**Specific cholesterol binding drives drastic structural alterations in Apolipoprotein A1**

Arjun Ray[1,2], Asmita Ghosh[1,2], Rahul Chakraborty[1,2], Santosh Kumar Upadhyay[1], Souvik Maiti[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**

Protein adopts multitude of flexible and rapidly interconverting conformers, many which are governed by specific protein-interaction domains. ApoA1, a key player involved in high-density lipoprotein (HDL) regulation exists in structurally diverse forms with varying degree of cholesterol association, and each state is associated with different functional properties. While disc-shaped HDL and its oligomeric ApoA1 protein components have been the focus of several investigations, structural properties of monomeric ApoA1 are poorly understood. Here, we undertook large-scale structural analysis of ApoA1 in apo and cholesterol-bound forms using tens of independent simulations with total computing time exceeding 50 $\mu$s. Examination of multiple lipid-free trajectories of monomeric ApoA1 revealed a common conformation, with distinct spatial proximity between N- and C-terminal domains. With incorporation of physiologically known cholesterol concentration ($\approx$100 cholesterol molecules) in ApoA1 simulations, the monomeric protein spontaneously formed an open circular topology. Remarkably, these drastic structural perturbations are driven by specific binding site at C-terminal and a novel cholesterol binding site at the N-terminal. We proposed a mechanism of stage-wise opening of ApoA1 and demonstrated that less cholesterol concentration around interaction sites or mutation within N-terminal binding sites does not lead to open bell-shaped topology. The kinetic barriers between open- and closed-states also showed an ensemble of loosely packed helix bundle (H1-H7; H4-H7) that posed as a slow-intermediate step. Lastly we performed complementary experiments, including ITC and CD measurements to confirm that structural changes are induced by ligand association and not driven by random hydrophobic effect. Collectively, our study suggests a previously unknown mechanism of cholesterol sequestering by ApoA1 that could directly aid in developing modulators for cholesterol efflux with chronic cardiovascular diseases.

# Introduction

The removal of cholesterol in humans is universally operated by reverse cholesterol pathway (RCT), that tightly controls lipid homeostasis via the transfer of cholesterol from peripheral cells to the liver.[1–3] A pivotal step in the formation of high-density lipoprotein (HDL) complex involves sequestering of cholesterol by apolipoproteins. Apolipoprotein A1 also commonly referred to as ApoA1, is the major protein constituent of HDL ($\approx$ 70%).[4] Although a relatively small protein of 243 amino acids, ApoA1 exhibits remarkable structural variability with varying degrees of cholesterol association[5](Figure 1A). These distinct structural states, range from lipid-free monomeric conformation in blood plasma cells,[6] to low-concentration cholesterol-bound monomeric conformation,[7,8] and higher oligomeric forms involved in belt-shaped HDL supramolecular assemblies.[9–12] Owing to the existence of this dynamic instability, attempts towards determining crystal structure of monomeric ApoA1 in lipid-free conditions have remained unsuccessful. However, experimentally determined oligomeric forms of ApoA1 are known.[13,14] Thus, given the association of elevated cholesterol with chronic cardiovascular diseases, understanding the structural basis of ApoA1 functional diversity is a long-standing open question in the field.

Although direct crystallographic evidence is lacking for monomeric ApoA1 protein, there exist several reports highlighting aspects of ApoA1 structure and function. The ApoA1 structure, suggested by proteolysis and mutagenesis studies is divided into two structural domains i.e., N- and C- terminal that are highly conserved across species (Figure 1B). Interestingly, the C-terminal -helix is known to directly mediate lipid efflux and also known to solubilize the membrane phospholipids. Early EPR studies[15,16] have also characterised the secondary structure segments of lipid-free ApoA1. A recent hydrogen exchange (HDX) study suggested the N-terminal containing stabilizing interactions in the amphipatic helical bundle and a disordered C-terminal.[17] Similar HDX study by Wu Z. et al. also presented a disordered C-terminal.[18] A putative close proximity of interaction between the N- and C- termini is also proposed.[19] Further the degree of total helical content of the protein in apo- and bound forms has been shown to vary using several biophysical approaches.[20,21] Recently, a chimeric ApoA1 monomeric structure has also been proposed based on consensus approach. While these studies have characterised many aspects of monomeric structure, conclusive model and direct structural evidence on interconverting lipid-free and bound states is largely missing.

