University of Alberta

Study of the phase behavior of triacylglycerols using molecular dynamics simulation.

by

Paulina Szewczyk

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
In
Chemical Engineering

Department of Chemical and Materials Engineering

Fall 2010
Edmonton, Alberta

Examining Committee

Dr. Phillip Choi, Professor, Department of Chemical and Materials Engineering

Dr. Hao Zhang, Assistant Professor, Department of Chemical and Materials Engineering

Dr. Tian Tang, Assistant Professor, Department of Mechanical Engineering

To my best friend and husband, Hubert.

Abstract

In the present work, we focused our attention on triacylglycerols. Their phase behaviour strongly influences production processes of products based on fats and oils. However, the mechanisms controlling such behaviour are not well understood. Hence, we decided to utilize computer simulation to gain more understanding in this matter. Using molecular dynamics simulation, we tried to mimic the thermal transition of triacylglycerols from liquid to solid state in order to shed more light on the crystallization of these species.

The main conclusions are that due to the large time scale of the crystallization process, computer resources and time available for this work were not sufficient to simulate the full phase transition of triacylglycerols. Nevertheless, we managed to observe the first structural changes that drive the crystallization, i.e., the strong tendency of oxygen atoms to group and isolate themselves from the hydrocarbon chains. The successfully created molecular models are now ready for further investigations.

Acknowledgements

I would like to take this opportunity to thank my supervisor Dr. Phillip Y.K. Choi for introducing me to the world of molecular modelling. I highly appreciate the guidance he provided in my research and also the trust he put in my ability to pursue my career as a MSc student in Chemical Engineering. Special thanks also go to Dr. Suresh Narine of Trent University for presenting me with the problem of fats crystallization and for his on-going support in that matter.

This work wouldn`t be possible without the kind help I received from Dr. Pieter Tieleman of the University of Calgary and his graduate students, who introduced me to the world of GROMACS software. Particularly, my thanks go to Cameron Laing for his advice, time and bottomless patience.

My last thanks go to the Westgrid consortium for providing the high performance computing, its facilities and great support service to my research project. Definitely, molecular modelling performed on the scale described in this work would not be possible without Westgrid resources.

Table of contents

List of tables

List of figures

List of symbols

α, β', β, β-2, β-3, β'_2	types of triacylglycerols polymorphic structures
σ	separation distance
λ	time-dependent factor
ξ	an improper torsional angle
φ	angle between *ijk* and *jkl* planes
ρ	density
γ	phase factor
ε	dielectric constant
π	Pi constant
θ	angle
τ	a constant
τ_p	pressure time constant
τ_T	temperature time constant
μ	scaling matrix
ω	torsion angle
Å	Angstrom
a, b	molecules
C_V	heat capacity
d	derivative
E	energy
i,j	particles
k	constant
K	Kelvin degrees
l	length
F	force
m	mass
M	diagonal matrix
n	multiplicity
N	number of point charge
P	pressure
q	atom`s charge
r	distance between particles
t	time
T	temperature
v	potential energy
x, y, z	coordinates
W^{-1}	coupling strength
%	percentage

List of abbreviations

cos	cosine
g/cm^3	density unit; gram per cubic centimetre
g(r)	radial distribution function
fs	femtosecond
ns	nanosecond
ps	picosecond
ref	reference
CG	coarse-grain
kJmol^{-1}nm^{-1}	force unit; kilo Joule per mol x nanometer
MD	molecular dynamics
MSD	mean square displacement
NPT	isothermal-isobaric ensemble
NVE	microcanonical ensamble
NVT	canonical ensamble
PPP	tripalmitin
RDF	radial distribution function
RMSD	root mean square deviation
SSS	tristearin
TAG	triacylglyceride, triacylglycerol

1. Introduction

Lipids, an abundant group of organic compounds, are essential constituents of life. They are the best energy source and storage and they help to shape bodies of all living organisms and the world they live in. The name *lipids* itself is often used interchangeably with the name *fats* [1]. This is actually a little misleading. The term "fats" refers to the edible fats. However, the group of lipids includes also a whole range of hydrocarbons from simple alkanes to phospholipids. For the purpose of this work, we turn our interest to the edible fats (i.e., solid fats at room temperature) and oils (i.e. liquid fats at room temperature). They are actually a much diversified subgroup of lipids themselves. They make up 99% of the lipids of animal and plant origin [2] and of these 95% consist of triacylglycerols (TAGs) [3].

Triacylglycerols, being naturally existing renewable resource of energy and matter, have been raising the interest of the industrial and scientific worlds for a long time. Our ancestors were using them as food and as fuel or even medicine, and this trend continues to grow at present. With the development of civilisations and technologies, fats have significantly gained on importance and application. Their production rate, following the consumption of TAGs mixtures, in recent years has dramatically overgrown the forecasts. It was predicted that in the period of 2003-2007 the average annual fat supply would be about 104 metric tonnes. However, only in the 2004-2005 period, it reached 136 million tonnes [4]. Today, natural edible fats and oils present a galore of applications and use. On a daily basis, throughout the world fats/TAGs are used in fat- and oil-based spreads, shortenings, baking fats, deep-frying fats and oils, dressings and many more [5]. They are widely used in cosmetics and medications [6], detergents, lubricants, bio-fuels and surface coatings [4]. Moreover, the industry has started recognizing structured lipids (i.e., chemically or enzymaticaly modified lipids), which provide special properties to their products, as food ingredients. Also, the nations-wide drive to improve human health via more natural than pharmaceutical ways and the recognition of the relation between fat-based diet

and health and disease, has given a birth to a new generation of lipids. They are called *the neutraceutical lipids*. According to C.C Akoh [7], they are species "isolated, concentrated or synthesized/modified from natural fats and oils and they provide medical or health benefits, including the potential for the prevention and/or treatment of disease". TAGs derivatives have also been discovered to be responsible for complex cell to cell interactions and to be the essential bio-membranes building-blocks in all living organisms. In addition, fats provide humans and animals with essential fatty acids (crucial for their development yet not synthesized by their bodies) and to consume vitamins soluble only in fats [4]. With growing interest and demand for fats and oils, there is definitely more interesting findings and applications to come.

All those TAGs-based products have something very important in common – their quality and properties are almost totally dependent on the TAGs structure. The reason for this lays in TAGs crystallization behaviour, which is controlled by the chemical composition and processing conditions of TAGs mixtures. A good example here are yellow spreads, i.e. butter, margarine etc. They must consist of enough oil at refrigerator temperature to make them spreadable. However, at ambient temperature they have to contain enough solid fat to keep them from pouring and oiling out. What is also important from consumer point of view is a good taste. It needs to be assured by making the spread to be liquid at mouth temperature [3]. Another example, that many of us are familiar with, is chocolate and its essential ingredient – the cocoa butter. This ingredient is naturally a complex TAGs mixture. To produce a good quality chocolate, the cocoa butter must have the most stable crystal packing. If this condition is not satisfied, the final product`s taste, mouth-feel, appearance and shelf-life will be highly compromised [8].

One may think that, since we know about and have worked with fats and oils for centuries, the knowledge about them should be well established. On the contrary, regardless a vast amount of studies to date, the nature of TAGs crystallization is far from being understood and it`s mechanism is still not clear. There simply had not been the right tool invented to look into phase changes and their mechanisms in lipids. Also,

the compounds themselves proved to be challenging to work with. The main reasons behind it have been the composition of TAGs (easy interestrification), the abundance of TAGs crystal structures (complex polymorphic behaviour) and its implications (i.e. viscosity changes, a melting range instead of sharp melting temperature), as well as the thermal instability of fats and oils [3]. In this information age, all sciences have flourished bringing huge changes in the way people think and live today. It's the technological revolution that sent humanity to the Moon and brought the "all-mighty" computers to our homes. Therefore, it would seem logical to think that the problems encountered in the science of lipids crystallization should be now possible to overcome. Unfortunately, it is not the case. The computational resources available, though vastly improved in the last several decades, still limit researchers in their explorations. There are now numerical methods to simulate almost up to 1 microsecond of lipids interactions [9], but they are still far from simulating the phenomena on fats and oils crystallization time scale. There is one branch of the field of lipid computer simulations that has been particularly successful, meaning the bio-membranes computer simulations [9-15]. However, it focuses on lipid species which compose biological cell membranes, such as dipalmitoylphosphatidylcholine, and is mainly of medical and pharmaceutical intrest. Hence, the edible fats and oils are still left on a side of this new emerging field of computational chemistry.

Luckily, W.F. van Gunsteren and his co-workers were probably the first people who showed interest in triacylglycerols simulations. They have taken on a challenge of implementing a method that had already been successfully used in simulations of bio-molecules [16] to aliphatic hydrocarbons. This later led them to molecular simulations of simple triacylglycerols in their solid phase [17]. Apart from this group, to my knowledge, there were very few scientists interested directly in computational calculations including TAGs [18-19].

Having said all the above, the goal of this work was to gain deeper understanding of TAGs phase behaviour on a molecular level by means of computer simulation. It was done to predict the state and properties of TAGs and their mixtures.

Following the work of van Gunsteren et al., the tool of choice was molecular dynamics simulation. The insight coming from below micro-scale simulations is hoped to shed more light onto order-disorder phase transformations of TAGs that are closely related to the onset of edible fats and oils crystallization phenomena.

2. Triacylglycerols

2.1 TAGs and their phase behaviour

Fats and oils, as previously stated, are versatile and interesting yet challenging subgroup of lipids to study and work with. The intricacies of fats and oils will be explained in this chapter.

Triacylglycerols, also known as triglycerides and TAGs, are built from three fatty acids (FA) esterified to a glycerol backbone (Fig.1) [20]. Common practice in the science of lipids is to name TAGs by using abbreviations for their FA constituents. For instance, the name PPP stands for tripalmitin – a triglycerol substituted with three palmitic acids [21]. The length of FA chains in most common fats and oils varies from 12 to 24 carbons. TAGs can consist of one type of FA (monoacid TAG) or of different FAs (mixed acid TAG). In both cases, saturated (no C=C bond in FA chain) and unsaturated (with a minimum of one C=C bond in the FA chain) fatty acids are possible. The C=C bond(s) allow the formation of *cis*- and *trans*- conformations [22]. Many of those complex mixtures, such as milk fat, are usually composed of more than 500 different types of TAGs [3].

It is worth mentioning here that the composition of fats and oils varies depending on their origin as well. The cocoa butter, for example, supplied from West Africa will have a different TAG composition than the one from India [4]. This obviously would add complexity to TAGs phase behaviours.

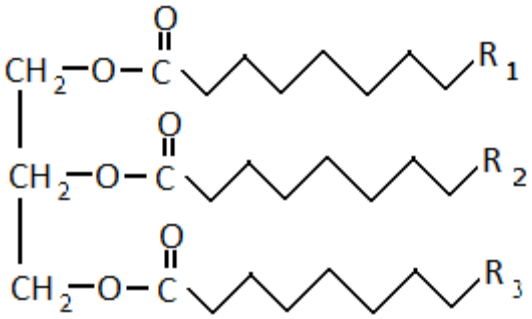

Fig.1. Triacylglycerol basic structure (TAG).

Furthermore, the diversity of fats and oils mixtures is being ascribed to an interestrification process to which TAGs are prone [1] [23]. Therefore, whenever two different TAGs are mixed together it is more than likely that the final product consist of a several new TAGs. Unfortunately, this reactivity becomes a significant obstacle when one wishes to analyze properties of pure TAG sample that is coming from a mixture. On this note, it becomes clear that when dealing with fats and oils being mixtures of interestrificated triacylglycerols, usually only determination of a melting range instead of melting temperature is possible. Even very small appearance of a different than intended TAG in a sample can change its crystallization behaviour [3]. This in turn makes further analysis and calculations less accurate.

On top of the above, individual TAG molecules like to organize themselves, meaning they crystallize. They pack themselves in two configurations (Fig.2):

- tuning fork – the acyl chain in position 2 is left alone and the chains in 1 and 3 positions pack alongside

- chair – the acyl chain in position 2 is alongside the chain on either 1 or 3 positions; those „chairs" can stack in a double or triple chain length structure.

Depending on their composition and various external conditions (e.g. temperature, pressure, time), TAGs crystallize in many ways, taking on different crystal forms i.e., expressing polymorphism. This polymorphic behaviour is the most challenging aspect of dealing with all fat- and oil-based products. Small changes in the composition and/or the conditions of crystallization affect strongly the polymorphic form of the fat, which then changes the final product properties. As a result, the characterization and/or production following the crystallization process become more difficult. The Fig.3 shows the structure hierarchy of fats along with their size scale.

To determine crystal structures of TAGs, the powder X-Ray diffraction analysis has been used [3]. The observed polymorphic structures are (Fig.4a):

- α - with hexagonal arrangement of acyl chains and XRD pattern near 0.42 nm; this is the least stable polymorph (<60 sec stability), occurring during process,

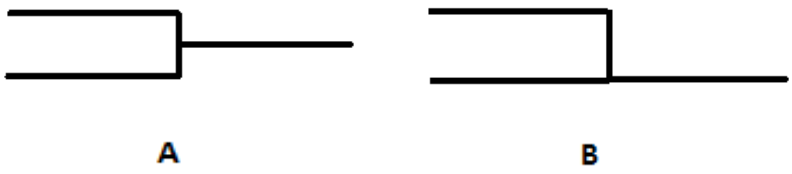

Fig.2 Packing conformations of TAGs:
a) tuning fork, b) chair.

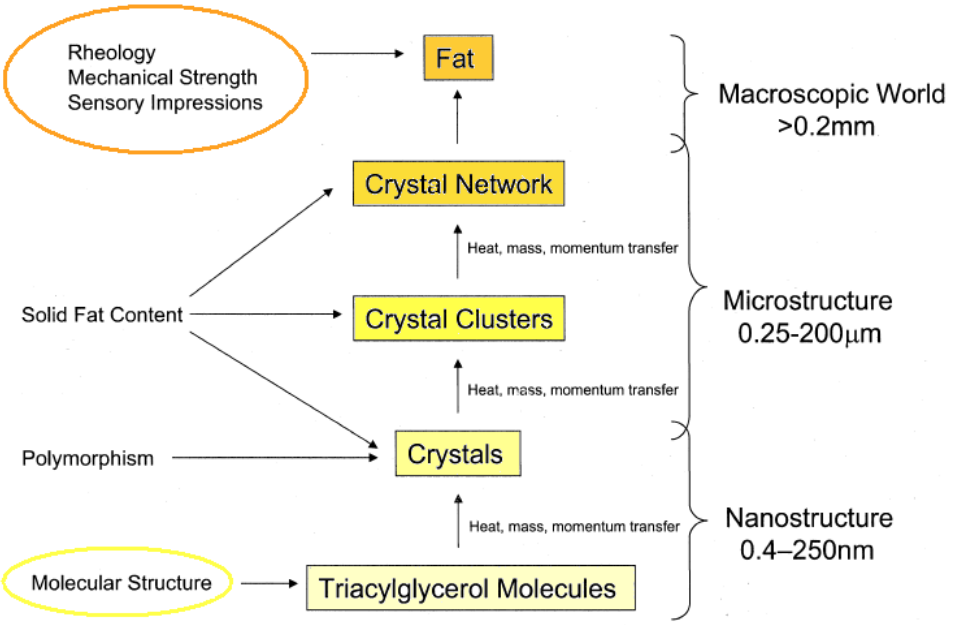

Fig.3. The structure hierarchy of fat crystal networks and their size scale [3].

- β'- with perpendicular orthorhombic acyl chain packing and XRD pattern at 0.42-0.43 nm and 0.37-0.40 nm; this is a metastable state (60 sec →years) occurring in products,
- β - having triclinic parallel acyl chain arrangement and strong XRD pattern at 0.46 nm; this is the most stable polymorph [3][24].

Also β-2, β-3 sub-modifications have been observed and are shown in Fig.4b [3][25]. However, these are definitely not the last ones reported. Many scientists keep working on the characterization of fats phase behaviour and more sub-modifications are expected to be discovered.

The phase transitions from the least to the most thermodynamically stable states are shown below (Fig.5). Their transition paths are analyzed with Differential Scanning Calorimetry tools and described in a form of cooling and heating thermo peaks (Fig.6). Also the relationship between the Gibbs energy and three most common TAG polymorphs can be seen in Fig.7.

The kinetics of TAGs transformations has been less explored. However, new methods of analysis, such as Synchrotron Radiation X-Ray Diffractometry (SR-XRD)(Fig.8) and molecular dynamics simulations, have been recently developed and more and more insights are being provided on a molecular level and kinetic characteristics of fats [25].

