J Am Oil Chem Soc
DOI 10.1007/s11746-016-2935-1

ORIGINAL PAPER

Estimating Thermodynamic Properties of Pure Triglyceride Systems Using the Triglyceride Property Calculator

Arun S. Moorthy[1,4] · Rong Liu[2] · Gianfranco Mazzanti[2] · Leendert H. Wesdorp[3] · Alejandro G. Marangoni[1]

Received: 3 August 2016 / Revised: 29 November 2016 / Accepted: 7 December 2016
© AOCS 2016

Abstract To date, the most comprehensive model for predicting thermodynamic properties of pure triglycerides was presented by Wesdorp in "Liquid-multiple solid phase equilibria in fats: theory and experiments" (1990). In this paper, we present (1) corrections to the published model, as well as (2) a software implementation of the model for numerical assessment. The software tool, Triglyceride Property Calculator (TPC), uses a semi-empirical model to estimate the enthalpy of fusion and melting temperature for a given triglyceride based on its molecular composition and polymorphic form. These estimates are compared to experimentally collected data when available. The web application is available at http://www.crcfoodandhealth.com (under research tools) and through the AOCS Lipid Library. The quality of estimates is characterized according to defined counting metrics and presented for TAG subcategories. Additionally, the extrapolative value of the TPC is assessed by checking for consistency with underlying thermodynamic constraints. The current TPC implementation is effective in describing experimentally collected melting point data, with greater than 91% of the fitted values falling within 10% of the actual data. The TPC is also very good at describing collected enthalpy data. The underlying semi-empirical model and parameter set perform well in ensuring enthalpy predictions are thermodynamically consistent, however, extrapolated melting temperatures appear unreliable. Developing models and parameter sets that ensure thermodynamic consistency is a priority with future TPC iterations.

Keywords Empirical modeling · Numerical simulation · Solid–liquid equilibrium · Triglyceride properties

Introduction

Understanding and controlling the melting and solidification behaviour of oils and fats is of importance for many applications. The complex compositions of fats, formed by a mixture of triglycerides exhibiting multiple physical states (liquid or solid) and polymorphisms within states, makes the prediction of melting and solidification of fats difficult. A welcome advance to understanding the behavior of mixed triglyceride mixed phase fat systems is the development of formal relationships, and thus predictive tools, describing the behavior of their constituent triglycerides.

King and Garner showed molar heat of enthalpy increased linearly with the number of CH_2 groups for many long chain compounds [1]. In particular, they showed that the enthalpy of fusion (H_f) for a fatty acid methyl or ethyl ester with chain length exceeding 8 can be modeled by the relationship:

$$H_f(n) = hn + b, \tag{1}$$

where h and b are constants and n is the number of CH_2 groups. Their model parameters were fitted with data collected for pure component triglycerides by Bailey [2], and then with data collected by Timms [3]. It should be noted that Timms also introduced a series of non-linear *correction* rules

✉ Alejandro G. Marangoni
amarango@uoguelph.ca

1 Department of Food Science, University of Guelph, Guelph, ON, Canada

2 Department of Process Engineering and Applied Science, Dalhousie University, Halifax, NS, Canada

3 Wesdorp R&D, Barendrecht, The Netherlands

4 Present Address: Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, MD, USA

to improve model predictions for specific triglyceride groups: (1) Saturated and trans-acids, (2) Cis unsaturated acids, (3) Stable β'-polymorphic forms, and (4) partial glycerides.

A study by Perron [4] generated improved correlation constants for a larger sample and variety of stable and unstable triglycerides. In 1990, Wesdorp [5] introduced a set of additive functions to improve the original linear model of King and Garner, extending on the strategy suggested by Timms. The same work by Wesdorp also included a large reference set of measured enthalpies of fusion and melting temperatures for a significant number of triglycerides and in many stable polymorphic forms. In a similar line of work, Hagemann and Rothfus used molecular modeling techniques to explore the existence of sub-modifications within a polymorphic form [6, 7]. These models explore the physics of phase behaviours in fully saturated lipids. In particular, Hagemann and Rothfus explored modes of packing geometry and molecular rotation freedom for various polymorphic forms, and the subsequent effect on thermal properties. Due to the complexity of this line of modeling, the range of triglycerides for which estimates are generated is more limited than what we can produce using the empirical methods described previously. However, it would be remiss not to point out that estimates achieved using packing geometry and molecular modeling are less prone to error due to their being derived from physical principles.

To our knowledge, the work of Wesdorp [5, 8] (1990, published with improvements in 2012) is the most comprehensive resource describing thermochemical properties of pure component triglyceride systems. The work of Wesdorp was reviewed by Liu [9] who questioned the published model parameters but was unable to find a satisfactory resolution. In this paper, we present a numerical tool built from the original mathematics of Wesdorp in the Triglyceride Property Calculator (TPC). The model parameters have been revised and corrected from the work published in [5, 8]. By designing, optimizing, and making this software tool available, we are able to:

(1) Provide a user-friendly software tool that can be employed by members of the oil chemistry community.
(2) Use the software tool for rigorous simulation investigations that will help us formalize gaps in our current knowledge and, subsequently, guide improvements in future model iterations.

This paper is structured to include a brief but necessary revision of Wesdorp's original work in modeling properties of pure component triglycerides, a detailed description of our simulation investigation and analytical techniques, a summary of numerical results, and an extended discussion about the general value of the existing TPC implementation and our plans moving forward with future TPC iterations.

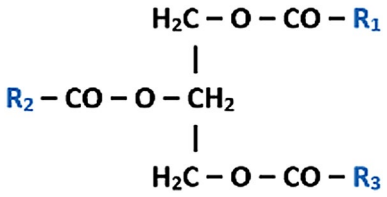

Fig. 1 Triglyceride stencil used for mathematical model description

Materials and Methods

Mathematical Model and Software Implementation

The underlying semi-empirical model, herein referred to as the Wesdorp model, for predicting thermodynamic properties of pure triglycerides was originally developed in [5], presented with improvements in [8] and further analysis in [9]. Readers should refer to the aforementioned references for a comprehensive description of the mathematics. In this section, we present a summarized representation of the Wesdorp model, with a focus on features of importance to the numerical calculator.

For convenience of description, we illustrate a simple sketch of a triglyceride as Fig. 1. As shown, there are three hydrocarbon chains attached to a glycerol head group. The length, level of saturation, and interaction between these hydrocarbon chains (referred to as R_1, R_2, and R_3 in Fig. 1) allow us to describe the triglyceride by defined attributes. These attributes are subsequently used to predict the thermodynamic properties of the triglyceride.

