Evolutionary and Structural Constraints Influencing Apolipoprotein A-I Amyloid Behavior

Gisonno RA[a,#], Masson T[b,#,*], Ramella N[a], Barrera EE[c], Romanowski V[b], Tricerri MA[a,*]

[a] Instituto de Investigaciones Bioquímicas de La Plata (INIBIOLP, CONICET-UNLP), Facultad de Ciencias Médicas, Universidad Nacional de La Plata, La Plata, Argentina

[b] Instituto de Biotecnología y Biología Molecular (IBBM, CONICET-UNLP), Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina

[c] Group of Biomolecular Simulations, Institut Pasteur de Montevideo, Montevideo, Uruguay

[#] Co-first authors

* Correspondence to Masson T and Tricerri MA: tomasmasson0@gmail.com aletricerri@yahoo.com

Highlights

- Aggregation-prone region 1 (APR1), comprising residues 14-19, is consistently conserved during the evolutionary history of Apolipoprotein A-I.
- APR1 contributes to thermal stability of the α-helix bundle in the full-length Apolipoprotein A-I model.
- Amyloid variants introduce a destabilizing effect on the monomer structure of Apolipoprotein A-I, in contrast to HDL-deficiency and naturally-occurring variants, which are nearly neutral.
- During molecular dynamics simulations, G26R amyloidogenic mutant lead to the partial unfolding of α-helix bundle and exposure of APR1.

Abstract

Apolipoprotein A-I (apoA-I) has a key function in the reverse cholesterol transport mediated by the high-density lipoprotein (HDL) particles. However, aggregation of apoA-I single point mutants can lead to hereditary amyloid pathology. Although several studies have tackled the biophysical and structural impacts introduced by these mutations, there is little information addressing the relationship between the evolutionary and structural features that contribute to the amyloid behavior of apoA-I. We combined evolutionary studies, *in silico* saturation mutagenesis and molecular dynamics (MD) simulations to provide a comprehensive analysis of the conservation and pathogenic role of the aggregation-prone regions (APRs) present in apoA-I. Sequence analysis demonstrated the pervasive conservation of an APR, designated here APR1, within the N-terminal α-helix bundle. Moreover, stability analysis carried out with the FoldX engine showed that this motif contributes to the marginal stability of apoA-I. Structural properties of the full-length apoA-I model suggest that aggregation is avoided by placing APRs into highly packed and rigid portions of its structure. Compared to HDL-deficiency or natural silent variants extracted from the gnomAD database, the thermodynamic and pathogenic impact of apoA-I point mutations associated with amyloid pathologies were found to show a higher destabilizing effect. MD simulations of the amyloid variant G26R evidenced the partial unfolding of the α-helix bundle and the occurrence of β-strand secondary elements at the C-terminus of apoA-I. Our findings highlight APR1 as a relevant component for apoA-I structural integrity and emphasize a destabilizing effect of amyloid variants that leads to the exposure of APRs. This information contributes to our understanding of how apoA-I, with its high degree of structural flexibility, maintains a delicate equilibrium between its native structure and intrinsic tendency to form amyloid aggregates. In addition, our stability measurements could be used as a proxy to interpret the structural impact of new mutations affecting apoA-I.

Keywords: aggregation, amyloidosis, apolipoprotein, evolutionary-conserved, variants.

Introduction

Apolipoprotein A-I (apoA-I) is the most abundant protein component of high-density lipoproteins (HDL) and is responsible for the reverse cholesterol transport from extracellular tissues back to the liver (1, 2), which has been associated with a protective function against cardiac disease and atherosclerosis (3, 4). The scaffolding functions of apoA-I in the HDL particle and its multiple protein-protein interactions, mainly with the lecithin:cholesterol acyltransferase and the ATP-binding cassette A1 transporter (5, 6), forces it to maintain a dynamic and flexible conformation (7).

In contrast to these physiological functions, several point mutations affecting apoA-I have been associated with hereditary amyloid pathology (8). These mutations are mainly distributed into two "hot spots", located in the N-terminal and the C-terminal regions of the protein, each one with a typical clinical manifestation (9). Mutations that occur in the N-terminal region (residues 26–100) are characterized by amyloid deposits in the liver and kidney (10, 11), while those located within a short C-terminal sequence (residues 170–178) are mainly associated with heart, larynx and skin deposits (12). In non-hereditary amyloidosis, full-length apoA-I is deposited in atherosclerotic plaques as fibrils or senile forms of amyloid. This process has been associated with aging, but it has also been described in chronic pathologies such as Alzheimer's disease and type 2 diabetes mellitus (13).

Amyloid behavior of apoA-I N-terminal fragment has been attributed to the presence of aggregation-prone regions (APRs) in its sequence and, specifically, to an APR located at the N-terminus (11). It has been hypothesized that amyloidogenic mutations or post-translational modifications could promote aggregation through the destabilization of apoA-I structure, described as a molten globular state, followed by the exposure of APRs. In this sense, most studies addressing the effect of amyloid variants have focused on the biophysical and physiological consequences of single point apoA-I mutants. However, the relationship between apoA-I amyloidogenic motifs and its aggregation process remains unclear.

In this study, through an extensive evolutionary analysis, we characterized the conservation of aggregating regions in a broad dataset of apoA-I sequences. Using the recently described full-length consensus structure (14), we examined the structural properties of apoA-I that contribute to reduce the exposure of its constituent APRs. *In silico* saturation mutagenesis analysis of apoA-I demonstrated that an evolutionary conserved APR (residues 14-19) contributes to the thermodynamic stability of the N-terminus and reveals a common destabilizing effect for amyloid-associated variants. Using molecular dynamics simulations, we studied the conformational and dynamic impact of five different amyloid variants on the structure of full-length apoA-I. Altogether, our results suggest that APR1 is a structural component that contributes to the stability of apoA-I helix bundle and emphasizes the destabilizing effect of amyloid variants, which is linked to subsequent APRs exposure in the case of G26R variant. This information is relevant to understand how a marginally stable, but metabolically active protein manages to initiate the formation of an amyloid structure and develop a severe pathology.

Results

Molecular evolution of apoA-I in vertebrates

Because amino acid residues relevant for protein native structure tend to be maintained across its evolutionary trajectory, we conducted an analysis of apoA-I sequence conservation. As a first step, a maximum likelihood phylogeny reflecting the evolutionary relationships between sequences was reconstructed from a dataset comprising 215 species (Figure 1A). Our phylogenetic tree segregated apoA-I sequences into three major groups: a distant clade including ray-finned fishes that diverged first from the common ancestor of apoA-I and two clades comprising tetrapods, one integrated by birds, reptiles, and frogs sequences and the remaining encompassing mammals. Using this phylogeny as a framework, we inferred site-wise evolutionary rates (ER) at both codon (*dN/dS*, the ratio of nonsynonymous to synonymous mutations) and amino acid level, both with similar results. The ER profile of apoA-I protein sequence (Figure 1B) revealed that most of the N-terminal region of the protein is relatively conserved, while a higher level of variability is present in the C-terminal region. Overall, the *dN/dS* profile indicated that a major portion of the protein sequence is evolving under purifying selection, suggesting the presence of functional and structural constraints acting on apoA-I (Supplementary Figure 1). ApoA-I sequence is organized into an N-terminal domain followed by ten 11/22-mer tandem repeats. Given the relevance of apoA-I repeats in the mature form of HDL particles, we decided to investigate the relationship between the ER and the residue identity inside these repeats (Figure 1C). From these results, the proline and positively charged residues (arginine and lysine) were consistently more conserved when compared against other residues, a result that is also supported by the sequence logos of the tandem repeats (Supplementary Figure 2). Although the ER values corresponding to leucine residues were similar to the mean ER of apoA-I, we noticed several highly conserved leucine residues among apoA-I orthologs (e.g. L46, L163, L200, L214, and L218).

Lastly, we tried to uncover possible patterns of coevolution between pairs of residue positions using the RaptorX server to predict a contact map for apoA-I. Intriguingly, the predicted contact pairs, located mostly in the N-terminus (Supplementary Table 1), do not correlate with contacts proposed in apoA-I structure (14). This inconsistency could reflect the coexistence of different conformational states where these contacts effectively occur.

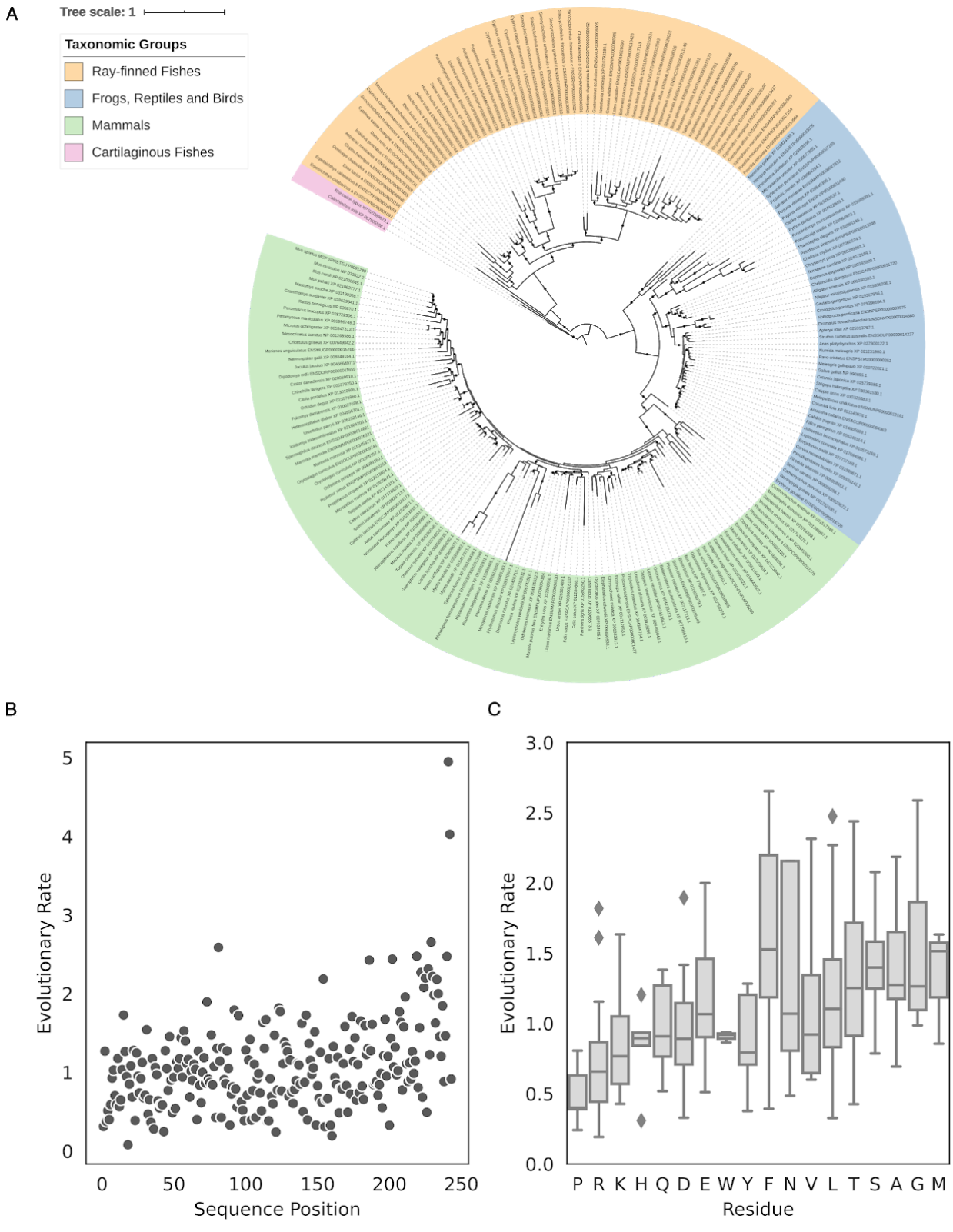

A
Tree scale: 1
Taxonomic Groups
Ray-finned Fishes
Frogs, Reptiles and Birds
Mammals
Cartilaginous Fishes
B
Evolutionary Rate
Sequence Position
0
50
100
150
200
250
0
1
2
3
4
5
C
Evolutionary Rate
Residue
P R K H Q D E W Y F N V L T S A G M
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Figure 1 ApoA-I Molecular Evolution in Vertebrates

A Maximum likelihood phylogeny of apoA-I orthologs. Major taxonomic groups are colored according to the legend. Branches with an ultra-fast bootstrap support value greater than 90 are marked with a black dot. Cartilaginous fishes were chosen as outgroup for phylogeny rooting.