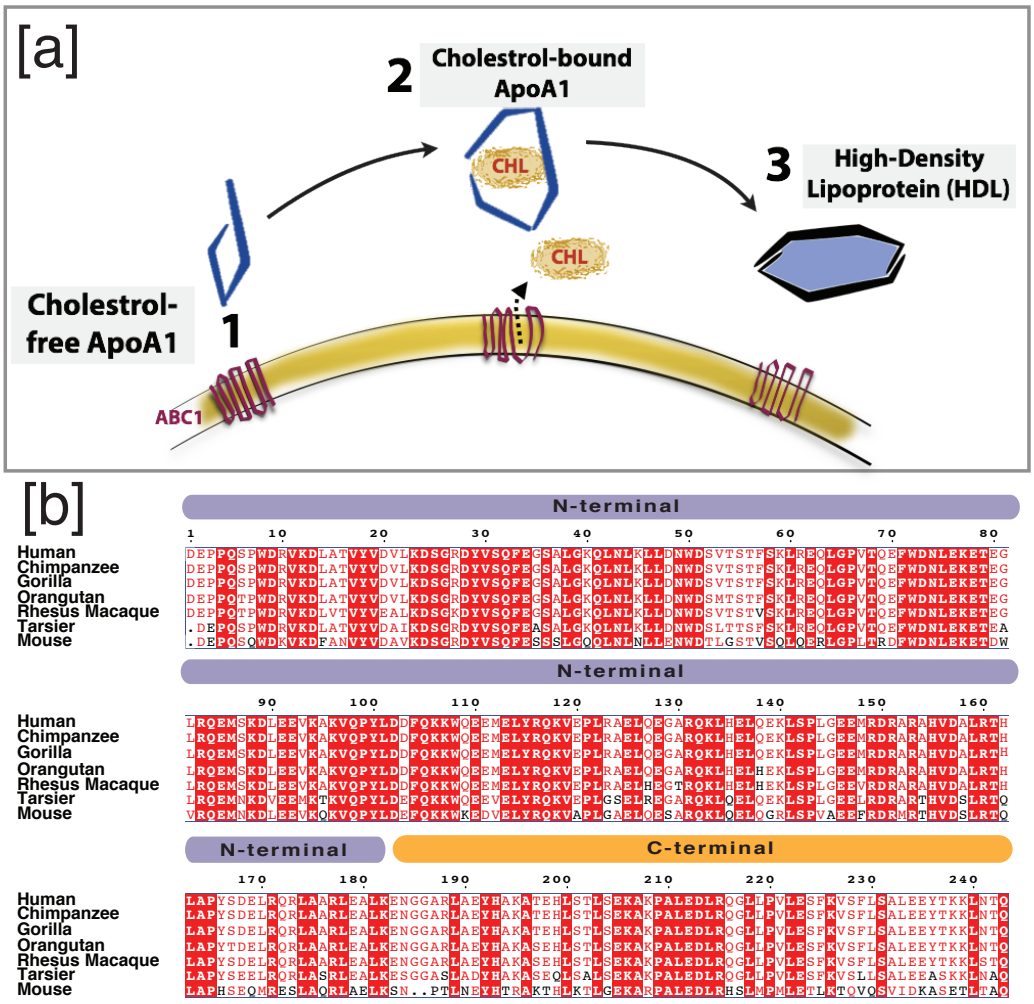


Figure 1: **The Biology of ApoA1 and Protein Domain Architecture.** A) Schematic diagram of reverse cholesterol pathway. The role of ApoA1 is highlighted in various stages. In the first step, ApoA1 exists in a monomer state with no association with cholesterol in blood plasma. The cholesterol is then 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.

In recent years, technological advancement has brought molecular dynamics (MD) simulations to $\mu$s-timescale. These developments have had a particular impact in biology, where experimental investigations of single molecule were prohibited by the fast-timescale dynamics, can now be studied. Previous computational reports have proposed a structure of lipid-free ApoA1 using iterative MD simulations. The model building was guided by each secondary structure information of the human $\Delta$1–43 4-mer ApoA1 crystal structure[13] and other previously reported experimental data.[22] Given the distance-restraint on N- and C-terminal topology based on earlier experiment, the short 150 ns MD simulation explored limited conformational landscape. In a different study, a chimeric ApoA1 structure was prepared. The model was built by repositioning key residues by bending and stretching the backbone to align and fitting the hydrophobic cleft between N- and C-terminal.[23] The final model was subjected to a very high-temperature MD simulation (500K) for 30 ns. They further performed cross-linking reactions and liquid chromatography- electrospray ionization mass spectrometry (LC-ESI-MS) to support the proposed residue contacts. Due to limited computation time, complete ApoA1 structural transitions have not yet been captured.

Here we employed multi-$\mu$s timescale MD simulations in explicit solvent to study cholesterol-free and -bound ApoA1 monomer dynamics at atomic scale. The 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. We also found the spontaneous formation of belt or open shaped ApoA1 topology upon cholesterol addition, as opposed to bundled or closed shaped topology in the lipid-free state. The analysis of trajectories also pinpoint molecular mechanism of structural opening and propose putative cholesterol binding sites. In particular, dynamic interaction of protein domain elements illustrated specific functional regions that act in concert to enable drastic conformational transitions between closed and open topology of the protein. Complementary experiments validated some of our findings, including ligand-induced structural changes and secondary structural content in varying degree of cholesterol concentrations. 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.

# Results

Our study was aimed at investigating the structural properties of ApoA1 upon it's interaction with cholesterol. This is undoubtedly the most important step in biogenesis of high-density lipoprotein (HDL) formation. Our approach was twofold. We probed ApoA1 protein structural 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 $\mu$s), leading to determination of structure and dynamics with statistical reliability. The mapped structural dynamics at $\mu$s timescale is three order magnitude longer than previous reports.[16,24]

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 $\mu$s, and remained stable for last 2 $\mu$s 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 Figure S2.

