1   Research Article

2 Revisiting a model to predict pure triglyceride
3 thermodynamic properties: parameter optimization and
4 performance

5   *Julia Seilert[1,3]*, Arun S. Moorthy[2], Anthony J. Kearsley[3], Eckhard Flöter[1]*

6   [1] Present Address: Department of Food Process Engineering, Technical University Berlin, Berlin,
7   Germany

8   [2] Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg,
9   MD, USA

10   [3] Applied and Computational Mathematics Division, National Institute of Standards and
11   Technology, Gaithersburg, MD, USA

12   ***Corresponding author:** Julia Seilert, Technical University Berlin, Department of Food Process
13   Engineering, Seestr. 13, Berlin, 13353, Germany; Email: julia.seilert@tu-berlin.de

14   **Running title:** Review of a model to predict TAG properties

21

2

Abstract

In 1990, a well-known model to predict pure component properties of triglycerides was presented by Wesdorp in "Liquid-multiple solid phase equilibria in fats: theory and experiments" and has been shown to perform well despite making thermodynamically inconsistent predictions for certain test cases. In this study, the underlying parameter set is improved to deliver more physically consistent predictions, i.e., increasing melting point and enthalpy of fusion with increasing stability of the polymorphs, without deterioration of the primary model quality to describe the available experimental data. Interestingly, when a curated dataset containing only thermodynamically consistent data is compared to a broader dataset, it appears that the model's efficacy is highly dependent on the quantity of data, specifically the number of unsaturated triglycerides data. Quality and thermodynamic consistency of model predictions and the condition of a reliable description of monoacid triglycerides as a subset is discussed, addressing a potential interdependence.

1. Introduction

Understanding phase behavior of fats remains important for many food applications. For example, the melting and solidification of fats in multi-component systems directly impacts product characteristics as melting range and solid-phase composition. The solid-liquid phase behavior and, thus, the product properties are determined by the behavior of triglycerides (TAGs) which are the major components of fats and oils (Flöter 2009). Different approaches to model the phase behavior of triglyceride mixtures were proposed to reduce extensive experimental studies. One approach to estimating the crystallization behavior of fats (TAG mixtures) is to utilize basic thermodynamical equilibrium calculations. This requires the availability of enthalpy change upon melting and melting point data of pure TAGs and mathematical models for non-ideal mixing in the solid phase. In this approach, significant data is required for a variety of pure TAGs and all polymorphic forms in which they occur. Acquiring experimental data for all polymorphs of all known TAGs is an ambitious if not impossible task. Thus, mathematical modeling provides an excellent alternative.

Several models have been introduced that predict thermodynamic properties of pure triglycerides (Zéberg-Mikkelsen and Stenby 1999; Wesdorp et al. 2013; Timms 1978; Hagemann and Rothfus 1983; Wesdorp 1990; Ollivon and Perron 1982) and for binary mixtures (Hjorth et al. 2015; Teles dos Santos et al. 2011; Coutinho et al. 2006; Coutinho et al. 2004). To date, the most comprehensive model for predicting the thermodynamic properties of pure components and binary systems was proposed by Wesdorp (Wesdorp 1990; Wesdorp et al. 2013). The pure component model was revisited in (Moorthy et al. 2017). It was observed that some model

56 predictions lacked thermodynamic consistency, i.e., strictly increasing enthalpy of fusion and

57 melting point with increasing stability of the polymorph within a triglyceride.

58 This study investigates model fit, performance, and predictive power when imposing

59 thermodynamic constraints on the predictions. Two approaches to ensure thermodynamic

60 consistency were evaluated. Firstly, a parameter set obtained by using a constrained optimization

61 procedure enforcing physically reasonable predictions was evaluated. Secondly, model

62 parameters were derived by optimization to a sub-dataset containing exclusively

63 thermodynamically sound data. In any case, the model fits were performed on a broad dataset that

64 contains data spanning nearly three decades and originating from various references leading to

65 variable data quality. The well-studied subset of monoacid saturated TAGs offers another

66 constraint, that is, prediction should possibly satisfy the evolution of the incremental enthalpy of

67 fusion and entropy of fusion per carbon. The effect satisfying this constraint has on the model

68 performance is evaluated. The model presented in (Wesdorp 1990; Wesdorp et al. 2013) is briefly

69 reviewed, and parameter sets derived from new optimization are described. The numerical results

70 are discussed with emphasis on the effects constraining has on the resulting parameters and

71 model performance.

2. Material and methods

2.1. <u>Mathematical model</u>

74 Wesdorp proposed a semi-empirical model for estimating the enthalpy of fusion and melting

75 points of saturated and unsaturated TAGs (Wesdorp 1990; Wesdorp et al. 2013; Moorthy et al.

76 2017). The thermodynamic properties are primarily defined as a function of total carbon number,

77 degree of saturation, and chain length differences. The enthalpy and entropy of fusion are posed

7 4

8

78   as linear combinations accounting for different sub-categories of TAGs, starting with monoacid

79   TAGs. The modifiers account for mixed-acid TAGs, odd-numbered TAGs, and unsaturated

80   TAGs. For mixed-acid TAGs, so-called chain length differences x and y were defined, where x

81   denotes the difference between the shortest outer chain length (P) and the middle chain length

82   (Q), and y is defined as the chain length difference between the two outer chain lengths (P and

83   R). A geometrical interpretation is depicted in Figure 1. The model equations, internal function

84   variables, and model parameters are summarized in Figure 2.

85   The model provides two distinct approaches for describing the coefficients A^{sat} and B^{sat} in eq. [11]

86   (Figure 2). The details of the rationale behind these two approaches can be found in (Wesdorp

87   1990). The first approach describes the coefficients solely as a quadratic function of the chain

88   length differences x and y, eq. [12] and eq. [13] in Figure 2. This approach is mentioned for

89   completeness and will not be pursued. The second route is inspired by the equilibrium condition

90   that the melting point temperature of a pure component equals the ratio of changes of enthalpy

91   and entropy on fusion, eq. [14] and eq. [15]. In contrast to the first approach, which introduced

92   additional parameters without any physical meaning, the parameters of the second approach

93   appear already in the computation of enthalpy of fusion eq. [1] and can be assigned to specific

94   contributions. Consequently, the second approach (eq. [14] and [15]) for estimating parameters

95   A^{sat} and B^{sat} in eq. [11] is considered in this work. Further, TAGs with odd carbon numbers are not

96   considered in this study due to the lack of reliable experimental data. Fatty acid moieties with odd

97   numbers of carbons in the alkyl chain do not naturally occur, which justifies omitting these

98   species in this work. As a result, the number of model parameters is reduced from 43 to 28 per

99   polymorph leading to a reduction from 129 to 84 parameters in the model.