Defining Triglyceride Attributes

A triglyceride can be discussed in terms of molecular composition using three letter acronyms. For example, tristearin is referred to as SSS. For a given three letter description, we define the following triglyceride attributes:

n_1 = number of carbons in R_1,

n_2 = number of carbons in R_2,

n_3 = number of carbons in R_3,

u_1 = number of double bonds in R_1,

u_2 = number of double bonds in R_2,

u_3 = number of double bonds in R_3,

where R_1 through R_3 are as referenced in Fig. 1, and values for n_1 through u_3 can be found in Table 1. The Wesdorp model requires the total number of carbons (n) and total number of double bonds (u), computed as follows:

Table 1 Symbols used in description of triglyceride composition

Symbol	Fatty acid	*nC*	*nU*	Symbol	Fatty acid	*nC*	*nU*
U	Butyric	4	0	G	Lignoceric	24	0
K	Caproic	6	0	F	Ceric	26	0
R	Caprylic	8	0	T	Palmitoleic	16	1
C	Capric	10	0	O	Oleic	18	1
L	Lauric	12	0	J	Linoleic	18	2
M	Myristic	14	0	N	Linolenic	18	3
P	Palmitic	16	0	D	Elaidic	18	1
S	Stearic	18	0	E	Erucic	22	1
A	Arachidic	20	0	H	Arachidonic	20	4
B	Behinic	22	0	#	# (odd chain)	#	*0*

Columns "*nC*" and "*nU*" describing the number of carbons and number of double bounds, respectively, in the fatty acid chain

$$n = n_1 + n_2 + n_3,$$

$$u = u_1 + u_2 + u_3.$$

From these primary attributes, many secondary and tertiary attributes can be defined. To classify families of triglycerides, we compute the differences in length of fatty acid chains. We define three lengths, the length of the shortest side chain (1 or 3 position) fatty acid in the triglyceride as P, the length of the middle side chain (position 2) fatty acid in the triglyceride as Q, and the length of the longest side chain (1 or 3 position) fatty acid in the triglyceride as R. Using computational syntax and the primary length attributes n_1 through n_3, we can define lengths P, Q and R as:

$$P = \min\,(n_1, n_3),$$

$$Q = n_2,$$

$$R = \max\,(n_1, n_3).$$

Subsequently, we can define the difference in chain lengths of the 3 fatty acid side chains using:

$$x = Q - P, \text{ and,}$$

$$y = R - P.$$

When describing unsaturated triglycerides (when at least one chain includes a double bond), the enthalpy of fusion and melting temperature are modified by the presence of particular unsaturated fatty acid chains: (1) oleic acid, (2) elaidic acid, (3) linoleic acid, and (4) linolenic acid. We can thus define the following variables:

n_O = number of oleic chains in the TAG,

n_E = number of elaidic chains in the TAG,

n_J = number of linoleic chains in the TAG,

n_N = number of linolenic chains in the TAG.

If the unsaturated triglyceride has multiple unsaturated chains, we also define the number of pairs of unsaturated fatty acids as follows:

n_{OO} = number of oleic − oleic pairs,

n_{EE} = number of elaidic − elaidic pairs,

n_{JJ} = number of linoleic − linoleic pairs,

n_{NN} = number of linolenic − linolenic pairs,

n_{OJ} = number of oleic − linoleic pairs,

n_{ON} = number of oleic − linolenic pairs,

n_{JN} = number of linoleic − linolenic pairs.

Using computational syntax, these pairs can be calculated as:

$$n_{OO} = \max\,(0, n_O - 1),$$

$$n_{EE} = \max\,(0, n_E - 1),$$

$$n_{JJ} = \max\,(0, n_J - 1),$$

$$n_{NN} = \max\,(0, n_N - 1),$$

$$n_{OJ} = \max\,(0, n_O + n_J - 1),$$

$$n_{ON} = \max\,(0, n_O + n_J - 1),$$

$$n_{JN} = \max\,(0, n_J + n_N - 1),$$

where the MAX function is used to maintain non-negativity during calculations in cases where any or all of the

unsaturated fatty acid combinations are not present in the triglyceride. For example, if there is are no oleic fatty acid in the triglyceride system, the number of pairs of oleic acids will be 0 and not −1. Using these attributes, the enthalpies of fusion and melting temperature of pure triglycerides can be calculated.

Predicting Properties of Saturated Triglycerides

In [8], the authors describe the enthalpy and entropy of fusion for saturated triglycerides using:

$$\Delta H_f^{\text{sat}}(n) = hn + h_o + h_{xy} f_{xy} + h_{\text{odd}} f_{\text{odd}} \cdot f_\beta, \tag{2}$$

$$\Delta S_f^{\text{sat}}(n) = sn + s_o + s_{xy} f_{xy} + s_{\text{odd}} f_{\text{odd}} \cdot f_\beta + R \ln 2 \cdot f_{\text{asym}} \cdot f_\beta, \tag{3}$$

where h and s are the enthalpy and entropy contributions, respectively, of the total carbon number in the triglyceride, h_o and s_o are the contributions due to the glycerol head group, h_{xy} and s_{xy} are the contributions due to difference in chain length between fatty acid chains, h_{odd} and s_{odd} are the contributions due to an odd number of total carbons in the fatty acids, R is the universal gas constant, and functions f_{xy}, f_{odd}, f_{asym} and f_β are defined as:

$$f_{xy}(x, y) = 2 - \exp\left(-\left(\frac{x - x_o}{k_x}\right)^2\right) - \exp\left(-\left(\frac{y}{k_y}\right)^2\right), \tag{4}$$

$$f_{\text{odd}}(n_1, n_2, n_3) = \begin{cases} 1, \text{if } n_1 \text{or} n_2 \text{or} n_3 \text{ is an odd number,} \\ 0, \text{if all fatty acid chains are even,} \end{cases} \tag{5}$$

$$f_{\text{asym}} = \begin{cases} 1, \text{if } y \neq 0, \\ 0, \text{if } R_1 = R_3, \end{cases} \tag{6}$$

$$f_\beta = \begin{cases} 1, \text{if TAG is } \beta \text{ polymorph,} \\ 0, \text{ if not } \beta \text{ polymorph,} \end{cases} \tag{7}$$

where parameters x_o, k_x and k_y are used for data fitting. Input TAG attributes (n_1, n_2, n_3, x, y, etc.) are as defined in the previous section. For ease of reading, we will not explicitly describe the input TAG attributes in the remainder of the paper. The melting point of a triglyceride can be defined as:

$$T_f = \frac{\Delta H}{\Delta S}, \tag{8}$$

and so for a saturated triglyceride, based on the descriptions of enthalpy and entropy described in Eqs. 2 and 3:

$$\begin{aligned} T_f^{\text{sat}} &= \frac{\Delta H_f^{\text{sat}}}{\Delta S_f^{\text{sat}}} \\ &= \frac{hn + h_o + h_{xy} f_{xy} + h_{\text{odd}} f_{\text{odd}} f_\beta}{sn + s_o + s_{xy} f_{xy} + s_{\text{odd}} f_{\text{odd}} f_\beta + R \ln 2 \cdot f_{\text{asym}} f_\beta} \\ &= \frac{hn + h'_o}{sn + s'_o}, \end{aligned} \tag{9}$$

where

$$h'_o = h_o + h_{xy} f_{xy} + h_{\text{odd}} f_{\text{odd}} f_\beta, \tag{10}$$

$$s'_o = s_o + s_{xy} f_{xy} + s_{\text{odd}} f_{\text{odd}} f_\beta + R \ln 2 \cdot f_{\text{asym}} f_\beta. \tag{11}$$

