Article

The structure of apolipoprotein B100 from human low-density lipoprotein

https://doi.org/10.1038/s41586-024-08467-w

Received: 31 January 2024

Accepted: 28 November 2024

Published online: 11 December 2024

Open access

Check for updates

Zachary T. Berndsen[1]✉ & C. Keith Cassidy[2]✉

Low-density lipoprotein (LDL) has a central role in lipid and cholesterol metabolism and is a key agent in the development and progression of atherosclerosis, the leading cause of mortality worldwide[1,2]. Apolipoprotein B100 (apoB100), one of the largest proteins in the genome, is the primary structural and functional component of LDL, yet its size and complex lipid associations have posed major challenges for structural studies[3]. Here we present the structure of apoB100 resolved to subnanometre resolution in most regions using an integrative approach of cryo-electron microscopy, AlphaFold2[4] and molecular-dynamics-based refinement[5]. The structure consists of a large globular N-terminal domain and an approximately 61-nm-long continuous amphipathic β-sheet that wraps around the LDL particle like a belt. Distributed quasi-symmetrically across the two sides of the β-belt are nine strategically located interstrand inserts that extend across the lipid surface to provide additional structural support through a network of long-range interactions. We further compare our structure to a comprehensive list of more than 200 intramolecular cross-links and find close agreement between the two. These results suggest a mechanism for how the various domains of apoB100 act in concert to maintain LDL shape and cohesion across a range of particle sizes. More generally, they advance our fundamental understanding of LDL synthesis, form and function, and will help to accelerate the design of potential therapeutics.

Lipoproteins (LPs) are complex and heterogeneous macromolecular nanoparticles that facilitate the transport of fat-soluble molecules through aqueous extracellular fluids[1,6]. They consist of a monolayer of phospholipid and free cholesterol with embedded apolipoproteins, all surrounding a core of primarily triglycerides (TGs) and cholesteryl esters (CE)[1]. ApoB100, an approximately 550 kDa glycoprotein, is the main structural and functional component of all non-high-density LPs, including very-low-density lipoprotein (VLDL), intermediate density lipoprotein (IDL), LDL and lipoprotein(a)[1,7]. Most apoB100 is secreted by the liver bound to large (around 30–80 nm) TG-rich VLDLs, which contain other exchangeable apolipoproteins such as apoE and apoC[1]. Once secreted, VLDL undergoes a dynamic catabolic journey, delivering its TG, CE and apolipoprotein cargo to peripheral tissue and other LPs. It is sequentially catabolized first to IDL and then to small (around 18–30 nm), TG-poor, CE-rich LDL, which is cleared from circulation by LDL receptor (LDLR)-mediated endocytosis[1]. During this catabolic conversion, a single copy of apoB100 remains bound to the LP particle to coordinate key intermolecular interactions and maintain LP structure, requiring apoB100 to be both mechanically robust and conformationally dynamic.

Beyond their metabolic functions, apoB100-containing LPs, particularly LDL, are key agents in the development and progression of atherosclerosis, the leading cause of mortality worldwide[2]. According to the cumulative exposure hypothesis, the progression of atherosclerosis is proportional to the circulating LDL concentration and duration of exposure, making LDL reduction a therapeutic focus[8]. Although diet and lifestyle are the primary drivers of dyslipidaemia, there are numerous mutations in the *APOB* gene that can disrupt proper LP synthesis and metabolism, with some leading to early-onset disease and mortality[9]. Determining the structure of apoB100 promises to reveal important insights in the molecular mechanisms of these disorders while providing a foundation for the development of novel lipid modulating therapeutics.

Despite its importance, apoB100 has long resisted structural characterization owing to its size and complexity as well as the extensive heterogeneity in LP preparations[10–16]. The past decade has seen the emergence of cryo-electron microscopy (cryo-EM) as a viable approach for solving high-resolution membrane protein structures[17]. In recent years, it has also become possible to extend the molecular details available from mid-resolution cryo-EM maps using the artificial intelligence encoded in programs such as AlphaFold[4], which are often applied in tandem with structural refinement techniques involving molecular dynamics (MD) simulation[18,19]. Thus, with improvements in sample preparation and extensive computational sorting and classification of cryo-EM images, along with a robust modelling and flexible refinement protocol, a detailed understanding of the apoB100 molecular structure has now been achieved.

Cryo-EM reconstruction of human LDL

In preparation for structural characterization, LDL isolated from human serum by ultracentrifugation was purchased from a commercial

[1]Department of Biochemistry, University of Missouri, Columbia, MO, USA. [2]Department of Physics, University of Missouri, Columbia, MO, USA. ✉e-mail: zberndsen@missouri.edu; ckcassidy@missouri.edu

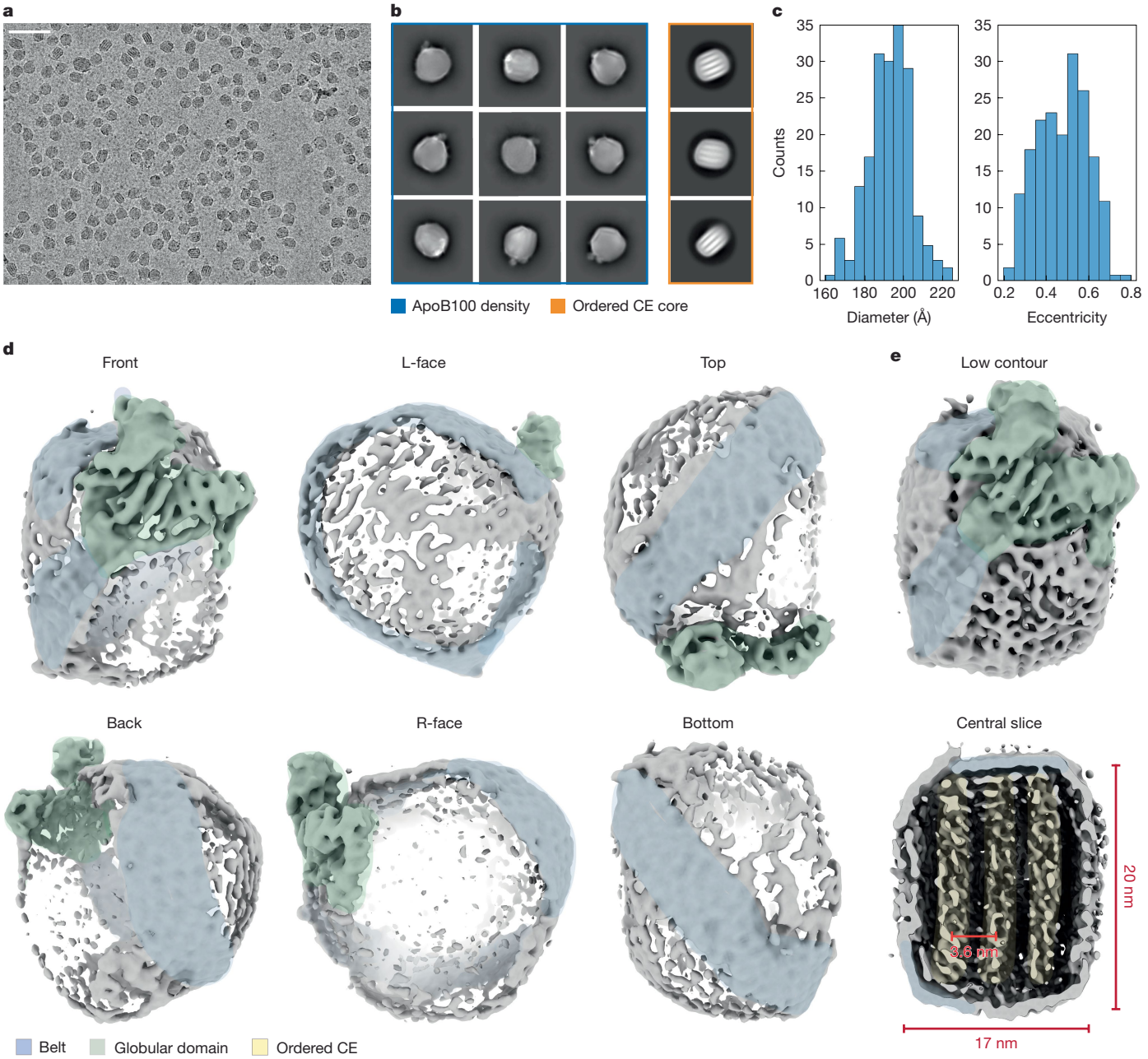


Fig. 1 | Cryo-EM reconstruction of human LDL. a, Representative cryo-electron micrograph showing LDL particles. Scale bar, 80 nm. **b**, Representative 2D class averages showing a mixture of classes with clear apoB100 density around the particle exterior (blue) and particles with ordered CE cores (orange). **c**, LDL particle diameter and eccentricity for 200 2D class averages. **d**, Three-dimensional reconstruction of the LDL particle with the globular domain and belt domain highlighted in green and blue, respectively, viewed from six different directions at a high contour level. **e**, Low-contour-level cryo-EM map viewed from the 'front' along with a central cross-section. Particle dimensions and spacing between the CE planes are indicated.

vendor and further purified by size-exclusion chromatography (SEC) (Extended Data Fig. 1a). In addition to apoB100, gel electrophoresis revealed several smaller bands possibly corresponding to other exchangeable apolipoproteins as well as albumin, most of which were excluded by the SEC step (Extended Data Fig. 1b,c). The slowest eluting approximate quarter of the LDL peak was pooled and prepared for cryo-EM imaging (Extended Data Fig. 1a). The rationale for selecting the smallest apoB100-containing LP particles from the tail end of the LDL size range was fourfold: (1) the initial set of particle images is more uniform, requiring less computational sorting; (2) smaller particles have a higher protein-to-lipid ratio; (3) the smaller diameter and lower lipid content results in less signal attenuation through the particle; and (4) apoB100 will be more compact on the surface, presumably leading to less dynamic motion and heterogeneity. We collected around 3.6×10^3 micrographs and selected about 600,000 LDL particles for further processing (Fig. 1a, Extended Data Fig. 2a and Extended Data Table 1). The particle diameters ranged from 16.2 nm to 22.4 nm (mean, 19.3 nm), suggesting that the larger LDL species were successfully excluded, and their mean eccentricity was 0.48 (Fig. 1c and Extended Data Fig. 3). Of these particles, around 30% contained ordered stacks of high-density features in their core corresponding to the liquid crystalline phase of CE[10,20] (Fig. 1b and Extended Data Fig. 2a). These particles were excluded from the dataset after two-dimensional (2D) classification as they biased the alignment unfavourably towards the core and away from the weaker peripheral features. A subset of particles that showed high-intensity features on their surface, interpreted as apoB100, were

selected for further processing (Fig. 1b and Extended Data Fig. 2a). Extensive 2D and three-dimensional (3D) classification resulted in a final reconstruction of the entire LDL particle resolved around 6–14 Å (global resolution of about 9 Å) from around 53,000 particle images (Fig. 1d, Extended Data Figs. 2a,b and 4a and Supplementary Video 1). At high contour levels, the map reveals a well-resolved (Extended Data Fig. 4a) globular protein domain connected to a continuous belt of high density that wraps around the particle's circumference (Fig. 1d). This belt divides the particle into two sides that when viewed from the 'front' can be labelled the left (L) and right (R) faces (Fig. 1d). At lower contours, additional features appear on the surface, until ultimately most of the exterior and interior are fully occupied (Fig. 1e). The large-scale morphology of the particle is semi-discoidal with polygonal features and dimensions of around 17 nm × 20 nm (mean diameter, 18.5 nm) (Fig. 1e), placing it within the small dense LDL subclass[21], near the smallest LDL particle size recorded (Fig. 1c and Extended Data Fig. 3). The core is mostly disordered but retains a hint of liquid-crystalline CE packing, with the planar stacks oriented perpendicular to the curved face of the particle (Fig. 1e). Alternative reconstructions without ordered CE were also obtained, yet no appreciable differences in peripheral density features were observed (Extended Data Fig. 2d). Lastly, we generated a soft mask around the globular domain for focused classification and refinement, achieving a reconstruction with a resolution of around 5–7 Å (global resolution of 5.8 Å) in this region (Extended Data Figs. 2c and 4c).

Atomic model of apoB100

We next endeavoured to construct an all-atom model of full-length apoB100 guided by our cryo-EM data. Given the lack of existing structural data, AlphaFold2 (AF2)[4] was used to predict the complete 4,536-residue structure as three contiguous fragments, which each exhibited reasonably high predicted local distance difference test (pLDDT) confidence scores overall (Extended Data Fig. 5b,d). When combined, the structure could be divided grossly into two sections: (1) a globular N-terminal domain (NTD) around 1,000 residues in size; and (2) a large C-terminal domain (CTD) consisting of a continuous β-sheet (approximately 61 nm in length by 4 nm width) that we term the β-belt, interspersed with nine interstrand inserts ranging in length from around 30 to 700 residues (Fig. 2a,b and Extended Data Fig. 6). The overall topology of the NTD and β-belt structures clearly matched the well-resolved regions of density in the cryo-EM map (Fig. 1d and Extended Data Fig. 7a,b). However, in the absence of the lipid particle, the predicted AF2 models were collapsed into compact structures inconsistent with our cryo-EM data (Extended Data Fig. 5c). We therefore used MD flexible fitting[5] (MDFF) to refine the fragment conformations sequentially using a cascading approach starting from lower-resolution Gaussian-filtered maps followed by progressive refinement into higher-resolution maps, beginning with the NTD and then extending the model from the N to C terminus into the β-belt density (Supplementary Video 2). These targeted regions matched our data very well as evidenced by a high corresponding map-to-model cross-correlation (Extended Data Fig. 7a,b) and provided a solid basis for interpreting the nine more sparsely resolved inserts. Of these, the small inserts, 3, 5 and 7 (around 30–40 residues in length), which are predicted to be mostly unstructured, required little modification to account for the surrounding density, and the intermediate-sized inserts, 1, 2, 6 and 8 (around 90–170 residues in length), primarily required the use of MDFF to 'flatten' them out onto the particle surface. The two largest inserts, 4 and 9, which are 695 and 493 residues long, respectively, required a more sophisticated fitting strategy. Both are predicted to form primarily strings of amphipathic helices that fold back on themselves into pairs separated by flexible loops, akin to the helical apolipoproteins such as apoA-1 (refs. 22,23) (Extended Data Figs. 6a and 8a,b). Despite the helical bundles being largely folded in on themselves in the AF2 predictions, several prominent regions of density in the cryo-EM map suggested that both inserts extend along the particle faces to potentially mediate long-range interactions (Extended Data Fig. 7c,d). We therefore used a series of distance-based biasing potentials to guide inserts 4 and 9 along the distinctive paths observed in our map, followed by local MDFF refinement. This enabled the systematic fitting of these extended inserts while preserving their overall topology and predicted secondary structure (Fig. 2b and Extended Data Fig. 6b). Owing to reduced resolution in certain regions of the map (Extended Data Fig. 4a), only truncated fragments of inserts 2, 4, 6 and 9 are displayed in Fig. 2b; however, the results of our full modelling efforts are presented in Extended Data Fig. 6 and discussed in the Supplementary Information. To further validate our model, we compared it to an extensive list of more than 200 intramolecular cross-links and observed close agreement between the two (Supplementary Information, Supplementary Tables 1–8 and Supplementary Figs. 2–4). Our results confirm much of the previous research while substantially expanding upon the prevailing consensus models (Supplementary Information), establishing the quintessential structural characteristics of full-length apoB100 in its native LP environment (Fig. 2b and Extended Data Fig. 6b).

