ANTONIJEVIC, TODOR, Ph.D. Theoretical Modeling of Physical Processes in Low Density Lipoprotein Nanostructures (2015)
Directed by Dr. Joseph M. Starobin. 79 pp.

The research work aims to analyze LDL structure and study the structural influence on physical properties of the particle by means of computer simulations. Our approach is based on theoretical statistical mechanics and modern computational molecular dynamics. We analyzed theoretically molecular interactions to identify the major statistical factors contributing to the ordering and stability of LDL particle. Combining available experimental data with developed statistical model allowed us to predict thermodynamic properties of LDL. Conversely, dynamical properties of LDL particles were investigated employing methods of molecular dynamics. Protein moiety was obtained using homology modelling, fold recognition, and *ab-initio* structure prediction.

The obtained heat capacity does not have sharp peak typical for a bulk sample of material. Instead, we observed curve which spreads over a finite temperature interval as for a sample consisting of small subsystems with finite number of molecules. Our dynamical model of LDL particles readily supports experimental findings that below biological temperatures laminar structures observed in these particles could indeed form due to ordering of cholesterol esters.

THEORETICAL MODELING OF PHYSICAL PROCESSES IN LOW DENSITY LIPOPROTEIN NANOSTRUCTURES

by

Todor Antonijevic

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2015

Approved by

_____
Committee Chair

APPROVAL PAGE

This dissertation written by TODOR ANTONIJEVIC has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair_____

Committee Members_____

_____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

In the United States 29% of all deaths in the year 2002 were caused by cardiovascular disease (CVD) or sometimes called heart disease. The same trend is observed in the world with striking increase predicted in the future. The reason lies in continuous world industrialization [1]. Although genesis and progression of atherosclerosis, the major cause of CVD [2], are affected by Low Density Lipoproteins (LDL) [3] many questions related to molecular details of this particle are still unanswered. The reason for this is the complex nature of LDL particle [4]. Even though numerous experimental approaches were conducted in the past few decades to reveal the structure of LDL, still challenging particle allows only experimental resolution of 16Å which uncovers merely the overall shape of the LDL [5].

Low Density Lipoproteins is a macro molecule with hydrophobic core surrounded by phospholipid monolayer and Apolipoprotein B-100 (apoB-100), which is on the surface and is partially dipped into the monolayer. The main function of LDL is to provide transport of cholesterol esters to cells through vascular system. ApoB-100 protein plays essential role in recognition mechanisms which allow binding of LDL to the receptor of particular cell. Specifically, the absorption of LDL by the cells depends on configuration of apoB-100. In spite of its importance in molecular absorption, the structure of the apoB-100 protein is poorly understood. So far, the structure for only first 1150 amino acids was modeled.

Thus, evaluating molecular structure and dynamics of this protein is crucial not only for understanding the linking system between LDL and cells, but also its implications for development of heart diseases. Better understanding of apoB-100 protein will lead to better comprehension of LDL as a whole: its structural heterogeneity, complex behavior, and intrinsic dynamics of this particle.

Our strategy for clarification of the LDL structure is to employ statistical mechanics, structural bioinformatics, and molecular dynamics. The synergy of these approaches should lead to more detail theoretical analysis of this complex particle. Molecular interactions will be analyzed through computational simulations. Continuous development of computational power today allows one to deal with problems that were not possible a decade ago [6].

Analysis of the molecular interactions suggests that identification and isolation of the major contribution to the stability of the particular phase allows formation of the statistical model. Combining available experimental data with the statistical model allows calculation of thermodynamic properties of the system. Thus, our first goal is to use statistical methods to simulate the phase transition processes within the LDL and assess its thermodynamic behavior. This is important because thermodynamic behavior of the sample could be potentially beneficial for current estimations of LDL cholesterol.

Our second goal includes clarification of the complex structure of LDL, with emphasis on the apoB-100 moiety of the LDL using structural bioinformatics and molecular dynamics. Our detailed model of LDL heterogeneous structure is consistent with experimental observations. Detailed structural model of apoB-100 can be used in future for evaluating the recognition mechanism between receptor and lipid droplet with possible medical applications.

CHAPTER II

LITERATURE REVIEW

LDL is made of ApoB-100 protein and ~3000 lipids. These molecules are structured into LDL's hydrophilic outer shell and hydrophobic core. While ApoB-100, phospholipids and unesterified cholesterol constitute outer shell, LDL's core contains cholesteryl esters (CE), triglycerides (TG) and some of the unesterified cholesterol (UE).

The main purpose of LDL is to deliver cholesterol to the cells. Cholesterol is mostly used as a building block in cell membranes and as precursor for certain steroid hormones and bile salt. When cell needs cholesterol it expresses LDL receptors on the surface of the cell wall. These receptors will bind LDLs and carry them into the cell. Thus, uptake of LDL particles is based on receptor-mediated endocytosis. Finally, LDL is delivered to lysosomes where cholesteryl esters are hydrolyzed [7].

The uptake and the size of the LDL particles are governed by genetic and environment factors. For example, genetic factors may influence LDL receptors by lowering their function. This will cause buildup of cholesterol in the plasma [7]. In addition, genetic factors may influence the size of the particle as observed, for example, in the Québec Family Study where significant correlations were observed in the size of the LDL particle between parents and offspring and between siblings [8]. It was estimated that 40% of influences on the size of LDL are due to genetic factors [9]. Significance of this estimation is that it also emphasizes importance of environmental factors such as smoking, abdominal obesity, low fat diet and diet high in carbohydrates on the particle size [10].

As a result of all these influences, LDLs are heterogeneous particles that vary in size (18-25nm), surface charge, chemical composition, buoyant density, etc. [11]. These properties influence LDL's behavior in the subendothelial space.

Since highly heterogeneous, LDL particles are divided in subclasses. NMR measurements distinguish four LDL subclasses: small LDL particles (Pattern B) with diameter between 18.3 nm and 19.7 nm, intermediate LDL particles with diameters 19.8-21.2 nm, large LDL particles (Pattern A) with diameter in the range from 21.3 up to 22.7 nm, and Intermediate-density lipoprotein ILD with diameter between 23 nm and 27 nm [12]. Concentration between small and large LDL particles is inversely correlated [13]. In addition, different subclasses have different binding affinities to the receptor [14], susceptibility to oxidative modifications, etc. For example, extremely interesting is a subclass of small and dense LDL particles (sdLDL). When compared with large LDL, this subclass is more atherogenic and susceptible to modifications. Thus it has been identified as additional risk factor for atherosclerosis [15]. There are few unique characteristics of sdLDL that contribute to their atherogenicity. These particles have lower affinity for the LDL receptor which allows them to stay longer in plasma and expose longer to the modification factors. Also, they have increased susceptibility to oxidation and their size allows them to easily penetrate the endothelium.

Remodeling of LDL subclasses depends on plasma TG levels and it is reflected through changes in the core lipid composition and the particle size. If plasma TGs are elevated then core lipid composition is altered through reaction catalyzed by cholesterol ester transfer protein (CETP). CETP activity will cause exchange between TG molecules that originate from the very-low density lipoprotein (VLDL) particle with cholesteryl ester (CE) molecules from the core of LDL. Here VLDL is a donor of TGs and acceptor of CEs, and LDL is a donor of CEs and acceptor of TGs [16]. This will cause formation of CE-depleted and TG-enriched LDL particle

which then becomes a substrate for hepatic lipase (HL) enzyme. HL has high binding affinity for the LDL and, as a multifunctional protein, hydrolyzes TGs and phospholipids. This leads to structural remodeling and transformation of large TG-rich LDL particle into a small LDL [17]. Final core composition of small LDL is dependent on CETP activity and plasma triglycerides levels. Based on these influences core composition of small LDL particle may end up with normal CE content (CE/TG>4) or depleted (CE/TG<4) [18]. LDL particle that has 3nm smaller diameter, compared to the large particle, will also transport 40% less CE cargo. Thus, an increase of 70% in the number of particles is needed in order to carry the same amount of CE [19].

CEs in the core of the LDL particle undergo phase transition if the amount of TGs is not sufficiently high. Despite decades of research still there is no clear evidence for functional importance of this structural transition. It has been proposed that this phase transition has effect on LDL recognition process. In that way it may have a role in the progression of atherosclerosis. Regardless of the unknown physiological role of the phase transition, structural organization of the core lipids below phase transition is still matter of debate. First reports indicated that CEs are packed into the smectic-like concentric layers that give an overall spherical shape of the particle [20]. Later on, ellipsoidal shape was reported with dimensions 25x21x17.5nm based on cryo-EM images of the LDL [21]. This research also challenged previously established model of radial quasi-smectic structure of the lipid core for temperatures that are below phase transition.

Cryo-EM images show layers with interlayer distance of 3.6nm which corresponds to the molecule length. For temperatures above phase transition core lipids are in the liquid phase and overall shape is roughly spherical. From these examples it can be speculated that temperature does not affect only the shape of the particle but also the structure of ApoB-100. We have to note that a problem of TGs packing is also unanswered. Different explanations were offered from compartments inside the core rich in TGs to the random distribution of TGs.

Beside the problem of the LDL particle size, which is determined by the core lipids, uptake of the LDL is strongly dependent on the conformation of ApoB-100. This conformation affects protein binding capabilities to the LDL receptor and thus it is closely related to various diseases (atherogenesis, hypocholesterolemia, etc.). Due to complex and heterogeneous nature of LDL, experimental approach did not succeed to decipher ApoB-100 crystalline structure. Also, despite several experimental attempts to elucidate the secondary structure of the protein it still remains a matter of debate. Nevertheless, experimental approach was able to put forward the global layout of ApoB-100 as a thick ribbon that penetrates phospholipid monolayer and wraps around the LDL particle [22] and to identify LDL-receptor binding sites [23]. In order to fulfill the missing gaps, a theoretical prediction of the ApoB-100 structure was proposed [24]. This model consists of eight structural domains. For each domain 3D structural prediction was proposed and lipid-domain interactions were suggested. This allowed an overall shape prediction of ApoB-100 at the surface of LDL. Although this model is very detailed resent improvements in crystallography and threading algorithms propose new template models.

CHAPTER III

STATISTICAL MODEL OF ORDERING

Planar LDL Structures

Before proposing planar lattice model, it is important to identify major interactions that occur between CE molecules, mainly cholesteryl oleate (CO) and cholesteryl linoleate (CL). In this research we will focus on CO since its crystal structure is known.

At low temperatures pure isolated CO is in the crystal state. Gao and Craven were able to determine the structure of CO crystal at *123K* [25]. They observed monoclinic unit cell with dimensions: $a=12.356(2)$ Å, $b=8.980(3)$ Å, $c=18.382(2)$ Å, $\beta=85.49(2)^{\circ}$. Each cell has two molecules of the same type packed in antiparallel arrangement (Figure 1). This type of molecular packing results in formation of *monolayer type II* crystal structure with significant: (a) ring-ring interactions between antiparallel ester pairs, and (b) chain-chain interactions between acyl chains from opposing monolayers (Figure 2) [26].

Heating process of pure sample of CO will result in melting of solid CO. Close to the melting temperature a large pre-transitional effect is detected. This pre-transitional effect can be observed ,for example, on heat capacity curve [27]where it spans over *4°C* before melting and it is attributed to larger and larger molecular motion. Interesting to note is the fact that in the isotropic state ring-ring and chain-chain interactions still exists as a prevailing type of interactions [28]. Even more, these fundamental interactions are, as suggested by ^{13}C NMR spectroscopy, responsible for short range ordering within microdomains of liquid CO in vicinity of the phase transition [29]. In particular, short range ordering is attributed to strong hydrocarbon chain associations present in molten liquids of CO and CE.

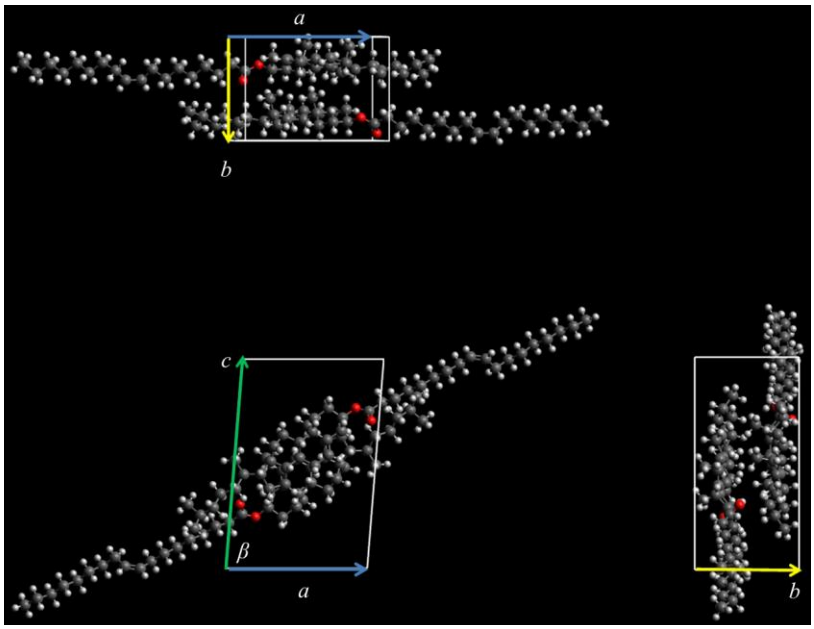

Figure 1. Top, Front and Right Side Projections of CO Unit Cell at 123K.

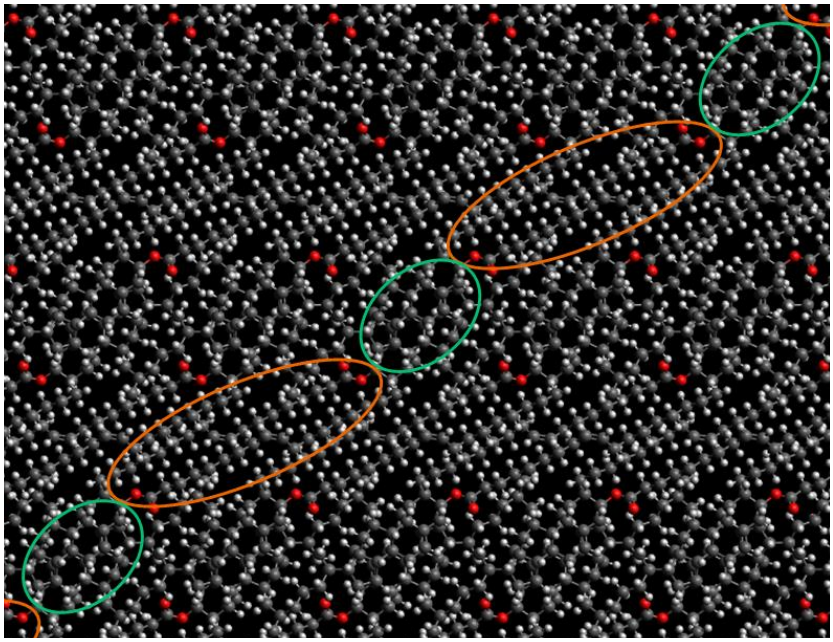

Figure 2. Molecular Interactions between CO Molecules in the Crystal State. Orange: Chain-Chain Interactions between Acyl Chains from Opposing Monolayers; Green: Ring-Ring Interactions between Antiparallel Ester Pairs.

During cooling process CO molecules undergo two additional phase transitions, liquid to cholesteric transition followed by cholesteric to smectic transition. While ring-ring interactions are important in ordering of the cholesteric phase, chain-chain interactions are responsible for formation and stability of the smectic phase. Noticing that the same fundamental interactions are responsible for stability of the liquid-crystal phase of CE, Ginsburg was able to propose a model for the smectic phase of CE (Figure 3). This model is based on *monolayer type II* crystal structure and it assumes "an antiparallel arrangement of cholesteryl ester molecules in which adjacent cholesteryl ester molecules contribute fatty acyl chains to adjacent layers resulting in inter digitation of acyl chains" [26].

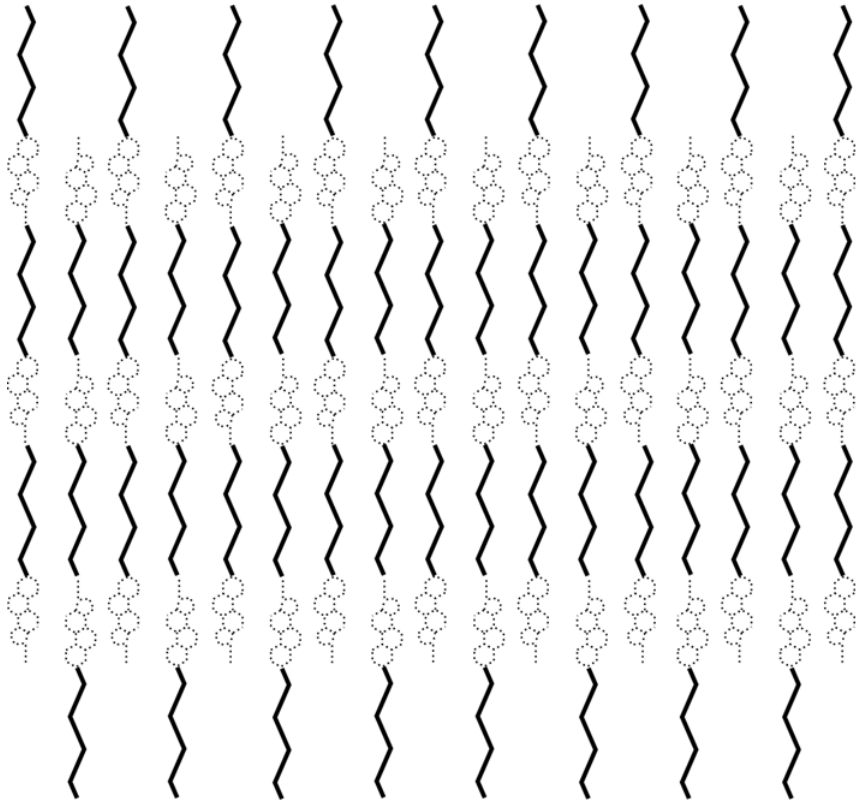

Figure 3. Ginsburg Model for CE Smectic Phase.

Thermodynamic behavior of CE in LDL is somewhat different. When a sample of LDL particles is heated CE will undergo smectic to liquid phase transition. This process is perfectly reversible during cooling [20]. Sometimes, this phase transition is described as order-disorder transition, referring to order as a smectic phase and to disorder as a liquid state. Furthermore, cholesteryl esters in intact LDL do not form cholesteric or crystal phase. From the foregoing, it is reasonable to assume that chain-chain interactions are the major type of interactions between CE molecules in the LDL and that this type of fundamental interactions is responsible for the observed structural transition between smectic and isotropic states.

Previous analysis is in compliance with well-known fact that molecular end chains take part in the ordering process of the liquid-crystalline state. This is especially true for the research of Marčelja who investigated a single chain in the molecular field due to the neighboring molecules and was able to evaluate ordering process by summing all conformations of the chain [30]. He noticed that in the liquid-crystalline state and in the liquid state different parts of the molecule are in continuous contact with neighboring molecules. These contacts may be represented in the Hamiltonian of the system in terms of three types of interactions:

$$\mathcal{H} = \mathcal{H}_{internal} + \mathcal{H}_{dispersive} + \mathcal{H}_{steric} \quad (1)$$

The first term takes into account conformational energy of all molecules in the system. The second term describes dispersive Van der Waals interactions between the neighboring molecules and the third term covers steric repulsions between different molecules. Steric contributions can be neglected when a large number of conformations are available to the molecule [30].

By lumping together all possible excited states of the chain in one representative state Doniach was able to propose simplified (coarse-grained) model for chain-chain interactions. In this model two states are allowed, the low state of the molecular chain and the excited state.

Applying simplification to previous work of Marčelja, Doniach was able to obtain model for lipid melting that is analog to the Ising magnet model [31]. This model was successfully applied to describe transitions in lipids [32], small unilamellar vesicles (SUV) [33], lipid-protein systems [34], etc.

Our previous analysis of the molecular interactions between CE molecules in the LDL, Ginsburg model for the smectic phase of CE and experimental suggestion that "strong hydrocarbon chain associations may take place perhaps within microdomains in the liquid phase" of cholesteryl esters [26] persuade us to accommodate previous two state model for the purpose of description of phase transition in the core of the LDL particle.

Statistical Mechanics Method for the Phase Transition Description

A smectic mesomorphic phase is characterized by well-defined layers. Depending on the orientation of the molecules to the layer's plane one can distinguish few smectic types. Major CE of the LDL particle form so called smectic-A type [27] [35]. In the smectic-A phase the long axis of the molecule is perpendicular to the layer's plane (Figure 3). Layers are flexible and glide freely over one another with low viscosity [36]. This means that layers are uncoupled and that they could be treated independent of each other. In other words, instead of treating a phase transition of three dimensional core we can concentrate only on one layer of interactions.

Coarse-grained models do not consider atomistic details of the molecule rather they are interested only in the major contributions to the particular transition. In case of two state coarse grained models each molecule is allowed to occupy in one of two possible states.