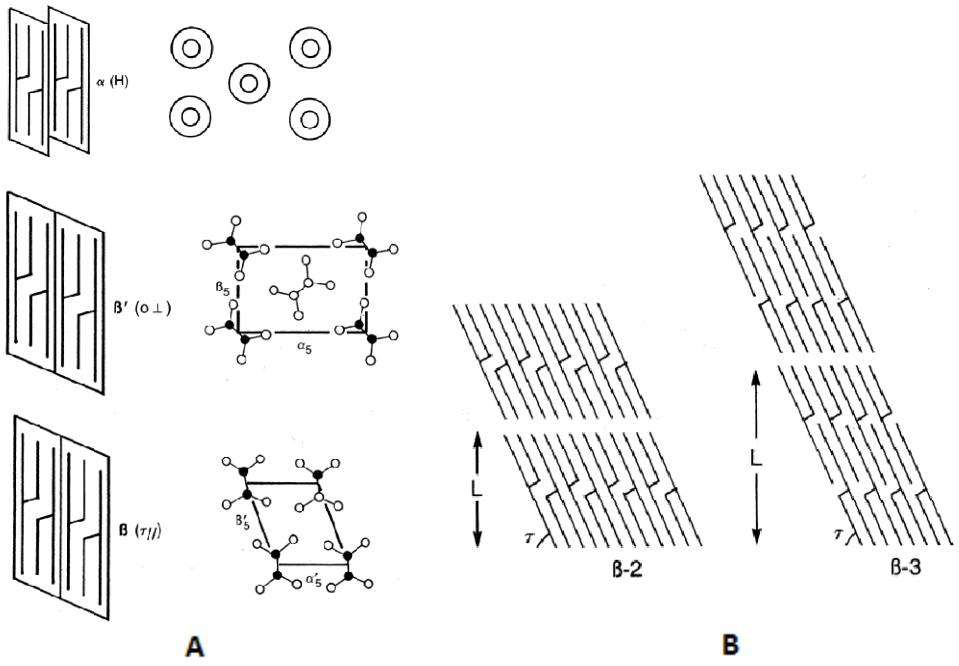

Fig. 4. a) Three basic polymorphic crystal structures of TAGs, b) β-2 and β-3 arrangements of TAGs [3].

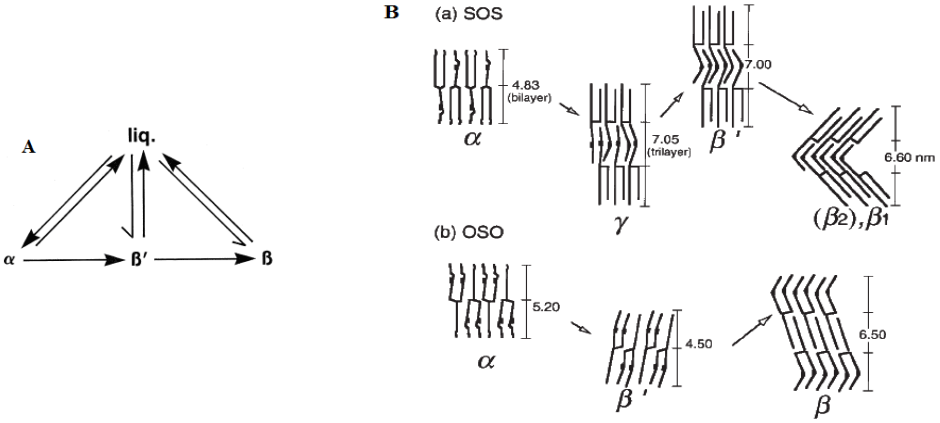

Fig.5. a) Possible polymorphic phase changes in fats [3], b) Polymorphic transformations of SOS and OSO; the lamellar thickness of each polymorph is given, based on X-ray diffraction data [26].

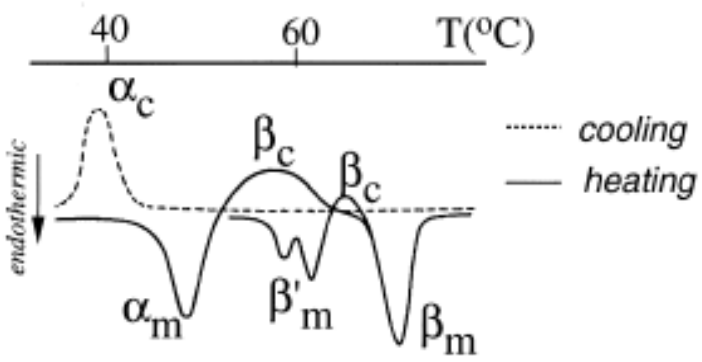

Fig.6. DSC thermo peaks for tripalmitin (PPP) [25].

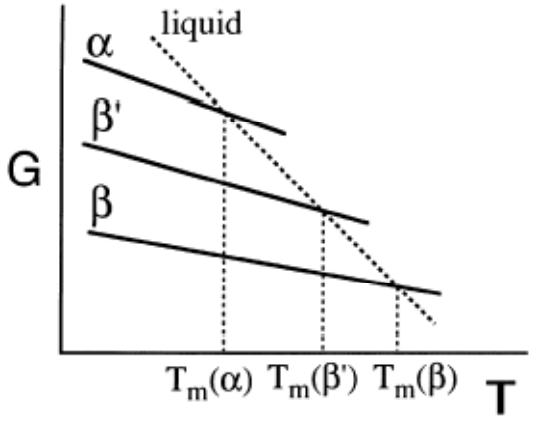

Fig. 7. Gibbs free energy relationship with temperature for α, β` and β polymorphs [25].

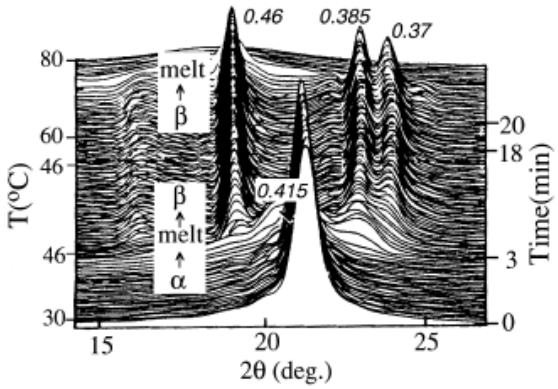

Fig.8. SR-XRD spectra of tripalmitin (PPP) α→β phase transition [25].

2.2 Tripalmitin and tristearin as model triacylglycerols.

In this work, as the subject of investigation two TAGs have been chosen – tripalmitin (Fig.9) and tristearin (Fig.10) – two TAGs abundantly occurring in almost all oils and fats anywhere in the world [4]. Our choice was based on their relative structural simplicity, high importance and vast occurrence in nature [4] and scientific literature.

Tripalmitin ($C_{51}H_{98}O_6$), or 1,2,3-trihexadecanoyl-sn-glycerol, also called PPP, with molecular weight of 807.32 $gmol^{-1}$ [27].

PPP crystallizes in three main forms: α, β', β. However for the β' and β-polymorphs many sub-modifications have been discovered. For instance, the β'_2 appears both on cooling and heating a melt, at 49 ^oC and 52 ^oC respectively. Also, it has been reported a sub-α-polymorph, appearing only at extremely low temperature of -100 ^oC [28].

Tristearin ($C_{57}H_{110}O_6$), or 1,2,3-propanetriyltrioctadecanoate, also called SSS, with molecular weight of 891.48 $gmol^{-1}$ [27]. It crystalizes in the α, β'_2 and β-polymorph. Due to its longer chain and hence slower crystalization kinetics, the β'_2 modification occures only on heating at 56 ^oC [28].

Both, PPP and SSS have been thoroughly analyzed via most of the analytical methods available. Also, since there was already some computational data on PPP and SSS molecular interactions available [18-19], we decided to use it for validation of obtained results in chapter 4.

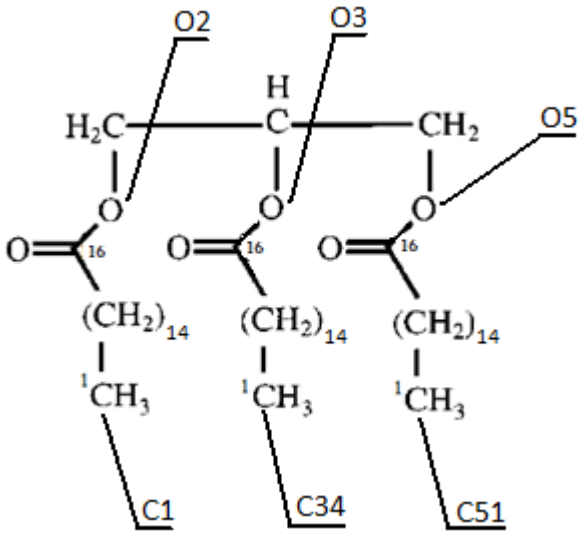

Fig.9 Tripalmitin (PPP, $C_{16}C_{16}C_{16}$);
O2, O3, O5, C1, C34, C51 – atom names used in MD simulations.

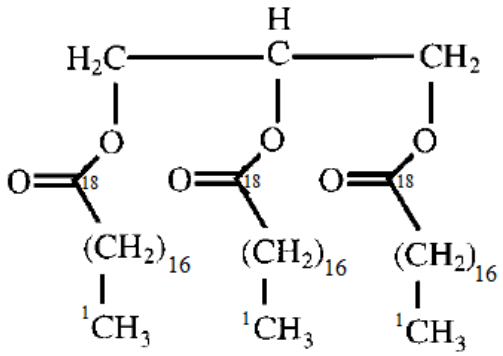

Fig.10 Tristearin (SSS, $C_{18}C_{18}C_{18}$).

3. Molecular dynamics simulation

3.1. General concepts in the computational approach

With the development of computational resources and numerical and statistical methods in the 20th century, molecular modelling has evolved into a separate, well established branch of modern chemistry. New powerful computers have increased the range of models and phenomena that now can be investigated, becoming practically irreplaceable tool of modern chemical analysis in the meantime. Molecular modelling is a discipline aiming in mimicking the behaviour of molecules and molecular systems subjected to various conditions and calculating and/or predicting their properties [29]. There are two groups of methods available for analysis in molecular modelling:

- *quantum mechanical* – treats electrons explicitly thereby enabling calculation of properties depending on the electronic distribution (e.g. properties determining and/or relaying upon forming and breaking of chemical bonds),
- *molecular mechanics* – calculates the energy of a molecular system as a function of nuclear positions and is based on a concept of a force field that will be discussed in more detail later in this chapter; there are two ways of calculations available:
  - *molecular dynamics* – based on time-averaging of the properties of a system and the idea that the state of the system at any time can be predicted from its current state; includes contributions from both the kinetic and the potential energy functions; this approach has been applied in computer simulations of this work and will be described further below;
  - *Monte Carlo (Metropolis)* simulation methods – successive steps of simulation depend only on the proceeding step; the total

energy of a system is directly calculated from the potential energy function; if the energy of a new randomly generated configuration is lower than the energy of the one from the immediate previous step, then the new configuration is accepted; however, if the energy is higher than the one of its predecessor, then the Boltzman factor of the energy difference is calculated and compared with randomly generated number from 0 to 1- if the random number is higher than the Boltzman factor, then the move is rejected and the original move is retained for the next simulation step [29]; This approach will not be discussed in more detail as it is not the subject of this work.

3.2. Molecular modelling with MD

Many of the problems of interest to scientific minds in the field of molecular modelling are on a nanoscopic scale (e.g. protein folding [30], bio-membrane phase behaviour [31][32], micelle formation [33]). This means that information coming from electronic interactions would be too detailed and too computationally expensive if quantum mechanical methods were used to solve them. By choosing molecular dynamics techniques we do scarify the information on the electronic level, but yet we gain a whole range of data on a nanoscopic scale, such as properties of a large collection of molecules, representing a bulk system. Naturally, expressions like "a large collection of molecules" or "a bulk system" are relative but they will be explained shortly.

3.2.1. The size and time-scales of MD simulations

Atomistic molecular dynamics simulations have started over 40 years ago and were first used in the analysis of simple liquids such as argon [34]. Back then, systems

consisting of about 864 atoms and lasting for up to 100 ps could be investigated by means of computers. In the nineties, the situation improved and calculations were conducted on collections of molecules in the order of 1500 lipid atoms. Today, on a daily basis, researchers can simulate systems consisting of several hundreds of lipid molecules (10 nm length scale, approx. 150 atoms/lipid) for tens of nanoseconds [35]. However, it is important to remember that many of the bio-membrane simulations include lipid interactions with water molecules presented in their surroundings. Simulation of that kind of system proves to be computationally costly. Furthermore, to mimic the behaviour of a bulk system use of small unit cells subjected to periodic boundary conditions is preferred (see section 3.2.6.).

3.2.2. Molecular model description

Apart from the all-atom approach, two other modelling approaches are also implemented. First, it is the **united-atom model**, where non-polar hydrogen atoms are included into carbon atoms creating united CH, CH_2 and CH_3 atoms (Fig.11). The second one is the **coarse-grain model**. Here, those united atoms are collected in groups of 3 or 4 called „beads" (Fig.12). This way, researchers save up on the computational time and resources, still being able to obtain reliable data for bigger systems [35]. It is clear that for many cases, the coarse-grain model is too general for the calculation of certain properties, but it is up to the researcher to recognize which model description is accurate enough.

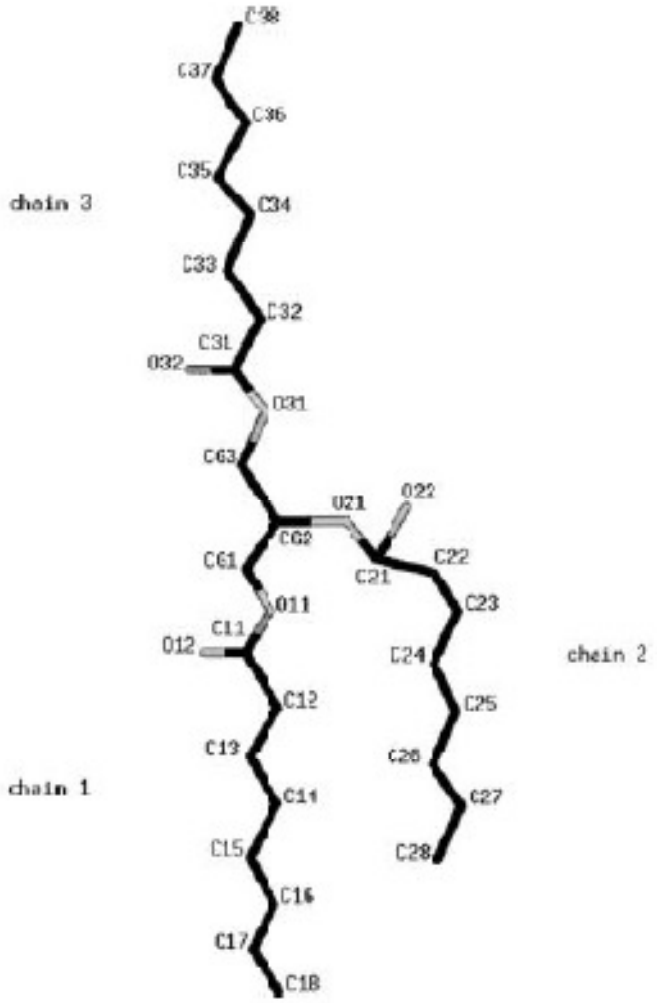

Fig. 11. A representation of trioctanoin in united-atom model [16].

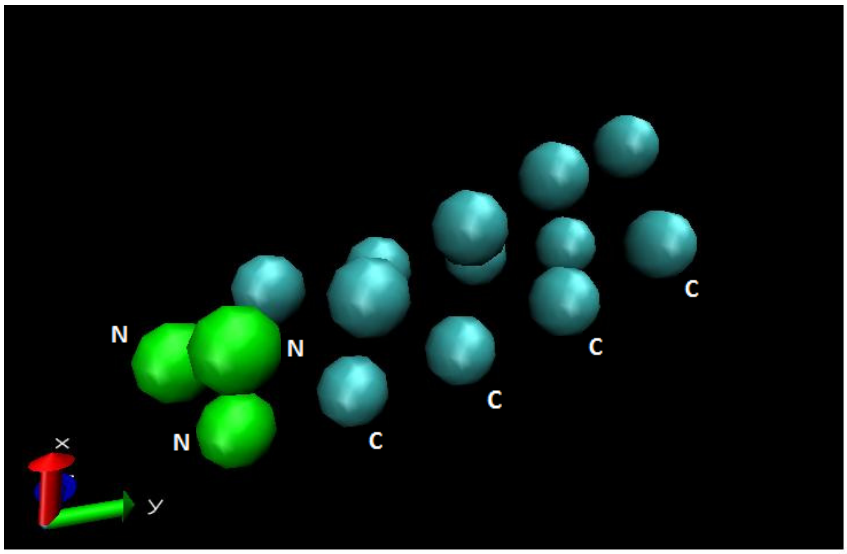

Fig. 12. A representation of one PPP molecule in the coarse-grain model.

