

Supporting Information

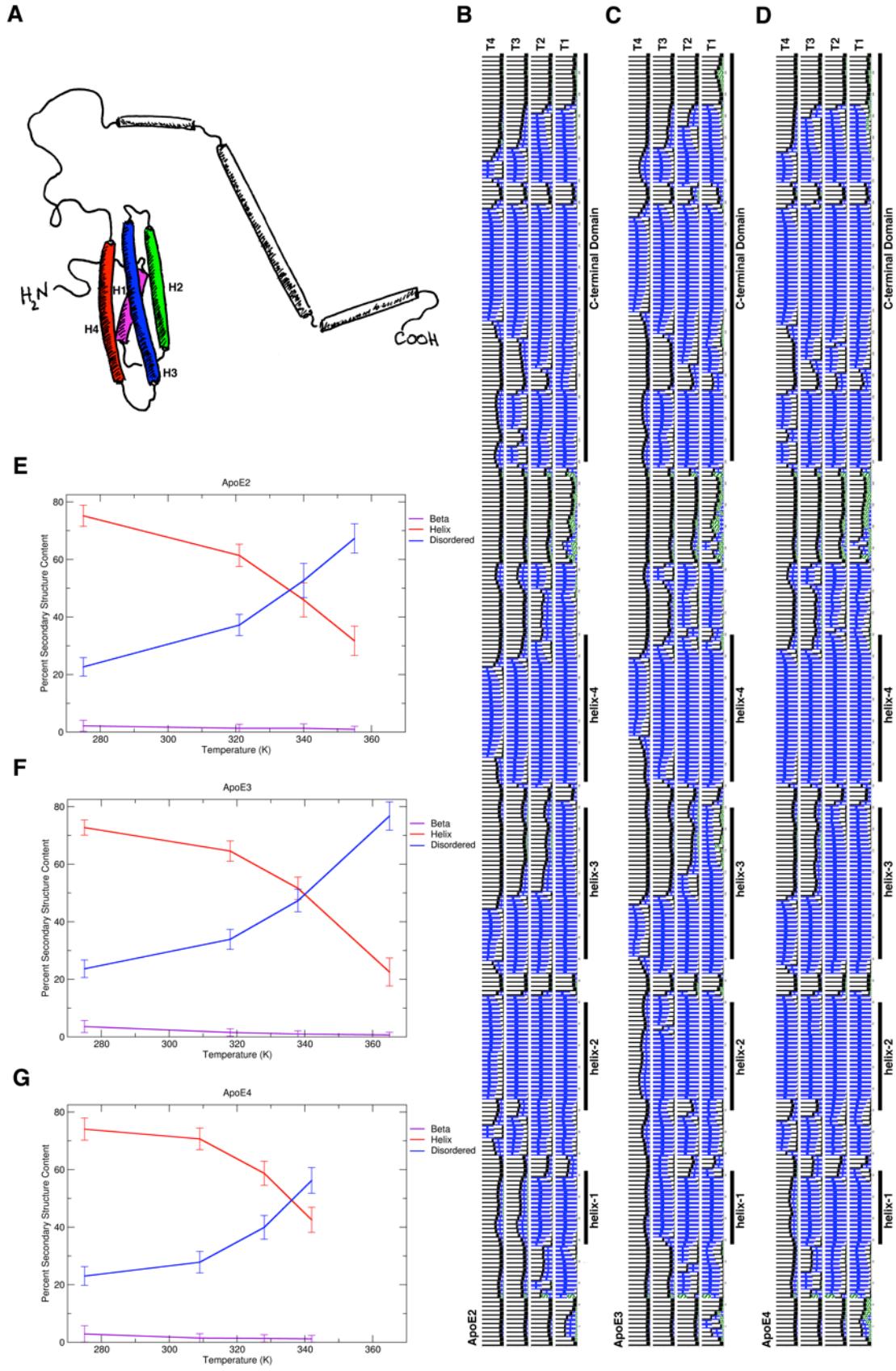
ApoE4-specific misfolded intermediate identified by molecular dynamics simulations

Benfeard Williams II, Marino Convertino, Jhuma Das, and Nikolay V. Dokholyan*

Biochemistry and Biophysics Department, University of North Carolina, Chapel Hill, NC, United States of America

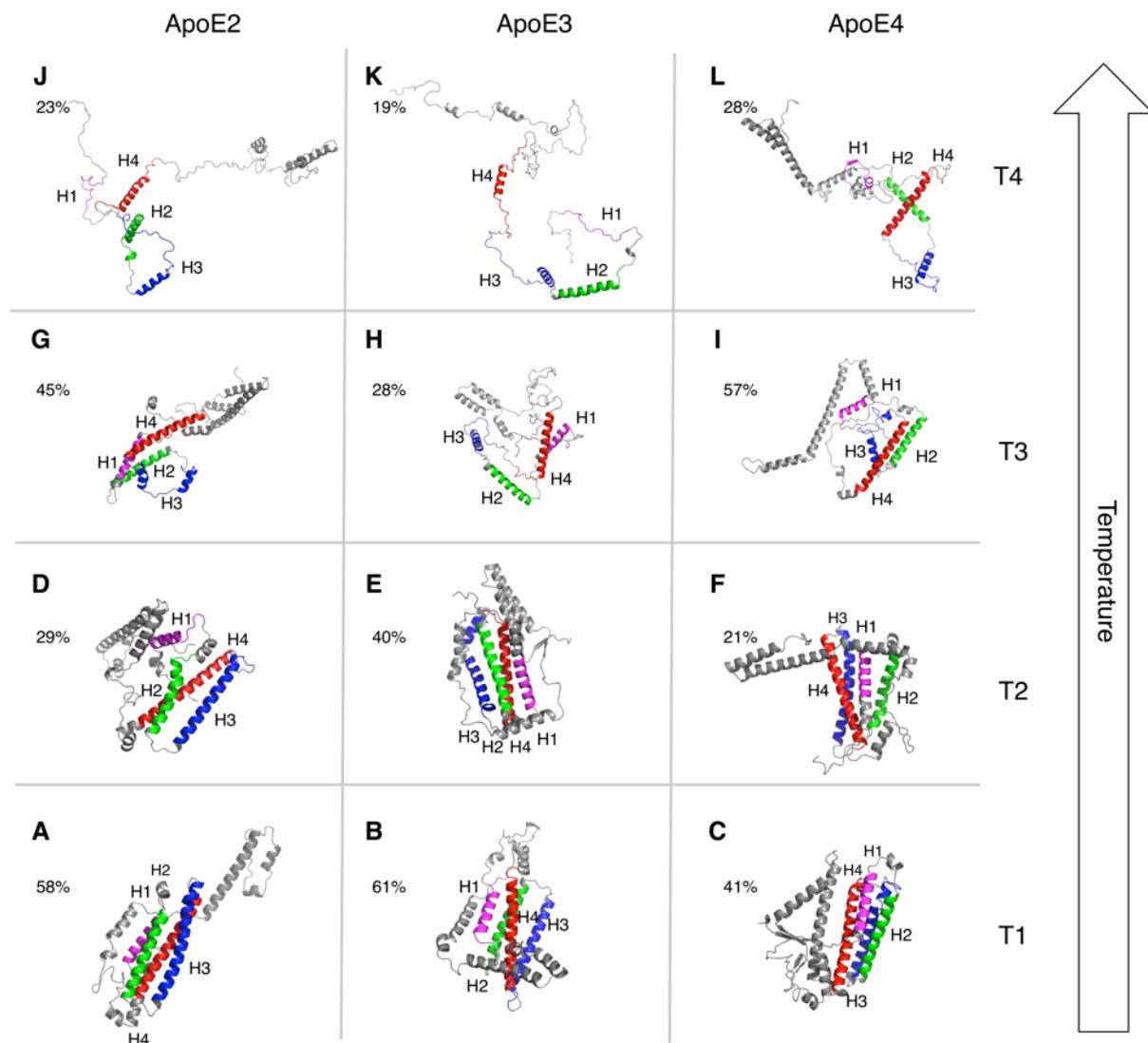
*To whom correspondence should be addressed: Nikolay V. Dokholyan, email: dokh@unc.edu

Running Title: ApoE4-specific misfolded intermediate



S1 Figure. Secondary structure analyses of ApoE isoforms.

(A) Cartoon representation of ApoE. The same structural features are applicable for all three ApoE isoforms: helix-1 (H1, residues 24 to 41), helix-2 (H2, residues 55 to 80), helix-3 (H3, residues 90 to 125), helix-4 (H4, residues 131 to 165), hinge region (residues 166 to 205) and C-terminal domain (residues 206 to 299) are represented in purple, green, blue, red, and grey, respectively. Single residue secondary structure analysis of ApoE2 (B), ApoE3 (C), and ApoE4 (D). The probability of secondary structure content (indicated in the plot as H, B, and L for alpha-helix, beta strand, and disordered respectively) at a specific residue is proportional to the relative height of the letter at that site. The same color code of (A) is used to indicate the helices position in (B-D). The average percentage of secondary structure content (alpha helix, beta strand, and disordered) as a function of temperatures T1 (~275 K for all three ApoE isoforms), T2 (~321 K, ~318 K, and ~309 K for ApoE2 (E), ApoE3 (F) and ApoE4 (G) respectively), T3 (~340 K, ~338 K, and ~328 K for ApoE2, ApoE3 and ApoE4 respectively) and T4 (~355 K, ~365 K, and ~342 K for ApoE2, ApoE3, and ApoE4, respectively) reveal that all of the three isoforms lose secondary structure without significant differences.