Here, we consider that molecule may: a) form the smectic phase with enthalpy H_s and entropy S_s, or b) it will contribute to the liquid phase with enthalpy H_i and entropy S_i.

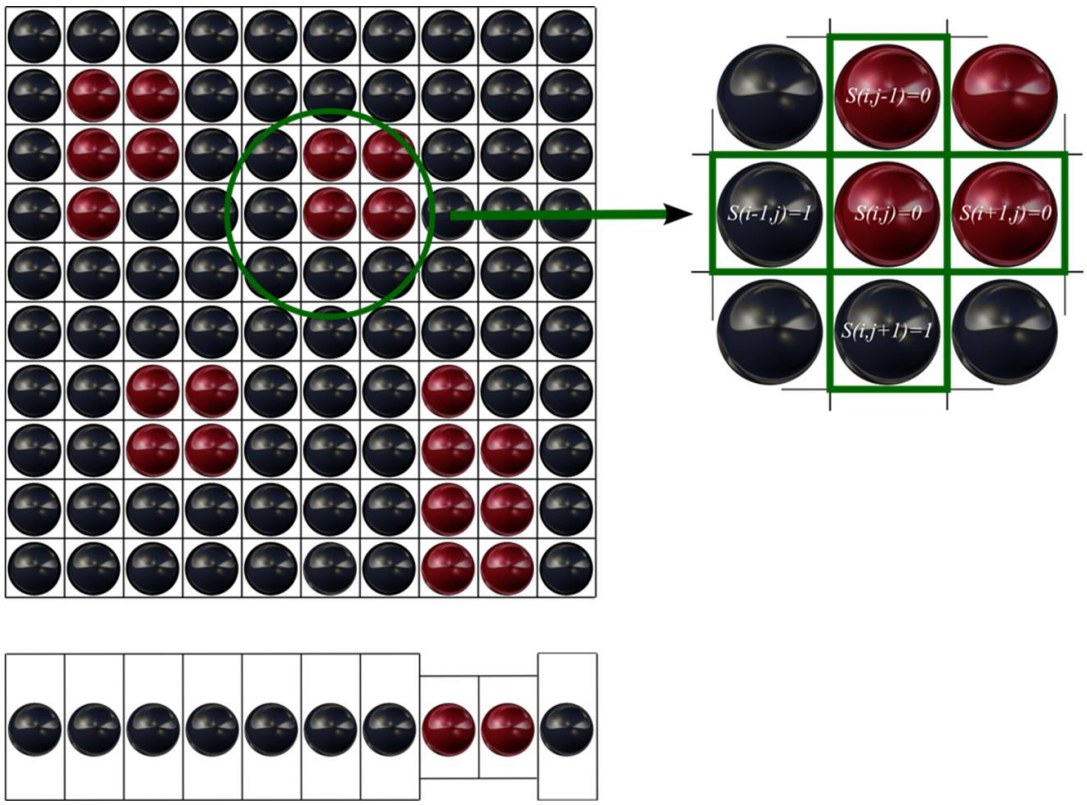

Figure 4. Left: Top and Front View of Two State Coarse Grained Model; Right: Numbering.

Each molecule is surrounded by a constant number of nearest neighbors. The number of nearest neighbors is sometimes called the coordination number '*z*'. In this research we investigated two cases, z=4and z=6. This will allow us to express actual conformation of the system in terms of matrix $\boldsymbol{S} = [s_{ij}]$ where each element s_{ij} represents molecule state. Matrix element may take value $s_{ij} = 0$ in order to represents molecule in the smectic phase or $s_{ij} = 1$ when molecule contributes to the liquid state.

In this statistical model we will assume NpT ensemble with periodic boundary conditions at the edge of the system. The reasoning for this is based on the fact that we are interested in effective interactions between CE molecules that lead to the phase transition, and because all available experimental results were measured at constant number of particles, pressure and

temperature. This means that if sample is in the thermal equilibrium with the surrounding, its equilibrium state is defined by the minimum of the Gibbs free energy. In accordance with equation (1) Gibbs energy of a particular configuration can be expressed as:

$$G(S) = N_s G_s + N_i G_i + N_{ss} G_{ss} + N_{si} G_{si} + N_{ii} G_{ii} \tag{2}$$

The first two terms cover internal energy of the system. G_s is the internal energy of molecule in the smectic phase and G_i is the internal energy of molecule in the isotropic state. N_s is the number of molecules in the smectic phase, and N_i is the number of molecules in the isotropic phase. Last three terms cover dispersive van der Waals interactions. G_{ss} is the interaction energy between two nearest neighbor molecules in the smectic phase and N_{ss} is the number of nearest neighbor smectic-smectic contacts. G_{si} is interaction energies between two molecules where one is in the smectic and the other one in the isotropic state. G_{ii} is interaction between two molecules in isotropic state. N_{si} is the number of nearest neighbor contacts between molecules in smectic and isotropic state, and N_{ii} is the number contacts between two isotropic nearest neighbor molecules.

Taking into account relationships between number of molecules in smectic and isotropic state and their mutual contacts [37]:

$$N = N_i + N_s, \; N_{ss} = \frac{zN_s - N_{si}}{2} \text{ and } N_{ii} = \frac{zN_i - N_{si}}{2} \tag{3}$$

It is possible to rewrite previous free energy of the system as:

$$G(S) = N\left(G_s + \frac{zG_{ss}}{2}\right) + N_i\left(\Delta H - T\Delta S\right) + N_{si} w \tag{4}$$

Here, ΔH is the change in enthalpy associated with transition and ΔS is the change in entropy. These two parameters are measured during phase transition. The only unknown parameter is so called cooperativity parameter *w*. One has to fit this parameter in order to describe the curve of the phase transition. If cooperativity parameter is greater than zero then molecules in similar states tend to form cluster, and when $w = 0$ cooperative effect between molecules will vanish. In the case when cooperativity parameter is below zero system will tend to bring molecules with opposite states in contact. Finally it has to be noted that cooperativity parameter is the sum of an enthalpic and entropic term: $w = w_H - Tw_S$.

Metropolis Monte Carlo Simulations

All steps in the Metropolis Monte Carlo (MMC) algorithm that were applied in the study of the proposed model are shown in Figure 5. In the first step one has to generate an initial configuration. In the next step random number generator (RAN) is employed to generate two random numbers. These numbers will define a particular matrix element s_{ij}. Then, trial configuration is formed by changing the value of the selected element from *0* to *1* if the initial state was *0*, or from *1* to *0* in the case when initial state is *1*. In the next step Gibbs energy difference is calculated between trial and initial state. The trial configuration is accepted if randomly chosen number is lower or equal to the Boltzmann factor, otherwise trial configuration is rejected and initial configuration is kept.

Data collection begins after reaching the equilibrium. Data collection implies recording the number of molecules that form isotropic state N_i and the number of unlike nearest neighbor contacts N_{si} for each configuration. In this way, running the production process for few million steps, it is possible to explore the phase space of the system. Averaging N_i and N_{si} values allowed us to find thermodynamic parameters such as heat capacity:

$$C_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{k_B T^2} \tag{5}$$

Now, one can estimate correct cooperative parameter *w* by running previous algorithm for different values of *w* and fitting the heat capacity curve. The best fit to the curve indicates appropriate cooperativity between molecules.

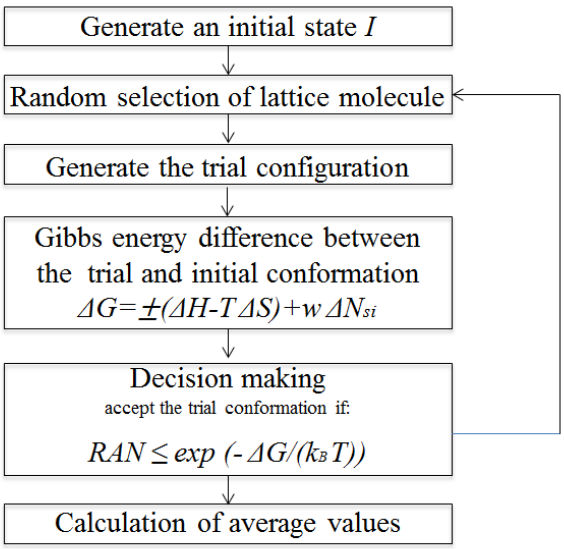

Figure 5. Metropolis Monte Carlo Algorithm.

Results

Using previous model and Monte Carlo approach we obtained heat capacity curve shown in Figure 6. Here we present graphs for heat capacity due to structural transition between smectic and isotropic states for two different coordination numbers. Upper graph represents the system where each molecule is surrounded with six nearest neighbors, and the lower curve is for the system where each molecule is surrounded with four nearest neighbors. Both graphs show good temperature fit to the experimental results.

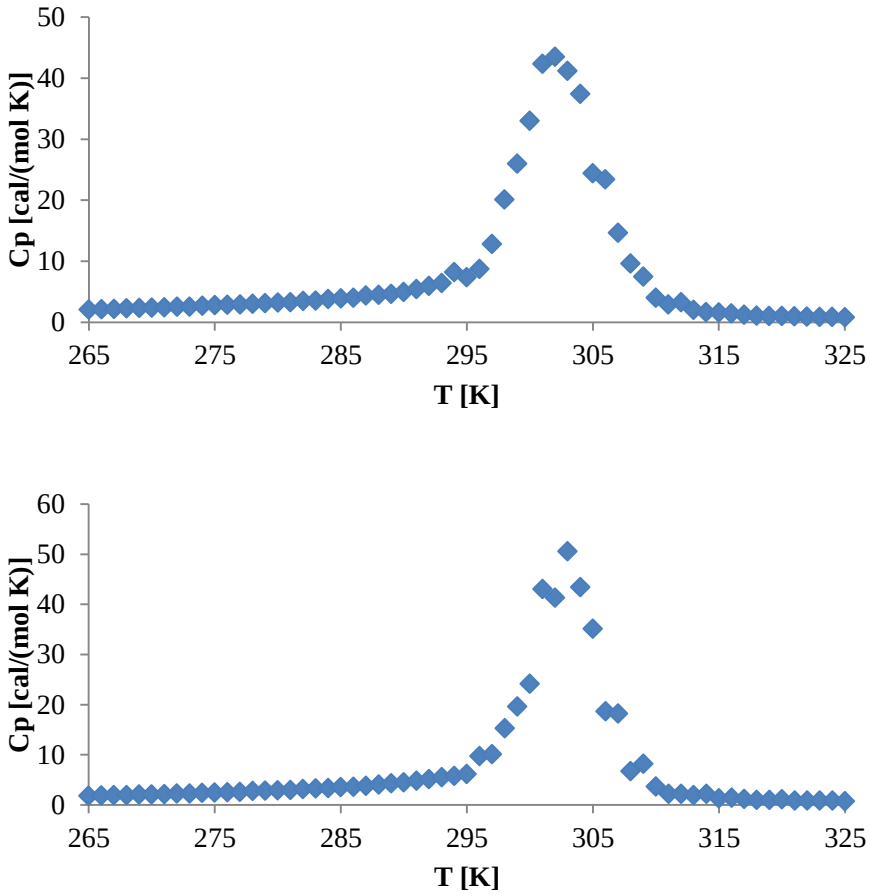

Figure 6. Heat Capacity Curves - Upper: Six Nearest Neighbors, w_H=320 *cal/mol*; Lower: Four Nearest Neighbors w_H=520 *cal/mol*.

Reported mean onset of the phase transition is 17.3±3.3°*C* (290.45±3.3 *K*). Temperature range for the starting point of the phase transition is from 11-23°*C*. Reported peak temperature is 30.3±2.3°*C* (303.45±2.3°*K*) in the temperature range from 26°*C* to 36.5°*C*. Temperature of the end of the transition is in a range between 36°*C* and 44°*C* with average of 41±1.7°*C* (314.15±1.7°*K*) [20]. Magnitude of the phase transition can be confirmed only by comparing our theoretical result with normalized heat capacity curve.

Fractional analysis can give us an average number of molecules in particular state. We extended our calculation to investigate molecules in isotropic state.

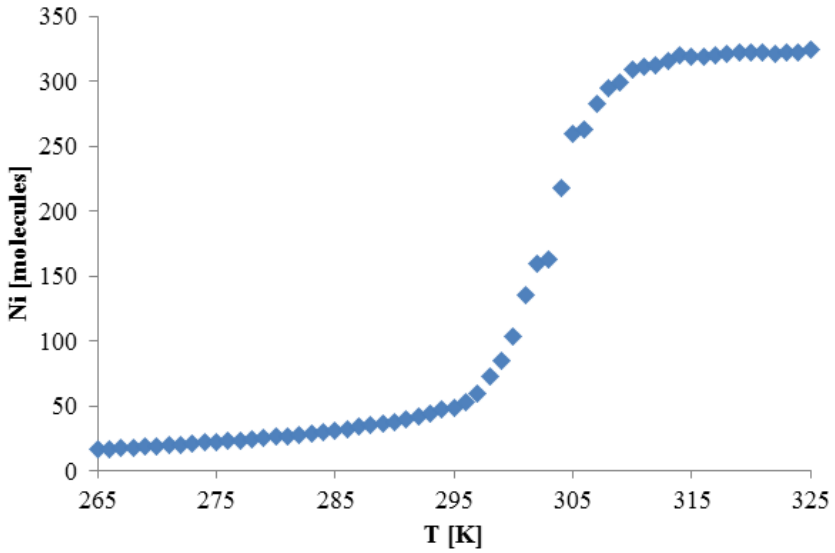

Figure 7. Average Number of CE in Isotropic State as a Function of Temperature. Number of Nearest Neighbors and Cooperativity are: z=6 and w_H=320 *cal/mol*.

Curve shown in Figure 7 represents the fractional completion of the phase transition and it reveals continuous broad transition commonly observed in small systems.

Comparing two graphs on Figure 6 one may observe that cooperative parameter in the case of four nearest neighbors is much higher than in the case of six nearest neighbors. We explain this effect in the following manner: since in the case of four nearest neighbors the number of contacts is smaller, than in order to describe the same thermodynamic behavior of the system cooperativity between these molecules has to be higher. We can apply similar principle to explain the effect of triglycerides in the system.

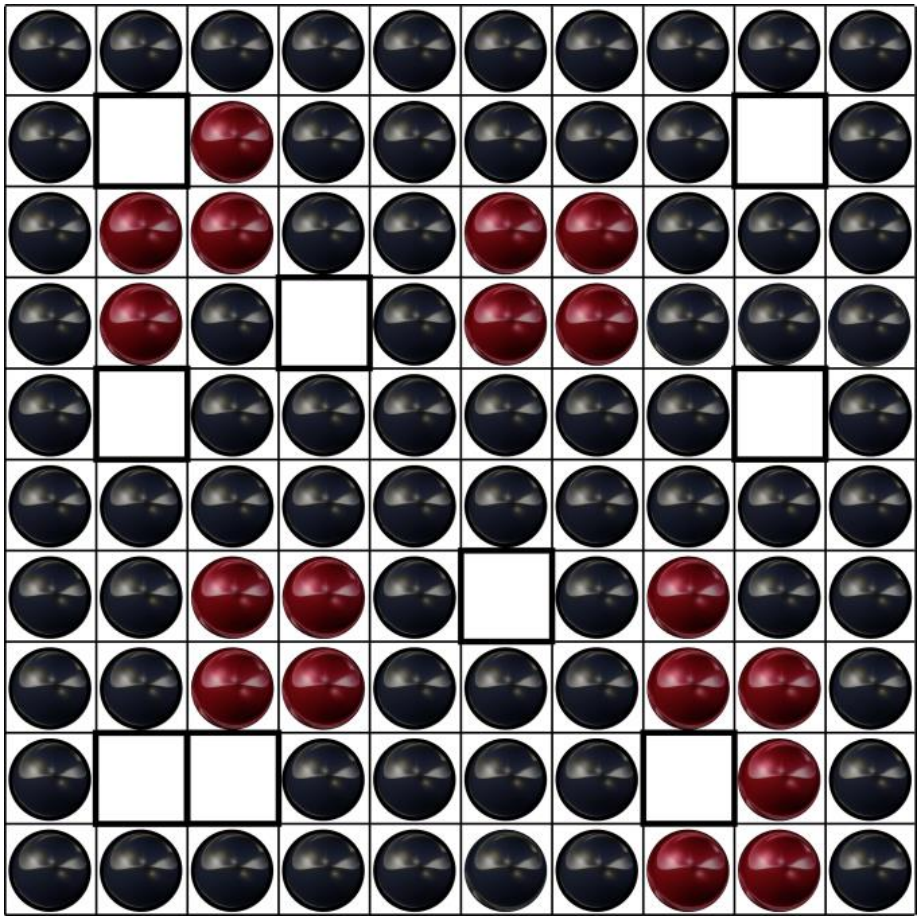

Figure 8. Incorporation of Triglycerides in the Two State System. Empty Squares Represent Cells Occupied by Triglycerides Molecules.

It is known that these species do not exhibit liquid crystal transition and if they are added to the system of CE they will result in a broad phase transition with lower peak. Also, the peak of the phase transition will be shifted to the lower temperatures. Broadening of the phase transition, lowering in the peak of the heat capacity and its displacement toward lower temperatures is related to the amount of triglycerides. In extreme cases when the triglycerides content is sufficiently high phase transition will be totally removed [38]. In other words, increased content in TGs lowers the number of CE contacts. Lower number of contacts can be expressed through

smaller cooperativity between CE molecules. Influence of increased triglycerides reflected through cooperative parameter is shown on Figure 9.

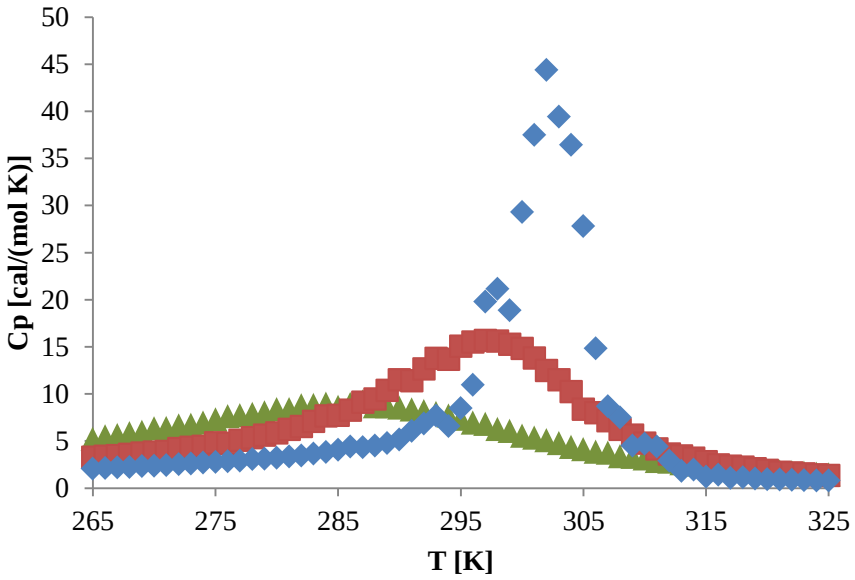

Figure 9. Influence of Increased Triglycerides Content Reflected Through Drop in Cooperativity between CE Molecules. Blue Diamonds represent Heat Capacity Curve Calculated with w_H=320 *cal/mol*, Red Squares are for w_H=290 *cal/mol*, and Green Triangles are For w_H=260 *cal/mol*.

In addition, proposed model is able to distinguish different sizes of the LDL cores. Blue dots on previous figure represent system of 400 CE molecules arranger on 20x20 lattice and black dots represent phase transition in system of 900 molecules in the layer arranged in 30x30 lattice.

Finally, we may derive conclusion that proposed model is capable to explain variety of different effects on thermodynamic behaviors of the LDL. Capability to distinguishing different particle sizes and to assess TG content in the LDL core based on thermodynamic behavior of the sample could be potentially beneficial for current estimations of LDL cholesterol. Particularly in

the case when sample contains a large number of small particles with TG-rich and CE-poor core. In these cases, traditional lipid profile may estimate values of the cholesterol contained in the LDL (LDL-C) that are 50 mg/dl lower compared with people who have large particles with normal composition. In other words, patient's CVD risk is not estimated correctly [39].