Equation 9 can be re-written as a series, which will prove valuable in describing unsaturated triglycerides, following a Taylor series expansion around $\left(\frac{1}{n}\right)$ resulting in:

$$T_f^{\text{sat}} = \frac{h}{s}\left(1 + \left(\frac{h'_o}{h} - \frac{s'_o}{s}\right)\frac{1}{n} - \frac{s'_o}{s}\left(\frac{h'_o}{h} - \frac{s'_o}{s}\right)\frac{1}{n^2} + \ldots\right). \tag{12}$$

Truncating the series as to carry second order errors and making substitutions for convenience, we can re-write Eq. 12 as:

$$T_f^{\text{sat}} \cong T_\infty\left(1 + \frac{A_s}{n} - \frac{A_s B_s}{n^2}\right), \tag{13}$$

where

$$T_\infty = \frac{h}{s}, \tag{14}$$

$$A_s = \frac{h'_o}{h} - \frac{s'_o}{s}, \tag{15}$$

$$B_s = \frac{s'_o}{s}. \tag{16}$$

Thus, with appropriate parameters, the enthalpy of fusion and melting point of a saturated TAG can be estimated using Eqs. 2 and 13, and their dependencies (Eqs. 10, 11, 14, 15, and 16).

During initial model fitting, the authors of [8] observed that Taylor series coefficients A_s and B_s were best described as a functions of chain length differences x and y, thus introducing the following Taylor series parameters:

$$T_\infty = T_{\infty,e}, \tag{17}$$

$$A_s = A_o + A_{\text{odd}} f_{\text{odd}} + A_x x + A_{x^2} x^2 + A_{xy} xy + A_y y + A_{y^2} y^2, \tag{18}$$

$$B_s = B_o + B_{\text{odd}} f_{\text{odd}} + B_x x + B_{x^2} x^2 + B_{xy} xy + B_y y + B_{y^2} y^2, \tag{19}$$

where parameters of the form C_i are fitted using experimental data. For completeness, the current implementation of the TPC estimates two melting temperatures, one using Eqs. 13–16, and the other using Eqs. 13, 17–19. The melting temperature estimated using Eqs. 13–16 is indicated with an "[a]" on the user interface, and the temperature estimated using Eqs. 13, 17–19 is indicated with a "[b]". For ease of discussion, we refer to the varied implementations by these indication letters: Model A and Model B.

Predicting Properties of Unsaturated Triglycerides

In [8], it was assumed that the thermodynamic properties of an unsaturated triglyceride can be estimated using the thermodynamic properties of the equivalent saturated system (based on the number of carbons and difference in chain length) and corrected for using the number of unsaturated chains in the triglyceride. The work done for predicting unsaturated triglycerides is far less developed than is for saturated triglycerides. To summarize, the enthalpy of fusion for an unsaturated triglyceride is estimated using:

$$\Delta H_f^{\text{unsat}} = \Delta H_f^{\text{sat}} + \hat{h}_O n_O + \hat{h}_E n_E + \hat{h}_J n_J, \tag{20}$$

where $\hat{h}_O$, $\hat{h}_E$, $\hat{h}_J$ are the contribution to enthalpy due to oleic, elaidic and linoleic acids, respectively, and the melting point of an unsaturated triglyceride is estimated using:

$$T_f^{\text{unsat}} = T_\infty \left(1 + \frac{A_u}{n} - \frac{A_u B_u}{n^2}\right), \tag{21}$$

where

$$\begin{aligned} A_u = {} & A_s + \hat{A}_O n_O + \hat{A}_E n_E + \hat{A}_J n_J \\ & + \hat{A}_N n_N + \hat{A}_{OO} n_{OO} + \hat{A}_{EE} n_{EE} \\ & + \hat{A}_{JJ} n_{JJ} + \hat{A}_{NN} n_{NN} + \hat{A}_{OJ} n_{OJ} \\ & + \hat{A}_{ON} n_{ON} + \hat{A}_{JN} n_{JN}, \end{aligned} \tag{22}$$

$$B_u = B_s + \hat{B}_O n_O + \hat{B}_J n_J + \hat{B}_N n_N, \tag{23}$$

and parameters of the form $\hat{C}_i$ and $\hat{C}_{ij}$ are fitting parameters describing the contribution of unsaturated fatty acid indexed by i and unsaturated pairs by i,j. As a reminder, triglyceride attributes of the form n_i and $n_{i,j}$ were defined earlier in this paper (see the section "Defining Triglyceride Attributes"). The parameters T_∞, A_s and B_s are defined as in Eqs. 14–16 or 17–19.

The described system of algebraic equations relies on the estimation of 28 parameters to estimate enthalpies of fusion and melting temperatures using Model A, and 42 parameters to estimate the melting temperature using Model B. These parameters were determined using experimentally collected data and a least squares regression by the authors of [8]. The authors of [9] challenged some of the published coefficient estimates and presented edited values. Further investigation has now made clear that many parameters in both [5] and [9] can be further improved, and so we present the best known parameters as Table 2. Continued collection of pure triglyceride enthalpies and melting points is being undertaken to improve the quality of parameter estimates. Experimental data presented in this article was extracted directly from [10].

Software Implementation Details

Numerical estimates of enthalpy of fusion for saturated triglycerides were obtained by solving (2) and its dependents (4–7). Melting points estimates for saturated triglycerides were obtained by solving (13) with either parameters as described in (14–16) or (17–19). Numerical estimates of enthalpy of fusion for unsaturated triglycerides were obtained by solving (20) and its dependents. Melting point estimates for unsaturated triglycerides were obtained by solving (21) and its dependents. This solution system was implemented using R [11] and the TPC web application which interfaces the Wesdorp model and experimentally available values was created using the R Shiny Package [12]. Complete access to the solution scheme implementation, parameter sets, and experimental data is available for download and review through https://www.crcfoodandhealth.com. The web application can also be retrieved through https://www.crcfoodandhealth.com or following links through the AOCS lipids library.

Simulation and Assessment Strategy

We characterize the functionality of the TPC by investigating two questions:

(1) How well does the predicted enthalpy and melting temperature of a triglyceride compare to the available experimental data?
(2) Do the predictions "make sense" in terms of the thermodynamic expectations?

In the case of the first question, we will be limited to the set of experimentally collected results that are available. However, we can make use of the complete spectrum of triglyceride property estimates that can be generated using the TPC while investigating question two.