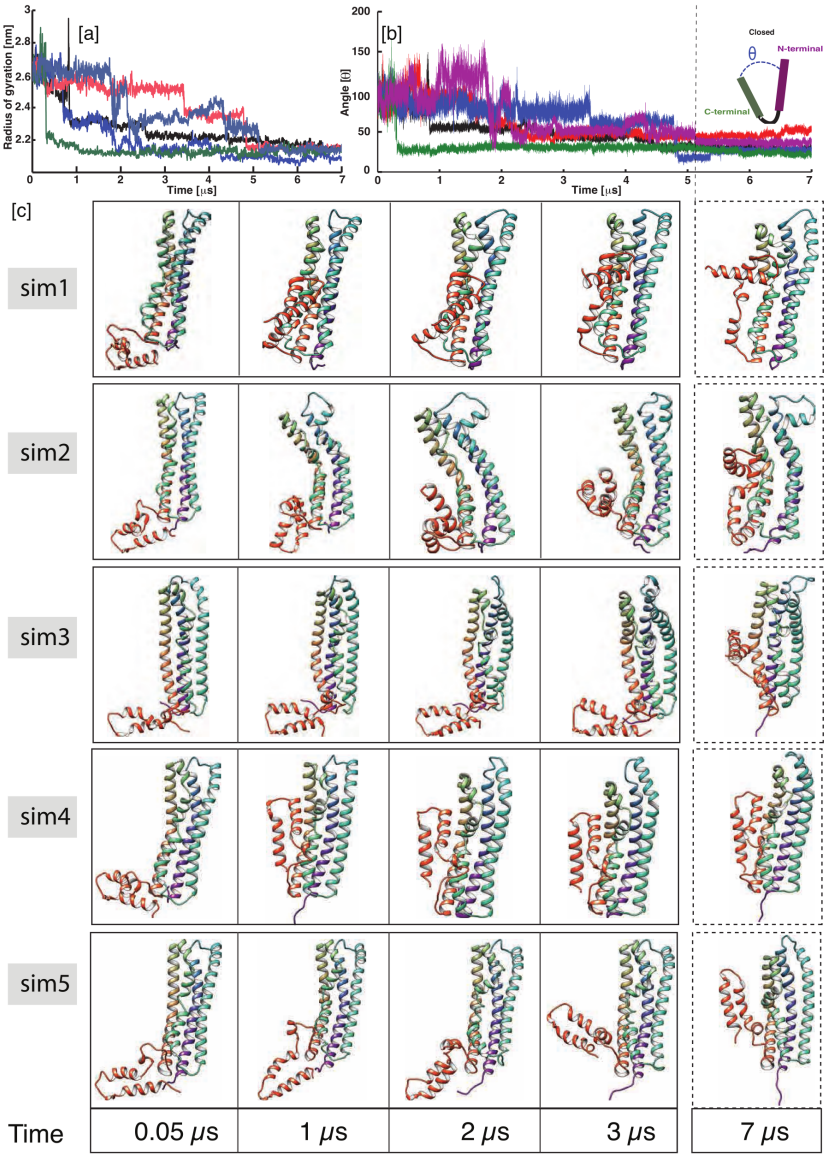


Figure 2: **ApoA1 Lipid-free Structure and Dynamics.** A) Time evolution of ApoA1 structure as a function of radius of gyration (Rg) for five independent molecular dynamics trajectories, shown in different colors. The Rg value shows drop in all simulations, indicating increase in the compactness of the structure. B) The evolution of angle between C- and N- terminal domain. The angle $\theta$ is defined by constructing the vector between residues 171, 182 and 214 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.

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.[15] In accord, we used similar physical probes to validate our data (Figure 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.

## 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.

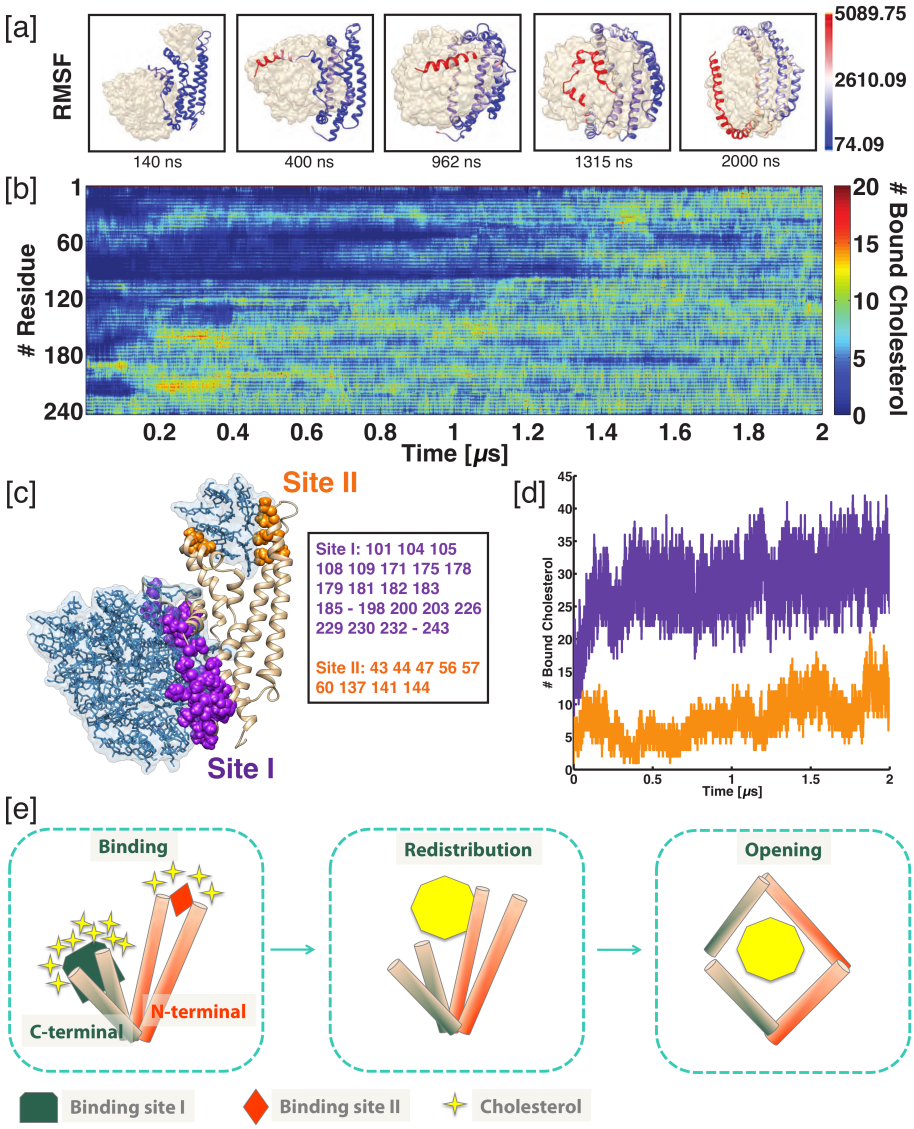


Figure 3: **Dual topological organization of ApoA1 upon cholesterol binding.** A) The representative snapshots are shown as a function of time for ApoA1-cholesterol simulations. The residue-based fluctuations (rmsf) are mapped onto the structural variations and the corresponding colour bar denotes the 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.

