Specific cholesterol binding drives drastic structural alterations in apolipoprotein A1

Arjun Ray[1,2], Shantanu Sengupta[1,2*] and Lipi Thukral[1,2,3*]

[1] CSIR-Institute of Genomics and Integrative Biology, South Campus, Mathura Road, New Delhi, 110025, India
[2] Academy of Scientific and Innovative Research (AcSIR), CSIR- Institute of Genomics and Integrative Biology, Mathura Road Campus, New Delhi, 110025, India
[3]Interdisciplinary Center for Scientific Computing, University of Heidelberg, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

∗ E-mail: Correspondence to Lipi Thukral (lipi.thukral@igib.res.in) and Shantanu Sengupta (shantanus@igib.res.in)

Abstract

Aberrant regulation of the cholesterol transport leads to several chronic metabolic disorders. Apolipoprotein A1 (ApoA1) is a key player involved in removing excess cholesterol from the cells, and is also a major constituent of high-density lipoprotein (HDL) that regulates lipid metabolism. While mechanisms governing HDL formation have been the focus of several investigations, structural properties of monomeric ApoA1 are poorly understood. Here, we report cholesterol-free and bound states of monomeric ApoA1 using μs-long (>50 μs in total) atomistic simulations in explicit solvent. Our results indicate that lipid-free ApoA1 exhibits spatial proximity between N- and C-terminal domains, resulting in a closed conformation. However, upon cholesterol addition, ApoA1 spontaneously forms open circular topology. Remarkably, these drastic structural perturbations are driven by specific binding sites and, high C-terminal mobility. Simulations reveal two distinct cholesterol binding sites, one at the C-terminal and a novel cholesterol binding site at the N-terminal. Furthermore, low concentration of cholesterol, leading to less N-terminal binding does not manifest in open topology, highlighting the importance of the N-terminal binding site. A population density map generated from tens of simulations with different starting configurations revealed the existence of distinct populated states, indicating stable protein regions and an obligatory intermediate with distinct N-terminal helix pairs (H1-H7;H4-H7) participating in the opening process. Collectively, this study suggests a previously unknown mechanism for sequestering of cholesterol by ApoA1 that explains its functional diversity.

Introduction

In all cells, cholesterol is removed through reverse cholesterol pathway (RCT). This process is of particular importance in ensuring lipid homeostasis, via the transfer of cholesterol from peripheral cells to the liver. The first pivotal step is the formation of high-density lipoprotein (HDL) complex, that involves sequestering of cholesterol by apolipoproteins. Apolipoprotein A1 also commonly referred to as ApoA1, is the major protein constituent of HDL ($\approx$ 70%). Although a relatively small protein of 243 amino acids, ApoA1 exhibits remarkable functional variability.[1] Given the association of dysfunctional cholesterol efflux with chronic cardiovascular diseases, understanding the structural basis of ApoA1 functional diversity has been a long-standing goal of many previous studies.[2,3] ApoA1 shows a great number of structural alterations with distinct level of complexity and selectivity.[4] These conformational changes are associated with different stages of HDL regulation (Figure 1A). Efforts to determine experimental structures have been hindered by the cholesterol driven aggregation and structural dynamics displayed by ApoA1. Thus, establishing structures and understanding dynamic interconversion between cholesterol-free and -associated monomeric ApoA1 may shed light on critical initial process of RCT.

The ApoA1 structure, suggested by proteolysis[5] and mutagenesis studies,[6,7] is divided into two structural domains i.e., N- and C- terminal that are highly conserved across species (Figure 1B). Interestingly, the C-terminal α-helix is known to directly mediate lipid efflux[8] and also known to solubilize the membrane phospholipids.[9,10] While the role of N-terminal is to impart protein stability. A major question arising in this context is what are the binding sites and how does ApoA1 monomer achieve these large conformational changes upon cholesterol addition. In recent years, technological advancement has brought molecular dynamics (MD) simulations to μs-timescale. These developments have had a particular impact in biology, where experimental investigations of single molecule that were prohibited by the fast-timescale dynamics, can now be studied. Here we employed long-timescale (>50 μs) MD simulations in explicit solvent to study cholesterol-free and cholesterol-bound ApoA1 monomer dynamics. Multiple independent simulations allowed us to study ApoA1 dynamics with good statistical reliability and to further our understanding of molecular mechanism underlying its conformational transitions.

All lipid-free ApoA1 trajectories converged to a common topology with juxtaposed N- and C-terminal domains, in accord with previous experimental findings.[11] We present cholesterol-mediated ApoA1 dynamics, addressing an open question in the field of reverse cholesterol pathway. Our results show

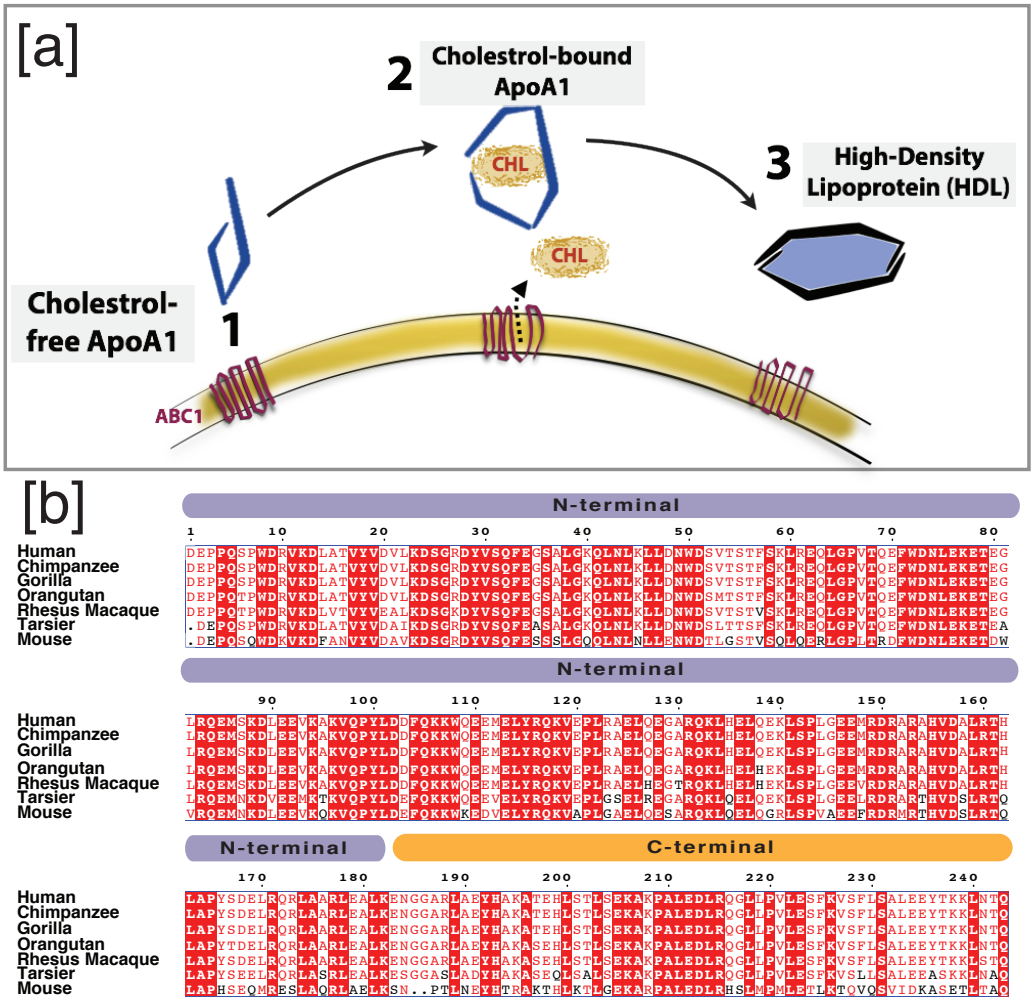

Figure 1: **The Biology of ApoA1 and Protein Domain Architecture.** A) Schematic diagram of reverse cholesterol pathway. The role of ApoA1 is highlighted in various stages. In the first step, ApoA1 exists in a monomer state with no association with cholesterol in blood plasma. The cholesterol is then transported via membrane receptor and, subsequently ApoA1 engages with cholesterol to form ApoA1-cholesterol associated monomeric state. Lastly, other proteins and phospholipids come together to form HDL. While the details of HDL are known, the aim of the present study is to reveal the structural details and dynamic rearrangement of ApoA1 during the first two steps which is limited in literature. B) Sequence alignment between human ApoA1 and other species. The N- and C-terminal domains are denoted as bars at the top of the alignment.

spontaneous drastic belt or open shaped ApoA1 transition upon cholesterol addition as opposed to bundled or closed shaped topology in the lipid-free state. Molecular mechanism underlying these structural arrangements was investigated to understand regions driving these large structural perturbations. We reveal cholesterol binding sites and showed that they are evolutionarily conserved across species. Our results resolve the ambiguity regarding the role of N-terminal in protein structural changes coupled to ApoA1 function.