B Evolutionary rate (ER) profile for apoA-I protein sequence, estimated with LEISR. A value of 1 represents the average apoA-I ER, while those rates below and above 1 indicate sites evolving at a slower and a faster rate than the average, respectively.

C Evolutionary rate for each residue type within apoA-I tandem repeats. Proline and positively charged residues (arginine (R) and lysine (K)) display values consistent with stringent evolutionary conservation.

ApoA-I has retained an APR during its evolutionary history

Protein aggregation through short regions has been associated with a wide range of human conformational diseases (15), notably amyloidosis. In the case of apoA-I, an N-terminal fragment has been linked to the formation of amyloid and atherosclerotic plaques, but a detailed mechanism of the amyloidogenic process remains elusive. To better understand how apoA-I sequence determines its aggregation properties, we performed a comprehensive evaluation of APRs present across vertebrates and mammals. We annotated as APR the regions with a TANGO score above 5% over a stretch of 5 or more residues and established the APR position as the centroid of its sequence range. From the distribution of APRs position across apoA-I sequences (Figure 2A), we evidenced that APRs in vertebrates are concentrated in two hotspots (residues 15-20 and 220-240) but mammalian APRs are almost exclusively located in the N-terminal region. Supported by our TANGO predictions and previous reports (16), we identified three APRs in the human apoA-I sequence, designated here APR1 (residues 14-19), APR2 (residues 53-58), and APR3 (residues 227-232). While the consensus sequences of APR1 and APR3 constitute amyloid peptides verified experimentally in the Waltz database (17) (Figure 2B), APR2 consensus does not represent an amyloid sequence and seems to be specific to the human species. Overall, the pervasive conservation of the N-terminal APR1 suggests an important role in apoA-I structure for this region, as reported for other protein families (15).

Given the fact that several natural variants of apoA-I have been reported to promote an increase in amyloid aggregation, we decided to investigate the impact of these mutations on its aggregation propensity. Although some of these variants have been proven experimentally to form amyloid fibers, we did not observe any increase in the intrinsic aggregation tendency for apoA-I amyloid variants (Supplementary File 1). Moreover, point mutations did not introduce novel APRs on apoA-I sequence.

Based on the human full-length structure of apoA-I (14), we decided to characterize the structural properties of its APRs. We complemented our previous TANGO aggregation measurements with solubility scores provided by the CamSol method, hexapeptide β-zipper tendency available at ZipperDB, residue mean square fluctuation (MSF) calculated with the Gaussian Network Model (GNM) implemented in ProDy and packaging level as measured by the weighted contact number (WCN). ApoA-I APRs showed low solubility (Figure 3A) and a high propensity to form aggregates (β-zipper structures with low rosetta energy) (Figure 3B), which are in agreement with our TANGO results and with previous observations (16). Interestingly, these aggregating regions are located in portions of apoA-I structure with low mobility (Figure 3C) and highly packaged environments, preventing its exposure to the solvent (Figure 3D). This is especially evident for APR1, which is buried inside the N-terminal α-helix bundle (Figure 3E).

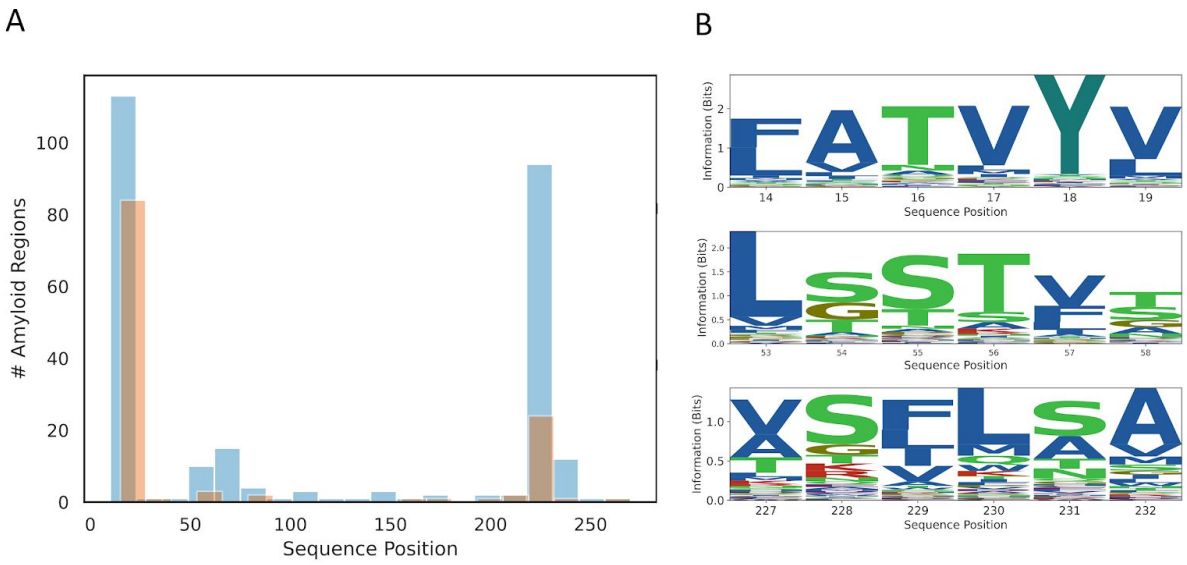


Figure 2 N-terminal APR of apoA-I is highly conserved in mammals

A Histogram showing the sequence distribution of all APRs detected across Vertebrata (blue) and Mammalia (orange). The height of the bars represents the number of APR instances detected by TANGO.

B Sequence logo of mammalian APRs. The size of each position inside the sequence logo is proportional to its conservation (shown as information content).

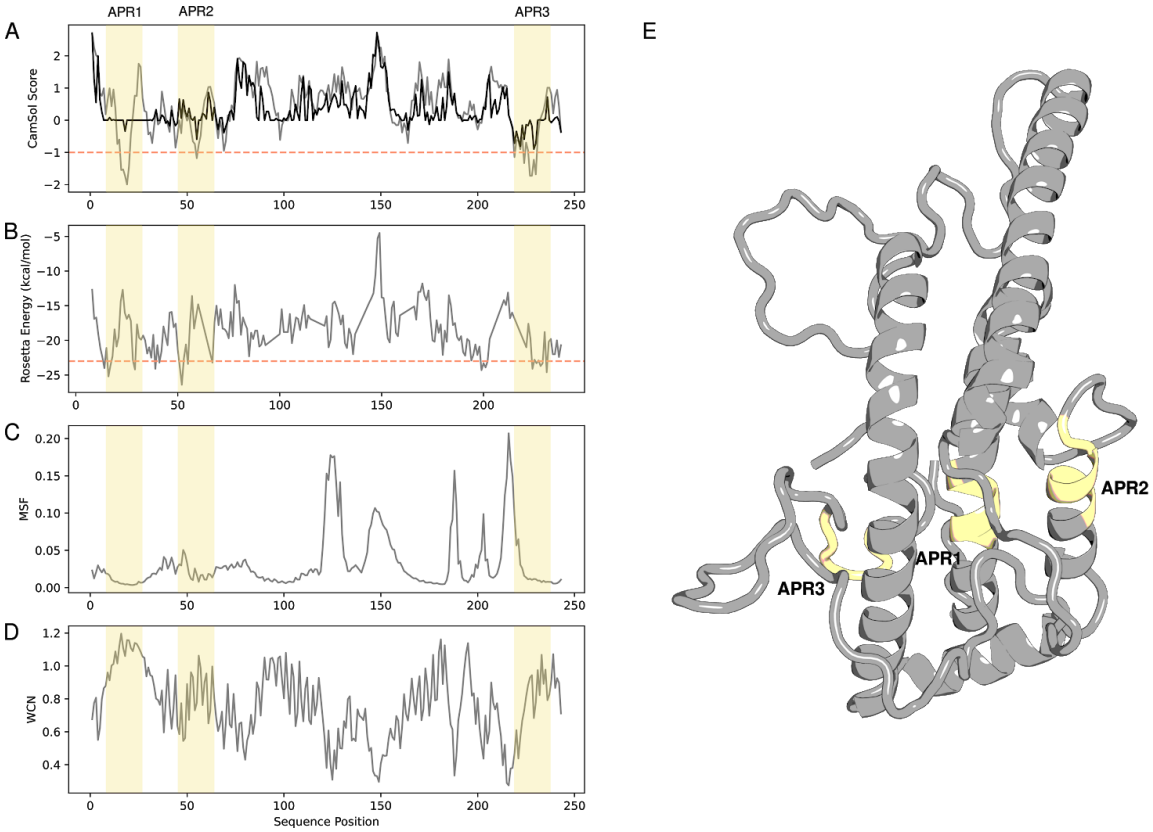


Figure 3 ApoA-I aggregation profile characterized in a structural context.
A-**D** CamSol solubility, β-zipper propensity (expressed as the stabilization energy calculated by Rosetta software), mean square fluctuation (MSF) and weighted contact number (WCN) profiles across the apoA-I sequence, respectively. Aggregating segments are highlighted in orange. For CamSol scores, we included calculations based on sequence-only (grey line) and structural (black line) information (negative scores indicate low solubility). Dashed lines indicate the threshold values used to identify protein sites with low solubility/high β-zipper propensity.
E APRs mapped on the full-length apoA-I 3D structure.

Amyloid-associated variants have a destabilizing effect on apoA-I monomer structure

In order to better understand the structural consequences of amyloid variants on apoA-I monomer, we explored their thermodynamic and pathological effects using *in silico* saturation mutagenesis. Destabilizing effect of each possible mutation in apoA-I sequence, represented by the difference in free energy (ΔΔG) between wild type and mutant structures, was measured using the FoldX empirical force field and the MutateX automation pipeline. To complement this approach, variant pathogenicity probability was estimated using Rhapsody. We noticed from the ΔΔGs distribution that most of the variants had a moderate impact on apoA-I stability (-1 kcal/mol < ΔΔG < 1 kcal/mol) (Figure 4A, complete FoldX results are available with Supplementary Figure 3). Further examination revealed that apoA-I structure is highly sensitive to mutations in the region of residues 7-28, which comprises the APR1 (Figure 4B). Pathogenicity probabilities also support this region as a mutation-sensible segment of apoA-I structure (Supplementary Figure 4). This result suggests that the conservation of APR1 in apoA-I could be necessary to maintain the marginal thermodynamic stability of the α-helix bundle despite the risk to undergo aggregation. In line with our observations, APRs have been recently proposed to play a stabilizing role in protein structure (18).

We used ΔΔG values to highlight differences between pathogenic variants associated with amyloidosis or HDL deficiencies (19), and natural variants reported by the gnomAD project (20). Our results evidenced that amyloid mutations had a destabilizing effect and a pathogenicity probability significantly greater when compared with natural or HDL-deficiency variants (Figure 5A and 5B), emphasizing the relationship between structural destabilization and amyloid pathology onset. An interesting observation from this result is that HDL-deficiency mutations have similar effects compared with natural variants, suggesting that this type of mutations could exert its pathogenic effect with little consequences on apoA-I monomer stability. Given the fact that a small group of variants in the gnomAD database showed an elevated impact on protein stability (> 2 ΔΔG kcal/mol), we decided to investigate how frequently they occur at population level. Frequency spectrum (Figure 5C) showed that variants with a severe impact on protein stability were present at low frequencies, thus minimizing their deleterious effect on the population. In contrast, variants with a higher frequency in our dataset had a nearly neutral effect on stability. It is worth noting that although gnomAD excluded subjects with mendelian and pediatric diseases from its cohorts, we cannot rule out the possibility that some of these destabilizing variants correspond to non-diagnosed pathologies.