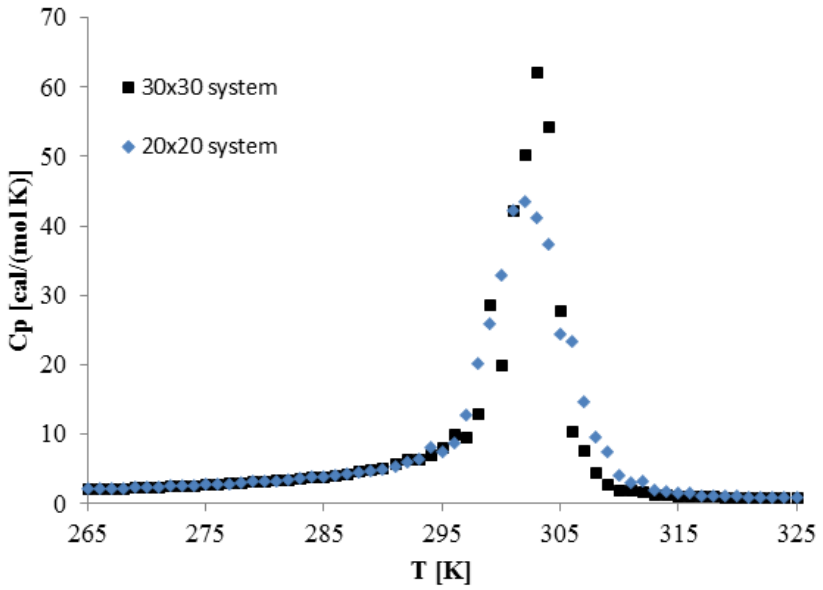

Figure 10. Heat Capacity Curves for Two State System that Differ in Size (All Other Parameters are the Same z=6, w_H=320 *cal/mol*, etc.). Black Squares represent Planar System of 30x30 Molecules and Blue Rhombuses represent System of 20x20 Molecules.

CHAPTER IV

PROTEIN MOIETY

Apolipoprotein B-100 is one of the largest known monomeric glycoproteins with molecular mass of 550 kDa. It is composed of 4536 amino-acid residues [40]. Apolipoprotein B 100 is moderately hydrophobic (3.82 kJ/residue) and due to this hydrophobicity it is insoluble in aqueous environment [41]. Thus, it is the only non-transferable apolipoprotein between lipoprotein particles and stays bounded to the LDL particle throughout its whole life. However, molecular structure of the apolipoprotein B-100 remains one of the major unanswered questions.

Currently, it is accepted that apolipoprotein B-100 can be thought of as a thick ribbon that wraps around the LDL with last 11% of the protein being represented as a "bow" that stretches back and crosses the ribbon somewhere between residues 3000-3500 [42]. Segrest et al. proposed pentapartite model of the protein based on computational analysis of secondary structure of apoB-100 [40]. The first 1000 amino acids form so called $\beta\alpha_1$ domain. This domain is followed by alternative consecutive sequences of amphipathic β strands and α helices. Thus, the final pentapartite structure can be represented as NH_3 - $\beta\alpha_1$ - β_1 - α_2 - β_2 - α_3 - COOH. According to this model, domains β_1 and β_2 are roughly 1000 and 1500 amino acid sequences long respectively, while α_2 and α_3 are roughly 500 amino acids long. The importance of the domain β_2 lays in the fact that it contains a receptor binding site. This site is named as site B and consists of 11 amino acid residues from positions 3359 to 3369: Arg-Leu-Thr-Arg-Lys-Arg-Gly-Leu-Lyc-Leu-Ala.

Structural conformation and access of this domain on the surface is of the highest importance since it influences interactions with cellular receptor and affects recognition mechanism that are responsible in LDL uptake or development of various diseases, such as atherogenesis and hypercholesterolemia. Although this model serves as a current standard, it provides little information on structural motifs [41].

Protein Primary Structure Prediction

Computational tools were user in order to evaluate structural representation of apoB-100. The first step in theoretical analysis of the protein is to determine structural domains, since apoB-100 is characterized as multi-domain protein. Domains are separate units along the polypeptide chain that form around a hydrophobic core and often have distinct function. Domains are connected by linkers to form multi-domain protein [43]. Regarding this, I analyzed the primary structure of the protein by searching against the Conserve Domains Database (CDD). CDD is highly respectful collection of well annotated multiple-sequence alignment models (MSA) for protein domains. Search against CDD v3.13 – 46918 PSSMs was performed using CD-search tool at National Center for Biotechnology Information (NCBI) using accession number P04114 available at Swiss-Prot. I have applied low complexity filter and the remaining options were left with default settings. As a result we have four significant hits:

- Vitellogenin_N with E-value 6.5 10^{-116},
- Domain of Unknown Function DUF 1943 (E-value 4.6 10^{-54}),
- Domain of Unknown Function DUF 1081 (E-value 3.22 10^{-32}),
- and ApoB-100_C (E-value 2.77 10^{-26})

The E-value stands for Expect Value and represents “a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size” [44].

From these four hits there are only two known structures. While DUF 1943 is a member of a family with no exact function, and Vitellogenin_N is well known protein family involved in lipid transport and has two structural domains. This clearly suggests that the first part of a protein has known homologue with known 3D structure.

Protein Secondary Structure Prediction

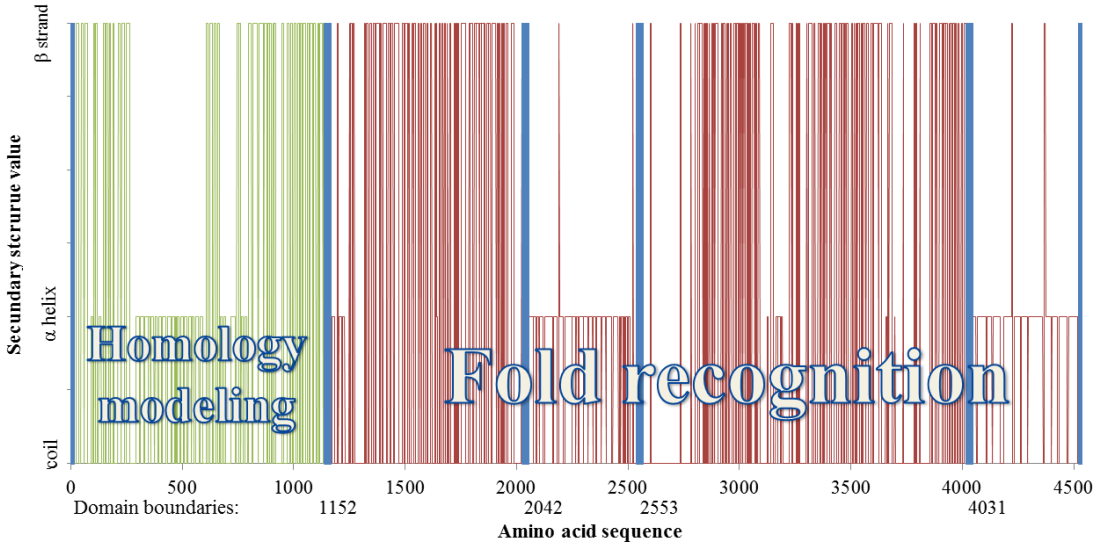

Figure 11. Results from the PSIPRED Secondary Structure Prediction; Highest Value Represents β Strands; Shorter Value Represents Helices and Coils are Represented on Horizontal Line. Blue Vertical Lines Mark Domain Borders. Green Color Indicates that Homology with Known 3D Structure may be Applied for Protein Predication Modeling.

Hydrogen bonds formed between the backbone amide and the carboxyl groups define secondary structure. In terms of computational modelling secondary structure can be helix (α), β-strand or coil. The secondary structure prediction is used as an intermediate step during protein

structure prediction when homologous proteins are not available in protein data bank [45]. I used PSIPRED [46] secondary structure prediction tool to estimate the secondary structure of the apoB-100. In this case mature protein was submitted to the PSIPRED webserver. Mature apoB-100 differs from a precursor P04114 by lacking the first 27 amino acids.

From Figure 11 domain borders were derived. The first domain has already been modelled for the first 1150 amino acids using PDB 1LSH crystal structure of lipovitellin as a template [47]. Analyzing the rest of the secondary structure we defined boundaries for the domain β_1 from 1153 to 2042, domain α_2 spreads between amino acid 2043 to 2553, domain β_{2-} will be considered between amino acids 2554 and 4031, and finally, domain α_3 will cover the rest of the sequence. Since the length of β domain is much larger than median length of proteins in eukaryotic organisms, we decided to split these regions into subdomains, again according to secondary structures. First domain of β_1 spans up to amino acid 1248 whereas the rest of the domain will be treated as a subdomain 2. In a case of domain β_2 we define 3 domains: the first one mainly contains coils and it covers protein sequence from 2554 to 2813; second domain spreads from 2814 to 3193; the third domain spreads from 3194 to 4031.

Protein Tertiary Structure Prediction

Domains previously obtained from the secondary structure analysis were treated by two protein structure prediction tools, homology modeling and tread recognition. The first domain has known homology template with 3D structure in the protein databank, and thus it was treated with *MODELLER* [48]. Homology modelling, or sometimes called comparative modelling, is based on construction of a full 3D atomic model of unknown protein, so called target, using already known 3D structure of related homologue. In this case template was recognized as lipovitellin with

1LSH entry in Protein Data Bank. Since seven out of eight disulfide bonds are present in this area, we were careful to include all of them during modelling. Modeled target is shown below.

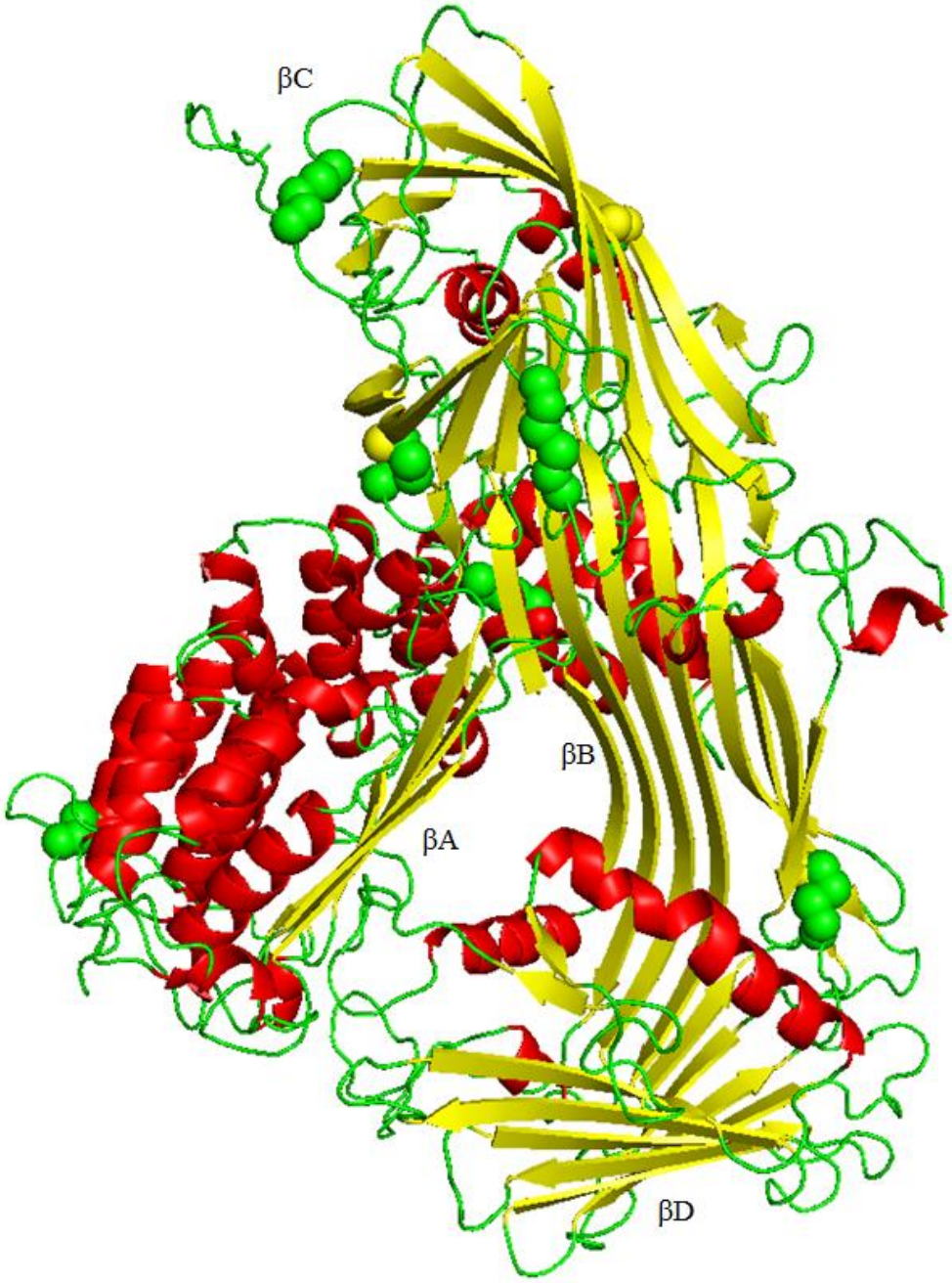

Figure 12. 3D Structure of $\beta\alpha_1$ Domain.

Figure 12 shows three-dimensional structure of $\beta\alpha_1$ domain with a "lipid pocket". "Lipid pocket" is formed from three antiparallel β sheets (shown in yellow). These β sheets are amphipathic and marked on figure as βB, βA, and βD. Double layer of antiparallel α – helices (shown in red) connect this lipid pocket with top amphipathic β sheets. In this way top β barrel, designated as βC, is connected with a lipid pocket. Green and yellow spheres represent seven disulphide bridges.

As previously seen from primary structure predication, the rest of the apoB-100 does not have known homologue in the PDB. In this case fold recognition technique (sometimes called "threading") is used. Justification for this technique is based on the fact that although two proteins do not share significant sequence identity, they can share a similar fold. Currently 1393 protein folds are known and since 2008 no new folds were identified. We decided to work with I-TASSER (Iterative Treading ASSembly Refinement) server because it was ranked as a number one server for protein structure prediction in CASP7 (Critical Assessment of Protein Structural Prediction), CAP8, CAP9, CASP10, and the latest CASP11 experiment [49]. I-TASSER procedure involves template identification by LOMETS (Local Meta-Threading-Server) followed by fragment structure re-assemble using replicate-exchange Monte-Carlo simulations. In the next step atomic structure is refined by REMO (algorithm for protein atomic structure construction) and FG-MD (Fragment Guided Molecular Dynamics Simulation). Finally, structure-based function prediction is done by COFACTOR. The estimation of model's quality is given by a confidence score (C-score) which estimates the quality of prediction. In general, if C-score is greater than -1.5 it is expected that predicted model has a correct fold.

We used I-TASSER to predict subdomain 2 of a β_1 domain, domain α_2, domain β_2, and α_3. As previously mentioned, subdomain 2 of β_1 contains sequences between 1249 and 2042 amino acids. In this case the best identified threading template was 4OM9 from PDB database. This

template is unique since a part of its chain forms β-strands in helical manner and into a trigonal prism. Figure 13 represents a model based on previously described template. Reported confidence score by I-TASSER for this model is -1.36, which means that predicted model has a correct fold.

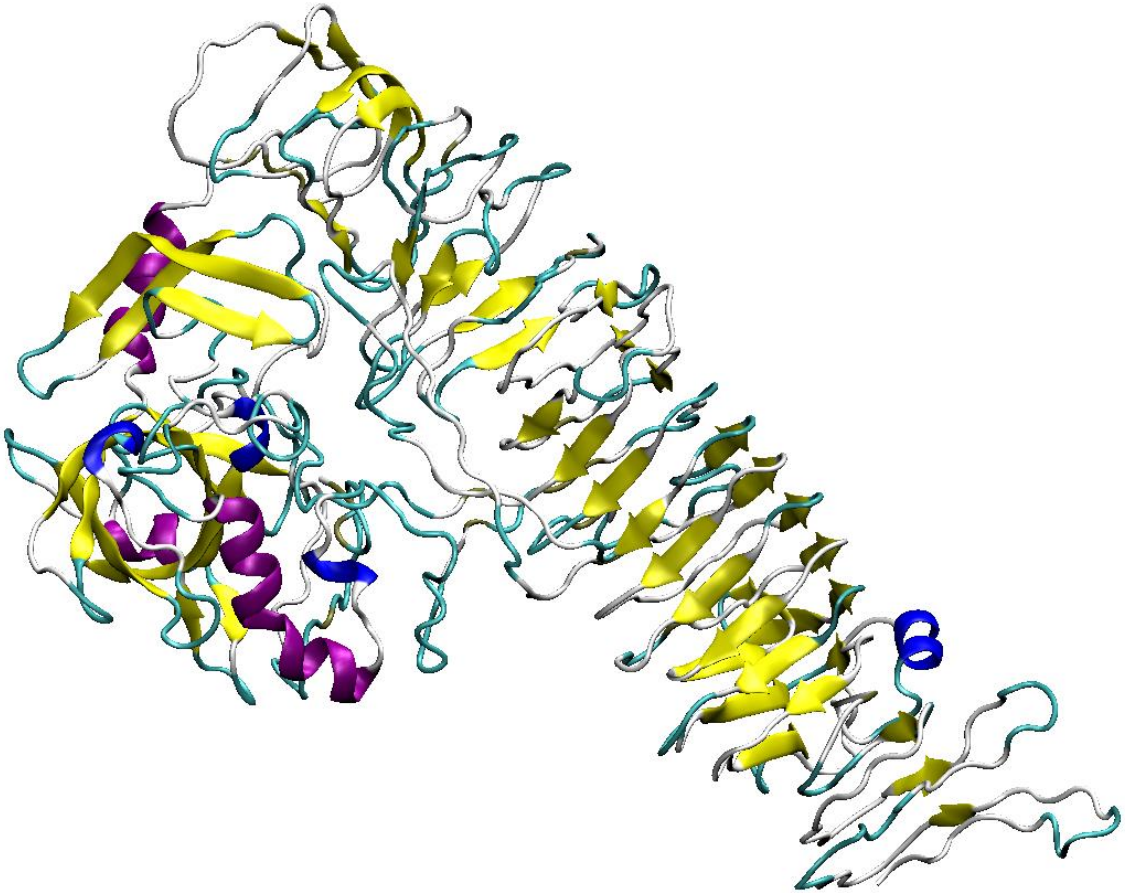

Figure 13. Subdomain 2 of Domain β_1. β structures are colored yellow; purple are α helices.

The best template recognized by the I-TASSER for domain α_2 in PDB database is 1B3U. It is a left handed superhelix with 15 repeats of HEAT motifs. Heat motif is formed by two rod-like antiparallel α helices with two turns that share a common axis. Loops that connect HEAT repeats are flexible and thus it was proposed that they may provide certain flexibility of this

domain [24]. Model based on this template reveals a double layer of 29 antiparallel α helices (Figure 14). Reported C score is -1.14, which again means that predicted model has a correct fold.

Figure 14. Domain α_2. α helices are shown in red.

Figure 15 shows that the first 260 amino acids of domain β_2, subdomain 1, are mostly coiled structures. In addition, disorder analysis of this subdomain reveals that this region is almost completely disordered. It is well known that disordered regions produce models with low C-score

and most often they are excluded from model process. However, since we know that this region is disorder, we will use the best predicted model as a guess for initial structure. The best template predicted by I-TASSER is 4NL16A with a C-score of -4.06. It is worth mentioning that recent experimental observations show that at 6℃ ApoB-100 has regions of high variance [5]. Since our analysis reveals subdomain 1 of domain β_2 as the largest disordered region in ApoB-100, one can speculate that this region is the region of high variance.

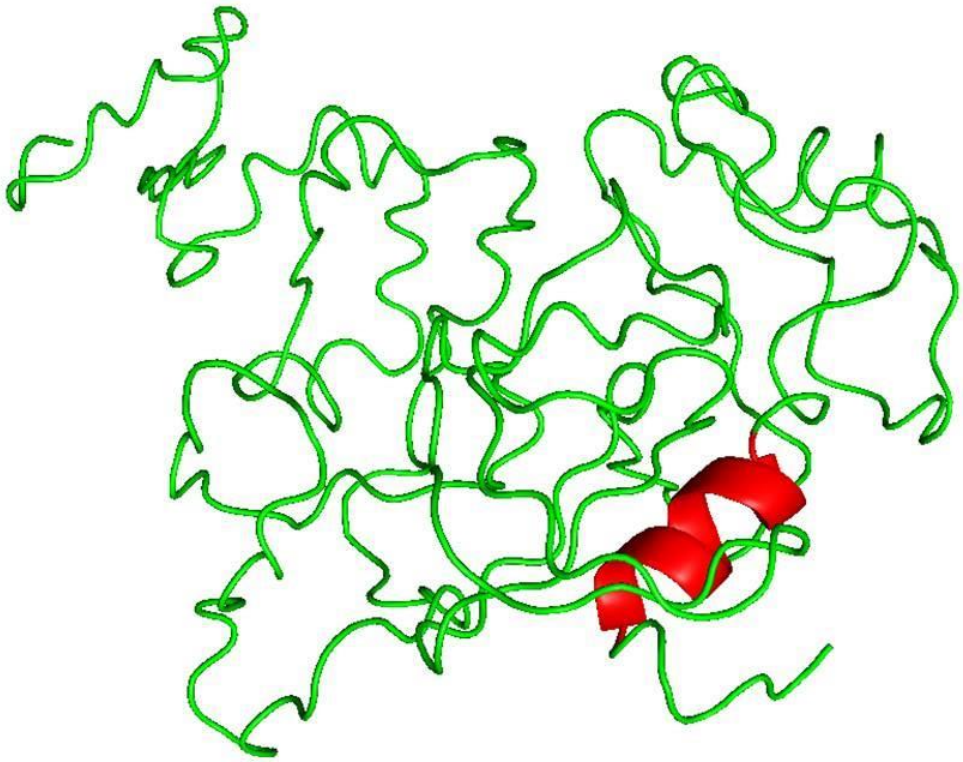

Figure 15. Subdomain 1 of Domain β_2. Highly disordered coiled region of 260 amino acids.