Furthermore, the conformational landscape of structural transition shows distinct populated states, indicating stable protein regions. We found that dynamic interaction of protein domain elements illustrates specific functional attributes that act in concert to enable drastic conformational transitions between closed and open topology of the protein. To our knowledge, this is the first time a robust monomer ApoA1 dynamics study has been conducted *in silico* with particular attention to dynamic conversion of configurations.

Materials and Methods

Three-Dimensional Model Building of ApoA1

There is no complete crystal structure of monomeric ApoA1 lipid-free state. In order to investigate the structural properties, we have used a hybrid protocol to generate a complete ApoA1 monomeric lipid-free structure. In this multi-step approach, we utilised experimentally resolved truncated structure of ApoA1 followed by loop modeling and extensive μs-timescale simulations to generate a complete atomistic model of ApoA1. The detailed protocol is shown in Figure S1. In the first step, we built a homology model of ApoA1 structure using a mouse C-terminal-truncated ApoA1 protein (PDBID: 2LEM). The initial structure showed compact N-terminal helices and structural overlap of 0.5 $\AA$ RMSD with mouse ApoA1 structure. To obtain an accurate C-terminal domain, we performed extensive loop-refinement targeted majorly at the C-terminal region followed by molecular dynamics refinement (384 ns) for the complete structure. The output of the short-MD refined structure served as a prelude to our final extensive set of MD simulations, from here referred to as the starting structure.

Molecular dynamics simulations

The MD simulations were performed with the program GROMACS 4.6.1.[12] The system was represented by CHARMM36 all-atom force-field.[13–15] The

water was modeled using the TIP3P representation.[16] Each of the five starting conformations was placed in a dodecahedral water box (volume = 2718.19 nm^3) large enough to contain the peptide and at least 1.0 nm of solvent on all sides. Periodic boundary conditions were used, and the long-range electrostatic interactions were treated with the particle mesh Ewald method[17] using a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation to compute the potential and forces in-between grid points. The real space cutoff distance was set to 0.9 nm and the van der Waals cutoff to 1.4 nm. The bond lengths were fixed,[18] and a time step of 2 fs for numerical integration of the equations of motion was used. Coordinates were saved every 200 ps. The pressure coupling was done by employing a Parrinello-Rahman barostat using a 1 bar reference pressure and a time constant of 2.0 ps with compressibility of 4.5e-5 bar using isotropic scaling scheme. Nine positive counter-ions (Na+) were added, replacing nine water molecules so as to produce a neutral simulation box. All the starting structures were subjected to a minimization protocol for 50000 steps using the steepest descent method followed by an equilibration run of 100 ps in NVT ensemble with restraints on the protein atoms.

The simulations were performed using the starting structure of ApoA1. Five independent trajectories, each 7 μs at 310 K were carried out. The combined time-scale of our simulations is 35 μs. In addition to ApoA1 protein simulations, simulation of ApoA1 protein in the presence of cholesterol were also performed. The closed topology of ApoA1 structure derived from lipid-free simulations was taken as the starting structure. The cholesterol molecules were randomly distributed around ApoA1 protein. Various simulations with different stoichiometric ratios of cholesterol were performed. The detailed summary of simulations, is shown in Table T1. All the structures were subjected to similar MD protocol as described above, before the final production run was started.

Analysis

The hydrophobic patch analysis was calculated using the tool Quilt.[19] It provides atomistic description of surface non-polar atoms that are contributing to a contiguous hydrophobic patch using a dot representation of a solvent accessible surface area. Using the polar expansion radius of 0.14 nm, the area and the corresponding number of atoms constituting the patch(es) are reported. The secondary structure analysis was performed with do_dssp script of GROMACS. It provides secondary structure of each residue of protein across the simulation length. We further converted these into probability distribution for the secondary structure elements using in-house developed

script.

Finally, the conservation analysis was performed by Consurf webserver.[20,21] We performed multiple sequence alignment with ApoA1 sequences from 37 species to reveal its evolutionary conservation profile. The web-based analysis generates conservation score in the range of 0 to 9. From this list, we selected 34 highly conserved residues.

Probability Distribution Contour Map

To enhance the sampling, multiple short-timescale (250 ns) simulations of ApoA1-cholesterol were initiated from different starting configurations. These starting or seed structures were obtained from pilot 2 μs trajectory of ApoA1-cholesterol. In total, 6 simulations of 6 structures were simulated for 250 ns, totalling 11 μs of computing time.

$$P_{ref \rightarrow i} = \frac{p_i}{p_{ref}} \tag{1}$$

In equation (1) where p_i and p_{ref} are the probabilities of finding the system in state i and state ref, respectively. We describe the probability distribution as a function of two order parameters, namely RMSD and the Rg parameter.

A grid 30X30 was used to divide this plane in 900 cells and for every cell the number of points counted and the relative probability calculated, allowing $P_{ref \rightarrow i}$ to be determined. Overall the landscape described three populated states (State I, II and III). To dissect their structural characteristics, we dissected the contour maps into three ensembles. The State I, containing the initial trajectory population, was defined by RMSD, Rg pair values of (0.4, 2.1) - (0.8, 2.2). State II contained intermediate structures was defined as (1.2, 2.5) - (1.5, 2.6) of RSMD, Rg pair value while state III, containing the final conformation, was defined between (1.7, 2.7) - (2.0, 2.8) of RMSD, Rg pair value. Out of the 4,88,161 structures considered, 40.15%, 33.13% and 26.70% resided in State I, II and III, respectively, illustrating a fairly even sampling for all the three states. In order to quantify kinetic transitions between states, we considered a stable transition, in both forward and reverse direction between states, with a residence time of at least 1 ns.

Results

Our study was aimed at investigating the structural properties of ApoA1 upon cholesterol binding. This is undoubtedly the most important step in biogenesis of high-density lipoprotein (HDL) formation. Our approach was twofold. We probed ApoA1 protein structure variations with and without cholesterol in all atomistic representation and, highlighted observed dynamic differences within protein domain elements that may be intricately coupled with its biological function.

Identification of lipid-free monomer ApoA1 structure reveals a closed topology.

To understand the functional significance of lipid-free ApoA1, we generated the ApoA1 starting structure using a hybrid modeling and simulation protocol (see Methods). Further, we performed multiple independent trajectories (5X7 μs), leading to determination of structure and dynamics with statistical reliability. The mapped structural dynamics at μs timescale is three order magnitude longer than previous reports.[11,22]

To fully describe the conformational changes, we utilised two parameters, namely radius of gyration (Rg) and angle calculation between N- and C-terminal domain (Figure 2A-B). As a function of time, the Rg plummeted from approx. 2.7 nm to 2.1 nm within 5 μs, and remained stable for last 2 μs in all simulations. Remarkably, in each of five simulations, ApoA1 converged to a common closed topology. Figure 2C shows the representative images of protein dynamics evolution and, a direct observation of the final structure completely showed a closed conformation, with intact N- and C-terminal domains. The time progress also showed that C-terminal region (residues 194-200) undergoes complete rearrangement to be in contact with N-terminal region residues. In particular, residues 108, 112-113, 116, 171 and 175 of the N-terminal helix are in close proximity to the C-terminal. The details of secondary structure assignment is shown in Fig S2.