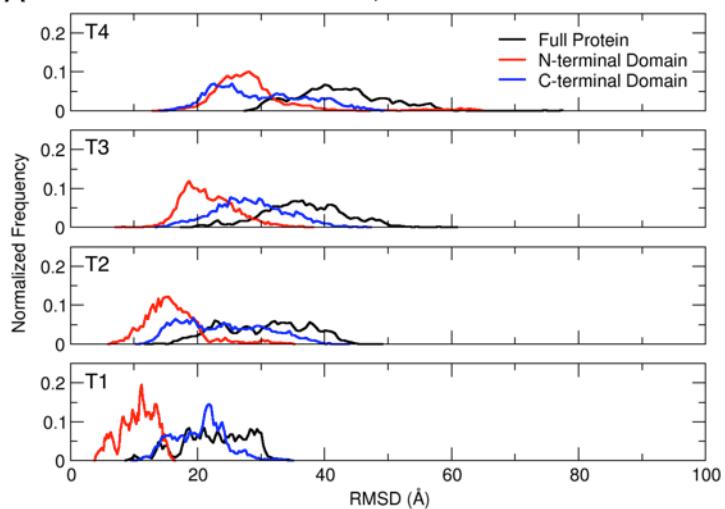


S2 Figure. Representative structures of ApoE isoforms.

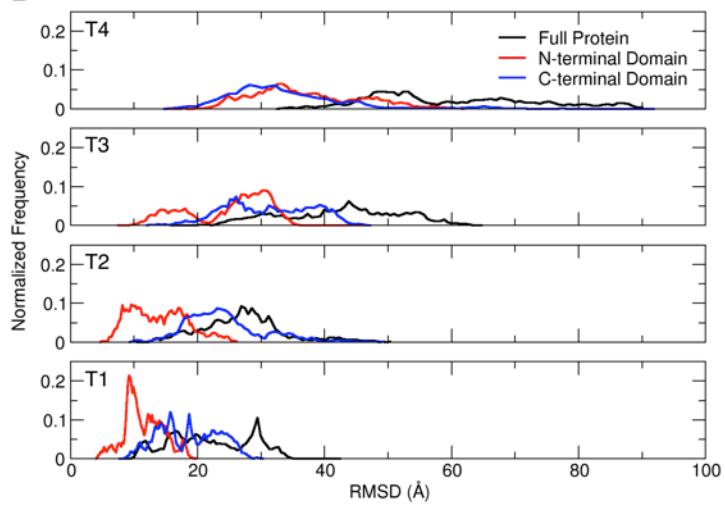
Representative structures (i.e., centroid of the most populated cluster) from clustering analysis of the ApoE isoforms' conformations extracted from the free energy basin (A-C) at T1 (~275 K for all three ApoE isoforms) and (D-F) at T2 (~321 K, ~318 K, and ~309 K for ApoE2, ApoE3 and ApoE4 respectively). All three isoforms exhibit native-live conformations with compact N-terminal domains. (G-I) at T3 (~340 K, ~338 K, and ~328 K for ApoE2, ApoE3 and ApoE4 respectively), the intermediate states of each ApoE variant represent the dominant conformations in the free energy landscape (see Fig. 2 in the main text). (J-L) at T4 (~355 K, ~365 K, and ~342 K for ApoE2, ApoE3, and ApoE4, respectively), the tertiary contacts are lost with the complete unfolding of the proteins. The size of the most populated cluster is reported in each panel. For all structures, helix-1 (H1), helix-2 (H2), helix-3 (H3), and helix-4 (H4) are represented in purple, green, blue, and red, cartoon respectively. The rest of the protein is represented in grey cartoon.