9   5

10

100   **2.2. <u>Parameter identification</u>**

101   *2.2.1. Formulation of the optimization problem*

102 Model equations can be summarized as a function of the TAG for which the thermodynamic

103 properties are predicted. The parameter set $P^{\{k\}}$ denotes the respective polymorph k as input,

104 where $k = \{\alpha,\ \beta',\ \beta\}$ and contains 28 individual parameters, respectively. The 3-tuple TAG

105 contains the fatty acids composition (including the degree of saturation and carbon number per

106 chain) and the respective position of the fatty acid (FA) in the TAG i,

$$f_{WS}^{\{k\}}\left(TAG_i, P^{\{k\}}\right) = PRED_i^{\{k\}}. \tag{21}$$

107 The model output contains the melting point and enthalpy of fusion predictions,

108 $PRED_i^{\{k\}} = \left[T_i^{\{k\}}; \Delta H_i^{\{k\}}\right].$

109 The three functions, $f_{WS}^{\{\alpha\}}$, $f_{WS}^{\{\beta'\}}$, and $f_{WS}^{\{\beta\}}$ can be rewritten as a single function, f_{WS}, predicting the

110 enthalpy of fusion and the melting points of a TAG i for every polymorph simultaneously:

$$f_{WS}\left(TAG_i, P\right) = PRED_i; \tag{22}$$

111 where

112 $P = \left[P_j^{\{\alpha\}}; P_j^{\{\beta'\}}; P_j^{\{\beta\}}\right]$ for j = 1, …, 28

113 and

114 $PRED_i = \left[T_i^{\{\alpha\}}; \Delta H_i^{\{\alpha\}}; T_i^{\{\beta'\}}; \Delta H_i^{\{\beta'\}}; T_i^{\{\beta\}}; \Delta H_i^{\{\beta\}}\right].$

11   6

12

115 It should be noted that all physical properties are predicted simultaneously in this function but

116 that computations of parameter sets and experimental data per polymorph can be regarded as

117 isolated problems. More precisely, $\Delta H_i^{\{\alpha\}}$ is a function of $P^{\{\alpha\}}$ exclusively.

118 Employed here is the often-used output least-squares approach to fitting model parameters to

119 experimental data leading to the problem, find P that minimizes

$$J(P)=\sum\left(f_{WS}\left(TAG_i,P\right)-EXP_i\right)^2; \qquad\qquad [23]$$

120 where $\qquad EXP_i=\left[T_i^{\alpha};\Delta H_i^{\alpha};T_i^{\beta'};\Delta H_i^{\beta'};T_i^{\beta};\Delta H_i^{\beta}\right] \qquad$ for i = 1,…,N,

121 and N denotes the number of TAGs in the dataset.

122 Additionally, the simple *bound constraints*:

$$lb_j \leq P_j \leq ub_j \qquad\qquad [24]$$

123 are imposed where *lb* and *ub* are constant vectors of lower and upper bounds. The parameters *h*

124 and *s* were constrained by literature values. Again, the model approximates the enthalpy and

125 entropy of fusion by additive functions starting from monoacid TAGs. The entropy and enthalpy

126 data suggest a linear dependency from the carbon number where the parameters *h* and *s* represent

127 the slopes of a linear fit of the respective properties, Figure 3. These parameters can be

128 interpreted as incremental hydrocarbon chain contribution and solely depend on the alkyl chain

129 packing and, thus, the polymorphic form (Wesdorp 1990). Furthermore, the processing of the

130 experimental data on enthalpy and entropy of fusion also confirms that these values evolve

131 according to thermodynamically given constraints following the sequence $\alpha < \beta' < \beta$. Thus, these

132 slopes and their evolution can be considered as observations or "processed" experimental data.

13 14

133   Based on this, one could either choose to ignore this information for parameter optimization,

134   consider the slopes are derived as additional experimental data to preferentially be met, or

135   constrain the model parameters to meaningful ranges around the values based on observations.

136   Consequently, the parameters h and s were set to the values derived from experimental data

137   previously reported in Wesdorp (1990), namely 2.5 for $h^{\{\alpha\}}$, 6.5 for $s^{\{\alpha\}}$, 3.87 for $h^{\{\beta'\}}$, 9.8 for $s^{\{\beta'\}}$

138   and 4.2 and 10.5 for $h^{\{\beta\}}$ and $s^{\{\beta\}}$, with relative deviation constrained to be no more than 10 %.

139   Further, the contribution of odd-numbered TAGs is not considered separately. The respective

140   parameter h_{odd} is bounded by a small number representing floating-point zero. The parameter T_{inf}

141   is bounded by 380 K and 410 K, the melting range of polyethylene. Here polyethylene represents

142   an infinitely long hydrocarbon chain and also TAGs composed from such chains (Wesdorp et al.

143   2013). Remaining parameters are unconstrained. Minimizing eq. [23] subject to the bound

144   constraints eq. [24] is referred to as *Problem I* and the solution is denoted as P^I.

145   Explicitly, underlying thermodynamic fundamentals state strictly increasing melting points and

146   enthalpies of fusion with increasing stability of the polymorph:

$$T_i^\alpha < T_i^{\beta'} < T_i^\beta \quad \text{and} \quad \Delta H_i^\alpha < \Delta H_i^{\beta'} < \Delta H_i^\beta. \qquad [25]$$

147   Accordingly, to enforce consistency regarding the underlying thermodynamic fundamentals

148   during parameter fitting, a set of linear inequality constraints is applied. In the implementation,

149   the constraints in eq. [25] were replaced with less-than-or-equal formation:

$$c_i(P) \leq 0; \qquad [26]$$

150   where

151
$$c_i = \begin{cases} PRE\,D_i^{\alpha} - PRE\,D_i^{\beta'} \\ PRE\,D_i^{\beta'} - PRE\,D_i^{\beta} \\ PRE\,D_i^{\alpha} - PRE\,D_i^{\beta} \end{cases}$$

152 and were formulated in a way such that all predictions of TAGs are considered regardless of

153 available experimental data. Minimizing eq. [23] subject to eq. [24] and eq. [25] with the solution

154 P^{II} is referred to as *Problem II* in the following.

155 *Initial guess*

156 As with virtually all techniques for solving output least-squares problems, an initial guess P^0 is

157 required. Here a variant of the starting point given by (Moorthy et al. 2017) is used but modified

158 to ensure feasibility, in particular, that the starting point was in the interior of the feasible regions

159 determined by the respective constraints to *Problem I* and *Problem II*.

160 *Search algorithms*

161 Two different numerical algorithms are employed to identify optimal parameter sets P^I and P^{II} for

162 formulations *Problem I* and *Problem II*, respectively. The first algorithm used is a Sequential

163 Quadratic Programming (SQP) method which approximates the first derivatives of the objective

164 function and constraints and makes a quasi-Newton approximation to the Hessian of the

165 Lagrangian function. A description of SQP methods is beyond the scope of this paper, but

166 interested readers are referred to (Boggs et al. 1999a) and (Boggs et al. 1999b) for a description

167 of theoretical properties. In the spirit of comparison, a multi/parallel-dimensional search

168 algorithm is used (Torczon 1997), which uses no gradient information, calculated nor

169 approximated. While several variants of these methods are implemented across different software

170 platforms, MATLAB (2020) has efficient implementations of both classes of methods and was

171 used to generate the numerical results presented in this contribution.

