A consensus model of human apolipoprotein A-I in its monomeric and lipid-free state

John T Melchior[1] , Ryan G Walker[2], Allison L Cooke[1], Jamie Morris[1], Mark Castleberry[1],
Thomas B Thompson[2], Martin K Jones[3], Hyun D Song[3], Kerry-Anne Rye[4], Michael N Oda[5], Mary G Sorci-Thomas[6],
Michael J Thomas[7], Jay W Heinecke[8], Xiaohu Mei[9], David Atkinson[9] , Jere P Segrest[3], Sissel Lund-Katz[10],
Michael C Phillips[10] & W Sean Davidson[1]

Apolipoprotein (apo)A-I is an organizing scaffold protein that is critical to high-density lipoprotein (HDL) structure and metabolism, probably mediating many of its cardioprotective properties. However, HDL biogenesis is poorly understood, as lipid-free apoA-I has been notoriously resistant to high-resolution structural study. Published models from low-resolution techniques share certain features but vary considerably in shape and secondary structure. To tackle this central issue in lipoprotein biology, we assembled a team of structural biologists specializing in apolipoproteins and set out to build a consensus model of monomeric lipid-free human apoA-I. Combining novel and published cross-link constraints, small-angle X-ray scattering (SAXS), hydrogen–deuterium exchange (HDX) and crystallography data, we propose a time-averaged model consistent with much of the experimental data published over the last 40 years. The model provides a long-sought platform for understanding and testing details of HDL biogenesis, structure and function.

Apolipoprotein (apo)A-I is the most abundant protein in HDL. Although HDL contains up to 95 'accessory' proteins (http://homepages.uc.edu/~davidswm/HDLproteome.html) that modulate particle function, apoA-I is central to mediating HDL's structure, interactions and function. Indeed, it is a cofactor for lecithin–cholesterol acyl transferase, mediating HDL maturation via the esterification of free cholesterol in plasma[1]. Lipid-free (or lipid-poor) apoA-I is also critical for HDL biogenesis via the cell-surface transporter ATP-binding cassette A1 (ABCA1)[2]. Thus, there has been a sustained effort to understand the apoA-I structure in both its lipid-free and its lipid-bound states.

Lipid-free monomeric apoA-I represents the protein in its most nascent form. As such, it is a logical starting point for the investigation into the structural steps of HDL biogenesis. However, apoA-I has proven to be difficult to study by traditional protein structural analyses like X-ray crystallography or NMR. It has a tendency to self-associate and bind hydrophobic substances, and it has exceptional structural flexibility. Indeed, multiple laboratories have been trying to crystallize the full-length protein for decades without success. Two groups published crystal structures of tetrameric[3] and dimeric[4] truncation

mutants of lipid-free apoA-I. The former, lacking the N-terminal 44 amino acids, was widely thought to mimic the apoA-I configuration on a nascent, discoidal HDL particle. The latter structure lacked the C-terminal 58 amino acids and may be more reflective of lipid-free apoA-I. Independent studies of the same truncation mutant showed that most of the structural features of the crystal structure were borne out in solution[5]. These reports have provided critical insights into structural details of lipid-free apoA-I, but they must still be extrapolated to the *in vivo* condition, in which the N and C termini are both present and the protein is in aqueous solution.

The difficulty of crystallizing full-length apoA-I has led to many studies using clever applications of lower-resolution techniques, such as circular dichroism (CD)[6], analytical ultracentrifugation[7], NMR[8], fluorescence spectroscopy[9], molecular dynamics[10], chemical cross-linking[11], HDX[12], electron paramagnetic resonance spectroscopy (EPR)[13,14] and others[15]. Cross-linking has been particularly useful for 3D models, because it provides distance-constraint data under physiological conditions in solution. To date, there have been eight models proposed for full-length monomeric apoA-I from four laboratories[10,11,16,17]. Examples are shown in **Figure 1**. They all convey

[1]Department of Pathology and Laboratory Medicine, University of Cincinnati, Cincinnati, Ohio, USA. [2]Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati, Cincinnati, Ohio, USA. [3]Department of Medicine, Division of Cardiovascular Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [4]School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, New South Wales, Australia. [5]Children's Hospital Oakland Research Institute, Oakland, California, USA. [6]Department of Medicine, Section on Endocrinology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. [7]Department of Pharmacology and Toxicology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. [8]Department of Medicine, University of Washington, Seattle, Washington, USA. [9]Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts, USA. [10]Division of Translational Medicine and Human Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. Correspondence should be addressed to W.S.D. (Sean.Davidson@uc.edu).

the generally held view that lipid-free apoA-I is a four-helix bundle in a pseudo-stable state poised to interact with lipids. However, looking more deeply, the models differ widely in shape, secondary structure placement and the location of the N and C termini. Thus, the field still lacks the structural detail to understand how apoA-I binds lipids, interacts with ABCA1 to form mature HDL, exchanges between HDL and mediates particle interactions with plasma-borne remodeling factors.

In the absence of a high-resolution structure of lipid-free apoA-I but with the benefit of nearly five decades of spectroscopic and cross-linking data, we assembled a group of apoA-I structural investigators, the ApoA-I Working Group, to derive a consensus model that consolidates the apoA-I structural literature. In addition, we report new cross-linking data to round out a universal list of 95 distance constraints from three laboratories, using multiple cross-linking reagents. We also included new molecular-shape data from SAXS and incorporated secondary structural information from HDX[12]. Using the crystal structure of dimeric apoA-I (ref. 4) as a template, we applied iterative computer modeling techniques to derive the new model.

RESULTS

Cross-linking and SAXS analyses

We used the homobifunctional cross-linkers CBDPS (cyanur-biotin-dimercapto-propionyl-sulfo-succinimide) and BS3 (bis(sulfosuccinimidyl) suberate), both NHS esters that cross-link lysine and, to a lesser extent, serine residues[18,19]. Wild-type (^{14}N) and isotopically labeled (^{15}N) apoA-I were mixed 1:1 under denaturing conditions. This allowed unambiguous peptide identification and confirmation that cross-links were not due to oligomerization[5]. After cross-linking, monomeric apoA-I was isolated by size-exclusion chromatography (SEC). We routinely obtained >95% purity of cross-linked monomeric apoA-I, as shown using SDS–PAGE (**Supplementary Fig. 1**). **Supplementary Tables 1** and **2** list the 65 cross-linked peptide pairs that were identified (criteria in Online Methods). All cross-links exhibited a dual-peak pattern, indicating intramolecular span[5,20]. 45 linkages were identified in both BS3- and CBDPS-treated samples, with 15 unique to BS3 and five unique to CBDPS.

We used SAXS to assess the shape of monomeric apoA-I in solution. ApoA-I's propensity for concentration-dependent self-association posed a challenge, because reliable SAXS data collection requires concentrations up to 4 mg/ml. We circumvented this issue by performing SAXS on apoA-I that had been locked into the monomeric state by cross-linking with BS3 or CBDPS (Online Methods). Careful evaluation using SEC showed that this method prevented further oligomerization. The SAXS parameters are shown in **Supplementary Table 3**.