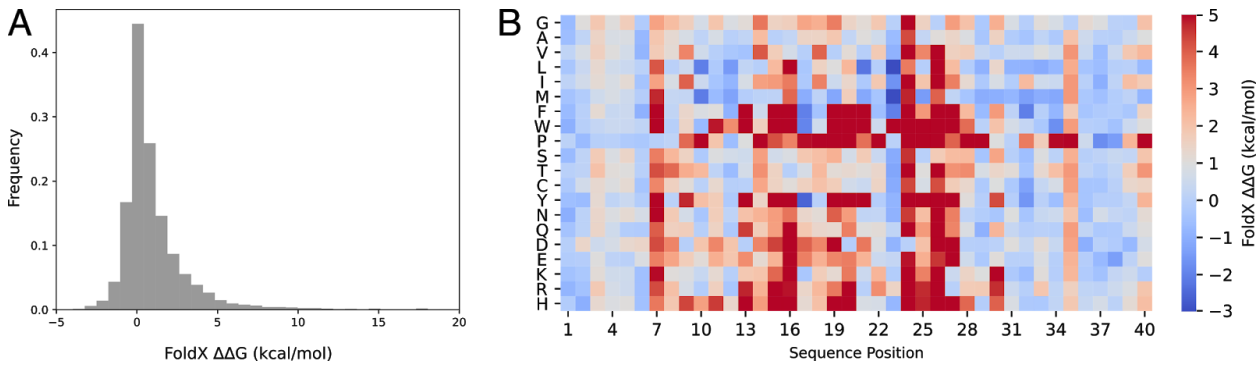


Figure 4 APR1 contributes to the stability of the α-helix bundle in apoA-I

The protein structural stability was quantified using the FoldX engine. The free energy difference (ΔΔG) was calculated by comparison between the ΔG of the mutant and wild type sequence

A ΔΔG values distribution corresponding to all possible mutations.

B Heatmap of ΔΔG values for the first 40 residues of apoA-I N-terminal region.

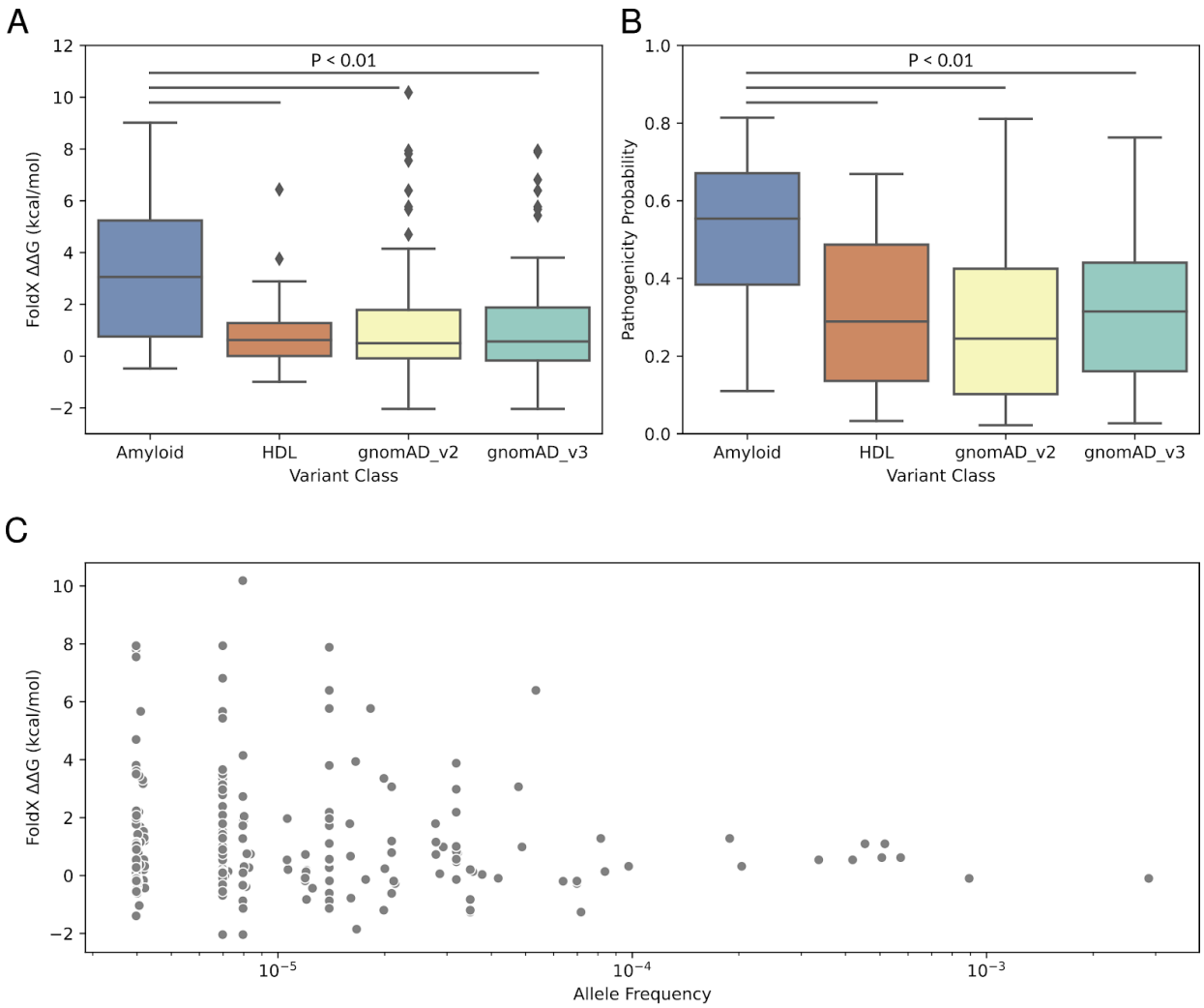


Figure 5 Impact of apoA-I variants on protein stability and pathogenicity probability
A-B Free energy difference (ΔΔG) and pathogenicity probability distributions for each variant class (amyloid, HDL-deficiency and natural, as reported by gnomAD project). Statistical significance was determined using the Mann-Whitney U Test.
C Allele frequency distribution for gnomAD variants against its predicted effect on protein stability.

Molecular dynamics simulations of apoA-I mutants

To complement our previous results showing the destabilizing effect of amyloid variants, we decided to study the dynamic properties of apoA-I amyloid mutants by conducting coarse-grain molecular dynamics simulations under the SIRAH force field. We selected four amyloid mutants (G26R, L60R, Δ107 and R173P) previously characterized by our group (21–24), plus the wild type protein, to prepare our simulation systems. Our selection also ensured that mutations were distributed throughout the apoA-I sequence.

In the first place, we explored the overall dynamics of our system by means of its root mean square deviation (RMSD). The recently described consensus structure for apoA-I was used as reference coordinates for RMSD calculations. We observed a great variability in the RMSD values for all simulated systems (5.4-10 Å) during our 1 μs simulation, which could be related with the highly dynamic and marginally stable structure proposed for apoA-I. We did not find evidence for any significant differences between systems (Supplementary Table 2), suggesting that the impact of point mutations is negligible when compared against the intrinsic backbone dynamics.

Given the structural variability evidenced by RMSD, we decided to compute MD observables over the last 100 ns of the simulations. Position-specific root mean square fluctuations (RMSF) for each of the systems studied showed that loop regions 120-150 and 180-200 are the most flexible regions in apoA-I, while the N-terminal α-helix bundle maintained a more compact structure during the simulation time (Supplementary Figure 5). These results are in good agreement with the MSF values computed by the GNM model (Figure 3C), reinforcing the dynamic profile obtained for apoA-I. The similar fluctuation profiles between the wild type apoA-I and the above-mentioned mutants suggest that mutations do not introduce major structural changes, at least during the simulation time frame.

To further characterize the structural impact of single point mutations on each system we measured the gyration radius (Rg) as a general descriptor of the protein shape for each system. When mutants were compared against the wild type system (Figure 6A), only the L60R system displayed a significantly higher Rg value, indicating a more extended conformation for this mutant. Mutants G26R and R173P also showed a tendency to present greater Rg values when compared against wild type, but they were statistically not significant, in part due to the highly variable nature of the apoA-I system.

We explore the possible role of mutations in amyloid aggregation of full length apoA-I by analyzing the solvent accessible surface area (SASA) of each APR in our five systems. We noted a significant increase in the solvent exposure of the APR1 in the G26R system when compared against the wild type, while the other systems did not exhibit a significant increase of SASA values for any of the APRs (Figure 6B). Visualization of the final time frames of the trajectory corresponding to the G26R system showed a partial unfolding of the α-helix bundle, which explains the increased exposure of APR1 (Figure 6C). Additionally, the G26R mutant evidenced the transitory formation of β-strand secondary structures at the APR3. The low impact of the L60P, Δ107 and R173P variants on APRs exposure suggests that these mutants could require further post-translational modifications in order to undergo native structure unfolding.

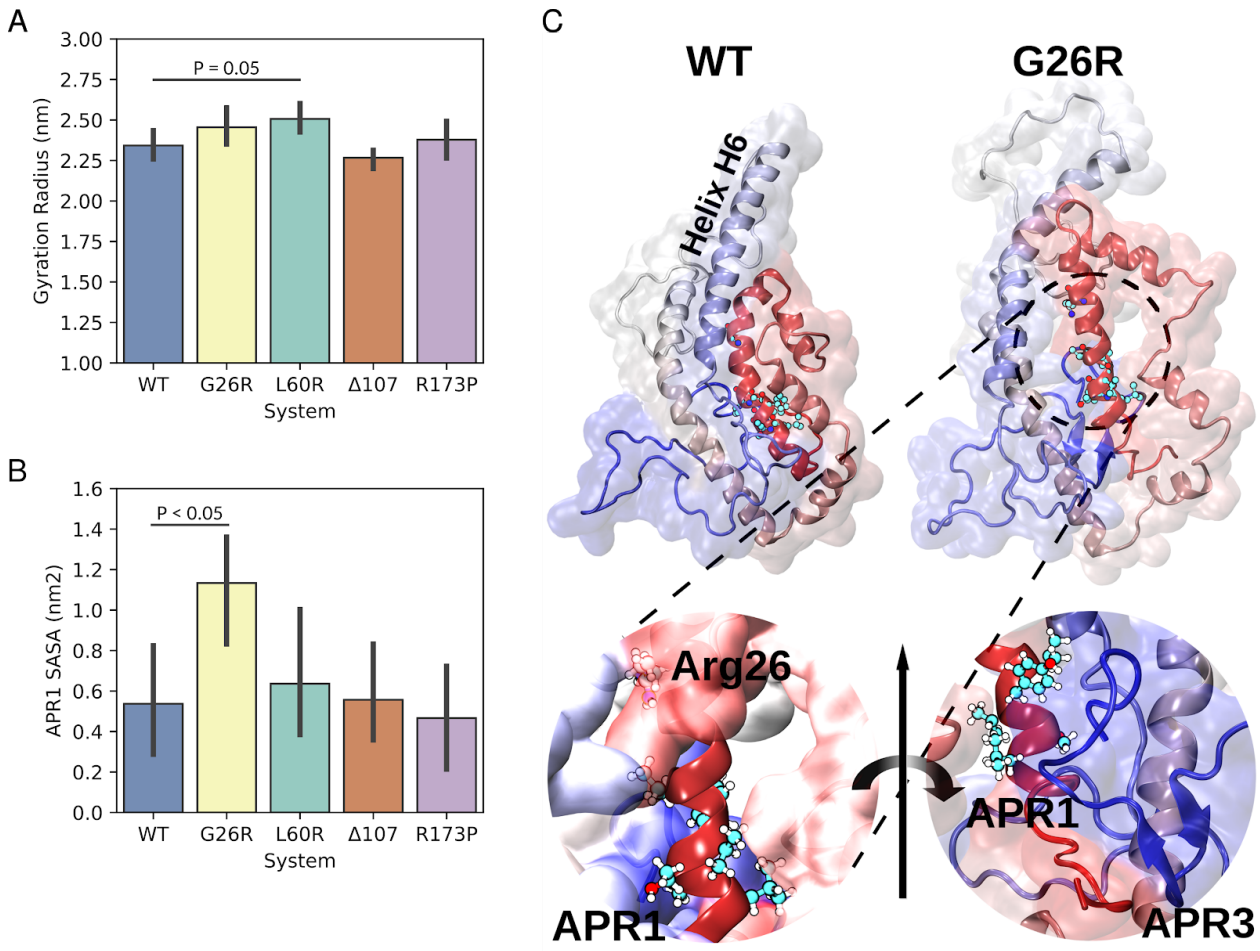


Figure 6 Molecular dynamics simulations of full-length apoA-I mutants

A Gyration radius (Rg) of each system computed over the last 100 nanoseconds from five independent simulations. The L60R mutant showed a higher Rg when compared against the wild type system (p-value ≤ 0.05, Student's Test).