2.2.2. Curation of dataset

173 Experimental data was gathered from public-domain literature (Zéberg-Mikkelsen and Stenby

174 1999; Ueno et al. 1997; Bayés-García et al. 2013b, 2015; Kodali et al. 1990; Hagemann and

175 Tallent 1972; Bayés-García et al. 2013a; Ghazani and Marangoni 2018; Baker et al. 2014a; Baker

176 et al. 2014b; Takeuchi et al. 2002; Bouzidi et al. 2010; Boodhoo et al. 2008; Lu et al. 2019;

177 Elisabettini et al. 1998; Wesdorp et al. 2013). The review revealed more melting point data than

178 enthalpy of fusion data. The accumulated data contained less data of unsaturated TAGs over a

179 wide range of carbon numbers compared to saturated TAGs. The resulting dataset D^L comprises

180 282 TAGs, from which 157 are saturated TAGs and 125 are unsaturated TAGs. The predictive

181 power of the model was assessed by performing a fit to a smaller dataset, D^C, and testing on the

182 greater dataset, here D^L. The strategy will be referred to as the train-and-test approach. A second

183 dataset was compiled by curating dataset D^L with a focus on a possible violation of underlying

184 thermodynamic constraints within one TAG. For reassessing the conditions given above, eq. [25],

185 at least two temperatures and two enthalpies of fusion must be prescribed. The resulting curated

186 dataset, D^C, consists of 134 saturated TAGs and 43 unsaturated TAGs. Details on both datasets

187 are given in Table 1. The experimental data are given in the *Supporting Information*.

188 The parameter estimation was conducted solving *Problem I* and *Problem II* for dataset D^L and

189 dataset D^C, respectively, resulting in four parameter sets. The thermodynamic consistency of the

190 model predictions was assessed for both datasets. Finally, an answer was sought as to whether

19

10

20

191 using a consistent (curated) dataset, D^C, for model fit is sufficient to achieve physically reliable

192 predictions or imposing thermodynamic constraints is required.

193 Additionally, the parameter estimation was conducted using only experimental melting point data

194 D^{MP} (a variant of D^L) as more of these than the enthalpy of fusion data is given in the literature.

195 The model equations for the enthalpy of fusion and melting point estimations for saturated TAGs,

196 eq. [1] and [11], are fully coupled. The model extension to unsaturated TAGs resulted in

197 incomplete coupling of eq. [18], eq. [19], and eq. [20], that is to say, not all parameters are

198 equally relevant to both model outputs. For example, the parameter h_O is necessary to estimate

199 the enthalpy of fusion for a TAG containing oleic acid but does not influence the melting point

200 estimation. Therefore, performing the parameter estimation using only melting point data is not

201 sufficient for estimating the thermodynamic properties of unsaturated TAGs, and the model fit

202 exclusively on melting point data was conducted for saturated TAGs only.

203 **2.3. <u>Numerical assessment strategy</u>**

204 *2.3.1. Prediction quality*

205 To assess the prediction quality of the model, evaluating measures were defined. Metrics were

206 computed for sub-categories of triglycerides: (1) saturated TAGs in α-polymorph, (2) saturated

207 TAGs in β'-polymorph, (3) saturated TAGs in β-polymorph and analogously for unsaturated

208 TAGs, resulting in 6 sub-categories of respective size N_{cat}, given Table 1. For the two predicted

209 quantities, the Root-Mean-Square Error (RMSE) was tabulated:

$$RMSE_l = \sqrt{\frac{\sum \left(l¿¿PRED - l_{\exp}\right)^2}{N_{cat}}} ¿; \qquad [27]$$

22

210 where l is either the melting point or enthalpy of fusion and N_{cat} is the size of the respective

211 category. The ratio of under predictions U to over predictions O was defined,

$$\frac{U}{O} = \frac{number\ of\ under\ predictions}{number\ \frac{of}{predictions}} \quad ; \tag{28}$$

212 where over predictions are those greater than the respective experimental data point, conversely,

213 under predictions are less than the experimental data point.

214 The thermodynamic consistency of the predictions was expressed by a "score". If the predictions

215 for a single TAG in all three possible modifications (α, β', β) were consistent, which means the

216 underlying thermodynamic constraints, eq. [24], are satisfied, the TAG was assigned a score of 1.

217 Conversely, if any constraint was violated, the TAG resulted in a score of 0. The overall score of

218 thermodynamic consistency was introduced,

$$TC = \frac{number\ of\ TAGs\ assigned\ score\ 1}{total\ number\ of\ TAGs \in category} * 100; \tag{29}$$

219 where, unlike measures for the prediction quality that require the experimental data, the score of

220 thermodynamic consistency could be determined for all TAGs in the dataset.

221 The described counting measures were tabulated for five parameter sets: the reference parameter

222 set taken from (Moorthy et al. 2017), parameter estimates derived in this work, $P^{I,L}$ and $P^{I,C}$ of

223 *Problem I*, and estimates $P^{II,L}$ and $P^{II,C}$ of *Problem II* for both datasets respectively. Further, the

224 prediction quality was assessed regarding the dataset used for fitting, i.e., D^L and D^C, following

225 the train-and-test approach.

23

24

226   *2.3.2. **Model sensitivity towards parameters***

227   Optimization schemes identify minima and hence provide optimal parameter values as a function

228   of data, but do not address model sensitivity directly. Therefore, a sensitivity analysis on the

229   model using a derivative-based method was performed to explore function behavior at solutions,

230   i.e., parameter sets obtained solving *Problem I* and *Problem II*, respectively. Since derivative-

231   based methods explore the parameter space locally, they only give qualitative estimates of model

232   sensitivity. For each parameter, a sensitivity coefficient Φ_j was estimated,

$$\Phi_j = \frac{\delta J}{\delta P_j} \frac{P_j}{J} \qquad [30]$$

233   as the partial derivative of the sum of squared residuals, eq. [23], with respect to the considered

234   parameter, eq. [30]. To normalize for the sake of comparability of the parameters, the quotient $\frac{P_j}{J}$

235   is included (Hamby 1994). The parameters were ordered by the magnitude of normalized partial

236   derivative indicating model sensitivity with respect to each parameter. The sensitivity coefficients

237   can be interpreted as a qualitative measure of sensitivity.

238   # 3. Results & Discussion

239   ## 3.1. <u>Prediction quality and thermodynamic consistency</u>

240   Model parameter fitting was performed with dataset D^L and D^C solving *Problem I* to estimate 84

241   total parameters, $P^{I,C}$ and $P^{I,L}$. Thermodynamic consistency of the melting point and enthalpy of

242   fusion predictions were satisfied through applying additional constraints in *Problem II* resulting

243   in the parameter estimates, $P^{II,C}$ and $P^{II,L}$. Data for each category used for parameter fitting are

244   given in Table 1 and the estimated parameter sets are given in Table 2. In summary, two

25   13

26

245   approaches for ensuring the thermodynamic consistency of the enthalpy of fusion and melting

246   point predictions were evaluated. First, an optimization problem *Problem II* with additional

247   thermodynamical constraints enforcing physically reasonable predictions was solved and then

248   compared to *Problem I*. Second, a curated fitting dataset D^C was generated containing only data

249   for which the underlying thermodynamic fundamentals, eq. [25], are satisfied. In this context, the

250   predictive power of the model was assessed. The quality of fit associated with predictions was

251   assessed using the measures described in section *Numerical assessment strategy*. The tabulated

252   values are given in Table 3.

253   Figure 4 illustrates predictions for the enthalpies of fusion and melting points for TAGs in the α-

254   polymorph using the reference parameter set and the parameter sets obtained in this work by

255   applying *Problem I* and *Problem II*. The model fit was performed on dataset D^L and the curated

256   dataset D^C. The melting points and enthalpies of fusion were computed for dataset D^L. The

257   display of the α-polymorph was chosen to highlight the differences between the predictions of

258   unsaturated TAGs as they were most evident. Similar results were obtained for the β' and β-

259   polymorph.