A

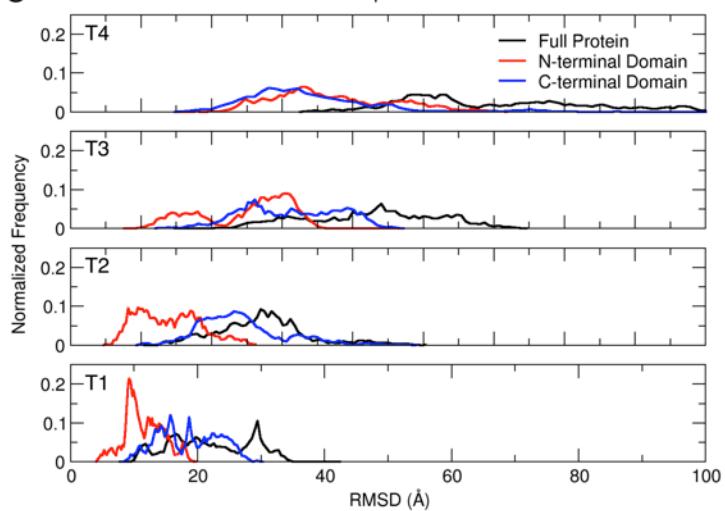
ApoE2

**B**

ApoE3

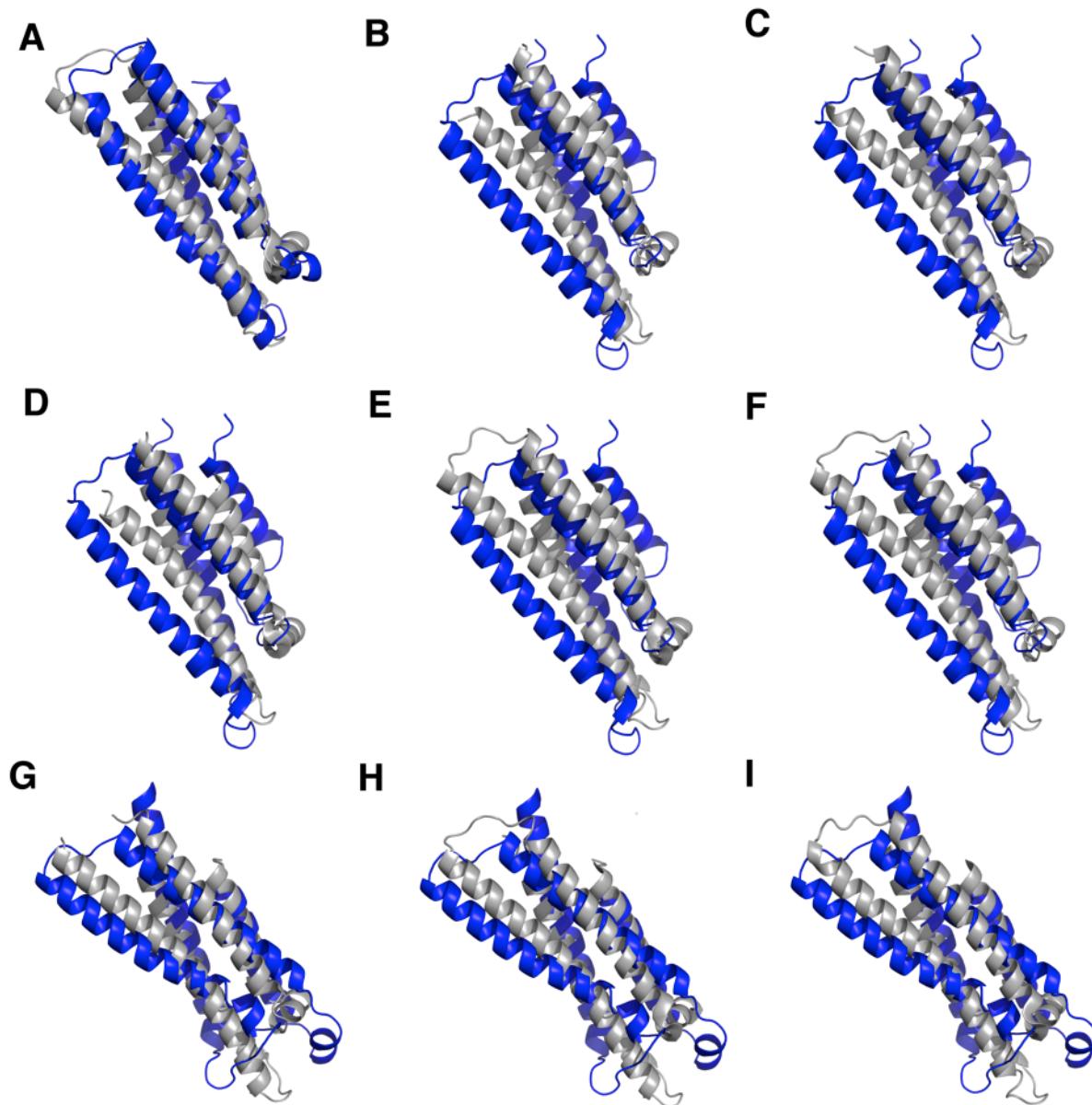
**C**

ApoE4



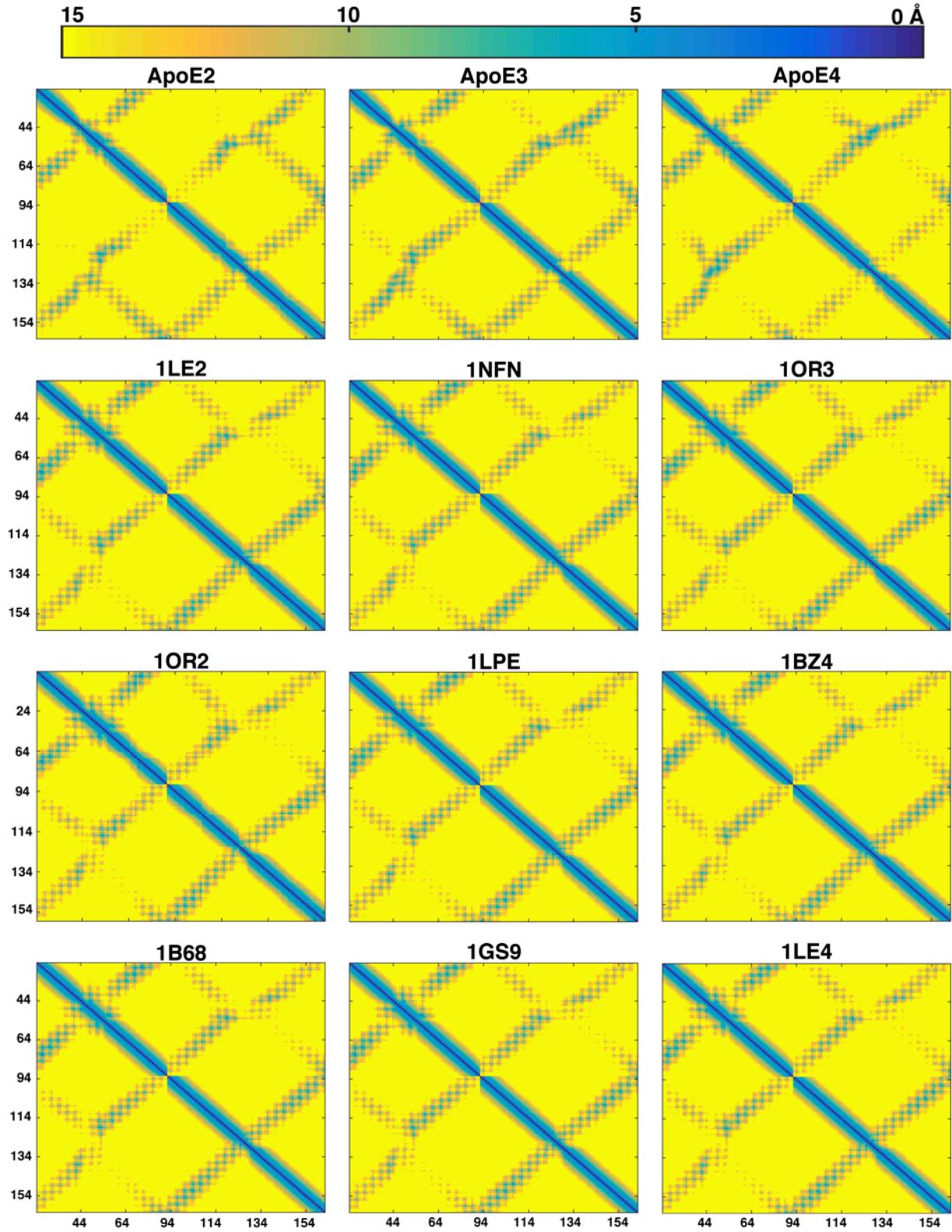
S3 Figure. RMSD distributions of ApoE isoforms' domains.

Distributions of the RMSD of the C α atoms for ApoE2 (A), ApoE3 (B), and ApoE4 (C) isoform at T1 (~275 K for all three ApoE isoforms), T2 (~321 K, ~318 K, and ~309 K for ApoE2, ApoE3 and ApoE4 respectively), T3 (~340 K, ~338 K, and ~328 K for ApoE2, ApoE3 and ApoE4 respectively) and T4 (~355 K, ~365 K, and ~342 K for ApoE2, ApoE3, and ApoE4, respectively) for the full protein (residues 1 to 299 in black), the N-terminal domain (residues 1 to 165 in red), and the C-terminal domain including the hinge region (residues 166 to 299 in blue). For temperatures ranging from T1 to T3, the C-terminal domain exhibits larger RMSD values than the N-terminal domain. At temperature T4 the RMSD values for N-terminal domain are much larger and comparable with the C-terminal domain indicating the overall unfolding of the protein. Temperatures from T1 to T4 are reported as insets within each plot. For all histograms, the width of the bins corresponds to 1 Å.



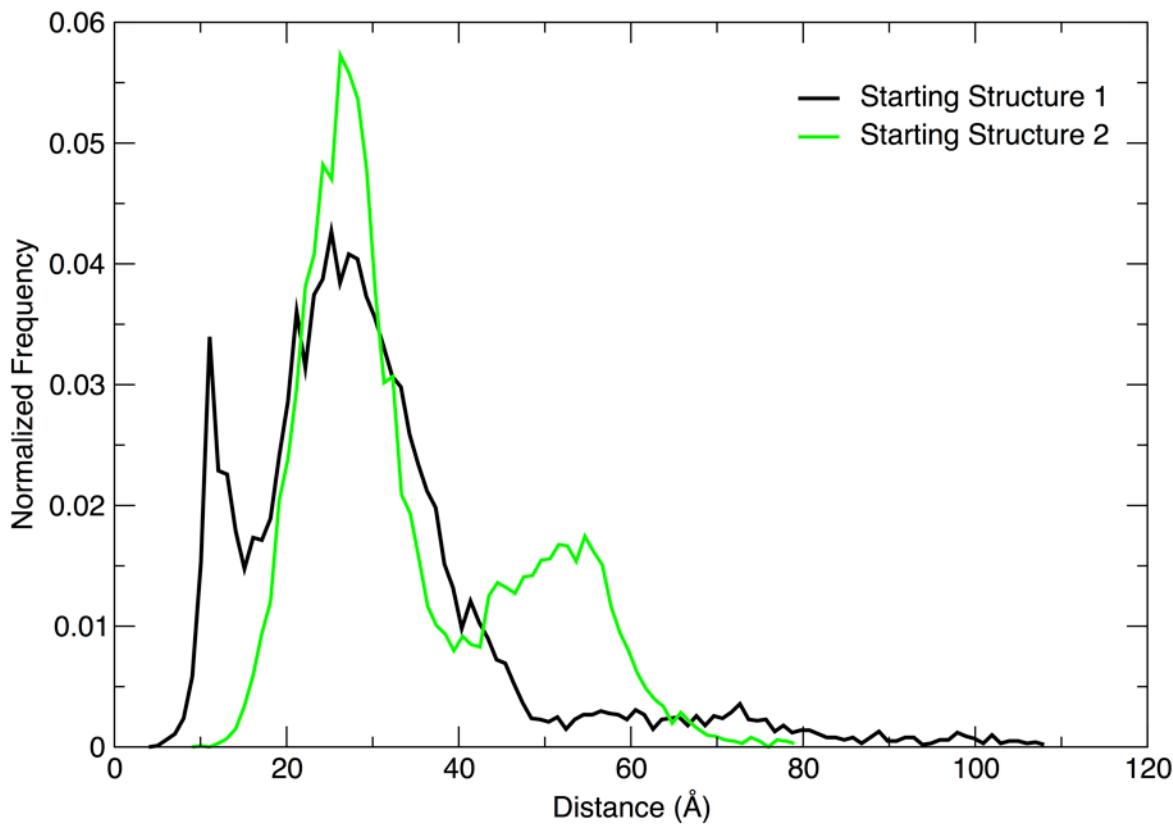
S4 Figure. N-terminal comparison of REX/DMD centroids with crystal structures.

Each panel represents an alignment between the centroids isolated through clustering analysis at 275 K from REX/DMD simulations with different previously solved crystal structures. (A) ApoE2 (1LE2) (B) ApoE3 (1NPN) (C) ApoE3 (1OR3) (D) ApoE3 (1OR2) (E) ApoE3 (1LPE) (F) ApoE3 (1BZ4) (G) ApoE4 (1B68) (H) ApoE4 (1GS9) (I) ApoE4 (1LE4). The corresponding RMSD for each alignment is presented in S3 Table. Structures in grey cartoon are crystallographic and structures in blue cartoon are the N-terminal domains as shown in Fig S2A-C.



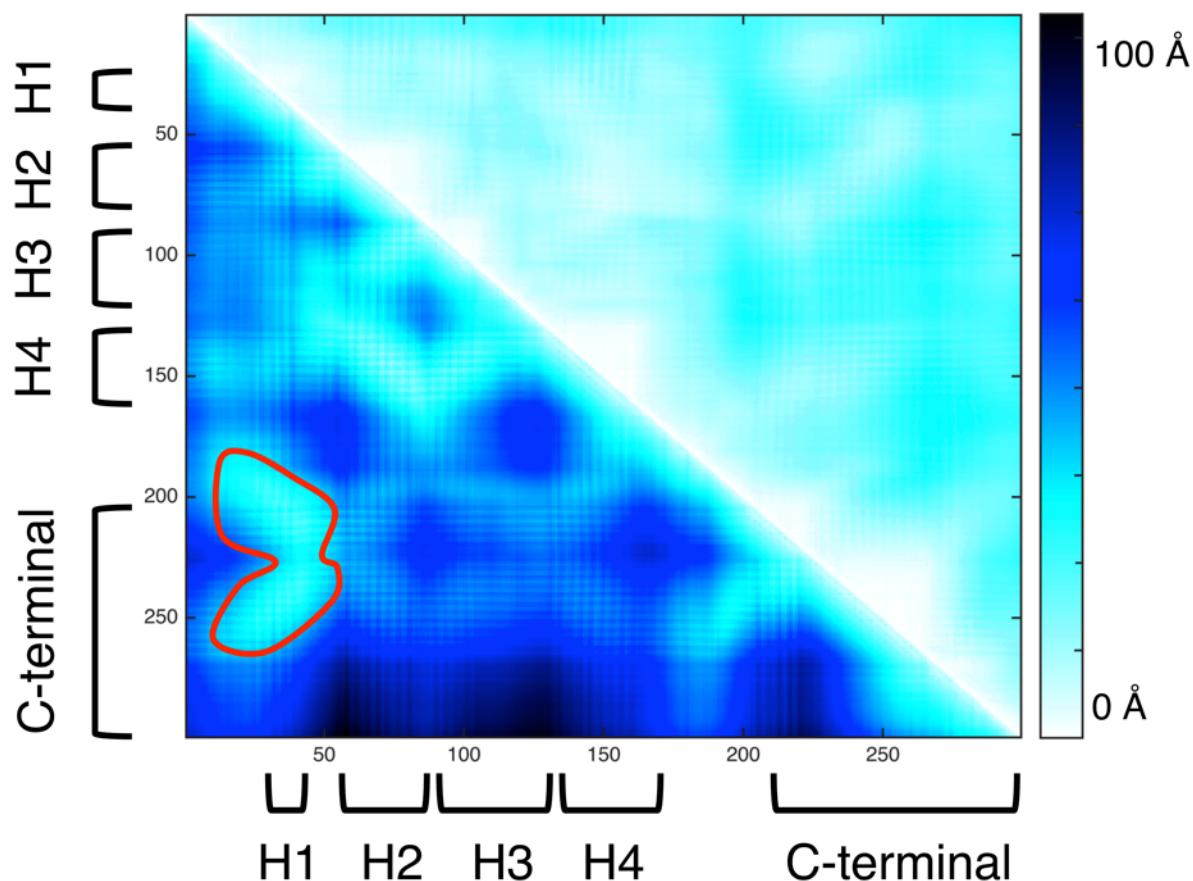
S5 Figure. ApoE native conformation contact maps.