The predicted dynamic properties are in good agreement with previously reported site-directed electron paramagnetic resonance spectroscopy, where the probes localised at residues 163, 217, and 226 revealed close association and parallel topology between N- and C-terminal.[23] In accord, we used similar physical probes to validate our data (Fig S3). The triad distance calculated from the probes showed concordance to our final conformation, indicating that ApoA1 monomer structure is in excellent agreement with previous experimental data. This structure thus enabled us to study its protein-cholesterol interaction with high reliability.

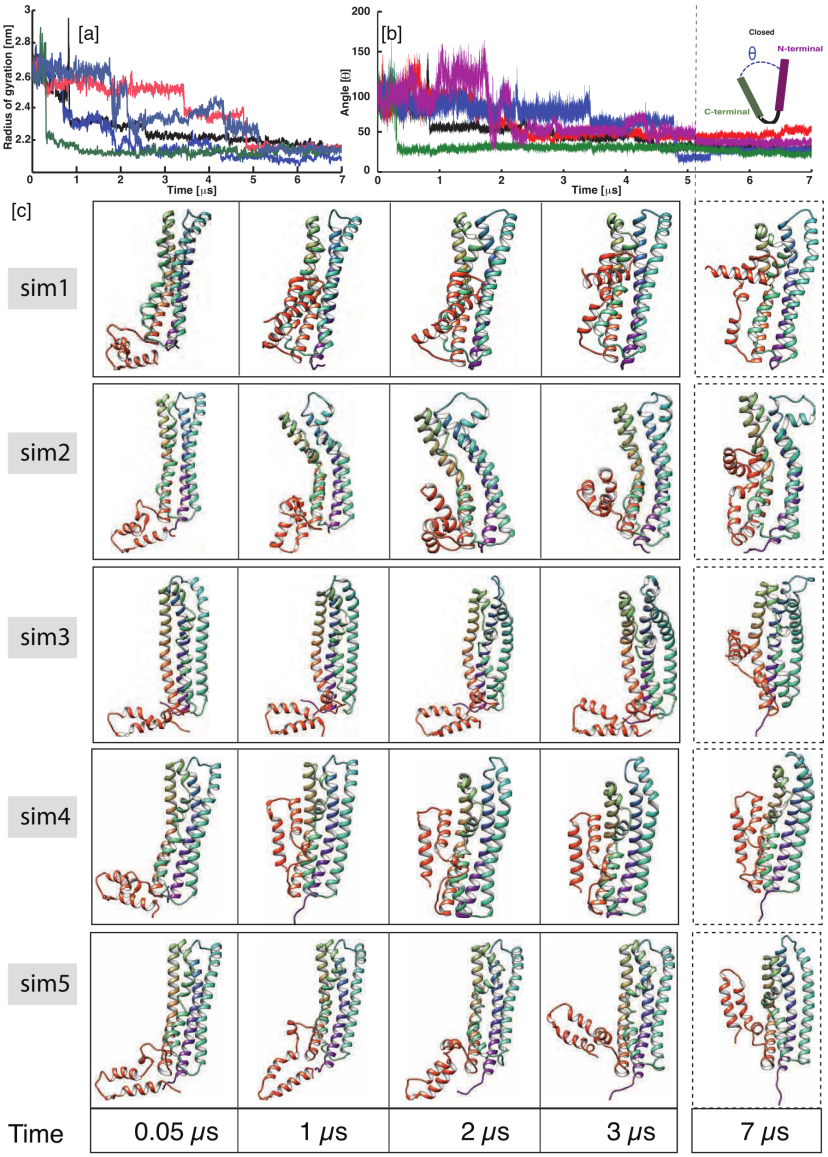

Figure 2: **ApoA1 Lipid-free Structure and Dynamics.** A) Time evolution of ApoA1 structure as a function of radius of gyration (Rg) for five independent molecular dynamics trajectories, shown in different colors. The Rg value shows drop in all simulations, indicating increase in the compactness of the structure. B) The evolution of angle between C- and N- terminal domain. The angle θ is defined by constructing the vector between 171, 182 and 214 residues across all simulations. C) Representative snapshots of all simulations showed closed topology towards the end of simulation time. The colour scale is denoted to represent different helices. The dotted line shows that the final structure in all trajectories represents the compact topology.

Spontaneous formation of ApoA1 open topology upon cholesterol binding and identification of preferential binding sites.

It is well characterized that ApoA1 exhibits protein regions that are likely to be essential for its physiological function. At the atomic level, however, our understanding is largely obscure due to fast-timescale conformational transitions that are experimentally difficult to characterize. Therefore, we assessed how cholesterol guides structural changes in monomeric ApoA1 protein. The evolution of dynamic structural changes with residue-based fluctuations are shown in Figure 3A. There are two major observations. First, cholesterol binds protein at two distinct sites that are subsequently merged in $\approx$ 400 ns, followed by opening up of ApoA1 structure to fully sequester the cholesterol. Second, the domain-wise protein mobility is highly variable. While the N-terminal is stable and well defined, the C-terminal region is highly mobile. At 1315 ns, the C-terminal twists and forms two short stretches of helix, before opening up and forming a circular or open topology. By the end of the simulation length, we found that all the residues interact with lipids, with the cholesterol in the center. These findings suggest that ApoA1 dynamically modulates its topology in the presence of cholesterol.

To quantitatively assess the cholesterol binding, we plotted the time occurrence of the cholesterol binding with ApoA1 (Figure 3B). During the initial simulation time ($\approx$140 ns), cholesterol specifically targets two binding sites. The C-terminal is the main site of cholesterol targeting and the second major binding pocket is at the top of the N-terminal helix bundle (Figure 3C). These two sites referred from here as Site I and Site II, respectively thus defines cholesterol binding sites with high propensity. Figure 3D shows time evolution of these two major cholesterol binding sites. There is a concomitant lipid binding at both sites followed cholesterol movement of N-terminal merging towards the core.

The results described above suggested that ApoA1 cholesterol association can be divided into following stages: a) Specific cholesterol binding at two sites, b) Redistribution with cholesterol merging from two sites, and c) Complete sequestering that involves N-terminal opening along with C-terminal plasticity(Figure 3E).

