leq 1$ ($\tau = 1$ for an intramolecular bond) and its likelihood of formation $W$. For each ring size, the corresponding $W$ can in principle be estimated from data derived from experiment or molecular simulation. Application of this theory to experimental systems of chemical and pharmaceutical interest will be the subject of future publications.

### Sites nomenclature

- $\alpha, \delta, \gamma, \psi$: subset of association sites
- $\Gamma$: set of association sites
- $a, b, c, \ldots$: association site types
- $s, s', s''$, $\ldots$: individual (generic) association sites
- $A, B, C, \ldots$: specific (labelled) association sites
- $N_{\text{ST}}$: number of site types
- $o_{ab}$: set of $ab$ pairs of association site ($ba$ pairs not included)
- $n_{a}^{\alpha}$: number of sites $a$ in set $\alpha$

### Notes

1. Note that $X_{\emptyset}$ does not correspond to $X_{0}$ because formally, $X_{\emptyset}$ represents the ‘fraction of molecules with no sites free (at least)’, which includes all molecules. The quantity $X_{\emptyset}$ is not seen in the expressions because the fraction of all molecules without restrictions always equals one.
2. Graph theory and lemmas for manipulations of graphs can be found in [72,73].
3. The value for the pair distribution function is typically taken at contact value since the association interactions are short ranged.
4. In Equation (53), the summation $\left(\sum_{\alpha' \in o_{ab}} X_{\alpha'}\right)$ can be written in terms of site types as $|o_{ab}|X_{ab}$. We are however not carrying out this change at this stage as this expression needs to be differentiated with respect to $X_{ad}$, and not site type, in the next section.
5. As an example, take the set of all sites $\Gamma = \{A_{1}, A_{2}, B\}$. In this model we would have $o_{AB} = \{A_{1}B, A_{2}B\}$, $o_{AA} = \{A_{1}A_{2}, A_{2}A_{1}\}$ and $o_{BB} = \emptyset$.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

S.A.F. is grateful to the Department of Chemical Engineering of Imperial College London for the award of a PhD Scholarship. The authors acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC) of the UK (Grants EP/E016340, EP/P006965 and EP/J014958) for support for the Molecular Systems Engineering group. C.S.A. is thankful to EPSRC for the award of a Leadership Fellowship (EP/J003840) and A.G. is thankful to the Royal Academy of Engineering and Lilly for support of a Research Chair (RCSRFP18193). It is a great pleasure to be able to contribute to this Special Issue in honour of Peter Cummings, a wonderful friend and mentor.

### ORCID

A. Aasen [http://orcid.org/0000-0002-7436-5825]
C. S. Adjiman [http://orcid.org/0000-0002-4573-7722]
G. Jackson [http://orcid.org/0000-0002-8029-8868]
A. Galindo [http://orcid.org/0000-0002-4902-4156]

### References

[1] E. Arunan, G.R. Desiraju, R.A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D.C. Clary, R.H. Crabtree, J.J. Dannenberg, P. Hobza, H.G. Kjaergaard, A.C. Legon, B. Mennucci and D.J. Nesbitt, Pure Appl. Chem. 83 (8), 1637–1641 (2011).
[2] Y. Maréchal, The Hydrogen Bond and the Water Molecule (Elsevier, Amsterdam, 2007), the nether edition.
[3] N.G. Almarza, Phys. Rev. E 86 (3), 030101 (2012).
[4] P.S. Burton, R.A. Conradi, N.F.H. Ho, A.R. Hilgers and R.T. Borchardt, J. Pharm. Sci. 85 (12), 1336–1340 (1996).
[5] U. Domańska, Fluid Phase Equilib. 55 (1), 125–145 (1990).
[6] U. Domańska, Ind. Eng. Chem. Res. 29 (3), 470–475 (1990).
[7] C.M. Dobson, Nature 426 (6968), 884–890 (2003).
[8] C.N. Pace, B.A. Shirley, M. McNutt and K. Gaijwala, FASEB J. 10 (1), 75–83 (1996).
[9] R.E. Hubbard and M.K. Haider, Hydrogen Bonds in Proteins: Role and Strength, 2010.
[10] A.N. Semenov, A.N. Semenov, I.A. Nyrkova and M.E. Cates, Macromolecules 28 (23), 7879–7885 (1995); I.A. Nyrkova, and M.E. Cates. Macromolecules, 28 (23), 7879–7885 (1995).
[11] M. Rubinstein and A.V. Dobrynin, Curr. Opin. Colloid Interface Sci. 4 (1), 83–87 (1999).
[12] P.J.M. Stals, M.A.J. Gillissen, R. Nicolaÿ, A.R.A. Palmans and E.W. Meijer, Polym. Chem. 4 (8), 2584–2597 (2013).
[13] C.J. Gregg, F.P. Stein and M. Radosz, Macromolecules 27 (18), 4972–4980 (1994).
[14] C.J. Gregg, F.P. Stein and M. Radosz, Macromolecules 27 (18), 4981–4985 (1994).
[15] F. Dolezalek, Zeitschrift für Phys. Chemie 64 (1), 727 (1908).
[16] H.C. Andersen, J. Chem. Phys. 59 (9), 4714–4725 (1973).
[17] H.C. Andersen, J. Chem. Phys. 61 (12), 4985–4992 (1974).
Appendix. Detailed expressions for selected models

Recall that the residual association contribution to the total free energy in our framework is obtained as

\[
A_{\text{assoc}} = A_{\text{Wert}} - \lim_{\rho \to 0} A_{\text{Wert}}. \tag{A1}
\]

Here we present the equations corresponding to specific associating model systems. These specific cases allow for a system of equations that does not involve the use of the partition notation and is therefore more straightforward to implement.