B Solvent accessible surface area (SASA) calculated for the APR1 (residues 14-19). The G26R system displayed a higher APR1 exposure when compared with the wild type system (p-value ≤ 0.05, Student's Test).

C Graphical representation of the consensus model (WT) and the final snapshot of one of the replicas simulated for the G26R mutant. The substitution of glycine by arginine in position 26 destabilizes the helix bundle, expelling helix H6 with the concomitant solvent exposure of APR1 (left inset). A 180º view rotation shows the β-sheet hairpin formed between residues S224 and A232, corresponding to the APR3 (right inset).

Discussion

Molecular mechanism of amyloid aggregation for apoA-I remains largely unknown, due in part to the limited structural information given its inherent conformational plasticity (7). This work builds upon evolutionary, dynamical and structural features of apoA-I in order to provide a comprehensive characterization of the amyloid phenomena in this protein, complementing the extensive experimental results available. Collectively, our results suggest an intimate relationship between aggregation-prone regions and structural stability in apoA-I. Additionally, MD simulations of full-length apoA-I mutants shed light on the first steps of the aggregation process in amyloid mutants.

First, we aimed to complement the few available studies that tackle the evolutionary history of apoA-I and their implications on its structure (25). An important observation that emerges from our results is that apoA-I evolution is tightly linked with the biophysical properties imposed by its constituent amphipathic α-helices. This is especially evident in the case of prolines and positively charged residues, which are critical for apoA-I function and structure. Prolines have been extensively characterized as a fundamental component for apoA-I flexibility and stability, as their positioning at the beginning of the 22-mers induces a relative break of one helical segment respect to the other, allowing the protein rearrangement required for lipid removal and dynamic interactions with membranes and proteins interactors (26). In the case of charged residues, a strong lipid affinity has been attributed to the cationic residues within the polar face of the amphipathic α-helices (27), as they interact with the negative heads of the phospholipids at the surface of lipid bilayers through a process designated "snorkeling" (28, 29). Arginine residues present at the helix 6 are also relevant in apoA-I-mediated activation of LCAT (30). In addition, the conservation of specific leucine positions observed from sequence logos could be driven by its involvement in lipid binding (31, 32) and stabilization of the hydrophobic clusters inside the helix bundle (33). Given that fast evolving regions of proteins have been associated with greater flexibility (34, 35), the higher evolutionary rate observed for the C-terminal region could be linked with the maintenance of the flexibility required for its lipid-binding properties.

The fact that apoA-I has consistently conserved an aggregation-prone segment (termed here APR1) along its evolutionary history raises questions about its structural relevance. Amyloid motifs have been proposed to contribute to protein structural stability through extensive interactions inside protein hydrophobic cores (18, 36), which establish a trade-off between protein environment, foldability and aggregation propensity (37, 38). Based on its conserved nature and FoldX stability results, it is possible to hypothesize that APR1 is necessary to ensure the marginal stability of apoA-I α-helix bundle, even though this region could trigger aggregation upon solvent exposure or proteolytic cleavage (39). Moreover, the presence of APR2 exclusively in human species represents a synergizing factor that could aggravate the amyloid behavior of apoA-I, as demonstrated recently for the N-terminal peptide (40, 41). In this context, the structural features of APR1 (low intrinsic flexibility and highly packaged environment) are likely to control its exposure to solvent and prevent aggregation events. Hydrogen-deuterium exchange experiments (42) supports this highly packaged nature of the α-helix bundle and the low solvent exposure of APR1 in apoA-I.

Amyloidogenic variants are primarily located in the N-terminal region of apoA-I, whereas variants associated with HDL deficiencies are clustered in the H5-H7 region (19, 43). Through a comprehensive evaluation of the destabilizing effect and pathogenicity of each possible mutation

affecting apoA-I we demonstrated that amyloid variants have a significant destabilizing effect on the monomer structure. The fact that TANGO aggregation tendency of APRs was not modified by the introduction of amyloid mutations, supports the hypothesis that aggregation propensity *per se* has a limited impact on the aggregation process of full-length apoA-I (44). On the other hand, variants associated with HDL defects had a minimal effect on structural instability, which provides evidence that this type of disorders could be caused by mechanisms less dependent on protein unfolding and probably involving the disruption of interaction sites with protein interactors during the reverse cholesterol transport pathway.

Taking advantage of the recently described consensus model of apoA-I (14), our MD simulations of mutant G26R revealed a partial unfolding of the N-terminal α-helix bundle and a significant increase in the exposure of APR1, which is also congruent with the destabilizing effect predicted from our ΔΔG calculations. This partial unfolding is in line with the experimental reports of increased susceptibility to proteases (45) and greater solvent exposure of the α-helix bundle (42) for this mutant. Moreover, β-sheet secondary structures present at APR3 could provide a template for the aggregation of full-length apoA-I (16).

Altogether, our results obtained from full-length protein information support the current hypothesis that unfolding of the helix bundle and exposure of aggregating regions represents the first steps of apoA-I-mediated amyloidosis (41). The mild effect of L60R, Δ107 and R173P variants on apoA-I structure and APRs exposure suggest that further modifications could be required to promote protein aggregation of these mutants, like oxidation or proteolytic cleavage (46, 47). Recently, the connection between the pro-inflammatory microenvironment and the formation of aggregation-prone species has been deeply characterized, reinforcing this hypothesis (21). Moreover, the presence of the N-terminal proteolytic fragment (residues 1-93) within patients' lesions raises the hypothesis that mutations may facilitate the cleavage of apoA-I by circulating proteases (48, 49).

Our work highlights the importance of an evolutionary-conserved aggregation-prone region for the stabilization of apoA-I α-helix bundle and suggests a general destabilizing effect for amyloidogenic variants. These results expand our knowledge of the sequence determinants involved in amyloid aggregation mediated by apoA-I. In addition, we hope that our *in silico* saturation mutagenesis results will be useful to guide further experimental explorations of novel apoA-I mutants by others.

Materials and Methods

Evolutionary analysis of apoA-I sequences

A comprehensive dataset of sequences was generated by collating apoA-I orthologs available at Ensembl and Refseq databases (50, 51). To exclude low quality data, only sequences which did not contain ambiguous characters, had a proper methionine starting codon and were longer than 200 amino acids were kept. Additionally, as both Ensembl and Refseq have overlapping data for some species, CD-HIT clustering tool (52) was employed to generate groups of similar sequences with an identity cut-off value of 0.98. Our final dataset comprised 215 protein sequences from vertebrate species.

In order to reconstruct a maximum likelihood phylogeny, a multiple sequence alignment (MSA) was built from the protein sequences using ClustalO with default parameters (53) and the phylogenetic inference was carried out with the IQ-TREE software (54). The substitution model was selected based on the ModelFinder evolutionary model fitting tool (55) and the ultrafast bootstrap implemented in IQ-TREE was used to calculate the support values for phylogeny branches (56). We rooted our phylogeny using cartilaginous fish species as outgroups (57). Visualization of the resulting phylogeny was carried out using the iTOL server (58).

Selective pressure acting on apoA-I sequence

Nucleotide coding sequences were retrieved for each protein in our dataset using the NCBI Entrez eutils tools for Refseq sequences and the Ensembl orthologs dataset. Because the evolutionary rate estimation requires a codon-level alignment, the software PAL2NAL was used to align codons in nucleotide sequence using a protein alignment as a guide (59). The Hypothesis Testing using Phylogenies (HyPhy) package was used to conduct evolutionary analysis on the codon-based alignment. Before testing for evidence of selective pressure, we conducted a recombination analysis using the Genetic Algorithm Recombination Detection (GARD) method (60), in order to screen for possible recombination events in our alignment; it is known that the presence of recombination leads to a larger number of false positives in selection analysis. We inferred the natural selection strength (Omega, *dN/dS*) for each alignment position using our phylogeny as framework. We employed the Fixed Effects Likelihood (FEL) (61) and the Fast Unconstrained Bayesian Approximation (FUBAR) methods (62) to quantify the *dN/dS* ratio for each codon in the alignment. Although both methods provide similar information, FEL provides support for negative selection ($dN/dS < 1$) whereas FUBAR has more statistical power to detect positive selection ($dN/dS > 1$). Because codon alignment positions are difficult to put in structural context, data were extracted for codons occurring in wild type human apoA-I. Additionally, we estimated evolutionary rates based on alignments at amino acid level using LEISR (63).

Coevolving residue pairs

Pairs of residues that are evolutionary correlated (coevolving sites) are useful to predict structural contacts. However, the repetitive structure of apoA-I and the lack of a large number of orthologs poses difficulties for this kind of analysis. To overcome these difficulties, putative coevolving residues were computed using the RaptorX server (64). RaptorX applies an ultra-deep convolutional residual neural network to predict contacts and distance and works particularly well on proteins without many sequence homologs. This method works by predicting the contact/distance matrix as a whole instead of predicting one residue pair independent of the others. RaptorX output represents the probability of two residues being in contact (i.e., their

distance falling in the range 0-8 Å). Only residue pairs with a contact probability greater than 0.5 were retained.

Structural features measurement

Residue solubility profile for apoA-I consensus structure was computed with the CamSol method (65). CamSol first calculates an intrinsic solubility score for each residue, based only on sequence information. Then, the algorithm applies a score correction to the solubility profile from the previous step to account for the spatial proximity of amino acids in the three-dimensional structure and for their solvent exposure.

Fibril-forming segments were identified with the ZipperDB resource (https://services.mbi.ucla.edu/zipperdb/). Fibrillation propensity is calculated as proposed by Thompson (66). Briefly, each hexapeptide not containing a proline from the query sequence is mapped onto the cross-beta crystal structure of the fibril-forming peptide NNQQNY. Energetic fit is evaluated with the RosettaDesign software (67). Hexapeptides with energies below the threshold of -23 kcal/mol were considered as highly propense to fibrillation.

Packaging level for residue *i* was represented by its Weighted Contact Number (WCN), which was calculated as follows:

$$WCN_i = \Sigma_{j \neq i} \frac{1}{r_{ij}^2}$$

Where, *rij* is the distance between the geometric center of the side-chain atoms for residue *i* and residue *j*. Calculations were carried out using a custom script (68).

Protein intrinsic dynamics was characterized using a coarse-grained simulation model based solely on protein topological information represented as a Gaussian Network Model (GNM). In this approach, protein structure is modelled as a network of nodes (alpha carbons) connected by springs. Numerical resolution of this model allows the calculation of the equilibrium displacement for all nodes (Mean Square Fluctuation, MSF), describing the global motions of the system. The ProDy package (69) was used to adjust a GNM to the apoA-I consensus structure. We selected the first ten slow modes for analysis and plotting, since they have been reported previously as the main determinants of the global dynamics of protein structure (70).