260   ### *3.1.1. Prediction quality*

261   A summary of the prediction quality measures is given in Table 3. A decreased RMSE indicates

262   an overall improvement of fit for the enthalpy of fusion of saturated and unsaturated TAGs for all

263   three polymorphs with respect to the reference parameter set when fitted on the dataset D^L. The

264   ratio of under predictions to over predictions U/O of 1.82 for P^I and P^{II} indicates an

265   underestimation of the enthalpy of fusion in β'-polymorph. The RMSE of the enthalpy of fusion

266   predictions of the α-polymorph of unsaturated TAGs relatively decreased by 18.66 % and

27   14

28

267   18.76 % for $P^{I,L}$ and $P^{II,L}$, respectively, in comparison to the reference. However, the ratio U/O of

268   0.5 suggests an underestimation of both quantities in the α-polymorph.

269   The RMSE of the melting point predictions of saturated TAGs in the β-polymorph relatively

270   increased by 20.13 % and 39.21 % for P^{I} and P^{II}, respectively. This is due to the exclusion of

271   fitting parameters addressing the lower melting points of odd TAGs s_{odd} and h_{odd}. However, the

272   ratio U/O of 1.19 and 0.96 for P^{I} and P^{II}, respectively, suggest a rather limited contribution from

273   odd-numbered TAGs on the parameter estimates. It should be noted that the dataset D^{L} contains

274   only 11 odd-numbered TAGs in β-polymorph of 157 saturated TAGs in total. The RMSE of

275   melting point predictions for unsaturated TAGs decreased relatively by 39.1 to 52.6 % for each

276   polymorph.

277   Using the adjusted parameters fit on the curated dataset D^{C} yielded comparable prediction quality

278   measures for both properties of saturated TAGs to the ones discussed above. This indicates an

279   overall improved fit independent from the amount of data used in the model fit procedure.

280   Similar results can be found for the enthalpy of fusion of unsaturated TAGs. However, a

281   worsening of the prediction quality for the melting point of unsaturated TAGs for all three

282   polymorphs is evident. Even when applying the thermodynamically constrained optimization

283   scheme of *Problem II,* no improvement of fit for unsaturated TAGs was achieved, suggesting the

284   curated dataset does not contain sufficient information on unsaturated TAGs despite satisfying

285   thermodynamic consistency.

286   Concluding, the prediction quality of saturated TAGs was neither affected by the applied

287   optimization scheme nor by the quality or size of the dataset used for fitting. This was expected

288   since the model equations for estimating the thermodynamic properties of saturated TAGs are

29

30

289   rather sophisticated and developed and improved by testing on a large dataset of 188 TAGs

290   comprising 92 experimental data of saturated TAGs in Wesdorp's work. Since the model

291   equations for estimating the thermodynamic properties of unsaturated TAGs are less developed,

292   they are subject to a strong dependence on the amount of data used for model fitting. This

293   became evident comparing the parameter fit on the large dataset D^L and the reduced dataset D^C.

294   Estimating the enthalpy of fusion for an unsaturated TAG requires only three additional

295   parameters to the model equation for saturated TAGs eq. [1], i.e., h_O, h_E, h_l. For the melting point

296   estimation, 17 new parameters are required (parameters 15 through 28 in Table 2). These address

297   not only the presence of oleic (O), elaidic (E), linoleic (l), and linolenic (le) acid but also the

298   interactions of pairs of unsaturated fatty acid chains in one triglyceride. These parameters are

299   only relevant to the enthalpy of fusion and melting point computations using eq. [18], eq. [11],

300   eq. [19], and eq. [20] when the respective internal function variables are non-zero. For example,

301   the parameters h_l, A_l and B_l are relevant to the estimated physical properties only if at least one of

302   the fatty acids in a TAG is linoleic acid and, hence, the internal function variable n_l is non-zero.

303   This emphasizes, for a fit of the full parameter set, the experimental dataset must contain at least

304   one TAG for each parameter. This requirement applies to all three polymorphs.

305   The coupling of model equations [1] and [11] for saturated TAGs was investigated by fitting the

306   parameters on the reduced dataset D^{MP} containing solely melting point data. As Table 4 reveals,

307   an improved fit of melting points is achieved as the RMSE indicate. In contrast to the melting

308   points, the enthalpy of fusion is poorly predicted. Consistent overestimation is expressed in

309   significantly increased RMSE and low U/O for the new dataset. Concluding, even reducing the

310   data set to saturated TAGs only, fitting the model parameters on melting point data only does not

311   result in a parameter set that predicts the enthalpy of fusion adequately.

31   16

32

3.1.2. Thermodynamic consistency

Unlike prediction quality measures (RMSE and U/O) which require the experimental data, the score of thermodynamic consistency can be determined for all TAGs in the dataset. Scores of melting points predictions for $P^{I,C}$, $P^{II,C}$, $P^{I,L}$, $P^{II,L}$ in comparison to the reference dataset are summarized in Table 5.

The thermodynamic consistency for the enthalpy of fusion was already safeguarded in the model using the reference parameters. However, melting point predictions using the reference set are not thermodynamically consistent (Moorthy et al. 2017). After refitting the model equations to the updated dataset D^L solving *Problem I*, the score of thermodynamic consistency of unsaturated TAGs increased from 37.6 to 48.8. Fitting on the sub-dataset D^C, curated for consistent data, but evaluating the score over the complete dataset D^L improved the score to a value of 69.6. However, in this case, the melting point predictions for unsaturated TAGs remained not satisfying the thermodynamic constraints, eq. [25], completely. This underpins the need for a constrained optimization scheme. Enforcing the physical constraints in *Problem II* yielded a significant increase of the thermodynamic score tested on D^L from 69.6 to 81.6 and 48.8 to 100 for parameter sets fit on dataset D^C and evaluated on D^L, respectively.

Inconsistent melting point predictions were found for poly-unsaturated TAGs which contain two linolenic acids (le) or a combination of linoleic (l) and linolenic acid (le). It must be noted that experimental data especially of poly-unsaturated TAGs are sparse.

From the thermodynamic scores obtained, it can be deduced that parameter sets yielding in the sense of thermodynamic consistency improved predictions can be identified when more consistent data are used during the optimization procedure. Conclusively, for predicting the

17

334 thermodynamic properties of saturated TAGs, no additional constraints addressing

335 thermodynamic consistency are necessary, and even fitting the model equations on fewer data

336 results in reasonable predictions. As for unsaturated TAGs, a broad dataset is needed to overcome

337 the lack of thermodynamic consistency. However, constrained optimization schemes deliver

338 parameter sets with improved consistency.

339 **3.2. <u>Model sensitivity</u>**

340 *3.2.1. Influence of constraints on model set-up conditions*

341 Next to satisfying a good prediction quality and established thermodynamic consistency of the

342 predictions generated by the model (discussed above), also intermediate characteristic data

343 generated, e.g., contribution per chain element, can be subject to meaningful relations that should

344 be satisfied, see Figure 3 in section 2.2.1. In this spirit, the bound constraints chosen for the

345 parameters h and s are discussed. To verify their effectiveness, the objective function, eq. [23],

346 was minimized with and without imposing said bound constraints, eq. [24]. The respective

347 parameter estimates were evaluated according to whether they satisfy the different conditions

348 mentioned above.