According to PSIPRED secondary structure analysis subdomain 2 of domain β_2 contains regions of disordered structure and thus one should expect that the confidence score of this

domain is lower. I-TASSER result identified the best template 3CM9S with reported confidence score -1.74. Together with subdomain 3 of β_2 it forms 8th disulfide bridge.

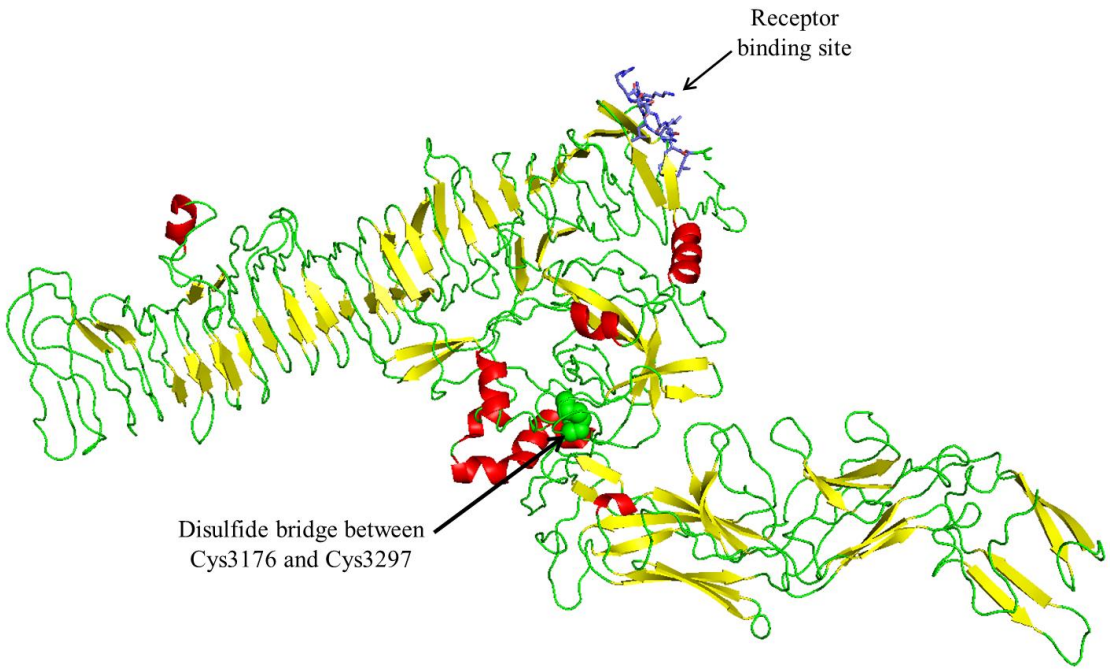

Figure 16. Subdomains 2 and 3 of β_2 Domain. We particularly note Receptor Binding Site (presented with Blue Color) and Disulfide Bridge (presented as Green Spheres).

Subdomain 3 of domain β_2 contains receptor binding site B. For this sequence span I-TASSER again reported 4OM9 with C-score of -1.04, which indicates correct fold. Figure 16 shows joined subdomains 2 and 3 of β_2. Receptor binding site shown in blue color forms a loop on the corner of the triangular prism that forms β helix topology and sticks out from the rest of the protein. Thus, this positively charged loop is positioned at the surface of domain, and is easily accessible to the receptor. We note that the receptor binding site is structurally very close to disulfide bridge.

As suggested by secondary structure analysis, domain α_3 is constructed of longer α helices. For the best template I-TASSER recognized 4UXVA with C score of -1.07. This template is made of five linear repeats of triple helical bundle. It is highly curved and almost semi-circled (Figure 17). The elongated structure of this domain was previously suggested by immunoelectron microscopy where it was proposed that this bow may prevent binding of very low density lipoprotein to the LDL receptor [42].

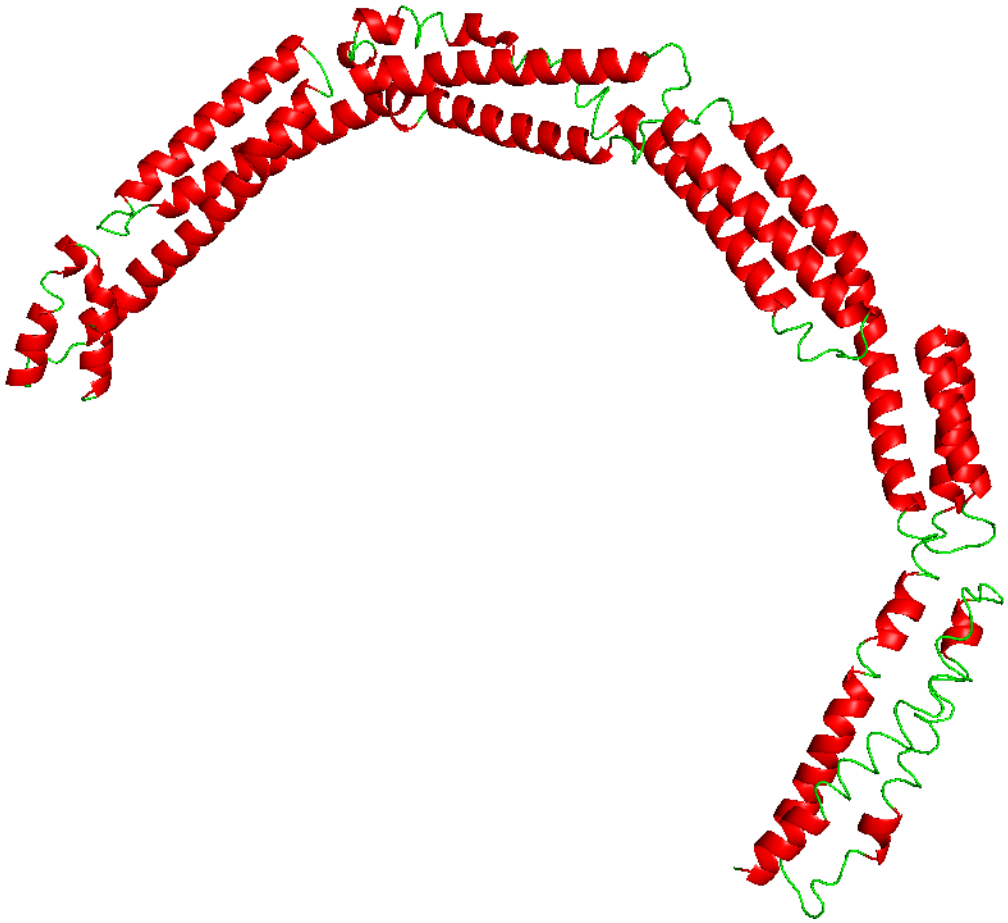

Figure 17. Domain α_3. A Helices are Depicted in Red, while Coils are shown in Green.

When there is no reliable template for construction of target model *ab-initio* protein prediction method is used. *Ab-initio* methods seek for protein structures that are favorable from thermodynamic and stereochemistry standpoint. We used QUARK *ab-initio* protein structure prediction tool in order to model subdomain 1 of domain β_1 since it was ranked as the best server for free-modeling (FM) in CASP9 and CASP10. QUARK uses a template modeling score (TM-score) in order to assess obtained results. TM-score measures topology similarity between two proteins with different tertiary structures. TM-score greater than 0.5 assume the same fold with higher score indicates a better match. The modelled region covers amino acid sequence from 1173 to 1228 position. Reported TM-score for the first model is 0.5281±0.0833. This indicates correct fold between native structure and model. Terminal regions were left un-modeled since they are mostly coils and may influence modeling process.

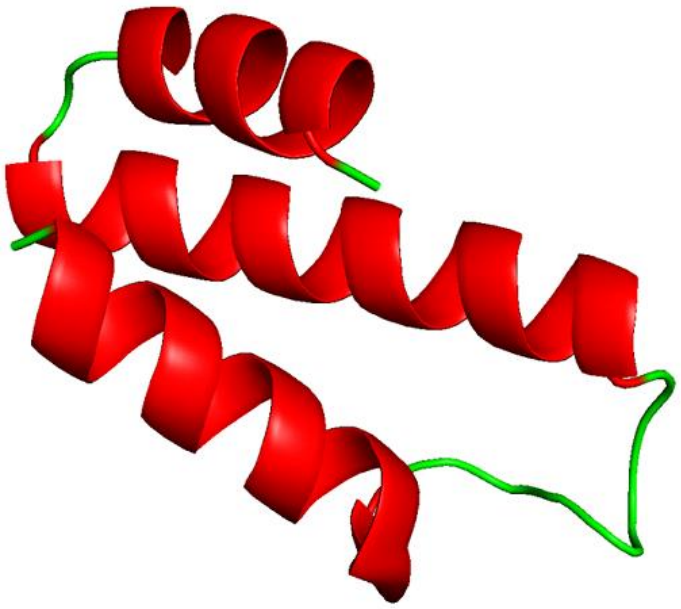

Figure 18. Subdomain 1 of Domain β_1.

Domain Assembly

The first step in assembly of domain was to place separate domains on a sphere with a diameter of 20nm according to positions reported by immunoelectron microscopy [42]. In these experiments monoclonal antibodies specific to ApoB-100 were attached on the surface of LDL and mapped.

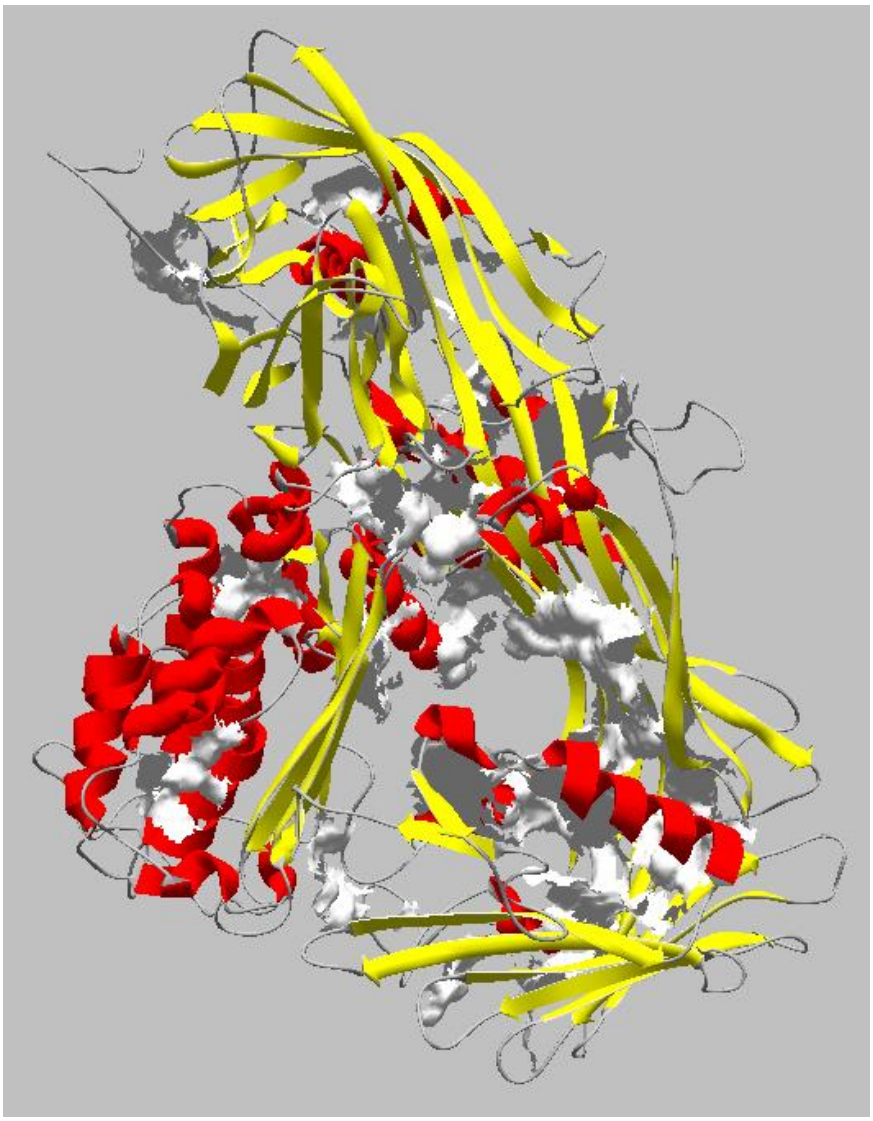

Figure 19. Hydrophobic Patches of Domain $\beta\alpha_1$.

In this way one is able to determine positions of specific regions of ApoB-100 on the intact LDL. Before placing domains in vicinity of epitopes, domains were oriented according to their hydrophobic patches. This is done because hydrophobic patches tend to face toward hydrophobic area and they are rarely exposed to hydrophilic area. We detected hydrophobic patches using Swiss-PDB Viewer 4.1.0 [50] with default values. We recognized three types of hydrophobic patches, two on opposite sides of peptide and one in the middle (equatorial) area. Then, we calculated the hydrophobic area for each side considering only the largest hydrophobic areas. Side with greater hydrophobic area was oriented toward the hydrophobic core.

Figure 19 shows that the greatest number of hydrophobic patches is in the lipid pocket. Thus, in our analysis, it was orientated toward the core of the LDL. A bundle of α helices was oriented toward the aqueous environment.

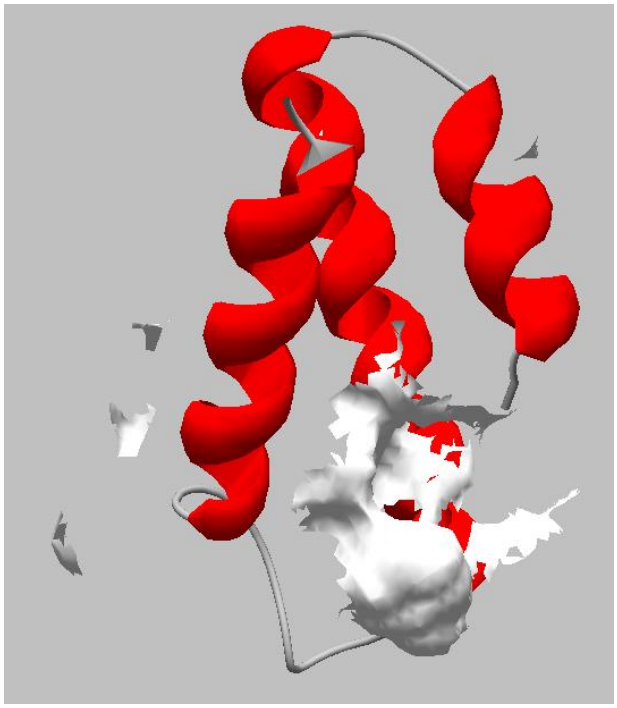

Figure 20. Hydrophobic Patches of Subdomain 1 of Domain β_1.

Subdomain 1 of domain β_1 is made of three α helices that have only one significant hydrophobic patch and, accordingly, this domain was oriented so that the hydrophobic patch faces the neutral LDL core. These hydrophobic patches are shown on Figure 20.

Figure 21 shows hydrophobic surfaces of subdomain 2 of domain β_1 that are oriented toward the core of lipid droplet. This area has almost twice as many of hydrophobic patches than the opposite side. In addition, lateral side also contributes with a significant number of hydrophobic patches. We suggest that this domain is immersed in a phospholipid monolayer due to such high number of hydrophobic patches.

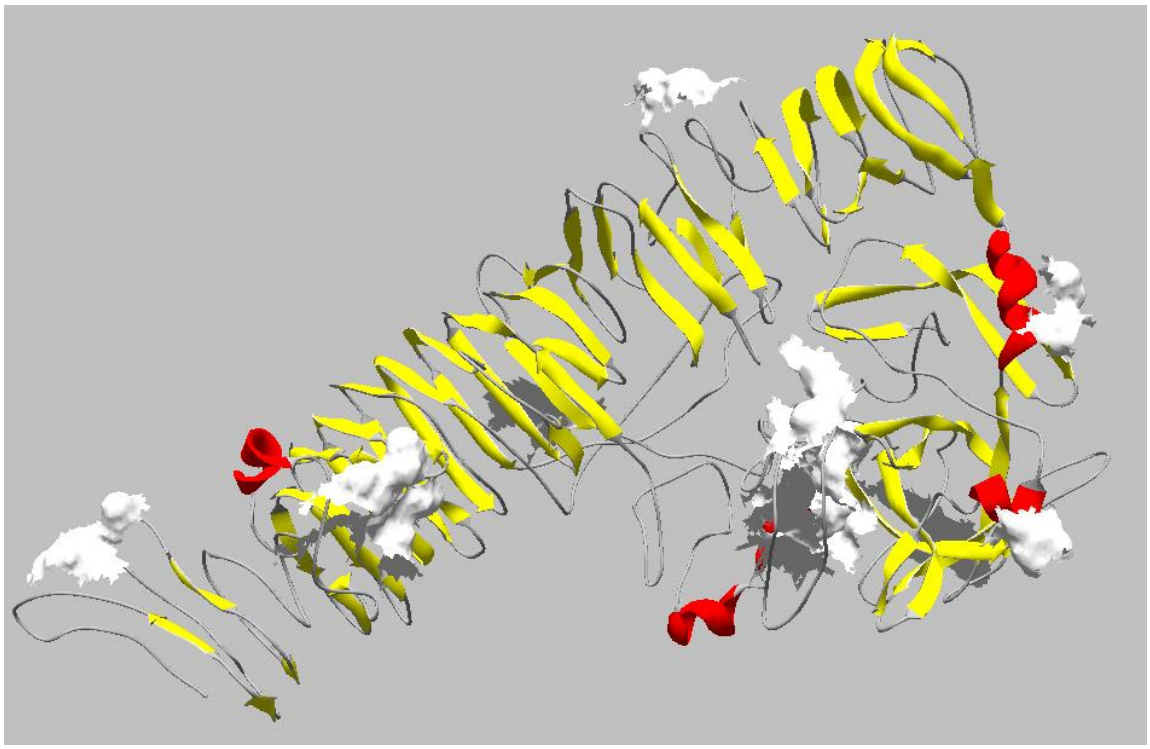

Figure 21. Hydrophobic Patches of Domain β_1 – Looking From a Core of LDL Particle.

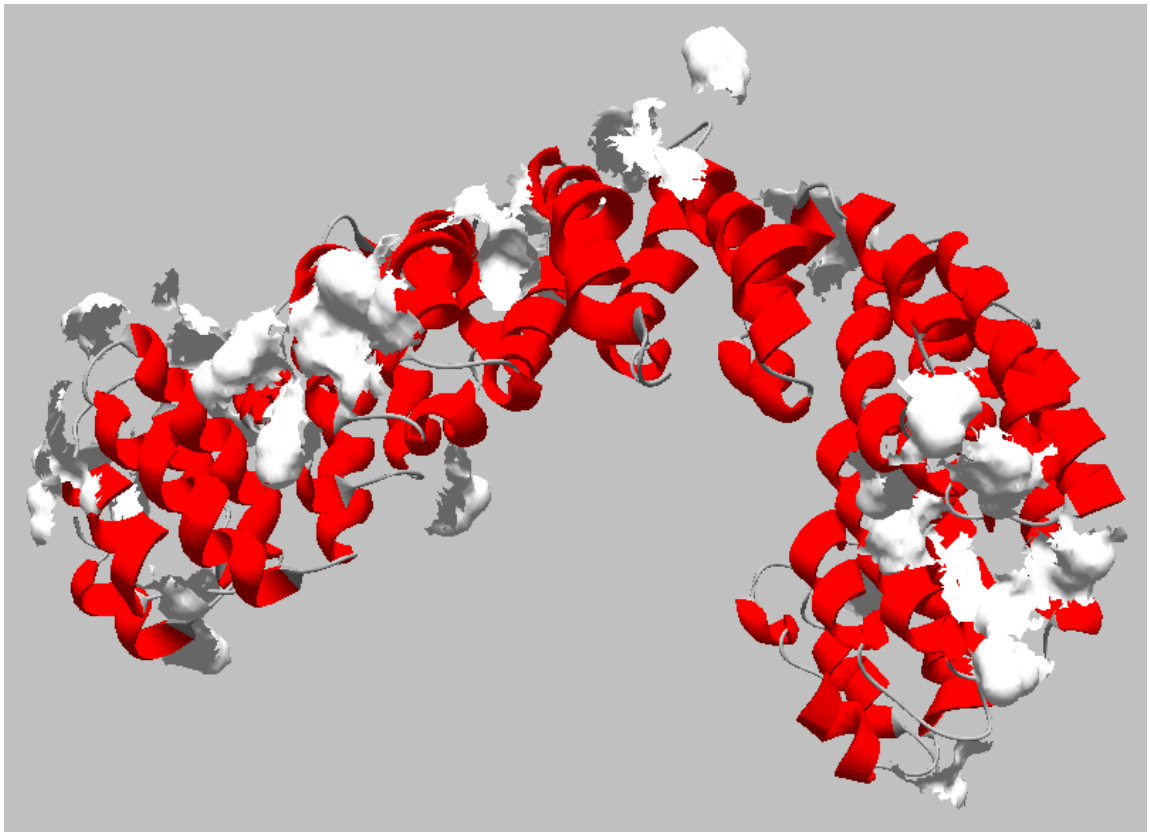

Figure 22. Major Hydrophobic Patches of Domain α_2.