Case I – Pure fluid of chain molecules with two sites A and two sites B; i.e., \( \Gamma = \{ A, B \} \). The formation of intermolecular rings and intramolecular hydrogen bonding is promoted by AB association, in addition the formation of open-chain aggregates is promoted by AA, BB, and AB association. The corresponding contribution to the Helmholtz free energy due to association can be developed as follows:

\[
A_{\text{Wert}} = NkT \left[ \ln X_0 + 2 - X_A - X_B \right.
\]

\[- \frac{1}{2} \sum_{R=1}^{N_{\text{RS,AB}}} \frac{1}{\tau_R} \left( 2X_{\text{AA}}\rho \Delta_{\text{AA}} \right)_{\text{tr}} \frac{W_{\text{AB,R}}}{\rho} \right], \tag{A2}
\]

\[
\lim_{\rho \to 0} A_{\text{Wert}} = NkT \left[ \ln X_0^{\text{ideal}} + 2 - X_A^{\text{ideal}} - X_B^{\text{ideal}} \right.
\]

\[- X_A^{\text{ideal}} \Delta_{\text{AA}} W_{\text{AA,intra}} \right], \tag{A3}
\]

\[
(X_0)^{-1} = (1 + c_A)^2 (1 + c_B)^2
\]

\[- + 4c_A (1 + c_A) (1 + c_B) + 2 (c_A)^2, \tag{A4}
\]

\[
X_{\text{AB}} = X_0 \left[ (1 + c_A) (1 + c_B) + c_A \right], \tag{A5}
\]

\[
X_A = \frac{1 - 2c_A X_{\text{AB}}}{1 + c_A}, \quad X_B = \frac{1 - 2c_B X_{\text{AB}}}{1 + c_B}, \tag{A6}
\]

\[
(X_0^{\text{ideal}})^{-1} = 1 + 4c_A^{\text{ideal}} + 2 (c_A^{\text{ideal}})^2, \tag{A7}
\]

\[
X_A^{\text{ideal}} = X_0^{\text{ideal}} \left[ 1 + c_A^{\text{ideal}} \right], \tag{A8}
\]

\[
X_A^{\text{ideal}} = 1 - 2c_A^{\text{ideal}} X_A^{\text{ideal}}, \quad X_B^{\text{ideal}} = 1 - 2c_B^{\text{ideal}} X_B^{\text{ideal}} \tag{A9}
\]

\[
c_A = 2\rho X_A \Delta_{\text{AA}} + 2\rho X_B \Delta_{\text{AB}},
\]

\[
c_B = 2\rho X_A \Delta_{\text{AB}} + 2\rho X_B \Delta_{\text{BB}}, \tag{A10}
\]

\[
c_{\text{AB}} = \sum_{R=1}^{N_{\text{RS,AB}}} (4X_{\text{AB}})_{\text{tr}}^{-1} \left( \rho \Delta_{\text{AB}} \right)_{\text{tr}} \frac{W_{\text{AB,R}}}{\rho}, \tag{A11}
\]

\[
c_{\text{ideal}} = \Delta_{\text{AB}} W_{\text{AB,intra}}. \tag{A12}
\]

Case II – Pure fluid of chain molecules with two sites A and two sites B; i.e., \( \Gamma = \{ A, A, B, B \} \). The formation of intermolecular rings and intramolecular hydrogen bonding is promoted by AA association, in addition the formation of open-chain aggregates is promoted by AA, BB, and AB association. The corresponding contribution to the Helmholtz free energy due to association can be developed as follows:

\[
A_{\text{Wert}} = NkT \left[ \ln X_0 + 2 - X_A - X_B \right.
\]

\[- \frac{1}{2} \sum_{R=1}^{N_{\text{RS,AA}}} \frac{1}{\tau_R} \left( 2X_{\text{AA}}\rho \Delta_{\text{AA}} \right)_{\text{tr}} \frac{W_{\text{AA,R}}}{\rho} \right], \tag{A13}
\]

\[
\lim_{\rho \to 0} A_{\text{Wert}} = NkT \left[ \ln X_0^{\text{ideal}} + 2 - X_A^{\text{ideal}} - X_B^{\text{ideal}} \right.
\]

\[- X_A^{\text{ideal}} \Delta_{\text{AA}} W_{\text{AA,intra}} \right], \tag{A14}
\]