349 From Table 6, which gives the RMSE of monoacid TAGs, it becomes evident that imposing

350 bound constraints on the parameters h and s does not negatively affect the overall prediction

351 quality. For each case, parameter fit with and without bound constraints, the RMSE is smaller

352 than the RMSE using the reference parameter set. A closer examination of the parameters h and s

353 reveals that the parameter h of the (thermodynamically) unconstrained model fit ($P^{l,L*}$) differs

354 relatively from the literature values by approx. 6.15 %, 24.92 %, and 15.11 % in the α, β', and β-

355 polymorph, respectively. The estimates for s differed from the values given in the literature by

35   18

36

356 approx. 0.24 %, 11.91 %, and 18.11 %, Table 7. When additionally thermodynamic constraints

357 were enforced on the model predictions ($P^{II,L*}$), values obtained for parameter h are close to the

358 literature values. Interestingly, for parameter s, relative deviations of 97.63 %, 94.92 %, and

359 52.81 % for the α, β', and β polymorph, respectively, were found. This indicates that it is

360 necessary to add bound constraints for these parameters to preserve the model's physical

361 justification, i.e., the parameter values match the "processed" experimental values (see section

362 2.2.1). Table 8 summarizes the satisfied and unsatisfied conditions for each case.

363 Concluding, for every case tested in Table 8, one obtains a satisfactory prediction quality.

364 However, it was found that constrained optimization needs to be performed to obtain model

365 predictions that satisfy two necessary conditions: thermodynamic consistency of model

366 predictions and parameter values of h and s close to the "processed" experimental values for

367 monoacid TAGs. From the cases studied, the conclusion can be drawn that these conditions do

368 not act contradictory within the model set-up but rather independently. One might speculate that

369 the described independence is due to a lack of data and low quality of those available, structural

370 discrepancies within the model, or the large number of parameters required.

371 ### *3.2.2. Gradient-based study on parameter sensitivity*

372 An assessment of Wesdorp's model for predicting pure component properties regarding

373 parameter sensitivity was conducted. For each parameter, a sensitivity coefficient Φ_j was

374 estimated at the initial guess P^0 and the estimated parameter sets obtained, $P^{I,L,}$ and $P^{II,L}$,

375 respectively. A ranking based on the magnitude of the normalized partial derivatives indicated

376 the model most sensitive towards six parameters, namely h_0, h, s, h_{xy}, s_{xy}, and T_{inf}. No parameter

377 required for estimating the thermodynamic properties of unsaturated triglycerides was identified

378 as a sensitive parameter. This is due to their special nature compared to the other universal

379 parameters: e.g., h and s are relevant for estimating the thermodynamic properties for every TAG,

380 whereby, e.g., h_O is only relevant to calculate the enthalpy of fusion of TAGs containing oleic

381 acid.

4. Conclusion

383 The well-established model by Wesdorp (1990) for predicting the thermodynamic properties of

384 pure triglycerides was investigated regarding the thermodynamic consistency of the model

385 output. For alternative parameter fitting, a constrained optimization scheme was applied to

386 enforce meeting thermodynamic constraints. The model fit was performed on two datasets, one

387 containing trustworthy data from literature without considering their consistency. Secondly, a

388 subset containing only data satisfying underlying thermodynamic fundamentals was used to

389 generate parameter sets. Accordingly, updated parameter sets ensuring thermodynamically

390 consistent predictions were presented. During the model fitting, enthalpy of fusion and melting

391 point predictions for saturated triglycerides were found to be thermodynamically consistent

392 without imposing additional thermodynamic constraints. The dependence on the type of dataset

393 used for fitting was found to be small. In contrast, melting point predictions of unsaturated

394 triglycerides revealed a significant lack of model predictive power. The most important limitation

395 lies in the amount of data of unsaturated triglycerides available for fitting. The effect of enforcing

396 thermodynamic constraints on the model output was further examined in terms of a preserved

397 reliable description of monoacid TAG data. This revealed that the model set-up conditions, i.e.,

398 bound constraints, linear inequality constraints, and prediction quality, are not interdependent.

39

40

399 This research has given rise to questions regarding model robustness and model sensitivity. A

400 comprehensive sensitivity could provide more information on the underlying model. Therefore,

401 the sources and magnitudes of uncertainty in model parameters and model output need to be

402 investigated. Knowledge of these would provide valuable means to improve model robustness.

403

41

42

References

Baker, Mark; Bouzidi, Laziz; Garti, Nissim; Narine, Suresh S. (2014a): Multi-Length-Scale Elucidation of Kinetic and Symmetry Effects on the Behavior of Stearic and Oleic TAG. I. SOS and SSO. In *J Am Oil Chem Soc* 91 (4), pp. 559–570. DOI: 10.1007/s11746-013-2404-z.

Baker, Mark R.; Bouzidi, Laziz; Garti, Nissim; Narine, Suresh S. (2014b): Multi-Length-Scale Elucidation of Kinetic and Symmetry Effects on the Behavior of Stearic and Oleic TAG. II: OSO and SOO. In *J Am Oil Chem Soc* 91 (10), pp. 1685–1694. DOI: 10.1007/s11746-014-2518-y.

Bayés-García, Laura; Calvet, Teresa; Àngel Cuevas-Diarte, Miquel; Ueno, Satoru; Sato, Kiyotaka (2013a): In situ observation of transformation pathways of polymorphic forms of 1,3-dipalmitoyl-2-oleoyl glycerol (POP) examined with synchrotron radiation X-ray diffraction and DSC. In *CrystEngComm* 15 (2), pp. 302–314. DOI: 10.1039/C2CE26522B.

Bayés-García, Laura; Calvet, Teresa; Cuevas-Diarte, Miquel Àngel; Ueno, Satoru; Sato, Kiyotaka (2013b): Crystallization and Transformation of Polymorphic Forms of Trioleoyl Glycerol and 1,2-Dioleoyl-3- rac -linoleoyl Glycerol. In *J. Phys. Chem. B* 117 (31), pp. 9170–9181. DOI: 10.1021/jp403872a.

Bayés-García, Laura; Calvet, Teresa; Cuevas-Diarte, Miquel Àngel; Ueno, Satoru; Sato, Kiyotaka (2015): Phase Behavior of Binary Mixture Systems of Saturated-Unsaturated Mixed-Acid Triacylglycerols: Effects of Glycerol Structures and Chain–Chain Interactions. In *J. Phys. Chem. B* 119 (12), pp. 4417–4427. DOI: 10.1021/acs.jpcb.5b00673.

Boggs, Paul T.; Kearsley, Anthony J.; Tolle, Jon W. (1999a): A Global Convergence Analysis of an Algorithm for Large-Scale Nonlinear Optimization Problems. In *SIAM Journal on Optimization* 9 (4), pp. 833–862. DOI: 10.1137/S1052623497316026.

Boggs, Paul T.; Kearsley, Anthony J.; Tolle, Jon W. (1999b): A Practical Algorithm for General Large Scale Nonlinear Optimization Problem. In *SIAM Journal on Optimization* 9 (3), pp. 755–778.

Boodhoo, M. V.; Kutek, T.; Filip, V.; Narine, S. S. (2008): The binary phase behavior of 1,3-dimyristoyl-2-stearoyl-sn-glycerol and 1,2-dimyristoyl-3-stearoyl-sn-glycerol. In *Chemistry and Physics of Lipids* 154 (1), pp. 7–18. DOI: 10.1016/j.chemphyslip.2008.04.001.