3.2.3. Force field

Some say that mathematics is the mother of all sciences. This I think, in particular, is true in the field of chemistry. Mathematical modelling (i.e., describing molecules and their behaviours via numbers), has already been proven to deliver many interesting insights.

The **classical MD** involves the stepwise integration of **Newton's equation of motion** (F=ma) from a given starting point. The resulting trajectory specifies how the positions and velocities of the molecules in a system vary with time. According to Newton's law's of motion, the force acting on atom *i* can be computed directly from the derivative of the force *F* with respect to the coordinate x_i:

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

(Eq. 1)

where m_i is the mass of a particle, and x_i is one coordinate along which the particle is moving [29]. When the initial coordinates and velocities are known, the classical equation of motion allows determining the coordinates and velocities at a later time. It is important to remember that many problems treated with classical MD, including the subject of this work, are the simplest case to which Newton's equation of motion can be applied [29].

Molecular mechanics, and so molecular dynamics, analyzes interactions between molecules based on the stretching of bonds, opening and closing of angles and rotations about a bond. To do that, a concept of a **force field** was created. It is an equation calculating the total energy of a whole system as a sum of energies coming from different interactions in the system (Eq.2). Each energy contribution is described as a function of nuclear positions of atoms. It includes the energy terms from all **interactions** in a system including **bonded** (bond stretching, angle bending, torsions),

non-bonded ones (i.e., electrostatics, van der Waals) and other terms (i.e. cross-terms) describing the molecular interactions characteristic to certain systems.

$$E_{total} = E_{bon} + E_{angle} + E_{dihedral} + E_{Lennard\text{-}Jones} + E_{van\ der\ Waals} + E_{cross\text{-}terms} \quad \text{(Eq.2)}$$

Use of such an equation requires its parameterization and validation against experimental data available and it is usually a representative of a certain type of molecules, such as lipids. However, once it is developed and tested on a small number of cases, it is then ready to be used to many more different problems and for bigger molecules than the tested ones. One should still bear in mind that the force fields are empirical and there is no one correct form of a force field. They are often a compromise between computational efficiency and the accuracy of calculations [29].

Most force fields use an **atom type** idea. It is needed to describe the molecule of interest, because it contains information about atom`s atomic number, hybridisation state and its surroundings. It also determines a parameter set that will be assigned to an atom. For example, the carbon atom in a carbonyl group is distinguished by a force field from the carbon in a carboxylic group and it is described by a different atom type and set of parameters.

Bond stretching, in a basic approach, is described by a Hooke`s law where the energy varies with the square of the displacement from the reference bond length l_0:

$$v(l) = \frac{k}{2}(l - l_0)^2 \quad \text{(Eq.3)}$$

The reference bond length l_0 is the length of the bond when all other terms in the force field are set to 0. The potential energy curve for a typical bond is shown in Fig.13.

Angle bending (Eq.4) (Fig.14) is also often described with Hooke`s law or a harmonic potential:

$$v(\theta) = \frac{k`}{2}(\theta - \theta_0)^2 \quad \text{(Eq.4)}$$

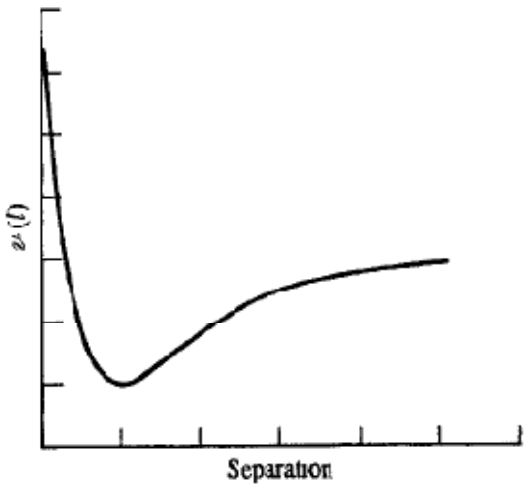

Fig.13. The potential energy curve of a bond with inter-atomic separation.

Here, the interaction contribution of each angle θ is described by a force constant k` and a reference value θ_0.

Higher order terms can be included in both the bond stretching and the angle bending for certain cases. This can serve improving the accuracy of the force field and then the general form for the angle bending term (Eq.5), for instance, becomes:

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2[1 - k'(\theta - \theta_0) - k''(\theta - \theta_0)^2 - k'''(\theta - \theta_0)^3 \dots]$$

(Eq.5)

Most of the impact on a structure and relative energies comes from the relationship between the torsional and the non-bonded contributions. The existence of barriers to rotation about chemical bonds is crucial to understanding the structural properties and conformational analysis. The energy changes caused by a bond rotation are often explained with the three minimum energy staggered conformations and three maximum energy eclipsed structures of ethane. The energy barrier to rotation is believed to arise from anti-bonding interactions between the hydrogen atoms on opposite ends of the molecule; they are minimised when the conformation is staggered and are at the maximum when the conformation is eclipsed [29].

Torsional (dihedral) potentials (Eq.6)(Fig.15) are usually defined with a cosine expansion:

$$v(\omega) = \sum_{n=0}^{N} \frac{Vn}{2}[1 + \cos(n\omega - \gamma)]$$

(Eq.6)

where ω is the torsion angle, n is the multiplicity - gives the number of minimum points in the function when the bond is rotated through 360°, γ is the phase factor determining where the torsion angle passes through its minimum value.

In certain cases, such as simulations of TAGs molecules, it is necessary to include more bending terms. The TAG molecule has three ester carbons which have been shown experimentally to remain in the plane of their ester moiety. To keep them planar in MD simulation it is vital to include „improper" torsion angle term.

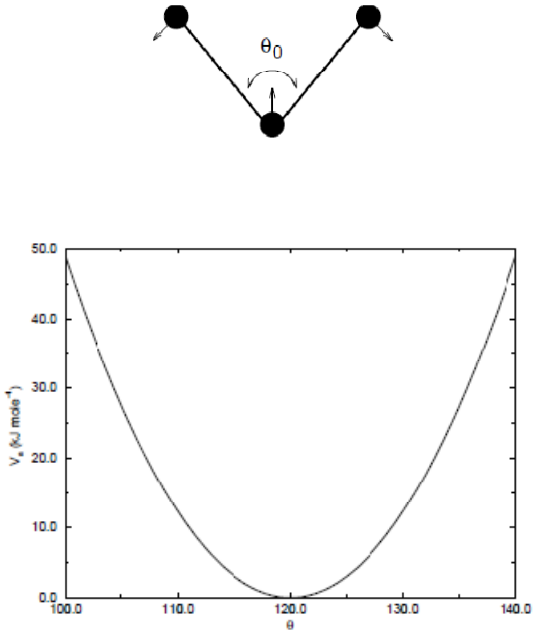

Fig.14. The bond angle vibration principle (top) and angle potential curve (bottom)[36].

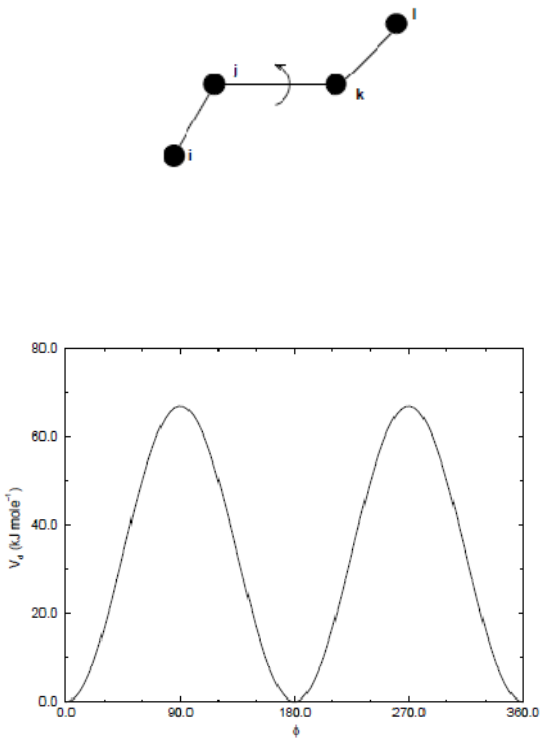

Fig.15. The principal of proper torsional angle and its potential [36].

The **improper torsional (dihedral) angle ξ** (Fig.16-17) is defined as the angle between planes (i,j,k) and (j,k,l) (Fig.16) and described by a harmonic potential (Eq. 7)[36]. The torsional potential is then used to maintain the improper torsion angle at 0^o.

$$v_{id}(\xi_{ijkl}) = k_{\xi}(\xi_{ijkl} - \xi_0)^2$$

(Eq.7)

Finally, the force field often includes so called **cross terms**. They account for coupling between the internal coordinates. They were found to be particularly important in cases of predictions of vibration spectra. Other example of using a cross term is to keep a benzene ring planar. The cross terms are expressed as functions of two or more internal coordinates (e.g. stretch-stretch, stretch-bend, stretch-torsion, bend-bend-torsion) [29].

Non-bonded interactions depend on forces between atoms from independent atoms and molecules and are described via functions of distance between them.

The **electrostatic interactions** between two molecules (or two different parts of one molecule) are represented by Coulomb`s law (Eq.8) as a sum of interactions between pairs of point charges. Fractional point charges are arranged throughout the molecule and they reproduce the electrostatic properties of the molecule. If the charges are assigned only to nuclear centers, they are referred to as partial atomic charges or net atomic charges [29].

$$V = \frac{1}{2}\sum_{i=1}^{Na}\sum_{j=1}^{Nb}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

(Eq.8)

Here, N_a and N_b are the numbers of point charges; ε_0 is the dielectric constant, q_i and q_j are charges of atoms *i* and *j*, and *r* is the distance between them.

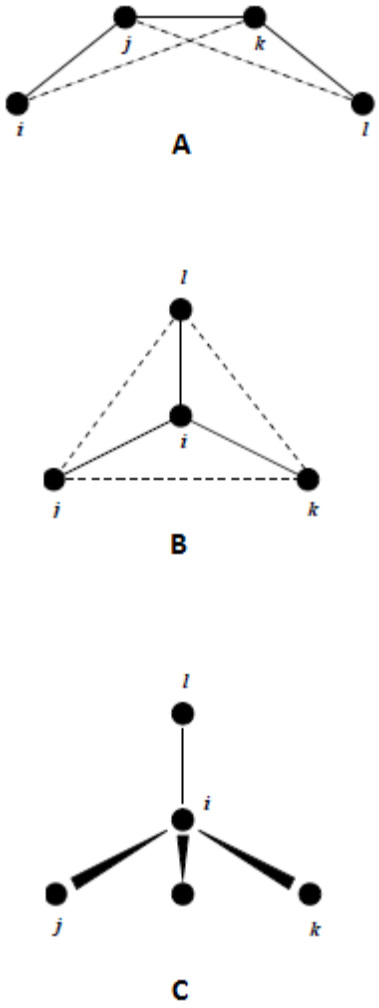

Fig.16. The improper torsional angle concept: a) out-of-plane bending for rings, b) substituent of rings, c) out of tetrahedral [36].

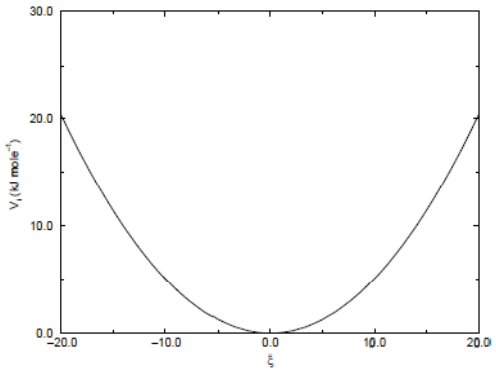

Fig.17. The improper torsional potential curve [36].

The partial atomic charge model is widely used in molecular dynamics calculations. Depending on the type and size of the molecules there are several methods available to calculate partial atomic charges. For simple species, calculation of partial atomic charges is possible from their geometry. In most cases, however, there is no direct method available to obtain partial atomic charges despite various options proposed. Often it is possible to calculate them from the charge model and then comparing results with experiment or quantum mechanics. For larger systems, with thousands of atoms, it is not possible to use quantum mechanical calculations to obtain the partial atomic charges. In such situation, molecules can be divided into smaller pieces suitable to recreate close local environment of the fragment in the larger molecule. Most of the charge models have atom-centered charges where the electrostatic forces due to charge-charge interactions act directly on the nuclei. This is important if one needs to calculate the forces on the nuclei as is required for energy minimization or a molecular dynamics simulation.

The dielectric constant used in the above equation is with the reference to a vacuum. If one wants to include solvent effects in the simulations without adding any solvent molecules, they can dampen the electrostatic interactions. This is often done by increasing the dielectric constant via using the appropriate value of the relative permittivity in the Coulomb`s equation (i.e. $\varepsilon=\varepsilon_0\varepsilon_r$) [29].

Electrostatic interactions do not account for all non-bonded interactions in a system. This can be clearly explained using the rare gas example. There is no dipole-dipole or dipole-induced dipole interactions between rare gas atoms because all their multiple moments are 0. It doesn`t mean, however, that the atoms do not interact with each other. They can exist in liquid and solid phases so the interactions have to be there. This kind of non-bonded interactions is due to **van der Waals forces**. The interaction energy between the two atoms being at infinite distance is equal to zero but when the separation is decreased the energy decreases passing a minimum at a distance of about 3.8 Å for argon atoms. Then the energy increases rapidly as the

separation decreases even more. What happens here is that at long distances (long-range) the attractive forces act upon the atoms, but at short distances the acting forces are repulsive. The attractive behaviour is due to dispersive forces occurring due to instantaneous dipoles which arise during the fluctuations in the electron clouds. This dipole can in a molecule induce a dipole in neighbouring atoms creating the attractive effect. Below 3 Å, however, even a small decrease in the distance between two argon atoms raises the energy significantly. That interaction is caused by electrons having the same spin. For this reason, short-range repulsive forces are often called exchange forces. The effect of exchange is reducing the repulsion between two electrons by not allowing them to occupy the same part of space.

In molecular modelling, the inter-atomic potential curve has been described by a simple empirical equation called the **Lennard-Jones 12-6 function** (Eq.9) illustrated in Fig.18:

$$v(r) = 4\varepsilon[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}]$$

(Eq.9)

The Lennard-Jones potential consists of the collision parameter σ which is the separation at which the energy is 0, and the well depth ε.

From the (Eq.9) we can distinguish the long-range and short-range interactions. The attractive force varies as r^{-6} and the r^{-12}. Both of them are depicted in Fig.19. In some force fields different powers for the repulsive term can be used (i.e. 9-10) [29].

The attractive and repulsive forces can be determined from crystal packing analysis or derived using liquid simulations in which the parameters are adjusted to reproduce thermodynamic properties, like densities and enthalpies of vaporisation of certain liquids.

The van der Waals and electrostatic interactions that are separated by three bonds are often described differently and called 1, 4 non-bonded interactions. They are scaled down by an empirical factor, for instance, to reduce the error associated with the use of repulsion (r^{-12}) term [29].

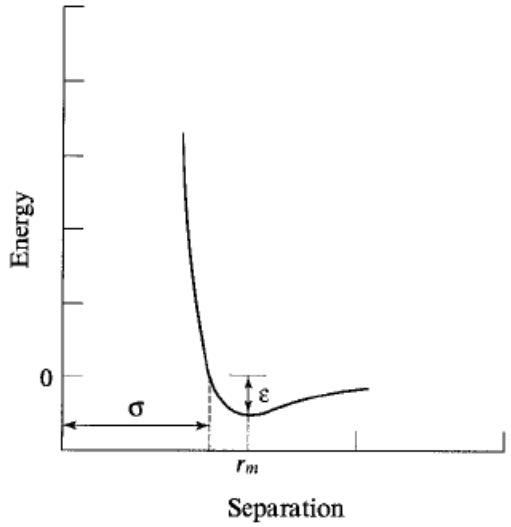

Fig.18. The Lennard-Jones 12-6 potential energy curve.

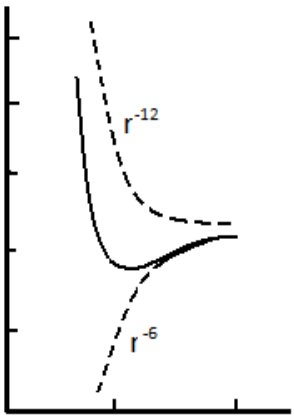

Fig. 19. The attractive term (r^{-6}) and the repulsive term (r^{-12}) of the Lennard-Jones pair potential.