To quantitatively assess the interaction, we plotted the time occurrence of the cholesterol interaction with ApoA1 (Figure 3B). During the initial simulation time (≈140 ns), cholesterol was found to specifically target two binding sites. The C-terminal is the main site of cholesterol targeting while 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 by 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).

## 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,[25,26] 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).[27] 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 $Å^2$) (Figure 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).

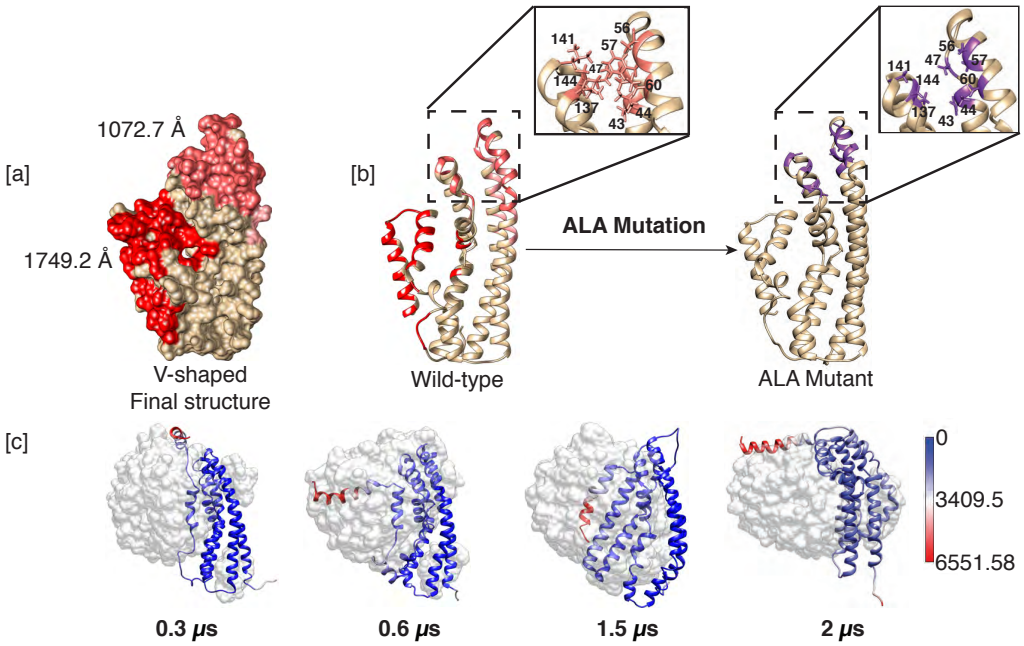


Figure 4: **Mutational analysis to validate the role of N-terminal cholesterol binding.** A) (left) Final structure of ApoA1 lipid-free simulation, showing increased exposed hydrophobic patches. The major patch includes the residues from a large C-terminal region along with few residues from N-terminal. 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.[27] B) Hydrophobic and polar residues (43, 44, 47, 56, 57, 60, 137, 141 and 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.

In mutant simulation, we observed significant changes compared to wild-type trajectory. Figure 4C shows that the cholesterol does not bind at the N-terminal site, thereby limiting the mobility of C-terminal and does not lead to N-terminal opening. This study provides further 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. Figure S5 shows the conservation scores mapped onto the ApoA1 structure and we found from our simulation studies that several of these highly conserved residues are directly involved in the initial lipid interaction. 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.[28] 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. 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).

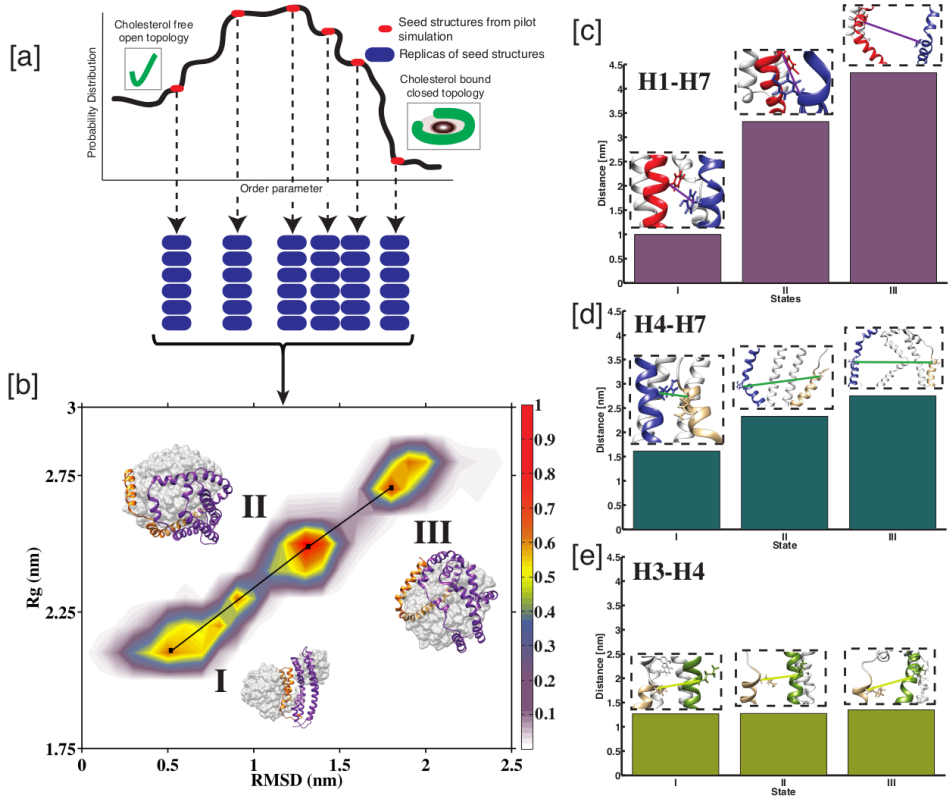


Figure 5: **Mapping for conformational space of ApoA1 during closed to open structural transition.** A) Schematic illustration depicting the strategy of enhanced sampling from initial pilot simulation. In total, six replicas were started from each seed structures, to obtain increased sampling. See methods 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 (I,II and III) predominantly populate the

conformational landscape (Figure 5B). To quantify the population distribution, we determined the properties of structural ensemble in each of the three states. Clearly, state I and III represent closed and open topology, respectively. Intriguingly, state II was found to be one of the major cluster and multiple interconversion of states, 1 $\leftrightarrow$ 2, 2 $\leftrightarrow$ 3 were observed in our simulations while 1 $\leftrightarrow$ 3 is never observed (Figure 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.[29] 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, in addition to 99:1 simulations described above (referred to as 50 cholesterol (50CHL), 75 cholesterol (75CHL) and 99 cholesterol (99CHL) henceforth). The time evolution of simulations using the three cholesterol concentrations 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 (Figure S7).