For any triglyceride with available experimental data, we can describe the quality of each individual prediction using a relative error (RE) measure:

$$RE = 1 - \frac{y_p}{y_e}, \tag{24}$$

Table 2 Parameters used for model predictions. Adapted and corrected from Wesdorp *et al.* (2012) and Liu (2014)

#	Parameter	Units	α—modification	β'—modification	β—modification
1	h_o	kJ/mol	−31.95	−35.86	−17.16
2	h	kJ/mol · nC	2.7	3.86	3.89
3	s_o	J/K · mol	−19.09	−39.59	31.04
4	s	J/K · mol · nC	6.79	10.13	9.83
5	h_{xy}	kJ/mol	−13.28	−19.35	−22.29
6	s_{xy}	J/K · mol	−36.7	−52.51	−64.58
7	$k_x = k_y = k$	nC	4.39	1.99	2.88
8	x_o	nC	1.25	2.46	0.77
9	$T_{\infty,e}$	K	401.15	401.15	401.15
10	h_{odd}	kJ/mol	–	–	2.29
11	s_{odd}	J/mol	–	–	–
12	A_o	$1/nC$	−9.058	−8.454	−8.048
13	A_{odd}	$1/nC$	−0.196	−0.308	−0.019
14	A_x	$1/nC$	0.003	−0.104	0.074
15	A_{x^2}	$1/nC$	−0.062	−0.019	−0.035
16	A_{xy}	$1/nC$	0.115	0.074	0.008
17	A_y	$1/nC$	−0.453	−0.497	−0.404
18	A_{y^2}	$1/nC$	−0.006	0.012	0.011
19	B_o	$1/nC$	−4.484	−0.265	2.670
20	B_{odd}	$1/nC$	−0.003	0.005	0.008
21	B_x	$1/nC$	−0.001	0.550	−0.317
22	B_{x^2}	$1/nC$	0.149	0.074	0.086
23	B_{xy}	$1/nC$	−0.366	−0.341	0.041
24	B_y	$1/nC$	1.412	2.342	0.550
25	B_{y^2}	$1/nC$	−0.002	−0.136	9e-4
26	$\hat{A}_O$	$1/nO$	−3.46	−2.2	−2.93
27	$\hat{A}_E$	$1/nE$	−1.38	−1.34	−1.68
28	$\hat{A}_J$	$1/nJ$	−3.35	−2.51	−4.69
29	$\hat{A}_N$	$1/nN$	−4.2	−2.23	−5.18
30	$\hat{A}_{\text{OO}}$	$1/nOO$	−0.01	0.27	0.89
31	$\hat{A}_{\text{EE}}$	$1/nEE$	0.01	0.04	0.4
32	$\hat{A}_{\text{JJ}}$	$1/nJJ$	−3.68	0.55	1.21
33	$\hat{A}_{\text{NN}}$	$1/nNN$	−0.98	1.51	1.38
34	$\hat{A}_{\text{OJ}}$	$1/nOJ$	0.53	−1	0.71
35	$\hat{A}_{\text{ON}}$	$1/nON$	0.83	0.76	0.69
36	$\hat{A}_{\text{JN}}$	$1/nJN$	−2.97	1.12	0.73
37	$\hat{B}_O$	$1/nO$	0	−4.3	−3.7
38	$\hat{B}_J$	$1/nJ$	5.4	−7.8	−1.5
39	$\hat{B}_N$	$1/nN$	2.6	−13.7	−1.8
40	$\hat{h}_O$	kJ/mol · nO	−31.7	−28.3	−30.2
41	$\hat{h}_E$	kJ/mol · nE	−11.7	(−15.9)	−15.9
42	$\hat{h}_J$	kJ/mol · nJ	(−37.7)	(−37.7)	−37.7

Unit "nC" indicates number of carbons. For example, the units for parameter (h) are kilojoules per mole per number of carbons in the TAG

Units "nX" indicate the number of molecules X. For example, the units of (A_O) are the inverse number of Oleic chains

Similarly, units "nXY" indicate the number of pairs of molecules X and Y. For example, the units of (A_{OJ}) are inverse number of Oleic-Linoleic pairs

Parameter values in parenthesis indicate those that were guessed due to inadequate data for fitting

Table 3 Summary of possible scoring codes for test of triglyceride predictions

Scenario	$y_\beta > y_{\beta'}$	$y_{\beta'} > y_\alpha$	$y_\beta > y_\alpha$
$y_\beta < y_{\beta'}; y_{\beta'} < y_\alpha; y_\beta < y_\alpha$	**0**	**0**	**0**
$y_\beta > y_{\beta'}; y_{\beta'} < y_\alpha; y_\beta < y_\alpha$	**1**	**0**	**0**
$y_\beta < y_{\beta'}; y_{\beta'} > y_\alpha; y_\beta < y_\alpha$	**0**	**1**	**0**
$y_\beta > y_{\beta'}; y_{\beta'} < y_\alpha; y_\beta > y_\alpha$	**1**	**0**	**1**
$y_\beta < y_{\beta'}; y_{\beta'} > y_\alpha; y_\beta > y_\alpha$	**0**	**1**	**1**
$y_\beta > y_{\beta'}; y_{\beta'} > y_\alpha; y_\beta > y_\alpha$	**1**	**1**	**1**

Binary scoring of the conditionals tested in each column are represented as 0 for false and 1 for true. Code [1,1,1] represents a thermodynamically feasible prediction

where y_p is the predicted measure (enthalpy or melting point), and y_e is the experimental value. In a perfect prediction, $y_p = y_e$ and so the relative error will be zero. If $y_p > y_e$ ($RE < 0$), we consider that an over prediction, and if $y_p < y_e$ ($RE > 0$), we count that as an under prediction.

For our analysis, we look beyond individual triglycerides and attempt to categorize the prediction quality for categories of triglycerides. In particular, we classify the triglycerides into 6 categories: (1) Saturated triglycerides in α polymorphic form, (2) Saturated triglycerides in β' polymorphic form, (3) Saturated triglycerides in β polymorphic form, (4) Unsaturated triglycerides in α polymorphic form, (5) Unsaturated triglycerides in β' polymorphic form, and (6) Unsaturated triglycerides in β polymorphic form. As noted previously, we distinguish an unsaturated triglyceride as including at least one unsaturated fatty acid chain. Conversely, a saturated triglyceride only includes fully saturated fatty acid chains.

We determine prediction quality within the category by looking at a variety of counting measures across the entire collection of triglycerides in each sub category. For example, we consider the ratio of over predictions to under predictions within a category. In an ideal scenario, this ratio of over predictions to under predictions will be approaching 1 (or undefined if all predictions are equal to experimentally collected values). We define three such metrics that we tabulate for each category:

$$\gamma_{o/u} = \frac{\text{number of overpredictions in a category}}{\text{number of underpredictions in a category}}, \tag{25}$$

$$RE_c = \frac{\text{number of predictions with } |RE| < 0.1}{\text{size of category}}, \tag{26}$$

$$RE_f = \frac{\text{number of predictions with } |RE| > 0.5}{\text{size of category}}, \tag{27}$$

where relative error (RE) is as tabulated in Eq. 24.

To investigate whether estimates "make sense", we consider the intra-estimate consistency within a single triglyceride. That is, do the estimates (melting temperature or enthalpy of fusion) within a triglyceride make sense given the triglycerides polymorphic form? We simulated enthalpies and melting points for all possible combinations and polymorphs of triglycerides available in the TPC. At present, the TPC includes the option to select between 19 fatty acids. This results in 4774 *unique* triglyceride compositions.

To assess intra-estimate consistency, we ask:

(1) Is the predicted $T_\alpha < T_{\beta'} < T_\beta$?
(2) Is the predicted $H_\alpha < H_{\beta'} < H_\beta$?