3.2.4. Energy minimization methods

They are an integral part of molecular dynamic calculations. Their aim is to prepare the system of interest for other types of calculations by decreasing the potential energy of the system. It is especially important in predictions of crystal configurations and in simulations of very large molecules and complex systems.

For a system with N atoms the energy is a function of 3N-6 internal degrees of freedom or 3N Cartesian coordinates used to describe the positions of atoms at different times, leading to a **potential energy surface**. In MD, researchers are particularly interested in minimum points on the energy surface, because they refer to stable states of the system. There can be many points on the potential energy surface with low energy. The one of the lowest energy is called the global energy minimum. The highest energy point on the way from one minimum to the next one is called the saddle point. It corresponds to a structure being in the transition state [29].

To minimize the energy of a molecule or a whole collection of them, one can choose between non-derivative and derivative minimisation methods. In this work, however, we will take a closer look only at the second group of calculations as they are the most popular ones and also relevant to this study.

The two most widely used derivative-based methods are the **steepest descent** and **conjugant gradient** methods. The direction of the first derivative of the energy (the gradient) localizes the minimum and its magnitude indicates the steepness of the local surface shape. By moving each atom in response to the force acting on it, we can lower the energy of the system. Second derivatives deliver information on where the function will change its direction. Both the steepest descent and conjugant gradient approaches are based on gradual changing of the coordinates of the atoms as they bring system closer and closer to the minimum point. The starting point of each integration is the molecular structure from the previous step.

The choice of the minimisation algorithm depends on factors such as the relative speed with which the various parts of the simulation can be performed and

the properties of the analyzed system. In general, the steepest descent works better if the structure is in some way from the minimum but the conjugant gradient is more efficient once the initial strain on the system has been removed [29].

3.2.5. Constant temperature and pressure conditions

In computational chemistry, it is extremely important to reproduce proper conditions for all steps of each simulation, because only this way the calculated properties can represent a real system. The reproduction of the conditions of interest during simulations can be achieved by using so called coupling methods. They simulate "an external bath" in which a system is being "submerged" in order to keep a certain property constant during the simulation run.

Temperature coupling to an "external bath"

There are two main methods here: the Berendsen coupling and the Nose-Hoover coupling. They control the temperature imposed on the system of interest during the simulation time and differ in the strength of coupling and accuracy.

The Berendsen coupling

It is based on an algorithm that is able to control the systems temperature fluctuations and slowly correct them to the reference temperature T_0:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}$$

(Eq.10)

This means that the temperature deviation decreases exponentially with the time constant τ determining the coupling time. The heat flow of the system is affected by scaling the velocities of each particle every step with a time-dependent factor λ, described by:

$$\lambda = [1 + \frac{\Delta t}{\tau_T}\{\frac{T_0}{T\left(t - \frac{\Delta t}{2}\right)} - 1\}]^{\frac{1}{2}}$$

(Eq.11)

where: $$\tau = 2C_V\tau_T(N_{df}k)^{-1}$$

(Eq.12)

C_V is the total heat capacity of the system, k is the Boltzmann`s constant, and N_{df} is the total number of degrees of freedom [36].

Nose-Hoover coupling

This method is appropriate to use at the constant NPT equilibration runs, after the system was pre-equilibrated with the Berendsen method, and in production runs. It is believed to be more computationally expensive, but more accurate, hence better for the data collection runs. The Nose-Hoover algorithm is expressed by Eq.13:

$$\frac{d^2r_i}{dt^2} = \frac{F_i}{m_i} - \xi\frac{dr_i}{dt}$$

Eq.13

$$\frac{d\xi}{dt} = \frac{1}{Q}(T - T_0)$$

Eq.14

The T_0 is the reference temperature and T is the temperature at a certain moment of time.

$$Q = \frac{\tau_T^2 T_0}{4\pi^2}$$

Eq.15

In this method, the oscillations produced by the canonical ensemble can be very large and it takes several times longer to relax them.

Pressure coupling to an "external bath"

The idea of pressure coupling is similar to the temperature coupling. Again, two methods are available: the Berendsen coupling and the Parrinello-

Rahman coupling. When the pressure coupling is imposed, there is no velocity scaling involved [36].

Berendsen coupling

It is a weak coupling method (Eq.16). It scales coordinates and the box vectors every step with a matrix μ, which has the effect of first-order kinetic relaxation of pressure to its reference value P_0.

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p}$$

(Eq. 16)

The scalling matrix μ can be described as:

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_p}\beta_{ij}\{P_{0ij} - P_{ij}(t)\}$$

(Eq.17)

where β is the compressibility of the simulated system.

Parrinello-Rahman coupling

When this type of barostat is used, the simulations box vectors are described with the matrix b according to the equation:

$$\frac{db^2}{dt^2} = VW^{-1}b'^{-1}(P - P_{ref})$$

Eq.18

$$(W^{-1})_{ij} = \frac{4\pi^2\beta_{ij}}{3\tau_p^2 L}$$

Eq.19

W^{-1} is the strength of the coupling, P_{ref} is the reference pressure, P is the pressure at a point of time and the τ_p is the pressure time constant.

In this approach the equations of motion are described as follows:

$$\frac{d^2r_i}{dt^2} = \frac{F_i}{m_i} - M\frac{dr_i}{dt}$$

Eq.20

$$M = b^{-1}[b\frac{db'}{dt} + \frac{db}{dt}b']b'^{-1}$$

Eq.21

As this method of calculations is more precise, also the τ_p needs to be 4-5 times larger than for a weak coupling, to bring the system to the reference pressure.

3.2.6. Periodic boundary conditions

In our computational experiment, in order to calculate any macroscopical properties from a system containing a small and constant number of molecules, we needed to create a box or "an imaginary flask" which the lipid would be "poured" into. That box was surrounded by images of the same box in all 3 dimensions. By multiplying the box images i.e. by setting the periodic boundary conditions (Fig.20), we were able to create a bulk system. This way all united-atoms experience forces like if they would be in a bulk sample [29]. The coordinates of the atoms in the image boxes are easily calculated. The atoms, however, are allowed to cross the boundaries. In a case when an atom leaves the box, an imaginary atom replaces it by entering the box from the opposite side and the requirement of constant number of molecules is still being satisfied.

3.2.5. The minimum image convention.

It is a time-saving way of dealing with the calculation of the energies and forces coming from non-bonded interactions which are the most time-consuming part of molecular dynamics simulation. In the minimum image convention, each atom "sees" only one image of every other atom in the system. It means that the interactions are calculated only for the closest atom or image (Fig.21) and the rest of them are set to zero. The distance at which an atom

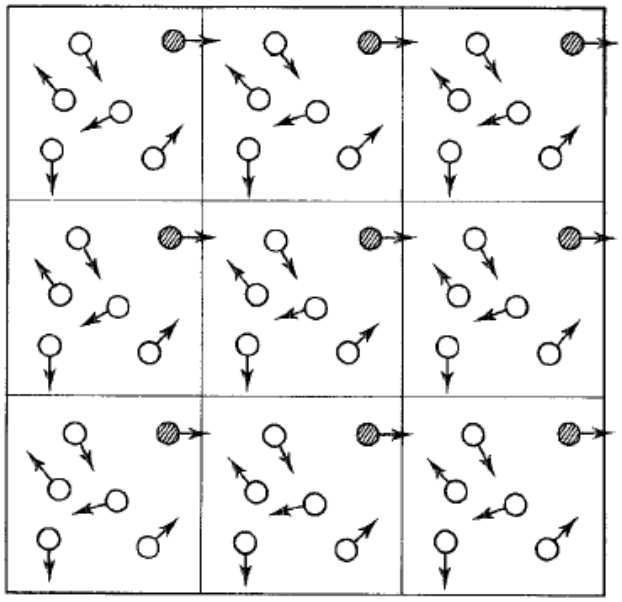

Fig.20. Periodic boundary conditions in two dimensions [29].

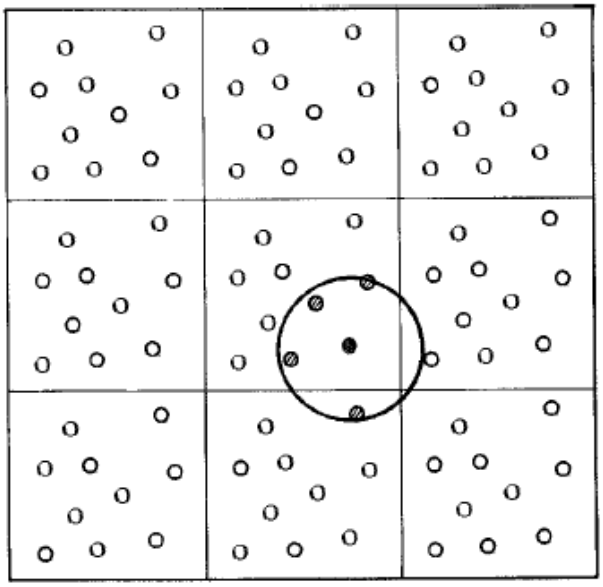

Fig. 21. The spherical cut-off and the minimum image convention [29].

"sees" another atom or image is set by a non-bonded cut-off length. In a system with periodic boundary conditions (described below), like ours, the non-bonded cut-off had to be no more than half the length of the simulation box [29].

3.2.5. Calculation of properties

Simple thermodynamic properties, such as energy, temperature, pressure, heat capacity, phase space and radial distribution function, can be relatively easy to obtain via molecular dynamics. Predictions of systems properties can also be done. This is especially important for systems for which it is very hard or impossible to collect experimental data. Molecular dynamics is also a good tool for gathering information about conformations of molecules and their changes, and about distribution of molecules in a system.

MD calculations are based on **statistical mechanics** and center around ensembles averages to calculate the properties over a period of time. The most commonly sampled ensembles are: the microcanonical (NVE), the canonical (NVT) and the isothermal-isobaric ensemble (NPT) [29].

Let`s not forget though that experimental data are crucial to validate the used model and estimate the accuracy of the calculations.

3.3. MD Simulations with GROMACS

3.3.1. Setting up the simulations

The crystallization phenomena of TAGs can take up anywhere from few seconds to few weeks or even years. Unfortunately, simulating just one second from that process is not possible yet, even with today`s state of the art computer science and technology. For this reason, this work was an attempt to capture the phase transformations of TAG molecule starting at several nanoseconds time-

scale. These are thought to be the first important changes in self-ordering of molecules and they are laying down the foundations to the further first-order crystallization events.

The state and basic properties of several systems of TAGs in the united-atom representation were predicted using the Westgrid platform and the free of charge GROMACS simulation package with the GROMOS-96 force field [37]. In the coarse-grained model we also performed calculations using the Westgrid platform, but the force field was the ff_v2.32 from Marink et al.[38].
Fig.22 presents the general steps undertaken in our simulations.

3.3.2. The united-atom model of TAGs

The atomistic molecular dynamics was performed by utilizing a simplified united-atom model which was proved to reproduce well the α-polymorphic state of triacylglycerides[16]. Following the work of W.F. van Gunsteren and I. Chandrasekhar, we used their parameters for molecules construction, but some of the details of simulations were changed to adjust them to our needs.

When one wants to use the GROMACS simulation package, he/she needs to generate a .pdb format file containing a molecule structure with initial coordinates of its all atoms. To do that, we used Materials Studio simulation software (Accelrys Inc.) which allows drawing the molecule of interest using the Cartesian coordinates and saves it in the .pdb format further read by GROMACS software. Unfortunately, the Materials Studio does not allow constructing the united-atom model and hence an all-atom model was constructed for coordinate's generation. The .pdb file was then translated to Gromacs format (.gro) with the GROMACS software and the structure was modified manually by

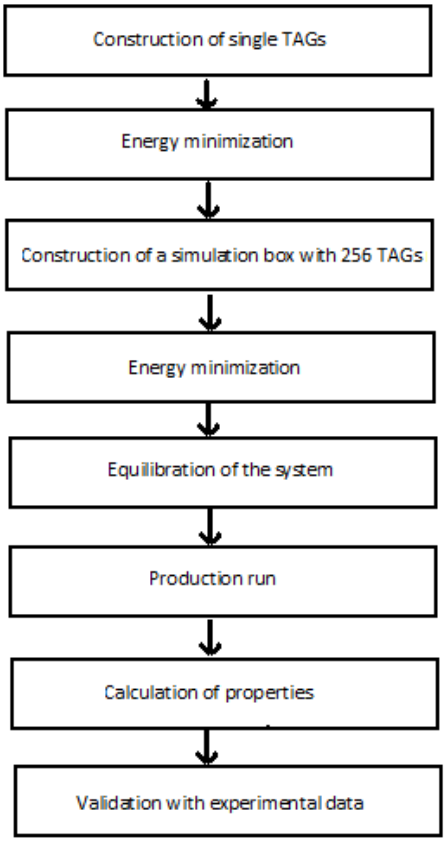

Fig. 22. The molecular dynamics simulations procedure.

removing all hydrogen atoms and correcting all bonded interactions for new united atoms, according to the Gromacs Mannual [36].

In the united-atom model, atoms of the CH, CH_2 and CH_3 groups were substituted with one united atom for each group. The oxygen atoms and ester carbons were left as single atoms and were parameterized with a special attention to planarity of ester moiety in triglycerides. The force field of choice was GROMOS-96. It was validated through experimental and *ab initio* calculations for lipid molecules in water and without addition of any solvent [17]. The subset of parameters chosen from that force field was 53a5.

Both single molecules of tripalmitin and tristearin were submitted to energy minimization. The potential energies converged with the steepest descent method, used with the force tolerance of 10 $kJmol^{-1}nm^{-1}$.

Next, we built a cubic simulation box with a single TAG molecule - the final structure from the energy minimization step (Fig.23a). Then, a system of 8 molecules was created by multiplying the initial single TAG twice in x, y and z directions. It was then multiplied to 128 molecules and further to 256 molecules (Fig.23b). Enlarging the systems in 3 stages allowed for better energy minimization. This way we obtained 3 systems submitted to further calculations (Table 1).

For all our simulations, we chose a cubic shape of the box, because it is a good representation of a bulk system containing pure lipid molecules, having no solvent and no surface effects to account for. All box dimensions were adjusted according to the minimum image convention. Also the periodic boundary conditions were imposed in all simulations.

The densities of each system were adjusted to known experimental values where possible. After each increase in size, systems were subjected to energy minimization with conjugate gradient method. The force tolerance was set to 10 $kJmol^{-1}nm^{-1}$. When the 256 TAGs systems satisfied the potential

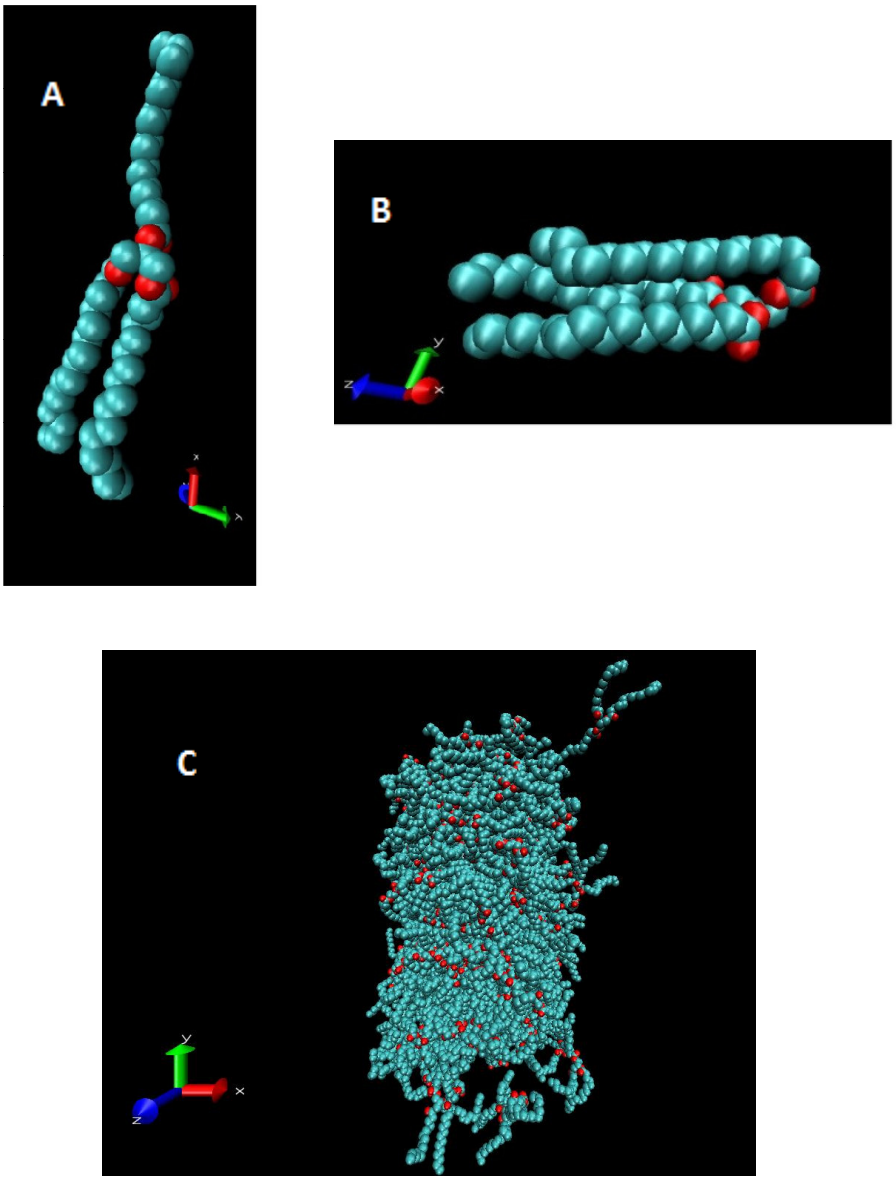

Fig.23. The Gromacs representation of united-atom particles:
a) a single PPP molecule, b) a single SSS molecule,
c) a system of 256 PPP molecules.

energy convergence requirement, they were ready for equilibration simulations.