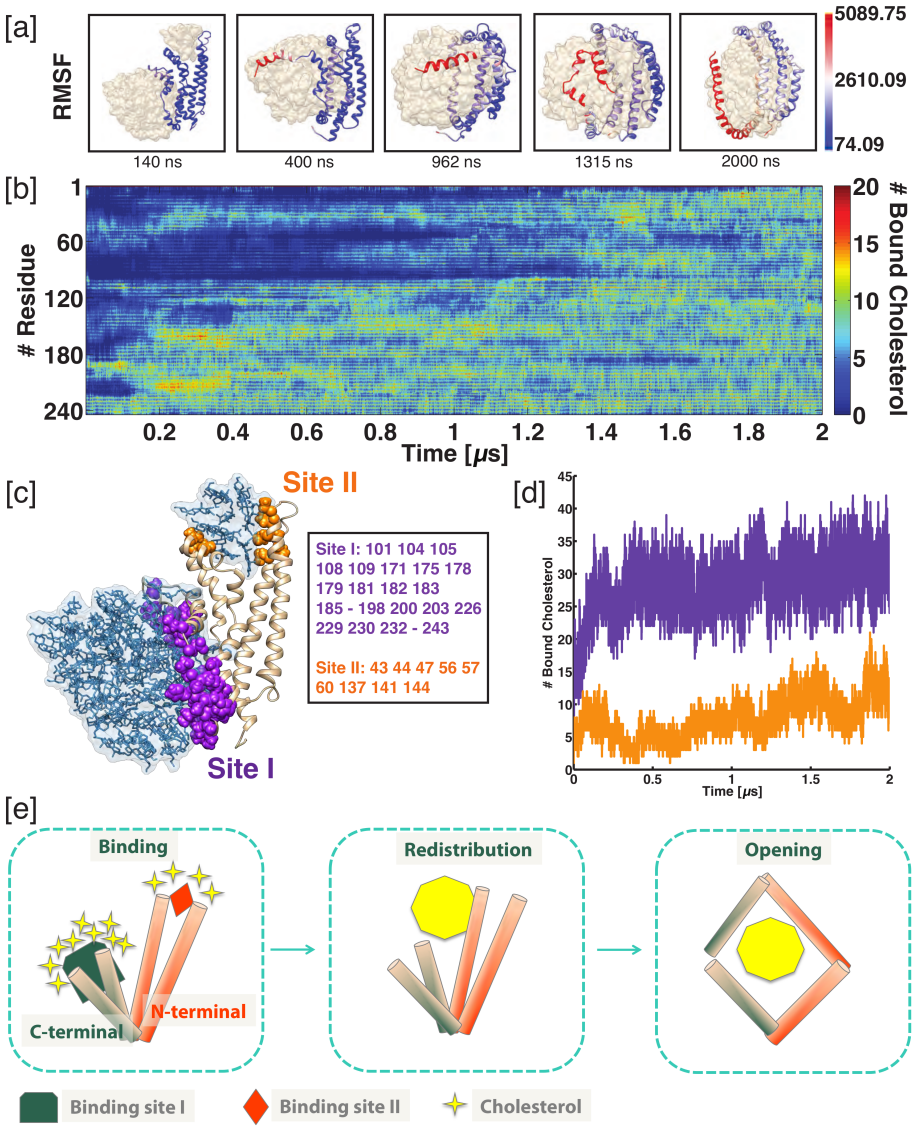

Figure 3: **Dual topological organization of ApoA1 upon cholesterol binding.** A) The representative snapshots are shown as a function of time for ApoA1-cholesterol simulations. The residue-based fluctuations (rmsf) are mapped onto the structural variations and the corresponding colour bar denotes the rmsf values in angstrom. B) The cholesterol distribution on each protein residue is shown along the entire simulation length. The color bar represents the number of bound cholesterol molecules. The trajectory movie in mpeg format is available. C) Representation of two major cholesterol binding sites and the participating residues. Site I refers to the binding site encompassing C-terminal while Site II refers to the binding site at N-terminal, shown in purple and orange, respectively. D) Time evolution of the two major cholesterol binding sites showed specific initial binding. E) Collectively, the events during ApoA1 cholesterol association can be divided into three stages shown in schematic representation.

Mutations in N-terminal negatively affect ApoA1 opening

Having identified probable mode of action for ApoA1, we next sought to validate the residues involved in the binding. Although the importance of C-terminal binding of cholesterol has been reported,[24,25] the role of N-terminal binding has not been established. We postulated that N-terminal binding sites are critical to attain the belt-like conformation of lipidated ApoA1. Since, cholesterol is expected to bind to the hydrophobic regions of the protein, we assessed exposed contiguous hydrophobic surface area using QUILT (see Methods).[19] Figure 4A showed that proximity of two terminal domains in lipid-free ApoA1, result in a drastic increase of contiguous hydrophobic patch composed of 40 residues (1749.2 $Å^2$) (Fig S4 and Table T2). The juxtaposition of domains in lipid-free ApoA1 structure, in principle, is the key factor that leads to an increased exposed hydrophobic surface area, thus directly facilitating cholesterol recognition at two binding sites. To further prove the role of N-terminal binding, the hydrophobic residues at Site II were mutated to Alanine residues (Figure 4B).

In mutant simulation, we observed significant changes compared to wild-type simulation. Figure 4C shows that the cholesterol does not bind at the N-terminal site, thereby limiting the mobility of C-terminal. Subsequently, the non-specific binding of cholesterol does not lead to N-terminal opening. These studies further provide evidence that N-terminal binding is critical for selectively recognizing cholesterol.

We also explored the evolutionary importance of lipid binding sites. For this, we performed multiple sequence alignment with ApoA1 sequences from 37 species and found that the cholesterol binding sites (Site I and Site II) are highly conserved in ApoA1 across species. Fig S5 shows the conservation scores mapped onto the ApoA1 structure and we found that several residues that are highly conserved are directly involved in lipid binding. This gives direct evolutionary evidence towards the conservation of these residues to maintain the biological function of the protein.

Inferring stable protein elements within N-terminal

There is compelling evidence to suggest that N-terminal is critical for maintaining the stability of the protein in solution.[26] We reasoned that the binding specificity of cholesterol followed by N-terminal opening could be exploited to reveal quantitative aspects of stable protein elements. To assure maximum sampling available to the system, we leveraged above mentioned simulation for selecting six conformations from various time points of the simulation.

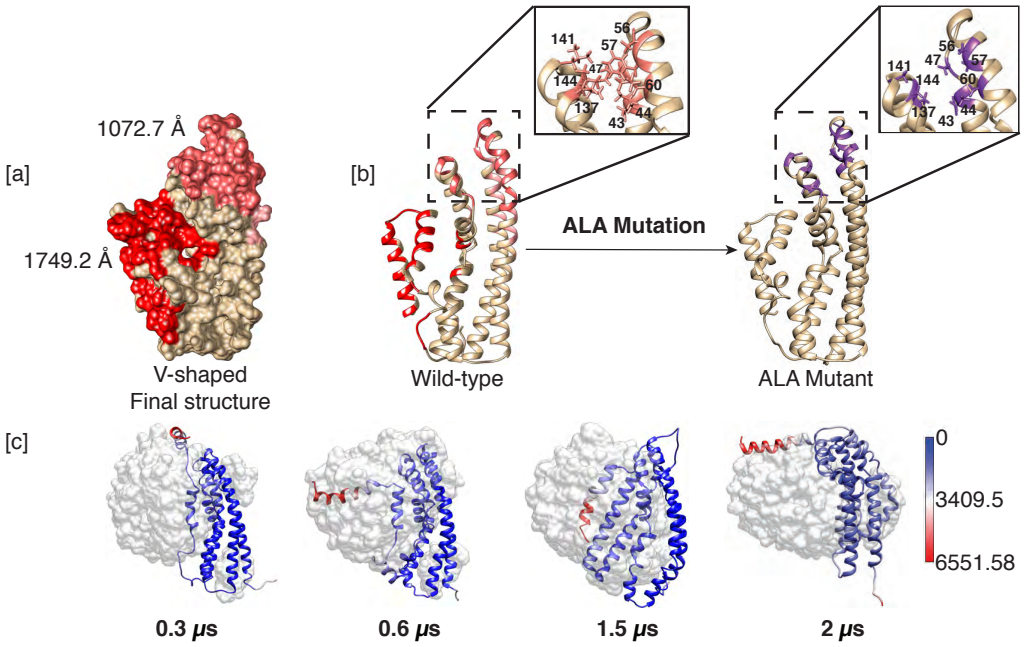

Figure 4: **Mutational analysis to validate the role of N-terminal cholesterol binding.** A) (left) Final structure of ApoA1 lipid-free simulation, showing increased exposed hydrophobic patches. The major patch, includes the residues from a large C-terminal region along with few residues from N-terminal. The color from dark to light shades of red corresponds to hydrophobic surface area. The data values for are 1749.2, 1072.7, 531.1 Å^2 and data values for starting structure are: 1349.5, 655.4, 581 Å^2. The hydrophobic patches were analyzed using Quilt software.[19] B) Hydrophobic and polar residues (43,44,47,56,57,60,137,141,144) residing at the top of the N-terminal domain (left) were mutated into alanine residues (right). C) Representative snapshots as a function of time showed limited conformational diversity in N-terminal region. The residue-based fluctuations are mapped onto the structural variations and the corresponding colour bar denotes the rmsf values in angstrom.