The apoB100 NTD

The NTD can be divided into three subdomains that we term (1) the β-barrel (residues 28–319); (2) the helical bundle (residues 320–634); and (3) the baseplate (residues 635–1010) (Fig. 2a–c). Residue 1010 was chosen to mark the end of the NTD because it corresponds to a small loop that separates the baseplate from the C-terminal β-belt. The first 27 residues correspond to a signal peptide that is cleaved from the mature protein[7]; however, these residues are still included in our numbering. The β-barrel, which projects away from the surface of the particle at around 35°, is composed of 11 strands and 3 helices, with one plugging the core of the β-barrel and another contacting the lipid surface (Fig. 2d). The helical bundle is composed of 17 short helices and its arrangement can be described as a super-helical right-handed coiled-coil with a two-helix repeating unit[24]. It lies partially on top of the baseplate with only the central section contacting the lipid surface (Fig. 2d). The baseplate lies underneath the β-barrel and helical bundle and is composed of two amphipathic β-sheets separated by an amphipathic helix (Fig. 2d). The first β-sheet has a 75-residue insert between strands 4 and 5 composed of 3 amphipathic helices that project down from the baseplate. The NTD contains seven out of the eight disulfide bonds in apoB100: four in the β-barrel, two in the helical bundle and one in the baseplate (Extended Data Fig. 9a).

Homology with the LLTP superfamily

ApoB100 belongs to the large lipid transfer protein (LLTP) superfamily along with vitellogenin/lipovitellin, which is the main apolipoprotein of egg yolk, and the chaperone microsomal triglyceride transfer protein (MTP), among others[25,26]. Both lipovitellin and MTP share around 30–40% sequence similarity with the NTD and both have been previously crystallized, displaying similar conformations[24,27,28]. At the domain level, our structure largely confirms previous homology models of the NTD based on these crystal structures[29,30]; however, it exhibits a unique conformation not yet seen in this class of proteins. Comparison of the relaxed NTD structure with lipovitellin and MTP yielded root mean squared deviation values of around 35 Å (Extended Data Fig. 9b), with the largest deviations occurring in the helical bundle and baseplate (Fig. 2e). In both lipovitellin and MTP, these regions exhibit a much tighter radius of curvature, which creates a lipid-binding cavity that is critical to their function. Consistent with this observation, the initial AF2 prediction also exhibited a tighter radius of curvature, which relaxed into a more planar conformation during refinement (Fig. 2e and

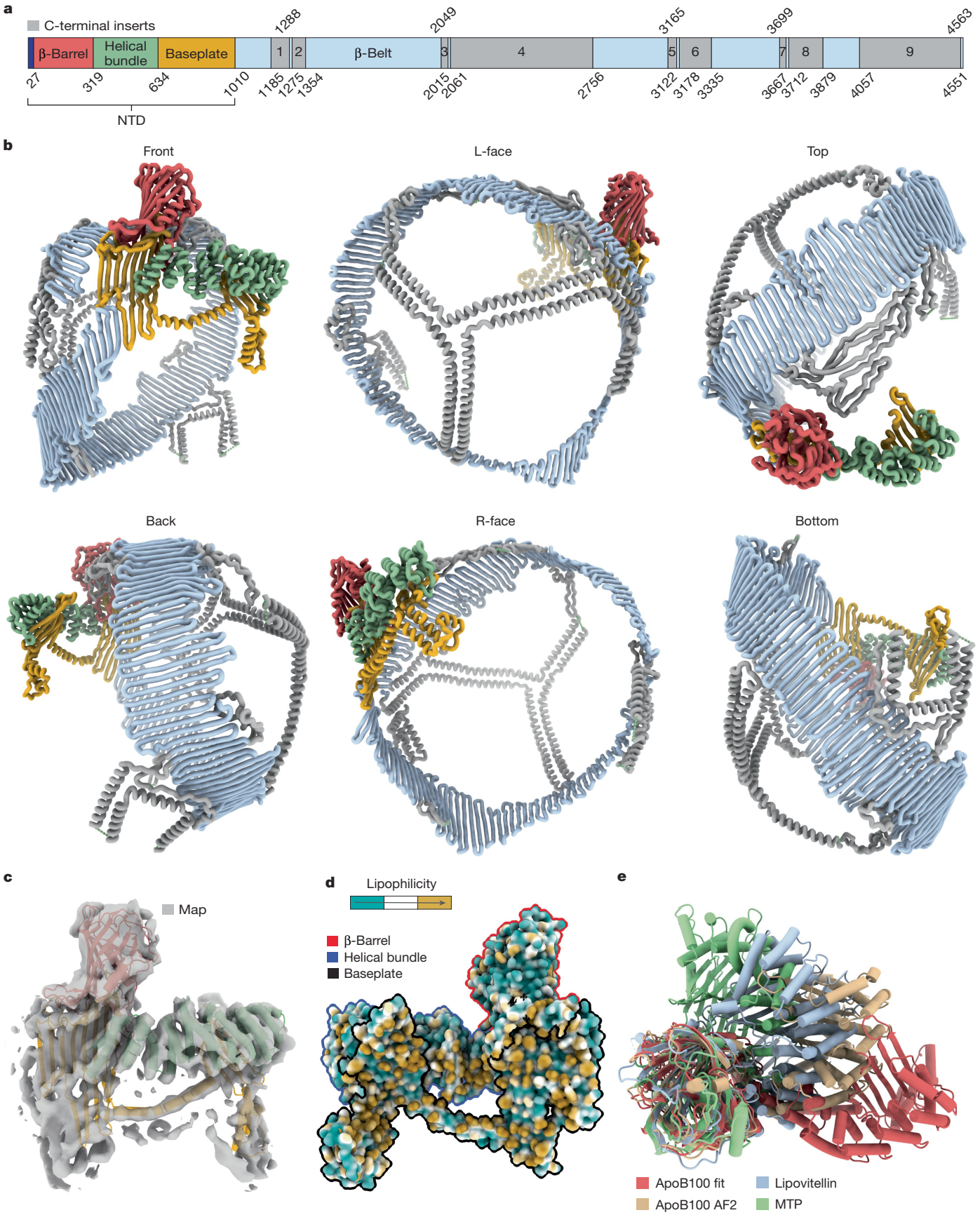


Fig. 2 | Atomic model of apoB100. a, Gene diagram of apoB100. **b**, Atomic model of apoB100 viewed from six different directions and coloured according to the gene diagram in **a**. Truncated domains are marked by green pseudobonds. **c**, Focused refinement of the NTD with relaxed atomic model viewed from the front. **d**, Atomic surface of the NTD viewed from the back and coloured by lipophilicity with the baseplate outlined in black, the helical bundle in blue and the β-barrel in red. **e**, Structural alignment of the relaxed NTD structure, the unrelaxed AF2 predicted structure and the crystal structures of lipovitellin (PDB: 1LSH) and MTP (PDB: 6I7S).

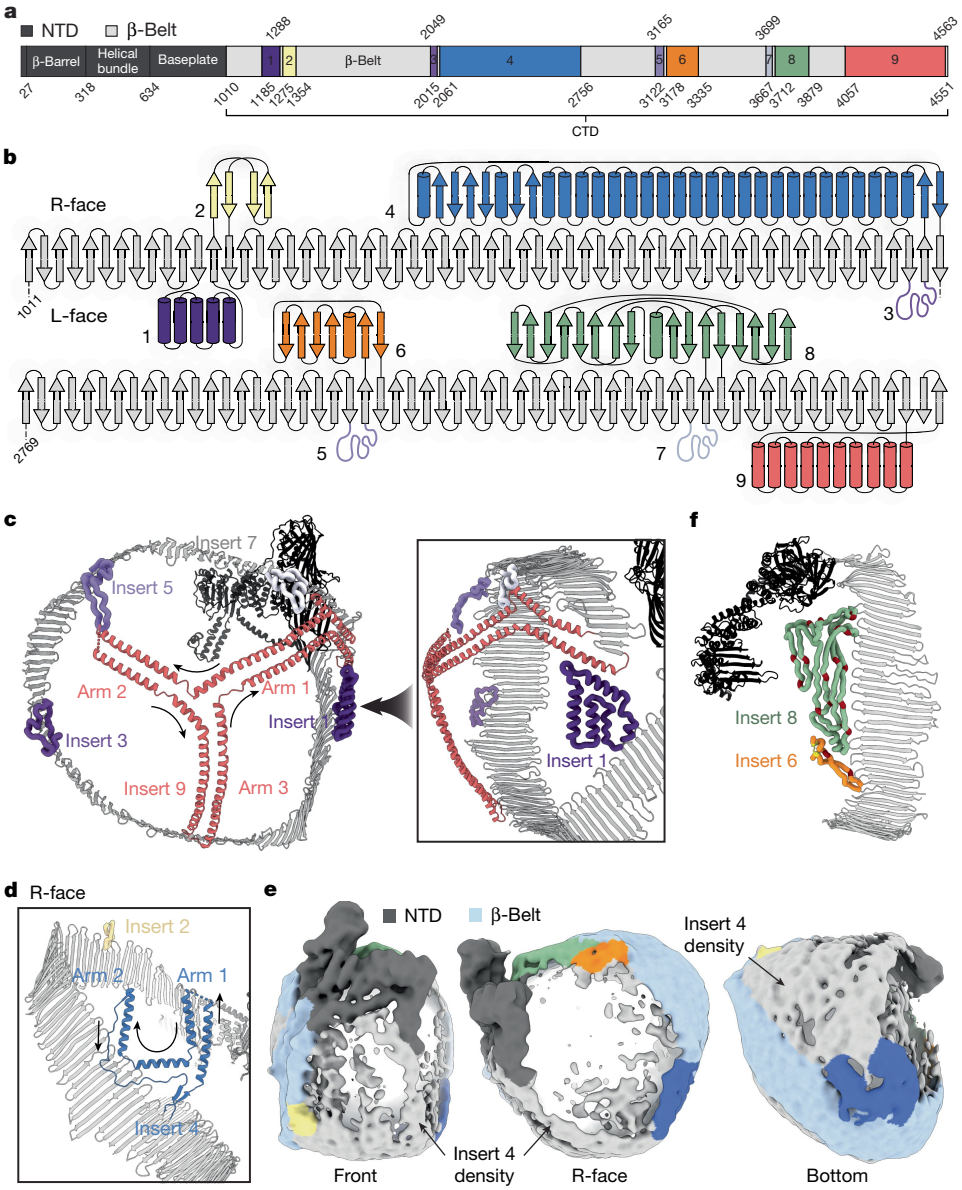


Fig. 3 | ApoB100 C-terminal inserts. a, Reproduction of the apoB100 gene diagram from Fig. 1 showing the NTD in dark grey, C-terminal β-belt in light grey and each interstrand insert coloured separately. **b**, Protein topology diagram of the CTD showing the continuous 120-strand β-belt and interstrand inserts with the L- and R-faces of the particle indicated. **c**, View of the particle L-face with insert 9 displayed as ribbons and inserts 1, 3, 5 and 7 as ropes. The black arrows point in the N to C direction of the protein chains. The location of the truncation in arm 2 of insert 9 is indicated by a pseudobond. **d**, Insert 4 depicted as ribbons and insert 2 depicted as a rope with the domain truncations indicated by pseudobonds. **e**, Gaussian-filtered cryo-EM map showing the particle R-face. Map density corresponding to the β-belt is coloured light blue, the NTD is coloured dark grey and the R-face inserts are coloured as in **a** and **b**. The arrows indicate lower-resolution density corresponding to the truncated segments of insert 4. **f**, The truncated structure of insert 6 and the full structure of insert 8 are depicted as ropes with the proline residues and disulfide bond coloured maroon and yellow, respectively.

Extended Data Fig. 9b). This predicted flexibility is further illustrated when the structure is segmented by the AF2 predicted aligned error (PAE) matrix[4], which divides the NTD down the centre of the helical bundle and baseplate (Extended Data Fig. 9c). Closer inspection of 3D classification results revealed some flexion in the NTD about this axis (Supplementary Video 3), and anisotropic network modelling produced a lowest-energy mode corresponding precisely to this same motion (Extended Data Fig. 9d), suggesting that this motion is intrinsic to the structure. It is therefore plausible that the NTD of apoB100 could also adopt a more curved conformation with a lipid-binding cavity, for example, during the early stages of protein synthesis and particle formation, as previously recognized[27,29,31,32], but which can flatten out as the particle matures.

The β-belt

The CTD of apoB100 can be subdivided into the β-belt (1,774 residues) and 9 interstrand inserts (1,779 residues) (Fig. 3a,b). The β-belt is the largest domain and is highly conserved (Extended Data Fig. 10). It is composed of 120 strands between 3 and 16 residues long (mean, 11), with an average width of around 4 nm (Fig. 3b). After looping around the particle, the β-belt passes behind the NTD on the L-face (Fig. 2b). Although a few potential interactions are observed between the NTD and β-belt, as well as between the NTD and insert 8, these regions do not mesh tightly together and appear largely distinct within our cryo-EM map (Extended Data Fig. 7e). In light of this, we propose a mechanism whereby the β-belt can either tighten or loosen around the particle

surface, sliding past the NTD when the end-to-end length of apoB100 exceeds the circumference of the particle. In addition to possessing this necessary flexibility, the amphipathic β-sheets of apoB100 are known to be critical structural domains that function as sites for lipid incorporation during particle synthesis[33,34] and are strongly associated with the hydrophobic core of particle[3,35]. This is consistent with the approximately 55,720 Å^2 of buried surface area observed in our structure. Along these lines, our cryo-EM data contain some evidence of coordinated lipids on the inner face of the β-belt (Extended Data Fig. 2d). Taken together, these results suggest that the β-belt is the primary structural domain of apoB100 responsible for maintaining LP shape and cohesion.

