

# Recent Advances in Industrial Coarse-Grained Method Development and Modeling

J.-W. Handgraaf<sup>1</sup> and J.G.E.M. Fraaije<sup>1,2</sup>

<sup>1</sup>Culgi B.V., P.O. Box 252, 2300 AG Leiden, The Netherlands, janwillem.handgraaf@culgi.com

<sup>2</sup>Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

## ABSTRACT

In present day the crux of performing coarse-grained simulations is the actual parameterization of the (molecular) interactions and the validation of the outcome of a simulation. In a recent effort we have work on the former and have devised a data-driven scheme, termed Automated Fragmentation and Parameterization (AFP) that uses Monte Carlo based fragmentation techniques in combination with excluded volume and free energy information of molecular pairs to obtain coarse-grained parameter sets. In this work we show in some detail how AFP could be employed and validated in an industrial setting by computing the wet octanol/water partitioning of a small set of industrial relevant fragrance molecules.

**Keywords:** Coarse-Graining, Multiscale Modeling, Parameterization, Simulation

## 1 Automated Fragmentation and Parameterization Method

The Automated Fragmentation and Parameterization (AFP) method has been proposed by Fraaije et al.[1] as a robust and fast method to obtain coarse-grained parameter sets for large set of molecules in a semi-automated fashion. For full details of AFP and a discussion on the accuracy we refer to the paper cited above. Here we very briefly introduce the method.

Molecules are fragmented, or "mapped", according to a scoring function, through a simulation annealing function that cuts through bonds; the optimal bond fission pattern is preserved and the fragments are stored. The scoring function is defined as:

$$S = \left(1 - \frac{V}{V_0}\right)^2 \quad (1)$$

with  $V$  the volume of the fragment and  $V_0$  the volume of a cluster of water molecules in its lowest energy conformer (i.e. the reference volume). In this approach we use the molecule-unique fragmentation in order to preserve as much as possible of the properties of the reference molecule. This means that the fragments are not database-unique, as is customary in coarse-grained simulations, but completely specific to a given molecule.

A key parameter is the mapping factor,  $N_m$ , which sets the (average) number of heavy atoms per coarse-grained particle, typically ranging between three to five heavy atoms. Previous versions of the protocol has been successfully applied to a diverse set of molecular systems, ranging from amorphous polymer systems [2] to surfactant brine/oil interfacial systems [3]. Very recently, the AFP method was applied to compute diffusion coefficients of small organics and drug-like molecules in water [4].

The thermodynamic calibration is through a semi-empirical rule for the  $a_{ij}$  interaction parameter using:

$$a_{ij} = \alpha_{EV} \nu_i \nu_j \alpha_{res} \sqrt{\nu_i \nu_j} \beta \Delta G_{res,ij} \quad (2)$$

with  $\nu_i = V_i/V_0$  the scaled volume of fragment  $i$ ,  $\beta = 1/k_B T$  and  $\Delta G_{res,ij}$  the excess Gibbs energy of mixing of two (molecular) fragments  $i$  and  $j$ . Note that the  $a_{ij}$  interaction, or repulsion parameter was originally introduced by Groot and Warren in their seminal work [5]. Very briefly, the Gibbs free energy of mixing is calculated through COSMO-RS calculations [6, 7], using the charge envelope of the fragments (the so-called sigma profiles). The charge profile is obtained from Density Functional Theory quantum-chemical calculations using NWChem, version 6.3 [8], from geometry-optimized molecules in a vacuum, using the BP86 [9, 10] exchange-correlation functional and the def2-TZVPD basis set. The AFP model contains only two adjustable parameters, i.e.  $\alpha_{EV}$  based on excluded volume and  $\alpha_{res}$  that captures the residual contribution from the free energy of mixing. Full details can be found in Fraaije et al. [1]. All simulations were performed with the Culgi software package [11].