Each of these six conformations served as starting structure for further simulation (upto 250 ns each). Six replicates of each were simulated to obtain a sizeable pool of conformations (Figure 5A).

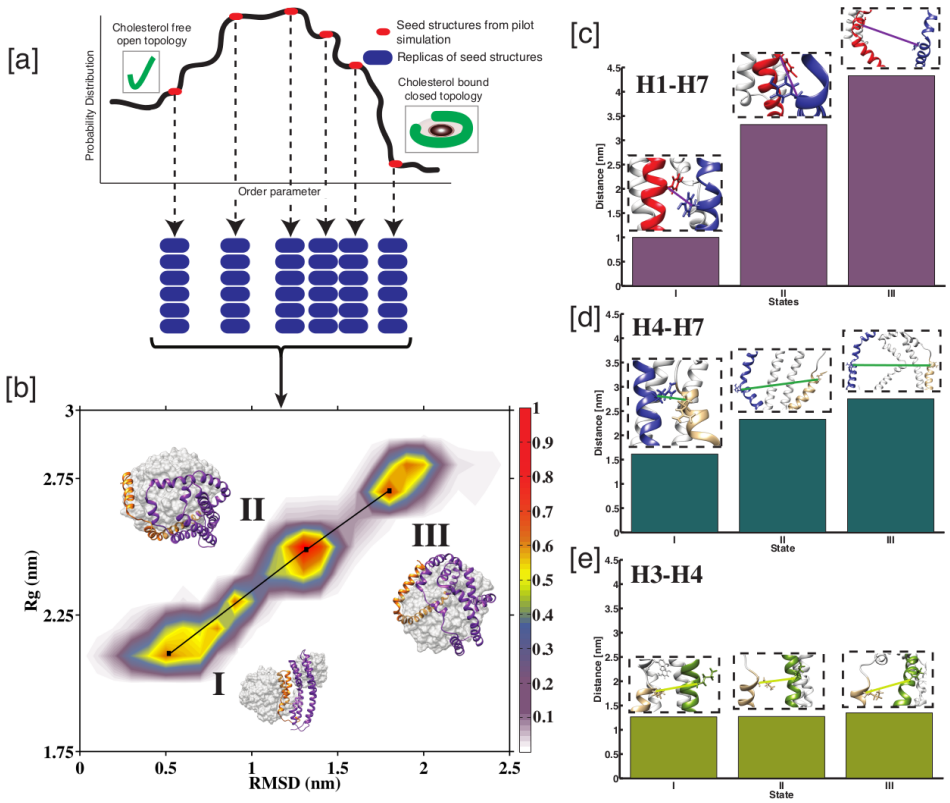

Figure 5: **Mapping for conformational space of ApoA1 during closed to open structural transition.** A) Schematic illustration depicting the strategy of enhanced sampling from initial pilot simulation. In total, six replicas were started from each seed structures, to obtain increased sampling. See methods for details. B) The population density map of conformational space is shown as a contour map projected onto two order parameters, RMSD and Rg. Clearly, three major states (State I-III) are populated in the map, whereby multiple transitions of closed to open topology were observed. C) Distances between helices H1-H7 (calculated between residues 29 and 166) and D) H4-H7 (calculated between residues 116 and 171) depict the expansion of the N-terminal domain; while E) H3-H4 (calculated between residues 77 and 118) helix pair distance was found to be constant in all three states. The bars represent the mean of the population for each state.

We found that there is a clear distribution of population density in the contour map plotted as a function of two order parameters, namely, RMSD and Rg (Figure 5B). Three states are predominantly populating the conformational landscape state I, II and III (Figure 5B). To quantify the population distribution, we determined the properties of structural ensemble in each of the three states. Clearly, the state I and III represent closed and open topology, respectively. Intriguingly, state II was found to be one of the major cluster. Multiple interconversion of states, $1 \leftrightarrow 2$, $2 \leftrightarrow 3$ were observed in our simulations while $1 \leftrightarrow 3$ is never observed (Fig S6). During transition from state I to III, a common state II was observed in all instances. In addition, there were multiple fast- and slow-transitions between respective states: 41% (I $\rightarrow$ II), 40% (I $\leftarrow$ II), 6% (II $\rightarrow$ III) and 11% (III $\leftarrow$ II). However, we never observed I $\leftrightarrow$ III transition, indicating that State II is an obligatory "intermediate state" in reaching the final open conformation.

Specific pair of helices showed that N-terminal is partially open in state II (Figure 5C-D). This was an interesting observation since it has been previously reported that the entire N-terminal stabilise the ApoA1 protein. However, our study indicates that specific pairs of N-terminal helices (H1-H7 and H4-H7) are prone to opening up. On the other hand, helix H3 and H4 in the N-terminal were found to be tightly packed, indicating that these helices are the stable regions prior to complete opening of ApoA1 topology.

ApoA1 engages with cholesterol in a concentration-dependent manner.

In all the simulations mentioned above, we considered the binding of 99 cholesterol molecules to ApoA1 monomer, since experimentally it has been shown earlier that 190-200 cholesterol molecules bind to ApoA1 dimer order.[27] Given the fact that the N-terminal binding of cholesterol is important in the subsequent opening up of the closed conformation of ApoA1, we checked if the opening up is stoichiometrically dependent. We systematically explored this by conducting additional molecular dynamic experiments with two concentrations of cholesterol, 50:1 and 75:1, apart from the previous 99:1 (referred to as 50CHL, 75CHL and 99CHL respectively). The time evolution of simulations using the three cholesterol concentrations of cholesterol are shown in Figure 6A. The representative snapshots showed clear distribution of cholesterol to our previously proposed binding sites, Site I and Site II in all simulations. A quantitative analysis of residue-wise binding of cholesterol with protein was calculated (Fig S7).