For both cross-linking reagents, scattering intensity increased proportionally with sample concentration, and the Guinier range was linear at low scattering angles (not shown), indicating no concentration-dependent effects and good data quality. No significant differences were observed in R_g (Guinier and real space) and D_{max}, and comparable molecular volumes were found across all concentrations. Additionally, scattering profiles and pairwise distribution plots for both samples were highly related (**Fig. 2a,b**). There may be small differences in the flexibility of differentially cross-linked apoA-I species. ApoA-I cross-linked with CBDPS showed a dip in the $q^3 \times I(q)$ plot (**Fig. 2c**) and a plateau in the $q^4 \times I(q)$ plot (**Fig. 2d**), indicative of a folded, rigid structure. The BS3-treated sample exhibited a plateau in the $q^3 \times I(q)$ plot and a lack of plateau in the $q^4 \times I(q)$ plot, thus indicating a possible flexible domain. The deviation in dynamics is likely due to differences in cross-linking efficiency, as much less CBDPS is required for linking oligomers than BS3. The R_g of both samples
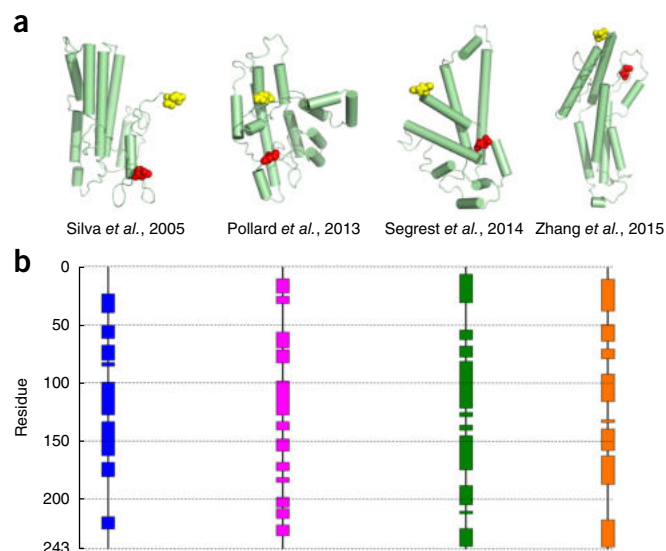
Figure 1 Published models of full-length lipid-free human apoA-I. There are eight published models of lipid-free monomeric apoA-I that have coordinates available, and the figure shows four of these models. Silva et al.[11] proposed a model based on chemical cross-linking and homology-sequence threading techniques. Pollard et al.[16] generated their model using chemical cross-linking and cysteine point mutations. Segrest et al.[10] used chemical cross-linking and molecular dynamics simulations using information from the crystal structures of apoA-I truncation mutants as templates; this article also proposed three other models that differed in terms of length of simulation (one of which is shown here, 15 ns). Zhang et al.[17] produced molecular simulations using information from the crystal structures as initial templates. (**a**) The models are shown in cartoon form with cylinders representing helical domains. The N-terminal amino acids are shown in red, and the C-terminal amino acids are shown in yellow. (**b**) The α-helical content of each model (as assessed using PyMOL) is shown on the linear diagram for each model; colored blocks represent helical domains, and black lines represent random-coil or loop regions.

were similar (25.35 ± 0.15 Å and 25.34 ± 0.17 Å for CBDPS and BS3, respectively) and, taken with the pairwise distribution plots, suggested that apoA-I had characteristics of a globular protein. Twenty-three independent envelope reconstructions were performed and averaged using DAMMIF[21] to generate a composite. Both DAMMIF reconstructions had normalized spatial discrepancies (NSD) between 0.5 and 0.7 (**Supplementary Table 3**), indicating good convergence of the independent reconstructions. The envelopes (**Fig. 2e,f**) serve as a low-resolution representation of the overall shape of the molecule, which can be used as a rough tool for visualizing the fit of a model. ApoA-I cross-linked with CBDPS had a calculated volume of ~69,400 Å[3] compared to ~79,500 Å[3] for that cross-linked with BS3, consistent with the more flexible nature of the BS3 sample. Taken together, the data suggest minimal structural differences between the samples.

Generation and evaluation of the model

The starting construct of the model was derived from the crystal structure of apoA-I[1–184], as shown in **Supplementary Figure 2**. As Mei and Atkinson[4] proposed, an inflection point was first placed in helix H5, near the center of the long helix shared by two protomers (A and B) in the dimer. However, guided by cross-linking constraints and SAXS, the inflection point was shifted more C terminal to residue 139 in H5. The C-terminal portion of monomer A was then folded back, resulting in the juxtaposition of H6 (A) and H5 (A) (**Supplementary Fig. 2b**). Using PyMOL (http://www.pymol.org) and Modeller (v9.14), the
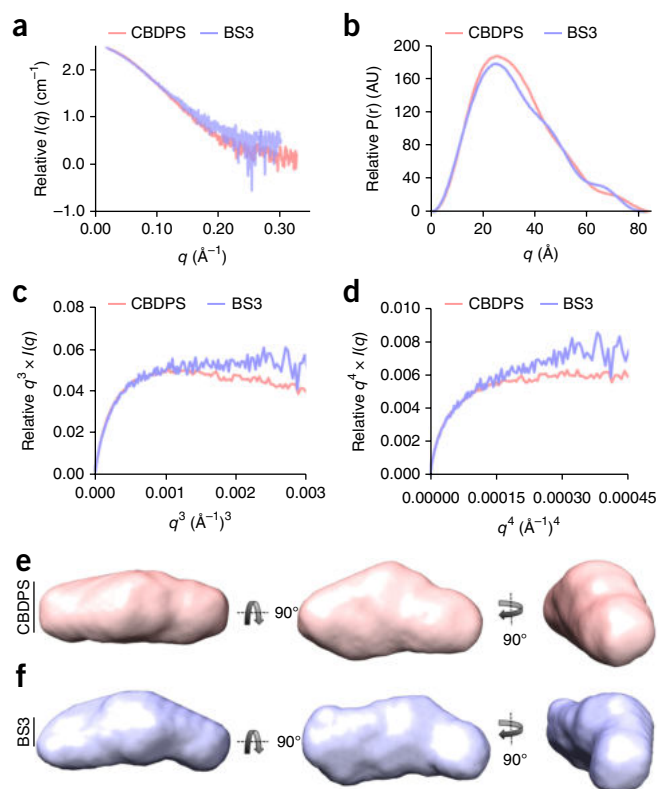
Figure 2 SAXS plots and molecular envelopes of monomeric apoA-I. Plasma and recombinant apoA-I were cross-linked with CBDPS and BS3, and monomeric species were isolated by gel filtration chromatography and subjected to SAXS. (**a**) Scattering profiles of monomeric apoA-I cross-linked with CBDPS and BS3. (**b**) P(r) curves of monomeric apoA-I cross-linked with CBDPS and BS3. AU, arbitrary units. (**c**) SIBYLS ($q^3 \times I(q)$ vs. q^3) and (**d**) Porod–Debye ($q^4 \times I(q)$ vs. q^4) flexibility plots of monomeric apoA-I cross-linked with CBDPS and BS3. (**e**,**f**) DAMMIF *ab initio* molecular envelopes of recombinant apoA-I cross-linked with CBDPS (**e**) and BS3 (**f**).

Figure 3 All-atom model of full-length, lipid-free monomeric apoA-I. The final time-averaged model is shown, in which dark blue represents N-terminal α-helical domains H1 and H2, light blue represents α-helical domains H3, H4, H5 and H6, pink represents random coil, red represents the C-terminal 59 residues absent in the crystal structure and yellow indicates the positions of prolines on each model. NT, N terminal; CT, C terminal. The model coordinates are in **Supplementary Data Set 1**.

missing residues 185–243 were threaded in and positioned, guided by the relevant cross-linking constraints[10,11,16] (**Supplementary Table 4**). This construct was used as a base model for further refinement. Next, secondary structural data derived exclusively from HDX[12] was implemented onto the base model. This required conversion of helical segments observed in the crystal structure to random coil (and vice versa in some cases). Then, the entire model was simultaneously refined for fit to the SAXS scattering curves and the cross-linking data, while holding the secondary structure assignments as constant as possible. Lastly, the model was subjected to energy minimization using YASARA[22], and side chain rotamers were manually refined in COOT[23] (details of each step in Online Methods).