## 2 Results

To validate the AFP method we computed the partitioning coefficient between wet octanol and water ( $\text{Log}P$ ) for a set of twenty industrial relevant fragrance molecules, see Table 1 below. In order to compute  $\text{Log}P$  for our purposes we can use:

$$\log P = \log [\exp (\ln \gamma_W - \ln \gamma_O) V_W / W_O] \quad (3)$$

Table 1: Partitioning coefficient between wet octanol and water ( $\text{Log}P$ ) of the set of fragrance molecules. Mole fraction of water in the wet octanol phase is set to the experimental value of 0.24. In order to compute the partitioning we have use the Reference Interaction Site Model (RISM) solvent model as implemented in the Culgi software. AFP-RISM-Qeq is a variant where we derive the COSMO charge envelope directly from the charge equilibration (Qeq) method [12]. For full details see Fraaije et al. [1] and the Culgi manual [13]. The  $\text{Log}P$  values computed with the ChemAxon software suite [14] are based on a group-contribution method and take as input the SMILES of the respective molecule.

Fragrance Molecule	AFP-RISM	AFP-RISM-Qeq	ChemAxon
<b>ambro</b> nat	<b>3.64</b>	<b>3.82</b>	<b>3.98</b>
dimethyl-ph-eth-carbinol	1.89	2.09	2.64
eucalyptol	1.93	2.28	2.35
<b>flor</b> hydral	<b>2.03</b>	<b>2.20</b>	<b>3.14</b>
gamma-decalactone	1.91	2.79	2.87
<b>hiv</b> ernal-1	<b>2.59</b>	<b>2.71</b>	<b>3.49</b>
hivernal-2	2.67	1.81	3.49
hydroxycitronellal	0.73	0.92	1.64
iso-amyl-butyrate	2.71	2.74	2.68
lilial	2.73	0.99	3.98
linalyl-iso-butyrate	4.81	4.63	4.33
methyl-beta-naphthyl-ketone	1.51	3.36	2.52
nerolin	2.20	2.37	2.81
orcynyl-3	1.20	1.38	2.03
p-cresyl-methylether	1.89	2.16	2.33
para-hydroxy-phenyl-butanone	0.32	0.38	2.08
polysantol	2.74	3.23	3.67
pomarose-1	2.65	2.37	3.76
pyranol	1.88	2.21	1.39
tobacarol-1-and-2	3.17	3.68	3.97

where  $\gamma_W$  and  $\gamma_O$  are the activity coefficients in water and wet octanol and  $V_W$  and  $W_O$  are the COSMO volumes of wet octanol and water<sup>1</sup>. Figure 1 shows the molecular structures of three fragrance molecules and the corresponding mapped coarse-grained structures based on Eqn. 1 above. Figure 2 shows models of the atomic structures of the underlying fragments that correspond to the coarse-grained particles of ambronat. Note that the fragmentation is fully flexible in terms of number of atoms per fragments. As we indicated before, we opted for three heavy atoms per fragment to keep sufficient chemical specificity among the fragments. In addition, in this work we have decided to create new fragments for every fragrance molecule. However, one could re-use fragments to build up the coarse-grained parameter database, although in principal there is no need, as the fragmentation itself takes per molecule on average (much) less than 1 minute on one CPU.

Here we used the Reference Interaction Site Model (RISM) solvent model, in particular the 1-D, hypernetted chain (HNC) variant, see ref [1] and references therein. Using HNC-RISM speeds up the parameterization process, thereby opening up the way to generate and validate on the fly parameter sets for molecular databases containing tens of thousands of components. In Table 1 we have collected the computed  $\text{Log}P$  values for the fragrance molecule set, highlighting in bold ambronat, florhydrat and hivernal-1. Note first of all that subtle differences, such as difference between stereo isomers in the case of hivernal-1 and hivernal-2, is not captured by the group-based method of ChemAxon, whereas for AFP we do see a difference between the two stereoisomers. In general, the  $\text{log}P$  values computed with AFP are lower than the values computed with ChemAxon, with the Qeq variant in closer agreement. There are some clear outliers between AFP and ChemAxon, in particular para-hydroxyphenyl-butanone (i.e. raspberry ketone) that is according to AFP much more hydrophilic, possibly caused by the weak-acid character of the phenol moiety.