Contact maps were calculated for the ApoE native-like conformations of each isoform isolated from clustering analysis at 275 K. A comparison was made with X-ray crystal structures of ApoE isoforms seen in S2 Table and S4 Fig. The distances were measured in angstroms between C α atoms of residues 24-82 and residues 93-162, the residues common amongst all structures. The numbers along the x- and y-axes represent the residue numbers. The color bar represents the distance between the centroid computed over the residues' side chains in Å.

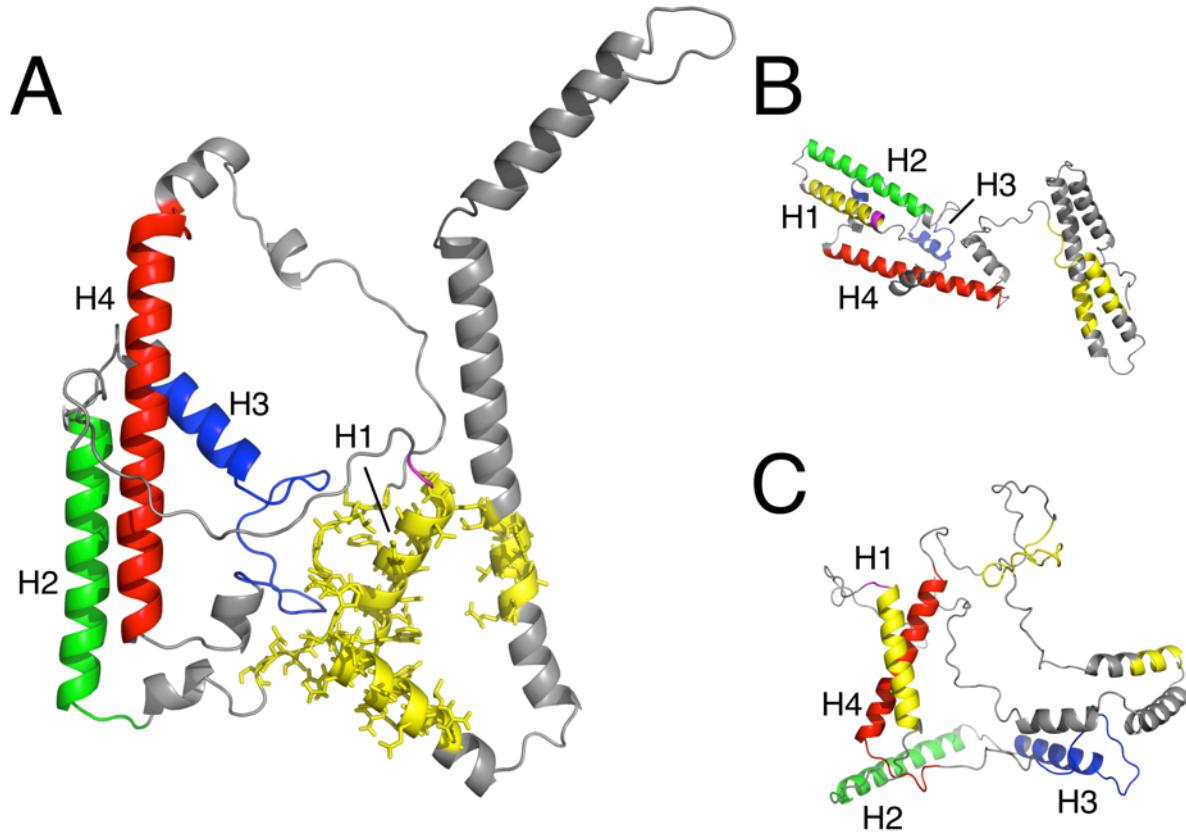


S6 Figure. End to end distances in ApoE4 misfolded intermediate state.

The histogram illustrates the distribution of distances measured between the C α atoms in terminal residues K1 and H299 in ApoE4 from serial DMD simulations. Two single temperature DMD simulations were performed at 309 K for 1 million time steps (~50 ns) using our ApoE4 misfolded intermediate state as a starting structure. The two different starting structures correspond to the centroid of the most populated cluster (S2I Fig.) and the lowest energy structure from the same cluster. The histogram reveals that the two termini in the ApoE4 misfolded intermediate conformation are below 60 Å for the majority of the simulations. This is a rough estimate for potential to generate FRET signals. The width of the histogram bins corresponds to 1 Å.

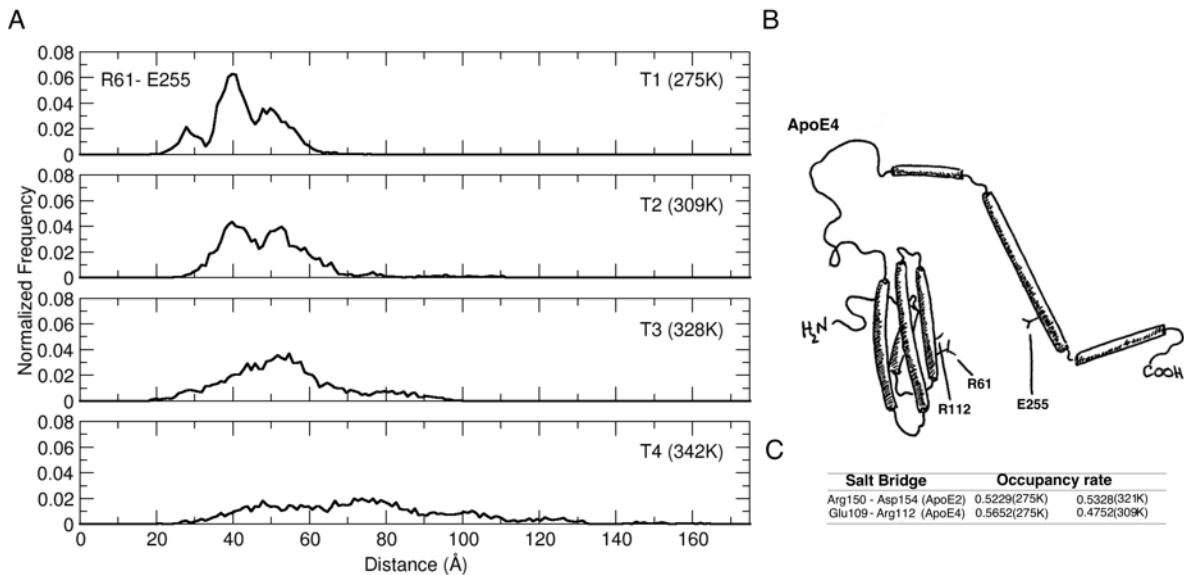


S7 Figure. Pair-wise inter-residue distances in the ApoE4 misfolded intermediate state. Inter-residue distance analysis of the most populated cluster of ApoE4 conformations extracted from the free energy basin at T3 (~342 K, see Fig. 2I) reveals a unique series of contacts (outline in red) and reported in S8A Fig. The upper and lower triangular matrices represent respectively the average and the standard deviation of the pair-wise inter-residue distance in Å. The color bar represents the distance between the centroid computed over the residues' side chains in Å.



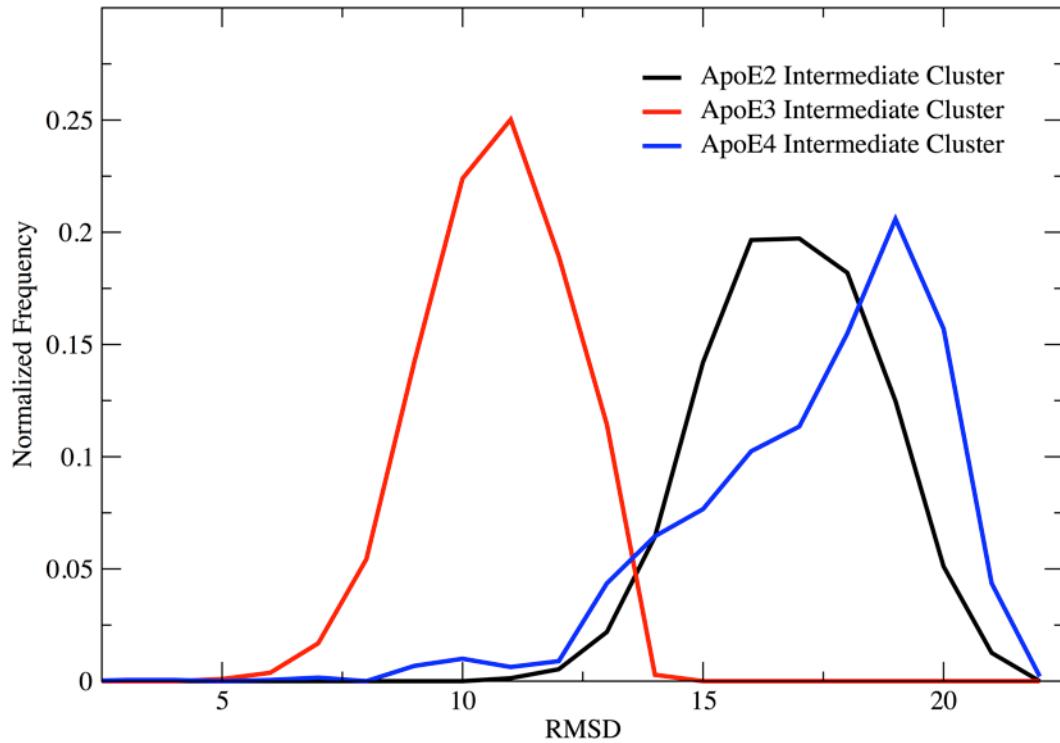
S8 Figure. Domain-domain interaction as a unique feature of the ApoE4 misfolded intermediate state.