The final model (**Fig. 3**) depicts monomeric, lipid-free apoA-I as a helical bundle composed of three main helices: H1 (residues 8–35), H5 (81–115) and H6 (148–179). Three shorter helical regions are interspersed between these helices: H2 (37–45), H3 (54–64) and H4 (68–78). The remainder is random coil with two main stretches spanning residues 116–147 and the entire C terminus, residues 180–243. The C terminus forms a globular domain that sits on the bundle at the N-terminal end of H5 and the C-terminal end of H6. The N and C termini are located near each other at the base of the bundle and opposite the bulk of the C-terminal globular domain. Proline residues at positions 7 and 66 reside in regions of transition between helix and random coil, with residue 66 clearly mediating a turn structure.
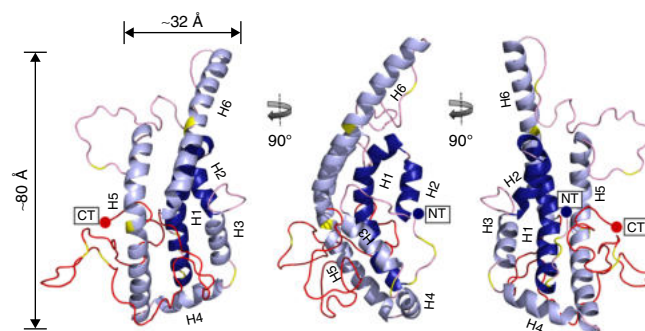
Although most of the other prolines are located in random-coil regions, prolines 99 and 165 occur in the middle of H5 and H6, respectively, without apparent disruption.

Compatibility of the model with old and new experimental data
A contact plot of the final structure is shown in **Figure 4a**, with experimental cross-links superimposed in blue and violations in red. Overall, 89 of 95 cross-links (both previously published and newly reported here) were consistent with the model. It is worth noting that this is 3–5 times the number of constraints used in previous modeling attempts, significantly reducing the conformational possibilities that can fit the data. The six violations were, on average, within ~2 Å of the allowable span. **Figure 4b** shows the model agreement with the HDX data of Chetty *et al.*[12]. The new model has 235 of 243 residues (97%) assigned in keeping with the HDX data. The small differences resulted from tradeoffs made for optimal cross-link and SAXS data fitting. These differences occurred near residue 36, which is predicted to form the minor H2 helix (also observed in the crystal structure[4]). The remaining differences involved 1–2 amino acids bookending α-helical regions. **Figure 4c** shows the model superimposed onto both DAMMIF *ab initio* molecular envelopes generated from recombinant apoA-I cross-linked with CBDPS and BS3. The primary measure of agreement to SAXS was the model fit to the scattering curves. **Figure 4d,e** shows that the model was in good agreement with both SAXS scattering curves ($\chi = 1.13$ and $\chi = 0.81$ for CBDPS and BS3, respectively). As a point of reference, our previous work compared a crystal structure of a truncation mutant of apoA-IV with similarly obtained SAXS data with a $\chi = 0.97$ (ref. 20).

We also reviewed the literature for additional structural data for lipid-free apoA-I collected under conditions in which it was monomeric. **Supplementary Table 5** summarizes numerous studies that used far-UV CD to estimate α-helix and random-coil contents. α-helical content ranged from 40% to 68% across 27 studies, for an average of $52 \pm 6\%$ or 126 ± 14 residues. The new model, driven primarily by the HDX restraints, matched exactly, with 126 α-helical residues. Another technique for identifying areas of low structural organization is limited proteolysis. **Supplementary Table 6** summarizes two studies that applied this technique to primarily monomeric apoA-I. Proteolytically susceptible sites (highlighted on the new model in **Fig. 5a**) tended to occur in random-coil areas near the molecule surface. The solvent accessibility of these sites in the new model, as assessed using VADAR[24], was highly consistent with the proteolysis data, with only
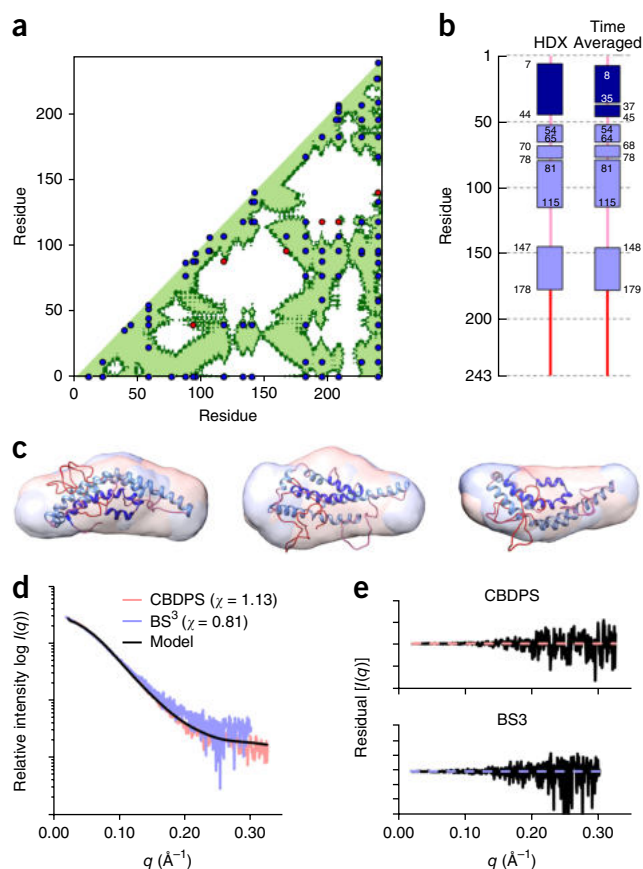
Figure 4 Evaluation of experimental cross-links, secondary structure and SAXS data with respect to the model. (**a**) Cross-linking data **Supplementary Table 4** containing all published cross-links reported by Melchior, Silva[11], Pollard[16] and Segrest[10]. The x and y axes of the plots indicate the residue number of apoA-I from 1–243. The diagonal line bisects the figure and the bottom and right show intramolecular interactions. Dark green represents residues within 28 Å (upper limit for C_α cross-linked with CBDPS or BS3) and light green represents residues within 23 Å (upper limit for C_α cross-linked with MDA). Blue dots represent experimental cross-links that fall within the upper-limit distance constraint, and red dots represent those that exceed the allowable distance on the time-averaged model. (**b**) Comparison of the secondary structure reported in the HDX study by Chetty et al.[12] and the time-averaged model. Rectangles represent α-helical segments and the lines represent random coil. (**c**) Superposition of the model on molecular envelopes derived from monomeric apoA-I cross-linked with CBDPS and BS3. (**d**) The theoretical scattering curve of the proposed model (black line) superimposed on the experimental scattering profile of apoA-I cross-linked with CBDPS and BS3. The fit to the respective scattering profiles are shown in parenthesis. (**e**) The residual plot of the model (black) to experimental scattering data from apoA-I cross-linked with CBDPS and BS3.

two of 13 reported cleavage sites (Tyr115, Phe57) showing solvent-accessible areas <20 Å[2].

Supplementary Figure 3 summarizes how the new model compares to the eight published models of monomeric apoA-I with respect to the experimental data outlined above. In most cases, the new model showed considerably improved compatibility with the data.

DISCUSSION

Lipid-free apoA-I is of direct physiological relevance because it and its related forms (such as preβ1-HDL) are essential to *de novo* HDL biogenesis via ABCA1-mediated efflux of phospholipid and cholesterol,
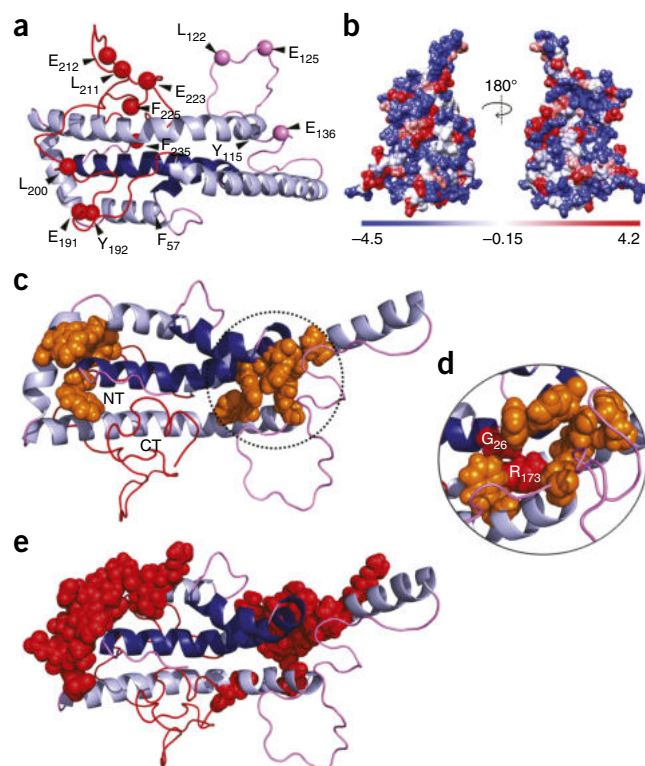