To assess these questions, we create an encoding system that describes the *type of error* where a label is assigned to a triglyceride based on which of the thermodynamic conditionals are not met during prediction. For example, assume we have a triglyceride with measurement y (enthalpy or melting point), a physically consistent estimation would lead to the following scenario:

$$y_\beta > y_{\beta'} \text{ and } y_{\beta'} > y_\alpha, \text{therefore } y_\beta > y_\alpha,$$

where the subscript symbol indicates the triglyceride polymorph. This scenario would be provided a score of [1,1,1], all conditionals are true. A completely incorrect scenario would be:

$$y_\beta < y_{\beta'} \text{ and } y_{\beta'} < y_\alpha, \text{therefore } y_\beta < y_\alpha.$$

This scenario would be provided a score of [0,0,0], all conditionals are not possible. A partially correct estimation might be of the form:

$$y_\beta > y_{\beta'} \text{ and } y_{\beta'} < y_\alpha, \text{therefore } y_\beta > y_\alpha.$$

This scenario would be provide a score of [1,0,1], the second conditional is not consistent with expectations.

The described true–false scoring system leads to six categories that the triglyceride estimates can fall in. These scenarios are summarized in Table 3. We compute errors codes for the two predicted measures (enthalpy and melting point) independently for saturated and unsaturated triglycerides.

Implementation validation and verification has been assessed through repetition and special case investigations. Due to the underlying experimental databases being a work in progress, any comments on the verifiability of the software tool would be speculation. We continue to run software tests and verification, but encourage users to share known implementation issues as they arise.

Results and Discussion

An experimental database of 188 triglycerides with full or partial information regarding enthalpies and melting temperatures for α, β', and β polymorphic forms of each triglyceride were extracted from [10]. In total, experimentally collected enthalpy data for 43 α—saturated

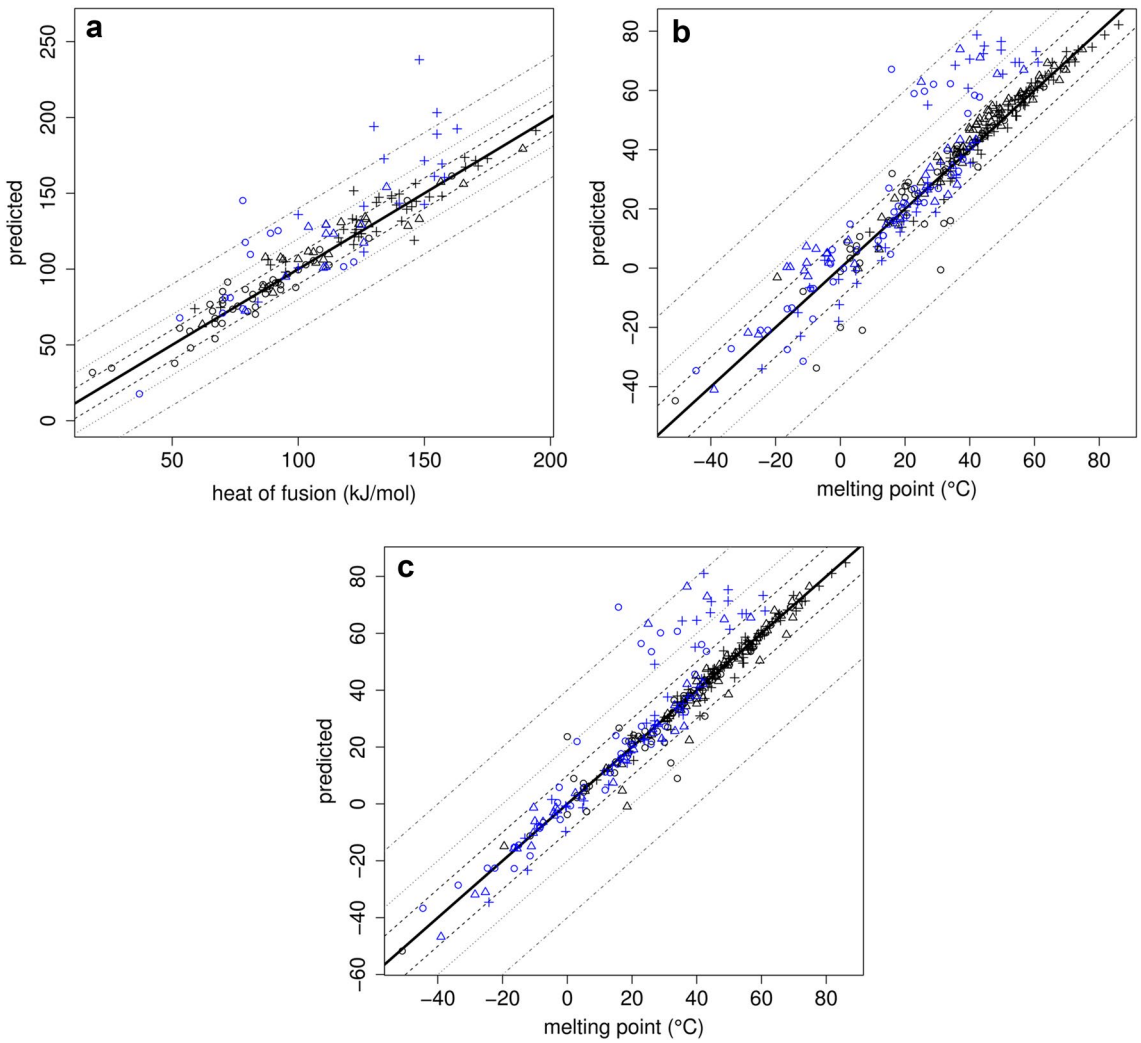

Fig. 2 Prediction quality as assessed for **a** enthalpy of fusion fit, **b** melting temperature fitted using model A—combined enthalpy/temperature fit, and **c** Melting temperature fitted using model B—temperature fit. *Black markers* indicate measures of saturated triglycerides, and *blue markers* indicate measures of unsaturated triglycerides. *Circle markers* (○) indicate triglycerides in α-polymorph, *triangle markers* (△) indicate triglycerides in β'-polymorph, and cross markers (+) indicate triglycerides in β-polymorph. The *solid black line* in either *plot* indicates "perfect prediction", with dashed lines representing bands of errors. In particular, the line closest to the *solid line* indicates a prediction off by ±10 units, followed by bands describing predictions off by ±20, and ±40 units, respectively, moving further from the "perfect fit" line (color figure online)

triglycerides, 24 β'—saturated triglycerides, 40 β—saturated triglycerides, 13 α—unsaturated triglycerides, 11 β'—unsaturated triglycerides, and 17 β—unsaturated triglycerides were available for analysis. Experimentally collected melting temperatures for 67 α—saturated triglycerides, 65 β'—saturated triglycerides, 77 β—saturated triglycerides, 43 α—unsaturated triglycerides, 35 β'—unsaturated triglycerides, and 44 β—unsaturated triglycerides were available for analysis. The extracted experimental data is accessible through the TPC calculator, but can also be retrieved as a data table by downloading the TPC application from http://www.crcfoodandhealth.com (under research tools).