Previous figure shows domain α_2 from the side that has higher number of hydrophobic patches. This area of double α helices has almost 60% greater hydrophobic surface then the opposite side. Interestingly, area in the middle section has the lowest hydrophobic surface.

The first part of domain β_2 consists of highly disordered coils. In hydrophobic analysis this domain is almost entirely surrounded by hydrophobic patches. Thus, we conclude that this area is most probably immersed in phospholipid monolayer. Hydrophobic patches of this domain are shown on figure below.

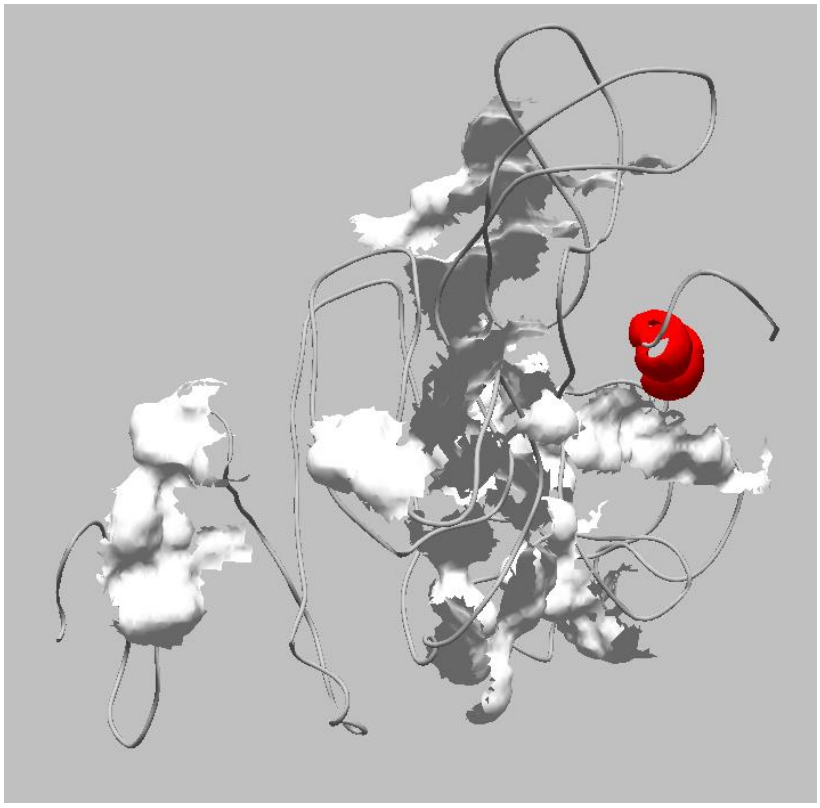

Figure 23. Hydrophobic Patches of Subdomain 1 of Domain β_2.

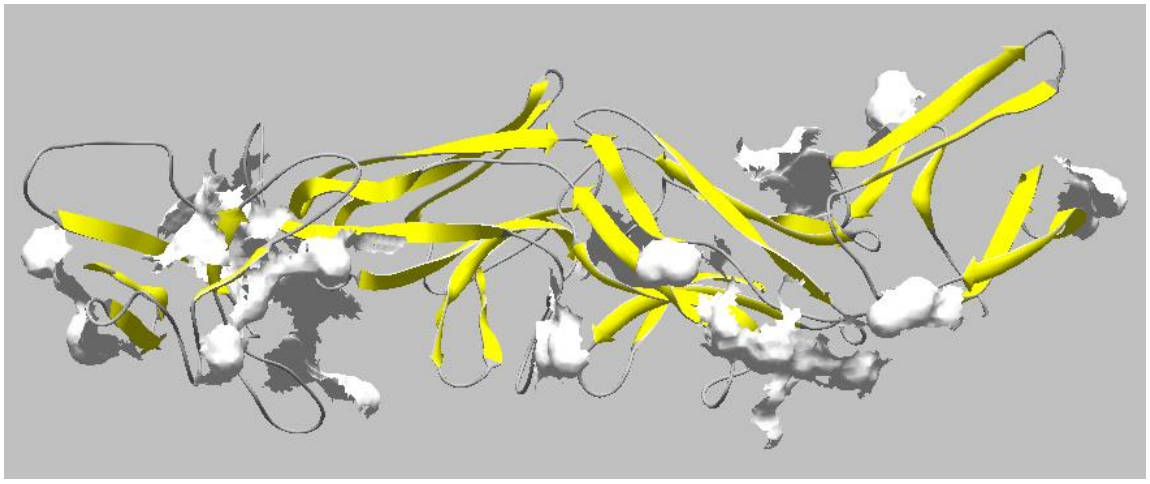

Figure 24. Side View of Subdomain 2 of Domain β_2.

Hydrophobic patches of subdomain 2 of domain β_2 are shown on Figure 24. This subdomain has high hydrophobic area in the middle portion of peptide. Upper and lower faces do not contribute too much to hydrophobic effect. Since this part of the protein is influenced by the disulfide bridge, we oriented it in such way that it can form disulfide bond with the following subdomain.

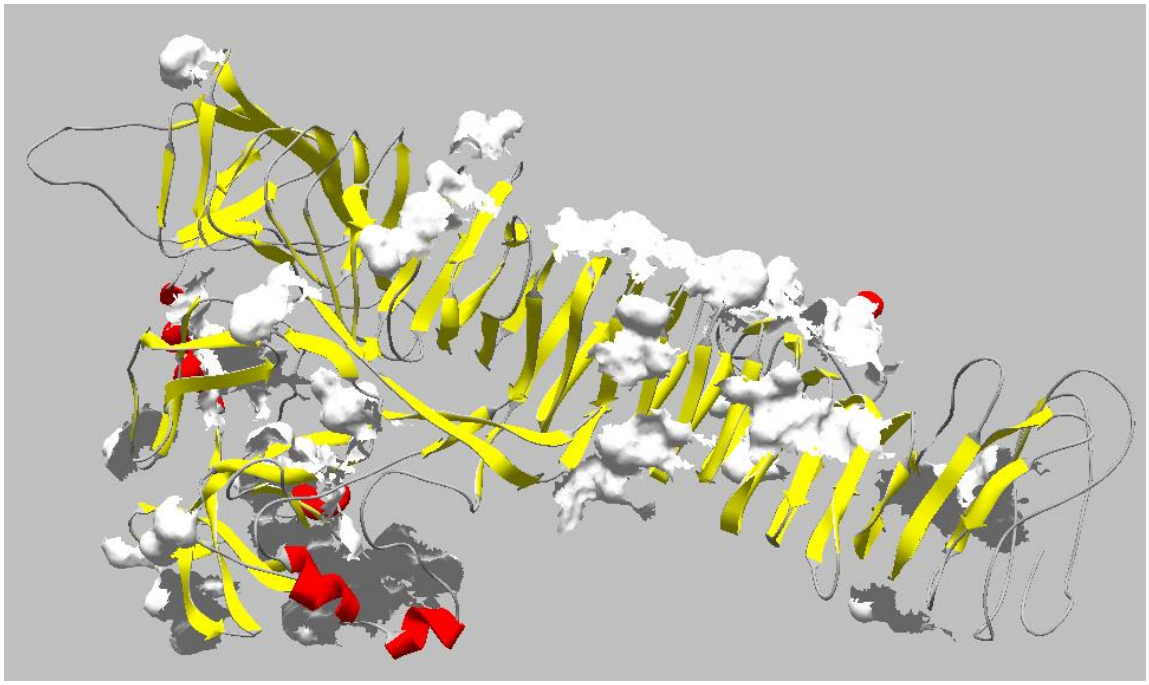

Figure 25. Hydrophobic Patches of Subdomain 3 of Domain β_2.

In subdomain 3 of domain β_2 side with receptor binding site has almost 50% lower content of hydrophobic patches that the opposite side. Most of these patches are in the area of disulfide bond. It is worth mentioning that the middle region has as much hydrophobic patches as the face that is oriented toward the center of the LDL. This allows us to guess that this area will be immersed into phospholipid monolayer and that it will most probably interact with core lipids.

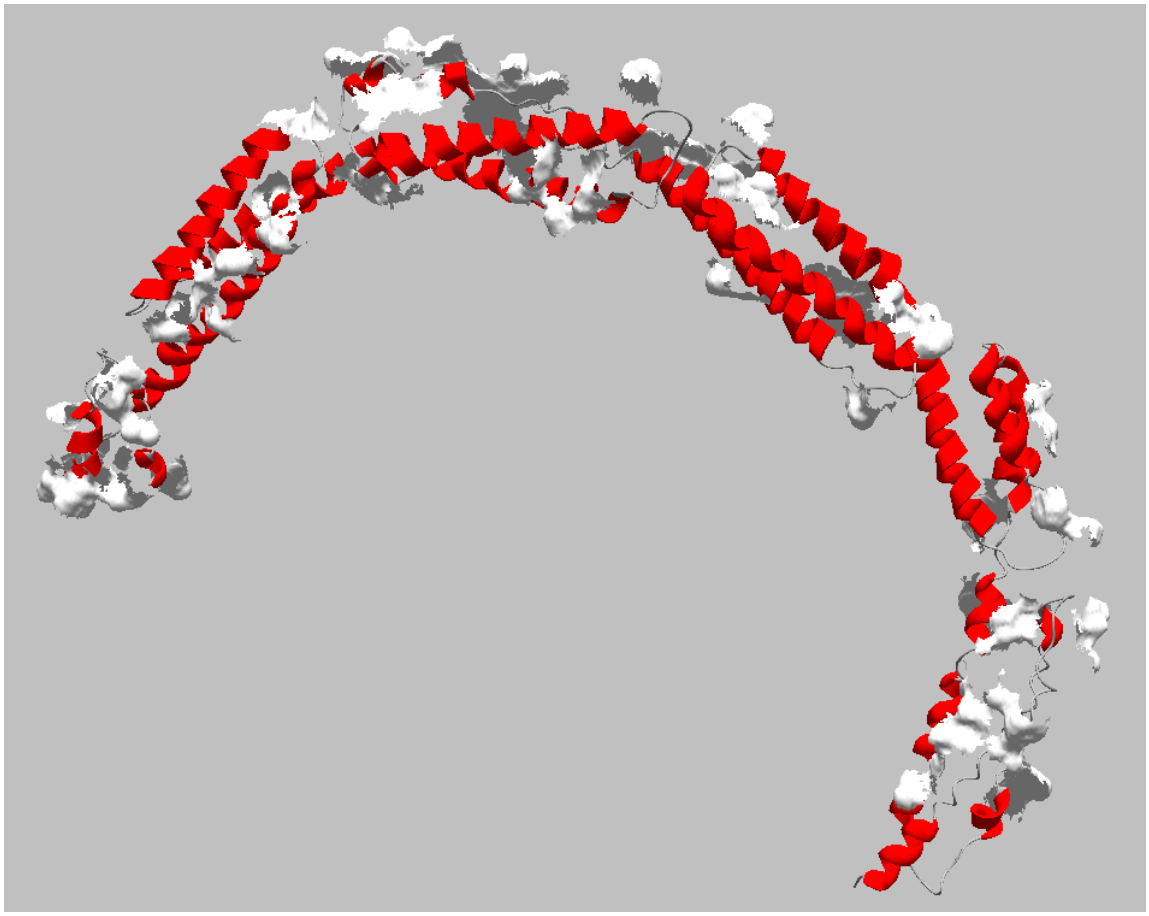

Figure 26. Hydrophobic Patches of Domain α_3.

Domain α_3 has almost equal hydrophobic area for both sides and twice as low hydrophobic patches in the middle region. The side that we oriented toward phospholipid monolayer has only 2% greater hydrophobic area than the other side. Since this domain constitutes so called “bow” that crosses β_2 domain, we speculate that the middle area of α_3 interacts with β_2 domain. Hydrophobic patches of this domain are shown on Figure 26.

Once hydrophobic profile of individual domain was known we were able to properly orient all peptides on the surface of the lipid droplet. After proper placement, domains were connected by submitted their coordinates to AIDA (*Ab Initio* Domain Assembly Server) server [43].

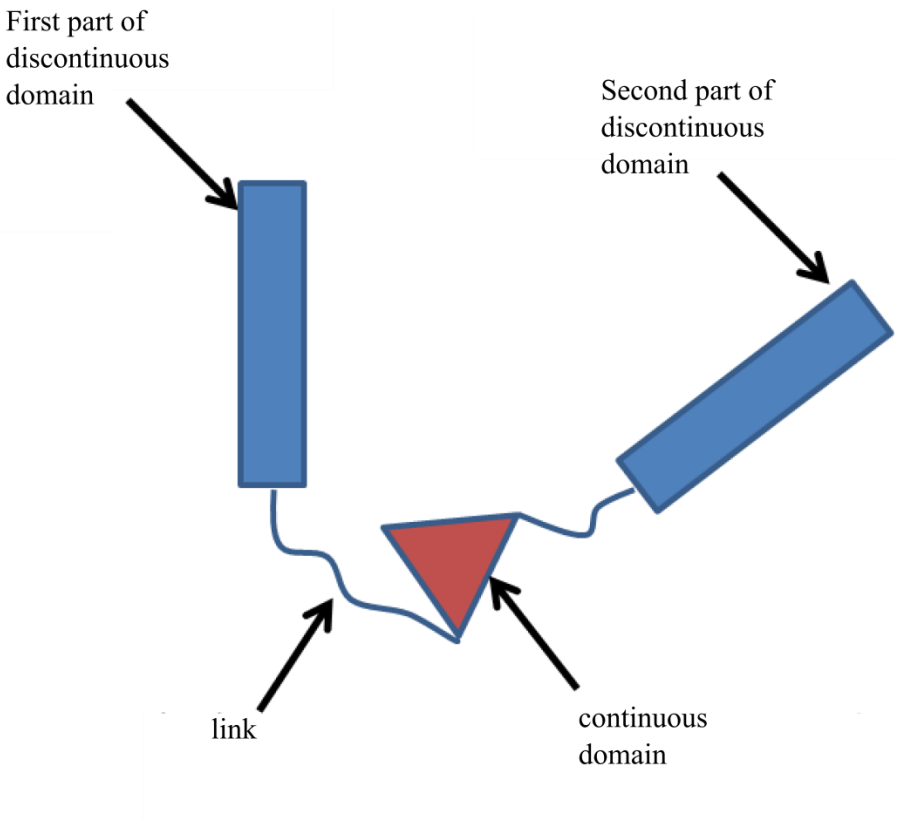

Figure 27. AIDA - Assembly of Discontinuous Domains Using Continuous Domain.

AIDA tries to find the best domain orientation while preserving connections between chains. During these processes domains are kept rigid and linkers between domains are left flexible. In this way it is possible to submit linkers to energy minimization while maintain the structure of domains unchanged. To assemble ApoB-100 domains we used assembly of discontinuous domains using continuous domains. This process is shown on Figure27. This is very useful tool since it allows connection between domains with predefined positions. Relative positions of the first and the second part of "discontinuous" domain are fixed during assembly and orientation of continuous domain, which is inserted in the middle of the discontinuous domain, is adjustable. In the final step, continuous domain is connected to the discontinuous domain via linkers.

During linking process AIDA program uses two residues at each end of domain as a linker. In some cases linkers are longer and most often their structure is not as accurate. In such cases it is advisable to remove those areas before submitting a structure file to AIDA. Program will treat these areas as linking region. Disorder analysis shows that almost at all domain boundaries amino acids are disordered and in coiled state. Thus, we decided to remove them from domain structure files and allow AIDA to remodel them.

Previously analyzed domains now have their 3D structure and known orientation on the LDL particle. According to AIDA assembly procedure, we need short peptide domains in between. Modeling of these domains was performed by submitting their sequence to QUARK *Ab initio* server. Domain $\beta\alpha_1$ and β_1 are connected with a short domain modelled from amino acid 1153 to 1172. Subdomain 2 of domain β_1 and domain α_2 were connected with short domain modelled between 2023 and 2042 amino acids. Domains α_2 and $\beta 2$ are connected with short domain modelled between 2554 and 2573 amino acid. Domains β_2 and α_3 are connected with short domain modelled between amino acids 4024 and 4043.

Assembled structure of apoB-100 is shown on Figure 28. It clearly indicates thick “ribbon” that spans until the last domain and the last domain is represented as a “bow” that crosses the ribbon.

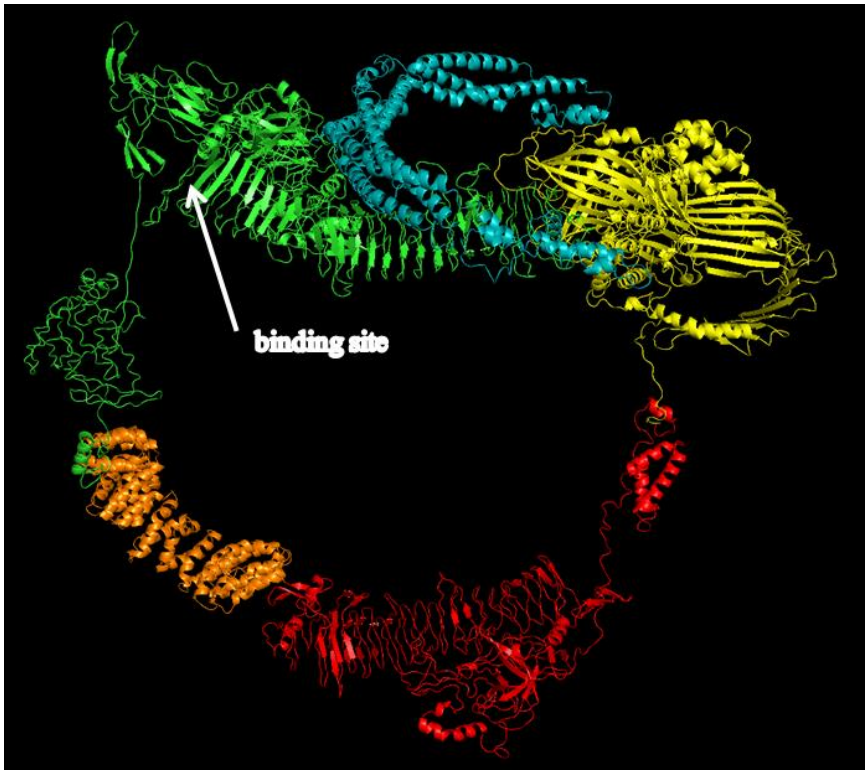

Figure 28. Assembled Structure of apoB-100. Yellow: $\beta\alpha_1$; Red: β_1;Orange: α_2; Green: β_2; blue: α_3.

CHAPTER V

DYNAMICS AND ORDERING IN 3D LDL PARTICLES

We conducted molecular dynamics investigation of LDL in order to evaluate dynamical properties of the LDL.

Force field (FF) methods are based on Born-Oppenheimer approximation where motion of light and fast electrons can be separated from the motion of heavy nuclei. This allows calculation of energy as a function of nuclear position only.

FF has both functional form and parametric form. An example of force field is shown below:

$$V(\vec{r}^{N}) = \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsion} \frac{V_n}{2}\left(1+\cos(n\omega - \gamma)\right) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left\{4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right\} \quad (6)$$

From this expression we see that potential energy $V(\vec{r}^{N})$ is a function of positions of all particles. In equation (6) potential energy consists of four terms. First three terms describe bonded interactions and the last term describes non-bonded interactions. Bond stretching is the first term. It is modeled as a harmonic potential between pairs of bonded atom, and summed over all bonds. Contribution to the system energy is obtained when the bond length l_i deviates from the equilibrium value $l_{i,0}$. Second summation describes energy penalties due to angle bending. Again, harmonic potential describes increase in system energy due to deviation of an angle between three atoms θ_i from its reference value $\theta_{i,0}$.

Third term describes bond rotation. In this case energy change due to bond rotation is modeled by torsion potential which is essentially a cosine series expansion. Here, V_n is a description of barrier height, n is the number of minimum points in the function as the bond is rotated through 360°, ω is torsion angle and γ is the phase. Non bonded interactions are all interactions between (a) two atoms belonging to different molecules, or (b) two atoms in the same molecule but separated by at least three bonds. A Lennard-Jones potential describes van der Waals interactions and Coulomb potential is responsible for electrostatic interactions.

Force fields are parameterized by fitting experimental and QM calculations to give appropriate results. As such, force fields can sometimes provide answers as accurate as highest-level QM calculations. Force fields are not parameterized to give just results for one individual molecule but instead to be applicable on a series of related molecules.