The initial velocities were calculated from the Maxwell-Boltzman distribution at 353 K to ensure liquid state of all systems.

Table 1. The numbers of molecules and united atoms in constructed systems.

Number of TAG molecules	TAG type	Total number of united atoms
1	PPP	57
1	SSS	63
256	PPP	14 592
256	SSS	16 128
2048	CG PPP*	30720

* CG - coarse-grained model

3.3.3. The force field

Atom types

According to the GROMOS-96, triacylglycerols were built of 4 different types of carbon atoms and two different types of oxygen. Below are the definitions of their united-atom representations:

O carbonyl oxygen (C=O; 15.99940 u.m.a.)

OM carboxyl oxygen (CO-; 15.99940 u.m.a.)

C bare carbon (C=O; 12.01100 u.m.a.)

CH1 aliphatic CH group (13.01900 u.m.a.)

CH2 aliphatic CH_2 group (14.02700 u.m.a.)

CH3 aliphatic CH_3 group (15.03500 u.m.a.)

Bonded interactions

The parameters for bond stretching interactions were adapted as pairwise potentials from GROMOS-96 (Table 2). The force constant was 2.5. The equation used for modelling of **bond stretching** was in the form:

$$v(l) = \frac{k'}{4}(l^2 - l_0^2)^2$$

(Eq.22)

The force constant in (Eq.22) is calculated from the following relationship of constants: $k=2k'l_0$, where k comes from the (Eq.3).

Table 2. Pairwise potential parameters for bond stretching from GROMOS-96 force-field [17].

Bond	l_0 [nm]	$10^7 k'$[kJmol^{-1}nm^{-4}]
CH_3-CH_2	0.153	0.715
CH_2-CH_2	0.153	0.715
CH_1-CH_2	0.153	0.715
C-CH_2	0.148	0.764
CH_2-O	0.147	0.871
C-O	0.133	1.180
C=O	0.123	1.660

The parameters for the **bond angle potentials** (Fig.18) are summarized in Table 3. For better computer efficiency the angle bonding interactions were calculated from a modified equation:

$$V_b(r_{ijk}) = \frac{1}{2}k_{ij}^b(r_{ijk} - b_{ijk})^2$$

(Eq.23)

with a corresponding force F_i:

$$F_i(r_{ij}) = k_{ij}^b(r_{ij} - b_{ij})\frac{r_{ij}}{r_{ij}}$$

(Eq.24)

The force constant k_{ij} is related to the harmonic force constant k_b in the following manner: $2k^b b_{ij}^2 = k^{b,harm}$. The subscripts *i, j, k* are in sequence of covalently bonded atoms with atom *j* is in the middle.

To parameterize the **proper torsional (dihedral)** interactions (Table 4) the expansion in powers of *cos φ* (Eq.25) was used. They were defined according to UPAC/IUB convention where the angle *φ* is defined as the angle between *ijk* and *jkl* planes (Fig.13) with 0 corresponding to the *cis*- configuration (*i* and *l* on the same side).

$$v(\varphi) = k_\varphi(1 + \cos(n\varphi - \varphi_0))$$

(Eq. 25)

We also decided to use another torsional interaction description (Eq.26) with the **Ryckaert-Bellmans** potentials (Fig.24), because it was proven to better mimic *cis*- conformations of hydrocarbon chains in triacylglycerides [39]. The constants C_0 to C_5 are given in a Table 5. It requires, however, the exclusion of 1,4-Lennard-Jones interactions from the non-bonded list.

$$v(\varphi) = \sum_n C_n(cos\varphi)^n$$

(Eq. 26)

To keep the planarity of ester moieties in TAGs, four **„improper" torsional angles** were defined with a harmonic function (Eq.27) (Table 6). This method also prevented the CH_2 united atom and the carboxyl oxygen linked to CH united atom from flipping over to their mirror images.

$$v_{id}(\xi_{ijkl}) = k_\xi(\xi_{ijkl} - \xi_0)^2$$

(Eq.27)

Table 3. Bond angle potentials parameter set [16].

Angle	θ_0 [deg]	k_θ'[kJmol^{-1}]
CH_2-CH_1-CH_2	109.5	517.0
CH_1-CH_2-O	111.0	527.0
CH_2-CH_1-O	109.5	517.0
CH_2-O-C	117.0	635.0
CH_1-O-C	117.0	635.0
O-C=O	124.0	730.0
O-C-CH_2	115.0	610.0
C-CH_2-CH_2	120.0	780.0
CH_2-CH_2-CH_2	111.0	527.0
CH_3-CH_2-CH_2	111.0	527.0
O=C-CH_2	121.0	685.0

Table 4. GROMOS-96 parameter set for proper dihedrals [16].

Proper dihedral	θ_0 [deg]	k_ϕ'[kJmol^{-1}]	n
CH_2-CH_1-CH_2 -O	0.0	5.85	3
CH_1-CH_2-O-C	0.0	3.76	3
CH_2-CH_1-O-C	0.0	3.77	3
CH_2-O-C=O	180.0	16.74	2
CH_1-O-C=O	180.0	16.74	2
O-C-CH_2-CH_2	0.0	0.42	6
C-CH_2-CH_2-CH_2	0.0	5.86	3
CH_2-CH_2-CH_2-CH_2	0.0	5.86	3
CH_3-CH_2-CH_2-CH_2	0.0	5.86	3
CH_2-O-C-CH_2	180.0	16.7	2
O=C-CH_2-CH_2	180.0	2.9	2
O-CH_1-CH_2-O	180.0	2.09	2

Table 5. Constants for Ryckaert-Bellmans potentials [kJ mol^{-1}] [36].

C_0	9.28
C_1	12.16
C_2	-13.12
C_3	-3.06
C_4	26.24
C_5	-31.5

For improper torsions (dihedrals), such as the above, it is recommended to keep the ξ_0 as far away from $\pm 180^o$ as possible [36].

Table 6. The improper torsional potentials for TAGs [36].

Improper dihedral	ξ_0 [deg]	k_ξ[kJ$mol^{-1}rad^{-2}$]
C O- CH2 =O	0.0	167.42
CH_1 –O CH_2 CH_2	180.0	335.00

Non-bonded interactions

The Lennard-Jones 12-6 potentials (Eq.9) were implemented. The L-J repulsion interactions are very strong up to the 3rd neighbouring atom* in the TAG molecule. For this reason, atoms that are close in the chain until three bonds are automatically excluded from the calculations. If they were accounted for, the molecule could break and the simulation could crash.

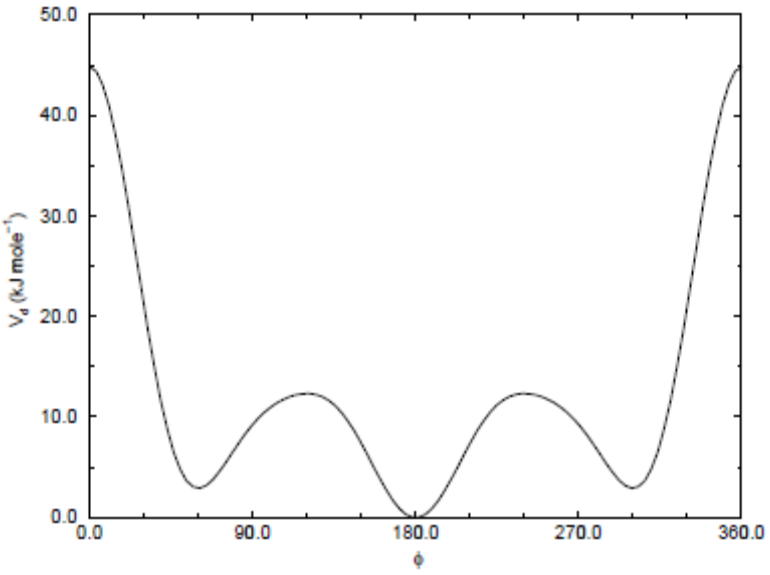

Fig.24. Ryckaert-Bellmans torsional potential [36].

The electrostatic interactions were calculated with the (Eq.7) from Coulomb`s law.

The L-J parameters along with partial charges of atoms are collected in the Table.7.

Twin-range cut-off for all non-bonded interactions was selected to avoid atom to interact with its own image.

The list of neighbouring atoms within the short-range cut-off, equal to 1 [nm], was updated every time-step.

Table7. The Non-bonded parameters for L-J interactions from GROMOS-96 force-field.

Particle	Charge [C]	$10^{-1}\sigma$[nm]	$10^{-1}\varepsilon$[kJmol^{-1}]
CH_3	0.00	3.748	8.6720
CH_2	0.200	4.070	4.1050
CH_1	0.200	5.019	0.9489
C	0.540	3.361	4.0590
O=	-0.380	2.626	17.2504
O-	-0.360	2.955	8.4960

*neighbouring atom is an atom that is far from a subsequent atom n bond.

3.3.4. Equilibration of the system

The molecular dynamics simulations started after obtaining the systems of 256 TAG molecules from energy minimizations. The purpose of equilibration simulations was to bring the systems close to their equilibrium states and assure the proper conditions imposed on the molecules, such as temperature and pressure. Equilibration of each system lasted until its potential energy converged

and the fluctuations of systems` temperature and pressure reached the reference values.

The MD calculations at this stage were conducted with the constant NVT ensemble first and then continued with the constant NPT ensemble. The temperature and pressure were set to their reference values and were allowed to fluctuate until the fluctuations were negligible and the average properties stabilized. To impose chosen values of temperature and pressure the „external bath" coupling approach was used (explained in chapter 3.2.6.). For equilibration simulations we used the Berendsen thermostat and the Berendsen barostat for the NVT part and the Nose-Hoover thermostat and the Parrinello-Rahman barostat for the NPT part.

The systems were equilibrated at 353 K to ensure their liquid state. The relaxation time was set to τ=0.1 ps and the initial density at this temperature was 0.868 g/cm^3 as reported by Sum et al.[18].

For the purpose of this work, we used the isotropic (i.e., equal in all three directions) pressure coupling. The reference pressure was set to 1 bar and relaxation time was 4 ps. When the pressure coupling was imposed there was no velocity scaling used.

The convergence of the temperature and pressure to their reference values, as well as the convergence of the potential energies of equilibrated systems were confirmed with g_energy tool from GROMACS package.

3.3.5. Simulated annealing

The aim of the study was to capture the first detectable phase transformations of a triacylglycerol sample. The phase changes were induced by lowering the temperature of the simulated system. To achieve that, we used the simulated annealing method developed by Kirkpatrick S. et al. [40]. It is based on an algorithm that slowly, but steadily, decreases the initial temperature, at a

chosen rate, to the temperature at which first crystallization changes are believed to occur. After reaching the final temperature, the system was subjected to the equilibration simulations, described above.

3.3.6. Constraints

To make simulations more computationally efficient without sacrificing their accuracy, one can use constraints imposed on their systems. This is done by constraining the internal Cartesian coordinates of the atoms to specified values so that bonds or angles values do not change over the course of the simulation [29].

In our calculations, constraints were imposed on all bonds in each system via LINCS algorithm [36][Appendix 1]. First, after an unconstrained update, the projections of new bonds on the old bonds are set to 0. Then, in the following step a correction is applied for the lengthening of the bonds due to rotation. To achieve that, the equations of motion change to:

$$\frac{d^2r}{dt^2} = M^{-1}F$$

Eq.28

Where F is the force vector and M is a diagonal matrix of 3Nx3N, containing the masses of particles. The system is constrained by K time-independent equations:

$$g_i(r) = |r_{i1} - r_{i2}| - d_i = 0$$

Eq. 29

$$i = 1, \dots, K$$

3.3.7. Production runs

After a successful equilibration of each system, we continued MD simulations under the constant NPT ensemble. This stage generates the data for properties calculations and is called a production run.

For the constant NPT conditions we used the Nose-Hoover temperature coupling and the Parrinello-Rahman pressure coupling algorithms. In these methods, the properties oscillations can be very large and it takes several times longer to relax them. For this reason, the time constant was chosen to be τ=0.4 ps and the pressure time constant τ_p was set to 16 ps. All simulations were performed under the pressure of 1 bar.

3.3.8. The coarse-grained model of tripalmitin

This approach was utilized for comparison with the united-atom model. We wanted to compare both approaches in terms of accuracy of calculations and the time scale of simulations. For the model construction please see section 3.2.2.

As already mentioned, this approach demands a different kind of force field - ff_v2.32 from Marink et al. [38].

Interaction sites

Instead of atom types, only four main types of interaction sites are used here. These are:

P – polar – neutral groups that can easily dissolve in water

N – non-polar – mixed groups of polar and apolar moieties

C – apolar – hydrophobic groups

Q – charged – ionized groups

For N and Q types, there are four subgroups for fine-tunning of parameters, such as:

0 – no hydrogen bonding capabilities

d – hydrogen donor group

a – hydrogen acceptor group

da – groups with both donor and acceptor capabilities.

Non-bonded interactions

The van der Waals interactions are described by the Lennard-Jones 12-6 potential (Eq.9). Because there are no charged groups in our model, no other interactions need to be descriebed.

Bonded interactions

In coarse-grain representation only two- and three-bodies potentials are needed to be considered. These are described with the weak harmonic potential as in eq.22 and eq.23 respectively.

Simulation parameters

The coarse-grain model was designed for room or physiological temperatures. However, for our purpose we dared to use it with the initial temperature of 353 K at the pressure of 1 bar.

For the temperature and pressure control the same standard coupling are suitable (see section 3.2.5.). The time and pressure constants need to be adjusted to the values of 0.4 ps and 16 bar respectively.

Because this approach was created to save on the computational time and resources, the time step value can be of one magnitude higher than for the united-atom model. Here, we chose the time step of 40 fs.

No constraints are applied in CG simulations. For more specific details please refer to the publication of Marrink et al. [38].

3.4. Analysis methods

3.4.1. Density

It is one of the most basic yet very useful bulk properties of any substance. There are many experimental methods to determine density. However, there are still some cases (i.e. thermally unstable substance, too small quantity of material or hard to remove impurities) in which it is difficult to obtain its value experimentally. Thus, for estimation purposes, density prediction methods are helpful. Molecular dynamics can be one of them. It estimates density by means of unit cell type and its volume and molecular weight of a substance over the course of a simulation.

3.4.2. Root Mean Square Deviation

The root mean square deviation (RMSD) allows to find the difference in the orientation of atoms in a molecule between the initial state (e.g. after energy minimization) and the final state, as well as helps to determine at what point in time the system is at equilibrium. RMSD was calculated using the least-square fitting method (Eq.30) with a reference to a structure (t_2=0) obtained from the energy minimization.

$$RMSD(t_1, t_2) = [\frac{1}{M}\sum_{i=1}^{N} m_i \|r_i(t_1) - r_i(t_2)\|^2]^{\frac{1}{2}}$$

Eq.30

where $M = \sum_{i=1}^{N} m_i$ and $r_i(t)$ is the position of atom *i* at time *t* [36].