Bouzidi, Laziz; Boodhoo, Marc V.; Kutek, Tomas; Filip, Vladimir; Narine, Suresh S. (2010): The binary phase behavior of 1,3-dilauroyl-2-stearoyl-sn-glycerol and 1,2-dilauroyl-3-stearoyl-sn-glycerol. In *Chemistry and Physics of Lipids* 163 (6), pp. 607–629. DOI: 10.1016/j.chemphyslip.2010.05.002.

Coutinho, João A.P.; Mirante, Fátima; Pauly, Jerome (2006): A new predictive UNIQUAC for modeling of wax formation in hydrocarbon fluids. In *Fluid Phase Equilibria* 247 (1-2), pp. 8–17. DOI: 10.1016/j.fluid.2006.06.002.

Coutinho, João A.P.; Pauly, Jerome; Daridon, Jean-Luc (2004): Modelling Phase Equilibria in Systems with Organic Solid Solutions. In : Computer Aided Property Estimation for Process and Product Design, vol. 19: Elsevier (Computer Aided Chemical Engineering), pp. 229–249.

Elisabettini, P.; Lognay, G.; Desmedt, A.; Culot, C.; Istasse, N.; Deffense, E.; Durant, F. (1998): Synthesis and physicochemical characterization of mixed diacid triglycerides that contain elaidic acid. In *J Am Oil Chem Soc* 75 (2), pp. 285–291.

445 Flöter, Eckhard (2009): The role of physical properties data in product development. In *Eur. J.*

446 *Lipid Sci. Technol.* 111 (3), pp. 219–226. DOI: 10.1002/ejlt.200800260.

447 Ghazani, Saeed M.; Marangoni, A. G. (2018): New Insights into the β Polymorphism of 1,3-

448 Palmitoyl-stearoyl-2-oleoyl Glycerol. In *Crystal Growth & Design* 18 (9), pp. 4811–4814. DOI:

449 10.1021/acs.cgd.8b00598.

450 Hagemann, J. W.; Rothfus, J. A. (1983): Polymorphism and transformation energetics of

451 saturated monoacid triglycerides from differential scanning calorimetry and theoretical modeling.

452 In *J Am Oil Chem Soc* 60 (6), pp. 1123–1131.

453 Hagemann, J. W.; Tallent, W. H. (1972): Differential scanning calorimetry of single acid

454 triglycerides: Effect of chain length and unsaturation. In *J Am Oil Chem Soc* 49, pp. 118–123.

455 Hamby, D. M. (1994): A review of techniques for parameter sensitivity analysis of environmental

456 models. In *Environmental Monitoring and Assessment* 32, pp. 135–154. Available online at

457 https://link.springer.com/content/pdf/10.1007/BF00547132.pdf, checked on 6/4/2020.

458 Hjorth, Jeppe L.; Miller, Rasmus L.; Woodley, John M.; Kiil, Søren (2015): Thermodynamic

459 Modeling of Multi-phase Solid–Liquid Equilibria in Industrial-Grade Oils and Fats. In *J Am Oil*

460 *Chem Soc* 92 (1), pp. 17–28. DOI: 10.1007/s11746-014-2577-0.

461 Kodali, Darma R.; Atkinson, David; Small, Donald M. (1990): Polymorphic behavior of 1,2-

462 dipalmitoyl-3-lauroyl(PP12)- and 3-myristoyl(PP14)-sn-glycerols. In *Journal of Lipid Research*

463 31, pp. 1855–1864.

464 Lu, Chao; Zhang, Bo; Zhang, Hua; Guo, Yun; Dang, Leping; Liu, Zhengan et al. (2019): Solid–

465 Liquid Phase Equilibrium and Phase Behaviors for Binary Mixtures Composed of

466 Tripalmitoylglycerol (PPP), 1,3-Dipalmitoyl-2-oleoyl-glycerol (POP), and 1,2-Dioleoyl-3-

48

467 palmitoyl-glycerol (POO). In *Ind. Eng. Chem. Res.* 58 (23), pp. 10044–10052. DOI:

468 10.1021/acs.iecr.9b01947.

469 MATLAB (2020): version 9.9.0 (R2020b). Natick, Massachusetts.

470 Moorthy, Arun S.; Liu, Rong; Mazzanti, Gianfranco; Wesdorp, Leendert H.; Marangoni, A. G.

471 (2017): Estimating Thermodynamic Properties of Pure Triglyceride Systems Using the

472 Triglyceride Property Calculator. In *J Am Oil Chem Soc* 94 (2), pp. 187–199. DOI:

473 10.1007/s11746-016-2935-1.

474 Ollivon, Michel; Perron, Roger (1982): Measurements of enthalpies and entropies of unstable

475 crystalline forms of saturated even monoacid triglycerides. In *Thermochimica Acta* 53, pp. 183–

476 194.

477 Takeuchi, M.; Ueno, S.; Sato, K. (2002): Crystallization kinetics of polymorphic forms of a

478 molecular compound constructed by SOS (1,3-distearoyl-2-oleoyl-sn-glycerol) and SSO (1,2-

479 distearoyl-3-oleoyl-rac-glycerol). In *Food Research International* 35, pp. 919–926.

480 Teles dos Santos, Moises; Le Roux, Galo A. C.; Gerbaud, Vincent (2011): Phase Equilibrium and

481 Optimization Tools: Application for Enhanced Structured Lipids for Foods. In *J Am Oil Chem*

482 *Soc* 88 (2), pp. 223–233. DOI: 10.1007/s11746-010-1665-z.

483 Timms, R. E. (1978): Heats of Fusion of Glycerides. In *Chemistry and Physics of Lipids* 21,

484 pp. 113–129.

485 Torczon, Virginia (1997): On the convergence of pattern search algorithms. In *SIAM Journal on*

486 *Optimization* 7 (1), pp. 1–25.

49

50

487   Ueno, S.; Minato, Akiyoshi; Seto, Hideki; Amemiya, Zoshizuki; Sato, Kiyotaka (1997):

488   Synchrotron Radiation X-ray Diffraction Study of Liquid Crystal Formation and Polymorphic

489   Crystallization of SOS (sn-1,3-Distearoyl-2-oleoyl Glycerol). In *J. Phys. Chem. B* 101, pp. 6847–

490   6854.

491   Wesdorp, Leendert H. (1990): Liquid-multiple solid phase equilibria in fats: theory and

492   experiments. Doctoral Dissertation. TU Delft, Delft University of Technology, Delft.

493   Wesdorp, Leendert H.; van Meeteren, J.; Jon, S. de; Giessen, R.; Overbosch, P.; Grootscholten,

494   P. et al. (2013): Liquid-multiple solid phase equilibria in fats: theory and experiments. In A. G.

495   Marangoni, Leendert H. Wesdorp (Eds.): Structure and Properties of Fat Crystal Networks, vol.

496   2. 2nd ed. Boca Raton: CRC Press, pp. 241–418.

497   Zéberg-Mikkelsen, Claus K.; Stenby, Erling H. (1999): Predicting the melting points and the

498   enthalpies of fusion of saturated triglycerides by a group contribution method. In *Fluid Phase*

499   *Equilibria* 162, pp. 7–17.