Once initial positions, velocities and the potential are known, forces acting on all particles can be determined as negative gradient of the potential. Thus, force on *i-th* particle is $\boldsymbol{F}_i = -\nabla V(\boldsymbol{r}^N)$. After all forces are known, one may proceed with integration of Newton's equation of motion $\boldsymbol{F}_i = m_i (d^2\boldsymbol{r}_i(t)/dt)^2$. This will give new positions and velocities. Previous two steps are repeated for the desired length of time and finally average quantities are computed. For molecular dynamics calculations we used "GROMACS" [51] since it is high performance software and well established program for treatment of biomolecules in sciences.

Lipid Droplet

LDL is the only macro molecule in our cardio-vascular system that supports phase transition in vicinity of body temperature. X-rays scattering differential scanning calorimetry (DSC) and ^{13}C NMR experiments show that this transition is related to cooperative melting of CE molecules.

These experimental results require region in the core of the LDL that is rich in CE molecules and capable of supporting smectic ordering at low temperatures.

We decided to create CE core based on the principles suggested by Ginsburg et al. [28]. This model is derived from the monolayer type 2 crystal structure of CE and assumes that CE molecules are placed in an antiparallel arrangement and that they contribute their chains to adjacent layers. Coordinates for a cholesteryl oleate in a smectic phase will be taken from known crystalline structure of this molecule [25]. In total, 1650 CO molecules will be treated in our lipid droplet. Free cholesterol will be constructed from the CO molecule by subtracting aliphatic hydrocarbon tail. In total, 600 free cholesterol molecules will be present in the system. In addition, lipid droplet will have 170 glyceryl trioleate, 635 molecules of palmitoyl oleoyl phosphatidylcoline (POPC), and 80 molecules of palmitoyl phosphatidylcholine (PPC).

CE form monoclinic unit cell that belongs to space group $P2_1$ with unit cell dimensions 12.356(2), b=8.980(3) and c = 18.382(2) Å, β = 85.49(2)° at 123K with two molecules in the unit cell. Atomic position parameters for this crystal are given in fractional coordinates with respect to crystal axes. It is possible to convert fractional coordinates to Cartesian coordinates using transformation matrix:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & c\dfrac{\cos\alpha - \cos\beta\cos\gamma}{\sin\gamma} \\ 0 & 0 & c\dfrac{v}{\sin\gamma} \end{bmatrix} \begin{bmatrix} \vec{\hat{a}} \\ \vec{\hat{b}} \\ \vec{\hat{c}} \end{bmatrix} \tag{7}$$

Here v is the volume of a cell:

$$v = \sqrt{1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma} \tag{8}$$

Once we obtained the unit cell (see Figure 1), we took coordinates of one molecule and applied Martini force field representation on it [52]. This force filed is parametrized for biomolecules such as lipids and proteins, including sterols. It is a coarse grained force field where four atoms are mapped to one coarse grain bead and in case of ring particles (for example cholesterol) two or three atoms are mapped to one course grain bead. The use of coarse grained model is very useful in case of molecules such as CE because they have low diffusion and thus they need large timescales to probe the system. These timescales are not feasible with current all-atom models. For example, integration time steps in case of Martini force field is between 5 fs and 40 fs. Moreover, comparison of diffusion constants with all-atom models showed that effective time sampling is 2- to 10-fold larger. For example, aggregation of lipids into vesicles occurs with 4-fold increase in time scale. We will adopt this value and report on all graphs and in the text only effective time since it is comparable with atomistic simulations. The force field was carefully parameterized for temperature range from 270 to 330° K, as the majority of biologically important and interesting phenomena occur at these temperatures. The topologies of the molecules were already present and available from Martini force fields download web page [53]. Strictly following these procedures we were able to build all molecules in our system. Here we present mapping of the most difficult molecule in our set of molecules. Inside of the red circles (Figure 29) are atoms that form a particular coarse grain bead. Center of mass for each bead was calculated according to:

$$\vec{R} = \frac{1}{M}\sum_{i=1}^{n} m_i \vec{r}_i \tag{9}$$

Here, $\boldsymbol{R}$ is a position of bead center of mass, M is the total mass of the atoms that form a bead, m_i is the mass of atom that forms a bead and $\boldsymbol{r_i}$ is its position. Each bead has its name and type that is used later for connection with parameters stored in a force filed.

It's well known that current methodologies to obtain thermodynamic properties, such as Monte Carlo simulations and molecular dynamics, investigate only a small portion of a phase space. On the other hand, equilibration process may last more than computational time available. In our case we are working with lipid molecules and esters that diffuse very slowly, so it is vital to begin with a state as close as possible to the equilibrium. For example, the phase transition of core lipids happens in a time scale of milliseconds, which is out of reach of current methodologies [54]. With this goal we first generate the single cell consistent with experimental findings and then multiply it in order to get the closest state.

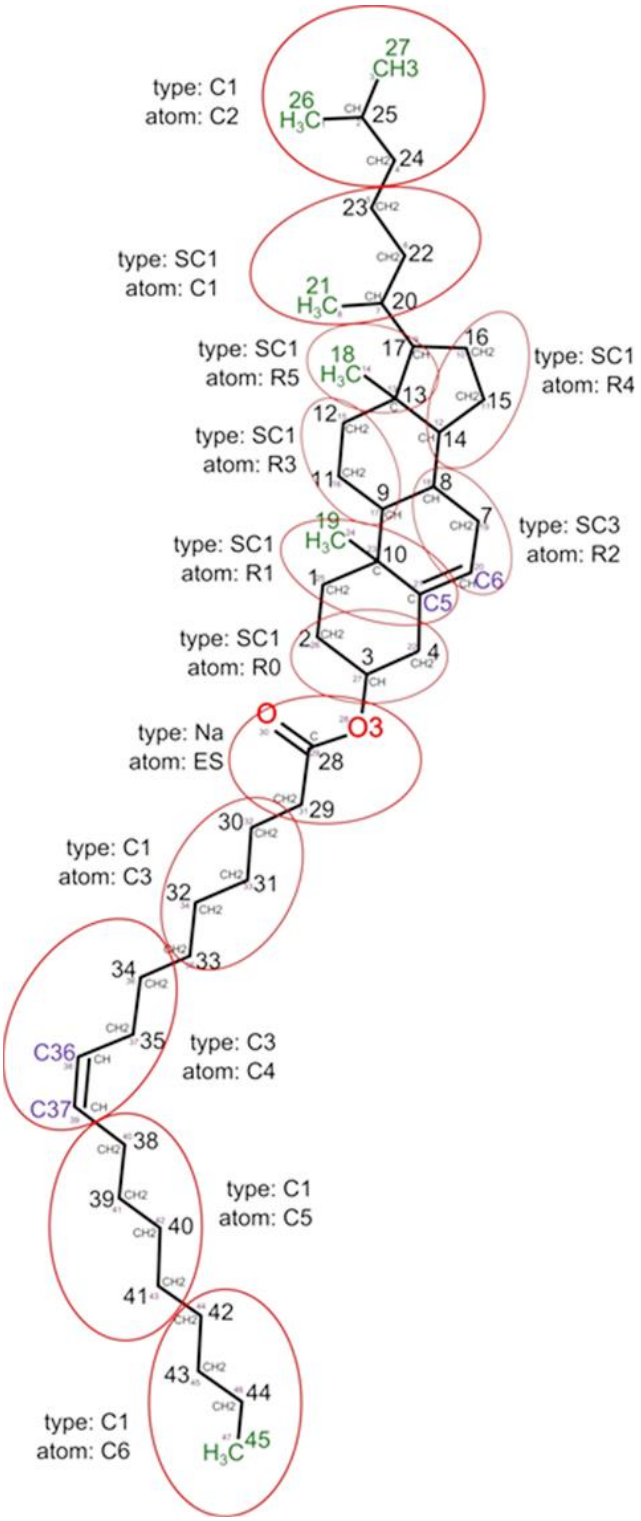

Figure 29. Mapping CO to Martini Representation.

Once we had Martini force field representation of the CO we started to build our Ginsburg's model for lamellar phase of CO. First we aligned CO molecule with principle axes and then we copied and rotated it for 180° using GROMACS package 4.5.4[51]. At this point we had two antiparallel molecules with aliphatic chains oriented toward adjacent sides.

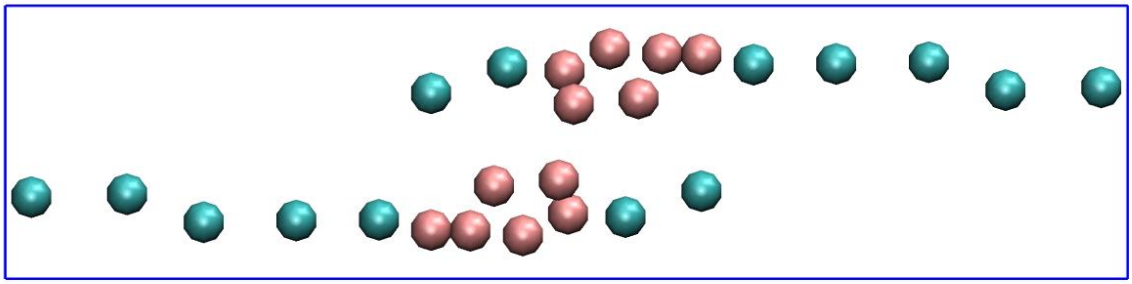

Figure 30. Building Block of the CO Layers.

Then we concatenated these files and placed everything in a box of dimensions 6.028, 1.47 and 0.578 nm . This box was then multiplied 3 times in X direction, 11 times in Y direction and 25 times in Z direction, which gave us the initial number of CO molecules of 1650. In the final step layers were indented for -2.138 nm in order to form correct distance. Figure 31 shows initial configuration of CO. System was then minimized in order to avoid any inappropriate geometries and steric clashes.

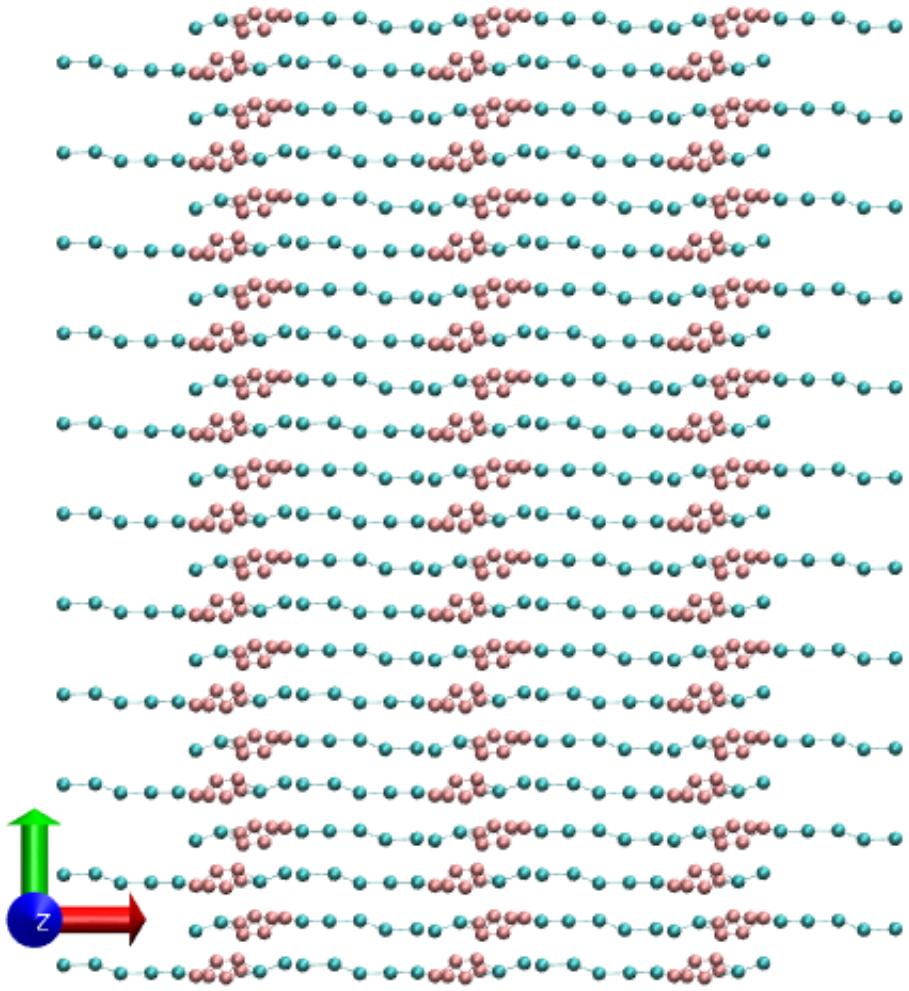

Figure 31. Replication of Building Blocks.

Steepest descends minimization procedure converged to machine precision in 1192 steps. The reported potential was -1.63316926771337e+05 kJ/mol and maximum force was found on atom 18530 as 9.60253403417682e+01 kJ/(mol nm). The reported values are very important since they tell us that our system reached minimum energy and that maximum force is lower than 1000kJ/mol nm, which tell us that the system is in a stable state and ready for further simulation.

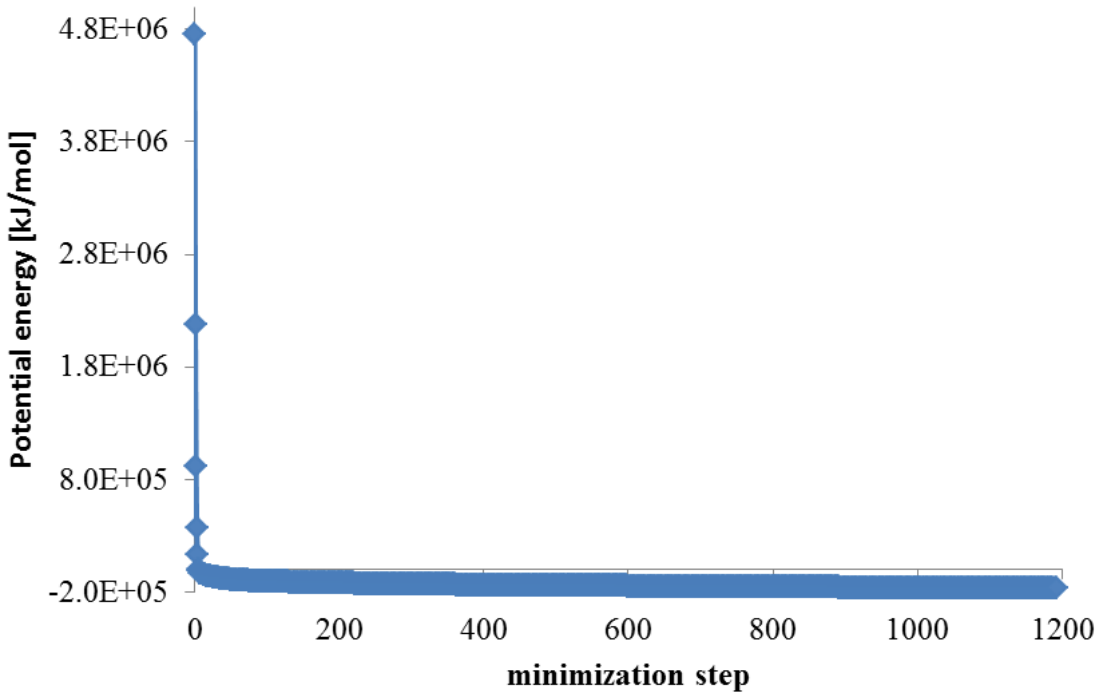

Figure 32. CO Potential Energy.

Once we had reasonable structure in terms of geometry, we equilibrated CO at temperature 279K and pressure 1bar by coupling it to Berendsen thermostat and barostat. Although we now that Berendsen thermostat does not represent exact thermodynamic ensemble, this coupling scheme is commonly used for equilibration as it is the most efficient way to scale simulation box and control the temperature. We decided to work at this temperature since CO molecules are in the smectic phase in both LDL particles and pure samples at these conditions. Obtained equilibrated temperature was 278.62K, pressure 1.017 bar and density 1069.11 kg/m^3 measured during last 2.4 ns and after stabilization of thermodynamic properties.

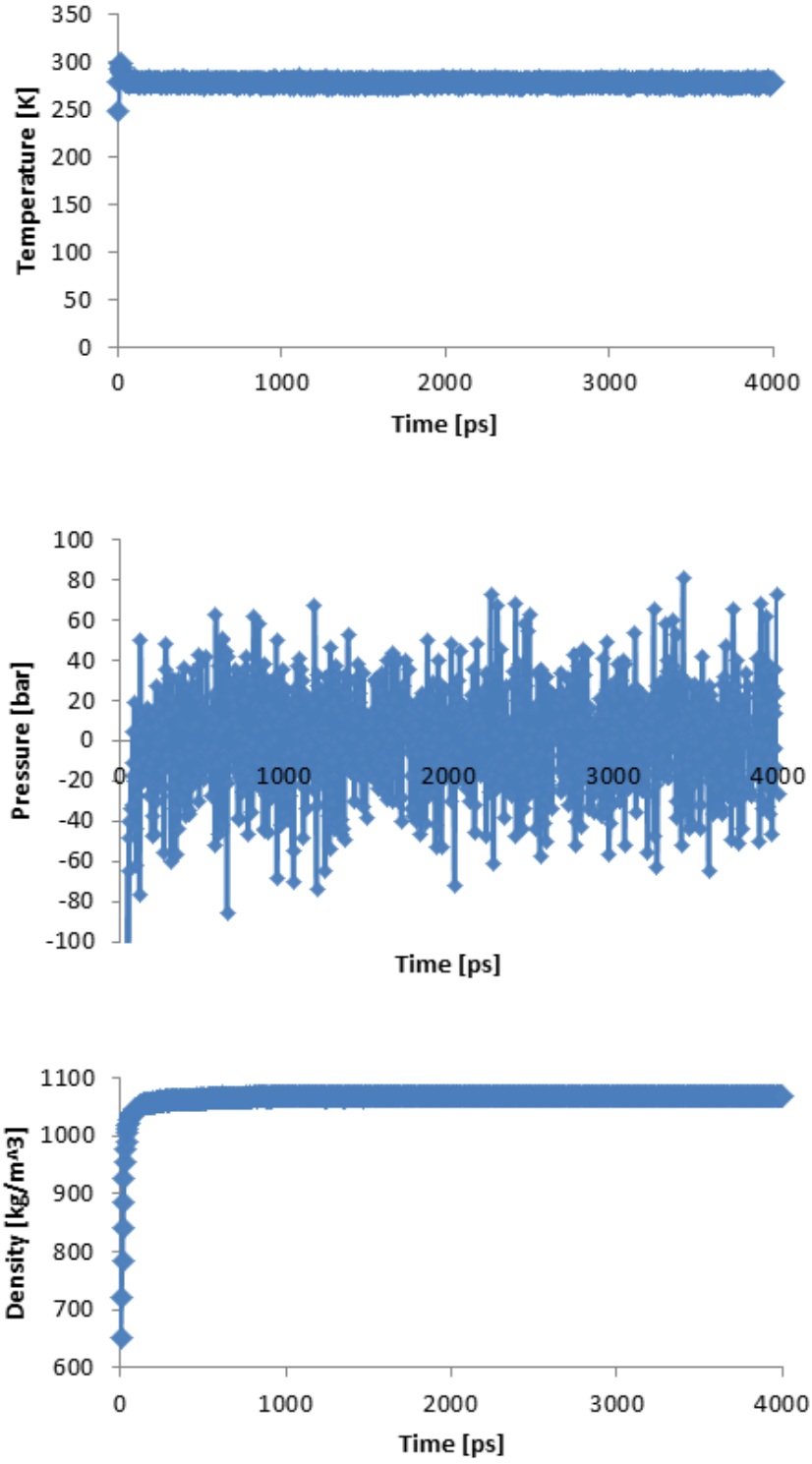

Figure 33. Temperature, Pressure and Density of Equilibrated CO System.

Formation of Lipid Droplet

Lipid droplet (LD) was built starting from cholesterol esters in the layered phase by adding other molecules to the simulation box. We added 1070 molecules of TO, 635 of POPC, 80 molecules of PPC, and 600 molecules of CHOL. Immediately after that we performed minimization to avoid steric clashes.