Surprisingly, the preferential binding does not lead to the opening of

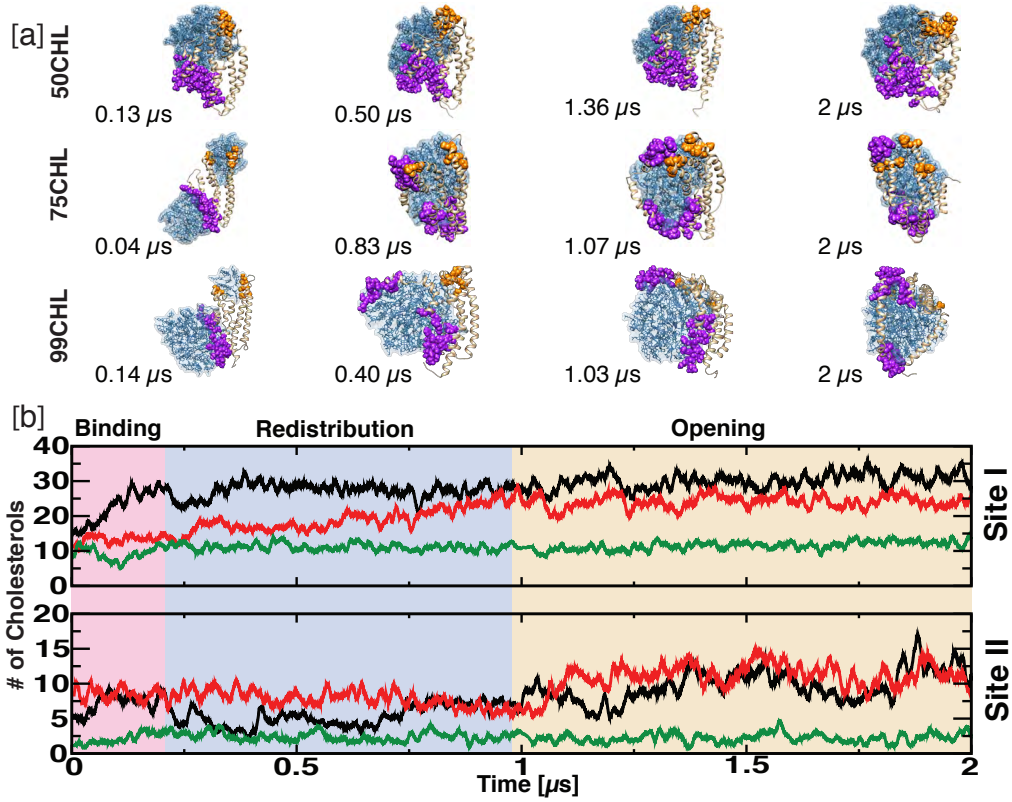

Figure 6: **Cholesterol concentration dependent structural changes** A) Representative snapshots as a function of time depicting the conformational transitions for three simulations, namely with 50, 75 and 99 cholesterol molecules. The cholesterol is shown in blue surface representation while the protein is coloured according to the Site I and II defined in Figure 3C. B) The kinetics of cholesterol binding is shown to occur in three stages in a concentration-dependent manner, namely, binding, redistribution, and opening. The 99CHL, 75CHL and 50CHL simulations are represented in black, red and green colour, respectively.

ApoA1 structure in 50CHL and 75CHL trajectories. We then further assessed the mode of binding and associated structural changes. We found that binding at preferential site is followed by a redistribution phase (Figure 6C). A characteristic feature emerges where the cholesterol molecules merge from both sites within 1 μs, in all simulations. The protein structure partially opens up with homogeneous lipid binding, however, N-terminal helical bundle remains intact in lower concentrations.

We further evaluated the biological significance of tight N-terminal packing in the presence of lower concentration cholesterol simulations. As shown before, the C-terminal domain was highly mobile during lipid engulfing (Figure 3A). The merging of the lipids is followed by active redistribution of the C-terminal residues, where it dynamically re-arranges to envelope the lipids with the residues displaying maximum fluctuation. This is accompanied by concomitant N-terminal opening of all helices, where the distance between parallel helix increases as shown in Figure 7A-B. While the C-terminal dynamics and preferential binding is similar in all instances, N-terminal helical bundle opening is highly dependent on the number of cholesterol molecules.

C-terminal dynamics predominantly dictates lipid-free and lipidated ApoA1 structural transition.

We then attempted to evaluate the changes in secondary structure of ApoA1 upon cholesterol addition and how secondary structure preferences of different segments of ApoA1 vary dynamically in lipid-free and cholesterol-associated simulations. In order to probe these differences, we computed the probability distribution for the secondary structural elements in both lipid-free and cholesterol associated simulations(Figure 8A). We found that the α-helical propensity of the N-terminal region (3-40, 48-90, 95-112, 118-124, 130-178) is highly conserved in both lipid-free and lipidated simulations. Furthermore, there is previous evidence suggesting that the N-terminal of the ApoA1 protein is relatively stable.[28] In contrast to N-terminal helicity, C-terminal region was found to be highly unstructured in lipid-free, with residues 183−186, 189, and 216−217 found to be mostly in loop conformations. These findings largely agree with previous low resolution structural data where the C-terminal has been suggested to exist in a combination of helical and coil structure.[7,11,28] To reveal how C-terminal residues drastically change the helix conformation, we looked for transition window in the narrowed residues within C-terminal (Figure 8B). In particular, in the presence of cholesterol, residues (182-190,198-199) dynamically interconvert between coil and helix (Figure 8C). Our data indicates that one of the factors driv-

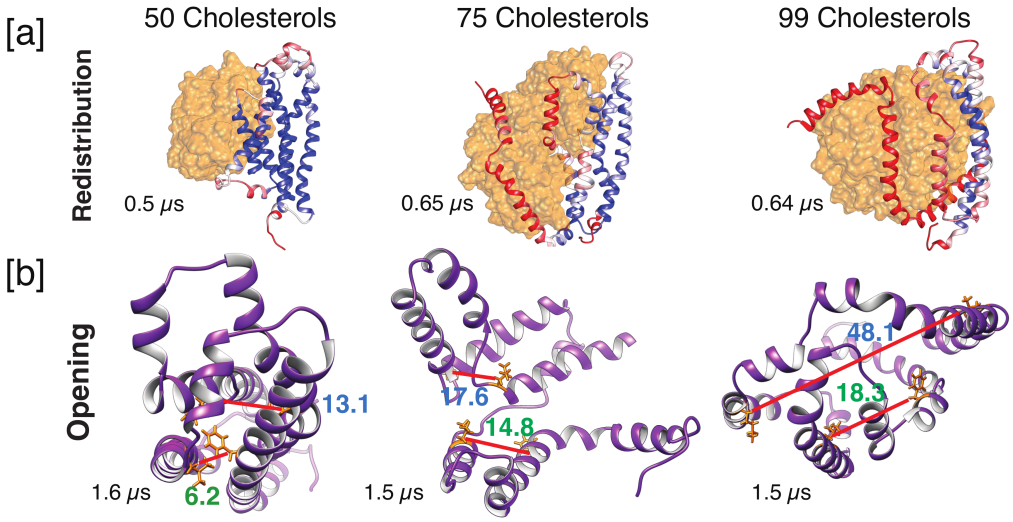

Figure 7: **Redistribution of cholesterol and opening of N-terminal domain.** A) The residue-based fluctuations of 50CHL, 75CHL, and 99CHL simulations to represent the protein mobility (red=high;white=medium;blue=low). The merging of the cholesterol is actively promoted by dynamical C-terminal residues (in red) in higher concentration simulation. B) The distance of residue pair (118-119; 75-33) were calculated to quantify the opening of helix bundle. The side chain of the residues and the corresponding distance is also highlighted.

ing lipid-free and lipid-associated ApoA1 structural transitions is increased C-terminal helicity (Fig S8).

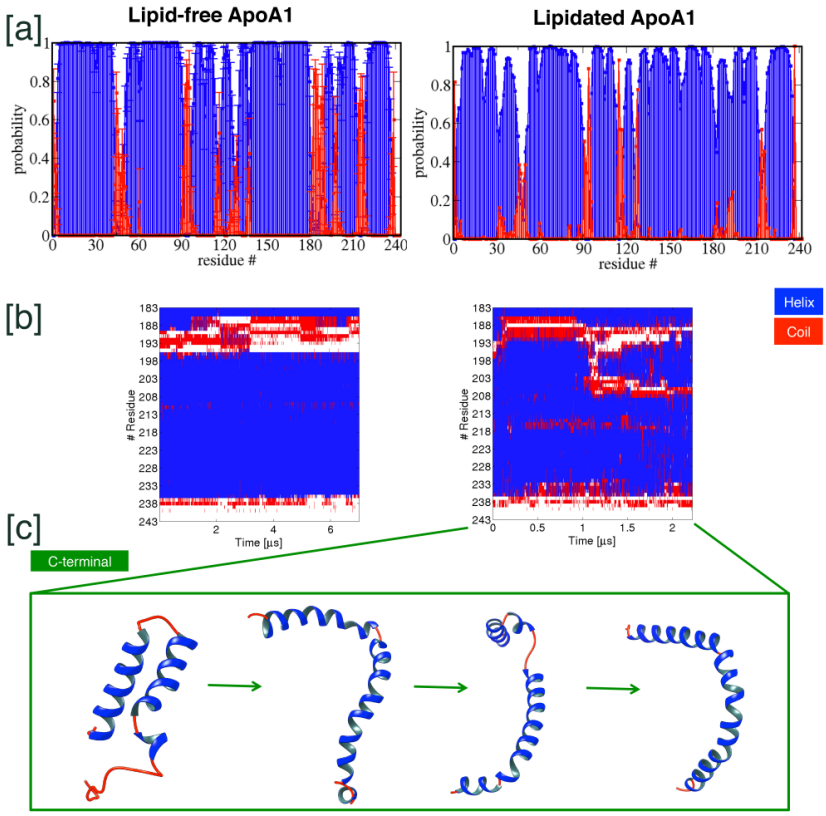

Figure 8: **Secondary structure characterization of the lipid-free and lipidated simulations.**A) Probability distribution for helix and coil secondary structure elements. The probability range is calculated on the basis of residue-based residence time, where 1 indicates the formation of helix/coil secondary structure throughout the simulation. The error bars denote the standard deviation computed from five independent trajectories for the native structure's simulations (left) and pilot lipidated simulation with ApoA1 (right). B) Time evolution of secondary structure change in the C-terminal for the representative lipid-free (left) and the lipidated simulation (right). C) Conformational changes in the C-terminal region, depicting the backbone variation across the simulation length, indicating spontaneous formation of α-helix.