Figure 5 Predictions of the new model with respect to proteolytic susceptibility and stability. (**a**) Proteolytic cleavage sites from **Supplementary Table 6**. (**b**) Space-filled rendering of the model using residue hydrophobicity, with hydrophilic residues in blue and hydrophobic residues in red. (**c**) Aromatic clusters (orange) that may participate in molecular stabilization and the proximity of the N and C termini in the time-averaged model. (**d**) Position of natural mutations in apoA-I_IOWA (Gly26) and apoA-I_MILANO (Arg173) within one aromatic cluster. (**e**) Critical stabilizing residues mapped by Gorshkova et al. for lipid-free apoA-I (refs. 33,42).

a critical anti-atherogenic pathway in human plasma[25]. This involves direct interaction of apoA-I with both ABCA1 and specialized domains in the plasma membrane[26,27]. The concentration of total apoA-I in human plasma is about 130 mg/dL (~50 μM). At 5–10% of total apoA-I (ref. 28), lipid-free apoA-I in plasma is around 2 μM and should be mostly monomeric. In peripheral lymph (interstitial fluid), the concentration is about ten-fold less concentrated (~0.2 μM or about 5 μg/ml), corresponding to the K_m for lipid efflux via ABCA1 (ref. 29). Thus, a detailed understanding of apoA-I in its most nascent biophysical state (lipid-free and monomeric) will aid in our understanding of HDL biogenesis and its eventual interaction with critical HDL mediators, such as plasma lecithin–cholesterol acyl transferase. A solution to this problem has been pursued intensively for decades, but a high-resolution structural model of apoA-I's native form has remained elusive. The model described here represents a comprehensive multitechnique attempt to address this issue, given apoA-I's recalcitrance to traditional high-resolution structural techniques. We first describe the features of the model, touch on its implications for understanding HDL biology and then discuss its limitations.

Figures 3 and **5** show that the model is consistent with previous proposals of a two-domain structure with a relatively organized N-terminal helical bundle and a mostly random-coil C-terminal domain from residues 180–243. With respect to the C-terminal domain, we point out that our methodology does not fully preclude the existence of some helical
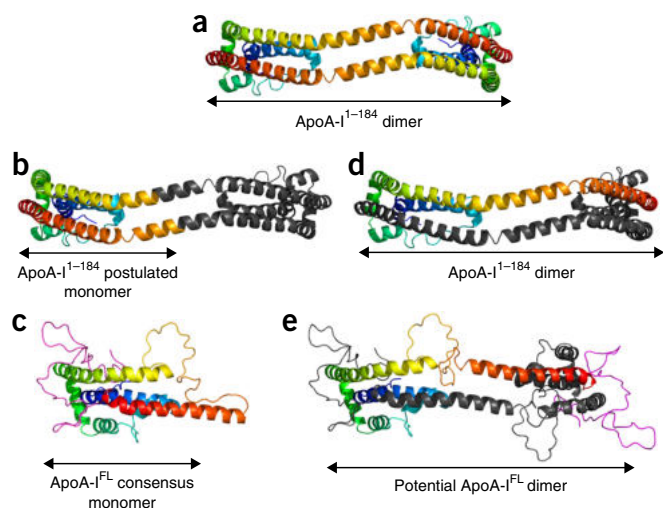
Figure 6 Comparison of the consensus apoA-I model to the apoA-I[1–184] crystal structure. (**a**) The dimeric apoA-I[1–184] crystal structure is shown in cartoon rendering. For each protomer, the N terminus is blue and gradates through cyan, green, yellow and orange to the C terminus (residue 184 in this case) in red. (**b**) A postulated monomeric form of apoA-I[1–184] as proposed by Mei *et al.*[4]. An inflection point was placed near the center of the long helix shared by two protomers (A and B) at approximately residue 126. Monomer A is rendered with the same color gradient as that in **a** until residue 126, at which point all remaining residues are colored gray. With its N-terminal portion colored gray, the C-terminal portion of monomer B is colored in a continuation of the gradient as though it was a single molecule with a turn introduced, as indicated by the arrow. (**c**) The consensus model of full-length apoA-I, colored scheme same as that in **a**, with residue 182 in red. The residues that are missing from the crystal structure are in magenta. (**d**) The apoA-I[1–184] crystal structure rendered with all of molecule B in gray. (**e**) Postulated model of dimeric full-length apoA-I. Molecule A is colored as is the monomer in **c**, with residues 183–243 in magenta. Molecule B is entirely gray.

structure, particularly highly unstable helices that cannot be detected at the timescale of the HDX experiments. Indeed, truncation of the C terminus from residue 243 to residues 231 and 221 reduced the CD-detectable-helix content by seven and 14 amino acid residues, respectively[30,31]. On the basis of such observations, Mei and Atkinson[31] suggested that the segment spanning residues 231–241 contains an α-helical structure. Such a C-terminal helix is further supported by preliminary molecular dynamics simulations of the new model (Segrest *et al.*, unpublished observation). Overall, the C-terminal domain sits on N-terminal α-helical domains H1 and H2 (**Fig. 5**, dark blue), with the termini residing close to each other. This theme was also observed in the solution structure for apoA-IV (ref. 20) and may explain structural and energetic studies indicating that the N- and C-terminal domains of apoA-I act cooperatively to modulate its stability[32] and lipidation kinetics[5,33–36]. The space-filling illustration in **Figure 5b** shows that the model is well packed, particularly in the N-terminal domain, with hydrophobic faces of the helices oriented toward the bundle core. However, there are clear instances of solvent-exposed hydrophobic residues, possibly contributing to the protein's low thermodynamic stability[37]. Another intriguing feature is two potential networks of aromatic residues (**Fig. 5c**). Aromatic residues play important roles in protein stability, protein–protein interactions and tertiary folding[38]. Indeed, the natural mutations G26R in apoA-I[IOWA][39] and R173C in apoA-I[MILANO][40] both result in substantial decreases in free energy of denaturation of lipid-free apoA-I (ref. 41), and both are central to an aromatic cluster that potentially stabilizes helices 1, 5, and 6 (**Fig. 5d**). Lastly, residues reported by

Gorshova and colleagues[33,42] to be critical for apoA-I stability (**Fig. 5e**, space filled in red) exhibit remarkable colocalization with the aromatic clusters highlighted in **Figure 5c**. The model predicts that additional mutations in these areas should affect stability and likely the propensity to oligomerize and interact with lipids.

We considered the new model in light of apoA-I's propensity to self-associate. **Figure 6** compares the consensus model to the crystal structure of dimeric apoA-I[1–184] and its monomer, as postulated by Mei *et al.*[4]. Comparing **Figure 6b** and **c**, the full-length model shares numerous features with the postulated monomer. Both models show an N-terminal helix between residues 7 and 36, situated alongside another helix composed of residues 80–116. Though these helices are bent more acutely in the full-length model, the interhelical interactions are similar. Also, the N termini of both models are in proximity to a short helix of residues ~66–76. The full-length model shows a large random-coil region between residues 116–147 that acts as a turn, allowing H6 (residues 148–178) to double back and participate in the helical bundle. Although more ordered in the crystal model, a similarly centered turn (near residue 129) generates an intramolecular helical bundle. The main difference is that in the full-length model, the position of H6 is shifted relative to that of the bundle, perhaps by the presence of residues 185–243 that sit against the bundle (note how the red helix in **Fig. 6c** is shifted to the right compared to that in **Fig. 6b**). Given the relative agreement with the monomeric model derived from the crystal structure, it follows that the full-length protein may dimerize similarly to the crystal structure dimer. **Figure 6d,e** show how this might occur. By placing a hinge point near residue 129, the C-terminal portion of full-length apoA-I can be extended to interact with a similarly extended apoA-I partner. Aside from the increased random-coil structure in the consensus-model dimer, many of the intermolecular alignments seen in the apoA-I[1–184] dimer are highly plausible. Thus, we propose that the dimer–monomer conversion scheme proposed by Mei *et al.* for apoA-I[1–184] is compatible with the new consensus model and may describe how the full-length protein dimerizes. Further experimental work is required to test this hypothesis.