500

51

26

52

501 **Tables**

502 Table 1. Number of TAGs in datasets D^L and D^C for which melting points (T_m) and enthalpies of

503 fusion (ΔH_f) are available in the literature

	dataset D^L						dataset D^C					
	ΔH_f			T_m			ΔH_f			T_m		
	α	β'	β	α	β'	β	α	β'	β	α	β'	β
saturated	72	48	67	128	127	141	62	42	57	127	123	125
unsaturated	21	19	30	74	42	68	18	15	18	42	27	35
total	93	67	97	202	169	209	80	57	75	169	150	160

504

54

Table 2. Parameter estimates of model fit *Problem I* and *Problem II*, parameter were fit on dataset D^L and D^C, compared to the reference

parameter set

	Parameter	Unit	Reference			$P^{I,L}$			$P^{II,L}$			$P^{I,C}$			$P^{II,C}$		
			α	β'	β	α	β'	β	α	β'	β	α	β'	β	α	β'	β
1	h_0	$kJ\ mol^{-1}$	-31.95	-35.86	-17.16	-25.58	-26.25	-22.13	-25.26	-26.36	-23.90	-24.53	-25.19	-24.11	-24.66	-25.24	-24.11
2	h	$kJ\ mol^{-1}\ nC^{-1}$	2.7	3.86	3.89	2.65	3.48	3.78	2.65	3.48	3.78	2.66	3.48	3.78	2.67	3.48	3.78
3	s_0	$J\ mol^{-1}\ K^{-1}$	-19.09	-39.59	31.04	-3.82	-13.65	11.78	-2.47	-13.81	-7.62	-0.25	-11.64	4.18	-0.63	-11.56	4.33
4	s	$J\ mol^{-1}\ K^{-1}\ nC^{-1}$	6.79	10.13	9.83	6.43	10.23	11.55	5.92	9.71	11.55	6.26	10.45	11.55	6.28	10.25	11.55
5	h_{xy}	$kJ\ mol^{-1}$	-13.28	-19.35	-22.29	-16.21	-22.33	-8.83	-16.97	-22.13	-19.96	-16.54	-24.77	-21.11	-16.45	-24.75	-21.09
6	s_{xy}	$J\ mol^{-1}\ K^{-1}$	-36.7	-52.51	-64.58	-46.95	-71.58	-31.43	-45.51	-67.35	-66.92	-45.98	-80.33	-72.23	-45.78	-78.75	-72.11
7	k	nC	4.39	1.99	2.88	3.67	2.31	0.01	3.78	2.28	3.92	3.21	2.32	-4.44	3.17	2.33	-4.44
8	x_0	nC	1.25	2.46	0.77	2.68	3.30	2.01	2.68	3.28	1.45	2.55	3.21	1.69	2.54	3.21	1.69
9	T_{inf}	K	401.15	401.15	401.15	388.69	381.51	389.85	388.89	380.66	380.00	390.30	381.39	390.40	390.19	381.39	390.58
10	h_{odd}	$kJ\ mol^{-1}$	-	-	2.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	s_{odd}	$kJ\ mol^{-1}\ K^{-1}$	-	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	h_O	$kJ\ mol^{-1}\ nO^{-1}$	-31.7	-28.3	-30.2	-32.73	-21.44	-25.49	-32.82	-21.44	-27.49	-33.01	-23.25	-25.10	-33.01	-23.28	-25.10
13	h_E	$kJ\ mol^{-1}\ nE^{-1}$	-11.7	-15.9	-15.9	-7.52	-6.86	-11.60	-7.55	-6.86	-13.56	-8.35	-8.09	-13.37	-8.37	-8.06	-13.37
14	h_l	$kJ\ mol^{-1}\ nl^{-1}$	-37.7	-37.7	-37.7	-70.12	-37.70	-32.86	-70.24	-53.42	-34.57	-37.7*	-37.7*	-37.7*	-37.7*	-37.7*	-37.7*
15	A_O	nO^{-1}	-3.46	-2.2	-2.93	-1.40	-2.36	-1.28	-1.39	-2.08	-1.26	-3.35	-0.90	-1.37	-3.34	-1.03	-1.49
16	A_E	nE^{-1}	-1.38	-1.34	-1.68	-0.43	-1.16	-1.47	-0.42	-1.13	-1.30	-0.81	-0.91	-1.28	-0.81	-0.91	-1.29
17	A_l	nl^{-1}	-3.35	-2.51	-4.68	-4.33	-3.98	-5.66	-2.82	-3.97	-4.40	-5.21	-1.69	-8.65	-5.78	-1.68	-8.35
18	A_{le}	nle^{-1}	-4.2	-2.23	-5.18	-7.28	1.78	-5.68	-5.06	-4.55	-12.37	-4.93	-2.23	-5.68	-4.85	-3.01	-5.25
19	A_{OO}	nOO^{-1}	-0.01	0.27	0.89	-3.15	-0.74	-2.03	-2.14	-0.65	-2.91	0.07	-0.13	0.56	-0.25	0.18	-0.94
20	A_{EE}	nEE^{-1}	0.01	0.04	0.4	-1.58	0.10	0.27	-1.59	0.06	0.21	-0.97	-0.31	-0.02	-0.97	-0.31	-0.01

Table 2. continued

Parameter		Unit	Reference			P^{I,L}			P^{II,L}			P^{I,C}			P^{II,C}		
			α	β'	β	α	β'	β	α	β'	β	α	β'	β	α	β'	β
21	A_{ll}	nll^{-1}	-3.68	0.55	1.21	-2.01	-5.93	0.46	-1.79	-1.26	-59.47	-3.68*	0.55*	1.21*	-3.68*	0.55*	1.21*
22	A_{lele}	$nlele^{-1}$	-0.98	1.51	1.38	-3.73	-5.44	-0.01	-0.54	0.55	-25.18	-0.58	1.51	0.70	-0.52	0.99	0.98
23	A_{Ol}	nOl^{-1}	0.53	-1	0.71	-0.12	-0.75	1.63	-0.12	0.06	0.74	-1.25	0.02	0.38	-0.64	0.02	3.10
24	A_{Ole}	$nOle^{-1}$	0.83	0.76	0.69	3.04	1.99	1.37	2.00	1.40	2.94	0.43	0.36	0.03	0.16	0.15	-1.19
25	A_{lle}	$nlle^{-1}$	-2.97	1.12	0.73	-3.87	2.43	-3.56	-1.28	-0.89	57.40	-3.46	1.12	0.40	-3.40	0.60	0.68
26	B_{O}	nO^{-1}	0	-4.3	-3.7	-9.11	-3.16	-14.33	-8.99	-6.55	-14.57	0.65	-17.75	-13.82	0.63	-16.74	-12.57
27	B_{l}	nl^{-1}	5.4	-7.8	-1.5	11.38	7.59	6.39	4.47	5.27	4.30	-2.51	-14.73	17.94	-2.56	-14.76	17.25
28	B_{le}	nle^{-1}	2.6	-13.7	-1.8	9.17	-100.00	5.98	2.83	0.30	30.99	6.02	-13.70	1.72	6.07	-13.75	1.72

508 *reference parameter, no refit was possible due to lack of data

57

58

509 Table 3. Summary of prediction quality results for the enthalpy of fusion (in kJ/mol) and melting

510 point (in K); model parameters estimated for *Problem I* and *Problem II* and the reference

511 parameter sets. Model fit was performed on datasets D^C and D^L and prediction quality assessed on