The CT inserts are arranged in pairs

Interspersed within the β-belt are the 9 interstrand inserts, which are composed of approximately 60% amphipathic α-helices, 30% coils and 10% amphipathic β-sheets (Fig. 3b). They are on average less conserved than the other domains in apoB100 (Extended Data Fig. 10). The smallest inserts, 3, 5 and 7, are mostly unstructured, inserts 1 and 9 are predominantly helical, inserts 2 and 8 are predominantly β-sheets, and inserts 4 and 6 are a mixture of both. Five of the inserts project onto the L-face of the particle and four project onto the R-face (Fig. 3b), and both faces contain one large helical insert (inserts 4 and 9) along with several smaller ones. Moreover, the first eight inserts are arranged in relatively evenly spaced pairs separated by a single β-strand such that they project onto opposite faces (Fig. 3b), suggesting a possible preference for symmetry across the two faces. Despite this paired arrangement, there exists a relative imbalance in both the size and structure of the inserts, suggesting that the two faces of the particle may have unique properties.

L-face inserts

The L-face of the particle contains the small unstructured inserts, 3, 5 and 7, the intermediate-sized helical insert, 1, and the large helical insert, 9. Insert 9, the second largest insert, is located at the C-terminal tip of the β-belt, originating on the third to last strand (Fig. 3b). The L-face of our cryo-EM maps shows an intermediate-density feature loosely in the shape of a three-point star with roughly linear segments separated by around 120° and large areas of low map density situated between the segments (Extended Data Fig. 7c). AF2 predicts that insert 9 folds into a trefoil with three arms composed of pairs of amphipathic helices separated by small breaks; however, the lipophilic surfaces are erroneously collapsed into a larger helical bundle (Extended Data Fig. 8a). Once relaxed into the map, insert 9 clearly fit the distinct star-shaped density feature just mentioned (Extended Data Fig. 7c), with the three arms meeting in the centre, creating the domed appearance of the L-face (Fig. 3c). Arm 1, which is formed by the first 2 and last 2 helices of insert 9, is initially oriented parallel to the β-belt, then turns inward at its 'elbow', contacting insert 7 (Fig. 3c and Extended Data Fig. 8a). The conformation adopted by the proximal segment of arm 1 and the upstream segment of the β-belt, including insert 7, fits into a well-resolved region of the cryo-EM and is predicted with relatively high confidence by AF2 (Extended Data Figs. 7c and 8a), requiring minimal adjustment during MDFF refinement, suggesting that it may represent a low-energy conformation. In addition to contacting insert 7, the bottom edge of arm 1 also contacts insert 1, which is situated just below the NTD (Fig. 3c). Arm 2 then extends toward the back of the particle, making potential contact with insert 5, which adopts an extended loop conformation (Fig. 3c). Owing to the reduced resolution in this region, the remainder of arm 2 beyond the elbow is excluded from the model presented in Fig. 3c; however, the full structure of insert 9 is presented in Extended Data Fig. 11. Finally, arm 3, which is composed of two continuous helices, extends downward toward the edge of the β-belt near residue 1719 (Fig. 3c). Taken together, our structure reveals that insert 9 is the main structural domain of the L-face, and the smaller inserts are positioned strategically to mediate long-range interactions with the three arms.

R-face inserts

The R-face of the particle contains the intermediate-sized inserts, 2, 6 and 8, and the largest insert, 4 (Fig. 3b). Insert 4 is predicted by AF2 to form an extended prong-shaped appendage with two highly articulated arms (Extended Data Fig. 8b). Arm 1 is the longest and is composed of numerous helical pairs separated by flexible linkers, whereas arm 2 is shorter and features a large β-sheet domain at its tip (Extended Data Fig. 8b). As with insert 9, the AF2 predicted structure of insert 4 is also erroneously collapsed onto itself, requiring extensive tertiary adjustment during MDFF refinement. Owing to the reduced resolution in this region (Extended Data Fig. 4a), the distal portions of both arms are excluded the model in Fig. 3d, but we present and discuss our full model in Extended Data Fig. 11 and the Supplementary Information. However, the base of insert 4, where the two arms meet near the edge of the β-belt, fits well into a prominent density feature in the map (Extended Data Fig. 7d). As with the proximal segment of arm 1 in insert 9, the conformation of insert 4 in this region is predicted by AF2 with relatively high confidence and required minimal adjustment during MDFF refinement (Extended Data Fig. 8b). From the structure of this segment, we see that both arms project upwards toward the NTD (Fig. 3d and Extended Data Fig. 11b). The R-face of the particle features a large flat low-density surface with distinct higher-density boundaries that differs from the domed shape of the L-face (Fig. 3e). Our full model suggests that this edge is formed by arm 1 of insert 4, which extends towards the NTD to contact the helical extension projecting from the bottom of the baseplate (Extended Data Fig. 11b), which is further supported by the chemical cross-linking data (Supplementary Information). The conformation of insert 2, which is positioned below the NTD, is also ambiguous; however, our initial modelling efforts (Extended Data Fig. 11b) as well as the chemical cross-linking data (Supplementary Information) suggest that it extends into the R-face, perpendicular to the β-belt, and interacts extensively with the compact β-sheet domain of arm 2. As with insert 9 on the L-face, we propose that insert 4 is the main structural domain of the R-face.

The last two R-face inserts, 6 (156 residues) and 8 (166 residues), occupy a region near the NTD in our model (Fig. 3b,f). Overall, they are similar in size and structure and are composed primarily of amphipathic β-hairpins enriched in proline residues, similar to the β-sheet domain of arm 2 in insert 4 (Fig. 3f and Extended Data Fig. 11c). Insert 6 contains the only other disulfide bond outside the NTD (Cys3194–Cys3324), which links the ascending and descending strands near the base of the domain (Fig. 3f and Extended Data Fig. 9a). Between the two, insert 8 is better resolved and fits clearly into a high-density feature positioned directly behind the NTD (Extended Data Fig. 7e). Its structure roughly matches the shape of the top boundary of the NTD, suggesting that it may serve as a buffer between the NTD and β-belt (Fig. 3f). Insert 6 on the other hand is the most poorly resolved insert in our cryo-EM map (Extended Data Figs. 4a and 7e), implying that it is dynamic, poorly folded and/or loosely associated with the lipid surface. Along these lines, this region exhibited lower (around 40–50) pLDDT scores (Extended Data Fig. 5d), alternative conformations among the top 5 ranking models (Extended Data Fig. 8c) and extensive intradomain cross-linking (Supplementary Information). Consistent with our data, the regions corresponding to inserts 6 and 8 were previously identified based on their high proline content and susceptibility to protease digestion[36], suggesting that these regions, in particular insert 6, may form compact structures that are detached from the particle surface.

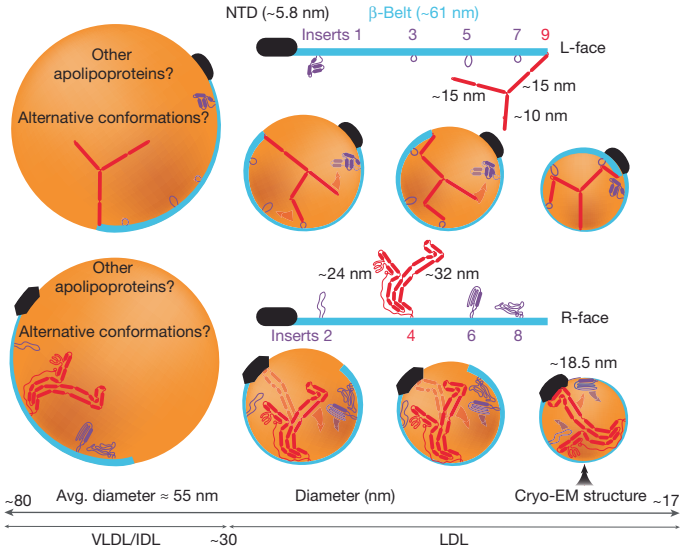


Fig. 4 | Model for apoB100 conformational change. A cartoon representation of different-sized apoB100-containing LPs drawn roughly to scale, from VLDL/IDL to LDL, depicting the possible conformational changes that would occur in apoB100 on the L-face (top) and R-face (bottom) of the particle. Transparent arrows and protein domains represent possible domain motions and alternative conformations.

Discussion

ApoB100-containing LPs vary substantially in size, shape and chemical composition; from the largest VLDLs (approximately 80 nm diameter) to the smallest LDL (approximately 17 nm diameter); however, the mechanism by which apoB100 accommodates this heterogeneity is unclear. Our complete model of apoB100 (Extended Data Figs. 6 and 11), along with the supporting chemical cross-linking data (Supplementary Information), suggests a mechanism for how the various domains might act together to maintain particle shape and cohesion across the full range of sizes (Fig. 4). We propose that the following four structural changes occur in unison: (1) the NTD can curl up or flatten out to match the curvature of the particle surface; (2) the β-belt can tighten and loosen, sliding past the NTD on the L-face when the particle circumference is smaller than the end-to-end length of apoB100; (3) insert 9 on the L-face can expand or compress by bending arms 1 and 2 about their flexible elbows while maintaining additional points of contact with inserts 1, 3, 5 or 7; and (4) insert 4 on the R-face can compress or extend by bending about its numerous flexible linkers all while maintaining its long-range contacts with the NTD and/or insert 6. Finally, when modelling the possible conformation of apoB100 on larger LPs such as VLDL, it is clear apoB100 would have less of an effect on the overall particle morphology and lose the ability to form many of the long-range interactions between inserts we observe here (Fig. 4). In support of this, no long-range chemical cross-links were observed within apoB100 on VLDL particles[30].

Although diet and lifestyle factors are the primary drivers of dyslipidaemia, LP levels can also be affected by genetic mutations in the *APOB* gene[9]. Some cause elevated LP levels and early onset atherosclerosis, as seen in familial hypercholesterolaemia (FH) and familial defective apoB100 (FDB), and others lead to abnormally low levels, as in familial hypobetalipoproteinaemia (FHBL)[37,38]. In the case of FH, the R3527Q mutation is the most common and widely studied and is known to reduce the affinity of apoB100 for the LDLR[38]. As such, it is widely believed that Arg3527 must either be within or near the receptor binding site (RBS)[3]. As a preliminary experiment, we used the newly released AlphaFold3 (ref. 39) server to model the potential complex of LDLR and apoB100 in this region, which predicts that Arg3527 is indeed part of the binding site, making direct interactions with one of the ligand-binding modules of LDLR (Supplementary Information and Supplementary Fig. 5). Moreover, the majority of other FH- and FDB-associated mutations also fall within this same stretch of β-belt[9] (Extended Data Fig. 12). Two other rare FH-associated mutations map to insert 9, suggesting that disruption of its structure or long-range interactions could affect receptor binding by indirectly altering the conformation of the receptor binding site (Extended Data Fig. 12). In the case of FHBL, mutations in the *APOB* gene result in premature stop codons that create C-terminally truncated proteins of different sizes[40]. It has been shown that LPs containing apoB89, a common FHBL mutant resulting in the loss of most of insert 9 (Extended Data Fig. 12), are more rapidly cleared than wild-type particles, leading to the characteristic hypolipidaemia[41]. It was hypothesized that the loss of the C terminus may lead to enhanced presentation of the RBS[41,42]. Considering our structure, it is plausible that the loss of insert 9 could affect the conformation of the β-belt in a way that alters the presentation of the RBS.

There are several limitations of the results presented here. First, owing to the extreme heterogeneity of LP preparations, the dynamic nature of apoB100 and its exclusive location on the periphery of the LP particle, achieving near-atomic resolution of apoB100 is very challenging. Even with additional biochemical homogenization, our final reconstruction was derived from less than 10% of the total particle images due to the need for extensive computational homogenization. Much larger datasets will therefore be necessary to improve the resolutions we obtained here, although this is not guaranteed. As such, our modelling, particularly of the C-terminal inserts, relied heavily on the assumption of accuracy in the AF2 predictions of secondary and local tertiary structure. Although AF2 boasts state-of-the-art accuracy, it may perform less reliably on proteins that interact extensively with lipids. ApoB100 is the primary protein component of VLDL, IDL and LDL, which vary in their lipid composition; we therefore cannot rule out that structural changes may occur with changes in lipid composition. However, the strong agreement between our cryo-EM map (Extended Data Fig. 7a,b), particularly within the well-resolved NTD and β-belt domains (Extended Data Fig. 4a), as well as the close agreement between the predicted local tertiary structure and chemical cross-linking data (Supplementary Information), lends strong support to the accuracy of the AF2 predictions.

The results presented here will enable highly precise genetic, structural, functional and computational studies of apoB100-containting LPs. This newly acquired knowledge will not only advance our understanding of LP synthesis and metabolism but also accelerate the design of potential next-generation LDL-modulating therapies for the treatment and prevention of atherosclerosis, such as monoclonal antibodies that can target LDL, or even specific subclasses of LDL directly, to achieve more controlled treatment with potentially fewer side effects.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08467-w.

1. Feingold, K. R. *Endotext: Introduction to Lipids and Lipoproteins* (eds Feingold K. R. et al.) (MDText, 2024).
2. Linton, M. F. et al. *Endotext: The Role of Lipids and Lipoproteins in Atherosclerosis* (eds Feingold, K. R. et al.) (MDText, 2019).
3. Segrest, J. P. et al. Structure of apolipoprotein B-100 in low density lipoproteins. *J. Lipid Res.* **42**, 1346–1367 (2001).
4. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
5. Trabuco, L. G. et al. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
6. Berneis, K. K. & Krauss, R. M. Metabolic origins and clinical significance of LDL heterogeneity. *J. Lipid Res.* **43**, 1363–1379 (2002).