### 3 Concluding Remarks

The Automated Fragmentation and Parameterization (AFP) method has been applied to a small library of industrial relevant fragrance molecules. The obtained coarse-grained parameter set was in turn used to compute the partitioning coefficient between wet octanol and water. Overall, despite the loss in chemical detail due to the coarse-graining, AFP is able to capture the solubility of the fragrance molecules in wet octanol and water correctly to a large extend. A major potential of AFP is the possibility of fast screening of large molecu-

<sup>1</sup>A coarse-grained water particle is described by a cluster of three water molecules in its lowest energy conformer. This gives an additional correction to  $\text{Log}P$  of  $\ln(3) = 1.0986$ .

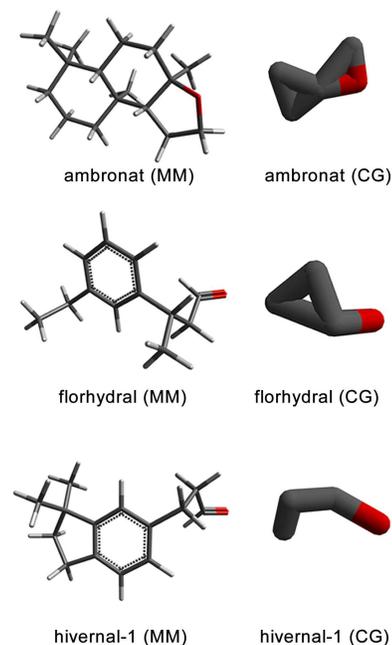


Figure 1: Automated fragmentation in action: Molecular models (MM) and corresponding coarse-grained (CG) models of ambronat, florhydrat and hivernal-1.

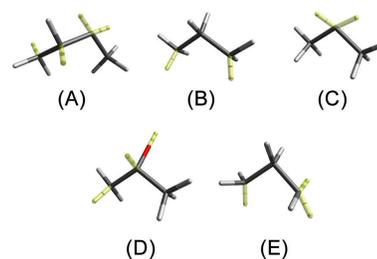


Figure 2: The five atomic fragments of ambronat. Yellow tubes indicate the connectors between the fragments.

lar databases, which is currently under investigation by us.

### REFERENCES

- [1] Fraaije, J. G. E. M.; van Male, J.; Becherer, P.; Serral Gracià, R. *J. Chem. Inf. Model.* **2016**, *56*, 2361-2377.
- [2] Handgraaf, J.-W.; Serral Gracià, R.; Nath, S. K.; Chen, Z.; Chou, S.-H.; Ross, R. B.; Schultz, N. E.; Fraaije, J. G. E. M. *Macromolecules* **2011**, *44*, 1053-1061.
- [3] Fraaije, J. G. E. M.; Tandon, K.; Jain, S.; Handgraaf, J.-W.; Buijse, M. *Langmuir* **2013**, *29*, 2136-2151.
- [4] Fraaije, J. G. E. M.; van Male, J.; Becherer, P.;

- Serral Gracià, R. *J. Chem. Theory Comput.* **2017**, *14*, 479-485.
- [5] Groot, R. D.; Warren, P. B. *J. Chem. Phys.* **1997**, *107*, 44234435.
- [6] Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224-2235.
- [7] Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, 2005.
- [8] Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; van Dam, T. P. S. H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.
- [9] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100.
- [10] Perdew, J. P. *Phys. Rev. B* **1988**, *33*, 8822-8824.
- [11] The Chemistry Unified Language Interface (Culgi), version 11.0, Culgi B.V., The Netherlands (2004-2017).
- [12] Rappe, A. K.; Goddard III, W. A. *J. Phys. Chem.* **1991**, *95* 33583363.
- [13] Johannes G.E.M. Fraaije, Shyamal K. Nath, Jan van Male, Paul Becherer, Joanne Klein Wolterink, Jan-Willem Handgraaf, Fiona Case, Catalin Tanase, Rubn Serral Graci, Culgi Scientific Manual, version 11.0, [www.culgi.com](http://www.culgi.com) (2017).
- [14] Calculator Plugins were used for structure property prediction and calculation, Marvin 17.2.27.0, 2017, ChemAxon (<http://www.chemaxon.com>)