(A) Inter-residue contacts (residues from 26 to 44, from 196 to 215, and from 235 to 243) identified in the ApoE4 misfolded intermediate state at temperature T3 (~342 K). The relative positions of the same residues are reported for ApoE2 (B) and ApoE3 (C) at temperature T3 (~340 K and ~338 K, respectively). For all structures, helix-1 (H1), helix-2 (H2), helix-3 (H3), and helix-4 (H4) are represented in purple, green, blue, and red, cartoon respectively. Residues in contacts are reported as yellow cartoon (side chains represented as sticks only for ApoE4), while the rest of the protein is represented in grey cartoon.



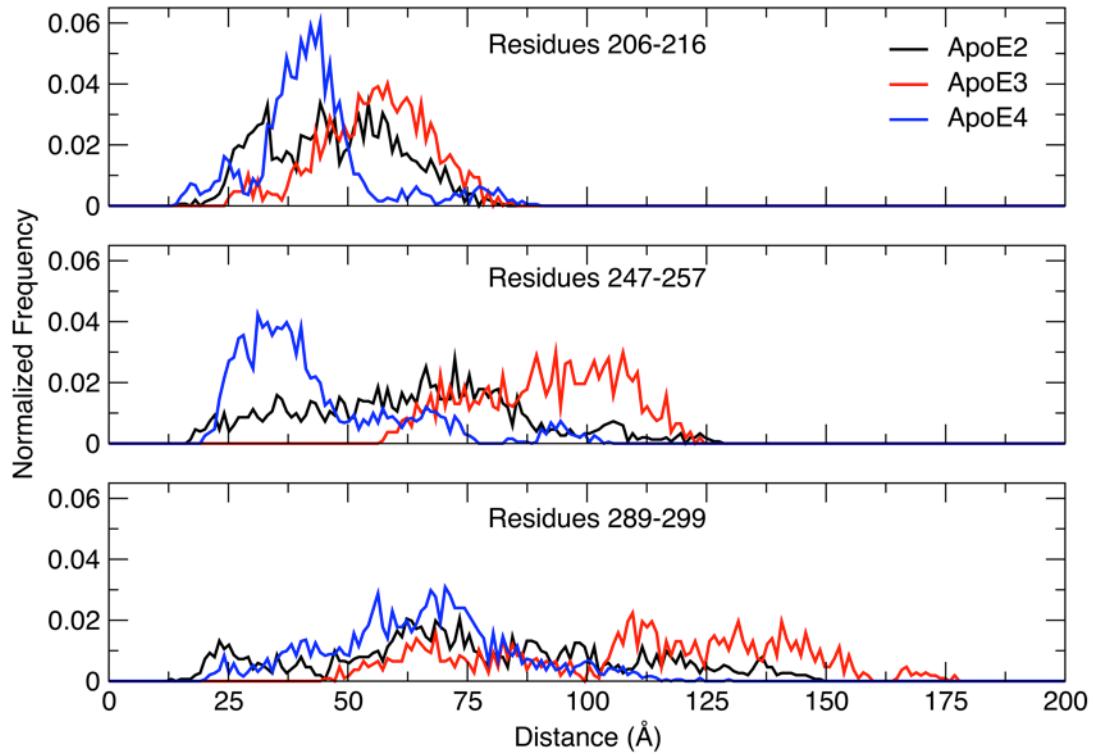
S9 Figure. Putative salt bridge distance in ApoE4.

(A) The distribution of distances between residues R61 and E255 as explored in ApoE4 isoform REX/DMD simulations is consistently greater than 20 Å regardless of the simulation temperature, and thus, not compatible with the existence of a salt bridge between the two residues. The width of the histogram bins corresponds to 1 Å. (B) A cartoon representation of the putative salt bridge interaction adapted from Mahley RW, Huang Y (2012) J Med Chem, 55: 8997-9008. doi:10.1021/jm3008618. (C) Occupancy rate of isoform-specific salt bridges.



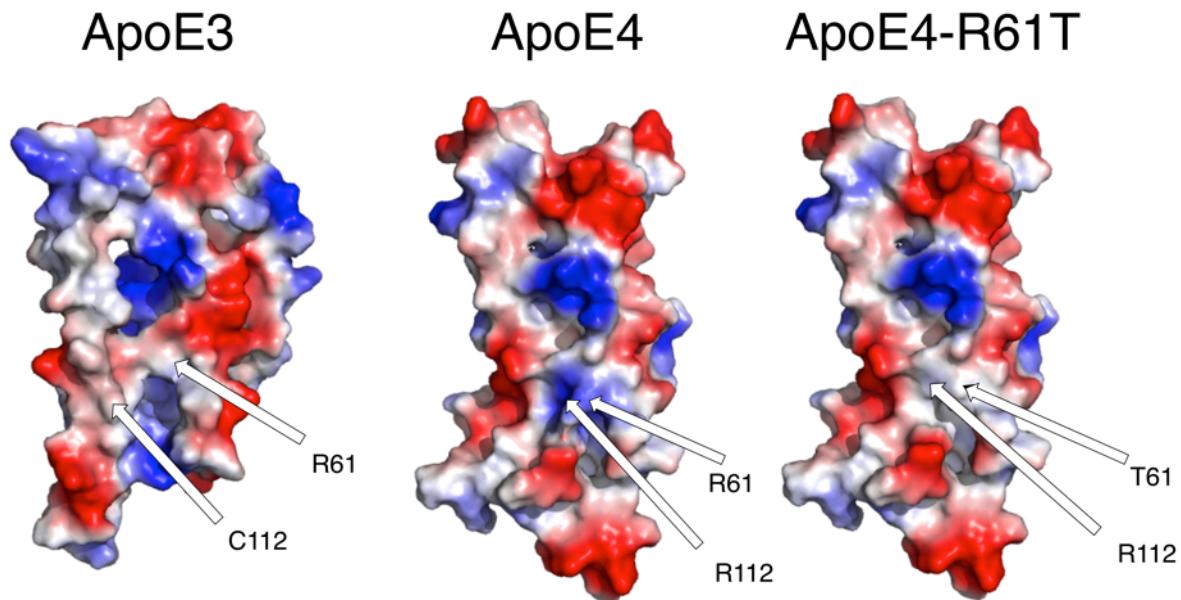
S10 Figure. RMSD distributions of ApoE Intermediate state clusters.

For the most populated cluster of each ApoE isoform at T3 (~340 K, ~338 K, and ~328 K for ApoE2, ApoE3 and ApoE4, respectively), the RMSD between each member of the cluster and the centroid was calculated. The width of histogram bins corresponds to 1 Å RMSD of the C α atoms in the N-terminal domain helices.



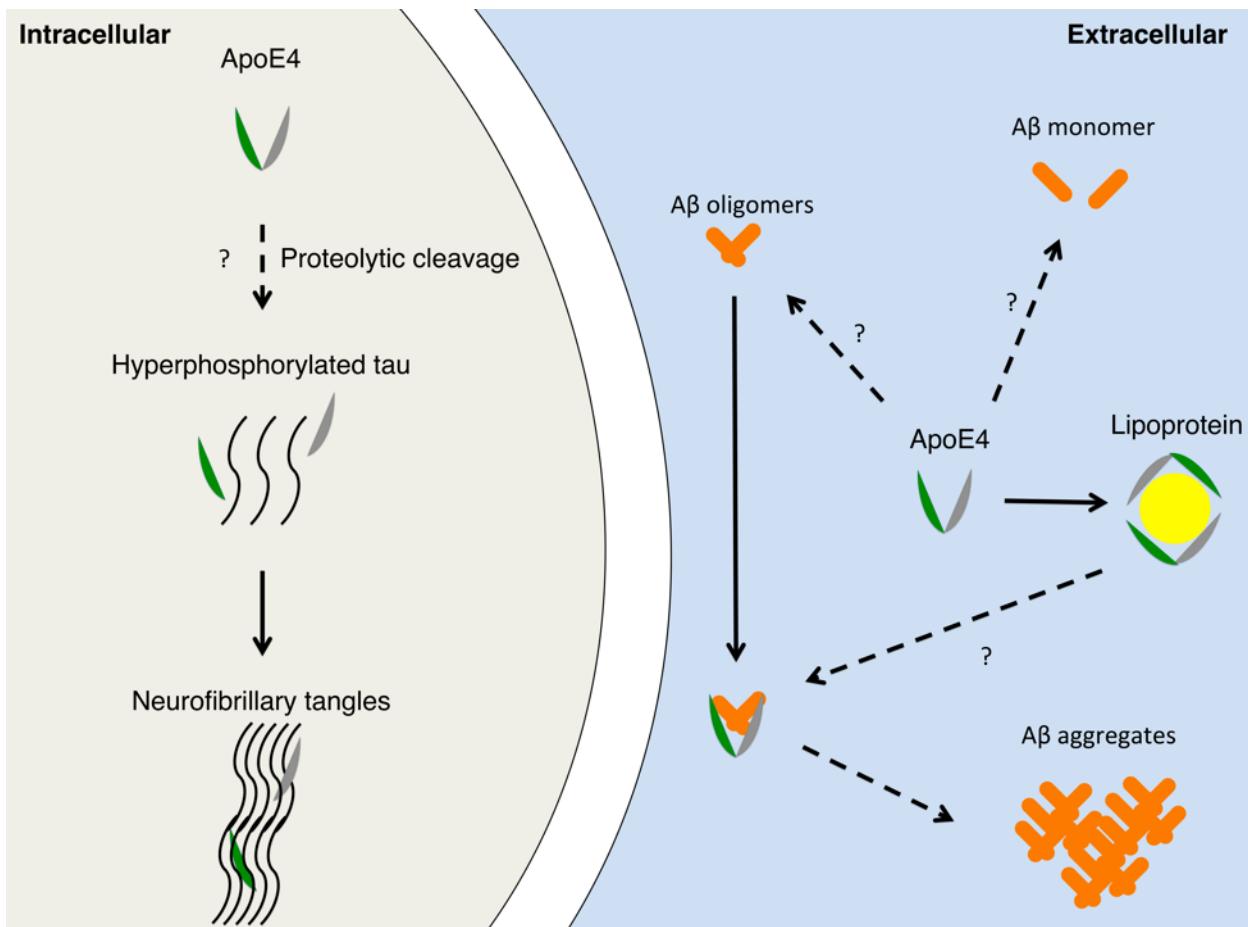
S11 Figure. C-terminal flexibility within ApoE intermediate states.