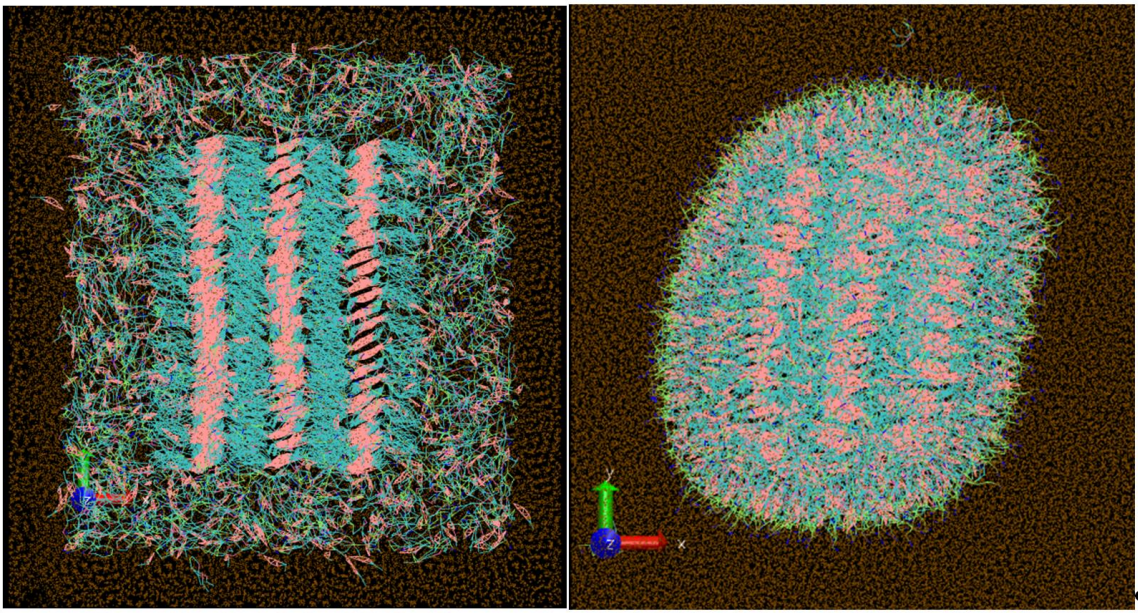

Figure 34. Formation of LD. Initial random placement of molecules in the simulation box and self-assembly of LD after 200ns of equilibration.

Once minimized, system of lipids was then solvated with water. File "water.gro" was downloaded from Martini force filed website. Here we have six CG water molecules per lipid and this is equivalent to 24 all atom water molecules per lipid. Once again, we performed energy minimization.

Before we start with equilibration process, it is important to couple lipids together in one temperature coupling group and separate them from aqueous solvent. Thus, in the next phase

lipids were coupled together. Once again, Berendsen thermostat and barostat were used to equilibrate the system at 295°K and pressure of 1 bar for 200ns. Randomly added lipids self-assembled around CO core in a form of a monolayer shell. As a result of equilibration, obtained average temperature of the LD was 294.9478 K, and average pressure was 0.95984 bar.

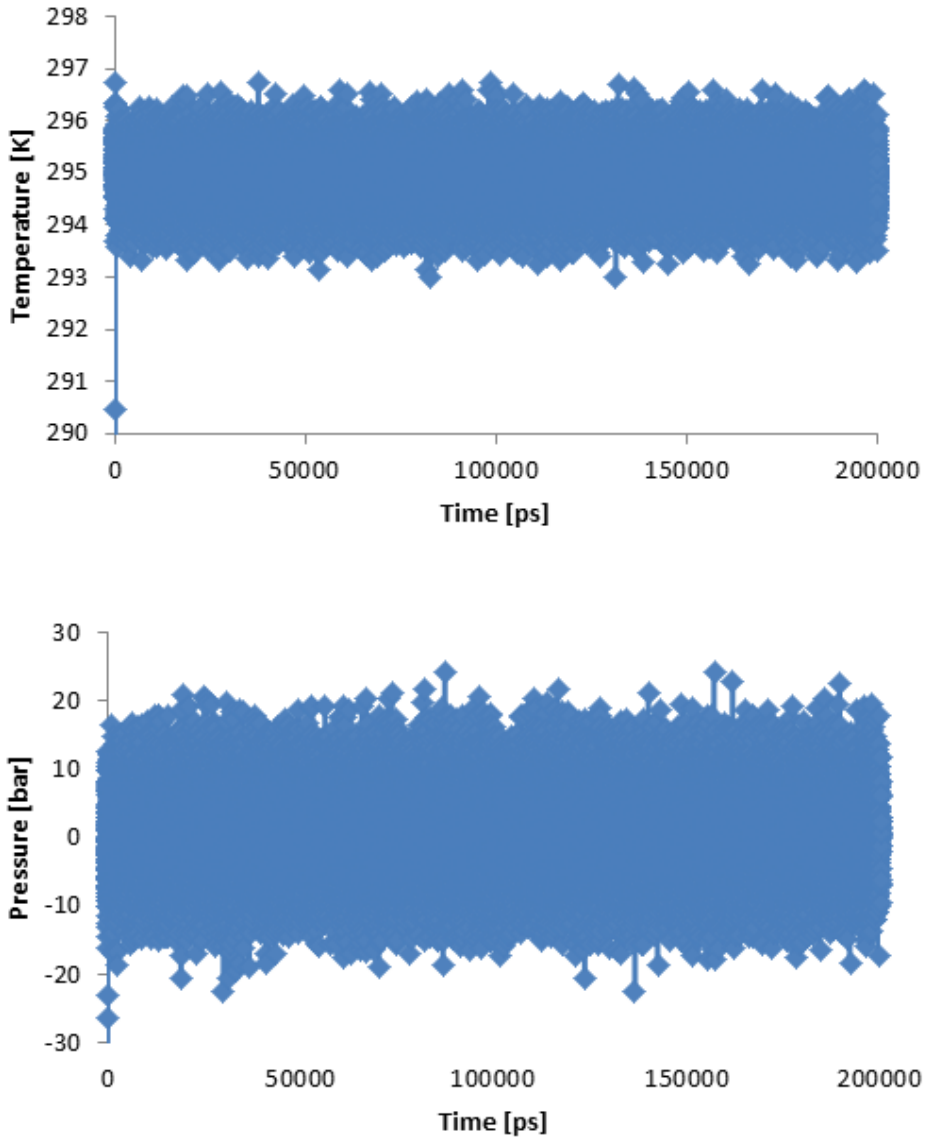

Figure 35. Temperature and Pressure During LD Equilibration.

Assembly of Lipid Droplet and apoB-100

As to the best of our knowledge there is no particular software that could assemble protein and lipids together. Thus we had to manually prepare molecules for the process of assembly. We also wrote a program that would test any overlap between LDL and protein. The procedure we performed consists of two main steps. In the first step protein and lipid droplet were placed in separate simulation boxes of the same size. Then, they were aligned with the principle axes. We proceeded by placing lipid droplet into the center of the box and we placed protein roughly in the vicinity of immunoelectron microscopy results. This brought apoB100 and LD in their relative position as shown on Figure 36. In this figure, protein and LD are in separate simulation boxes (separate files), and they were just overlaid one on top of another for the purpose of our program. If these two molecules were placed in the same simulation box (same file) it would result in serious steric clashes and we could not proceed any further.

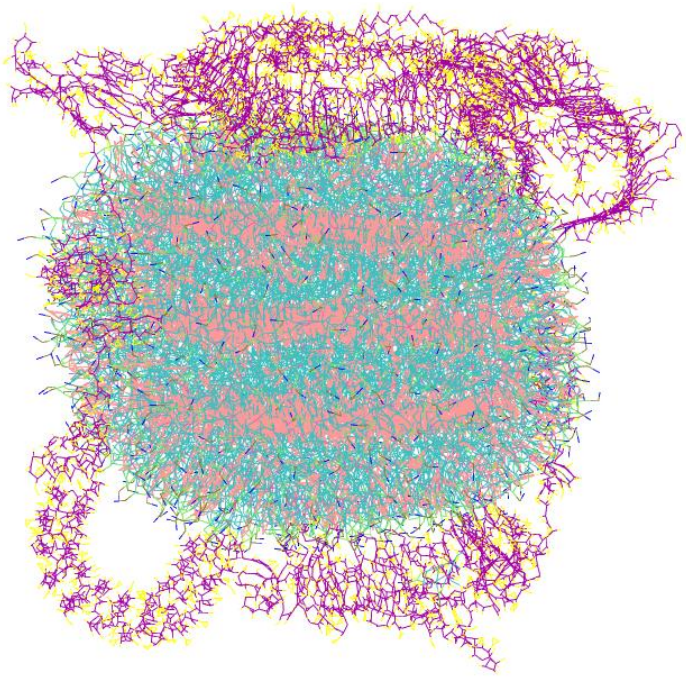

Figure 36. Alignment of Protein and LD. Protein and LD are in Separate Simulation Boxes.

To overcome previously described problem, the second step of assembling was treated by our program. Program tested the distance between each atom in the lipid droplet with all atoms in the protein for possible steric clashes and overlaps. If any atom of a lipid droplet was closer than 0.5 nm to any protein atom, it was removed from the system. The remaining structure was automatically written in the format of GROMACS structural file *.gro. In total, 25 of CO, 27 of TO, 119 of POPC, 23 of PPC, and 66 of CHOL molecules were removed due to overlap with protein. After this, the removed molecules were randomly placed back into the simulation box using GROMACS tools. Finally, the system was minimized to avoid any inappropriate geometry.

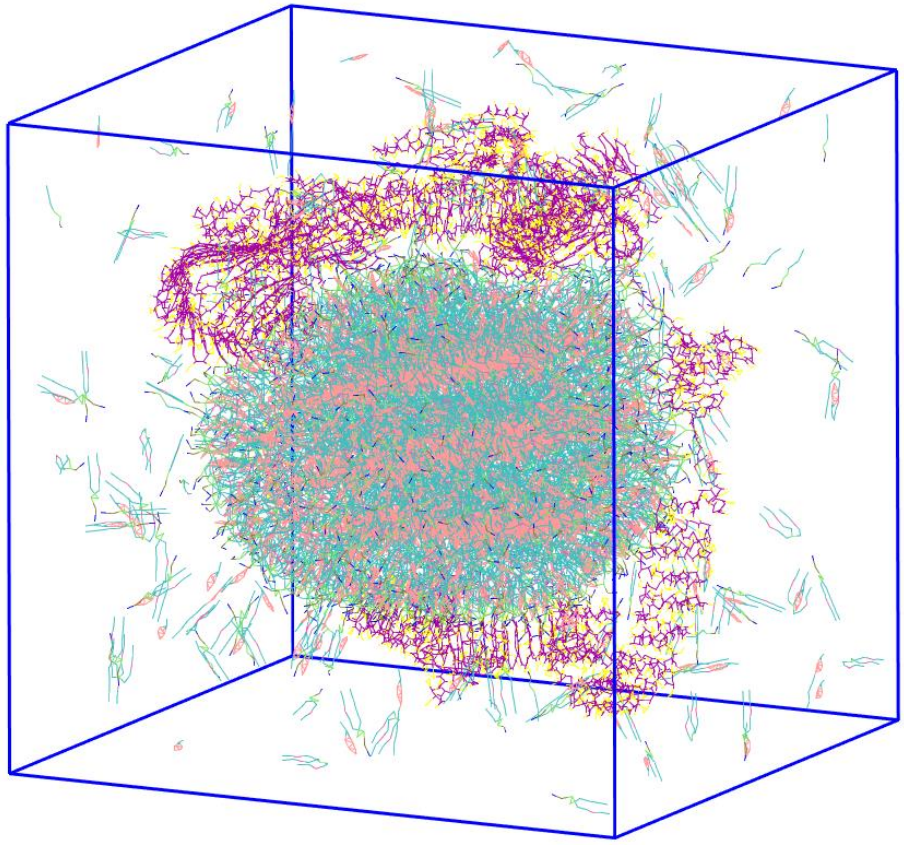

Figure 37. Lipid Molecules That had Clashes With ApoB100 Were Removed from the LD and Randomly Placed Back to the Simulation Box.

Now, it was time to bring these removed molecules back toward the lipid droplet. We investigated two cases: in the first case we solvated and neutralized the system with ions and equilibrated it for a long time (several microseconds). This system was slow to equilibrate and it failed to form good nanoparticle over a simulation span of few microseconds. Randomly oriented lipid molecules formed a number of micelles in simulation box and slowly reassemble on a surface of LD. Thus, in the second case, we performed a brief NpT simulation in implicit solvent. In this case lipids were rapidly attracted by the lipid droplet and the protein. Similar procedure was performed before [56].

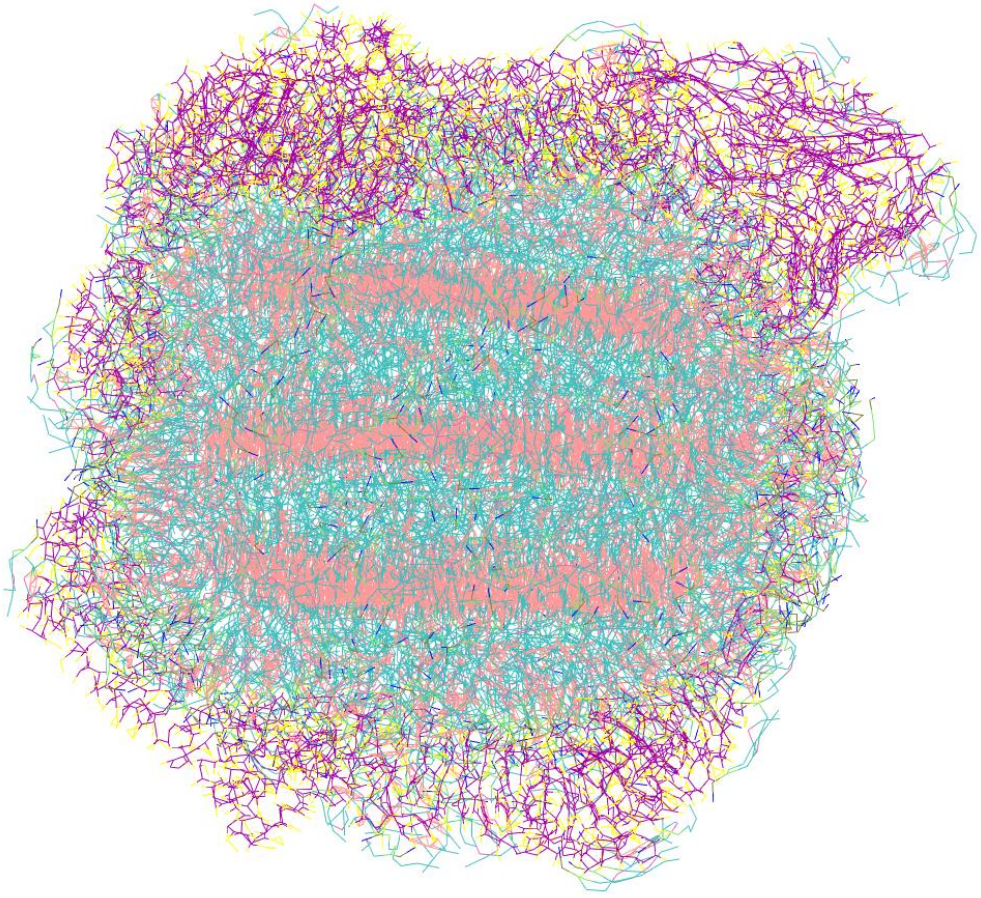

Figure 38. Assembled LDL Particle.

Assembled LDL Molecular Dynamics

In order to prepare our system for final equilibration, we placed LDL in new box that is 3nm larger than LDL and solvated it with water. After solvation system was minimized.

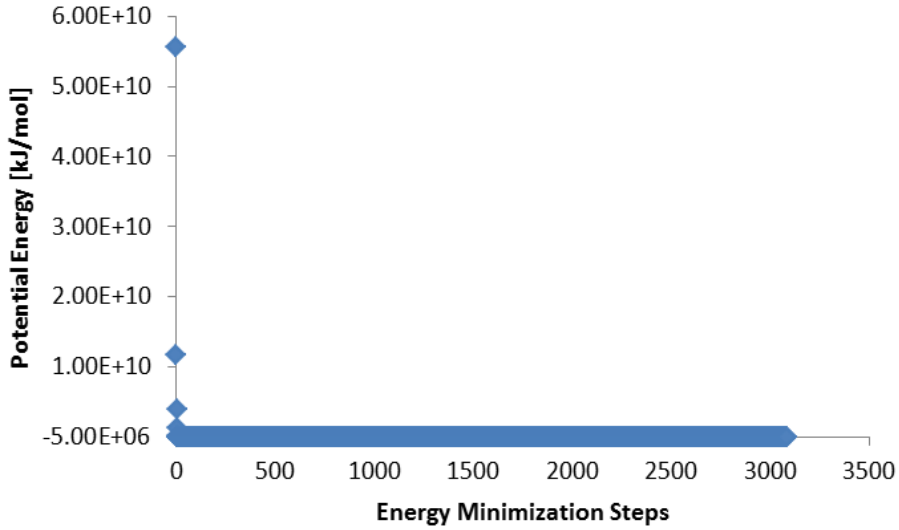

Figure 39. Energy Minimization of Solvated Assembled LDL Particle.

Ions were placed into the system by replacing the water molecules. Since normal concentration of Na^+ ions in blood is between 135 to 150 mmol/L, and Cl^- ions are in range between 95 to 110 mmol/L [57], we decided to work with concentration of ~140 mmol/L. Additional 26 positive Na^+ ions were added into the system to neutralize all the protein charges. Once again, solvated system with ions was minimized. This was short minimization of 64 steps with final potential energy of -4906324 kJ/mol and maximum force of 129.392 kJ/(mol nm). This indicates that system is properly minimized with force lower than 1000 kJ/(mol nm). In other words, these values suggest stable simulation.

Before starting equilibration process ions and water molecules were coupled into a separate temperature coupling group, lipids were coupled into the second temperature group and protein

was put into the third temperature coupling group. First we equilibrated system at 295°K in NVT ensemble for 100ns using Berendsen temperature coupling scheme. This allowed us to stabilize temperature of the system.

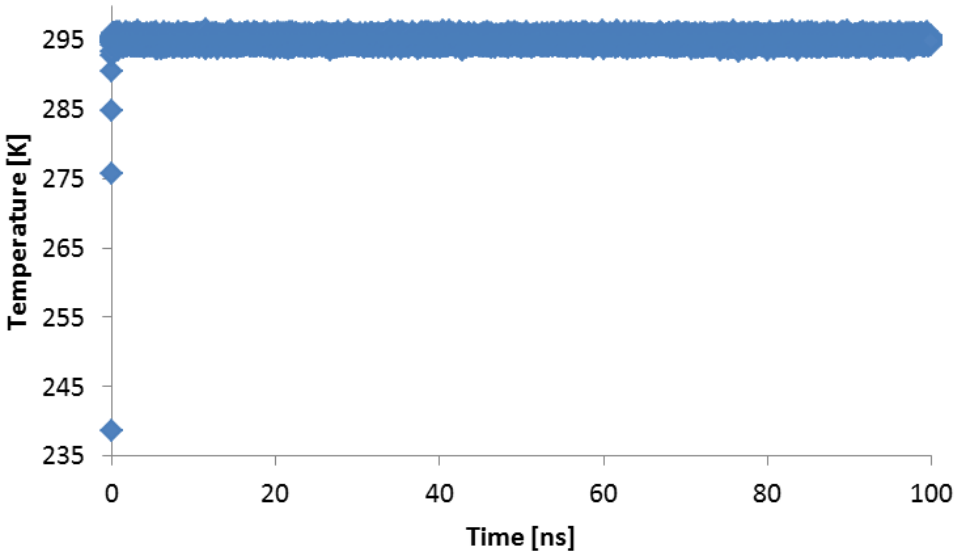

Figure 40. A Plateau at 295 K Implies Successful Stabilization of Temperature.

NpT ensemble was used to stabilize pressure at 1 bar and temperature at 295°K for additional 200ns. For the first 100ns Berendsen thermostat was used and it was changed in v-rescale in the last 100ns of equilibration. Pressure coupling was performed using Berendsen barostat in both cases. Average reported pressure was 0.956279 bar and average temperature was 294.9365 K. One can observe that value for pressure fluctuate widely about 1bar over the NpT equilibration as shown on next figure. This behavior is normal in molecular dynamics and it is rooted in small size of the investigated system.

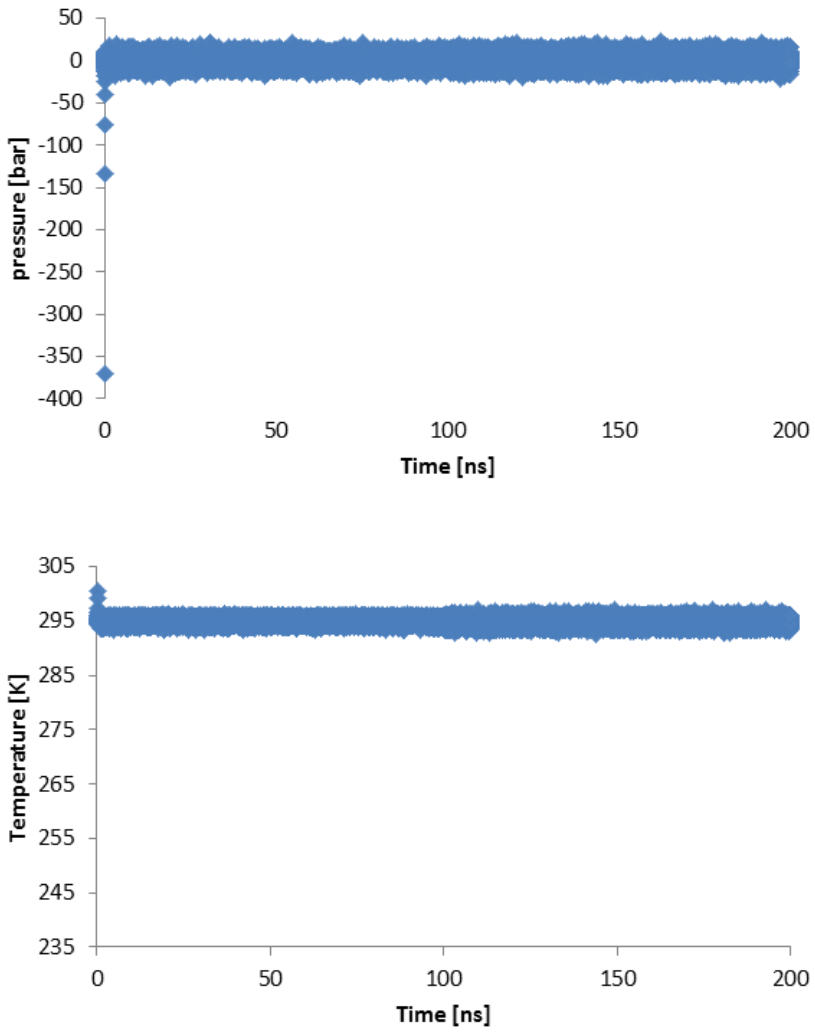

Figure 41. Plateaus at1bar and Temperature of 295K Show that After Our System Reaches Desired Pressure and Temperature it Stays Stable at Those Values Through the Rest of the Equilibration.