Conservation of Aggregation-prone Regions (APRs)

Signal peptide sequences were trimmed and removed from the MSA to retain only the mature protein sequence. TANGO software (71) was used to detect APRs in the protein sequences dataset. This algorithm predicts beta-aggregation using a space phase where the unfolded protein can adopt one of five states: random coil, alpha-helix, beta-turn, alpha-helical aggregation or beta-sheet aggregation. Importantly, TANGO is based on the assumption that the core regions of an aggregate are fully buried. Predictions were carried out using default settings: no protection for the C-terminus and N-terminus, pH 7, temperature of 310 K and ionic strength of 0.1.

Output files provide an aggregation score per position; as suggested in the TANGO manual, contiguous regions comprising five or more residues with a score of at least five were annotated as an APR. To address the impact of single point mutations in apoA-I aggregation tendency we ran TANGO for each mutant sequence and compared the scores profile against the wild type sequence. Sequence logos of each APR were plotted using the LogoMaker package (72). TANGO software was downloaded from http://tango.crg.es using an academic license.

Impact of missense variants on protein stability

The FoldX engine (73) implements an empirical energy function based on terms significant for protein structure stability. The free energy of unfolding (ΔG) of the protein includes terms for van der Waals interactions, solvation of apolar and polar residues, intra and intermolecular hydrogen bonds, water bridges, electrostatic interactions and entropic cost for fixed backbone and side chains. Changes in free energy of folding upon mutation is calculated as the difference between the folding energy (ΔΔG) estimated for the mutants and the wild type variants. Although FoldX seems to be more accurate for the prediction of destabilizing mutations and less accurate for the prediction of stabilizing mutations, in both cases it was shown that FoldX is a valuable tool to infer putative relevant sites for structural stability. FoldX 5 suite was downloaded from http://foldxsuite.crg.eu/academic-license-info.

We employed MutateX software (74) to automate the prediction of ΔΔGs associated with the systematic mutation of each available residue within apoA-I, by employing the FoldX energy function. MutateX automated pipeline engine handles input preparation and performs parallel runs with FoldX. Basic steps involve protein data bank (PDB) structure repair (involving energy minimization to remove unfavorable interactions), model building for the mutant variants, energy calculations for both mutant and wild type structures and summarizing the estimated average free energy differences.

Pathogenicity probability of missense variants

The Rhapsody prediction tool (75) consists of a random forest classifier that combines sequence, structure, and dynamics-based features associated with a given amino acid variant and is trained over a comprehensive dataset of annotated human missense variants. Dynamical features include: mean-square fluctuations of the residue at the mutation site, which estimates local conformational flexibility; perturbation-response scanning effectiveness/sensitivity, accounting for potential allosteric responses involving the mutation site, and the mechanical stiffness at the sequence position of the mutated residue. These properties are computed from Elastic Network Models (ENM) representations of protein structures that describe inter-residue contact topology in a compact and computationally-efficient format that lends itself to a unique analytical solution for each structure. The algorithm was recently upgraded to include coevolutionary features calculated on conserved Pfam domains, and the training dataset was further expanded and refined. The latter combines annotated human variants from several publicly available datasets (Humvar, ExoVar, predictSNP, VariBench, SwissVar, Uniprot's Humsavar and ClinVar). All analyses were performed using the Rhapsody server http://rhapsody.csb.pitt.edu/

Molecular Dynamics Simulations

Coarse grained Molecular Dynamics simulations were performed with the SIRAH force field (76) and GROMACS 2018.4 software package (77). We employed the consensus model of human apoA-I in its monomeric and lipid-free state (14). The PDB file was downloaded from Davidson Lab homepage (http://homepages.uc.edu/~davidswm/structures.html). Mapping atomic to coarse-grained representations was done with a Perl script included in SIRAH Tools (78). G26R, L60R, R173P and Δ107 mutants were generated with Chimera (79), editing the coordinates of the consensus model PDB file. For the case of the deletion mutant, we removed Lys107 and connected residues Lys106 and Trp108 with an unstructured segment using Modloop (80). Wild type apoA-I and the mutant systems were assembled using the following setup: The protein was placed inside an octahedron simulation box defined by setting a distance of 1.5 nm between the

solute and the edges of the box. Systems were solvated setting a 150 mM NaCl concentration following the protocol (81). Energy minimization and heating steps were done following the protocol recommended (76) using positional restraints in the protein backbone to ensure side-chain relaxation, especially in the mutant models. Production runs were performed by quintuplicate in the absence of any positional restraint, generating 1 μs trajectories at 310 K using a 1 bar NPT ensemble. Structural analysis was performed with GROMACS tools gmx rmsf, gmx gyrate and gmx sasa. Root mean square fluctuation was calculated for each residue aligning the full trajectory apoA-I coordinates with the initial models. Radius of gyration and Solvent accessible surface areas (SASA) were obtained averaging the values corresponding to the last 0.1 μs of simulation. The SASA calculations were measured over three amyloid prone regions, comprising residues 14-19 (APR1), 53-58 (APR2) and 227-232 (APR3).

Code Availability

All Python packages used were installed through the Conda environment manager into a single environment. The workflow manager Snakemake was used in the evolutionary analysis in order to gain reproducibility and consistency of the results (82). All data, Snakefile and Python scripts used in this work are available at https://github.com/tomasMasson/APOA1_evolution.

Statistical Analyses and Visualizations

Scipy Python libraries were used for data manipulation and statistical analyses (83). Statistical significance was determined using Mann-Whitney U Test for variant's impact comparison and Student's Test for MD observables. MD graphs are reported as means ± standard deviation derived from five independent experiments. All visualizations were prepared with the Seaborn library (84).

References

1. Lund-Katz S, Phillips MC. High Density Lipoprotein Structure–Function and Role in Reverse Cholesterol Transport. In: Cholesterol Binding and Cholesterol Transport Proteins: [Internet]. Springer Netherlands; 2010. p. 183–227. Available from: https://doi.org/10.1007/978-90-481-8622-8_7
2. Rader DJ, Alexander ET, Weibel GL, Billheimer J, Rothblat GH. The role of reverse cholesterol transport in animals and humans and relationship to atherosclerosis. J Lipid Res [Internet]. 2008;50(Supplement):S189--S194. Available from: https://doi.org/10.1194/jlr.r800088-jlr200
3. Navab M, Reddy ST, Lenten BJ Van, Anantharamaiah GM, Fogelman AM. The role of dysfunctional HDL in atherosclerosis. J Lipid Res [Internet]. 2008 Oct;50(Supplement):S145--S149. Available from: https://doi.org/10.1194/jlr.r800036-jlr200
4. Rosenson RS, Brewer HB, Ansell BJ, Barter P, Chapman MJ, Heinecke JW, et al. Dysfunctional HDL and atherosclerotic cardiovascular disease. Nat Rev Cardiol [Internet]. 2015 Sep;13(1):48–60. Available from: https://doi.org/10.1038/nrcardio.2015.124
5. Chroni A, Liu T, Gorshkova I, Kan H-Y, Uehara Y, Eckardstein A von, et al. The Central Helices of ApoA-I Can Promote ATP-binding Cassette Transporter A1 (ABCA1)-mediated Lipid Efflux. J Biol Chem [Internet]. 2002;278(9):6719–30. Available from: https://doi.org/10.1074/jbc.m205232200
6. Manthei KA, Patra D, Wilson CJ, Fawaz M V, Piersimoni L, Shenkar JC, et al. Structural analysis of lecithin:cholesterol acyltransferase bound to high density lipoprotein particles. Commun Biol [Internet]. 2020;3(1). Available from: https://doi.org/10.1038/s42003-019-0749-z
7. Gursky O, Atkinson D. Thermal unfolding of human high-density apolipoprotein A-1: implications for a lipid-free molten globular state. Proc Natl Acad Sci U S A [Internet]. 1996 Apr 2;93(7):2991–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=39748&tool=pmcentrez&rendertype=abstract
8. Sipe JD, Benson MD, Buxbaum JN, Ikeda S, Merlini G, Saraiva MJM, et al. Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines. Amyloid [Internet]. 2016 Oct;23(4):209–13. Available from: https://doi.org/10.1080/13506129.2016.1257986
9. Das M, Gursky O. Amyloid-Forming Properties of Human Apolipoproteins: Sequence Analyses and Structural Insights. In: Advances in Experimental Medicine and Biology [Internet]. Springer International Publishing; 2015. p. 175–211. Available from: https://doi.org/10.1007/978-3-319-17344-3_8
10. Mucchiano GI, Häggqvist B, Sletten K, Westermark P. Apolipoprotein A-1-derived amyloid in atherosclerotic plaques of the human aorta. J Pathol. 2001;193(2000):270–5.
11. Obici L, Franceschini G, Calabresi L, Giorgetti S, Stoppini M, Merlini G, et al. Structure, function and amyloidogenic propensity of apolipoprotein A-I. Amyloid [Internet]. 2006 Dec [cited 2012 Jun 26];13(4):191–205. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17107880
12. Gaglione R, Smaldone G, Di Girolamo R, Piccoli R, Pedone E, Arciello A. Cell milieu significantly affects the fate of AApoAI amyloidogenic variants: predestination or serendipity? Biochim Biophys Acta - Gen Subj [Internet]. 2018 Mar [cited 2019 Jun 20];1862(3):377–84. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0304416517303835
13. Westermark PP, Mucchiano GG, Marthin TT, Johnson KHKH, Sletten KK. Apolipoprotein A1-derived amyloid in human aortic atherosclerotic plaques. Am J Pathol [Internet]. 1995 Nov;147(5):1186–92. Available from: http://pubget.com/paper/7485381