Distances between the center of mass of three different segments of the C-terminal domains (i.e., residues 206 to 216, residues 247 to 257, and residues 289 to 299) and the center of mass of the helix-4 in the N-terminal domain for each ApoE isoforms. Being the most stable helix in all ApoE simulations (see Fig. S1B-D), helix-4 has been chosen as reference position in the N-terminal domain. All calculations performed on the most populated clusters at temperature T3 (~340 K, ~338 K, and ~328 K for ApoE2, ApoE3 and ApoE4, respectively). For all histograms, the width of the bins corresponds to 1 Å.



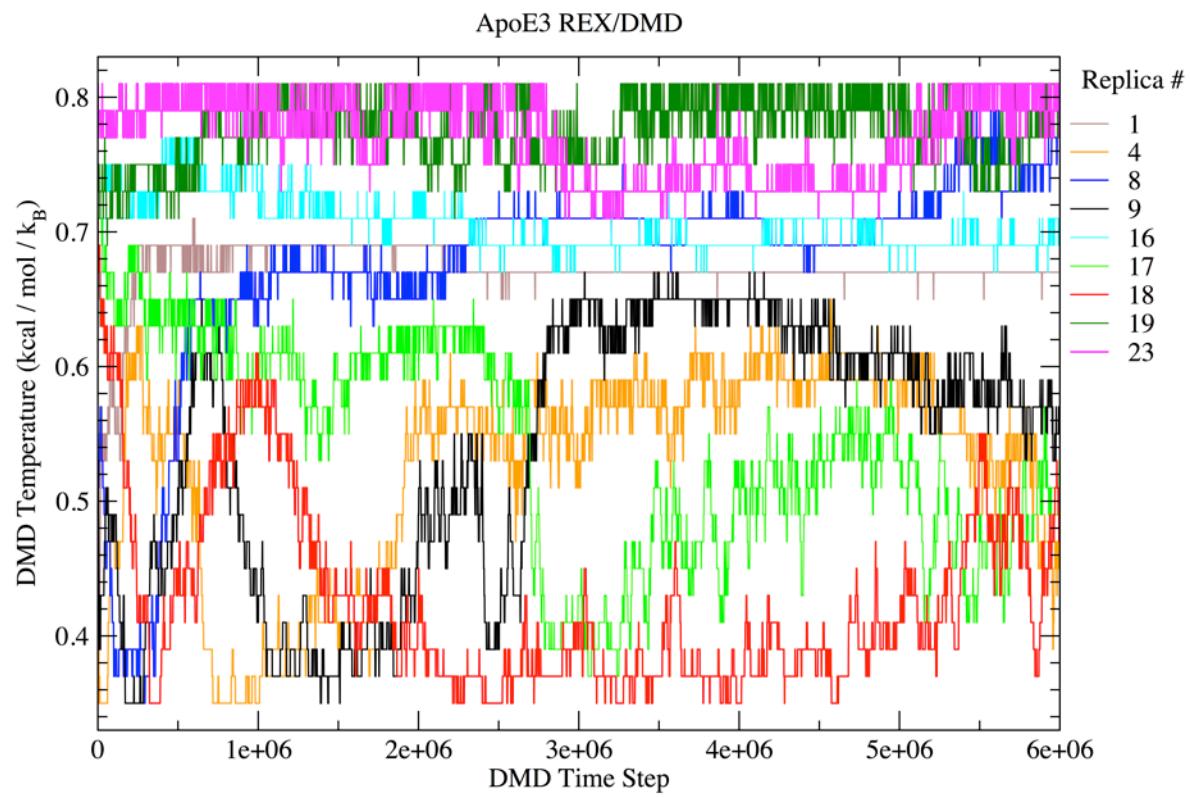
S12 Figure. Electrostatic surface potential in ApoE3 and ApoE4 isoforms.

Comparisons of the electrostatic surface potential of the N-terminal domain of ApoE3, ApoE4 and the ApoE4-R61T. Blue and red colors correspond to positively and negatively charged surfaces, respectively, and white color corresponds the neutral hydrophobic ones.



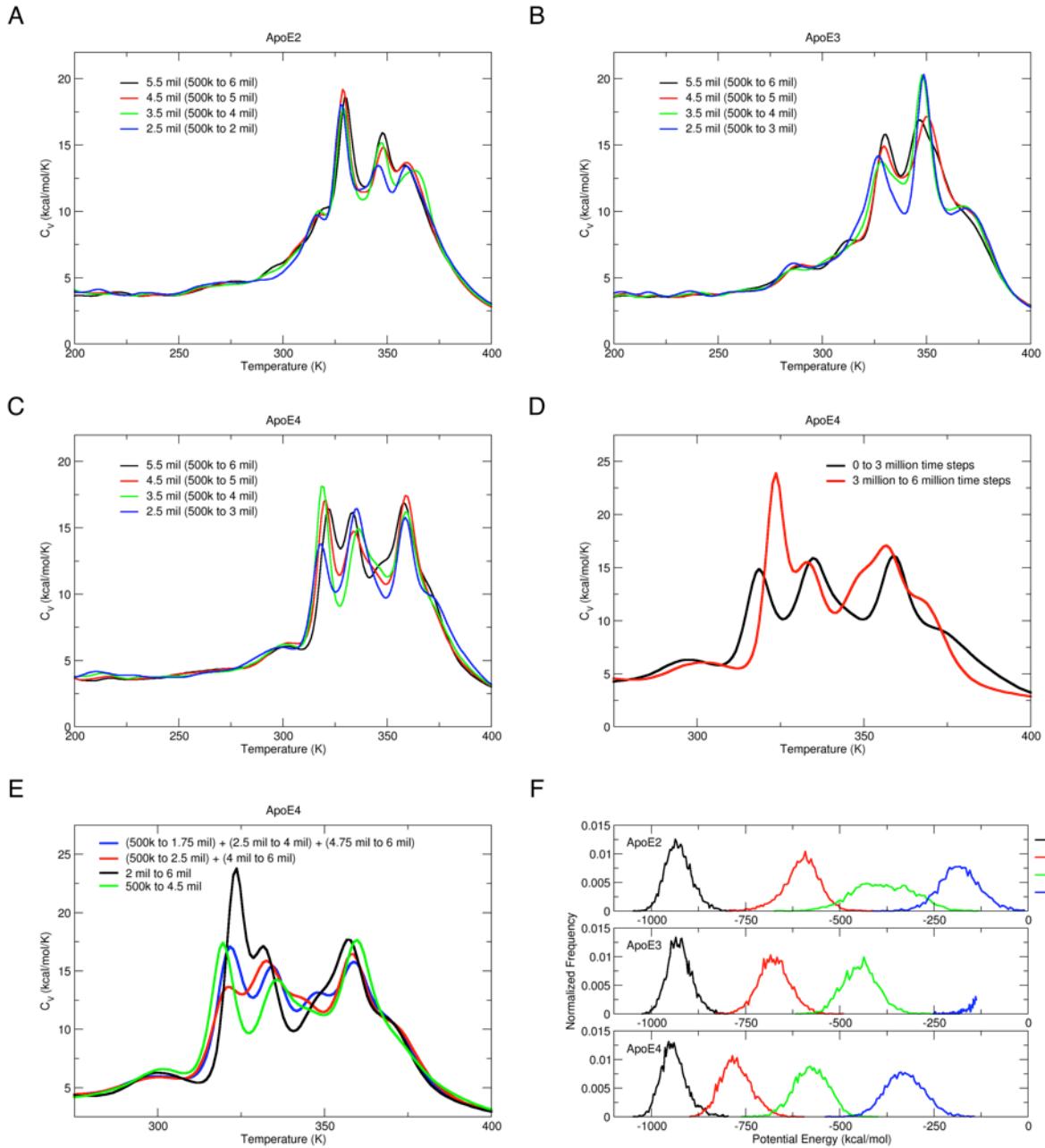
S13 Figure. Potential ApoE4-related pathological pathways.

We speculate that the ApoE4 misfolded intermediate state may affect the kinetics of A β peptides aggregation and clearance. Concurrently, it may favor the formation of intracellular neurofibrillary tangles. Further studies are required to elucidate the molecular mechanisms underlying the intracellular events leading to tau hyperphosphorylation and aggregation. Moreover, additional investigations are necessary to elucidate the molecular events at the basis of the interaction between ApoE4, lipids, and A β peptides.



S14 Figure. REX/DMD Replicas in temperature space.