7. Chen, S. H. et al. The complete cDNA and amino acid sequence of human apolipoprotein B-100. *J. Biol. Chem.* **261**, 12918–12921 (1986).
8. Ference, B. A., Braunwald, E. & Catapano, A. L. The LDL cumulative exposure hypothesis: evidence and practical applications. *Nat. Rev. Cardiol.* **21**, 701–716 (2024).
9. Burnett, J. R. & Hooper, A. J. Common and rare gene variants affecting plasma LDL cholesterol. *Clin. Biochem. Rev.* **29**, 11–26 (2008).
10. Orlova, E. V. et al. Three-dimensional structure of low density lipoproteins by electron cryomicroscopy. *Proc. Natl Acad. Sci. USA* **96**, 8420–8425 (1999).
11. Sherman, M. B. et al. Structure of triglyceride-rich human low-density lipoproteins according to cryoelectron microscopy. *Biochemistry* **42**, 14988–14993 (2003).
12. Ren, G. et al. Model of human low-density lipoprotein and bound receptor based on cryoEM. *Proc. Natl Acad. Sci. USA* **107**, 1059–1064 (2010).
13. Kumar, V. et al. Three-dimensional cryoEM reconstruction of native LDL particles to 16 Å resolution at physiological body temperature. *PLoS ONE* **6**, e18841 (2011).
14. Yu, Y. et al. Polyhedral 3D structure of human plasma very low density lipoproteins by individual particle cryo-electron tomography. *J. Lipid Res.* **57**, 1879–1888 (2016).
15. Lei, D. et al. Single-molecule 3D imaging of human plasma intermediate-density lipoproteins reveals a polyhedral structure. *Biochim. Biophys. Acta* **1864**, 260–270 (2019).
16. Cisse, A. et al. Targeting structural flexibility in low density lipoprotein by integrating cryo-electron microscopy and high-speed atomic force microscopy. *Int. J. Biol. Macromol.* **252**, 126345 (2023).
17. Kühlbrandt, W. Forty years in cryoEM of membrane proteins. *Microscopy* **71**, i30–i50 (2022).
18. Fontana, P. et al. Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science* **376**, eabm9326 (2022).
19. McGreevy, R. et al. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **100**, 50–60 (2016).
20. Deckelbaum, R. J. et al. Thermal transitions in human plasma low density lipoproteins. *Science* **190**, 392–394 (1975).
21. Ivanova, E. A. et al. Small dense low-density lipoprotein as biomarker for atherosclerotic diseases. *Oxid. Med. Cell. Longev.* **2017**, 1273042 (2017).
22. Frank, P. G. & Marcel, Y. L. Apolipoprotein A-I: structure-function relationships. *J. Lipid Res.* **41**, 853–872 (2000).
23. Li, L. et al. Double belt structure of discoidal high density lipoproteins: molecular basis for size heterogeneity. *J. Mol. Biol.* **343**, 1293–1311 (2004).
24. Raag, R. et al. Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 2.8 A resolution. *J. Mol. Biol.* **200**, 553–569 (1988).
25. Baker, M. E. Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? *Biochem. J.* **255**, 1057–1060 (1988).
26. Shelness, G. S. & Ledford, A. S. Evolution and mechanism of apolipoprotein B-containing lipoprotein assembly. *Curr. Opin. Lipidol.* **16**, 325–332 (2005).
27. Segrest, J. P., Jones, M. K. & Dashti, N. N-terminal domain of apolipoprotein B has structural homology to lipovitellin and microsomal triglyceride transfer protein: a "lipid pocket" model for self-assembly of apob-containing lipoprotein particles. *J. Lipid Res.* **40**, 1401–1416 (1999).
28. Biterova, E. I. et al. The crystal structure of human microsomal triglyceride transfer protein. *Proc. Natl Acad. Sci. USA* **116**, 17251–17260 (2019).
29. Richardson, P. E. et al. Assembly of lipoprotein particles containing apolipoprotein-B: structural model for the nascent lipoprotein particle. *Biophys. J.* **88**, 2789–2800 (2005).
30. Jeiran, K. et al. A new structural model of apolipoprotein B100 based on computational modeling and cross linking. *Int. J. Mol. Sci.* **23**, 11480 (2022).
31. Manchekar, M. et al. Apolipoprotein B-containing lipoprotein particle assembly: lipid capacity of the nascent lipoprotein particle. *J. Biol. Chem.* **279**, 39757–39766 (2004).
32. Dashti, N. et al. The N-terminal 1000 residues of apolipoprotein B associate with microsomal triglyceride transfer protein to create a lipid transfer pocket required for lipoprotein assembly. *Biochemistry* **41**, 6978–6987 (2002).
33. Spring, D. J. et al. Lipoprotein assembly. Apolipoprotein B size determines lipoprotein core circumference. *J. Biol. Chem.* **267**, 14839–14845 (1992).
34. McLeod, R. S. et al. Apolipoprotein B sequence requirements for hepatic very low density lipoprotein assembly. Evidence that hydrophobic sequences within apolipoprotein B48 mediate lipid recruitment. *J. Biol. Chem.* **271**, 18445–18455 (1996).
35. Wang, L., Walsh, M. T. & Small, D. M. Apolipoprotein B is conformationally flexible but anchored at a triolein/water interface: a possible model for lipoprotein surfaces. *Proc. Natl Acad. Sci. USA* **103**, 6871–6876 (2006).
36. Forgez, P. et al. Identification of surface-exposed segments of apolipoprotein B-100 in the LDL particle. *Biochem. Biophys. Res. Commun.* **140**, 250–257 (1986).
37. Whitfield, A. J. et al. Lipid disorders and mutations in the *APOB* gene. *Clin. Chem.* **50**, 1725–1732 (2004).
38. Bruikman, C. S., Hovingh, G. K. & Kastelein, J. J. P. Molecular basis of familial hypercholesterolemia. *Curr. Opin. Cardiol.* **32**, 262–266 (2017).
39. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
40. Schonfeld, G. Familial hypobetalipoproteinemia: a review. *J. Lipid Res.* **44**, 878–883 (2003).
41. Parhofer, K. G. et al. Lipoproteins containing the truncated apolipoprotein, Apo B-89, are cleared from human plasma more rapidly than Apo B-100-containing lipoproteins in vivo. *J. Clin. Invest.* **89**, 1931–1937 (1992).
42. Parhofer, K. G. et al. Positive linear correlation between the length of truncated apolipoprotein B and its secretion rate: in vivo studies in human apoB-89, apoB-75, apoB-54.8, and apoB-31 heterozygotes. *J. Lipid Res.* **37**, 844–852 (1996).

Methods

SDS–PAGE

SDS–PAGE was performed in the Bio-Rad Mini-Protean Tetra Cell chamber using 4–20% gradient precast Tris-glycine gels.

Cryo-electron microscopy

Sample preparation. Around 500 μg of human LDL purified by density gradient ultracentrifugation was purchased from Thermo Fisher Scientific (L3486) and further purified by SEC using the Superose 6 Increase column (Cytiva) and Akta Pure 25L system (Cytiva). To reduce heterogeneity in LDL size before cryo-EM, fractions corresponding to the slowest-eluting approximate quarter of the LDL peak (rightmost tail) were pooled and concentrated to around 1 mg ml^{-1} (Extended Data Fig. 1a). Four microlitres of this sample was then applied to plasma-cleaned holey carbon grids (1.2/1.3, 300 mesh, Quantifoil) without the addition of detergent and plunge-frozen using the FEI Vitrobot mark IV at 4 °C and 100% humidity using a blot force of 0 and blot time of 4 s.

Data collection. Imaging was performed on the FEI Titan Krios G4 Cryo-TEM (Thermo Fisher Scientific) equipped with the Gatan K3 direct detector operated in counting mode and a Gatan BioQuantum Imaging Filter. Automated data collection was controlled using SerialEM[43]. Video micrographs were collected with a frame rate of 100 ms until a total dose of approximately 50 e^{-} Å^{-2} was achieved. Detailed imaging parameters and data statistics are provided in Extended Data Table 1.

Data processing. All data processing was performed with CryoSparc (v.4.0)[44] installed on the University of Missouri high-performance computing cluster and a custom GPU workstation following a standard workflow (Extended Data Fig. 2a,b). In brief, after frame alignment, dose-weighting and contrast transfer function (CTF) estimation, micrograph curation was performed to remove images with poor CTF resolution estimates, thick ice and high drift. Next, around 600,000 LDL particles were picked using the blob picker algorithm and extracted as 2× binned particles. Numerous rounds of 2D classification followed by subset selection were performed, selecting for homogeneous particles based on size and particle integrity. In the initial round of 2D classification, the number of classes was kept at the default of 50 and a default class uncertainty factor of 0.2 was applied. For subsequent rounds of 2D classification, the number of classes was increased to either 100 or 200, the class uncertainty factor was increased to 0.6–0.8, the number of particles per batch was increased to 200–400, the number of EM iterations to 30–40 and the number of final full iterations to 5–10. These subsequent rounds of 2D classification were crucial for achieving better alignment and resolution of weak peripheral density features corresponding to apoB100 and for improved selection of only the highest-quality particles. Particles with ordered CE stacks in their core were excluded from further processing at this stage as they were found to bias the 3D alignment towards the core of the particle, thereby preventing alignment on peripheral protein features of interest. Numerous rounds of 3D classification, first including the entire particle, then with a soft mask around the globular NTD, were performed using different algorithms followed by subset selection. The general strategy adopted during 3D classification was to focus on selecting classes with the highest quality and most homogeneous density in the NTD and β-belt domain, as these are the most prominent and homogeneous protein features in the structure. Ab initio classification into four classes was performed first, followed by homogeneous refinement of any classes that showed hints of NTD density. After the initial round(s) of refinement, the 3D classification algorithm was used to sort into a large number of classes (30–50). This allowed for particles with a poorly ordered NTD/β-belt density to be removed. Three-dimensional variability analysis was also useful, primarily when focused on the NTD, to remove residual heterogeneity from continuous flexibility about the curved axis (Supplementary Video 3). After 3D classification, a final set of around 52,000 homogeneous particles was re-extracted as unbinned particles and refined to around 9 Å resolution with local resolutions of around 6–14 Å (Extended Data Figs. 2b and 4a). All final 3D refinement was performed with the non-uniform refinement algorithm without per-particle or global CTF refinement enabled. Lastly, the same soft mask around the NTD used for classification was used for local refinement of this domain, resulting in an approximately 6 Å-resolution local reconstruction of the NTD with local resolutions of around 4.5–6 Å (Extended Data Figs. 2c and 4c). Numerous alternative processing pipelines were explored, either including or excluding particles with ordered cores, and using different 3D classification strategies, all of which resulted in either worse or roughly equivalent final map quality (Extended Data Fig. 2d). Local-resolution analysis was performed in CryoSparc. All map and model visualizations were generated with ChimeraX[45,46].

Morphological analysis. Morphological analysis of 2D cryo-EM and negative-stain EM class averages was performed in MATLAB v.9.12.0.2009381 (R2022a) equipped with the Image Processing Toolbox v.11.5 (MathWorks) using a custom script.

Model building, refinement and analysis

The apoB100 structure predictions presented in the main body of this work were performed using AF2 (ref. 4) as implemented through ColabFold[47]. Owing to current single-prediction-size limitations, the apoB100 sequence (UniProt: P04114) was divided into three fragments (residues 1–1820, residues 1681–3500 and residues 3361–4563) with 140-residue overlaps to assist with subsequent model concatenation (Extended Data Fig. 5a–d). For each fragment, we chose from the top two predictions the one with the least compact tertiary structure. Together, these structures formed the basis of our atomic modelling. Note that, since initiating our study, AlphaFold3 (AF3) and an accompanying webserver has been released[39], which is capable of folding full-length apoB100 in a single run. However, the resulting structures are severely collapsed and considerably more entangled than the AF2 fragment predictions, displaying numerous severe clashes (Supplementary Fig. 1). Given that no significant differences were observed in the average pLDDT score or secondary structure content between the AF2 and AF3 predictions, we deemed it most appropriate to maintain AF2 as the basis for structure prediction.

The full-length apoB100 model was constructed progressively through a series of MD-based structural refinement simulations, using particularly the MDFF methodology[48]. We began by flexibly fitting the first AF2 fragment to our NTD-focused refined map, after which we sequentially attached fragments 2 and 3 and fit them to our global map using only the β-belt atoms and restraining the previously fit portion of the structure (Supplementary Video 2). We then refined each insert structure in turn where necessary. We note that, for all inserts, clear density was present where the insert was attached to the β-belt, demonstrating that our sequential MDFF protocol had appropriately threaded the β-belt through its density. When modelling inserts 4 and 9, the collective variables (colvars) module[49] was used to create additional attractive harmonic potentials between distant regions of the apoB100 structure within the simulation. Such potentials enabled us to roughly guide the positions of the various arms of inserts 4 and 9 according to the distinctive regions observed in our map, which were then locally fit using MDFF. Note that no cross-linking information was used during the modelling procedure, and it was compared only with the final refined atomic model. To refine the model stereochemistry, the entire apoB100 structure was first subjected to a series of restrained equilibration simulations and conjugate gradient energy minimizations in explicit solvent, followed by real-space refinement in Phenix (v.1.20)[50] and, finally, manual refinement using ISOLDE[51].

Model building and visualization were performed using a combination of VMD (v.1.9.4)[52] and ChimeraX (v.1.6.1)[45,53]. All simulations were performed using NAMD (v.2.14)[54,55] and the CHARMM36 force field[56]. All MDFF simulations were performed using standard settings[48] with a coupling factor of 0.3 applied to subsets of the protein backbone. To help to overcome regions of heterogeneous resolution in the cryo-EM data, we used a cascade approach in each MDFF simulation[19], first fitting the target region to four Gaussian-blurred maps (more blurred progressing to less blurred) before finally fitting to the sharpened map. Harmonic restraints were applied during all fittings to maintain the AF2-predicted secondary structure and to prevent the formation of *cis*-peptide bonds and chirality errors. Before fitting, the density from the particle centre was manually removed, preventing the potential for erroneous attraction to the CE stacks.

Normal mode analysis. The relaxed atomic model of the NTD of apoB100 (residues 28–1010) was loaded into ProDy[57,58] and an anisotropic network model was created from all backbone atoms, which was then used to calculate the ten lowest-energy normal modes. The normal mode wizard[57] implemented through VMD[52] was used to visualize the results.