\[(X_0)^{-1} = (1 + c_A)^2 (1 + c_B)^2 + 4c_A (1 + c_A) (1 + c_B) + 2 (c_A)^2, \tag{A15}
\]

\[
X_{\text{AA}} = X_0 (1 + c_B)^2, \tag{A16}
\]

\[
X_A = \frac{1 - c_A X_{\text{AA}}}{1 + c_A}, \quad X_B = \frac{1}{1 + c_B}, \tag{A17}
\]

\[
(X_0^{\text{ideal}})^{-1} = 1 + c_A^{\text{ideal}}, \tag{A18}
\]

\[
X_A^{\text{ideal}} = X_0^{\text{ideal}} \left[ 1 + c_A^{\text{ideal}} \right], \tag{A19}
\]

\[
X_A^{\text{ideal}} = 1 - 2c_A^{\text{ideal}} X_A^{\text{ideal}}, \quad X_B^{\text{ideal}} = 1 - 2c_B^{\text{ideal}} X_B^{\text{ideal}} \tag{A20}
\]

\[
c_A = 2\rho X_A \Delta_{\text{AA}} + 2\rho X_B \Delta_{\text{AB}},
\]

\[
c_B = 2\rho X_A \Delta_{\text{AB}} + 2\rho X_B \Delta_{\text{BB}}, \tag{A21}
\]

\[
c_{\text{AB}} = \sum_{R=1}^{N_{\text{RS,AB}}} (2X_{\text{AA}})_{\text{tr}}^{-1} \left( \rho \Delta_{\text{AA}} \right)_{\text{tr}} \frac{W_{\text{AB,R}}}{\rho}, \tag{A22}
\]

\[
c_{\text{ideal}} = \Delta_{\text{AB}} W_{\text{AB,intra}}. \tag{A23}
\]

Case III – Fluid mixture of \( N_C \) components. Each component \( i \) consists of chain molecules with any given set of sites \( \Gamma_i \). The formation of intermolecular rings and intramolecular hydrogen bonding is promoted by AB association, where each ring is only allowed to involve a single component. In addition formation of open-chain aggregates is promoted by any pair-combination of sites in \( \Gamma \), including AA, BB, AB, CC, AC, BC, . . . Here, \( n_{a_i} \) is used to represent the number of sites of type \( a \) in set \( \alpha_i \) in component \( i \). The corresponding contribution to the Helmholtz free energy due to association can be developed as follows:

\[
A_{\text{Wert}} = NkT \sum_{i=1}^{N_C} \sum_{a=1}^{N_{\text{RS}}} \left[ \ln X_{i,0} + \frac{1}{2} \sum_{a=1}^{N_{\text{RS}}} n_{a_i} (1 - X_{i,a}) \right.
\]

\[- \frac{1}{2} \sum_{a=1}^{N_{\text{RS}}} n_{a_i} (1 - X_{i,a}) \right] - s_{a_i} s_{B_i} X_{AB}^{\text{ideal}} \Delta_{\text{L}AB} W_{\text{L}AB,\text{intra}} \right], \tag{A25}
\]
Case IV – Fluid mixture of $N_C$ components. Each component $i$ consists of chain molecules with any given set of sites $I_i$. Formation of intermolecular rings and intramolecular hydrogen bonding is promoted by AA association, where each ring is only allowed to involve a single component. In addition formation of open-chain aggregates is promoted by any pair combination of $N_{RS,AA}$, including AA, BB, AB, CC, AC, BC, ... . Here, $n_i^n$ is used to represent the number of sites of type $a$ in set $n$, in component $i$. The corresponding contribution to the Helmholtz free energy due to association can be developed as follows:

$$A_{\text{Wert}} = NkT \sum_{i=1}^{N_C} \left[ \ln X_{i,0} + \frac{1}{2} \sum_{a=1}^{N_{ST,i}} s_{ia}(1 - X_{i,0}) \right]$$

$$- \frac{1}{2} \sum_{R=1}^{N_{RS,AA}} \left( s_A(s_A - 1)X_{i,AA} \rho \Delta_{i,AA} \right)^{r_R}$$

$$\times \frac{W_{i,AA,R}}{\rho},$$

$$\lim_{\rho \rightarrow 0} A_{\text{Wert}} = NkT \sum_{i=1}^{N_C} \left[ \ln X_{i,ideal} + \frac{1}{2} \sum_{a=1}^{N_{ST,i}} s_{ia}(1 - X_{i,ideal}) \right]$$

$$- \frac{1}{2} s_A(s_A - 1)X_{i,AA}^{\text{ideal}} \Delta_{i,AA} W_{i,AA,\text{intra}},$$

$$\lim_{\rho \rightarrow 0} A_{\text{Wert}} = NkT \sum_{i=1}^{N_C} \left[ \ln X_{i,ideal} + \frac{1}{2} \sum_{a=1}^{N_{ST,i}} s_{ia}(1 - X_{i,ideal}) \right]$$

$$- \frac{1}{2} s_A(s_A - 1)X_{i,AA}^{\text{ideal}} \Delta_{i,AA} W_{i,AA,\text{intra}}.$$