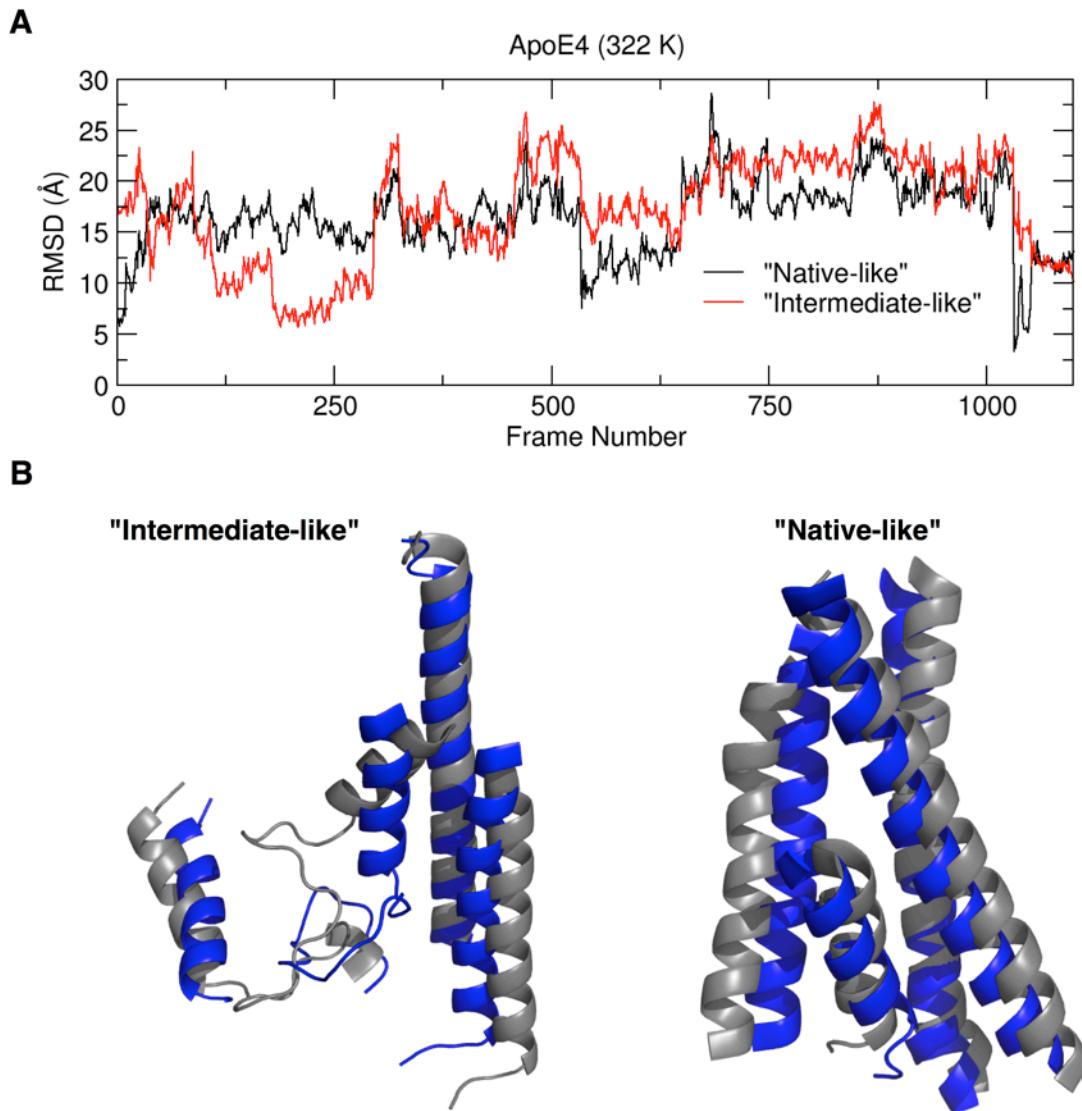
Representative replicas from our ApoE3 REX/DMD simulations are plots with their DMD temperature (kcal/mol/ k_B) as a function of time steps. There are replicas that travel through a wide range of temperature and those that are confined to a smaller spread in temperature space.



S15 Figure. Heat capacity convergence and energy distributions in REX/DMD simulations of ApoE isoforms.

The heat capacity (C_v) curves for ApoE2 (A), ApoE3 (B), and ApoE4 (C) isoforms computed using WHAM on REX/DMD trajectories in the range of 200 to 400 K including 5.5×10^6 , 4.5×10^6 , 3.5×10^6 and 2.5×10^6 time steps show the convergence of REX/DMD simulations. The heat capacity curves computed using WHAM on REX/DMD trajectories for ApoE4 in the range of 275K to 400K (D) using two independent and equal size windows from the same simulation. The peak positions in the two curves are slightly shifted revealing that 3×10^6 time steps is not sufficient to reach convergence. (E) The C_v curves computed using different windows corresponding to 4×10^6 time steps from the ApoE4 REX/DMD simulation. Segments of 4×10^6

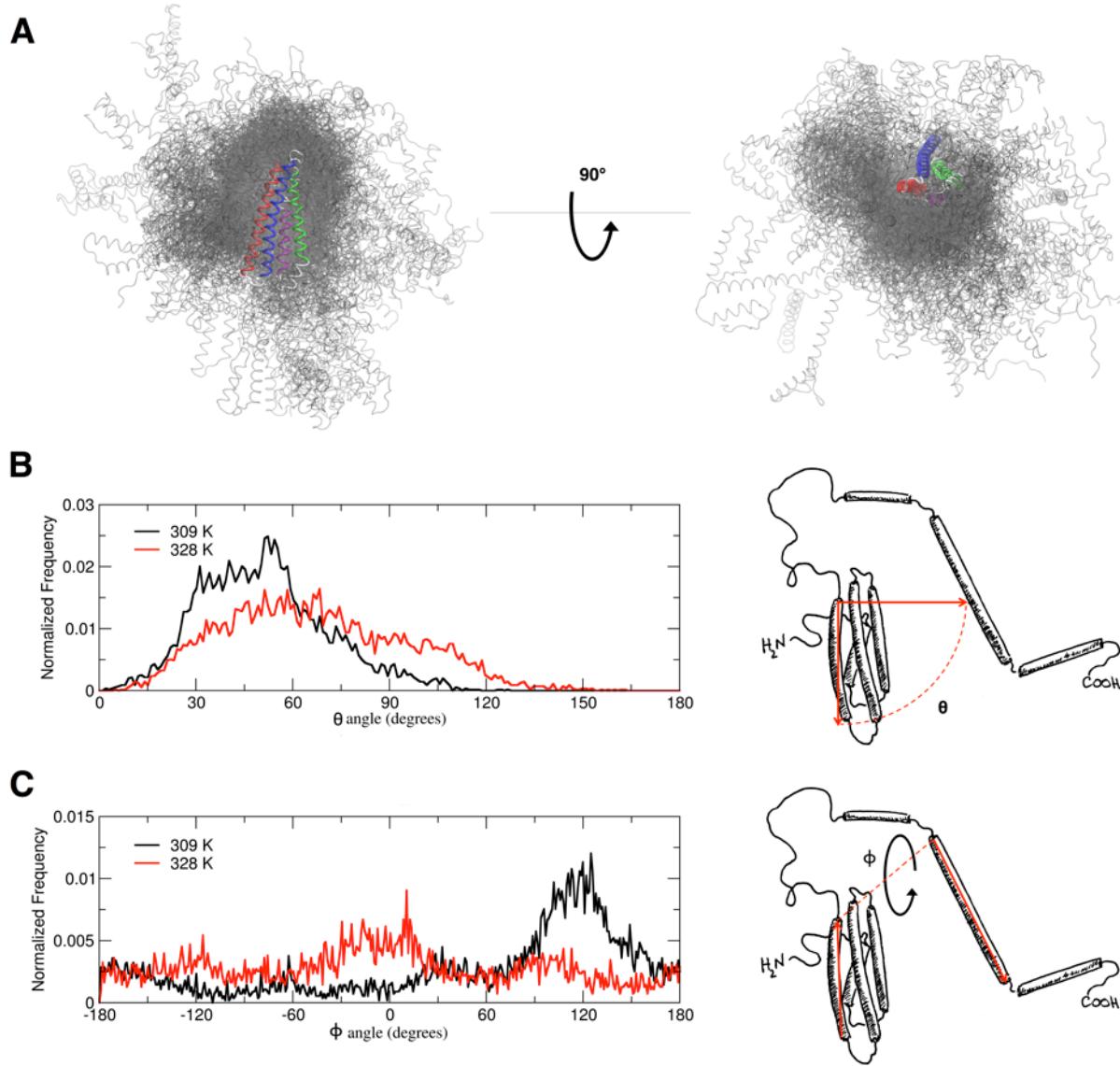
time steps in WHAM calculations allow for more consistent peak locations. (F) The three ApoE isoforms exhibit Gaussian distributions of potential energy supporting the treatment of REX/DMD simulations as partition functions at T1 (~275 K for all three ApoE isoforms), T2 (~321 K, ~318 K, and ~309 K for ApoE2, ApoE3 and ApoE4 respectively), T3 (~340 K, ~338 K, and ~328 K for ApoE2, ApoE3 and ApoE4 respectively) and T4 (~355 K, ~365 K, and ~342 K for ApoE2, ApoE3, and ApoE4, respectively). The width of the histogram bins corresponds to 1 kcal/mol.



S16 Figure. “Native-like” and “intermediate-like” N-terminal states of ApoE4.

(A) The RMSD of the C α atoms in the N-terminal domain helices was calculated between the trajectory of conformations at 322 K and the centroids found from clustering analysis at 309 K (T2) for the “native-like” state and at 328 K (T3) for the misfolded “intermediate-like” state. The RMSD values reveal that at the transition peak, ApoE4 visits both the “native-like” and misfolded “intermediate-like” N-terminal domain conformations (RMSD < 5 Å). (B) The conformations found at the transition state peak with the lowest RMSD values from (A) shown

in blue are aligned with their corresponding centroid structures colored in gray. The “native-like” conformation has an RMSD value of 3.28 Å with its respective centroid while the “intermediate-like” conformation has an RMSD value of 5.68 Å suggesting that the same conformation is found in the free energy basin at 328 K. The RMSD for both (A) and (B) is calculated using the N-terminal domain helices because the flexibility of the C-terminal domain adds considerable noise as discussed in the Methods section and as shown in S3 Fig.



S17 Figure. C-terminal conformations in ApoE4.