We also considered the implications of the new model for apoA-I lipid binding and HDL biogenesis. There is a general consensus that lipid binding is initiated when the relatively unstable C-terminal domain comes in contact with a lipid surface[31,43–45], either via a packing defect[46] or an ABCA1-mediated mechanism[26]. It is easy to visualize the periodic movement of the C terminus away from the helical bundle in the consensus model. If it encounters lipid, the formation and/or stabilization of C-terminal helices provides a favorable ΔG (ref. 30) that drives further lipid association. There is also evidence that the N terminus may play a role in initial lipid contact[35,47]. Freed of the C-terminal domain, the bundle helices can then unpack, reorienting the hydrophobic faces of the central part of apoA-I to penetrate the lipid surface[35]. One of the main benefits of having the consensus structure is that it gives a 'road map' to the initial position of the helices and offers a structural basis for how they may unpack during this transition. For example, Saito *et al.* speculated that H1 and H2, based on their similar low stabilities, may be in close opposition and unfold as a unit[30] early on in the lipid-binding process. Our model shows that H2 (in our case, residues 50–80) forms a bent helix that actually wraps around H1, thus supporting this idea.

There is considerable debate over the exact subsequent steps involved in the creation of a nascent HDL particle. Key questions remain as to the role and timing of apoA-I dimerization (Does it happen before, during or after lipidation?) and the exact mechanism and role of ABCA1 (Does it play a role in apoA-I dimer formation, and how does it physically load lipid onto apoA-I?). A recent cryo-EM
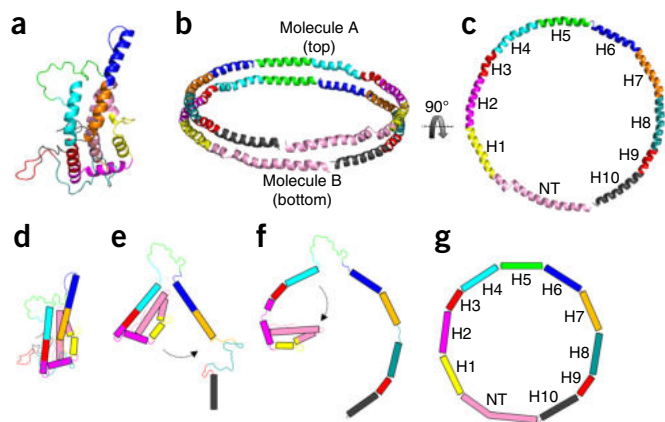
Figure 7 Hypothesis on lipid binding and the relationship to apoA-I arrangement on discoidal HDL. (**a**) The consensus model of full-length apoA-I. (**b**) Proposed arrangement of two molecules of apoA-I on discoidal HDL: the double-belt model. (**c**) Molecule B from the double-belt model rotated 90° (top view). (**d**–**g**) Cartoon representation, with rectangles representing α-helical segments. The model and cartoon is colored based on α-helical repeats identified and reported by Segrest *et al.*[49]. (**d**) Time-averaged model in its folded state. (**e**) Dissociation of helix 6 and the C terminus from the lipid-free model and formation of a helical region upon interaction of the C terminal with lipid. (**f**) Formation of additional α-helical segments upon lipid filling and unfurling of the N-terminal bundle. (**g**) Endpoint as a belt on an HDL particle.

structure of ABCA1 offers interesting possibilities[48], but a consensus on how this mechanism works has yet to materialize. In **Figure 7**, we show a simplified scheme consistent with that proposed by Mei *et al.*[31] of how the consensus model could be unpacked to generate a ring-like conformation, as proposed for the double-belt structure of apoA-I in discoidal HDL particles[49]. We caution that this scheme is only one of many possibilities, and the role of apoA-I dimerization and ABCA1 was intentionally not addressed. Now that we have good ideas of the structures of both apoA-I and ABCA1, we look forward to exciting new work on this critical mechanism.

Despite its consistency with much of the known structural data, we note that the model is still limited in resolution compared to those from NMR or X-ray crystallography. The general backbone configuration is likely correct, but more refined molecular interactions such as salt bridging and hydrogen bonding remain unclear. Additionally, one might question whether it is possible to represent a highly dynamic protein with a single, time-averaged model. ApoA-I has a ΔG of helix stabilization ranging from ~3–5 kcal/mol[12]. At neutral pH and room temperature, HDX studies have demonstrated complete hydrogen exchange into apoA-I α-helical segments occurring in ~10 min (**Supplementary Fig. 4**); all apoA-I helical segments unfolded and reformed at least once in this timeframe. To put this finding into context, a stable globular protein like cytochrome *c* has a ΔG of helix stabilization of 10 kcal/mol, with complete hydrogen exchange requiring ~10 weeks. Given this degree of dynamics, it seems likely that, at physiological temperatures, apoA-I exists in an ensemble of related structures at any point in time. Indeed, the fact that some cross-links failed to fit the model (six were close, but not strictly allowed) directly suggests differential conformations. Our model is a time-averaged structure derived from experimental data obtained on a longer time scale than that of typical secondary structure oscillations. For this reason, we think of it as a base model upon which hypothesized dynamics and conformational alterations can be further modeled and tested. However, because the majority of cross-links did fit the model and

the SAXS patterns showed consistent molecular shapes, we argue that the consensus model probably reflects a majority of the molecules existing at a given time or that any conformers exhibit related shapes and structures.

Another issue relates to the notion of solvent accessibility of the cross-linking reagents. Whereas most cross-links fit the model in terms of Euclidian distance ('as the crow flies'), many are impeded by some obstruction, like a side chain from a nonparticipating residue or an adjacent helical domain. Given that the model represents a time average of an ensemble of related apoA-I structures, cross-links that appear sterically hindered or solvent inaccessible could occur on an alternate conformation within the boundaries of the experimental system. Additional studies are needed to better define these boundaries and the extent of rearrangement apoA-I can achieve *in vivo* and *in vitro*. Additionally, the constant unfolding of helical domains over the course of the experiment could allow cross-links that might not be expected in a static structure. Nevertheless, previous reports have shown excellent consistency between observed cross-links in solution with crystal structures of apoA-I[1–184] (ref. 5) and apoA-IV (ref. 20) and non-apoliproteins[50–53] validating the approach. Importantly, instances of steric cross-linker hindrance were also routinely observed in these studies. A more detailed discussion of potential limitations of the model is presented in the **Supplementary Note** and **Supplementary Table 7**.

In summary, we propose a new time-averaged model of the structure of apoA-I in solution that is, by and large, consistent with decades worth of experimental data. Although still lacking amino acid level resolution, it provides a much-needed starting point upon which to test structural hypotheses with regard to loss-of-function mutations, interaction with cell-surface proteins and HDL formation.

METHODS
Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
J.T.M. and W.S.D. conceived and designed new experiments reported in this paper. J.T.M., R.G.W., A.L.C., J.M., M.C. and H.D.S. performed experiments. J.T.M., R.G.W., M.C., T.B.T., M.K.J., H.D.S., J.P.S., M.C.P. and W.S.D. analyzed data. J.T.M., M.C., T.B.T., M.K.J., H.D.S., K.-A.R., M.N.O., M.G.S.-T, M.J.T., J.W.H., X.M., D.A., J.P.S., S.L.-K., M.C.P. and W.S.D. derived the model. J.T.M., T.B.T., K.-A.R., M.N.O., M.G.S.-T., M.J.T., J.W.H., D.A., J.P.S., S.L.-K., M.C.P. and W.S.D. wrote the manuscript.

1. Soutar, A.K. *et al.* Effect of the human plasma apolipoproteins and phosphatidylcholine acyl donor on the activity of lecithin: cholesterol acyltransferase. *Biochemistry* **14**, 3057–3064 (1975).