Upon completion of the equilibration we preformed production molecular dynamics at temperature 295°K, and pressure 1bar for 100ns. Temperature coupling was performed using v-rescale thermostat and pressure was controlled via Berendsen barostat.

Analysis

Dynamical properties of LDL were investigated during 100ns of production run. Here we were interested in stability of the protein and diffusion of CO in lamellar conformation.

Root mean square deviation (RMSD) was used to compares two structures. RMSD of an atom in the molecule is calculated by least-square fitting of a structure with respect to the reference structure at the beginning of the production using GROMACS g_rms tool. Smaller RMSD indicates spatially similar proteins. In this work we measured RMSD of all atoms in the protein.

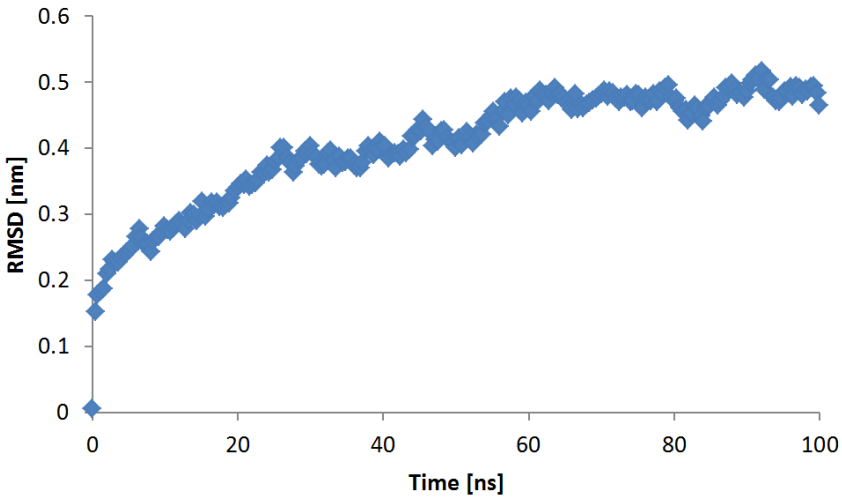

Figure 42. RMSD of All Atoms in the ApoB100.

Figure 42 shows that RMSD is almost constant at value of 0.5 after 60 ns of production simulation indicating that protein is in stable conformation. Recent development of elastic networks in combination with MARTINI FF showed that in some cases RMSD can get even smaller value. Since these methods are still in development we decided to apply only classical approach in our protein investigation. In addition, smaller value for RMSD may be obtained investigating RMSD of only $C_{\propto}$ atoms. The difference arise from the fact that RMSD of $C_{\propto}$

atoms treats only the main chain which is more constrained than the lateral side chains of amino acids.

Compactness of the protein can be assessed using the radius of gyration R_g. Stable proteins tend to have somewhat constant value of radius of gyration. If radius of gyration changes over time it means that protein is unfolding.

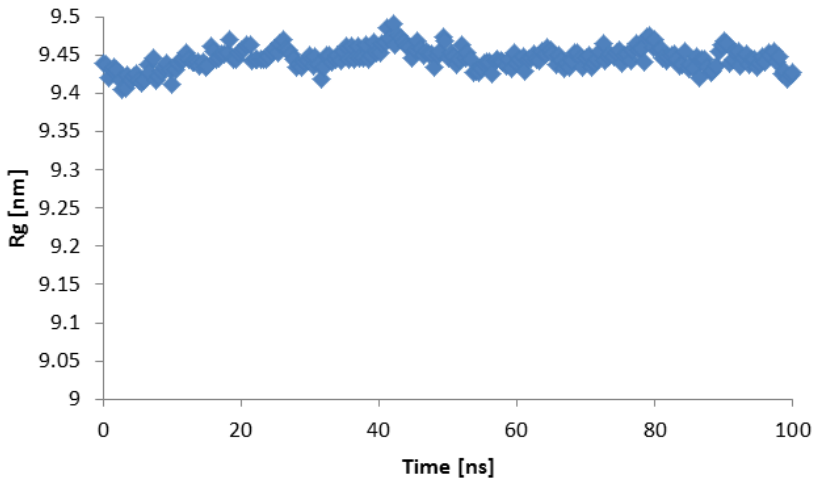

Figure 43. Radius of Gyration of ApoB100.

Radius of gyration indicates spatial distribution of the amino acids of a protein. In case of apoB100 it is reasonably constant with value that fluctuates between 9.4nm and 9.49nm. This means that protein remain stable during the course of simulation in its folded state.

Diffusion coefficient *D* was calculated using GROMACS g_msd tool. In a case of MARTINI FF a conversion factor of 4 must be applied to the result obtained by the GROMACS tool. This conversion factor takes in account for the faster diffusion of coarse-grained molecules. The reason for this faster diffusion lies in the smoothened free energy surface of coarse-grained force field [53].

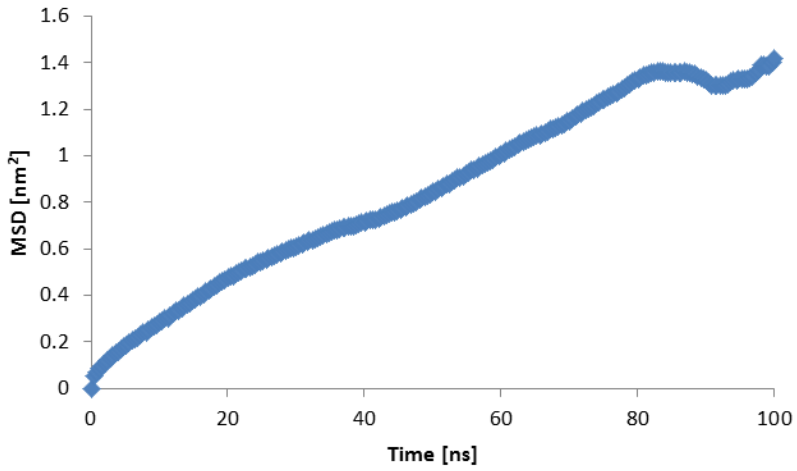

Figure 44. Mean Square Displacement of CO Molecules.

After conversion, diffusion coefficient obtained for CO molecules in the core of nanoparticle is D=0.23 10^{-7} cm^2/s. This result is order of magnitude smaller than result for diffusion coefficient obtained experimentally for lipids on the surface of the LDL and LDL-like lipid droplets [58], and it is also comparable with results obtained for CO in high density lipoproteins at 300K [59]. In that research reported value for diffusion coefficient of CO and TG molecules was 0.3 10^{-7} cm^2/s. Value obtained in the core may indicate that the shell of the particle slows down motions of the CE in the core.

In order to investigate influence of apoB100 on core molecule mobility, in particular on mobility of CE mobility, we calculated diffusion coefficient of CE molecules in LDL-like lipid droplet. Obtained value of D=0.2875 10^{-7} cm^2/s indicates weak influence of apoB100 on CE diffusion coefficient. The slight increase may be due to missing interactions of peripheral CE molecules with apoB100.

CHAPTER VI

CONCLUSIONS

In this work we investigated complex structure of Low Density Lipoprotein particle by employing statistical mechanics, structural bioinformatics, and molecular dynamics. Through series of computational simulations we were able to model structure of LDL.

In our evaluation of thermodynamic behavior of LDL, we developed a theoretical model using statistical mechanics. The proposed model is capable to explain variety of different effects, including the effects of triglycerides within the core of LDL. Capability to distinguish different particle sizes and to assess TG content in the LDL core based on thermodynamic behavior of the sample could be potentially beneficial for current estimations of LDL cholesterol. Particularly in the case when sample contains a large number of small particles with TG-rich and CE-poor core. In these cases, traditional lipid profile may estimate values of the cholesterol contained in the LDL (LDL-C) that are 50 mg/dl lower compared with people who have large particles with normal composition. In other words, patient's CVD risk is not estimated correctly [39].

Using structural bioinformatics, and molecular dynamics we were able to simulate lamellar structures inside LDL. Our detailed model of LDL structure is consistent with experimental observations. In this work we clarified complex structure of apoB100, consisting of 5 different domains. We showed how these domains are connected and how they interact with each other. We also showed the final LDL assemblage, pointing all the obstacles we found to create the final structure.

We elucidated the molecular structure of apoB-100. More future work is necessary on dynamical aspect of the apoB-100 and its role in recognition mechanisms. Detailed structural model of apoB-100 can be used in future for evaluating the recognition mechanism between receptor and lipid droplet with potential medical applications.

BIBLIOGRAPHY

1. Alpert, Joseph S. "A few unpleasant facts about atherosclerotic arterial disease in the United States and the world." The American journal of medicine 125.9 (2012): 839-840.
2. Frostegård, Johan. "Immunity, atherosclerosis and cardiovascular disease." BMC medicine 11.1 (2013): 117.
3. Stocker, Roland, and John F. Keaney. "Role of oxidative modifications in atherosclerosis." Physiological reviews 84.4 (2004): 1381-1478.
4. R.Prassl and P. Laggner. Lipoprotein Structure and Dynamics: Low Density Lipoprotein Viewed as a High Dynamic and Flexible Nanoparticle. In: Kostner G. and Frank S, editiors. Lipoproteins - Role in Health and Diseases. Rijeka, Croatia: InTech, 2012.
5. Kumar, Vibhor, et al. "Three-dimensional cryoEM reconstruction of native LDL particles to 16Å resolution at physiological body temperature." PloS one 6.5 (2011): e18841.
6. Murtola, Teemu, et al. "Low density lipoprotein: structure, dynamics, and interactions of apoB-100 with lipids." Soft Matter 7.18 (2011): 8135-8141.
7. Brown, Michael S., and Joseph L. Goldstein. "A receptor-mediated pathway for cholesterol homeostasis." Science 232.4746 (1986): 34-47.
8. Bossé, Yohan, et al. "Heritability of LDL peak particle diameter in the Quebec Family Study." Genetic epidemiology 25.4 (2003): 375-381.
9. Austin, Melissa A. "Genetic epidemiology of low-density lipoprotein subclass phenotypes." Annals of medicine 24.6 (1992): 477-481.

10. A. R. M. Barbagallo et al., "Low-density-lipoprotein peak particle size in a Mediterranean population.," *European journal of clinical investigation, 33(2),*, pp. 126-133, 2003.
11. Chapman, M. J., M. Guerin, and E. Bruckert. "Atherogenic, dense low-density lipoproteins. Pathophysiology and new therapeutic approaches." European heart journal 19 (1998): A24-30.
12. Blake, Gavin J., et al. "Low-density lipoprotein particle concentration and size as determined by nuclear magnetic resonance spectroscopy as predictors of cardiovascular disease in women." Circulation 106.15 (2002): 1930-1937.
13. Mora, Samia, et al. "LDL particle subclasses, LDL particle size, and carotid atherosclerosis in the Multi-Ethnic Study of Atherosclerosis (MESA)." Atherosclerosis 192.1 (2007): 211-217.
14. Nigon, Fabienne, et al. "Discrete subspecies of human low density lipoproteins are heterogeneous in their interaction with the cellular LDL receptor." Journal of lipid research 32.11 (1991): 1741-1753.
15. Galeano, Narmer F., et al. "Small dense low density lipoprotein has increased affinity for LDL receptor-independent cell surface binding sites: a potential mechanism for increased atherogenicity." Journal of lipid research 39.6 (1998): 1263-1273.
16. Deckelbaum, Richard J., et al. "Reversible modification of human plasma low density lipoproteins toward triglyceride-rich precursors. A mechanism for losing excess cholesterol esters." Journal of Biological Chemistry 257.11 (1982): 6509-6517.
17. Rizzo, Manfredi, and Kaspar Berneis. "The Clinical Significance of Small, Dense Low-Density Lipoproteins." Biochemistry of Atherosclerosis. Springer US, 2006. 168-185.

18. Otvos, James. "Measurement of triglyceride-rich lipoproteins by nuclear magnetic resonance spectroscopy." Clinical cardiology 22.S2 (1999): II-21.

19. Otvos, James D., Elias J. Jeyarajah, and William C. Cromwell. "Measurement issues related to lipoprotein heterogeneity." The American journal of cardiology 90.8 (2002): 22-29.

20. Deckelbaum, RICHARD J., G. GRAHAM Shipley, and D. M. Small. "Structure and interactions of lipids in human plasma low density lipoproteins." Journal of Biological Chemistry 252.2 (1977): 744-754.

21. Orlova, Elena V., et al. "Three-dimensional structure of low density lipoproteins by electron cryomicroscopy." Proceedings of the National Academy of Sciences 96.15 (1999): 8420-8425.

22. Chatterton, J. E., et al. "Immunoelectron microscopy of low density lipoproteins yields a ribbon and bow model for the conformation of apolipoprotein B on the lipoprotein surface." Journal of lipid research 36.9 (1995): 2027-2037.

23. Borén, Jan, et al. "Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100." Journal of Clinical Investigation 101.5 (1998): 1084.

24. Kriško, Anita, and Catherine Etchebest. "Theoretical model of human apolipoprotein B100 tertiary structure." Proteins: Structure, Function, and Bioinformatics 66.2 (2007): 342-358.

25. Gao, Qi, and B. M. Craven. "Conformation of the oleate chains in crystals of cholesteryl oleate at 123 K." Journal of lipid research 27.11 (1986): 1214-1221.

26. Ginsburg, Geoffrey S., David Atkinson, and Donald M. Small. "Physical properties of cholesteryl esters." Progress in lipid research 23.3 (1984): 135-167.

27. Albertini, G., et al. "Investigation of the Cholesteryl Oleate Mesophases by X-Ray Diffraction." Molecular Crystals and Liquid Crystals 70.1 (1981): 169-182.

28. Ginsburg, Geoffrey S., Donald M. Small, and James A. Hamilton. "Temperature-dependent molecular motions of cholesterol esters: a carbon-13 nuclear magnetic resonance study." Biochemistry 21.26 (1982): 6857-6867.

29. Hamilton, JAMES A., N. O. R. M. A. N. Oppenheimer, and E. H. Cordes. "Carbon-13 nuclear magnetic resonance studies of cholesteryl esters and cholesteryl ester/triglyceride mixtures." J Biol Chem 252.22 (1977): 8071-8080.

30. Marčelja, S. "Chain ordering in liquid crystals. I. Even-odd effect." The Journal of Chemical Physics 60.9 (1974): 3599-3604.

31. Doniach, S. "Thermodynamic fluctuations in phospholipid bilayers." The Journal of Chemical Physics 68.11 (1978): 4912-4916.

32. Ivanova, Vesselka P., and Thomas Heimburg. "Histogram method to obtain heat capacities in lipid monolayers, curved bilayers, and membranes containing peptides." Physical Review E 63.4 (2001): 041914.

33. Sugár, István P., Rodney L. Biltonen, and Neil Mitchard. "[25] Monte Carlo simulations of membranes: Phase transition of small unilamellar dipalmitoylphosphatidylcholine vesicles." Methods in enzymology 240 (1994): 569-593.

34. Heimburg, Thomas, and Rodney L. Biltonen. "A Monte Carlo simulation study of protein-induced heat capacity changes and lipid-induced protein clustering." Biophysical journal 70.1 (1996): 84.

35. Dubini, B., et al. "Study of Phase Transitions Involving the Mesophases of Cholesteryl Linoleate." Molecular Crystals and Liquid Crystals 95.1-2 (1983): 1-10.

36. McMillan, W. Lo. "Simple molecular model for the smectic A phase of liquid crystals." Physical Review A 4.3 (1971): 1238.

37. K. Huang, Statistical mechanics, New York: John Wiley & Sons, Inc, 1987.

38. Kroon, Paulus A. "The order-disorder transition of the core cholesteryl esters of human plasma low density lipoprotein. A proton nuclear magnetic resonance study." Journal of Biological Chemistry 256.11 (1981): 5332-5339.

39. Otvos, J. D. "Why Cholesterol Measurements May be Misleading about Lipoprotein Levels and Cardiovascular Disease Risk–Clinical Implications of Lipoprotein Quantification Using NMR Spectroscopy/Warum Cholesterin bei der Beurteilung von Lipoprotein-Konzentrationen und kardiovaskulärem Risiko in die Irre führt–Klinische Implikationen der Lipoproteinqualifizierung mittels NMR-Spektroskopie." LaboratoriumsMedizin 26.11/12 (2002): 544-550..

40. Segrest, Jere P., et al. "Structure of apolipoprotein B-100 in low density lipoproteins." Journal of lipid research 42.9 (2001): 1346-1367.

41. Prassl, Ruth, and Peter Laggner. "Molecular structure of low density lipoprotein: current status and future challenges." European Biophysics Journal 38.2 (2009): 145-158.

42. Chatterton, J. E., et al. "Immunoelectron microscopy of low density lipoproteins yields a ribbon and bow model for the conformation of apolipoprotein B on the lipoprotein surface." Journal of lipid research 36.9 (1995): 2027-2037.

43. Xu, Dong, et al. "AIDA: ab initio domain assembly server." Nucleic acids research 42.W1 (2014): W308-W313.

44. Johnson, Mark, et al. "NCBI BLAST: a better web interface." Nucleic acids research 36.suppl 2 (2008): W5-W9.

45. Pavlopoulou, Athanasia, and Ioannis Michalopoulos. "State-of-the-art bioinformatics protein structure prediction tools (Review)." International journal of molecular medicine 28.3 (2011): 295-310.

46. Buchan, Daniel WA, et al. "Scalable web services for the PSIPRED Protein Analysis Workbench." Nucleic acids research 41.W1 (2013): W349-W357.

47. Anderson, T. A., D. G. Levitt, and L. J. Banaszak. "The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein." Structure 6.7 (1998): 895-909.

48. Webb, Benjamin, and Andrej Sali. "Comparative protein structure modeling using Modeller." Current protocols in bioinformatics (2014): 5-6.

49. Yang, Jianyi, et al. "The I-TASSER Suite: protein structure and function prediction." Nature methods 12.1 (2015): 7-8.

50. Johansson, Maria U., et al. "Defining and searching for structural motifs using DeepView/Swiss-PdbViewer." BMC bioinformatics 13.1 (2012): 173.

51. Van Der Spoel, David, et al. "GROMACS: fast, flexible, and free." Journal of computational chemistry 26.16 (2005): 1701-1718.

52. Marrink, Siewert J., et al. "The MARTINI force field: coarse grained model for biomolecular simulations." The Journal of Physical Chemistry B 111.27 (2007): 7812-7824.

53. S. J. Marrink, "Martini Force Field," [Online]. Available: http://md.chem.rug.nl/cgmartini/index.php/downloads.

54. Prassl, Ruth, et al. "Low density lipoproteins as circulating fast temperature sensors." PloS one 3.12 (2008): e4079.

55. Pastor, Richard W., Richard M. Venable, and Scott E. Feller. "Lipid bilayers, NMR relaxation, and computer simulations." Accounts of chemical research 35.6 (2002): 438-446. *of chemical research*, pp. 35(6), 438-446, 2002.

56. Murtola, Teemu, et al. "Low density lipoprotein: structure, dynamics, and interactions of apoB-100 with lipids." Soft Matter 7.18 (2011): 8135-8141.

57. Oesch, Urs, Daniel Ammann, and Wilhelm Simon. "Ion-selective membrane electrodes for clinical use." Clinical Chemistry 32.8 (1986): 1448-1459.

58. VAUHKONEN, Matti, et al. "Lateral diffusion of phospholipids in the lipid surface of human low-density lipoprotein measured with a pyrenyl phospholipid probe." European Journal of Biochemistry 186.3 (1989): 465-471.

59. Vuorela, Timo, et al. "Role of lipids in spheroidal high density lipoproteins." PLoS computational biology 6.10 (2010): e1000964.