A broad view of the model fit quality to experimental data using TPC is shown through Fig. 2, panel (a) demonstrating enthalpy of fusions, (b) demonstrating melting temperature using Model A, and (c) demonstrating melting temperature using Model B. In total, the TPC performs well in predicting enthalpies of fusion, with only 4 of the 148 (<3%) of the predicted enthalpies diverging from experimental values by a relative error greater than ±50% (see Fig. 2a; Table 4). The enthalpies of saturated triglycerides are very well predicted, with only one of the 107 predictions falling outside of a ±50% error, and ≈ 74% of the tested predictions coming within ±10% of the experimentally collected value. Predictions for enthalpy of saturated

Table 4 Summary of prediction quality results for Enthalpy of Fusion predictions categorized by polymorph and whether the triglyceride is saturated or unsaturated

Model	Polymorph	Category	No. samples	$\gamma_{o/u}$	RE_c	RE_f
A/B	α	Saturated	43	0.95	0.67	0.02
	β'	Saturated	24	1.67	0.71	0
	β	Saturated	40	1.50	0.83	0
	α	Unsaturated	13	2.25	0.15	0.15
	β'	Unsaturated	11	1.20	0.64	0
	β	Unsaturated	17	4.67	0.41	0.06
		IDEAL	∞	*1*	*1*	*0*

Measures $\gamma_{o/u}$, RE_c, and RE_f describe the ratio of over-prediction to under-prediction, the proportion of samples with absolute relative error less than 10% and the proportion of samples with absolute relative errors greater than 50%, respectively, and are defined in Eqs. 25, 26, and 27. To be compared with Fig. 2a. Italic values indicate those that would results from perfect predictions (when predictions and collected measures are equivalent)

Table 5 Summary of prediction quality results for Melting Temperature predictions categorized by polymorph and whether the triglyceride is saturated or unsaturated

Model	Polymorph	Category	No. samples	$\gamma_{o/u}$	RE_c	RE_f
A	α	Saturated	64	2.37	0.98	0
	β'	Saturated	63	3.43	1	0
	β	Saturated	76	1	1	0
	α	Unsaturated	43	1.39	0.91	0
	β'	Unsaturated	35	4	0.94	0
	β	Unsaturated	44	1	0.95	0
B	α	Saturated	64	0.78	1	0
	β'	Saturated	63	0.88	1	0
	β	Saturated	76	1	1	0
	α	Unsaturated	43	1.32	0.93	0
	β'	Unsaturated	35	1.06	0.94	0
	β	Unsaturated	44	1.44	0.98	0
		IDEAL	∞	*1*	*1*	*0*

Measures $\gamma_{o/u}$, RE_c, and RE_f describe the ratio of over-prediction to under-prediction, the proportion of samples with absolute relative error less than 10% and the proportion of samples with absolute relative errors greater than 50%, respectively, and are defined in Eqs. 25, 26, and 27. To be compared with Fig. 2b and c. Of note, relative errors computed using temperature measures in degrees Kelvin. Italic values indicate those that would results from perfect predictions (when predictions and collected measures are equivalent)

triglycerides can be deemed unbiased based on the ratio of over predictions to under predictions being near 1:1.

The predicted enthalpies of unsaturated triglycerides are successful in the sense that only 3 of the 41 samples had errors greater than ±50%, however, only 16 of the 41 would be considered successful by the criteria that the prediction is within ±10% of the experimental value. A deeper investigation of the polymorphs within the unsaturated categories demonstrates that 2 of the 3 errors greater than ±50% originate from predictions of α polymorphs. In just considering β' and β polymorphs, only 1 of the 28 samples generated severely incorrect predictions. It is important to note that there is an apparent bias in predicting enthalpies of fusion for β polymorphs; there are almost 5 times as many over predictions as under predictions. Whether this bias is a consequence of the available experimental data or an intrinsic property of the mathematical model is not clear at the moment. Continued collection of experimental values will provide important insights. A summary of the quality of prediction measures for enthalpy of fusion data is provided as Table 4.

The quality of melting point predictions using the TPC for existing experimentally available data is very strong. Referring to Fig. 2b and c, there are a few predictions that fall beyond the ±20 °C bands. However, more than 98% of the predictions are within 10% of the experimental data and there are no severely incorrect predictions (Relative error > 50%). Using model A to generate melting