2. Phillips, M.C. Molecular mechanisms of cellular cholesterol efflux. *J. Biol. Chem.* **289**, 24020–24029 (2014).
3. Borhani, D.W., Engler, J.A. & Brouillette, C.G. Crystallization of truncated human apolipoprotein A-I in a novel conformation. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1578–1583 (1999).
4. Mei, X. & Atkinson, D. Crystal structure of C-terminal truncated apolipoprotein A-I reveals the assembly of high density lipoprotein (HDL) by dimerization. *J. Biol. Chem.* **286**, 38570–38582 (2011).
5. Melchior, J.T. *et al.* An evaluation of the crystal structure of C-terminal truncated apolipoprotein A-I in solution reveals structural dynamics related to lipid binding. *J. Biol. Chem.* **291**, 5439–5451 (2016).
6. Nolte, R.T. & Atkinson, D. Conformational analysis of apolipoprotein A-I and E-3 based on primary sequence and circular dichroism. *Biophys. J.* **63**, 1221–1239 (1992).
7. Barbeau, D.L., Jonas, A., Teng, T. & Scanu, A.M. Asymmetry of apolipoprotein A-I in solution as assessed from ultracentrifugal, viscometric, and fluorescence polarization studies. *Biochemistry* **18**, 362–369 (1979).
8. Okon, M., Frank, P.G., Marcel, Y.L. & Cushley, R.J. Heteronuclear NMR studies of human serum apolipoprotein A-I. Part I. Secondary structure in lipid-mimetic solution. *FEBS Lett.* **517**, 139–143 (2002).
9. Davidson, W.S. *et al.* Structural organization of the N-terminal domain of apolipoprotein A-I: studies of tryptophan mutants. *Biochemistry* **38**, 14387–14395 (1999).
10. Segrest, J.P., Jones, M.K., Shao, B. & Heinecke, J.W. An experimentally robust model of monomeric apolipoprotein A-I created from a chimera of two X-ray structures and molecular dynamics simulations. *Biochemistry* **53**, 7625–7640 (2014).
11. Silva, R.A., Hilliard, G.M., Fang, J., Macha, S. & Davidson, W.S. A three-dimensional molecular model of lipid-free apolipoprotein A-I determined by cross-linking/mass spectrometry and sequence threading. *Biochemistry* **44**, 2759–2769 (2005).
12. Chetty, P.S. *et al.* Helical structure and stability in human apolipoprotein A-I by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **106**, 19005–19010 (2009).
13. Lagerstedt, J.O. *et al.* The "beta-clasp" model of apolipoprotein A-I--a lipid-free solution structure determined by electron paramagnetic resonance spectroscopy. *Biochim. Biophys. Acta* **1821**, 448–455 (2012).
14. Oda, M.N., Forte, T.M., Ryan, R.O. & Voss, J.C. The C-terminal domain of apolipoprotein A-I contains a lipid-sensitive conformational trigger. *Nat. Struct. Biol.* **10**, 455–460 (2003).
15. Phillips, M.C. New insights into the determination of HDL structure by apolipoproteins: Thematic review series: high density lipoprotein structure, function, and metabolism. *J. Lipid Res.* **54**, 2034–2048 (2013).
16. Pollard, R.D., Fulp, B., Samuel, M.P., Sorci-Thomas, M.G. & Thomas, M.J. The conformation of lipid-free human apolipoprotein A-I in solution. *Biochemistry* **52**, 9470–9481 (2013).
17. Zhang, X., Lei, D., Zhang, L., Rames, M. & Zhang, S. A model of lipid-free apolipoprotein A-I revealed by iterative molecular dynamics simulation. *PLoS One* **10**, e0120233 (2015).
18. Swaim, C.L., Smith, J.B. & Smith, D.L. Unexpected products from the reaction of the synthetic cross-linker 3,3′-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. *J. Am. Soc. Mass Spectrom.* **15**, 736–749 (2004).
19. Leavell, M.D., Novak, P., Behrens, C.R., Schoeniger, J.S. & Kruppa, G.H. Strategy for selective chemical cross-linking of tyrosine and lysine residues. *J. Am. Soc. Mass Spectrom.* **15**, 1604–1611 (2004).
20. Walker, R.G. *et al.* The structure of human apolipoprotein A-IV as revealed by stable isotope-assisted cross-linking, molecular dynamics, and small angle x-ray scattering. *J. Biol. Chem.* **289**, 5596–5608 (2014).
21. Franke, D. & Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
22. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77** (Suppl. 9), 114–122 (2009).
23. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
24. Willard, L. *et al.* VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **31**, 3316–3319 (2003).
25. de la Llera-Moya, M. *et al.* The ability to promote efflux via ABCA1 determines the capacity of serum specimens with similar high-density lipoprotein cholesterol to remove cholesterol from macrophages. *Arterioscler. Thromb. Vasc. Biol.* **30**, 796–801 (2010).
26. Vedhachalam, C. *et al.* ABCA1-induced cell surface binding sites for ApoA-I. *Arterioscler. Thromb. Vasc. Biol.* **27**, 1603–1609 (2007).
27. Hassan, H.H. *et al.* Identification of an ABCA1-dependent phospholipid-rich plasma membrane apolipoprotein A-I binding site for nascent HDL formation: implications for current models of HDL biogenesis. *J. Lipid Res.* **48**, 2428–2442 (2007).
28. Rye, K.A. & Barter, P.J. Formation and metabolism of prebeta-migrating, lipid-poor apolipoprotein A-I. *Arterioscler. Thromb. Vasc. Biol.* **24**, 421–428 (2004).
29. Vedhachalam, C. *et al.* Influence of ApoA-I structure on the ABCA1-mediated efflux of cellular lipids. *J. Biol. Chem.* **279**, 49931–49939 (2004).
30. Saito, H. *et al.* Domain structure and lipid interaction in human apolipoproteins A-I and E, a general model. *J. Biol. Chem.* **278**, 23227–23232 (2003).
31. Mei, X., Liu, M., Herscovitz, H. & Atkinson, D. Probing the C-terminal domain of lipid-free apoA-I demonstrates the vital role of the H10B sequence repeat in HDL formation. *J. Lipid Res.* **57**, 1507–1517 (2016).
32. Koyama, M. *et al.* Interaction between the N- and C-terminal domains modulates the stability and lipid binding of apolipoprotein A-I. *Biochemistry* **48**, 2529–2537 (2009).
33. Gorshkova, I.N., Liadaki, K., Gursky, O., Atkinson, D. & Zannis, V.I. Probing the lipid-free structure and stability of apolipoprotein A-I by mutation. *Biochemistry* **39**, 15910–15919 (2000).
34. Gross, E., Peng, D.Q., Hazen, S.L. & Smith, J.D. A novel folding intermediate state for apolipoprotein A-I: role of the amino and carboxy termini. *Biophys. J.* **90**, 1362–1370 (2006).
35. Pollard, R.D., Fulp, B., Sorci-Thomas, M.G. & Thomas, M.J. High-density lipoprotein biogenesis: defining the domains involved in human apolipoprotein A-I lipidation. *Biochemistry* **55**, 4971–4981 (2016).
36. Panagotopulos, S.E. *et al.* The role of apolipoprotein A-I helix 10 in apolipoprotein-mediated cholesterol efflux via the ATP-binding cassette transporter ABCA1. *J. Biol. Chem.* **277**, 39477–39484 (2002).
37. Brouillette, C.G. *et al.* Förster resonance energy transfer measurements are consistent with a helical bundle model for lipid-free apolipoprotein A-I. *Biochemistry* **44**, 16413–16425 (2005).
38. Puchkaev, A.V., Koo, L.S. & Ortiz de Montellano, P.R. Aromatic stacking as a determinant of the thermal stability of CYP119 from Sulfolobus solfataricus. *Arch. Biochem. Biophys.* **409**, 52–58 (2003).
39. Nichols, W.C., Dwulet, F.E., Liepnieks, J. & Benson, M.D. Variant apolipoprotein AI as a major constituent of a human hereditary amyloid. *Biochem. Biophys. Res. Commun.* **156**, 762–768 (1988).
40. Franceschini, G., Sirtori, C.R., Capurso, A. II, Weisgraber, K.H. & Mahley, R.W. A-IMilano apoprotein. Decreased high density lipoprotein cholesterol levels with significant lipoprotein modifications and without clinical atherosclerosis in an Italian family. *J. Clin. Invest.* **66**, 892–900 (1980).
41. Chetty, P.S. Effects of the Iowa and Milano mutations on apolipoprotein A-I structure and dynamics determined by hydrogen exchange and mass spectrometry. *Biochemistry* **51**, 8993–9001 (2012).
42. Gorshkova, I.N. *et al.* Structure and stability of apolipoprotein a-I in solution and in discoidal high-density lipoprotein probed by double charge ablation and deletion mutation. *Biochemistry* **45**, 1242–1254 (2006).
43. Davidson, W.S., Hazlett, T., Mantulin, W.W. & Jonas, A. The role of apolipoprotein AI domains in lipid binding. *Proc. Natl. Acad. Sci. USA* **93**, 13605–13610 (1996).
44. Saito, H. *et al.* Alpha-helix formation is required for high affinity binding of human apolipoprotein A-I to lipids. *J. Biol. Chem.* **279**, 20974–20981 (2004).
45. Tanaka, M. *et al.* Influence of N-terminal helix bundle stability on the lipid-binding properties of human apolipoprotein A-I. *Biochim. Biophys. Acta* **1811**, 25–30 (2011).
46. Pownall, H.J., Massey, J.B., Kusserow, S.K. & Gotto, A.M. Jr. Kinetics of lipid-protein interactions: interaction of apolipoprotein A-I from human plasma high density lipoproteins with phosphatidylcholines. *Biochemistry* **17**, 1183–1188 (1978).
47. Palgunachari, M.N. *et al.* Only the two end helixes of eight tandem amphipathic helical domains of human apo A-I have significant lipid affinity. Implications for HDL assembly. *Arterioscler. Thromb. Vasc. Biol.* **16**, 328–338 (1996).
48. Qian, H. *et al.* Structure of the human lipid exporter ABCA1. *Cell* **169**, 1228–1239. e10 (2017).
49. Segrest, J.P. *et al.* A detailed molecular belt model for apolipoprotein A-I in discoidal high density lipoprotein. *J. Biol. Chem.* **274**, 31755–31758 (1999).
50. Huang, B.X., Kim, H.Y. & Dass, C. Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **15**, 1237–1247 (2004).
51. Jacobsen, R.B. *et al.* Structure and dynamics of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry. *Protein Sci.* **15**, 1303–1317 (2006).
52. Young, M.M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **97**, 5802–5806 (2000).
53. Peng, L., Rasmussen, M.I., Chailyan, A., Houen, G. & Højrup, P. Probing the structure of human protein disulfide isomerase by chemical cross-linking combined with mass spectrometry. *J. Proteomics* **108**, 1–16 (2014).