Discussion

The apolipoprotein ApoA1 governs numerous functional interactions during the complex assembly of high-density lipoprotein (HDL) particles,[1,29] but the initial mechanistic events leading to the formation and interconversion of diverse monomeric ApoA1 states have remained unknown. Four major configurations have been proposed in a recent review,[29] (i) Lipid-free monomeric ApoA1, (ii) Lipid-poor ApoA1 monomer, also known as pre-β HDL, (iii) Oligomeric ApoA1 in discoidal HDL (d-HDL); and (iv) Oligomeric ApoA1 in the final matured spherical HDL (s-HDL). Furthermore, there is evidence suggesting that the ApoA1 exhibits highly diverse conformations in all the above states.

Here, in this work we focussed on a particular aspect - initial steps of cholesterol uptake pathway i.e., the structure and dynamics of ApoA1 monomeric protein and its dependence on cholesterol. We constructed a full length ApoA1 monomer structure from the truncated N-terminal mouse structure and modelled the C-terminal based on loop modeling and further refinement by MD simulations. The starting structure was subjected to long time-scale five independent molecular dynamics trajectories, with an aggregate simulation time of 35 μs. Remarkably, all the five trajectories converge to a distinct closed structure, displaying a distinct N- and C-terminal proximity (Figure 2). This allowed us to examine the structural properties of ApoA1-lipid free structure with great statistical reliability. Significant increase in surface exposed hydrophobicity was observed in the final ensemble, with C-terminal contributing towards the largest hydrophobic patch. These findings agree with previous observations that the phospholipids probably binds at the hydrophobic pocket formed by the hydrophobic C-terminal domain and the helical bundle N-terminal domain.[30,31]

While the lipid-free ApoA1 monomer simulated here showed helix-bundled topology, cholesterol-bound ApoA1 structures have been reported to exist as open belt-shaped topology,[32–34] providing clear evidence for distinctive structure-function relationship. We performed cholesterol bound ApoA1 MD simulations, where indeed the transition of ApoA1 structure occurs upon cholesterol binding. This is attributed to highly flexible C-terminal domain and preponderance of two cholesterol binding sites on the ApoA1 structure referred to as Site I and Site II (Figure 3). In accord, previous investigations have found a strong association between conformational changes and functional binding sites.[35] Previous reports also suggest that the C-terminal α-helix has the ability to mediate lipid efflux[8] and solubilize membrane phospholipids.[9,10]

Interestingly, we found novel binding site at the N-terminal. The specific

residues at this site, is highly correlated with the contiguous hydrophobic surface area. Mutation of these regions showed closed structure, suggesting that N-terminal binding site may have direct functional relevance (Figure 4).

Thus so far, our detailed kinetic analysis of the cholesterol-dependent protein changes revealed spontaneous formation of closed to open structures. These dynamic differences in topology is finely modulated by molecular recognition events governed upon cholesterol binding. The question is how is this transition achieved and what are the stable protein elements during this large-scale conformational transition. In order to investigate this pathway, we initiated multiple trajectories from the seed structures obtained from the pilot 2 μs MD trajectory. The structural characteristics of the seed structures were drastically different, allowing us to sample large conformational space of cholesterol induced opening of ApoA1 structure (Figure 5A). This framework allowed us to obtain three populated states. Interestingly, state II was most populated with partial N-terminal opening. Further analysis suggested that specific pair of helices (H3-H4) contributed to N-terminal stability.

Earlier studies suggest that ApoA1 adopts different conformations depending on the extent of lipidation.[1,36] Our results showed that systematically changing the number of cholesterol molecules in ApoA1 monomeric simulation leads to correlated changes in the protein structure (Figure 6 and 7). In particular, increasing the number of cholesterol leads to high mobility at the C-terminal domain leading to concomitant concentration-dependent opening of the N-terminal helices. Interestingly, the binding of the cholesterol occurred at identical locations, demonstrating the inherent propensity of proposed Site I and Site II to target cholesterol. Residues such as 40, 59, 220−243 etc. have been experimentally found to have higher propensity for lipid binding.[11,22,25,37–40] Our simulations are in excellent match with all these previous results and shows new regions of lipid-binding initiation in concentration gradient manner.

In this study, we also established the location of α-helical segments in two monomeric ApoA1 states (Figure 8). Detailed comparison of lipid-free and lipidated simulations showed that N-terminal is highly helical, while C-terminal segments are a mix of helical and unstructured regions. A significant secondary structural changes bore a striking resemblance to previous experimental reports.[11,39–42] In particular, residues 14-19, 26-31, 55-64, 70-77, 83-85, 93-97, 102-115, 147-148, 158-178, 224-227. Region 115-150 was found to sensitive to proteolysis and hence thought to be flexible and unstructured.[5,43] We also observed several loops, which could impart flexibility to the region. In C-terminal, there is a sharp decrease in helical propensity from 181-192, 196-200, 212-221 while regions like 183-186, 189, 216-217, with high propensity of coil conformation. In particular, helical region 224-231,

proposed to have the highest lipid-binding property matches very well with previous experimental findings.[11,25,39,40]

Conclusions

Intense interest in cardiovascular diseases has spurred major interest in reverse cholesterol pathway (RCT). According to recent studies, novel markers have been proposed to capture the early signs of atherosclerosis.[44,45] Studying ApoA1 structural characteristics is challenging, because ApoA1 is highly dynamic with cholesterol molecules. While a dual topological organisation upon cholesterol binding is reported, there are no reports on dynamic transition between lipid-free and lipid-associated states. Thus the goal of the work was to understand the underlying molecular mechanism of ApoA1 cholesterol induced structural alteration. The protein domains (N- and C-terminal) elements of ApoA1 are predominant regions that are highly conserved across various species. We show that the lipid-free exhibits compact shape, with N- and C-terminal juxtaposed in closed topology.

Interestingly, upon cholesterol addition ApoA1 monomer spontaneously forms open-topology that engulfs cholesterol in a belt shape. We show that dynamics of protein domains is highly specific, indicating its link to distinct biological function. While, the role of C-terminal is already established in previous reports,[24,25] our simulations reveal a novel binding site at the N-terminal. Mutation at this site abrogates ApoA1 opening. Furthermore, conformational map of structural transitions confirmed that specific helices in the N-terminal largely contribute to protein stability. Thus, deciphering the mechanisms of ApoA1 induced cholesterol recognition, a major component in RCT, will provide the means to significantly advance our understanding in maintaining cardiovascular health.

Acknowledgements

AR acknowledges the financial assistance from Council of Scientific and Industrial Research (CSIR), Ministry of Science and Technology, Govt of India, India under the CARDIOMED-BSC0122 project. LT is funded by INSPIRE Faculty Fellowship from Department of Science and Technology.

References

[1] Vedhachalam, C.; Chetty, P. S.; Nickel, M.; Dhanasekaran, P.; Lund-Katz, S.; Rothblat, G. H.; Phillips, M. C. *Journal of Biological Chemistry* **2010**, *285*, 31965–31973.