AF2 PAE-matrix-based model segmentation. Automatic segmentation of the NTD structure by clustering of the AF2 PAE matrix as seen in Extended Data Fig. 9c was performed and visualized in UCSFChimeraX[46].

Sequence analysis

The apoB100 multiple sequence alignment was calculated using UniProt BLAST[59] and visualized with Jalview[60].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All cryo-EM maps and refined atomic models have been deposited into the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB), respectively, under the following accession codes: EMDB-47801 (primary 3D reconstruction of the full LDL particle, alternative 3D reconstruction and focused refinement of the NTD), PDB 9E9R (truncated atomic model of apoB100), PDB 9EA7 (primary full-length model of apoB100) and PDB 9EAG (alternative full-length model of apoB100).

Code availability

All custom data processing and analysis code is available at GitHub (https://github.com/ZTBioPhysics/The_Structure_of_ApoB100_From_Human_Low-Density_Lipoprotein).

43. Mastronarde, D. N. SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).
44. Punjani, A. et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
45. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
46. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
47. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
48. Trabuco, L. G. et al. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).
49. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).
50. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
51. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).
52. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
53. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
55. Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).
56. Huang, J. & MacKerell Jr, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
57. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
58. Zhang, S. et al. ProDy 2.0: increased scale and scope after 10 years of protein dynamics modelling with Python. *Bioinformatics* **37**, 3657–3659 (2021).
59. Zaru, R. & Orchard, S. UniProt Tools: BLAST, Align, Peptide Search, and ID Mapping. *Curr. Protoc.* **3**, e697 (2023).
60. Waterhouse, A. M. et al. Jalview version 2-a: multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

Acknowledgements Z.T.B. thanks the staff of the Mizzou Electron Microscopy Core, especially M. Su, and the Mizzou Research Support Solutions staff, especially M. Stanley and A. A. M. Ali, for their technical support with microscopy and high-performance computing, respectively, the staff of the Mizzou Gehrke Proteomics Center, in particular B. Mooney and T. Nguyen, for their help with cross-linking experiments, and A. Boosani for running SDS–PAGE gels. Computational aspects of this work were partially supported by the high-performance computing infrastructure operated by Research Support Solutions in the Division of IT at the University of Missouri (https://doi.org/10.32469/10355/97710). We also thank S. Van Doren, M. Chapman, G. King and G. Ren for reading the manuscript. The efforts of Z.T.B. and C.K.C. were supported by University of Missouri faculty startup packages from the Departments of Biochemistry and Physics, respectively.

Author contributions Z.T.B. conceived the experiments, prepared samples, performed EM data collection, data processing and model building, normal mode analysis, sequence analysis, chemical cross-linking data analysis, interpreted results, prepared figures and wrote the manuscript. C.K.C. performed model building and validation, MD and flexible fitting simulations, interpreted the results and wrote the manuscript.

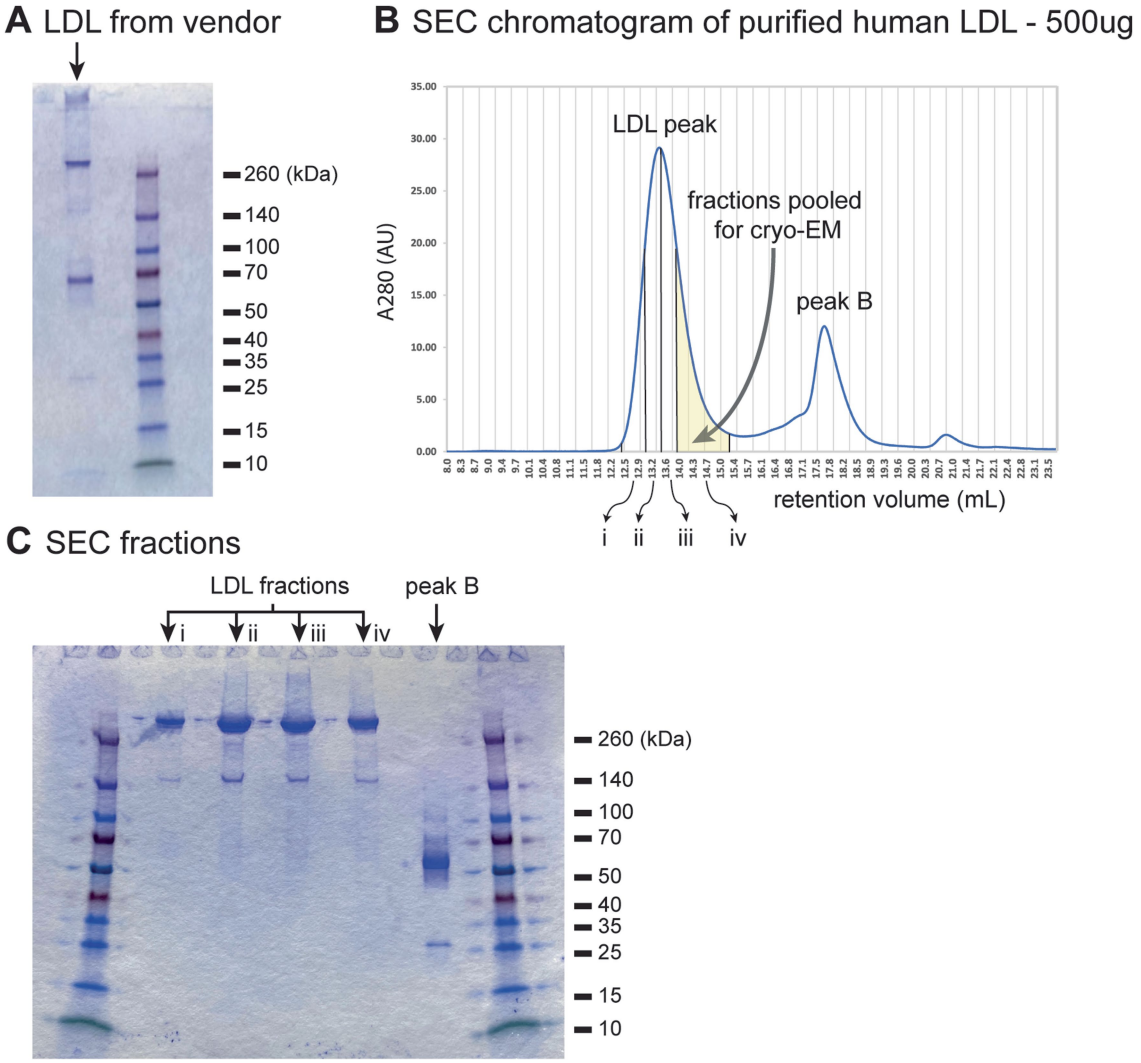
Competing interests The authors declare no competing interests.

Additional information

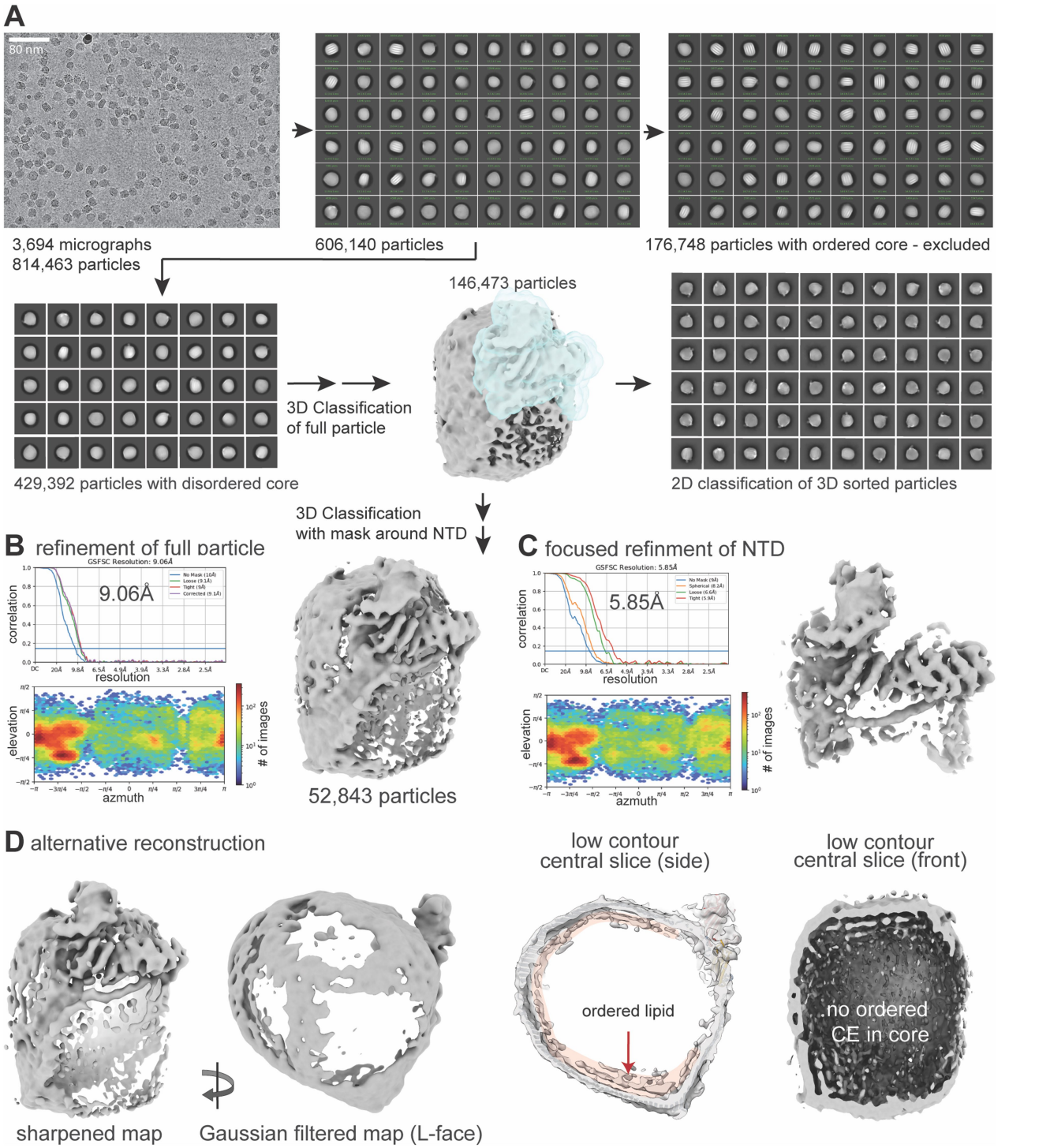
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08467-w.

Correspondence and requests for materials should be addressed to Zachary T. Berndsen or C. Keith Cassidy.

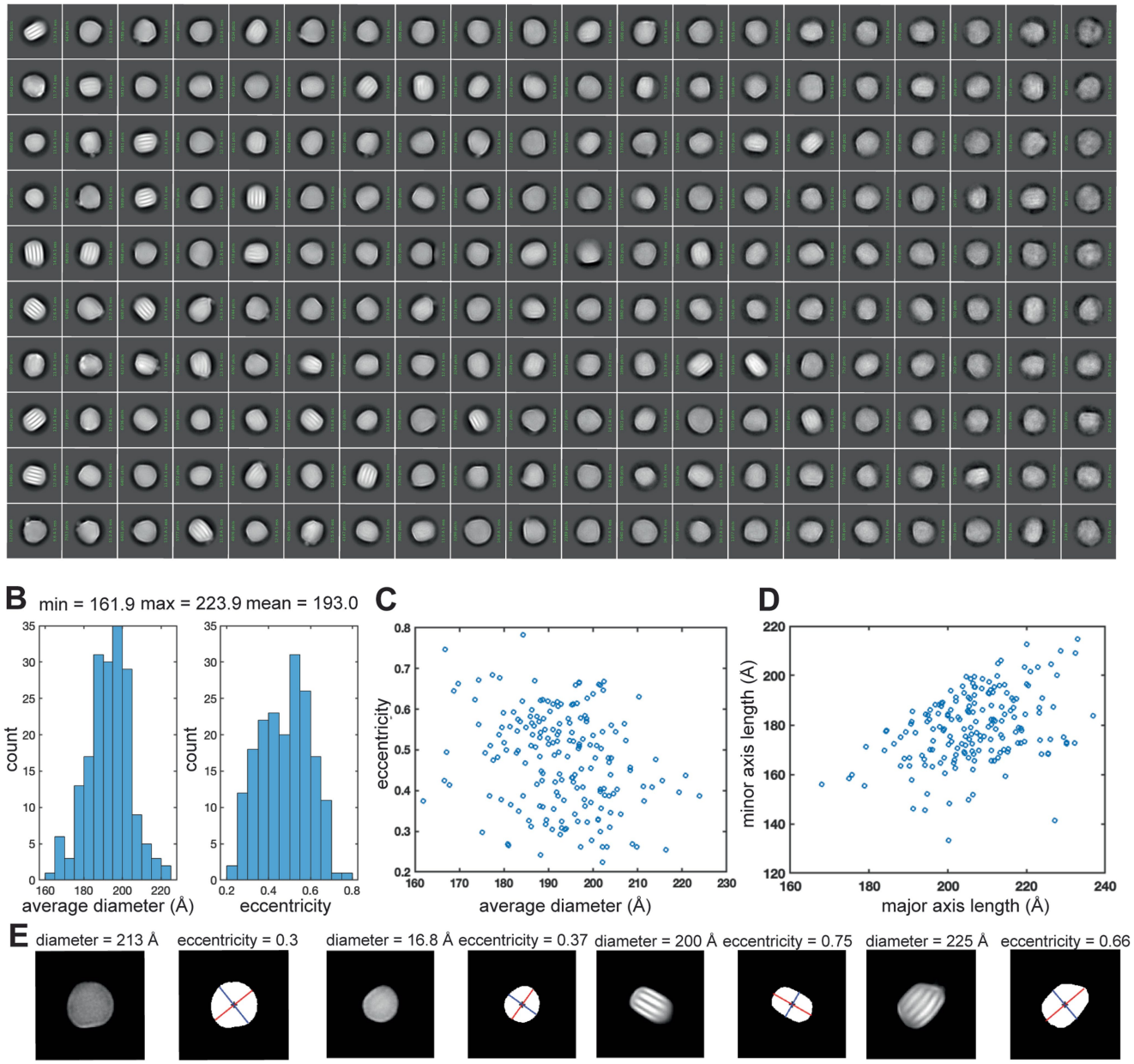
Peer review information *Nature* thanks Gang Ren and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