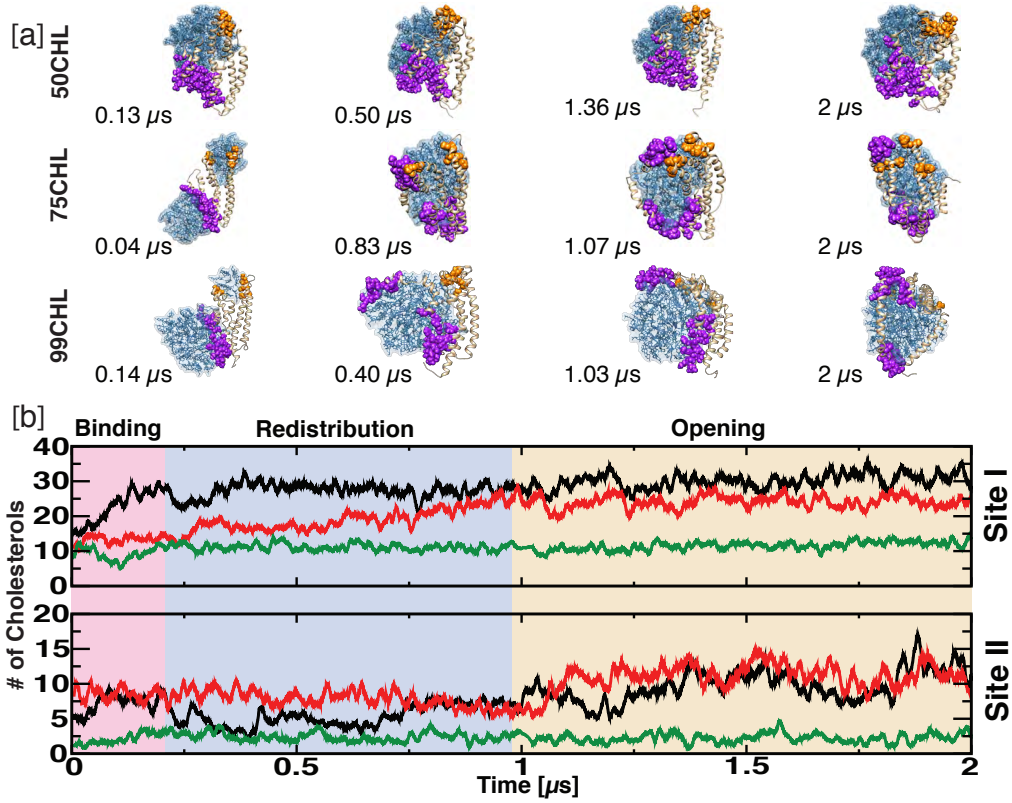


Figure 6: **Cholesterol concentration dependent structural changes** A) Representative snapshots as a function of time depicting the conformational transitions for three simulations, namely with 50, 75 and 99 cholesterol molecules. The cholesterol is shown in blue surface representation while the protein 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.

Surprisingly, the preferential binding does not lead to the opening of 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 6B). A characteristic feature emerges where the cholesterol molecules merge from both sites within 1 $\mu$s, 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.

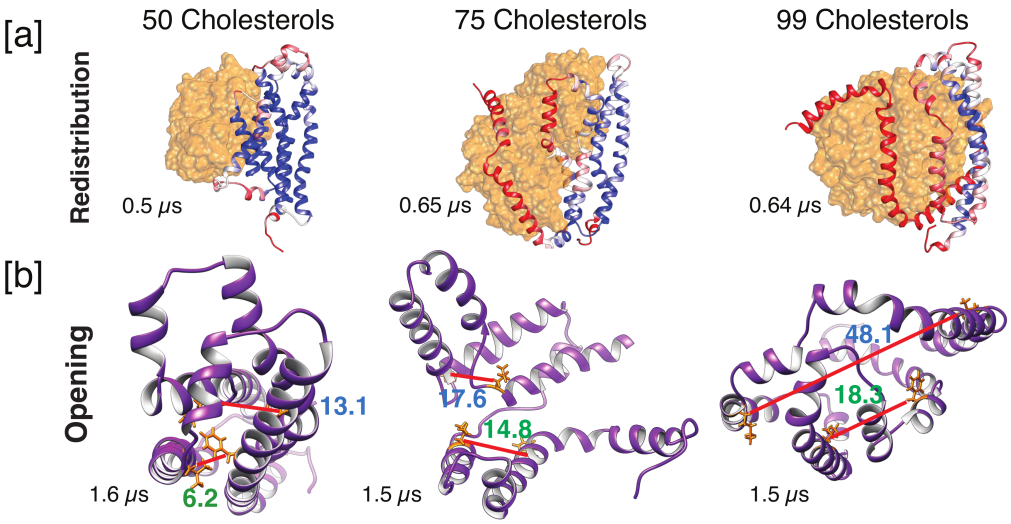


Figure 7: **Redistribution of cholesterol and opening of N-terminal domain.** A) The residue-based fluctuations of 50CHL, 75CHL, and 99CHL simulations to represent the protein mobility (red=high;white=medium;blue=low). The merging of the cholesterol is actively promoted by dynamical C-terminal residues (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 (in Å). The side chain of the residues and the corresponding distance is also highlighted.