[2] Assmann, G.; Brewer, H. B. *Proceedings of the National Academy of Sciences* **1974**, *71*, 1534–1538.

[3] Gursky, O. *Journal of molecular biology* **2013**, *425*, 1–16.

[4] Borja, M. S.; Zhao, L.; Hammerson, B.; Tang, C.; Yang, R.; Carson, N.; Fernando, G.; Liu, X.; Budamagunta, M. S.; Genest, J.; Shearer, G. C.; Duclos, F.; Oda, M. N. *PloS one* **2013**, *8*, e71541.

[5] Roberts, L. M.; Ray, M. J.; Shih, T.-W.; Hayden, E.; Reader, M. M.; Brouillette, C. G. *Biochemistry* **1997**, *36*, 7615–7624.

[6] Davidson, W. S.; Hazlett, T.; Mantulin, W. W.; Jonas, A. *Proceedings of the National Academy of Sciences* **1996**, *93*, 13605–13610.

[7] Saito, H.; Dhanasekaran, P.; Nguyen, D.; Holvoet, P.; Lund-Katz, S.; Phillips, M. C. *Journal of Biological Chemistry* **2003**, *278*, 23227–23232.

[8] Vedhachalam, C.; Liu, L.; Nickel, M.; Dhanasekaran, P.; Anantharamaiah, G.; Lund-Katz, S.; Rothblat, G. H.; Phillips, M. C. *Journal of Biological Chemistry* **2004**, *279*, 49931–49939.

[9] Vedhachalam, C.; Duong, P. T.; Nickel, M.; Nguyen, D.; Dhanasekaran, P.; Saito, H.; Rothblat, G. H.; Lund-Katz, S.; Phillips, M. C. *Journal of Biological Chemistry* **2007**, *282*, 25123–25130.

[10] Saito, H.; Dhanasekaran, P.; Nguyen, D.; Deridder, E.; Holvoet, P.; Lund-Katz, S.; Phillips, M. C. *Journal of Biological Chemistry* **2004**, *279*, 20974–20981.

[11] Oda, M. N.; Forte, T. M.; Ryan, R. O.; Voss, J. C. *Nature Structural & Molecular Biology* **2003**, *10*, 455–460.

[12] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *Journal of computational chemistry* **2005**, *26*, 1701–1718.

[13] Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. *Journal of computational chemistry* **2009**, *30*, 1545–1614.

[14] Huang, J.; MacKerell, A. D. *Journal of computational chemistry* **2013**, *34*, 2135–2145.

[15] Denning, E. J.; Priyakumar, U.; Nilsson, L.; Mackerell, A. D. *Journal of computational chemistry* **2011**, *32*, 1929–1943.

[16] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of chemical physics* **1983**, *79*, 926–935.

[17] Darden, T.; York, D.; Pedersen, L. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

[18] Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. *Journal of computational chemistry* **1997**, *18*, 1463–1472.

[19] Lijnzaad, P.; Berendsen, H. J.; Argos, P. *Proteins Structure Function and Bioinformatics* **1996**, 192–203.

[20] Glaser, F.; Pupko, T.; Paz, I.; Bell, R. E.; Bechor-Shental, D.; Martz, E.; Ben-Tal, N. *Bioinformatics* **2003**, *19*, 163–164.

[21] Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. *Nucleic acids research* **2010**, gkq399.

[22] Pollard, R. D.; Fulp, B.; Samuel, M. P.; Sorci-Thomas, M. G.; Thomas, M. J. *Biochemistry* **2013**, *52*, 9470–9481.

[23] Lagerstedt, J. O.; Budamagunta, M. S.; Oda, M. N.; Voss, J. C. *Journal of Biological Chemistry* **2007**, *282*, 9143–9149.

[24] Holvoet, P.; Zhao, Z.; Vanloo, B.; Vos, R.; Deridder, E.; Dhoest, A.; Taveirne, J.; Brouwers, E.; Demarsin, E.; Engelborghs, Y. *Biochemistry* **1995**, *34*, 13334–13342.

[25] Palgunachari, M. N.; Mishra, V. K.; Lund-Katz, S.; Phillips, M. C.; Adeyeye, S. O.; Alluri, S.; Anantharamaiah, G.; Segrest, J. P. *Arteriosclerosis, thrombosis, and vascular biology* **1996**, *16*, 328–338.

[26] Davidson, W. S.; Arnvig-McGuire, K.; Kennedy, A.; Kosman, J.; Hazlett, T. L.; Jonas, A. *Biochemistry* **1999**, *38*, 14387–14395.

[27] Jonas, A. *Methods in enzymology* **1986**, *128*, 553.

[28] Saito, H.; Lund-Katz, S.; Phillips, M. C. *Progress in lipid research* **2004**, *43*, 350–380.

[29] Gogonea, V. *Frontiers in pharmacology* **2015**, *6*, doi: 10.3389/fphar.2015.00318.

[30] Duong, P. T.; Weibel, G. L.; Lund-Katz, S.; Rothblat, G. H.; Phillips, M. C. *Journal of lipid research* **2008**, *49*, 1006–1014.

[31] Safi, W.; Maiorano, J. N.; Davidson, W. S. *Journal of lipid research* **2001**, *42*, 864–872.

[32] Borhani, D. W.; Rogers, D. P.; Engler, J. A.; Brouillette, C. G. *Proceedings of the National Academy of Sciences* **1997**, *94*, 12291–12296.

[33] Rogers, D. P.; Roberts, L. M.; Lebowitz, J.; Datta, G.; Anantharamaiah, G.; Engler, J. A.; Brouillette, C. G. *Biochemistry* **1998**, *37*, 11714–11725.

[34] Segrest, J. P.; Jones, M. K.; Klon, A. E.; Sheldahl, C. J.; Hellinger, M.; De Loof, H.; Harvey, S. C. *Journal of Biological Chemistry* **1999**, *274*, 31755–31758.

[35] Marsh, J. A.; Teichmann, S. A. *Structure* **2011**, *19*, 859–867.

[36] Siuda, I.; Tieleman, D. P. *Journal of chemical theory and computation* **2015**, *11*, 4923–4932.

[37] Frank, P. G.; Marcel, Y. L. *Journal of lipid research* **2000**, *41*, 853–872.

[38] Zhang, X.; Lei, D.; Zhang, L.; Rames, M.; Zhang, S. *PloS one* **2015**, *10*.

[39] Tricerri, M. A.; Behling Agree, A. K.; Sanchez, S. A.; Jonas, A. *Biochemistry* **2000**, *39*, 14682–14691.

[40] Lagerstedt, J. O.; Budamagunta, M. S.; Liu, G. S.; DeValle, N. C.; Voss, J. C.; Oda, M. N. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **2012**, *1821*, 448–455.

[41] Chetty, P. S.; Mayne, L.; Lund-Katz, S.; Stranz, D.; Englander, S. W.; Phillips, M. C. *Proceedings of the National Academy of Sciences* **2009**, *106*, 19005–19010.

[42] Okon, M.; Frank, P. G.; Marcel, Y. L.; Cushley, R. J. *FEBS letters* **2001**, *487*, 390–396.

[43] Ren, X.; Sivashanmugam, A.; Chen, B.; Li, C.; Yang, Y.; Neville, T.; Hoyt, D.; Jerome, W. G.; Sparks, D.; Li, Q. *Functional and Structural Biology on the Lipo-Network, eds. K. Morikawa and S. Tate, Transworld Research Network, Trivandrum, India* **2006**, 65.

[44] Basak, T.; Varshney, S.; Akhtar, S.; Sengupta, S. *Journal of proteomics* **2015**, *127*, 50–60.

[45] Basak, T.; Varshney, S.; Hamid, Z.; Ghosh, S.; Seth, S.; Sengupta, S. *Journal of proteomics* **2015**, *127*, 169–177.