(A) The superposition of ApoE4 C-terminal conformations from REX/DMD simulations at 309 K with an aligned N-terminal domain reveals that the C-terminal domain explores a variety of positions relative to the N-terminal domain. The C-terminal domain assumes conformations next to the helix-1/helix-2 side of the N-terminal domain as suggested previously in literature as well as in conformations next to the helix-1/helix-4 and helix-4/helix-3 sides. The superposition of every 25th frame of C-terminal conformations from the REX/DMD trajectory is represented in grey. A representative conformation of the N-terminal domain alignment is shown with helix-1

(H1), helix-2 (H2), helix-3 (H3), and helix-4 (H4) in purple, green, blue, and red, cartoon, respectively. (B) The angle between the N-terminal domain and C-terminal domain shows the relative closeness between the two domains. An angle of zero degrees represents the N-terminal and C-terminal next to each other, while an angle of 180 degrees represents conformations with the C-terminal away from the sides of the N-terminal domain. The angle θ is measured using residues L148 and G165 in the most stable N-terminal helix, helix-4, and L252 representing the center of the most stable C-terminal domain helix. Note that the magnitude of the angle does not always correspond to a similar magnitude in distances between the two domains. (C) The dihedral angle φ between the N-terminal domain and the C-terminal domain shows the relative orientation between the two domains. An angle of zero degrees represents an anti-parallel orientation between the two domain helices while an angle of -180 or 180 degrees represents a parallel orientation. N-terminus to C-terminus is used for directionality. The angle φ is measured between vectors defined by the center of mass of residues E131 and G165 in the most stable N-terminal helix, helix-4, and residues E238 and F265 in the most stable C-terminal domain helix. The width of the histogram bins corresponds to 1 degree.

Supplementary Tables

S1 Table. Human ApoE Sequence.

KVEQAVETEPEPELRRQQTEWQSGQRWELALGRFWDYLWVQTLSEQVQEELLSSQVT
QELRALMDETMKELKAYKSELEEQLTPVAEETRARLSKELQAAQARLGADMEDV**C**GR
LVQYRGEVQAMLGQSTEELRVRLASHLRKLRKRLRDADDLQK**R**LAVYQAGAREGAE
RGLSAIRERLGPLVEQGRVRAATVGSLAGQPLQERAQAWGERLARMEEMGSRTRDRL
DEVKEQVAEVRAKLEEQAQQIRLQAEAFQARLKSWFEPLVEDMQRQWAGLVEKVQAA
VGTSAAPVPSDNH

The amino acid sequences of the three ApoE isoforms are identical with the exception of C112R and R158C mutations found in ApoE4 And ApoE2 variants, respectively.

S2 Table. Structures found in our simulations have N-terminal helix conformations similar to solved crystal structures.

Ref. Protein (PDB ID)	RMSD
ApoE2 (1LE2) [1]	3.188 Å
ApoE3 (1NFN) [2]	2.361 Å
ApoE3 (1OR3) [3]	2.265 Å
ApoE3 (1OR2) [3]	2.096 Å
ApoE3 (1LPF) [4]	2.754 Å
ApoE3 (1BZ4) [3]	2.714 Å
ApoE4 (1B68) [5]	2.924 Å
ApoE4 (1GS9) [6]	2.809 Å

ApoE4 (1LE4) [1]	3.247 Å
------------------	---------

Root mean square deviation (RMSD) of the centroid the most populated clusters of ApoE isoforms' conformations from local minima on the PMF-derived free energy landscapes at 275K. RMSD is computed over the C α of the N-terminal domain helices. The C-terminal domain is not available in any of the ApoE crystal structures.

S3 Table. RMSD between native-like ApoE conformations at 275 K.

ApoE Isoform	E2	E3	E4
E2	0	5.84	6.47
E3	5.84	0	4.52
E4	6.47	4.52	0

RMSD values in Å for the C α atoms of the N-terminal helices for the centroids from clustering analysis at 275 K from REX/DMD simulations (Fig S2A-C).

S4 Table. Population of ApoE intermediates at physiological temperatures.

ApoE Isoform	Population of Intermediate from 300 to 310 K
E2	70.66 %
E3	0.05 %
E4	11.12 %

The populations of the ApoE intermediate states in the temperature range of 300 to 310 K were identified using RMSD against the centroid of the most populated cluster for intermediate state of each ApoE isoform (Fig 3BCD). Similarity was determined by an RMSD of the N-terminal domain helices with the centroids that is less than 15 Å.

S5 Table. EPR distances in REX/DMD simulations.

Frequency of distances at or below experimental measurements

Residue Pairs	76-241	76-263	76-264	77-241	77-263	77-264
Distance Cutoff	(14.1 +/- 5 Å)	(21.2 +/- 5 Å)	(22 +/- 3 Å)	(18.7 +/- 5 Å)	(22 +/- 3 Å)	(22 +/- 3 Å)
ApoE3 (318 K)	0.78%	14.64%	18.4%	3.4%	18.84%	24.46%
ApoE3 (338 K)	0.04%	1.09%	1.05%	0.16%	1.14%	1.07%

ApoE4 (309 K)	0.11%	19.77%	17.83%	1.17%	20.33%	21.67%
ApoE4 (328 K)	1.51%	27.08%	29.1%	3.14%	26.94%	30.1%
ApoE4 (328 K) Free Energy Basin	1.99%	22.55%	23.5%	2.92%	22.55%	24.16%

Six residue pairs within specific distance ranges were identified by Hatters et al. using electron paramagnetic resonance for 60% of ApoE4 tetramer conformations. Comparing these distance measurements with conformations extracted from our ApoE3 and ApoE4 monomer REX/DMD simulations reveals that we explore conformations that satisfy all six residue pairs with different frequencies. The low occupancy of experimental distance for residue pairs such as 76-241 and 77-241 may be explained by conformational differences between ApoE4 monomers and tetramers. In addition, at 338 K for ApoE3 and 328 K for ApoE4, corresponding to the temperature for their respective intermediate states, the distances are satisfied with a higher frequency in ApoE4 than in ApoE3 suggesting that these measurements may indeed be indicative of an ApoE4 intermediate conformation.

S6 Table. Hydrophobic Solvent Exposed Surface Area of ApoE intermediate states.

ApoE Isoform	Average Hydrophobic Surface Area (1000 Å²)
E2	10.13 +/- 0.74
E3	10.47 +/- 0.52
E4	8.63 +/- 0.72

The solvent exposed surface area was calculated for hydrophobic residues of the ApoE intermediate state clusters using a sphere with a 1.4 Å radius with Visual Molecular Dynamics software tools [7].

S7 Table. ApoE structural and thermodynamic insights. Summary of structural features and thermodynamic properties revealed by REX/DMD simulations for the three ApoE isoforms. ApoE4 exhibits multiple isoform-specific features related to the presence of a misfolded intermediate state.

Features	ApoE2	ApoE3	ApoE4	Reference
Mutation	C112/C158	C112/R158	R112/R158	Table 1
Hydrophobic core disruption*	330 K	331 K	322 K	Fig 1
Helix content at 275 K	75.2 ± 3.7 %	72.7 ± 2.6 %	74.1 ± 3.8 %	S1 Fig
Beta strand content at 275 K	2.2 ± 1.9 %	3.6 ± 2.1 %	2.9 ± 2.8 %	S1 Fig
Lowest RMSD with crystal structures	3.2 Å (PDB ID: 1LE2)	2.1 Å (PDB ID: 1OR2)	2.8 Å (PDB ID: 1GS9)	S2 Table

C-terminal domain range of motion at T3 [¶]	$28.2 \pm 5.7 \text{ \AA}$	$24.1 \pm 6.3 \text{ \AA}$	$30.9 \pm 7.0 \text{ \AA}$	S3 Fig
Distance between C-terminal segment (residues 206-216) and helix-4 at T3 [¶]	$46.6 \pm 13.5 \text{ \AA}$	$55.4 \pm 11.3 \text{ \AA}$	$42.2 \pm 12.7 \text{ \AA}$	S11 Fig
Distance between C-terminal segment (residues 247-257) and helix-4 at T3 [¶]	$62.9 \pm 23.4 \text{ \AA}$	$91.0 \pm 15.5 \text{ \AA}$	$42.9 \pm 17.8 \text{ \AA}$	S11 Fig
Hydrophobic surface area at T3 ^{¶\$}	10.1 ± 0.7	10.5 ± 0.5	8.6 ± 0.7	S6 Table
Salt bridge distance (R61 – E255) at T3 [¶]	N/A	N/A	$52.9 \pm 15.0 \text{ \AA}$	S9 Fig
Average RMSD distribution of protein conformations in the most populated cluster at T3 [¶]	$16.3 \pm 1.8 \text{ \AA}$	$10.2 \pm 1.5 \text{ \AA}$	$16.8 \pm 2.6 \text{ \AA}$	S10 Fig
Population of intermediate state at 300 to 310 K	70.7 %	0.1 %	11.1 %	S4 Table

*First peak in the specific heat plot (Figure 1). [¶]Analysis performed at temperature T3 (~340 K, ~338 K, and ~328 K for ApoE2, ApoE3 and ApoE4 respectively) correspond to the respective local minima of heat capacity curves for each ApoE isoform. ^{\$}Values expressed in the units of 1000 \AA^2 .

References

1. Wilson C, Mau T, Weisgraber KH, Wardell MR, et al. (1994) Salt bridge relay triggers defective LDL receptor binding by a mutant apolipoprotein. *Structure*. 2: 713–8.
2. Dong LM, Parkin S, Trakhanov SD, Rupp B, Simmons T, et al. (1996) Novel mechanism for defective receptor binding of apolipoprotein E2 in type III hyperlipoproteinemia. *Nat Struct Biol* 3: 718–722.
3. Segelke BW, Forstner M, Knapp M, Trakhanov SD, Parkin S, et al. (2000) Conformational flexibility in the apolipoprotein E amino-terminal domain structure determined from three new crystal forms: implications for lipid binding. *Protein Sci* 9: 886–897. doi:10.1110/ps.9.5.886.
4. Wilson C, Wardell MR, Weisgraber KH, Mahley RW, Agard DA (1991) Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. *Science* 252: 1817–1822.
5. Dong J, Peters-Libeu CA, Weisgraber KH, Segelke BW, Rupp B, et al. (2001) Interaction of the N-terminal domain of apolipoprotein E4 with heparin. *Biochemistry* 40: 2826–2834. doi:10.1021/bi002417n.
6. Verderame JR, Kantardjieff K, Segelke BW, Weisgraber KH, Rupp B. (2003) Crystal Structure of the 22K Domain of Human Apolipoprotein E4. To be published. doi:10.2210/pdb1gs9/pdb.
7. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14: 33–8– 27–8.