## Experimental validation of monomeric ApoA1 bound to the cholesterol and subsequent structural changes.

To validate the above findings observed from simulations, we performed quantitative experiments to study Apo1-cholesterol ligand-induced binding

changes. Purification of monomer lipid-free ApoA1 posses experimental challenges due to its inherent nature of aggregation in moderately high concentration. We purified and confirmed the monomeric full-length lipid-free ApoA1 by Matrix Assisted Laser Desorption/Ionization (MALDI), with absence of any dimeric or oligomeric species (see SI Figure S8C). Dynamic light scattering (DLS) with 1mg/ml of the protein showed significant loss of monomeric species while there was an abundant amount of dimeric and other oligomeric species that formed upon concentrating the protein (see SI Figure S8D). Therefore, we ensured that the protein is kept at substantially low concentration, to confirm the monomeric state of the protein (using MALDI), in all our experiments.

The isothermal titration (ITC) experiments were conducted by step-by-step injections of 1 mM cholesterol solution as shown in Figure 8A. Magnitude of the heat signals become close to the titration of cholesterol into buffer at 9th injection. The integrated enthalpic curves, shown in Figure 8B, contain observed enthalpy changes ($\Delta H_{obs}$ in kcal/mol of injectant) for each injection of 0.1 mM cholesterol into buffer. The dilution curve of cholesterol into buffer can be classified into two regions. In region I, the injectant containing concentrated cholesterol produced large exothermic heat effects due to the deaggregation process of the aggregated cholesterol. In this region, monomeric and aggregated cholesterol coexist. In region II, a portion of the aggregated cholesterol injected essentially gets diluted in the cell since the concentration of the injectant is higher than that in the cell. However, the integrated enthalpic curves of cholesterol titrating into ApoA1 follow different trend and two regions can be seen under this experimental condition.

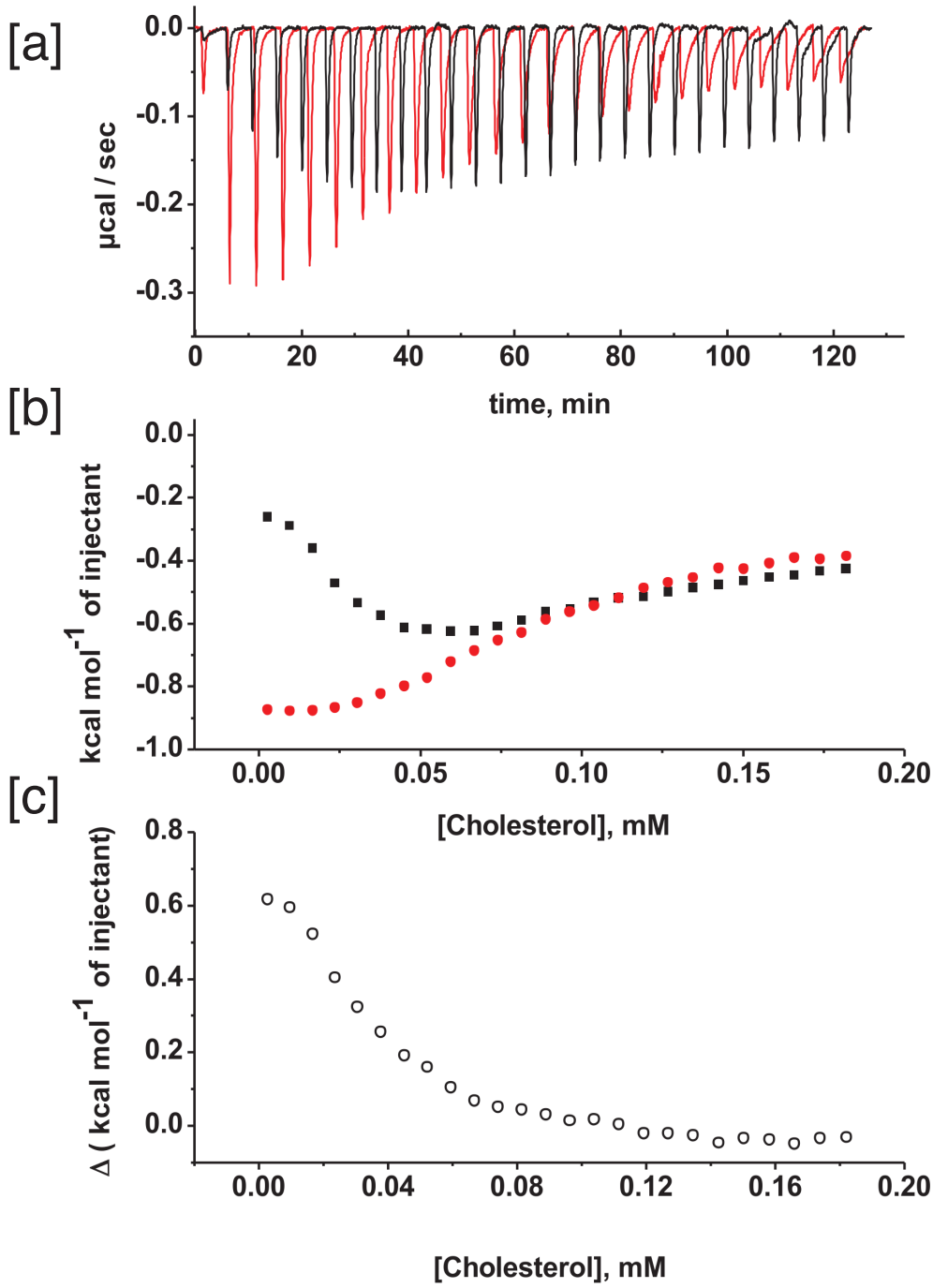


Figure 8: **Isothermal titration calorimetry (ITC):** A) The thermograms (cell feedback (CFB) in $\mu$cal/s) for titrations of 1mM choresterol into buffer (red) and 10 $\mu$M ApoA1 (black) B) $\Delta H_{obs}$ of each injection C) Enthalpic change due to the monomeric cholesterol binding to ApoA1 as a function of cholesterol concentration.