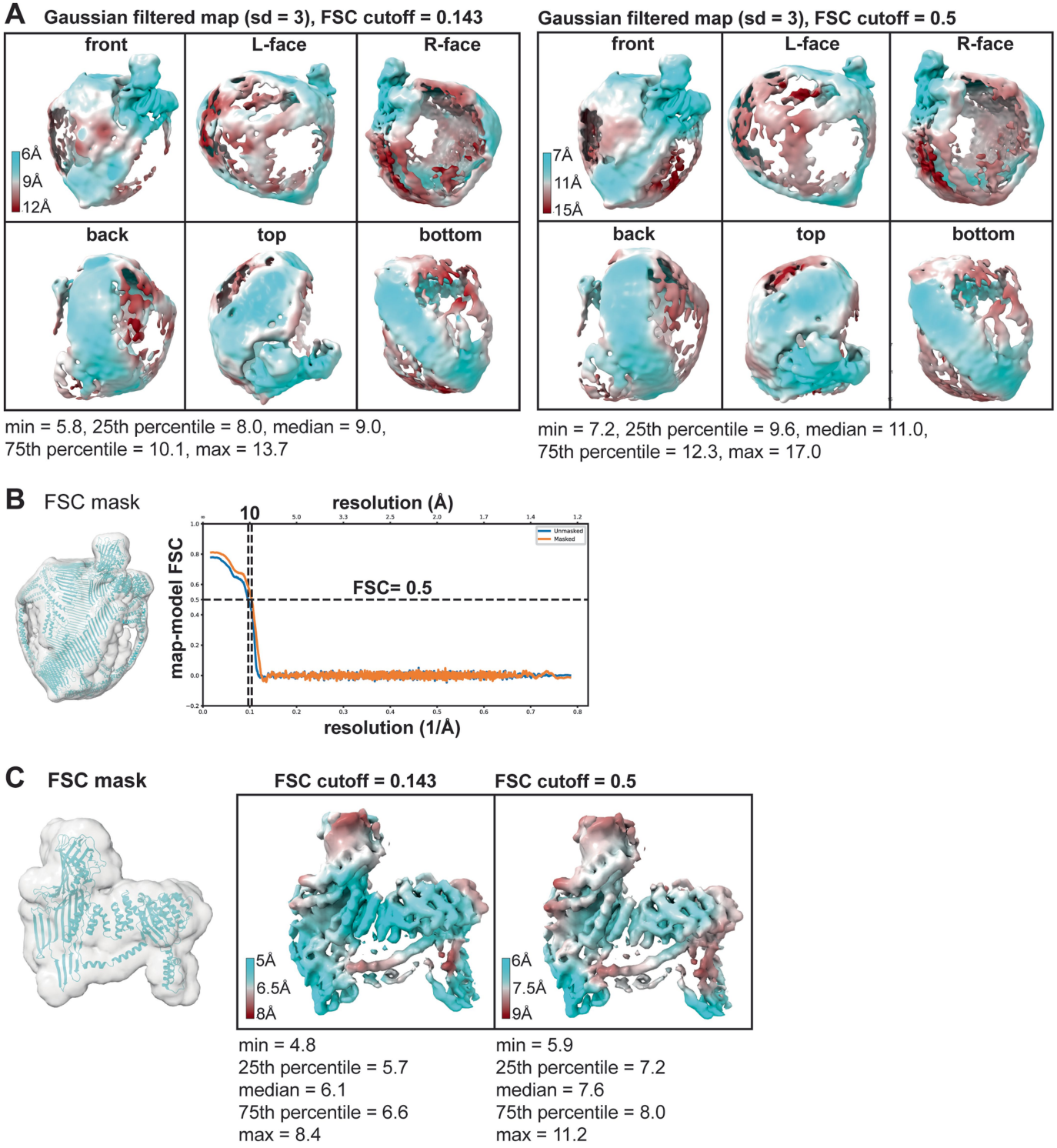
Extended Data Fig. 1 | SDS-PAGE and SEC of purified LDL subfractions. **A**. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel (**A**) and SEC chromatogram (**B**) of human LDL purified by density gradient ultracentrifugation (purchased from commercial vendor) with the LDL peak indicated as well as the fractions collected for cryo-EM imaging. **C**. SDS-PAGE gel of isolated SEC fractions.



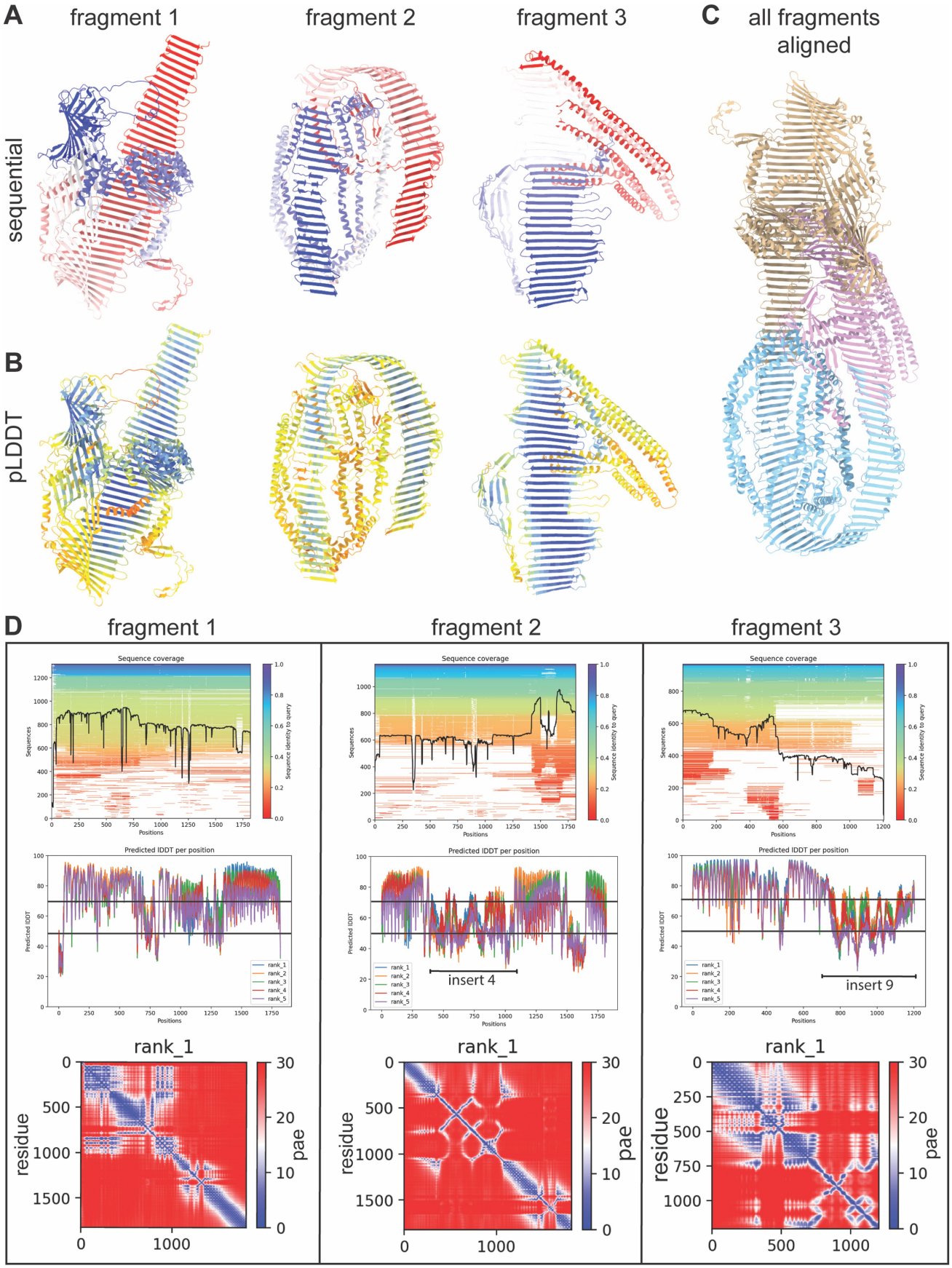
Extended Data Fig. 2 | Cryo-EM data processing workflow. A. Data processing workflow showing a representative micrograph with scale bar (80 nm), 2D class averages, and intermediate 3D reconstructions. Fourier shell correlation (FSC) curve and global resolution estimate for the final LDL particle refinement (**B**) and the local refinement of the NTD (**C**). **D.** Alternative refinement of the full LDL particle that does not contain any liquid crystalline CE in the core viewed from 2 directions and as central slices at low contour. Also shown is the density corresponding to coordinated lipids on the inner face of the β-belt.



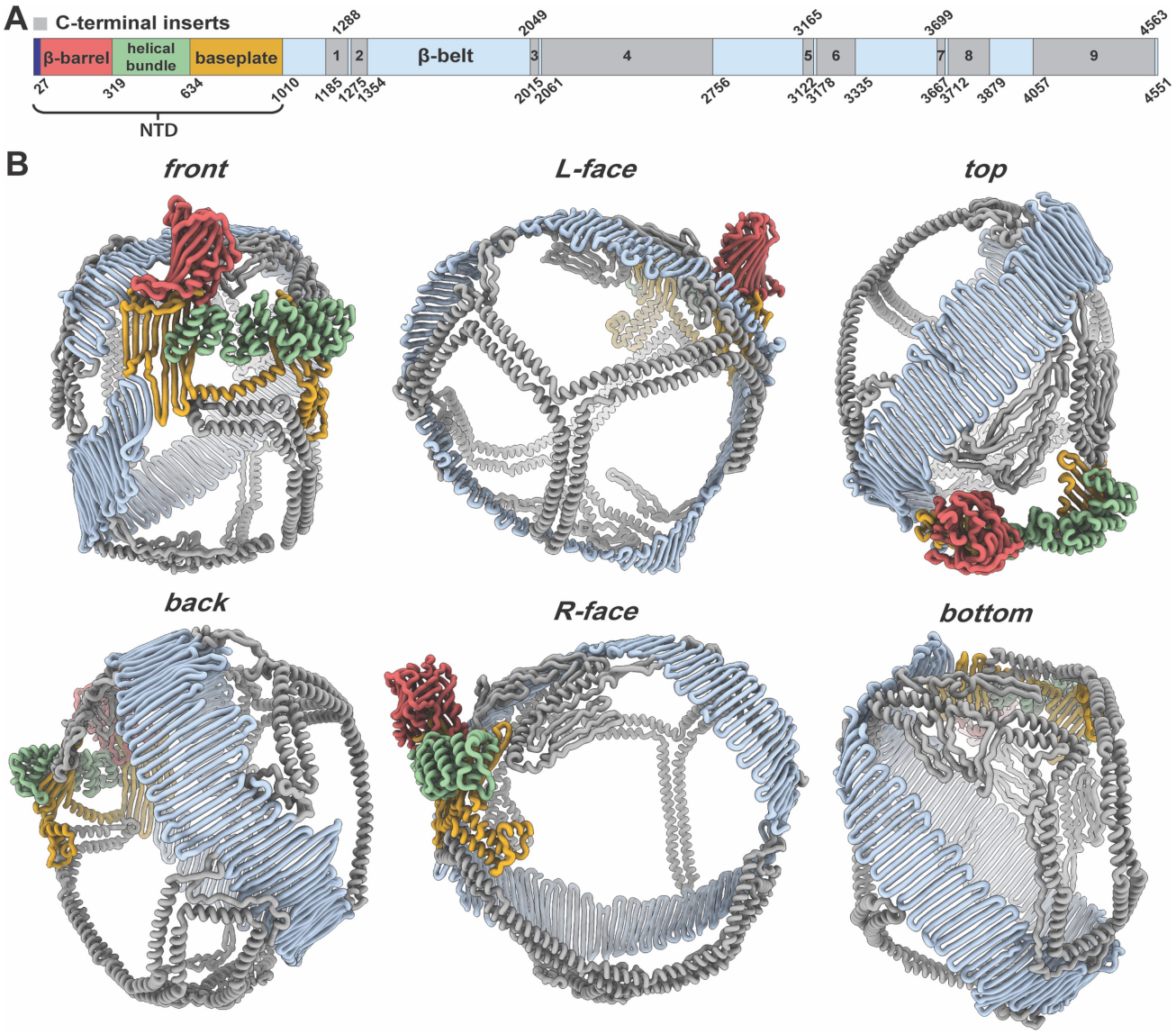
Extended Data Fig. 3 | Morphological analysis of cryo-EM 2D class averages. **A**. 200 2D class averages of LDL particles used for morphological analysis. **B**. Histogram of particle diameters and eccentricities. **C**. Scatter plot showing particle diameter vs. mean eccentricity and (**D**) major vs. minor axis length. **E**. Representative class averages from different extremes of the distribution.