ONLINE METHODS

Protein expression and purification. Recombinant apoA-I was expressed and purified as previously described[54] with minor modifications. Briefly, a pET30 vector (Novagen) containing mature apoA-I (sequence confirmed by the Cincinnati Children's Hospital Sequencing Core on an Applied Biosystems 3730 DNA Analyzer) was transformed into BL-21 *Escherichia coli* cells. Cells were grown at 37 °C in LB culture media containing kanamycin to select transformants. Protein expression was induced by the addition of IPTG at 0.1 mM followed by shaking at 225 r.p.m. for 2 h at 37 °C. Cells were pelleted by centrifugation and stored at −20 °C. Cells were thawed on ice, resuspended in binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7.9) and lysed at 4 °C by probe sonication. The insoluble cell components were pelleted by centrifugation, and supernatant was applied to His-bind columns (Novagen), which were extensively washed and eluted with elution buffer (1 M imidazole, 0.5 M NaCl, 20 mM Tris, pH 7.9). The His-tag was cleaved by tobacco etch virus (TEV) protease at a ratio of 20:1 for 2 h at room temperature. The sample was reapplied to the His-binding resin to remove uncut protein. Fractions with purified protein were pooled, dialyzed into 10 mM ammonium bicarbonate, pH 8.1, and lyophilized for storage at −80 °C until ready for use. Lyophilized protein was solubilized in a Tris buffer containing 3M guanidine HCl at concentrations below 0.1 mg/ml for 2 h at 4 °C and then exhaustively dialyzed into appropriate buffer for experimentation. Our recombinant protein contains an N-terminal glycine, which does not affect any protein structural or functional index that we have examined[54]. The ^{15}N-labeled version of apoA-I was expressed and purified with minor modifications to the system as previously described[20]. The experimental molecular weight of ^{15}N apoA-I was 28,483.7 Da, determined by direct injection MS. This translated to 342 of 348 successfully labeled nitrogen atoms, which was >98% labeling efficiency.

Mixing of wild-type and isotopically labeled protein. Recombinant ^{14}N and ^{15}N apoA-I were solubilized separately in 3M guanidine in Tris-HCl (1.5 M NaCl, 100 mM Tris-HCl, 10 mM EDTA, 0.2% Azide), pH 8.0, at a concentration below 0.1 mg/ml to ensure complete unfolding and dissociation of higher-order oligomers. The species were then mixed at a 1:1 molar ratio and dialyzed against PBS (140 mM NaCl, 2.7 mM KCl, 10.1 mM Na$_2$HPO$_4$, 1.8 mM KH$_2$PO$_4$, pH 7.4) to refold the protein. The [14/15]N recombinant apoA-I was further purified using size-exclusion chromatography using a Superdex 200 gel filtration column (10/300 GL; GE Healthcare) on a ÄKTA FPLC system (GE Healthcare) in PBS. Fractions containing pure apoA-I were pooled and concentrated to between 1–2 mg/ml for cross-linking. Protein concentration was determined using the Markwell modified Lowry assay[55]. Purity was routinely >99%, as determined by SDS–PAGE and MS.

Cross-linking and isolation of monomeric apoA-I. ApoA-I was cross-linked with bis(sulfosuccinimidyl) suberate (BS3) (Thermo Fisher Scientific) or cyanur-biotin-dimercapto-propionyl-sulfo-succinimide (CBDPS), a homobifunctional, biotin-tagged cross-linking agent (Creative Molecules Inc.)[56]. Samples were cross-linked at molar ratios of 50:1 (BS3) or 20:1 (CBDPS), which were determined based on optimal formation of oligomeric species. The reactions proceeded for 12 h at 4 °C and were quenched by excess Tris-HCl. Monomeric species of apoA-I were isolated by size-exclusion chromatography on three Superdex 200 gel-filtration columns (10/300 GL; GE Healthcare) in series on a ÄKTA FPLC system (GE Healthcare) in PBS at a flow rate of 0.3 ml/min. 0.25-ml fractions were collected and analyzed by SDS–PAGE. Fractions containing purified monomeric apoA-I were pooled, concentrated and quantified. Monomers undergoing MS analysis were dialyzed into 10 mM ammonium bicarbonate, pH 8.1. 50-µg aliquots were digested with sequencing grade trypsin (Promega) at a 1:20 mass ratio of trypsin to protein for 16 h at 37 °C. Peptides were lyophilized to dryness and stored at −20 °C until analysis by MS.

Peptide purification with CBDPS. We used the biotin tag on CBDPS to isolate peptides that had been modified by the cross-linker[56]. ApoA-I was cross-linked with CBDPS-H8/D8 at a molar ratio of 20:1 cross-linker to protein. Spin cups were loaded with 150 µg of Ultralink streptavidin resin (Pierce, 53117) and washed 5× with 300 µl PBS by centrifuging at 4,000*g* for 1 min at RT. 100 µg of digested peptides per 150 µl of slurry were incubated together on an orbital shaker for 1 h at RT. Unlabeled peptides were removed by centrifugation at 4,000*g* for 1 min at RT. The resin was washed once with PBS and then eluted

with two aliquots of 300 µl of 50% acetonitrile, 0.4% trifluoroacetic acid. The peptides were then lyophilized to dryness for MS analysis.