512 D^L

		dataset D^L				dataset D^C			
quantity		ΔH_f		T_m		ΔH_f		ΔH_f	
counting measure		*RMSE*	*U/O*	*RMSE*	*U/O*	*RMSE*	*U/O*	*RMSE*	*U/O*
saturated									
α	*Reference*	7.32	1.00	7.00	0.54	7.32	1.00	7.00	0.54
	P^I	6.96	1.32	6.96	0.28	7.07	1.32	6.94	0.28
	P^{II}	6.91	1.25	6.99	1.41	7.09	1.32	6.93	1.41
β'	*Reference*	21.42	0.37	4.58	1.21	21.42	0.37	4.58	1.06
	P^I	16.35	1.82	3.74	1.08	16.38	2.20	3.67	0.92
	P^{II}	16.36	1.82	3.76	1.08	16.37	2.20	3.67	1.21
β	*Reference*	14.69	0.46	2.68	1.21	14.69	0.46	2.68	1.06
	P^I	14.30	1.58	3.22	1.19	12.26	1.23	3.05	0.92
	P^{II}	12.26	1.31	3.73	0.96	12.26	1.23	3.05	1.21
unsaturated									
α	*Reference*	18.66	0.75	13.63	1.03	18.66	0.75	13.63	1.03
	P^I	15.18	0.50	7.49	0.38	16.66	0.50	41.51	0.38
	P^{II}	15.16	0.50	7.74	1.00	16.66	0.50	42.95	1.00
β'	*Reference*	13.74	0.50	11.67	0.97	13.74	0.50	11.67	2.04
	P^I	10.14	0.64	6.80	1.00	10.79	2.00	11.80	0.83
	P^{II}	10.14	0.64	7.11	1.34	10.79	1.00	10.47	1.00
β	*Reference*	11.12	0.27	6.27	1.03	11.12	0.27	6.27	2.04
	P^I	11.84	2.11	2.97	0.91	10.71	0.75	16.24	1.44
	P^{II}	9.91	1.15	13.09	0.74	10.70	0.75	15.70	0.70
	IDEAL	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*

59

60

Table 4. Summary of prediction quality results for the enthalpy of fusion (in kJ/mol) and melting point (in K); model parameters estimated for *Problem I* and *Problem II* and the reference parameter set. Model fit was performed and tested on dataset D^{MP} for saturated TAGs

quantity		dataset D^L				dataset D^C			
		ΔH_f		T_m		ΔH_f		T_m	
counting measure		RMSE	U/O	RMSE	U/O	RMSE	U/O	RMSE	U/O
saturated									
α	Reference	7.32	1.00	7.00	0.54	7.32	1.00	7.00	0.54
	P^I	16.72	0.16	5.13	0.26	16.72	0.16	5.13	0.26
	P^{II}	13.81	0.82	5.72	1.28	13.81	0.82	5.72	1.28
β'	Reference	19.83	0.36	4.63	1.44	19.83	0.36	4.63	1.44
	P^I	33.60	1.14	2.62	1.05	33.60	1.14	2.62	1.05
	P^{II}	43.14	0.22	2.81	1.21	43.14	0.22	2.81	1.21
β	Reference	14.61	0.45	2.70	1.23	14.61	0.45	2.70	1.23
	P^I	44.89	0.00	2.60	1.27	44.89	0.00	2.60	1.27
	P^{II}	33.13	0.00	2.46	1.01	33.13	0.00	2.46	1.01
	IDEAL	0	1	0	1	0	1	0	1

Table 5. Summary of the thermodynamic consistency score *TC* of melting point predictions using the parameters estimates $P^{I,C}$, $P^{II,C}$, $P^{I,L}$, $P^{II,L}$ compared to the reference parameters, predictions obtained for dataset D^L and D^C

	dataset D^L	dataset D^C
saturated		
Reference	73.89	75.4
$P^{I,C}$	100	100
$P^{II,C}$	100	100
$P^{I,L}$	100	100
$P^{II,L}$	100	100
unsaturated		
Reference	37.6	51.2
$P^{I,C}$	69.6	81.4
$P^{II,C}$	81.6	100
$P^{I,L}$	48.8	72.9
$P^{II,L}$	100	100
IDEAL	100	100

61

62

520 Table 6. Summary of RMSE for predictions of enthalpy of fusion (in kJ/mol) and melting point

521 (in K) of monoacid TAG data obtained using different parameter set $P^{I,L}$, $P^{II,L}$, $P^{I,L*}$, $P^{II,L*}$

Parameter set		ΔH_f	T_m
saturated			
α	*Reference*	6.95	2.69
	$P^{I,L}$	6.75	5.07
	$P^{II,L}$	6.75	5.12
	$P^{I,L*}$	6.75	5.07
	$P^{II,L*}$	6.78	5.22
β'	*Reference*	31.14	5.47
	$P^{I,L}$	19.39	2.90
	$P^{II,L}$	19.40	3.09
	$P^{I,L*}$	13.71	3.81
	$P^{II,L*}$	14.22	3.99
β	*Reference*	21.96	2.78
	$P^{I,L}$	17.08	4.19
	$P^{II,L}$	17.10	5.91
	$P^{I,L*}$	15.21	2.76
	$P^{II,L*}$	15.48	2.66
	IDEAL	*0*	*0*

522 *no bound constraints imposed

523

524 Table 7. Summary of estimated parameters *h* and *s*

	h			**s**		
Case	**α**	**β'**	**β**	**α**	**β'**	**β**
$P^{I,L}$	2.7	3.5	3.8	6.4	10.2	11.6
$P^{II,L}$	2.7	3.5	3.8	5.9	9.7	11.6
$P^{I,L}*$	2.7	2.9	3.6	6.1	8.6	12.4
$P^{II,L}*$	2.7	3.0	3.6	0.1	0.5	16.0
Literature values	2.5	3.9	4.2	6.1	9.8	10.5
Linear fit**	2.2	3.4	4.2	4.7	8.8	10.7

525 *no bound constraints imposed

526 **linear fit performed on the experimental enthalpy of fusion and entropy of fusion data of monoacid TAGs against
527 total carbon number (Figure 3)

528

64

529 Table 8. Comparison of the fulfillment of different model conditions for different model fit cases

	Optimization conditions		Model output conditions		
Case	Bound constraints $[lb_{h,s}\,;\,ub_{h,s}]$	Linear inequality constraints $c_i(P)$	Thermodynamic consistency $TC = 100$	Prediction quality RMSE	{h,s} within physical bounds
P^{I,L}	1	0	0	1	1
P^{II,L}	1	1	1	1	1
P^{I,L*}	0	0	0	1	1
P^{II,L*}	0	1	1	1	0

530 *no bound constraints imposed

531

65

66

Figure legends

Figure 1. Geometrical approach to describe the chain length differences in a TAG according to Wesdorp (1990)

Figure 2. Summary of the model for predicting the thermodynamic properties of pure triglycerides presented by Wesdorp (1990)

Figure 3. Enthalpy of fusion (a) and entropy of fusion (b) as a function of the total carbon number for monoacid TAGs in α-polymorph (squares), β'-polymorph (circles), β-polymorph (triangles)

Figure 4. Predictions for enthalpy of fusion and melting point for the α-polymorph using reference parameter set (squares), parameter set P^I (circles), and parameter set P^{II} (triangles); saturated TAGs (empty markers), unsaturated TAGs (filled markers). The solid black line indicates the 'perfect prediction', with dashed and dotted lines representing deviations of $\pm$ 10 units and $\pm$ 20 units, respectively. Parameter sets were fit on dataset D^L, (a, b). Parameter sets were fit on the curated dataset D^C (c, d)

67

68