With an increase in cholesterol, binding sites in protein are occupied by cholesterol and less cholesterol is bound in subsequent injections. Overall endothermic contribution to heat change was gradually decreased. In the second region, there are sites in protein available for binding, monomeric cholesterol is are under equilibrium with aggregated cholesterol continued to bind. The differential enthalpic values under each injection ($\Delta H_{(Cholesterol-ApoA1)} - (\Delta H_{(Cholesterol-buffer)}$ is the enthalpy change due to the monomeric cholesterol binding to ApoA1, which has been plotted as a function of cholesterol concentration in Figure 8C. As it is seen that enthalpy change for such interaction is endothermic in nature, the interaction is completely entropic driven.

These findings from ITC experiments suggest that ApoA1-cholesterol interaction is indeed cholesterol-concentration dependent and exhibits ligand-induced equilibrium shift between different conformational states.

***Subsequent secondary structural changes.*** 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 9A). We found that the $\alpha$-helical propensity of the N-terminal region (residues 3-40, 48-90, 95-112, 118-124, 130-178) is highly conserved in both lipid-free and lipidated simulations. Previous reports also suggest that the N-terminal of the ApoA1 protein is relatively stable.[30] In contrast, 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. In particular, in the presence of cholesterol, residues (182-190, 198-199) dynamically interconvert between coil and helix. 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.[16,20,30]

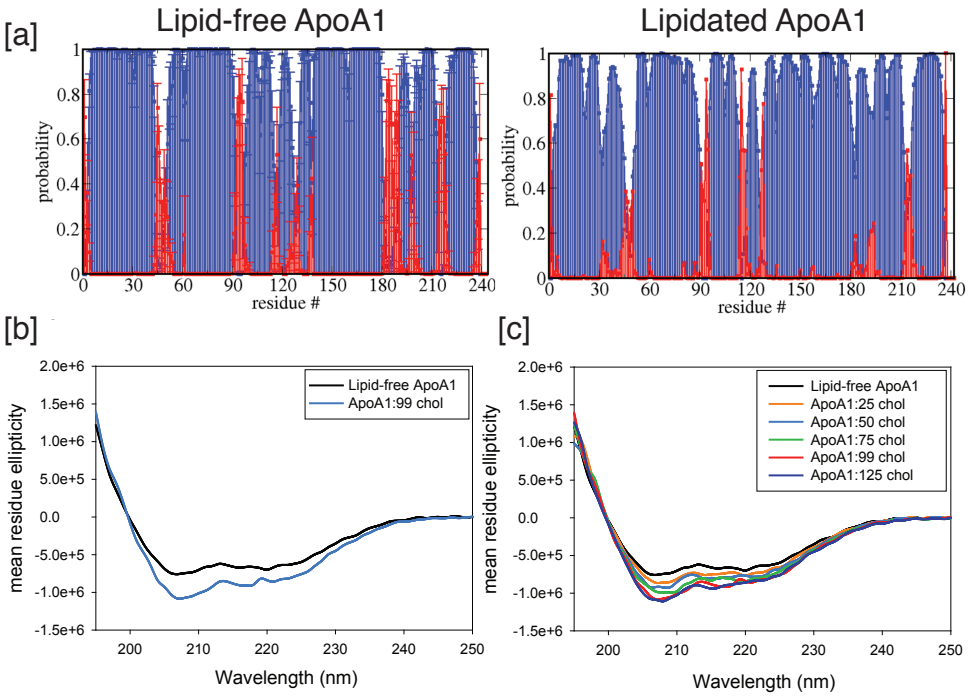


Figure 9: **Monitoring change in secondary structure of ApoA1 upon interaction with cholesterol:** A) Probability distribution for helix and coil secondary structure elements. The probability range is calculated on the basis of residue-based residence time, where 1 indicates the formation of helix (blue)/coil (red) 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) Mean residue molar ellipticity of monomeric ApoA1 with 99 molecules of cholesterol is plotted showing a dip at 208 nm and 222 nm, indicating increase in helicity upon cholesterol interaction. (C) Mean residue molar ellipticity of monomeric ApoA1 is plotted as compared to the wavelength with different ratios of cholesterol.

Further, we also performed CD spectroscopy of lipid-free monomeric ApoA1 and, cholesterol bound ApoA1 in increasing concentration (Figure 9B-C). The mean residue ellipticity drop at 208nm and 222nm, indicates increased alpha-helical region in the lipid-free monomeric state of ApoA1 compared to bound ApoA1. The delta change of mean residue ellipticity at 208nm further increases upon cholesterol interaction with increasing concentration of cholesterol, indicating an increase in the helical region of the protein with increase in the amount of cholesterol present in

the reaction. The increase in helicity of monomeric ApoA1 is proportional to the ratio of cholesterol to the protein in the reaction upto 99 cholesterol molecules per molecule of monomeric protein (Figure 9C). With 125 molecules of cholesterol, we do not find further increase in helicity of the monomeric protein. This corroborates with our simulations where we observe increase in helicity upon concentration-dependent lipid addition (Lipid-free<50CHL<75CHL<99CHL). Our data thus indicates that one of the key factors driving lipid-free and lipid-associated ApoA1 structural transitions is increased C-terminal helicity.

## Discussion

The Apolipoprotein ApoA1 governs numerous functional interactions during the complex assembly of high-density lipoprotein (HDL) particles,[5,31] 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,[31] (i) Lipid-free monomeric ApoA1, (ii) Lipid-poor ApoA1 monomer, also known as pre-$\beta$ 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.