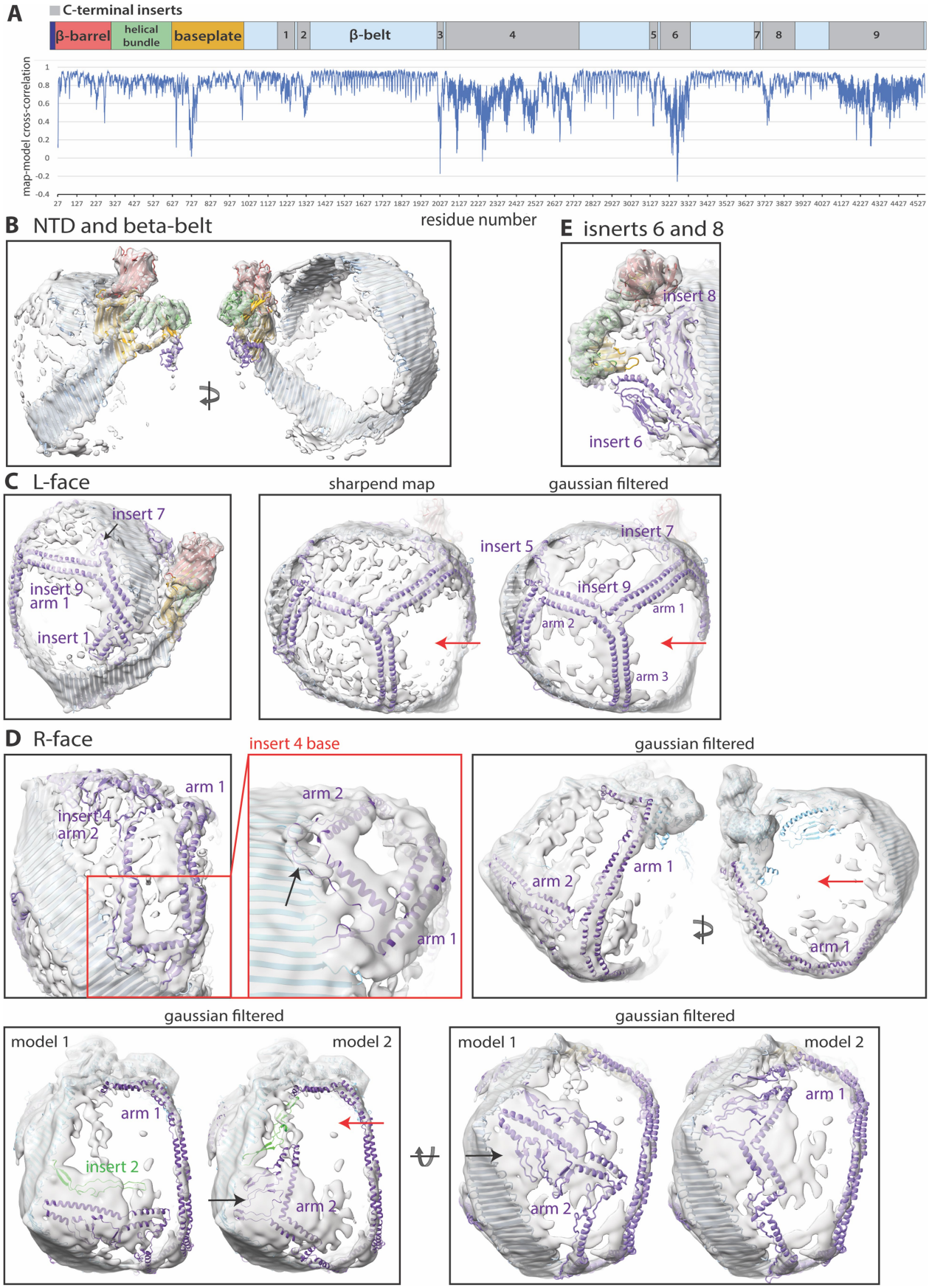
Extended Data Fig. 4 | Cryo-EM local and map-to-model resolution estimates. A. "Gold-standard" FSC local resolution estimates for the full LDL reconstruction using both the 0.143 and 0.5 FSC cutoffs. The map shown has been smoothed using a Gaussian kernel with a standard deviation of 3 for clarity. **B**. Map-to-model FSC for the full apoB100 structure using the soft mask shown on the left. **C**. Local FSC resolution estimates for the focused refinement of the NTD of apoB100 using the soft mask shown on the left.



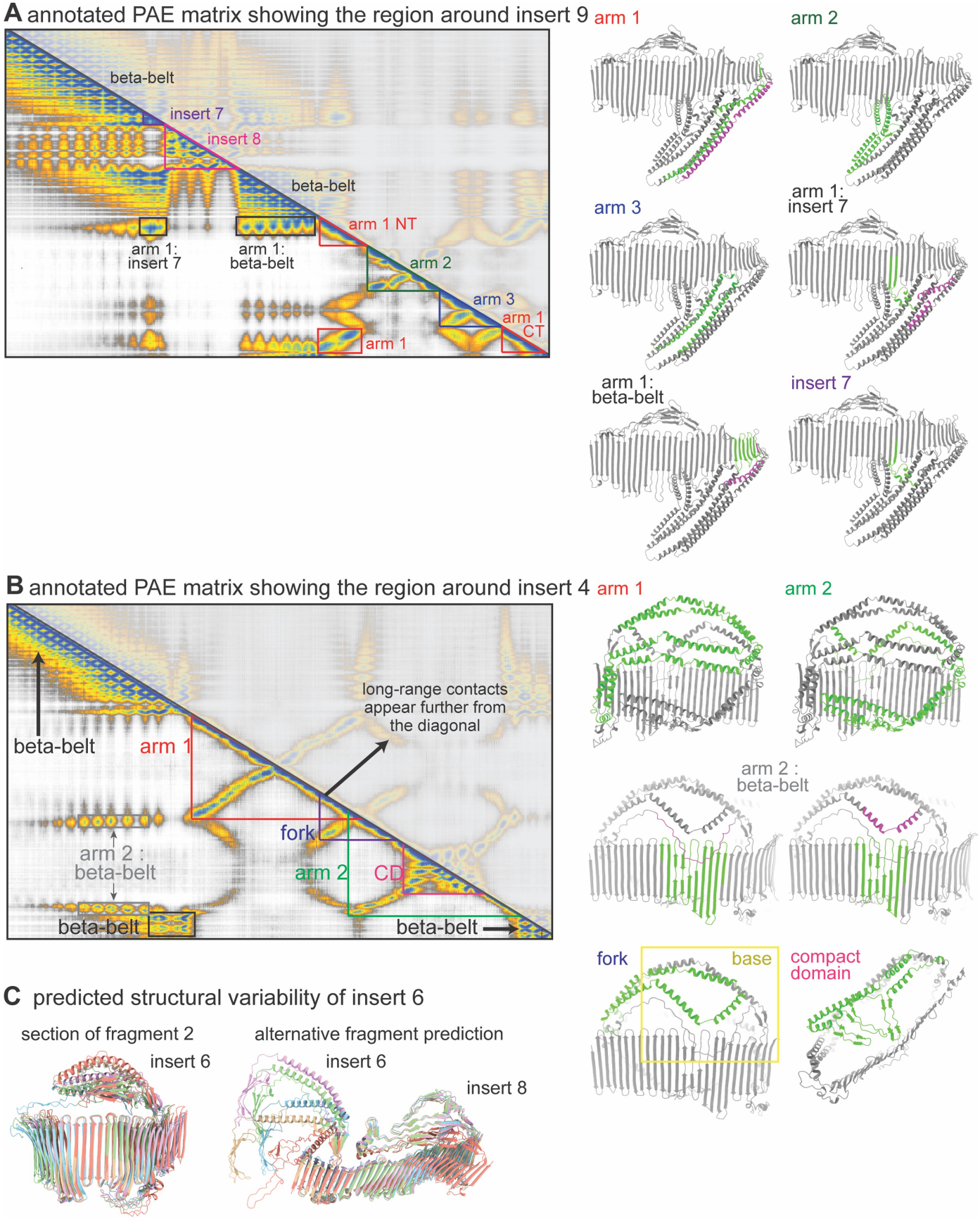
Extended Data Fig. 5 | AlphaFold2 structural modelling of apoB100. A. AF2 predicted structures for the 3 contiguous apoB100 fragments coloured sequentially (**A**) and by pLDDT score (**B**). **C**. All 3 AF2 fragments aligned to one another. **D**. Per-residue sequence coverage, pLDDT score, and PAE score for each fragment.



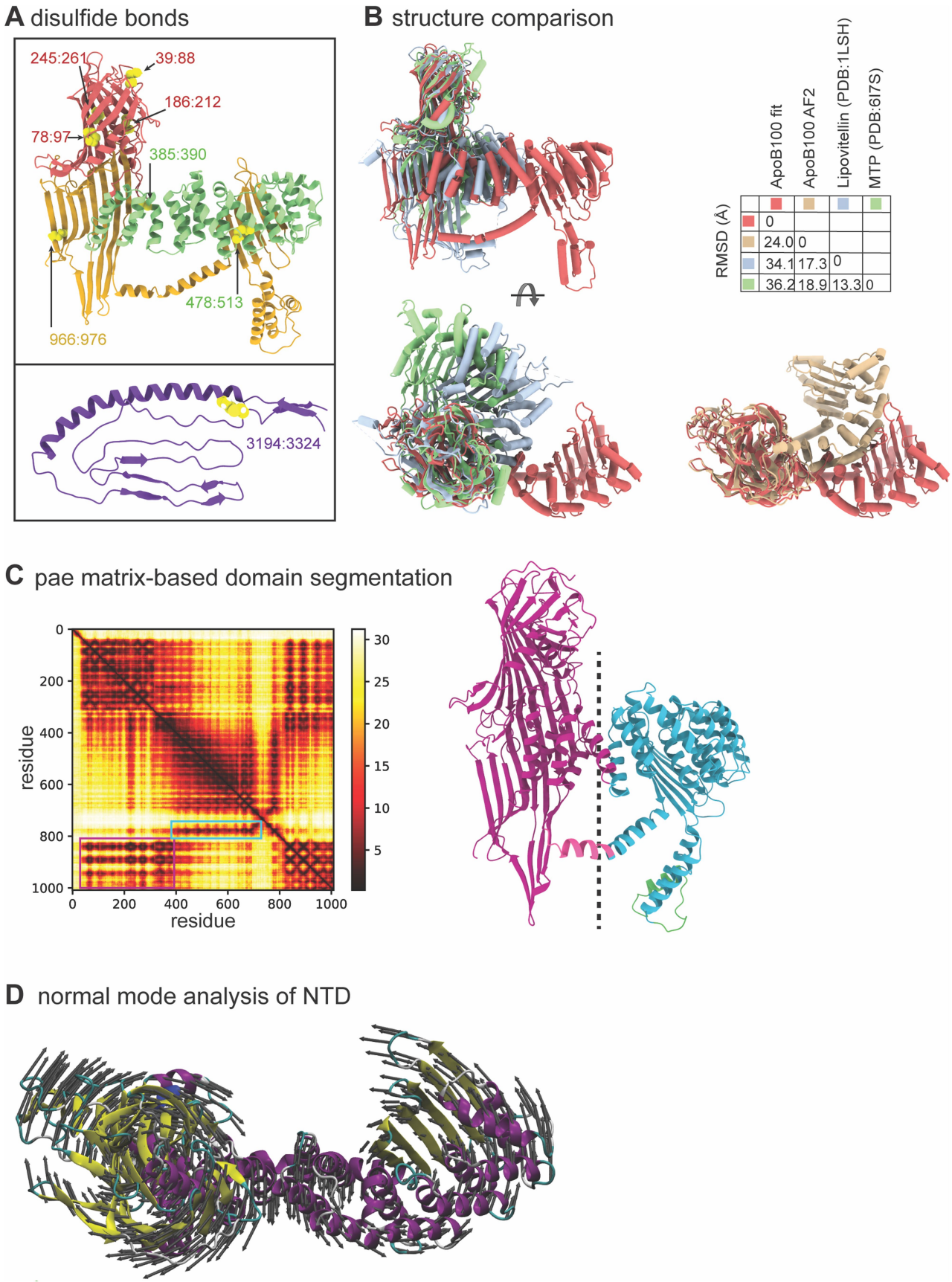
Extended Data Fig. 6 | Atomic model of full-length apoB100. A. Gene diagram of apoB100. **B**. Atomic model of full-length apoB100 viewed from six different directions and coloured according to the gene diagram in panel **A**.



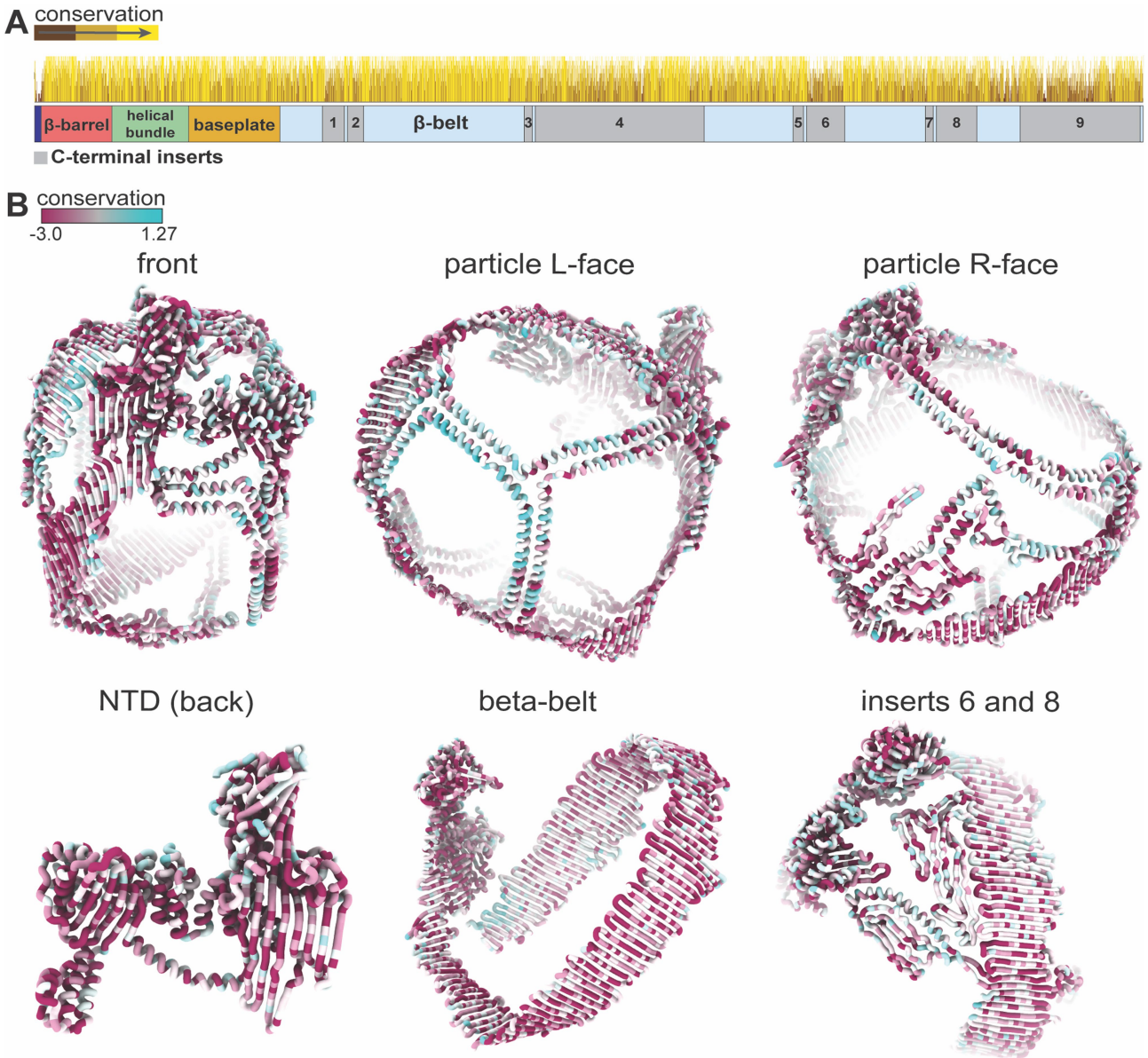
Extended Data Fig. 7 | Map-to-model fits. A. Per-residue map-to-model cross-correlation plot. Map to model fit of just the NTD and β-belt (**B**), L-face inserts (**C**), and R-face inserts (**D-E**). Red arrows point to regions without any high-density map features indicative of protein-free surface.