3.4.3. Radial Distribution Function

The radial distribution function (RDF), g(r), is a correlation function between a pair of particles A and B (Eq.31) accounting for non-bonded interactions. It allows calculating the probability of finding a pair of atoms a distance r apart, relative to the probability expected for a completely random distribution at the same density [41]. In the case of an ideal gas, which is described by the lack of molecular interactions, the RDF is equal to one at all distances r. In a real system, however, interactions will be characterized as peaks and values relating to the regions of high and low density, respectively. For this reason, the RDF can be used as a measure of structural organization of any given system.

$$g_{AB}(r) = \frac{<\rho_B(r)>}{<\rho_B>local} = \frac{1}{<\rho_B>local\ N_A}\sum_{i\in A}^{N_A}\sum_{j\in B}^{N_B}\frac{\delta(r_{ij}-r)}{4\pi r^2}$$

Eq.31

In the Eq.31 $< \rho_B(r) >$ is the atom B average density at distance r around atom A (Fig.25), $< \rho_B >_{\text{local}}$ is the average density of atoms B inside of the maximum radius, N_A and N_B are the total numbers of atoms A and B. The g(r) is dimensionless. In our work, A and B are the same.

3.4.4. Diffusion coefficient

Another useful property that can be obtained through molecular dynamics is the self-diffusion coefficient. It is extracted from the mean square displacement (MSD) of the centers of mass of all molecules in each system. Graphically, it is represented by a slope of the linear part of the MSD curve.

The self-diffusion coefficients were calculated on the basis of the Einstein relation (Eq. 32):

$$\lim_{t\to\infty}\langle\|r_i(t)-r_i(0)\|^2\rangle_{i\epsilon A}=6\,D_A t \qquad \text{(Eq.32)}$$

where D_A is the self-diffusion coefficient of particle A and r_i is the center of mass position of that particle [36]. To obtain more reliable self-diffusion coefficients, only the first 25 ns rather than the entire of MD trajectories was used to determine the mean square displacement.

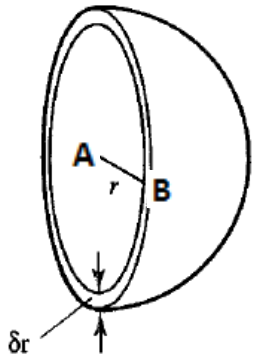

Fig 25. Spherical shell of thickness δr at a distance r from atom A to B.

4. Results and discussion

As stated in the last chapter, the production runs were performed under constant temperature and constant pressure conditions. The fluctuations of properties (e.g. density) with time were utilized to gather the information about the accuracy of our models and about the phase behaviour of triacylglycerols.

4.1. Density

To evaluate the accuracy of our models, it was important for us to start with the calculation of the densities for all systems. Density, being a basic property, is easy to compute and sheds some light on the state of a system at any given temperature.

From the beginning, density was set to experimental values and was allowed to spontaneously fluctuate in the course of equilibration and production runs. The final densities estimates were performed at three temperatures: 353 K, 338 K and 315 K. The choices of temperature were determined by transformation temperatures of tripalmitin.

As it was expected, the results (Table 8) show the increase of density with decreasing temperature (Fig.26). The liquid density of tripalmitin at 353 K increases to the value of 0.913 g/cm^3. The highest obtained density suggests that the system achieved a higher degree of organization and was undergoing the transition to a solid-like state (see fig. 27-31, section 4.2). It is expected that given more simulation time under chosen conditions, the system would reach the packing density of a crystal-like structure, or at least of a liquid crystal, i.e. around 1 g/cm^3.

As we can see from Table 8 the computed density values differ slightly from the experimental ones yet still being in good approximation to reality. We

believe that the observed discrepancies are between the calculated and experimental values are due to:

a) at the lowest temperature – the molecular movements were slowed down considerably; the timescale of simulations was not large enough to ensure more accurate result.
b) for the coarse-grain model – the difference between the experimental and calculated values comes from the nature of coarse-grain approach which makes calculations less accurate to gain on simulation time and efficiency.

Please note that density, being a bulk property, is influenced not only by temperature, but also pressure, unit cell volume and molecular weight. However, in regards to the early stage of our project and to simplify our models, we chose only the most influential variable (i.e. temperature) to take under consideration.

Table 8. The calculated and experimental values of densities computed through molecular dynamics simulations.

System	Temperature [K]	$\rho_{exp.*}$ [g/cm^3]	$\rho_{calc.**}$ [g/cm^3]	The difference between $\rho_{exp.}$ and $\rho_{calc.}$ [%]
128 PPP	353	0.868 [18]	0.877	1.04
	338	-	0.899	-
	315	approx. 1.0 [42]	0.913	8.70
256 PPP	353	0.868 [18]	0.877	1.04
	338	-	0.899	-
	315	approx. 1.0 [42]	0.913	8.70
2048 PPP CG	353	0.868 [18]	1.079	24.3
256 SSS	353	0.856 [43]	0.885	3.4

*experimental value based on literature
**the value calculated in our simulations.

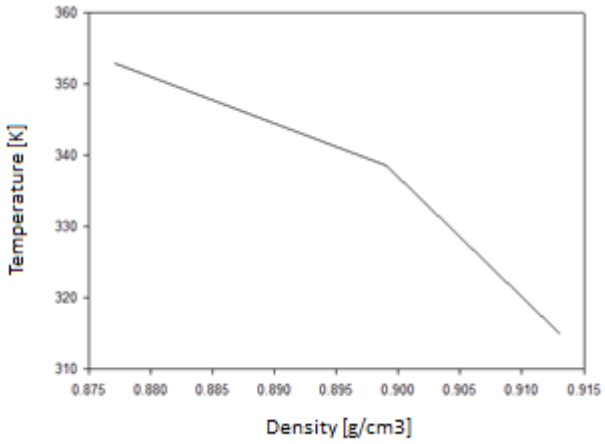

Fig. 26. The density change with temperature for 256 molecules of tripalmitin.

4.2. Root Mean Square Deviation

The physical states of our systems were also validated via the root mean square deviation (RMSD). The aim was to confirm their equilibration and to gain more insight into their possible structural changes while on their way to equilibrium. The calculations followed the method described in section 3.4.2.

As we can see from the first graph (Fig.27), for 256 tripalmitin molecules at 353 K, RMSD levels-off fast at the beginning of the production run suggesting that the particles oscillate around their equilibrium states. This situation looks stable for tens of nanoseconds. However, approximately between 80 and 150 ns, the RMSD decreases considerably and increases almost reaching the previous level. This is especially interesting, because in the same time there is no change in the average temperature or pressure. Therefore, there is no external "force" causing the system to re-structure. The situation repeats for the 256 molecules of tripalmitin at temperatures of 338 K and 315 K (Fig. 28-30), but it differs in time. This may probably be caused by the fact that at lower temperatures any changes in a system take longer to occur. We speculate that the decrease in the RMSD value is attributed to the fact that the tripalmitin undergoes a tentative re-organization to a meta-stable state, which has been proven experimentally. Many triacylglycerols show the tendency to organize themselves locally and/or to go through few polymorphic states, including the meta-stable ones, before they reach the most stable β-polymorph.

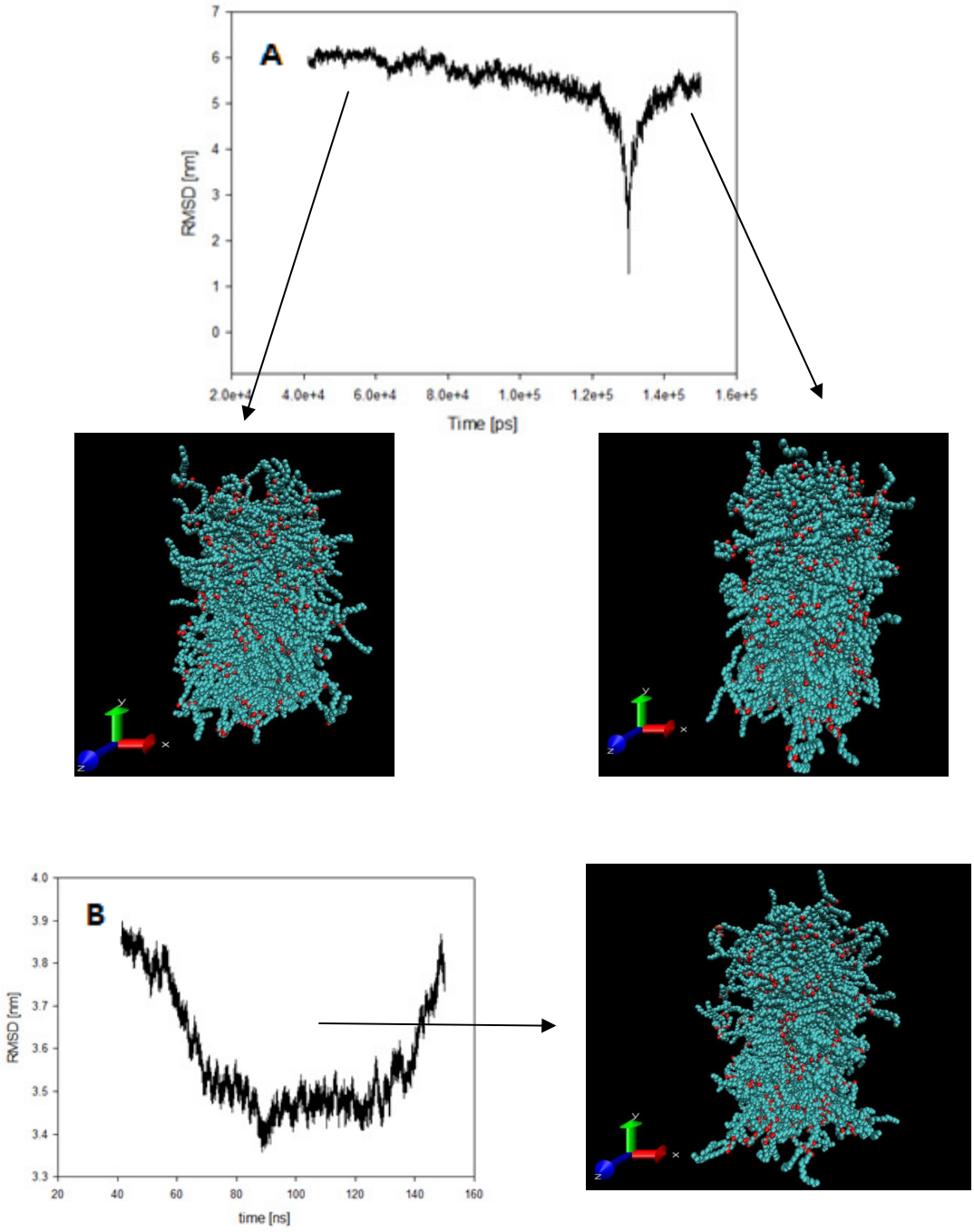

Fig.27. RMSD for the united-atom model at 353 K:
a) the whole production run, b) a close-up on the well area.

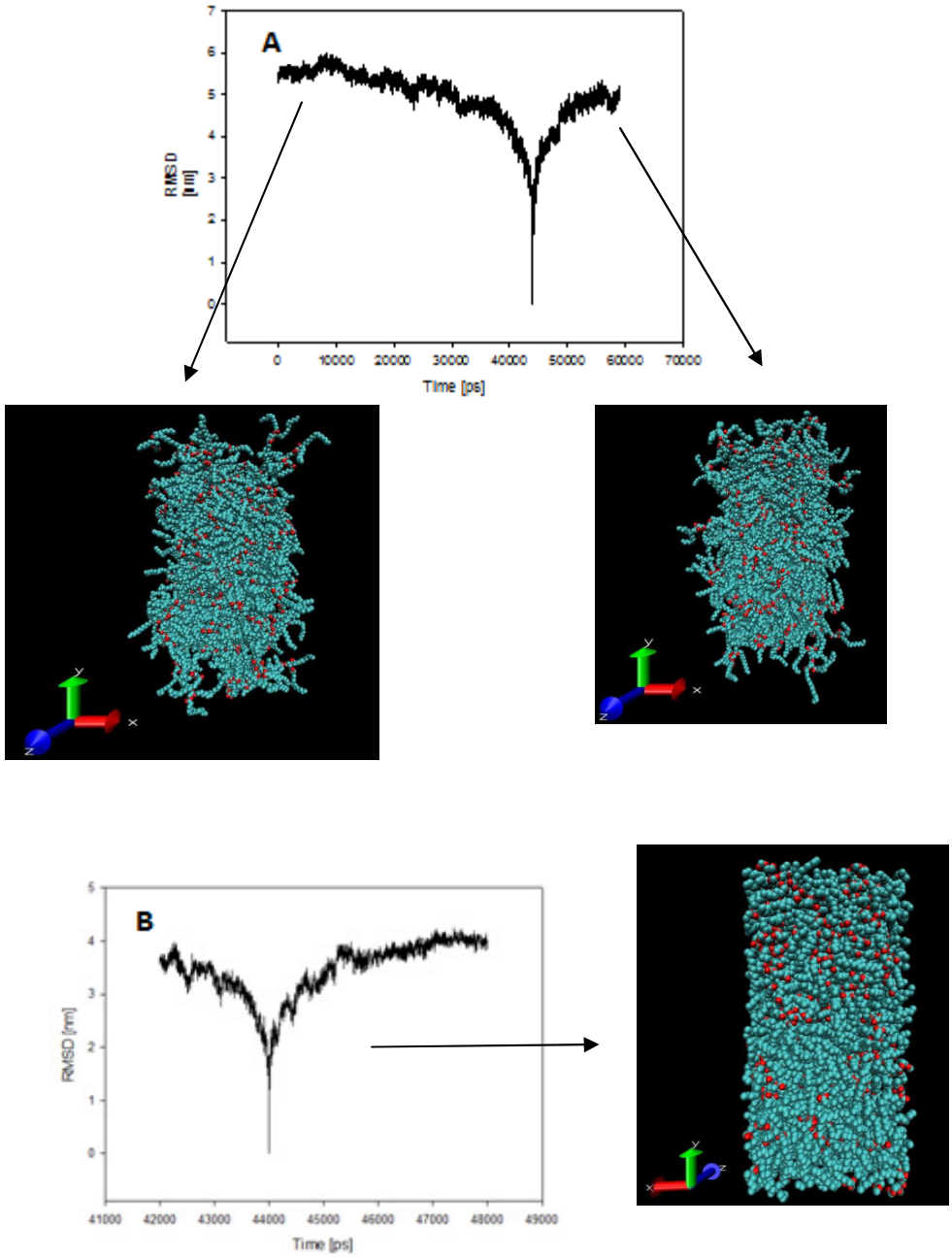

Fig. 28 RMSD for the 256 tripalmitin molecules at 338 K:
a) the whole production run, b) a close-up on the well area.

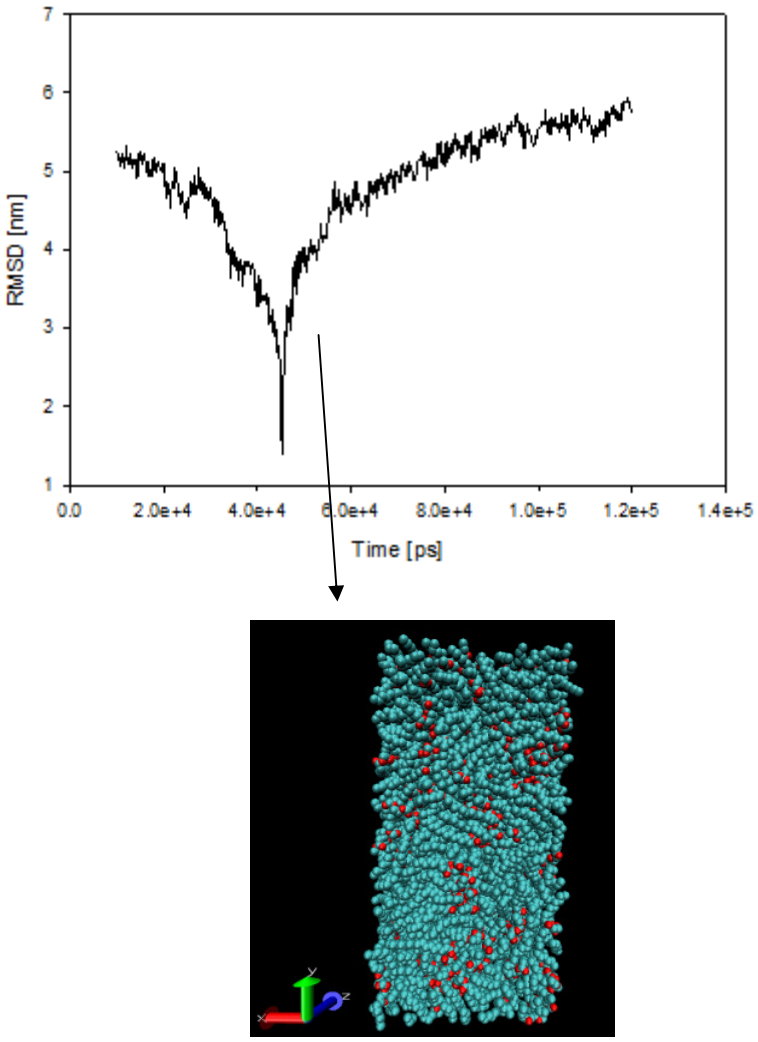

Fig.29. RMSD for the united-atom model at 315 K – the whole production run.