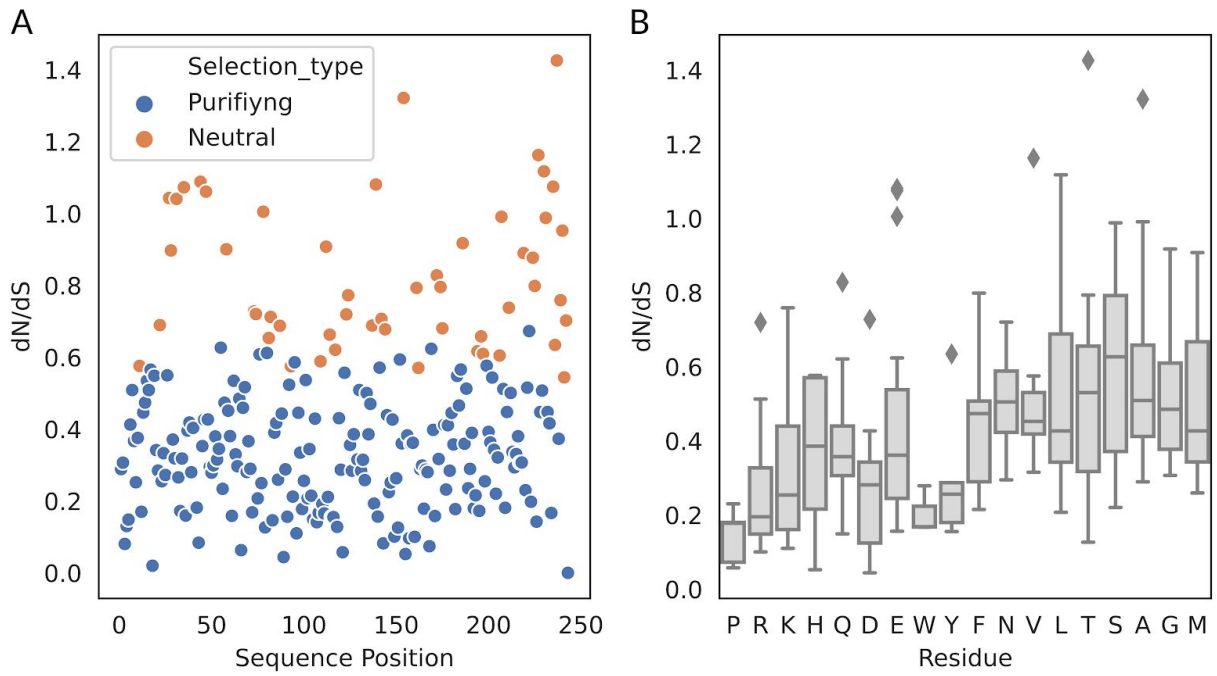
14. Melchior JT, Walker RG, Cooke AL, Morris J, Castleberry M, Thompson TB, et al. A consensus model of human apolipoprotein A-I in its monomeric and lipid-free state. Nat Struct Mol Biol [Internet]. 2017 Nov;24(12):1093–9. Available from: https://doi.org/10.1038/nsmb.3501
15. Buck PM, Kumar S, Singh SK. On the Role of Aggregation Prone Regions in Protein Evolution, Stability, and Enzymatic Catalysis: Insights from Diverse Analyses. Iakoucheva LM, editor. PLoS Comput Biol [Internet]. 2013 Oct;9(10):e1003291. Available from: https://doi.org/10.1371/journal.pcbi.1003291
16. Das M, Mei X, Jayaraman S, Atkinson D, Gursky O. Amyloidogenic mutations in human apolipoprotein A-I are not necessarily destabilizing - A common mechanism of apolipoprotein A-I misfolding in familial amyloidosis and atherosclerosis. FEBS J. 2014;281(11):2525–42.
17. Louros N, Konstantoulea K, De Vleeschouwer M, Ramakers M, Schymkowitz J, Rousseau F. WALTZ-DB 2.0: an updated database containing structural information of experimentally determined amyloid-forming peptides. Nucleic Acids Res [Internet]. 2019 Sep;48(D1):D389--D393. Available from: https://doi.org/10.1093/nar/gkz758
18. Langenberg T, Gallardo R, der Kant R van, Louros N, Michiels E, Duran-Romaña R, et al. Thermodynamic and Evolutionary Coupling between the Native and Amyloid State of Globular Proteins. Cell Rep [Internet]. 2020;31(2):107512. Available from: https://doi.org/10.1016/j.celrep.2020.03.076
19. Gogonea V. Structural insights into high density lipoprotein: Old models and new facts. Vol. 6, Frontiers in Pharmacology. Frontiers Media S.A.; 2016.
20. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature [Internet]. 2020 May;581(7809):434–43. Available from: https://doi.org/10.1038/s41586-020-2308-7
21. Gisonno RA, Prieto ED, Gorgojo JP, Curto LM, Rodriguez ME, Rosú SA, et al. Fibrillar conformation of an apolipoprotein A-I variant involved in amyloidosis and atherosclerosis. Biochim Biophys Acta - Gen Subj [Internet]. 2020;1864(4):129515. Available from: https://doi.org/10.1016/j.bbagen.2020.129515
22. Ramella NA, Schinella GR, Ferreira ST, Prieto ED, Vela ME, Ríos JL, et al. Human Apolipoprotein A-I Natural Variants: Molecular Mechanisms Underlying Amyloidogenic Propensity. Uversky VN, editor. PLoS One [Internet]. 2012 Aug 28 [cited 2018 Jan 28];7(8):e43755. Available from: http://dx.plos.org/10.1371/journal.pone.0043755
23. Rosú SA, Rimoldi OJ, Prieto ED, Curto LM. Amyloidogenic Propensity of a Natural Variant of Human Apolipoprotein A-I : Stability and Interaction with Ligands. 2015;1–17.
24. Gaddi GM, Gisonno RA, Rosú SA, Curto LM, Prieto ED, Schinella GR, et al. Structural analysis of a natural apolipoprotein A-I variant (L60R) associated with amyloidosis. Arch Biochem Biophys [Internet]. 2020 May;685:108347. Available from: https://doi.org/10.1016/j.abb.2020.108347
25. Bashtovyy D, Jones MK, Anantharamaiah GM, Segrest JP. Sequence conservation of apolipoprotein A-I affords novel insights into HDL structure-function. J Lipid Res. 2011;52(3):435–50.
26. Klon AE, Segrest JP, Harvey SC. Molecular Dynamics Simulations on Discoidal HDL Particles Suggest a Mechanism for Rotation in the Apo A-I Belt Model. J Mol Biol [Internet]. 2002;324(4):703–21. Available from: https://doi.org/10.1016/s0022-2836%2802%2901143-9
27. Fuentes LA, Beck WHJ, Tsujita M, Weers PMM. Charged Residues in the C-Terminal Domain of Apolipoprotein A-I Modulate Oligomerization. Biochemistry [Internet]. 2018 Mar;57(15):2200–10. Available from: https://doi.org/10.1021/acs.biochem.7b01052
28. Leman LJ, Maryanoff BE, Ghadiri MR. Molecules That Mimic Apolipoprotein A-I: Potential

Agents for Treating Atherosclerosis. J Med Chem [Internet]. 2013 Oct;57(6):2169–96. Available from: https://doi.org/10.1021/jm4005847
29. Oda MN. Lipid-free apoA-I structure - Origins of model diversity. Vol. 1862, Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids. Elsevier B.V.; 2017. p. 221–33.
30. Roosbeek S, Vanloo B, Duverger N, Caster H, Breyne J, De Beun I, et al. Three arginine residues in apolipoprotein A-I are critical for activation of lecithin:cholesterol acyltransferase. J Lipid Res. 2001 Jan;42(1):31–40.
31. Hovingh GK, Brownlie A, Bisoendial RJ, Dube MP, Levels JHM, Petersen W, et al. A novel apoA-I mutation (L178P) leads to endothelial dysfunction, increased arterial wall thickness, and premature coronary artery disease. J Am Coll Cardiol [Internet]. 2004;44(7):1429–35. Available from: https://www.onlinejacc.org/content/44/7/1429
32. Fotakis P, Kateifides AK, Gkolfinopoulou C, Georgiadou D, Beck M, Gründler K, et al. Role of the hydrophobic and charged residues in the 218-226 region of apoA-I in the biogenesis of HDL. J Lipid Res. 2013 Dec;54(12):3281–92.
33. Gursky O, Mei X, Atkinson D. The Crystal Structure of the C-Terminal Truncated Apolipoprotein A-I Sheds New Light on Amyloid Formation by the N-Terminal Fragment. 2012;(Table 1).
34. Tiwari SP, Reuter N. Conservation of intrinsic dynamics in proteins — what have computational models taught us? Curr Opin Struct Biol [Internet]. 2018 Jun;50:75–81. Available from: https://doi.org/10.1016/j.sbi.2017.12.001
35. Campitelli P, Modi T, Kumar S, Ozkan SB. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. Annu Rev Biophys [Internet]. 2020 May;49(1):267–88. Available from: https://doi.org/10.1146/annurev-biophys-052118-115517
36. Tartaglia GG, Vendruscolo M. Proteome-Level Interplay between Folding and Aggregation Propensities of Proteins. J Mol Biol [Internet]. 2010 Oct;402(5):919–28. Available from: https://doi.org/10.1016/j.jmb.2010.08.013
37. Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L. A Comparative Study of the Relationship Between Protein Structure and $p˘beta$-Aggregation in Globular and Intrinsically Disordered Proteins. J Mol Biol [Internet]. 2004 Sep;342(1):345–53. Available from: https://doi.org/10.1016/j.jmb.2004.06.088
38. Monsellier E, Ramazzotti M, Taddei N, Chiti F. Aggregation Propensity of the Human Proteome. Nussinov R, editor. PLoS Comput Biol [Internet]. 2008 Oct;4(10):e1000199. Available from: https://doi.org/10.1371/journal.pcbi.1000199
39. Arciello A, Piccoli R, Monti DM. Apolipoprotein A-I: the dual face of a protein. FEBS Lett [Internet]. 2016 Dec [cited 2019 Jun 20];590(23):4171–9. Available from: http://doi.wiley.com/10.1002/1873-3468.12468
40. Wong YQ, Binger KJ, Howlett GJ, Griffin MDW. Identification of an amyloid fibril forming peptide comprising residues 46-59 of apolipoprotein A-I. FEBS Lett [Internet]. 2012 Jun 21 [cited 2012 Nov 2];586(13):1754–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22609356
41. Mizuguchi C, Nakagawa M, Namba N, Sakai M, Kurimitsu N, Suzuki A, et al. Mechanisms of aggregation and fibril formation of the amyloidogenic N-terminal fragment of apolipoprotein A-I. J Biol Chem [Internet]. 2019 Jul;294(36):13515–24. Available from: https://doi.org/10.1074/jbc.ra119.008000
42. Das M, Wilson CJ, Mei X, Wales TE, Engen JR, Gursky O. Structural Stability and Local Dynamics in Disease-Causing Mutants of Human Apolipoprotein A-I: What Makes the Protein Amyloidogenic? J Mol Biol [Internet]. 2016;428(2):449–62. Available from: http://dx.doi.org/10.1016/j.jmb.2015.10.029
43. Matsunaga A, Uehara Y, Zhang B, Saku K. Apolipoprotein A-I mutations [Internet]. First Edit. The HDL Handbook. Elsevier Inc.; 2010 [cited 2014 Nov 17]. 133–151 p. Available

from: http://dx.doi.org/10.1016/B978-0-12-382171-3.10007-5
44. Raimondi S, Guglielmi F, Giorgetti S, Gaetano S Di, Arciello A, Monti DM, et al. Effects of the known pathogenic mutations on the aggregation pathway of the amyloidogenic peptide of apolipoprotein A-I. J Mol Biol [Internet]. 2011;407(3):465–76. Available from: http://dx.doi.org/10.1016/j.jmb.2011.01.044
45. Adachi E, Nakajima H, Mizuguchi C, Dhanasekaran P, Kawashima H, Nagao K, et al. Dual Role of an N-terminal Amyloidogenic Mutation in Apolipoprotein A-I. J Biol Chem [Internet]. 2012;288(4):2848–56. Available from: https://doi.org/10.1074/jbc.m112.428052
46. Witkowski A, Chan GKL, Boatz JC, Li NJ, Inoue AP, Wong JC, et al. Methionine oxidized apolipoprotein A-I at the crossroads of HDL biogenesis and amyloid formation. FASEB J [Internet]. 2018;32(6):3149–65. Available from: https://doi.org/10.1096/fj.201701127r
47. Chan GKL, Witkowski A, Gantz DL, Zhang TO, Zanni MT, Jayaraman S, et al. Myeloperoxidase-mediated Methionine Oxidation Promotes an Amyloidogenic Outcome for Apolipoprotein A-I *. 2015;290(17):10958–71.
48. Cavigiolio G, Jayaraman S. Proteolysis of Apolipoprotein A-I by Secretory Phospholipase A2. J Biol Chem [Internet]. 2014 Feb;289(14):10011–23. Available from: https://doi.org/10.1074/jbc.m113.525717
49. Kareinen I, Baumann M, Nguyen SD, Maaninka K, Anisimov A, Tozuka M, et al. Chymase released from hypoxia-activated cardiac mast cells cleaves human apoA-I at Tyr192 and compromises its cardioprotective activity. J Lipid Res [Internet]. 2018 Mar;59(6):945–57. Available from: https://doi.org/10.1194/jlr.m077503
50. O′Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res [Internet]. 2015 Nov;44(D1):D733--D745. Available from: https://doi.org/10.1093/nar/gkv1189
51. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res [Internet]. 2019 Nov; Available from: https://doi.org/10.1093/nar/gkz966
52. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics [Internet]. 2012 Oct;28(23):3150–2. Available from: https://doi.org/10.1093/bioinformatics/bts565
53. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol [Internet]. 2011;7(1):539. Available from: https://doi.org/10.1038/msb.2011.75
54. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Haeseler A von, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Teeling E, editor. Mol Biol Evol [Internet]. 2020 Feb;37(5):1530–4. Available from: https://doi.org/10.1093/molbev/msaa015
55. Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods [Internet]. 2017 May;14(6):587–9. Available from: https://doi.org/10.1038/nmeth.4285
56. Minh BQ, Nguyen MAT, Haeseler A von. Ultrafast Approximation for Phylogenetic Bootstrap. Mol Biol Evol [Internet]. 2013 Feb;30(5):1188–95. Available from: https://doi.org/10.1093/molbev/mst024
57. de Carvalho LL, Bligt-Lindén E, Ramaiah A, Johnson MS, Salminen TA. Evolution and functional classification of mammalian copper amine oxidases. Mol Phylogenet Evol [Internet]. 2019 Oct;139:106571. Available from: https://doi.org/10.1016/j.ympev.2019.106571
58. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res [Internet]. 2019;47(W1):W256--W259. Available from: https://doi.org/10.1093/nar/gkz239
59. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence

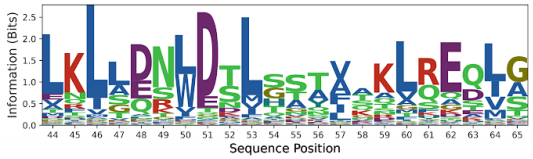
alignments into the corresponding codon alignments. Nucleic Acids Res [Internet]. 2006 Jul;34(Web Server):W609--W612. Available from: https://doi.org/10.1093/nar/gkl315
60. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. Mol Biol Evol [Internet]. 2006 Jul;23(10):1891–901. Available from: https://doi.org/10.1093/molbev/msl051
61. Pond SLK, Frost SDW. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol Biol Evol [Internet]. 2005 Feb;22(5):1208–22. Available from: https://doi.org/10.1093/molbev/msi105
62. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, et al. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Mol Biol Evol [Internet]. 2013 Feb;30(5):1196–205. Available from: https://doi.org/10.1093/molbev/mst030
63. Spielman SJ, Pond SLK. Relative evolutionary rate inference in HyPhy with LEISR. PeerJ [Internet]. 2018 Feb;6:e4339. Available from: https://doi.org/10.7717/peerj.4339
64. Wang S, Sun S, Li Z, Zhang R, Xu J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. Schlessinger A, editor. PLOS Comput Biol [Internet]. 2017;13(1):e1005324. Available from: https://doi.org/10.1371/journal.pcbi.1005324
65. Sormanni P, Aprile FA, Vendruscolo M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. J Mol Biol [Internet]. 2015;427(2):478–90. Available from: https://doi.org/10.1016/j.jmb.2014.09.026
66. Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, Eisenberg D. The 3D profile method for identifying fibril-forming segments of proteins. Proc Natl Acad Sci [Internet]. 2006 Mar;103(11):4074–8. Available from: https://doi.org/10.1073/pnas.0511295103
67. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci [Internet]. 2000 Sep;97(19):10383–8. Available from: https://doi.org/10.1073/pnas.97.19.10383
68. Sydykova DK, Jack BR, Spielman SJ, Wilke CO. Measuring evolutionary rates of proteins in a structural context. F1000Research [Internet]. 2018 Feb;6:1845. Available from: https://doi.org/10.12688/f1000research.12874.2
69. Bakan A, Meireles LM, Bahar I. ProDy: Protein Dynamics Inferred from Theory and Experiments. Bioinformatics [Internet]. 2011;27(11):1575–7. Available from: https://doi.org/10.1093/bioinformatics/btr168
70. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol [Internet]. 1999;9(2):164–9. Available from: https://doi.org/10.1016/s0959-440x%2899%2980023-2
71. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol. 2004 Oct 12;22(10):1302–6.
72. Tareen A, Kinney JB. Logomaker: beautiful sequence logos in Python. Valencia A, editor. Bioinformatics [Internet]. 2019;36(7):2272–4. Available from: https://doi.org/10.1093/bioinformatics/btz921
73. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. J Mol Biol. 2002 Jul 5;320(2):369–87.
74. Tiberti M, Terkelsen T, Cremers TC, Marco M Di, Piedade I da, Maiani E, et al. MutateX: an automated pipeline for in-silico saturation mutagenesis of protein structures and structural ensembles. 2019 Nov; Available from: https://doi.org/10.1101/824938
75. Ponzoni L, Peñaherrera DA, Oltvai ZN, Bahar I. Rhapsody: predicting the pathogenicity of human missense variants. Ponty Y, editor. Bioinformatics [Internet]. 2020 Feb;36(10):3084–92. Available from: https://doi.org/10.1093/bioinformatics/btaa127
76. Machado MR, Barrera EE, Klein F, Sóñora M, Silva S, Pantano S. The SIRAH 2.0 Force

Field: Altius, Fortius, Citius. J Chem Theory Comput [Internet]. 2019 Feb;15(4):2719–33. Available from: https://doi.org/10.1021/acs.jctc.9b00006
77. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX [Internet]. 2015 Sep;1–2:19–25. Available from: https://doi.org/10.1016/j.softx.2015.06.001
78. Machado MR, Pantano S. SIRAH tools: mapping, backmapping and visualization of coarse-grained models. Bioinformatics [Internet]. 2016;32(10):1568–70. Available from: https://doi.org/10.1093/bioinformatics/btw020
79. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera?A visualization system for exploratory research and analysis. J Comput Chem [Internet]. 2004;25(13):1605–12. Available from: https://doi.org/10.1002/jcc.20084
80. Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. Bioinformatics [Internet]. 2003;19(18):2500–1. Available from: https://doi.org/10.1093/bioinformatics/btg362
81. Machado MR, Pantano S. Split the Charge Difference in Two! A Rule of Thumb for Adding Proper Amounts of Ions in MD Simulations. J Chem Theory Comput [Internet]. 2020;16(3):1367–72. Available from: https://doi.org/10.1021/acs.jctc.9b00953
82. Koster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. Bioinformatics [Internet]. 2012;28(19):2520–2. Available from: https://doi.org/10.1093/bioinformatics/bts480
83. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods [Internet]. 2020;17(3):261–72. Available from: https://doi.org/10.1038/s41592-019-0686-2
84. Waskom M, the seaborn development team. mwaskom/seaborn [Internet]. Zenodo; 2020. Available from: https://doi.org/10.5281/zenodo.592845

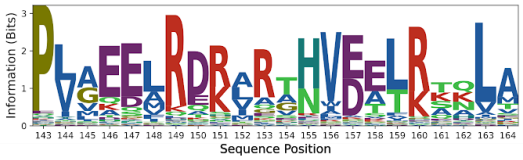


Supplementary Figure 1 Apoa-I evolutionary rates based on codon alignments
A Evolutionary rate (*dN/dS*) profile for apoA-I coding sequence, estimated with FEL (FUBAR results are highly similar). Points colored in blue indicate the presence of purifying selection, while orange indicates neutral selection at a residue position.
B Evolutionary rate (*dN/dS*) for each residue type inside apoA-I tandem repeats. Proline and positively charged residues (K and R) display values consistent with a more stringent conservation.