In this work we focused 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. Previous studies have reported two crystal structures of ApoA1 in oliogomeric states, one with truncated C-teriminal as a dimer,[14] and an oligomeric assembly structure with tetrameric assembly.[13] Few solution structure of ApoA1 also exists,[32,33] including recently reported discoidal high-density lipoprotein particles.[34] The inherent cholesterol association of ApoA1 induces technical challenges in ascertaining its monomeric structure. The ApoA1 structural model, proposed in this paper, is constructed from a truncated mouse NMR structure having a 68% sequence identity with ApoA1 human, with the helical bundle N-terminal conformation and the C-terminal modeled based on extensive 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 $\mu$s. 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.[35,36]

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,[10,13,37] 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 found a strong association between conformational changes and functional binding sites.[38] Previous reports also suggest that the C-terminal $\alpha$-helix has the ability to mediate lipid efflux[39] and solubilize membrane phospholipids.[40,41]

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, 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 $\mu$s 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.[5,42] 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 interaction 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.[16,22,24,26,43–45] Our simulations are in excellent match with all these previous results and show new regions of lipid-binding initiation in concentration gradient manner.

The biophysical experiments performed in this study addressed a pertinent issue in ApoA1 biology. Does cholesterol associate in specific or random hydrophobic-driven manner? To answer this, experiments of monomeric protein cholesterol association and concentration-dependent studies were performed (Figure 8). ITC experiments indeed confirmed that adsorption of cholesterol binding is dependent on availability of protein sites and the reaction is entropically driven. We also confirmed that there exists a differential secondary structural arrangement in the presence of cholesterol. In this study, we also established the location of $\alpha$-helical segments in two monomeric ApoA1 states (Figure 9). 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.[16,17,44–46] In particular, residues 14-19, 26-31, 55-64, 70-77, 83-85, 93-97, 102-115, 147-148, 158-178, 224-227. Region of residues 115-150 was found to sensitive to proteolysis and hence thought to be flexible and unstructured.[47,48] We also observed several loops, which could impart flexibility to the region. In C-terminal, there is a sharp decrease in helical propensity from residues 181-192, 196-200, 212-221 while regions like residues 183-186, 189, 216-217, with high propensity of coil conformation. In particular, helical region residues 224-231, proposed to have the highest lipid-binding property matches very well with previous experimental findings.[16,26,44,45]

# Conclusion

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.[49,50] 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.

Using combined experimental and simulation approach we hereby showed that ApoA1 cholesterol association is specific and lipid concentration-dependent. We show that dynamics of protein domains is highly specific, indicating its link to distinct biological function. Interestingly, upon cholesterol addition ApoA1 monomer spontaneously forms open-topology that engulfs cholesterol in a belt shape. Using combined experimental and simulation approach we also showed that this phenomenon was found to be lipid concentration-dependent. 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,[25,26] 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] Glomset, J. A. *Journal of lipid research* **1968**, *9*, 155–167.

[2] Linsel-Nitschke, P.; Tall, A. R. *Nature reviews Drug discovery* **2005**, *4*, 193–205.

[3] Wang, X.; Rader, D. J. *Current opinion in cardiology* **2007**, *22*, 368–372.

[4] Lewis, G. F.; Rader, D. J. *Circulation research* **2005**, *96*, 1221–1232.

[5] 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.

[6] Miyazaki, O.; Ogihara, J.; Fukamachi, I.; Kasumi, T. *Journal of lipid research* **2014**, *55*, 214–225.

[7] Mulya, A.; Lee, J.-Y.; Gebre, A. K.; Thomas, M. J.; Colvin, P. L.; Parks, J. S. *Arteriosclerosis, thrombosis, and vascular biology* **2007**, *27*, 1828–1836.

[8] Kunitake, S. T.; La Sala, K.; Kane, J. P. *Journal of lipid research* **1985**, *26*, 549–555.

[9] Wlodawer, A.; Segrest, J. P.; Chung, B. H.; Chiovetti, R.; Weinstein, J. N. *FEBS letters* **1979**, *104*, 231–235.

[10] 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.

[11] Shih, A. Y.; Sligar, S. G.; Schulten, K. *Journal of The Royal Society Interface* **2009**, rsif20090173.

[12] Barter, P.; Kastelein, J.; Nunn, A.; Hobbs, R.; Board, F. F. E. *Atherosclerosis* **2003**, *168*, 195–211.

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

[14] Mei, X.; Atkinson, D. *Journal of Biological Chemistry* **2011**, *286*, 38570–38582.

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

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

[17] 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.

[18] Wu, Z.; Wagner, M. A.; Zheng, L.; Parks, J. S.; Shy, J. M.; Smith, J. D.; Gogonea, V.; Hazen, S. L. *Nature structural & molecular biology* **2007**, *14*, 861–868.

[19] Silva, R. G. D.; Hilliard, G. M.; Fang, J.; Macha, S.; Davidson, W. S. *Biochemistry* **2005**, *44*, 2759–2769.

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

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

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

[23] Segrest, J. P.; Jones, M. K.; Shao, B.; Heinecke, J. W. *Biochemistry* **2014**, *53*, 7625–7640.

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

[25] 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.

[26] 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.

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

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

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

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

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

[32] Wang, G.; Sparrow, J. T.; Cushley, R. J. *Biochemistry* **1997**, *36*, 13657–13666.

[33] Wang, G.; Treleaven, W. D.; Cushley, R. J. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* **1996**, *1301*, 174–184.

[34] Bibow, S.; Polyhach, Y.; Eichmann, C.; Chi, C. N.; Kowal, J.; Albiez, S.; McLeod, R. A.; Stahlberg, H.; Jeschke, G.; Güntert, P.; Riek, R. *Nature Structural and Molecular Biology* **2017**, *24*, 187.

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

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

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

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

[39] 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.

[40] 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.

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

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

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

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

[45] 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.

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

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

[48] 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.

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

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