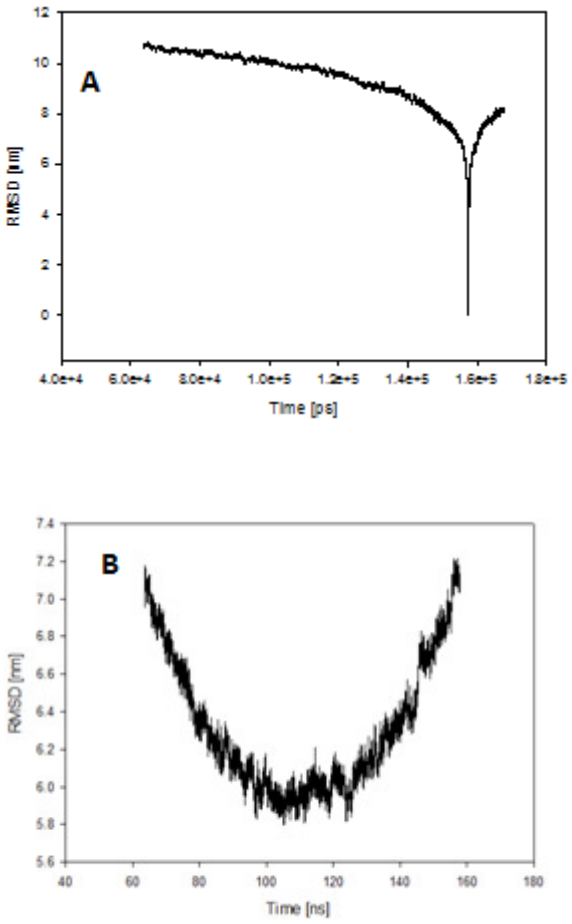

Fig. 30. RMSD for the coarse-grain model of 2048 tripalmitin molecules at 353 K: a) the whole production run, b) a close-up on the well area.

4.3. The Radial Distribution Function

The radial distribution functions (RDF) were calculated with the eq.31, and were described in section 3.4.3.

Figure 31a shows the radial distribution function of the PPP molecules at 353 K. In the short-range interaction region we can see a distinct peak at r=0.5 nm, followed by a smaller one at r=0.6 nm and another wider peak at 1 nm. This is consistent with a liquid-like structure with the interactions fading in the long-range region, after reaching 1.5 nm. Similar results have been reported for simulations of unsaturated TAGs [38]. This kind of behaviour is typical for fluids in a liquid state.

At the temperature of 338 K (Fig.31b), we see the same dependence. The system still reveals the liquid characteristics, despite lowering its temperature. There may be three reasons for this: a) changes in the liquid-like structure had not been as pronounced yet to be visibly detectable, b) still high temperature favours the liquid-like behaviour c) the system should be simulated longer for the structural changes to be easily visible on the graph.

Lastly, we analyzed the box of 256 molecules of tripalmitin at 315 K. On the third graph (Fig. 31c), we can see changes in the structure of the molecules being in the closest neighbourhood to one another. Judging the sharpness of the peaks in the 0.0-0.5 nm region, we can assume that much higher degree of phase organization has been achieved. It is hard to say exactly if the system reached the gel-like phase. We believe though that this graph indicates first signs of phase change that happened due to the decrease in temperature from 338 K to 315 K. The second temperature is the experimental crystallization temperature of tripalmitin, undergoing a change from α-phase to β-phase that has been imposed on the system to trigger the crystallization process. Again, as it was stated previously, the time scale of our simulations should be considerably longer to observe the structural changes also in the long range interaction region

of the RDF graph and witness the crystallization process of a mathematical model of TAGs.

Another way to obtain information about intermolecular structure is to compute the radial distribution functions for certain atoms within the molecules. As representative atoms we chose the oxygen atoms within the ester group (i.e., O2, O3, O5) of a TAG as well as the terminal carbon atoms (i.e., C1, C34, C51) (see Figure 8, section 2.2.; Fig.32-35).

It has been observed in the literature that lipids tend to self-assemble and create bilayered structures under favourable conditions. Hence, our next step was to examine simulated triacylglycerols in terms of their ability to express that kind of behaviour after being subjected to a cooling process. Figure 36 shows the resulting structures from production runs at the highest and lowest temperatures. After production runs at 353 K and at the 315 K no distinct layering structure was observed. It is understandable given the considerably short timescale of those simulations and high density of tripalmitin. However, there is an apparent strong tendency of the oxygen atoms to gather in groups, trying to isolate themselves from the hydrocarbon tails. For this reason, we believe that the correct inter- and intra-molecular interactions have been produced between the molecules, and given enough time to re-structure, the TAGs would be able to form layers.

The self-assembling activity is also noticeable locally for the hydrocarbon tails. This is especially visible for the system of 128 tripalmitin molecules at the end of the simulation at 315 K (Fig.36 c). For a smaller TAGs system, such as this one, shorter period of time is needed to undergo any phase change.

Here, though it is still hard to talk about a phase transformation, it is clear that the long hydrocarbon tails (in the center of the system) were able to layer themselves in a close to parallel way.

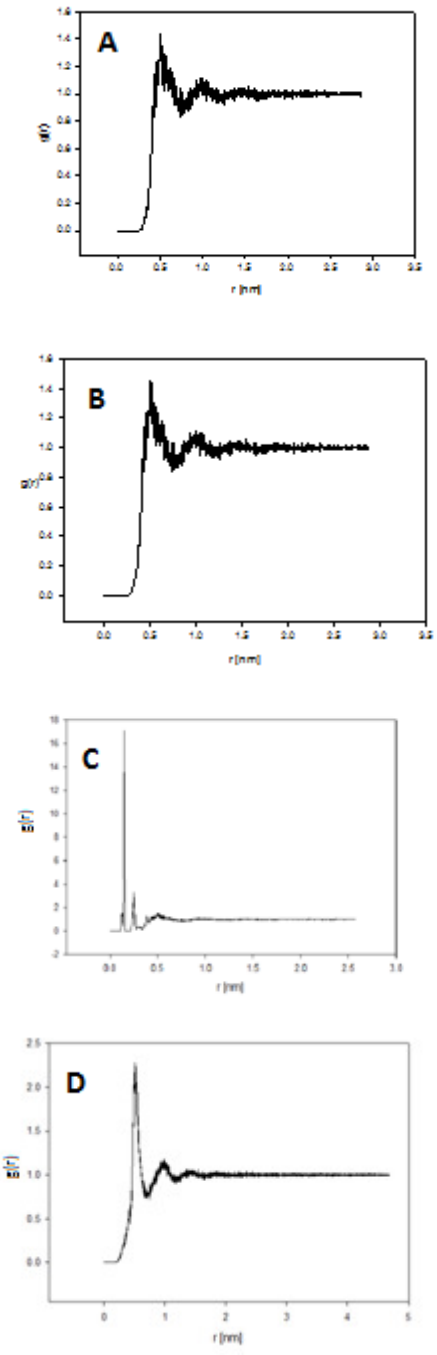

Fig.31. RDF for the center of masses of 256 PPP:
a) at 353 K, b) at 338 K, c) at 315 K, d) the 2048 CG PPP at 353 K

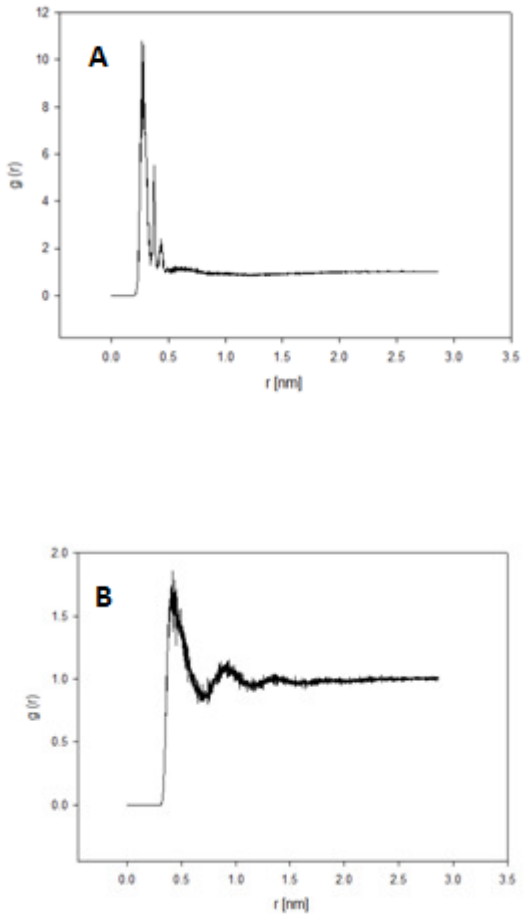

Fig.32. RDF for 256 PPP at 353 K:
a) ester oxygens, b) CH_3 groups at the end of hydrocarbon tails.

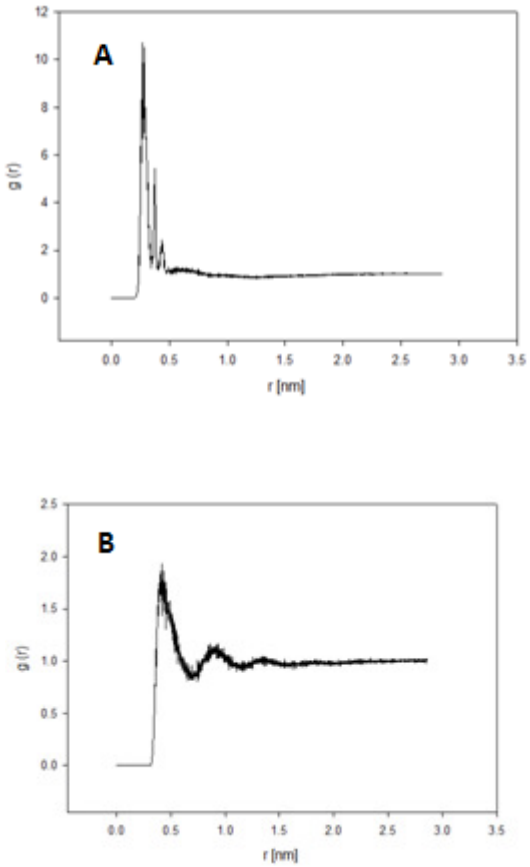

Fig.33. The RDF for 256 PPP at 338 K:
a) ester oxygens, b) CH_3 groups at the end of hydrocarbon tails.

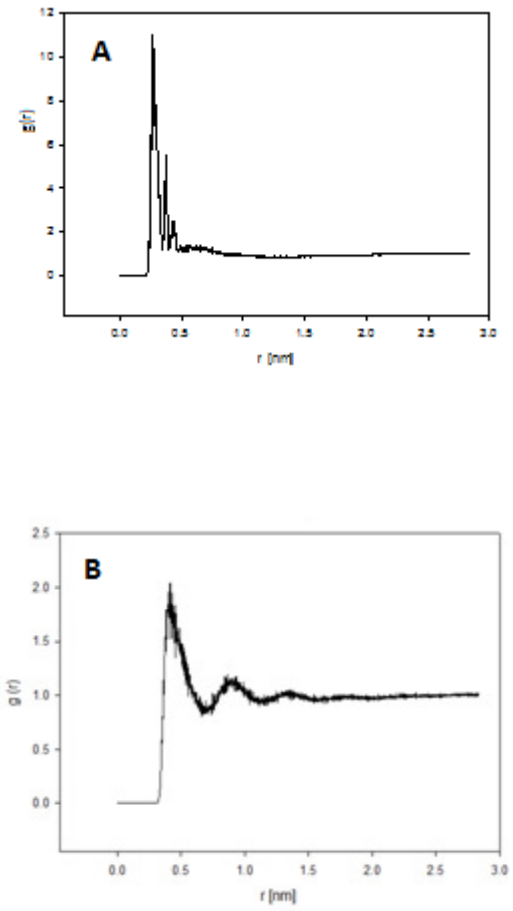

Fig.34. The RDF for 256 PPP at 315 K:
a) ester oxygens, b) CH_3 groups at the end of hydrocarbon tails.

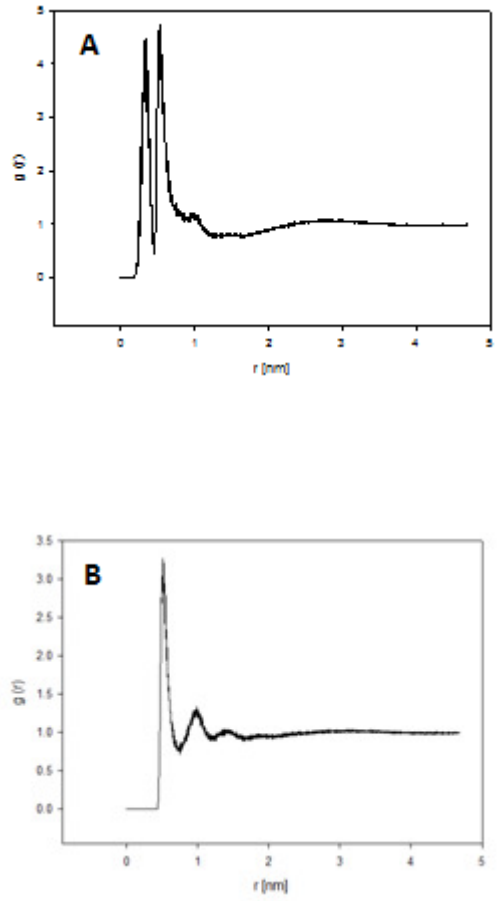

Fig.35. The RDF for the coarse-grain model:
a) for the three glycerol beads at 353K, b) CH_3 groups at the end of hydrocarbon tails.

At this stage, however, we can assume that the molecules mimic a gel-like behaviour, being the pre-crystallization stage for triacylglycerols. As far as the crystallization goes, we would expect to observe a layered sheet of molecules with their hydrocarbon tails stretching in opposite directions, and with a thin array of oxygen moieties in the middle. This kind of high degree of organization is characteristic to all lipids actually, and has been proved especially in the case of bilayer structures of lipids forming biomembranes [11].

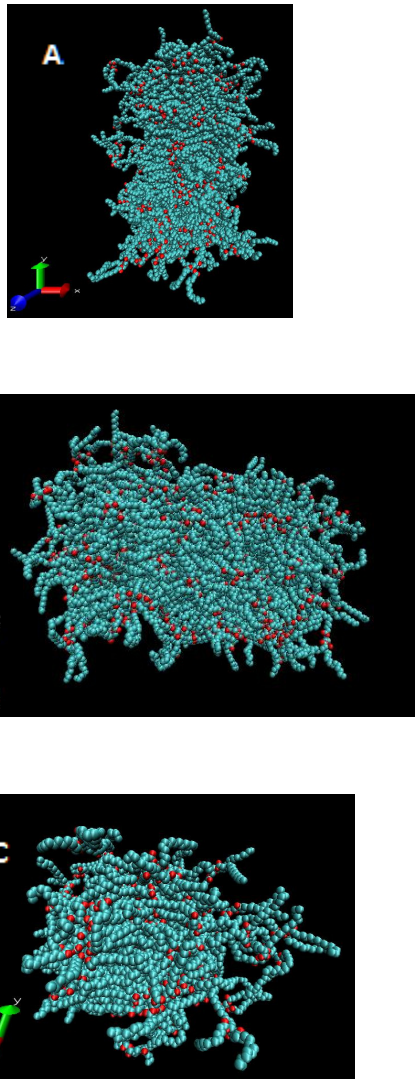

Fig.36. Organizational changes of two tripalmitin systems under different temperature conditions: a) 256 PPP at 353 K, b) 256 PPP at 315 K, c) 128 PPP at 315 K.

4.4. Diffusion Coefficient

In order to quantify the degree of ordering of liquid TAGs molecules during the cooling process, we used a self-diffusion coefficient. It was extracted from the mean square displacement (MSD) of the centers of mass of all molecules in each system, according to eq.32. The results are summarized in Table 9.

We can clearly see that with the decrease in temperature, the self-diffusion coefficients also decrease. This is expected as the movement of particles becomes more restricted at low temperatures and at higher degrees of organization. Also, the increasing friction between molecules causes them to slow down what can be observed in smaller diffusion rates. We would like to note here that no apparent phase change has been observed, in which case the self-diffusion coefficient would change on the scale of magnitude(s). The differences in our self-diffusion coefficients are small what we believe is due to relatively short times (approx. 100 ns), with respect to a real time crystallization process, over which the calculations were performed. Nonetheless, the expected qualitative trend in the TAGs movement has been proven. Given enough time to re-organize the TAGs would be expected to enter the gel-like phase which would be an onset of crystallisation process. Furthermore, at this point it is hard to compare the calculated values with ones from experiments, because of much longer time scales followed in experiments. Unfortunately, little information about calculated self-diffusion coefficients can be found for systems similar to ours hence the validation of the results is hard.