Mass spectrometry and identification of cross-linked peptides. MS analyses were performed as previously described[20]. Briefly, Nano-LC–MS/MS analyses were performed on a TripleTOF 5600+ (AB Sciex, Toronto, Canada) coupled to an Eksigent (Dublin, CA) NanoLC-Ultra nanoflow system. Dried samples were reconstituted in formic acid/H$_2$O 0.1/99.9 (v/v), and 5 µl (~1–2 µg of digest) was loaded onto a C18 IntegraFrit trap column (New Objective, Inc.) at 2 µl/min in FA/H$_2$O 0.1/99.9 (v/v) for 15 min to desalt and concentrate the samples. For the chromatographic separation, the trap-column was switched to align with the analytical column, Acclaim PepMap100 (Dionex-Thermo Fisher Scientific). Peptides were eluted at 300 nL/min using a mobile phase gradient from 95% phase A (FA/H$_2$O 0.1/99.9, v/v) to 40% phase B (FA/ACN 0.1/99.9 v/v) for 35 min (1% per min), then from 40% B to 85% B in 5 min with re-equilibration. Effluent was introduced to the mass spectrometer using a NANOSpray III Source (AB Sciex, Toronto, Canada). The instrument was operated in positive ion mode for 65 min, where each cycle consisted of one TOF-MS scan (0.25 s accumulation time, in a 350 to 1,500 m/z window) followed by 30 information dependent acquisition mode MS/MS-scans on the most intense candidate ions selected from initially performed TOF-MS scan during each cycle. Each product ion scan was operated under vendor-specified high-sensitivity mode with an accumulation time of 0.075 s and CE of 43 with an 8 unit scan range. The .wiff files were converted to Mascot generic files using PeakView v1.2.0.3 software (AB Sciex). Cross-links were identified using SimXL[57] as previously described[5]. A user-defined error of 20 ppm was used for mass detection in all experiments. Two independent experiments were performed for apoA-I cross-linked with CBDPS and three independent experiments were performed for apoA-I cross-linked with BS3. Positive identification of cross-linked peptides required confirmation from the MS/MS fragmentation pattern and visual identification of mass peaks independently by two experienced analysts. Additionally, a unique cross-link had to appear in a minimum of two experiments for inclusion.

Small angle X-ray scattering. SAXS data were collected using the SIBYLS beamline (Berkeley, CA)[58]. Cross-linked monomeric apoA-I samples were shipped overnight at 4 °C within 48 h of isolation to minimize aggregation artifacts. The cross-linked molecules were sampled at 4 mg/ml, 2 mg/ml, and 1 mg/ml at 10 °C with four exposure times; 0.5, 1.0, 2.0, and 5.0 s. Scattering profiles showing evidence of radiation damage or particle repulsion were discarded. ScÅtter (SIBYLS) and ATSAS program suite (EMBL) were used for data analysis. 23 independent *ab initio* molecular envelopes were generated, averaged, and superimposed using DAMMIF (ATSAS, EMBL-Hamburg)[21]. The averaged molecular envelope graphics were rendered using UCSF Chimera.

Iterative model generation and evaluation. A 3D composite homology model of monomeric mutant apoA-I[1–184] was generated from the dimeric crystal structure[4,59] using Modeller v9.14 (ref. 60) as previously described[5]. The C-terminal 59 residues were modeled in incremental segments guided by experimental cross-linking data for respective regions using Modeller v9.14. Models were constrained with previously published cross-links[10,11,16] and newly identified cross-links using an upper bound of 23.0 ± 0.001 Å or 28.0 ± 0.001 Å from C$_\alpha$ to C$_\alpha$ for cross-links identified with MDA and BS3, respectively. The model's primary and secondary structure were adapted and refined iteratively guided by HDX data published by Chetty and colleagues[12]. Models were evaluated against experimental data obtained from SAXS using the online server FoXS[61,62]. The final model was subjected to energy minimization using the YASARA server[22], and rotamers were manually refined using COOT[23]. Given the lack of a density map, rotamer position was determined using the most probable position without resulting in clashes. The reliability of the model was determined using MolProbity[63] and solvent accessibility to individual amino acids was determined using VADAR[24].

Data availability. The coordinates for the model is available as a pdb file in **Supplementary Data Set 1** and can also be downloaded from the Davidson Lab homepage (http://homepages.uc.edu/~davidswm/structures.html). All other source data supporting the findings in this study are available upon reasonable request. A **Life Sciences Reporting Summary** for this article is available.

54. Tubb, M.R., Smith, L.E. & Davidson, W.S. Purification of recombinant apolipoproteins A-I and A-IV and efficient affinity tag cleavage by tobacco etch virus protease. *J. Lipid Res.* **50**, 1497–1504 (2009).

55. Markwell, M.A., Haas, S.M., Bieber, L.L. & Tolbert, N.E. A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal. Biochem.* **87**, 206–210 (1978).

56. Petrotchenko, E.V., Serpa, J.J. & Borchers, C.H. An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Mol. Cell Proteomics* **10**, M110.001420 (2011).

57. Lima, D.B. *et al.* SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis. *J. Proteomics* **129**, 51–55 (2015).

58. Dyer, K.N. *et al.* High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol. Biol.* **1091**, 245–258 (2014).

59. Borhani, D.W., Rogers, D.P., Engler, J.A. & Brouillette, C.G. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation. *Proc. Natl. Acad. Sci. USA* **94**, 12291–12296 (1997).

60. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

61. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).

62. Schneidman-Duhovny, D., Hammel, M., Tainer, J.A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).

63. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

nature research

Corresponding author(s): Davidson, W.S.

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. **Sample size**

   Describe how sample size was determined.

   > While no statistical comparisons were made in this purely structural study, we point out that data from multiple laboratories and multiple protein preparations (native apoA-I and recombinant apoA-I, for example) along with multiple cross-linking agents.

2. **Data exclusions**

   Describe any data exclusions.

   > No data was excluded from the generation of the model. Indeed, we found some cross-links, for example, that did not strictly fit the final model. These were still included in the reporting. In some cases, SAXs samples that showed clear signs of aggregation were repeated with new preps.

3. **Replication**

   Describe whether the experimental findings were reliably reproduced.

   > This model was produced from multiple preparations of native and recombinant apoA-I with two separate cross-linking agents. This study is unique in that it utilized structural data from at least 4 different laboratories. Our comparisons showed the most of the reported chemical cross-links were indeed reproducible among these laboratories. This is despite differences in cross-linking agent, sample processing and mass spectrometry equipment. A strength of this study is that the data was not only produced in a single lab. Thus, by definition, it is reproducible across laboratories

4. **Randomization**

   Describe how samples/organisms/participants were allocated into experimental groups.

   > Not applicable: This structural study did not have an experimental group design.

5. **Blinding**

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   > Not applicable: This structural study did not have an experimental group design.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☒	☐	The <u>exact sample size</u> (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	This work primarily utilized protein structure modeling software platforms that are freely available on the Web. These include Modeler v9.14 (https://salilab.org/modeller/), FoxS (https://modbase.compbio.ucsf.edu/foxs/), YASARA (http://www.yasara.org/homologymodeling.htm), COOT (http://www.ysbl.york.ac.uk/~lohkamp/coot/wincoot.html), and MolProbity (http://molprobity.biochem.duke.edu/). For the cross-linking data analysis, we used SIM-XL, which we helped to develop (http://patternlabforproteomics.org/sim-xl/). For mass spectrometry, we used the manufacturer provided software to analyze the data (Agilent, Qualitative Analysis).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	There are no restrictions on the availability of the materials used here. We are happy to provide recombinant or purified human apoA-I to requesting laboratories (or the expression plasmids). In fact, we have done so for numerous labs over the years.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cells were used here.
b. Describe the method of cell line authentication used.	N/A
c. Report whether the cell lines were tested for mycoplasma contamination.	N/A
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	N/A

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

> No animals were used in the current study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

> Human donors were involved in providing plasma for purification of human plasma apoA-I. These were obtained from a local blood bank (expired units) and were fully de-identified prior to our use. No relevant medical information could be obtained or transmitted to outside entities based on our use of these samples.