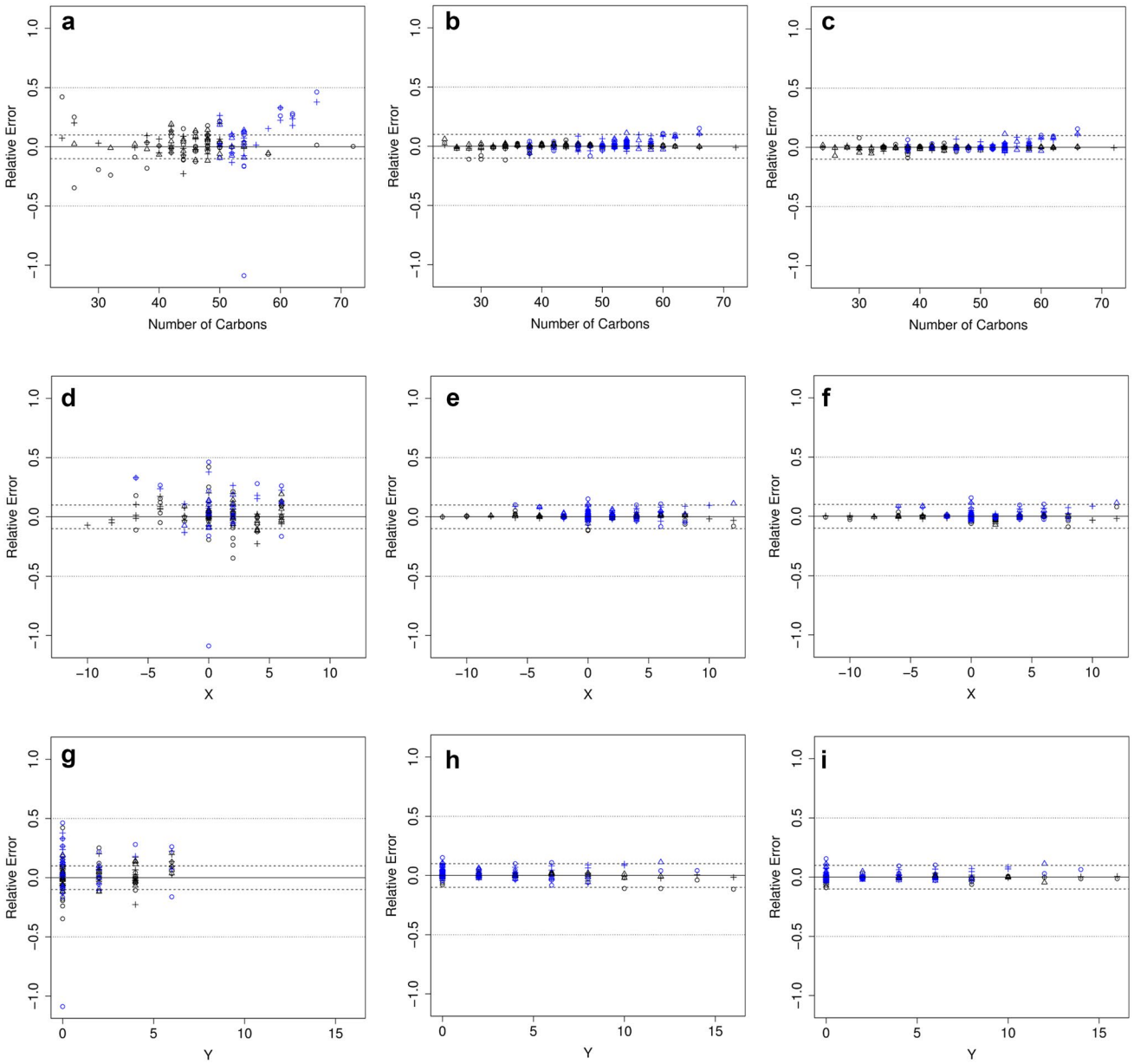

Fig. 3 Summary of fitted quality as a function of number of carbons (*row 1*), chain length difference between fatty acid at *sn*1 *versus* fatty acid at *sn*2 (*row 2*), and chain length difference between fatty acid at *sn*1 and *sn*3 (*row 3*) for predicted enthalpy of fusion (*column 1*), melting temperature predicted using model A (*column 2*), and melting temperature predicted using model B (column 3). *Black markers* indicate measures of saturated triglycerides, and *blue markers* indicate measures of unsaturated triglycerides. *Circle markers* (○) indicate triglycerides in α-polymorph, triangle markers (Δ) indicate triglycerides in β'-polymorph, and cross markers (+) indicate triglycerides in β-polymorph. The *solid black line* in either *plot* indicates "perfect prediction", with *dashed lines* representing bands of errors. In particular, the line closest to the *solid line* indicates a prediction off by $\pm 10\%$, followed by a bands describing predictions off by $\pm 50\%$ (color figure online)

Table 6 Summary of Thermodynamic consistency results

Model	Measure	Category	No. samples	Error coding					
				000	100	010	101	011	111
	Enthalpy	Saturated	1254	0	0	0	0.014	0	0.986
	Enthalpy	Unsaturated	3520	0	0	0	0	0	1
A	Melt. Temp.	Saturated	1254	0	0	0	0	0.314	0.686
	Melt. Temp.	Unsaturated	3520	0.017	0.017	0.187	~0.00	0.456	0.323
B	Melt. Temp.	Saturated	1254	0.093	0.029	0.065	0.081	0.207	0.525
	Melt. Temp.	Unsaturated	3520	0.122	0.065	0.039	0.099	0.218	0.458
		IDEAL	∞	*0*	*0*	*0*	*0*	*0*	*1*

Error code [1,1,1] represents physically expected behavior. Italic values indicate the error codes that can be expected if predicitions were perfectly correct

temperature estimate is more likely to produce over-estimates. A summary of the quality of prediction measures for melting point data is provided as Table 5.

Further evidence of the TPC's ability to characterize experimental data is provided in Fig. 3, where the relative errors associated with estimating enthalpy of fusion (column 1), melting temperature using model A (column 2) and melting temperature using model B, is plotted with respect to the total carbon in the triglyceride (row 1), chain length difference between the fatty acid at *sn*1 *versus sn*2 (row 2), and chain length difference between the fatty acid at *sn*1 *versus sn*3 (row 3). The effect of the noted model inputs (total carbon, chain length differences) is not overtly obvious, suggesting that the results are generally not biased by these attributes. However, we can see that errors in enthalpy estimates are greater for small triglycerides of low total carbon number (less than 30) or large triglycerides of high total carbon number (greater than 60).

When assessing the intra-estimate consistency of estimated enthalpies and melting temperatures of saturated and unsaturated triglycerides, we found that in total, 4756 of the 4774 (>99%) enthalpy estimates behaved in accordance with expectations, whereas only 41% of the melting point estimates using model A were feasible. Melting point estimates found using model B proved only slightly better with 47% of the estimates being consistent with thermodynamic expectation. A closer inspection of the types of errors and in the scope of distinct saturated *versus* unsaturated triglycerides reveals that ~86% of the failed melting point predictions using model A occurred with unsaturated triglycerides, and that the most common error coding is [0,1,1]—the predicted melting point of the β unsaturated triglyceride is less than the β' polymorph, but still greater than the α polymorph. Errors associated with melting points estimated with model B are less slanted towards any particular error coding, rather, showing results that fail in all possible combination of errors. The estimated enthalpies of fusion for all unsaturated triglycerides behaved consistent with thermodynamic expectations. A summary of error coding results for all possible saturated and unsaturated triglyceride combinations in the TPC is provided as Table 6.

To improve clarity of discussion, we begin by discussing the general quality of our findings, and then will continue to describe the clear development priorities identified through this analysis. We will provide insights into development approaches and expected challenges for each of the identified development priorities.

When discussing the "effectiveness" of the TPC in predicting the thermodynamic properties of triglycerides with available experimental values, it is important to recognize the limitation with respect to experimental data availability. Given the current list of fatty acids employed by the TPC, melting temperatures and enthalpies for three polymorphic forms of 4774 triglycerides can be predicted, 1254 of which are saturated and 3520 are unsaturated. In comparison, our experimental database consists of 185 triglycerides, most of which are saturated triglycerides and most of which do not contain complete information for all polymorphic forms of the triglyceride and/or both enthalpy and melting temperature information. The authors of [8] recognized the scarcity of experimental information as a limitation when developing their model, and this problem persists to date. With a growing database of experimentally collected enthalpies and melting temperatures, we can expect that future iterations of the TPC and underlying mathematical models will provide improved accuracy when comparing predictions to measurements.

That said, the value of the present implementation of the TPC is in exploration of its own model limitations as demonstrated through the previously described sequence of simulation experiments and analysis. From this assessment, we have identified numerous unresolved questions about the underlying mathematical model, and have generated priorities around which to focus our future investigations.

The most obvious limitation of the current implementation of the TPC is its inability to predict melting points. This failure is not seen in describing existing experimental data, rather, the melting point predictions fail to be

consistent with thermodynamic expectations. Based on the effectiveness of the TPC in estimating enthalpies for unsaturated triglycerides, we can speculate that the challenges are rooted in the choice of fitting parameters of the form A_i, B_i, A_{ij}, B_{ij}, defined in Eqs. 18, 19, 22 and 23, and listed with values as parameters 12-39 in Table 2.