Table 9. The diffusion coefficient values for all simulated systems.

System	Temperature [K]	$D_{calc.}$ [cm^2/s]	$D_{exp.*}$ [cm^2/s]
128 PPP	353	5.78 e-7 (+/- 2.52 e-7)	3.7 e-7 at 350 K [19]
	338	2.15 e-7 (+/- 1.55 e-8)	n/a
	315	1.84 e-7 (+/- 6.10 e-8)	1.0 e-9 at 310 K [19]
256 PPP	353	8.86 e-7 (+/- 4.41 e-8)	3.7 e-7 at 350 K [19]
	338	3.71 e-7 (+/- 5.77 e-8)	n/a
	315	11.93 e-7 (+/- 8.3 e-9)	1.0 e-9 at 310 K [19]
2048 CG PPP	353	11.33 e-7 (+/- 3.56 e-8)	3.7 e-7 at 350 K [19]
256 SSS	353	4.98 e-7 (+/- 1.11 e-7)	n/a

*experimental value based on literature

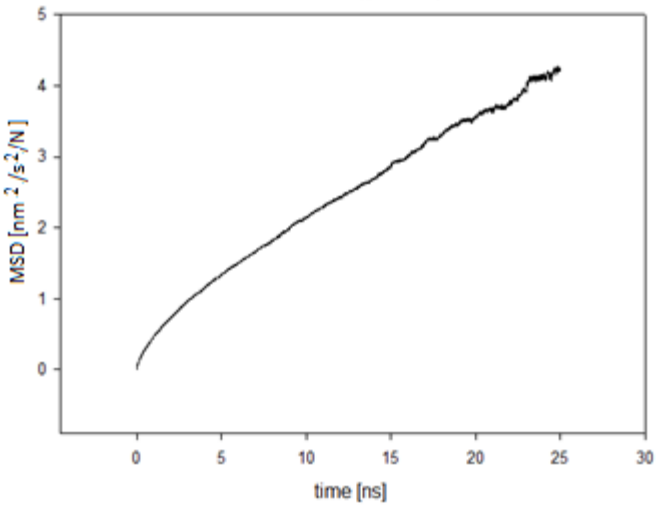

Fig.37. An example of a MSD plot for 256 PPP molecules at 338 K.

5. Summary

5.1. General conclusions

Tempted by the technological development and on-going exponential growth of computation resources, we decided to look into the possibility of using molecular dynamics simulation to gain more insight into the phase transformations of edible fats and oils. Their building blocks, triglycerides, express complex polymorphic behaviour which determines greatly process management of fat mixtures. Therefore, understanding fats interactions and behaviour on a molecular level promises better process operations and better quality products.

The aim of this work was to utilize molecular dynamics simulation to predict the state and properties of TAGs and their mixtures. It was only partially achieved, mainly due to the time consuming process of overcoming computational challenges of molecular dynamics. The objective has overgrown our time restrictions for this project; however, it laid down good foundations for future investigations. We were not able to finish our work with simulations of the tripalmitin-tristearin mixtures on time to include their findings in this thesis. However, the one-component triglyceride models of PPP and SSS proved to be stable enough to undergo lengthy but fruitful simulations. Now, they can be used further for construction and simulation of the mixture of thereof.

In order to foresee the microscopic state of TAGs and, on that basis, to draw conclusions on their macroscopic properties, we built several TAGs systems and treated them with simulated annealing to impose structural changes leading to their crystallization. To do that, we used molecular modelling tools from Materials Studio and GROMACS software and we adopted the united-atom and coarse grain models. For the united-atom model, the force field of choice was GROMOS-96 from the

GROMACS package, which was especially parameterized for lipid molecules and proved to produce reliable results. The coarse-grain model was built with the Martini force-field.

In the course of this work we had learnt that it is possible to predict the density of a TAG system that is very close to its experimental value (except the coarse-grain approach); hence, the molecular model can be considered reliable. When the temperature of the simulations was lowered, we were able to observe that the systems in liquid states started to become denser and the molecules started to re-organize themselves to reach more thermodynamically stable structure. While the time scale of those simulations was not long enough to observe the crystallization process, the witnessed changes can be considered as the onset of the crystallization process. Further visual analysis confirmed these findings.

Through the analysis of RMSD calculations we came to the conclusion that all of the subjected systems underwent a major re-reorganizational event, which was expressed by a large well on the RMSD graph. After that, they came back to their relatively stable state. The events were not caused by any change in the simulation conditions. Also, they were not significant enough to be considered phase changes, though they were significant enough to be registered.

Another interesting phenomenon observed was the grouping behaviour of the TAGs oxygen atoms. As the temperature was lowered, the oxygen atoms started gathering in larger groups, pushing away the hydrocarbon tails in opposite directions. This behaviour seems to show how triglycerides form layered structures which are widely observed in bio-membrane lipids MD simulations. It is also understandable that the negatively charged ester moieties within the centers of TAGs molecules try to isolate themselves from non-polar hydrocarbon chains.

On the basis of RDF calculations, it was possible to confirm the physical state of TAGs systems in different temperature conditions. At the temperatures of 353 K and 338 K, the tripalmitin systems were in liquid states. We might speculate though, that

the peaks registered for the oxygen moieties in the short-range interactions region (0.0-0.5 nm) indicate that the oxygen atoms tend to stay very close to each other. The CH_3 groups at the ends of hydrocarbon chains are allowed to move relatively freely ensuring the fluidity of a system. At the lowest temperature (315 K) we were able to observe more pronounced changes in the systems structure. We estimated that the system had started the phase transition though it was a very early stage of that process.

Lastly, we considered the diffusion coefficient as another way to analyze the phase behaviour of TAGs. In this case, we observed only slight changes in the diffusion coefficient after each temperature change. It`s value decreased as a result of tighter molecule packing and increased friction between molecules.

Summing up, our simulations were too short for the kind of phase behaviour we anticipated to observe with edible fats, but the qualitative trends were succesfully preserved. Given more time and even more powerful computational resources, our molecular models for triacylglycerols should be able to crystallize.

5.2. Future work

Our molecular models proved to be able to mimic first significant structural changes of triacylglycerols being on their way to crystallization. Next step should be the continuation of the above work on longer timescales in order to register more pronounced structural changes and actual phase transition from the liquid to solid states of fats.

What is more, it is also believed that now the above mathematical models can serve as building blocks of new mixed triacylglycerols systems. It is of outmost importance to further investigate the phase behaviour of mixtures of triacylglycerols, because these are the most realistic representations of fats and oils people deal with in real life situations and would therefore add the most meaningful insights from the industrial process management perspective. We

suggest starting with binary mixtures of TAGs as more challenges are expected to appear due to mixing and miscibility of two different TAG molecules.

6. Bibliography

[1] http://en.wikipedia.org/wiki/Lipid

[2] Humprey K.L., PhD thesis, University of Alberta, 2007

[3] Marangoni A.G., Fat crystal networks, New York: Marcel Dekker.

[4] Gunstone F.D, Harwood J.L., Dijkstra A.J., The lipid handbook, CRC Press, 2007.

[5] Bockish M., Fats and oils handbook, AOCS Press, 1998

[6] Rabasco Alvarez M.B., Rodriguez M.L.G., Grasas y Aceites, 51, Fasc. 1-2, 2000, 74-96

[7] Akoh C.C., Oi-Ming L., Healthful lipids, CRC Press, 2005.

[8] P. Lonchampt and R. W. Hartel, Eur. J. Lipid Sci. Technol. 106 (2004) 241–274

[9] Marrink S.J., Mark A.E., Biophys. J.,2004, 87(6) 3894–3900

[10] Marrink et al., J. Phys. Chem. B, Vol. 108, No. 2, 2004

[11] Tieleman et al., Biochimica et Biophysica Acta 1331, 1997, 235-270

[12] Tieleman et al., J. Phys. Chem. B 2000, 104, 6380-6388

[13] Bachar et al., J. Phys. Chem. B, Vol. 108, No. 22, 2004

[14] Schuler et al., Eur Biophys J., 2001, 30, 330-343

[15] Hyvonen et al., Eur Biophys J., 2005, 34, 294-305

[16] Chandrasekhar I., van Gunsteren W.F., Eur Biophys. J., 31 (2002), 89-101

[17] Schuler L., Daura X., van Gunsteren W.F., Journal of Computational Chemistry, Vol. 22, No. 11, 1205–1218 (2001)

[18] Sum A.K. et al., J. Phys.Chem. B, 2003, 107, 14443-14451

[19] Hall A., J.Phys.Chem. B, 2008, 112, 13772 -13782

[20] Small D. et al., Handbook of lipid research 4 –The physical chemistry of lipids, Plenum Press, Ny, 1986

[21] van Lavengelde A. et al., Acta Cryst. B 55, 1999, 114-122

[22] Himwan C. Et al., Advances in colloid science and interface science, 122, 2006, 3-33

[23] Sikorski Z.E., Kolakowska A., Chemical and functional properties of food lipids, 2003 by CRC Press LLC

[24] Sato K. et al., Progress in Lipid Research, 38, 1999, 91-116

[25] K. Sato, Chem. Eng. Sci., 56, 2001, 2255-2265

[26] Yano et al., J.Lipid Res., 40, 1999, 140-151

[27] http://www.chemicalbook.com

[28] Kellens M. Et al., Chemistry and Physics of Lipids, 58 (1991), 131-144

[29] Leach A.R., Molecular modelling – Principles and applications, Wesley Longman Ltd. 1996

[30] Krivov S., Karplus M., *Proc. Natl. Acad. Sci. USA,* 2004, 101, 14766-14770

[31] Lubartsev A.P., Eur. Biophys. J., 2005, 35, 53-61

[32] Heller H. et al., J.Phys. Chem., 1993, 97, 8343-8360

[33] Marrink et al., J. Phys. Chem. B 2000, 104, 12165-12173

[34] Rahman A., Phys. Rev. A, 136, 405-411

[35] Feller S.E., Computational modelling of membrane bilayers, Current topics in membranes, 60, 2008

[36] The Gromacs development team, The Gromacs Manual - Groningen Machine for Computer Simulations, v.3.2; University of Groningen, Netherlands

[37] Scott et al., J.Phys.Chem. A, 1999, 103, 3596-3607.

[38] S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. de Vries. The MARTINI forcefield: coarse grained model for biomolecular simulations. J. Phys. Chem. B, 111:7812-7824, 2007.

[39] J.P. Ryckaert *and A.* Bellmans, Molecular dynamics of liquid n-butane near its boiling point, Chem. Phys. Lett. 1975 (1975), 123

[40] Kirkpatrick S. et al., Optimization by simulated annealing, Science 220, 670-680, 1983

[41] M. P. Allen, D. J. Tildesley, Computer simulations in chemical physics, Springer, 1st. Ed., 1993.

[42] van Lavengelde et al., Journal of AOCS, Vol. 76, no. 5 (1999), 603-609

[43] van Lavengelde et al., *Chem. Mater.* 2001, *13,* 1089-1094

[44] Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M., LINCS: A linear constraint solver for molecular simulations. J.Comp.Chem., 18, 1463-1472, 1997.

7. Appendix 1

LINCS algorithm [36]

It is an algorithm which adjusts bonds to their actual lengths after an unconstrained update. It is considered to be more stable and faster than SHAKE algorithm (another algorithm option in GROMACS).

If we consider a system with N particles and positions given by a 3N vector r(t), for MD the equations of motion are given by Newton's law:

$$\frac{d^2r}{dt^2} = M^{-1}F$$

(Eq. 33)

where F is the 3N vector and M is a 3N x 3N diagonal matrix, which contains the masses of the particles. The system is constrained by K time-independent constraint equations:

$$g_i(r) = |r_{i1} - r_{i2}| - d_i = 0 \qquad i = 1, \dots, K$$

(Eq. 34)

The LINCS algorithm uses two steps. First, the projections of the new bonds on the old bonds are set to zero. In the second step, a correction is applied for the lengthening of the bonds due to rotation. For complete derivation, please refer to Hess et al. [44] (Fig.41).

A new notation is introduced for the gradient matrix of the constraint equations which appears on the right hand side of the equation:

$$B_{hi} = \frac{\delta g_h}{\delta r_i}$$

(Eq.35)

Here, B is a K x 3N matrix and contains the directions of the constraints. The constrained coordinates r_{n+1} are related to the unconstrained coordinates r^{unc}_{n+1} as follows:

$$r_{n+1} = (I - T_n B_n) r_{n+1}^{unc} + T_n d = r_{n+1}^{unc} - M^{-1} B_n (B_n M^{-1} B_n^T)^{-1} (B_n r_{n+1}^{unc} - d)$$

(Eq. 36)

In the first step, the projection of the new bonds onto the old directions is applied. To correct for the rotation of bond *i*, the projection of the bond on the old direction is set to:

$$p_i = \sqrt{2d_i^2 - l_i^2}$$

(Eq.37)

where l_i is the bond length after the first projection. The corrected positions are:

$$r_{n+1}^* = (I - T_n B_n) r_{n+1} + T_n p$$

(Eq.38)

This correction during MD is applied once. The relative constrained deviation after this process will be less than 0.0001 for every constraint.

In energy minimisation, however, this might not be accurate enough, so the number of iterations is equal to the order of expansion.

Half of the CPU time is spent on the inversion of the constraint coupling matrix $B_n M^{-1} B^T_n$, which has to be done every step. This K x K matrix has $1/m_{i1}+1/m_{i2}$ on the diagonal. The off-diagonal elements are only non-zero when two bonds are connected, then the element is $\cos\varphi/m_c$, where m_c is the mass of the atom connecting the two bonds and φ is the angle between them.

The matrix T is inverted through a power expansion. A K x K matrix S is introduced which is the inverse square root of the diagonal of $B_n M^{-1} B^T_n$. This matrix is used to convert the diagonal elements of the coupling matrix to one:

$$(B_n M^{-1} B_n^T) = SS^{-1}(B_n M^{-1} B_n^T)^{-1} S^{-1} S = S(SB_n M^{-1} B_n^T S)^{-1} S = S(I - A_n)^{-1} S$$

(Eq.39)

The matrix A_n is symmetric and sparse has zeros on the diagonal. Hence a simple trick can be used to calculate the inverse:

$$(I - A_n)^{-1} = I + A_n + A_n^2 + A_n^3 + \ \ldots$$

(Eq.40)

This is only valid when the absolute values of all eigenvalues of A_n are smaller than one. In molecules with only bond constraints the connectivity is so low that this will always be true. For molecules with all bonds constrained the eigenvalues of A are around 0.4, which means that with each additional order in the expansion eq.40 the deviations decrease by a factor of 0.4.

The accuracy of LINCS depends on the number of matrices used in the expansion eqn. 40. For MD calculations, an order expansion equal to 4 is enough. LINCS generates a warning when in one step a bond rotates more than a predefined angle. This angle is set by the user in the input file for mdrun tool of GROMACS.

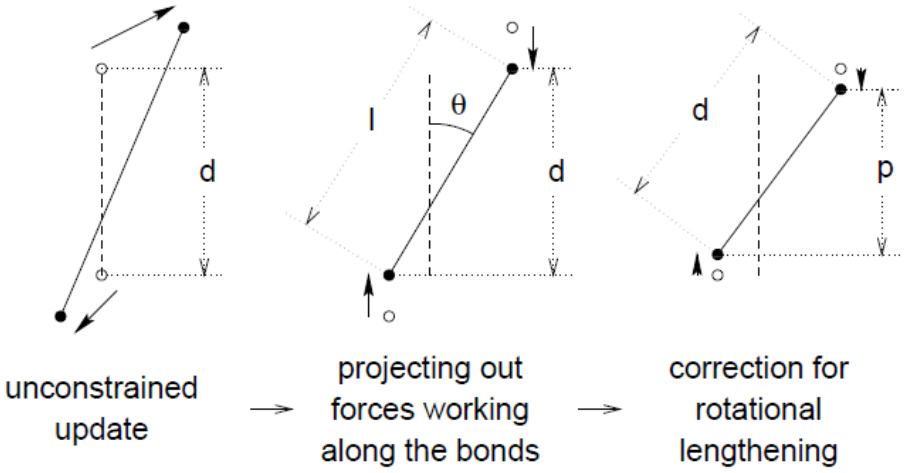

Fig. 38. The position updates needed for one time step. The dashed line is the old bond of length *d*, the solid lines are the new bonds, $l=d\cos\vartheta$ and $p=(2d^2-l^2)^{1/2}$