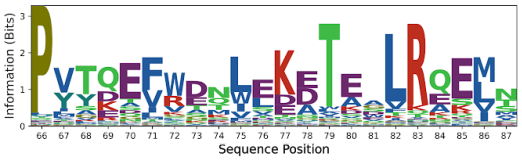
Repeat 1



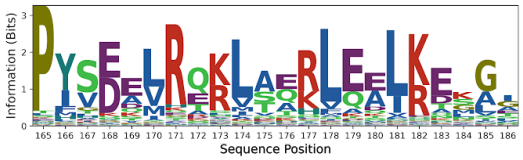
Repeat 6



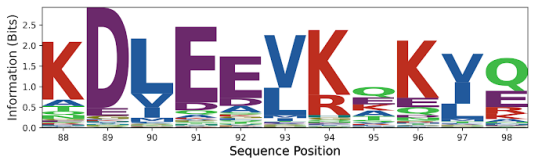
Repeat 2



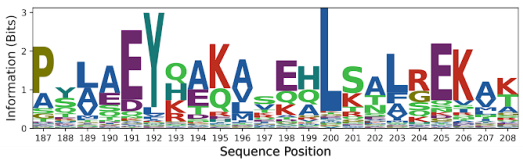
Repeat 7



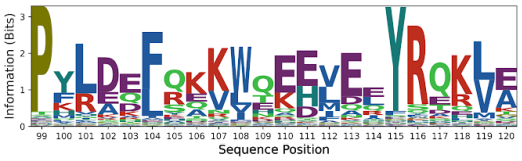
Repeat 3



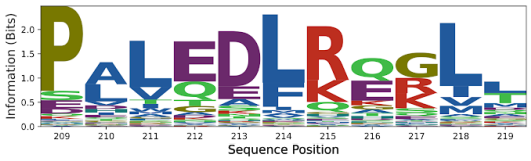
Repeat 8



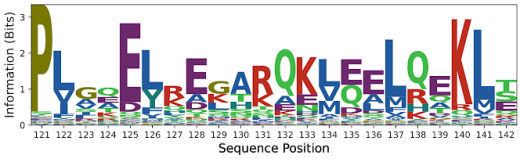
Repeat 4



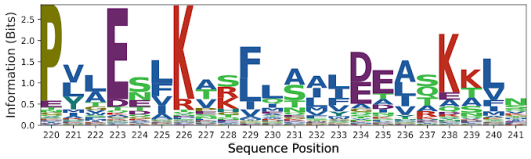
Repeat 9



Repeat 5

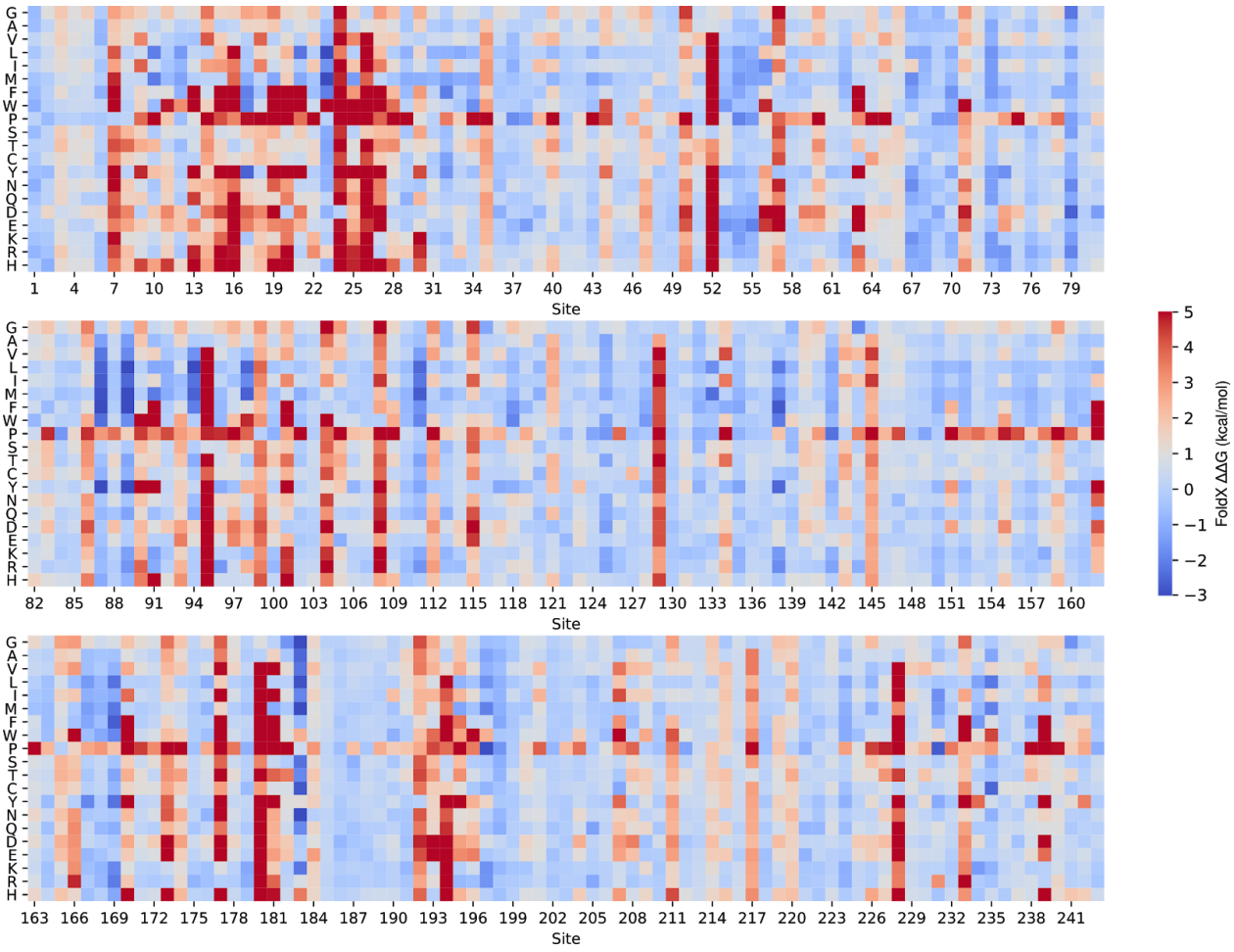


Repeat 10



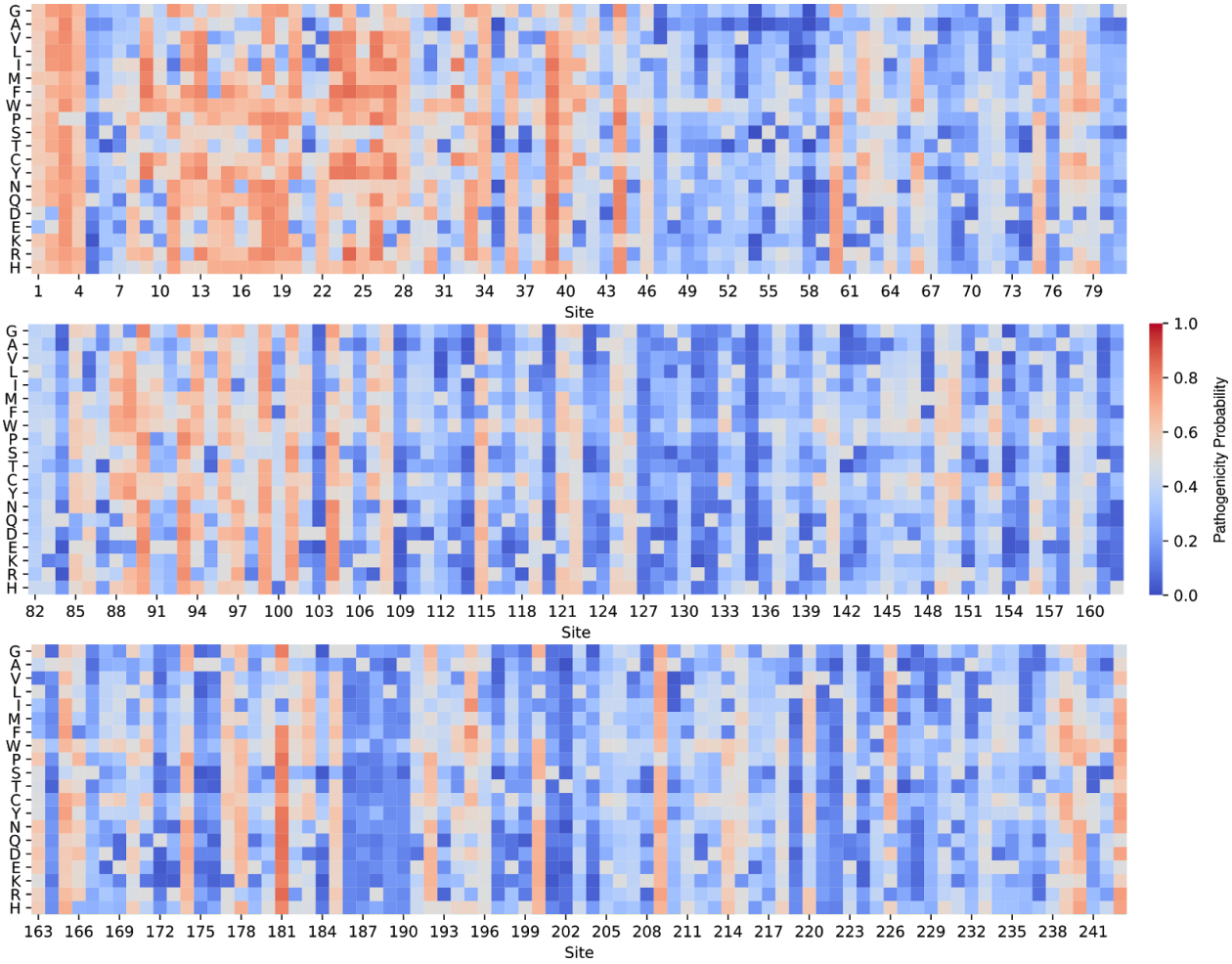
Supplementary Figure 2 Sequence logos for apoA-I tandem repeats

Sequence logos corresponding to each of the ten tandem repeats present in apoA-I (residues 48-243) were extracted from a multiple sequence alignment of mammalian sequence using the LogoMaker package. Position-specific conservation was calculated using information content (expressed in Bits).



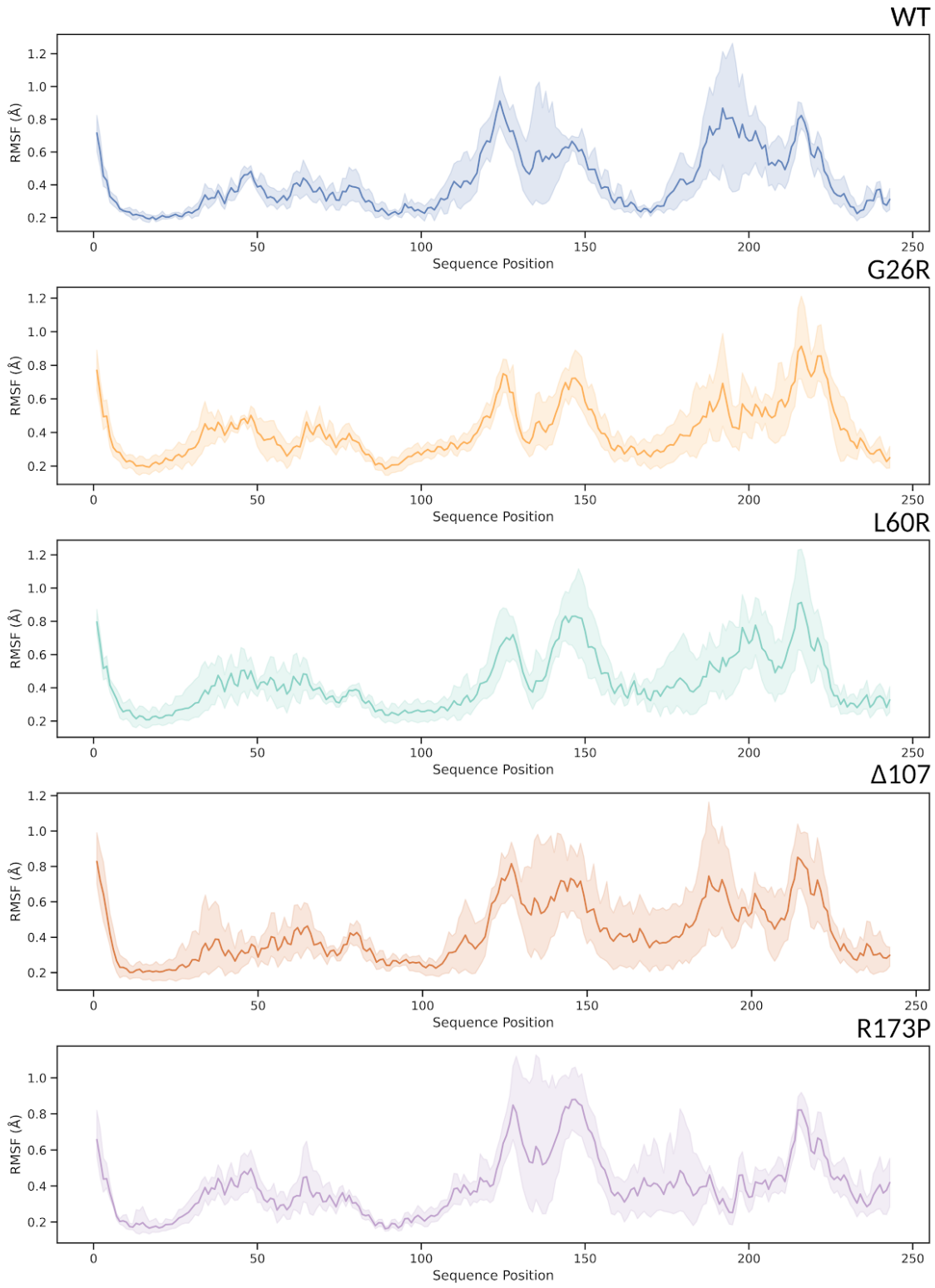
Supplementary Figure 3 FoldX Thermodynamic Destabilization landscape

ΔΔG values obtained by *in silico* saturation mutagenesis of apoA-I structure using the FoldX engine. Mutation introduced is depicted in the Y axis. Scales at the right indicate ΔΔG values expressed in kcal/mol.



Supplementary Figure 4 Rhapsody Pathogenicity landscape

Pathogenicity values obtained by *in silico* saturation mutagenesis of apoA-I structure using the Rhapsody package. Mutation introduced is depicted in the Y axis. Scales at the right indicate pathogenicity probability (1 more pathogenic, 0 less pathogenic and 0.5 for mutations introducing the same residue present in the wild type sequence).



Supplementary Figure 5 Root mean square fluctuation (RMSF) profiles for apoA-I mutants
RMSF values were computed for each protein position over the last 100 ns of the simulation. Mean values are depicted together with its standard deviation.

Supplementary Table 1 Coevolving residue pairs for apoA-I

res_i	res_j	Prob
32	38	0.9745979
32	41	0.9356469
29	42	0.8026068
32	42	0.7793443
10	63	0.7535808
21	52	0.7366209
33	42	0.7339497
29	45	0.7262684
14	59	0.6868986
28	45	0.6604719
25	45	0.6544870
18	56	0.6307588
18	52	0.6276968
25	49	0.6210243
14	56	0.6128104
21	49	0.6127903
14	63	0.6056398
33	39	0.6045238
7	70	0.6029568
10	66	0.6011602
10	67	0.5999863
22	49	0.5937240
7	67	0.5832257
3	9	0.5820144
7	63	0.5752648
29	46	0.5712630
18	53	0.5689167
6	66	0.5582736
10	70	0.5409693
14	228	0.5365205
6	70	0.5255757
10	59	0.5169616
28	41	0.5164070
11	60	0.5149656
32	45	0.5145751
7	66	0.5117174
11	63	0.5069553
17	59	0.5011830

Supplementary Table 2 Root mean square deviation values from molecular dynamics simulations

System	Root Mean Square Deviation (RMSD, Å)						
	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5	Mean	Standard Deviation
Wild type (WT)	7.91	7.24	8.04	5.39	7.92	7.30	1.11
G26R	6.75	10.33	7.54	7.65	6.10	7.67	1.61
L60R	6.61	6.11	10.28	8.30	7.87	7.83	1.63
Δ107	7.28	6.84	7.77	6.03	8.79	7.34	1.03
R173P	8.97	8.24	7.14	6.91	8.16	7.88	0.85