Let T_α be the melting temperature of an α-polymorph of an arbitrary saturated triglyceride and T_β be the melting temperature of the β-polymorph of the same saturated triglyceride. Our expectation from thermodynamics is that:

$$T_\alpha < T_\beta. \tag{28}$$

Using Eq. 13, we define:

$$T_\alpha = T_\infty\left(1 + \frac{A_\alpha}{n} - \frac{A_\alpha B_\alpha}{n^2}\right), \tag{29}$$

where A_α and B_α are fitting parameters from a Taylor Series expansion, and n is the number of carbons in the saturated triglyceride. Similarly,

$$T_\beta = T_\infty\left(1 + \frac{A_\beta}{n} - \frac{A_\beta B_\beta}{n^2}\right). \tag{30}$$

Substituting Eqs. 29 and 30 into 28 implies the following inequality should hold:

$$\frac{A_\alpha}{n} - \frac{A_\alpha B_\alpha}{n^2} < \frac{A_\beta}{n} - \frac{A_\beta B_\beta}{n^2}. \tag{31}$$

Following the same procedure except with a first order Taylor series approximation of melting temperature, we can show that,

$$A_\alpha - A_\beta < 0, \tag{32}$$

and therefore,

$$\frac{A_\alpha B_\alpha - A_\beta B_\beta}{A_\alpha - A_\beta} < n, \tag{33}$$

where constraint (33) is derived through algebraic manipulation of (31) with the additional constraint of (32). Using this simple example to re-examining the parameters listed in Table 2 and the properties of the triglycerides that failed to be consistent with thermodynamic expectation, we can confirm that the constraint in Eq. (33) was not maintained during original parameter fitting. We should note that maintaining these thermodynamic constraints during parameter fitting, in particular when considering the complexity included in the definitions of parameters A and B in Eqs. (18, 19, 22, 23), is not a trivial task. The challenges are compounded when considering the incomplete data set from which model parameters are found—we have very little data for multiple polymorphs from the same triglyceride. The subsequent use of a constrained optimization scheme to refit experimental data to the model equations may lead to lower quality fits with experimental data, but will be a suitable approach for ensuring predicted melting points within a triglyceride are consistent with thermodynamic expectation.

An alternate approach, which may be appropriate in combination with constrained fitting, would be to decompose the model parameter space and provide different fits for different types of triglycerides. As noted, the model of [5, 8] is the only comprehensive model for describing thermodynamic properties of pure triglycerides, however, expecting a single governing set of equations to successfully capture behaviour of both temperature and enthalpy of all polymorphic forms of triglycerides (saturated, unsaturated, monoacid, symmetric, asymmetric, etc.) may itself have been an overly ambitious task. Even more, sub-modifications within polymorphic forms are neglected completely, which itself may introduce further non-trivial biases in the fitting of model functions. For example, the effects of chain lengths x and y should be vastly different between a β-1 and β-2 polymorph, yet our coarse method treats both scenarios as equivalent. Perhaps fitting enthalpy of fusion measurements separate of melting temperature will itself be a sufficient approach to ensuring predicted melting temperatures maintain physical feasibility. The challenge with decomposing the fitting process will be having an adequate amount of data available for all of the resulting subcategories of triglycerides.

A final but important modeling question that must be addressed, both in the underlying mathematical modeling approach and in our post simulation analysis, is the scaling difference between melting temperatures and enthalpies of fusion. Due to the difference in magnitude (often a single order of magnitude) between enthalpy of fusion and melting temperature measurements, it is difficult to compare errors tabulated for each measurement with each other. For example, if we assume a triglyceride has a measured enthalpy of fusion of 100 kJ/mol and a measured melting point of 280 K, if the predicted enthalpy of fusion is 120 kJ/mol and the predicted melting point is 290 K, our analysis would suggest that the enthalpy prediction is weak (RE = 20%) and that the melting point is strong (RE = 3.5%). If we use another set of parameters to describe the same TAG which result in a predicted enthalpy of 110 kJ/mol and a melting temperature of 270 K, our analysis would suggest that both the enthalpy prediction (RE = 10%) and melting temperature prediction (RE = 7%) are fairly strong. The difference between both predictions may, for practical purposes, be trivial, but would have posed significant challenges during the initial parameter estimations in the work of Wesdorp *et al.* [5, 8]. We might speculate that the previously described decomposition of the fitting process to treat melting temperatures and enthalpies of fusion separately will sufficiently reduce scaling effects, however cannot be certain without a structured analytical approach. For now, we

suggest that further analysis of model output (melting temperature and enthalpy) sensitivity to parameters is needed immediately, as is a thorough review of prediction quality measures (relative error, quality of fit) as a result of model outputs.

Conclusions

The Triglyceride Property Calculator (TPC) is a software implementation of the semi-empirical model developed in [5] and described with improvements in [8, 9]. Rigorous testing of the calculator made apparent that the model is capable of estimating enthalpies of fusion with measurable success for both saturated and unsaturated triglycerides of all polymorphs, and particularly strong predictions of measurements for β saturated triglycerides. Testing also made evident that the current implementation of the Triglyceride Property Calculator has suspect value in predicting melting temperatures of triglycerides, often generating predictions that are physically unlikely. In spite of these issues, the TPC is an important preliminary advance in what will surely be a continued effort to generate quality estimates of thermodynamic properties for pure triglycerides. With continued, rigorous development, we expect to eventually build a tool, or tools, capable of estimating thermodynamic properties of mixed systems of multiple liquid and solid triglycerides; a development of significant value to scientists and engineers investigating naturally occurring fats and oils.

Acknowledgements This research program was funded by the Natural Sciences and Engineering Research Council of Canada.

References

1. King AM, Garner WE (1936) Melting points of long-chain carbon compounds. J Chem Soc 1936:1368–1372. doi:10.1039/JR9360001368
2. Bailey AE (1950) Melting and solidification of fats. Interscience Publishers, New York
3. Timms RE (1978) Heats of fusion of glycerides. Chem Phys Lipids 21:113–129. doi:10.1016/0009-3084(78)90059-2
4. Ollivon M, Perron R (1982) Measurements of enthalpies and entropies of unstable crystalline forms of saturated even monoacid triglycerides. Thermochim Acta 53:183–194. doi:10.1016/0040-6031(82)85007-7
5. Wesdorp L (1990) Liquid-multiple solid phase equilibria in fats: theory and experiments. Doctoral Dissertation. TU Delft, Delft University of Technology
6. Hagemann JW, Rothfus JA (1983) Polymorphism and transformation energetics of saturated monoacid triglycerides from differential scanning calorimetry and theoretical modeling. J Am Oil Chem Soc 60:1123–1131. doi:10.1007/BF02671340
7. Hagemann J (1988) Thermal behavior and polymorphism of acylglycerides. In: Garti N, Sato K (eds) Cryst. Polymorph. fats Fat. acids. Marcel Dekker, New York, pp 9–95
8. Wesdorp L, Van Meeteren J, De Jon S et al (2012) Liquid-multiple solid phase equilibria in fats: theory and experiments. In: Marangoni A, Wesdorp L (eds) Struct. Prop. Fat Cryst. Networks, 2nd edn. CRC Press, Boca Raton, pp 241–418
9. Liu R (2014) Thermodynamic calculations of multicomponent phase compositions. Masters Dissertation. Dalhousie University
10. Marangoni A, Wesdorp L (2013) Structure and properties of fat crystal, 2nd edn. CRC Press, Boca Raton
11. R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
12. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J (2016) shiny: Web application framework for R. R package version 0.14.2. https://CRAN.R-project.org/package=shiny