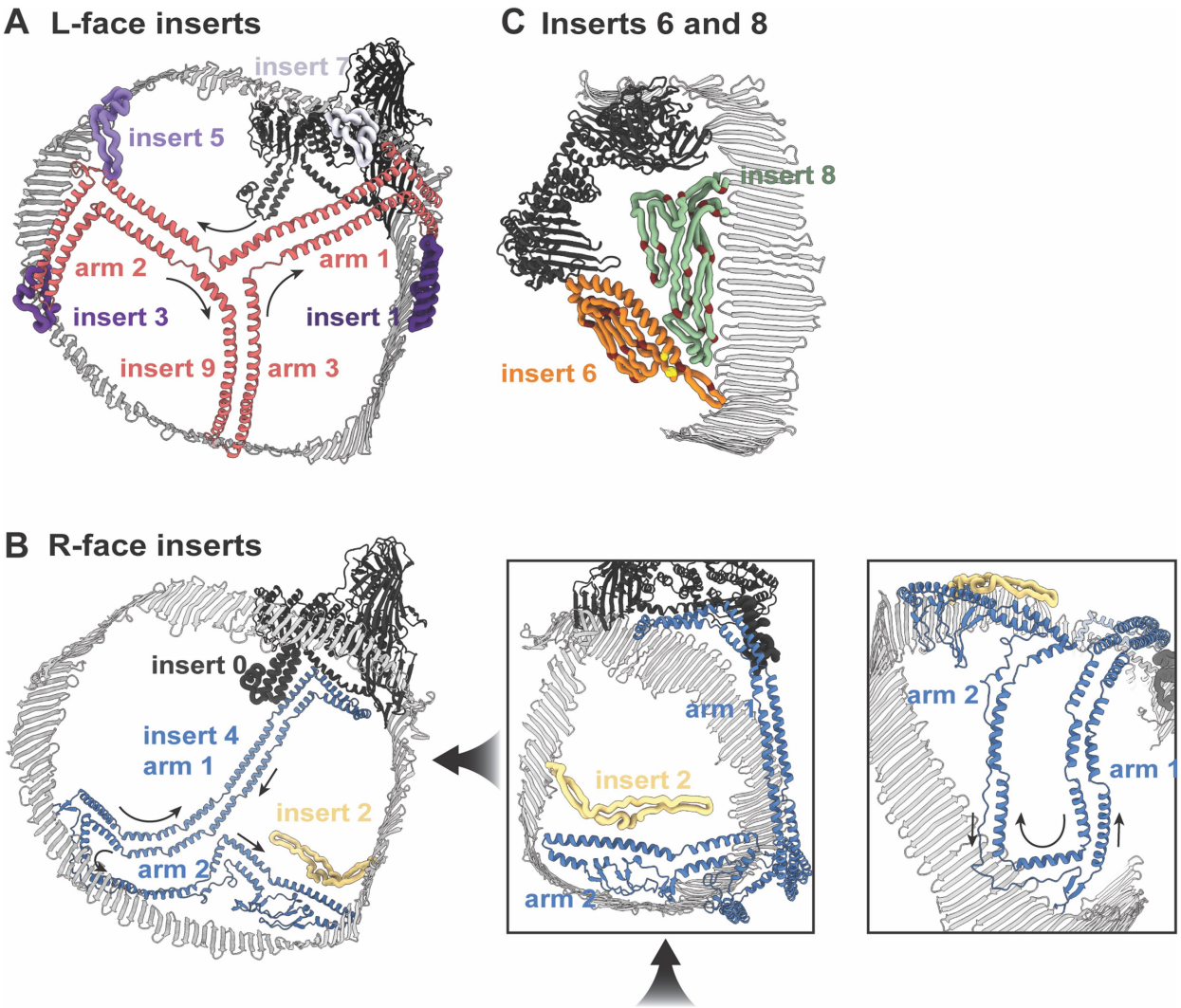
Extended Data Fig. 8 | Additional analysis of AlphaFold2 structure predictions. A. Annotated AF2 PAE matrix and segmented models for the region of apoB100 containing insert 9. **B**. Annotated AF2 PAE matrix and segmented models for the region of apoB100 containing insert 4. **C**. The 5 top-ranking AF2 models for the region containing insert 6 from fragment 2 along with the region containing inserts 6 and 8 from an alternative prediction with different boundaries.



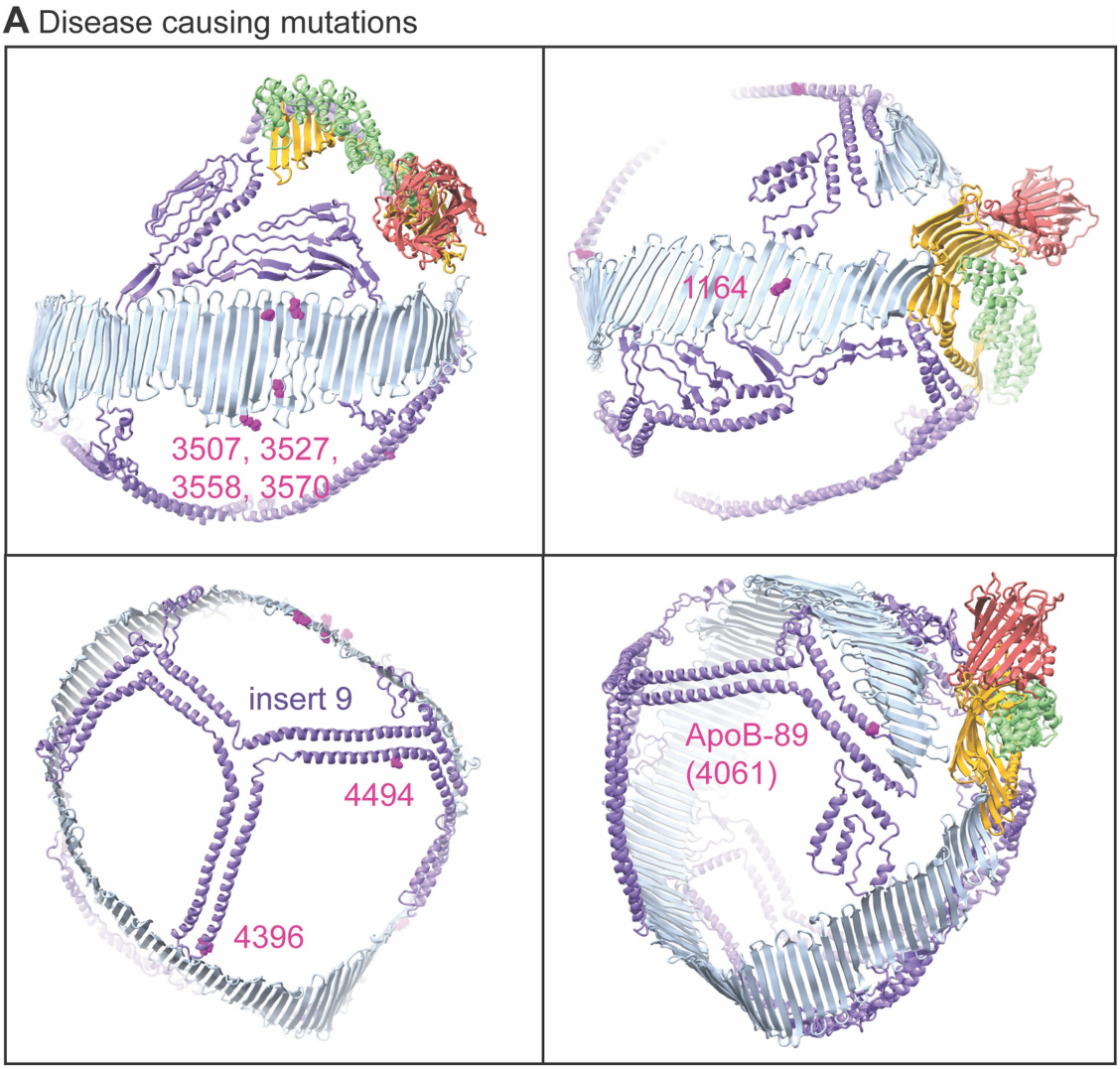
Extended Data Fig. 9 | Extended structural analysis of the apoB100 NTD. **A**. All disulfide bonds in apoB100 within the NTD (top) and insert 6 (bottom). **B**. Expanded structural comparison between the apoB100 NTD and the crystal structures of lipovitellin and MTP. **C**. AF2 PAE matrix for the apoB100 NTD and the PAE matrix-based segmentation. The dashed line shows the axis of flexibility between the two halves of the NTD. **D**. Top view of the NTD coloured by secondary structural elements showing the vectors of motion (grey arrows) corresponding to the slowest normal mode.



Extended Data Fig. 10 | Sequence conservation of apoB100. A. Per-residue sequence conservation score calculated from a multiple-sequence alignment of apoB100 variants plotted on top of the apoB100 gene diagram. **B.** Sequence conservation score mapped to the apoB100 structure.



Extended Data Fig. 11 | Atomic model of full-length apoB100 C-terminal inserts. A. View of the particle L-face with full-length insert 9 displayed as ribbons and inserts 1,3,5, and 7 as ropes. Black arrows point in the N to C direction of the protein chains. **B**. View of the particle R-face with full-length insert 4 depicted as ribbons and inserts 2 and 0 (NTD baseplate helical extensions) depicted as ropes. **C**. Structure of full-length inserts 6 and 8 depicted as ropes with the proline residues (maroon) and disulfide bond (yellow) highlighted. All colours correspond to the gene diagram in Fig. 3.



Extended Data Fig. 12 | ApoB100 disease-causing mutations. A. Common and rare disease-causing mutations and truncations mapped to the apoB100 atomic model.

Extended Data Table 1 | Statistics for cryo-EM data and refined atomic models

	Full LDL particle primary map (EMDB-47801) apoB100 - truncated (PDBID:9E9R)	apoB100 – full – primary (PDBID:9EA7)	apoB100- full – secondary (PDBID: 9EAG)	Local NTD refinement (EMDB-47801)
Data collection and processing				
Magnification	85,0000	N/A	N/A	85,0000
Voltage (kV)	300	N/A	N/A	300
Electron exposure (e–/Å^2)	~50	N/A	N/A	~50
Defocus range (μm)	-0.8:-2.8um	N/A	N/A	-0.8:-2.8um
Pixel size (Å)	1.09Å	N/A	N/A	1.09Å
Symmetry imposed	none	N/A	N/A	none
Initial particle images (no.)	814,463	N/A	N/A	814,463
Final particle images (no.)	52,843	N/A	N/A	60,333
Map resolution (Å) FSC threshold	9.06Å 0.143	N/A		5.85Å 0.143
Map resolution range (Å)	5.8-13.7Å	N/A	N/A	~4.8-8.4Å
Refinement				
Initial model used (PDB code)	AlphaFold2 prediction	AlphaFold2 prediction	AlphaFold2 prediction	AlphaFold2 prediction
Model resolution (Å) FSC threshold	N/A	N/A	N/A	N/A
Model resolution range (Å)	N/A	N/A	N/A	N/A
Map sharpening *B* factor (Å^2)	972.2	N/A	N/A	380.4
Model composition Non-hydrogen atoms Protein residues Ligands	 29417 3715 N/A	 36083 4526 N/A	 36083 4526 N/A	
B factors (Å^2) Protein Ligand	N/A	N/A		
R.m.s. deviations Bond lengths (Å) Bond angles (°)	N/A 0.009 0.939	N/A 0.004 0.924	N/A 0.004 0.964	
Validation MolProbity score Clashscore Poor rotamers (%) Cβ-outliers (%) CaBLAM outliers (%)	 1.8 9.4 0.0 0.0 2.06	 1.74 8.43 0.0 0.0 1.99	 2.1 17.92 0.1 0.0 3.07	
Ramachandran plot Favored (%) Allowed (%) Outliers (%)	 95.68 4.21 0.11	 95.89 4.02 0.09	 94.98 4.84 0.18	

Imaging and validation statistics for all cryo-EM maps and apoB100 atomic models.

nature portfolio

Corresponding author(s): C. Keith Cassidy
Zachary T. Berndsen
Last updated by author(s): Zachary T. Berndsen
C. Keith Cassidy 11-20-24

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
X	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
X	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
X	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
X	☐ A description of all covariates tested
X	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
X	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
X	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
X	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
X	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
X	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	all data collection software is published and/or publicly and/or commercially available
Data analysis	all data analysis software is published and/or publicly and/or commercially available

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

a data availability statement has been included with the manuscript

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	
Reporting on race, ethnicity, or other socially relevant groupings	
Population characteristics	
Recruitment	
Ethics oversight	

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n/a
Data exclusions	n/a
Replication	n/a
Randomization	n/a
Blinding	n/a

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	
Blinding	

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	
Location	
Access & import/export	
Disturbance	

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	
Validation	

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	
Authentication	
Mycoplasma contamination	
Commonly misidentified lines (See ICLAC register)	

Palaeontology and Archaeology

Specimen provenance	
Specimen deposition	
Dating methods	

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	
Wild animals	
Reporting on sex	
Field-collected samples	
Ethics oversight	

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	
Study protocol	
Data collection	
Outcomes	

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks

Novel plant genotypes

Authentication

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

Files in database submission

Genome browser session
(e.g. UCSC)

Methodology

Replicates

Sequencing depth

Antibodies

Peak calling parameters

Data quality

Software

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Instrument

Software

Cell population abundance

Gating strategy

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Design specifications

Behavioral performance measures

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

(See Eklund et al. 2016)

Correction

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis