TCABSE-J Research Article
ISSN: 2583- 2557, *Volume 1, Issue 7, pp26-31. 2024.*

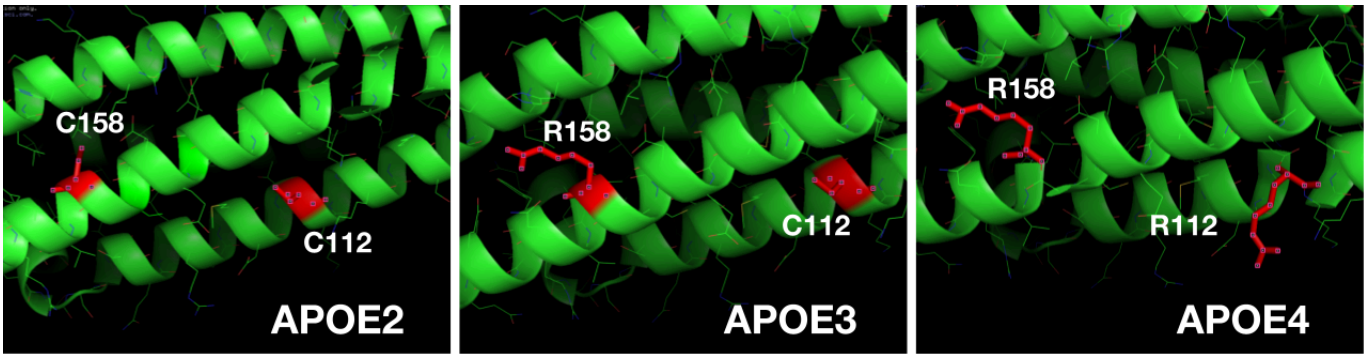
Comparative *in silico* analysis of the human apolipoprotein-E isoforms 2, 3 and 4 to elucidate their structural flexibility with implications in Alzheimer's disease.

Anupriya Yekula[1,2], Anitha Devandla[1,2], Sreenivasulu Bonala[1,2], Swarnalatha Gudapati[2] & Ravikiran S. Yedidi[1,3,*]

[1]Department of Intramural research core, The Center for Advanced-Applied Biological Sciences & Entrepreneurship (TCABS-E), Visakhapatnam 530003, A.P. India; [2]Department of Biomedical Sciences, Hindu College, Guntur 522002, A.P. India; [2]Department of Biotechnology, Andhra University, Visakhapatnam 530003, A.P. India. (*Correspondence to R.S.Y.: tcabse.india@gmail.com).

Keywords: Apolipoprotein, Alzheimer's disease, molecular modeling, molecular dynamics simulations, membrane protein.

The human apolipoprotein-E (ApoE) has four major isoforms of which the isoform4 has been implied to play a role in Alzheimer's disease. ApoE2 contains C112/C158, ApoE3 contains C112/R158 and ApoE4 contains R112/R158 in their N-terminal domains. In this study we evaluated the conformational flexibility of these residues in comparison with the overall root mean square deviations (RMSD) in the protein by performing molecular dynamics (MD) simulations. Based on our results, although the overall RMSDs of three isoforms were apparently similar, the RMSD of ApoE4 was slightly different. While ApoE2 and ApoE3 showed more steady equilibration of RMSDs from within 2 ns up to 5 ns, ApoE4 showed more fluctuations not only within the 5 ns but also up to 10 ns suggesting the overall structural flexibility is more. The root mean square fluctuations (RMSF) were different for all three isoforms especially around position 112. These results suggest that the overall structural flexibility of ApoE4 is higher than ApoE2 and ApoE3.



Citation: Yekula, A., Devandla, A., Bonala, S., Gudapati, S. and Yedidi, R.S. (2024). Comparative *in silico* analysis of the human apolipoprotein-E isoforms 2, 3 and 4 to elucidate their structural flexibility with implications in Alzheimer's disease. *TCABSE-J*, Vol. 1, Issue 7:26-31. Epub: July 10th, 2024.

Alzheimer's disease (AD) is characterized by gradual and progressive neurodegeneration caused by neuronal cell death [1, 2]. Dementia is a general term used to describe a significant decline in cognitive ability that interferes with a person's activities of daily living [1]. AD is the most prevalent type of dementia, accounting for at least two-thirds of cases in individuals aged 65 and older. AD is a neurodegenerative condition with insidious onset and progressive impairment of behavioral and cognitive functions [3]. These functions include memory, comprehension, language, attention, reasoning, and judgment. While AD does not directly cause death, it substantially raises vulnerability to other complications, which can eventually lead to a person's death. AD is a multifactorial condition associated with many known risk factors. The most significant factor is age, typically manifested after age 65, referred to as late-onset AD (LOAD). However, early-onset AD (EOAD), occurring before 65, is less common and seen in about 5% of AD patients. EOAD often exhibits atypical symptoms, and its diagnosis is usually delayed, leading to a more aggressive disease course.

TCABSE-J Research Article
ISSN: 2583- 2557, *Volume 1, Issue 7, pp26-31. 2024.*

Figure 1. Multiple sequence alignment of ApoE2, ApoE3 and ApoE4 using CLUSTAL Omega.

The symptoms of AD can vary depending on the stage of the disease. Memory loss is the key symptom of Alzheimer's disease. Early signs include difficulty remembering recent events or conversations. But memory gets worse and other symptoms develop as the disease progresses. AD is classified into different stages based on the level of cognitive impairment and disability experienced by individuals. These stages include the preclinical or pre-symptomatic stage, mild cognitive impairment and dementia stage. The dementia stage is further divided into mild, moderate, and severe stages.

Genome-wide association studies have confirmed that the ε4 allele of APOE is the strongest genetic risk factor for AD [4, 5]. The presence of this allele is associated with increased risk for both early-onset AD and LOAD [6, 7]. The ε2 allele of APOE has protective effects against AD: the risk of AD in individuals carrying APOE ε2/ε2 or ε2/ε3 are lower than those of ε3/ε3 [8]. APOE ε4 is associated with increased prevalence of AD and lower age of onset [9-11]. APOE ε4 experiences more-rapid decline in several cognitive and functional assessments and severity of the deficits is strongly associated with the APOE ε4 gene [12-14]. APOE ε4 is associated with enhanced amyloid pathology in cognitively normal people. The proportion of Pi B-positive individuals follows a strong APOE allele-dependent pattern (ε4 > ε3 > ε2) [15-17]. Diabetes also increases the risk of AD and the association is particularly strong among APOE ε4 carriers [18-20]. A truncated fragment of ApoE4, resulting from proteolytic cleavage of ApoE following stress or injury, increases tau hyperphosphorylation, cytoskeletal disruption and mitochondrial dysfunction [21-23].

Abnormal lipid metabolism is a strongly related reason to the pathogenesis of AD. In the CNS, ApoE mediates neuronal delivery of cholesterol, which is an essential component for axonal growth, synaptic formation and remodeling events that are crucial for learning, memory formation and neuronal repair [24, 25]. APOE4 carriers are more likely to have increased severity of Lewy body pathology, independently of AD pathology [26-28]. ApoE expression in the brain was first described in astrocytes. It has been found to be drastically up regulated in activated microglia and in stressed neurons under pathological conditions and injury. The exact causes of Alzheimer's disease aren't fully understood. Scientists believe that for most people, Alzheimer's disease is caused by a combination of genetic, lifestyle and environmental factors that affect the brain over time. Risk factors include age, family history/genetics, down syndrome, sex, mild cognitive impairment, head traumas, excessive alcohol consumption, poor sleep patterns, lifestyle, etc. [29].

In this study we evaluated the structural flexibility of the N-terminal domains from ApoE2, ApoE3 and ApoE4 that contain the receptor binding domains. Molecular dynamics (MD) simulations were performed in order to elucidate the structural flexibility of the three proteins. Post simulation structural analysis was performed.

TCABSE-J Research Article
ISSN: 2583- 2557, *Volume 1, Issue 7, pp26-31. 2024.*

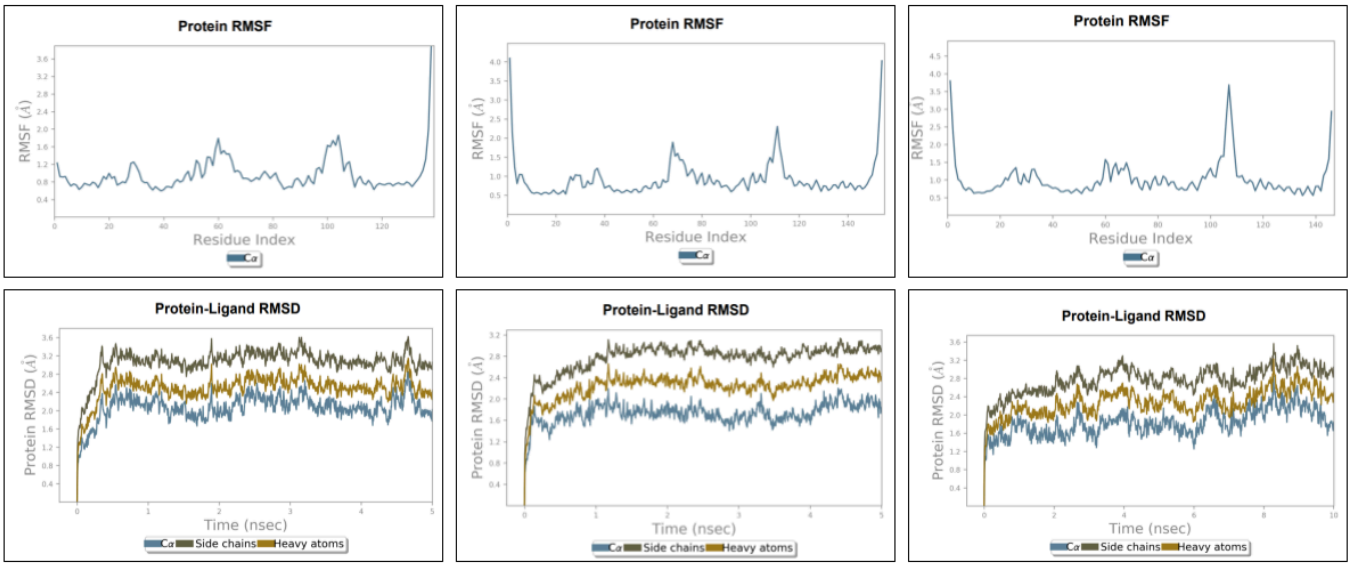


Figure 2. The overall RMSDs and RMSFs for ApoE isoforms.

Materials & Methods:

Multiple sequence alignment (MSA): The protein fasta sequences of ApoE2, ApoE3 and ApoE4 proteins were taken from the PDB IDs 1NFO, 1OR3 and 8AX8, respectively. The sequences were then pasted in the CLUSTAL Omega sequence alignment box online. MSA was performed and the result is given in Figure 1.

Preparation of ApoE structures: The 3-D structures of ApoE proteins were searched and downloaded from the protein data bank (www.rcsb.org). Structures of ApoE2, ApoE3 and ApoE4 proteins were taken from the PDB IDs 1NFO, 1OR3 and 8AX8, respectively. Upon examination of the structures using PyMOL, molecular graphics software, mainchain breaks were noticed in the structures. Hence, homology models were generated for all three proteins.

Homology model building for isoforms: Fasta sequences of ApoE2, ApoE3 and ApoE4 proteins were taken from the PDB IDs 1NFO, 1OR3 and 8AX8, respectively. Homology models were built using the SWISS-MODEL tool (https://swissmodel.expasy.org/). The final models with continuous main chain were chosen for further analysis.

Protein pre-processing: ApoE protein isoform models were pre-processed using the Maestro molecular graphics window from Schrödinger LLC, NY. Pre-processing of models includes multiple steps of optimization such as assignment of bond orders, capping the protein termini, addition of hydrogens, conversion of selenomethionine to methionine and generation of hydrogen bonds. The hydrogen atoms were then optimized by considering the water molecules in the structure. Any close contacts were rectified by relaxing the structure. The processed protein was then used to build the system for MD simulations.

System builder: The processed protein was used to set up the system with periodic boundaries containing the implicit solvent continuum using the SPC water model. The overall volume of the system was minimized by choosing the orthorhombic model. Ions were added and neutralized so that there is no net charge on the system. To mimic the physiological states, 0.15 M salt was added into the system. System builder option of Desmond (D.E. Shaw Research, NY) was used to build the systems for both CagA structure and its homology model.

MD simulations: The MD simulations were performed using Desmond on a 6th generation i5-quad core processor. The system built as described above is loaded into the Maestro molecular graphics window to display all the atoms. Periodic recording of the trajectory at fixed intervals along with the energy was set depending on the total length of the MD simulation. OPLS2005 force field was used to perform the current molecular mechanics-based simulations.

Trajectory analysis: The post MD simulation trajectory was loaded into the Maestro molecular graphics window and was analyzed frame by frame to evaluate any visible structural changes. The overall protein C_α root mean square deviations (RMSD) and per residue RMS fluctuations (RMSF) plots were saved. for each MD simulation. Coordinate files at intervals of 1 ns were saved as ".pdb" files to visualize the 3D conformational changes in the protein using Maestro.

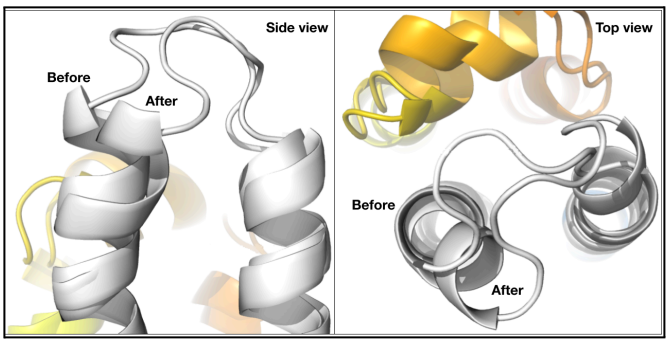


Figure 3. ApoE4 helix tilt (5 Å) after 5 ns MD simulation.

Results and Discussion:

Multiple sequence alignment: MSA of all three proteins revealed that the amino acid mismatch was found only at positions 112 and 158 as expected. As shown in Figure 1, an additional amino acid mismatch was seen at position 154. This is because the structure of ApoE2 (PDB ID: 1NFO) contains D154A amino acid substitution. Since homology models of the three proteins were used for MD simulations rather than their X-ray crystal structures, the D154A substitution is absent in the ApoE2 homology model and this mismatch can thus be ignored. This MSA analysis confirms that any structural flexibility or rigidity of the three proteins can be completely attributed to the amino acid differences at positions 112 and 158 only.

Post MD simulations trajectory analysis of ApoE proteins: The overall RMSD (y-axis) of C-alpha atoms, backbone atoms and side chains were plotted against the simulation time (x-axis). The RMSD curves for ApoE3 were parallel without any overlap over the 5 ns MD simulations while ApoE2 and ApoE4 showed some overlaps of the RMSD curves. The RMSF values of all three proteins were also plotted against the simulation time (Figure 1). All three proteins showed 3 major fluctuations FI, FII and FIII around the positions 20-40, 60-80 and 100 to 120, respectively. The FIII for ApoE4 was recorded as the highest among all reaching up to 3.0 Å while the FIII values of ApoE2 and ApoE3 were recorded as 2.0 Å and 2.5 Å, respectively, over 5 ns MD simulation. The FIII value for ApoE4 was recorded as 3.5 Å over a 10 ns MD simulation although the same fluctuation was recorded as 3.0 Å within the first 5 ns of the MD simulation for ApoE4. This high fluctuation suggests that the flexibility of ApoE4 is greater than the other two proteins, especially around the third fluctuation site. Contrarily, the FII values of ApoE2 and ApoE3 were recorded as 2.0 Å while the FII value of ApoE4 was <2.0 Å. These results suggest that the ApoE4 has highest flexibility at FIII probably due to the presence of two arginines.

Structural deviations in ApoE4: Based on the RMSF values, it was predicted that the ApoE4 may have some structural deviations around the FIII site. Hence structural analysis was performed by superposing the structure of ApoE4 before and after 5 ns MD simulations. As shown in Figure 3, deviation in the helices was clearly observed in both the side view and top view. A tilt of 5 Å was seen in the helix that can be attributed to the high RMSF at the FIII site of ApoE4 from the superposed structures. This structural deviation continued into the loop followed by the helix where the deviation was calculated to be 4 Å. We further investigated whether the presence of arginine, R158, may have any effect on this helix tilt.

Polar contact between R158 and Ra180 is broken during the MD simulation: Structural analysis of ApoE4 before the MD simulation showed that the side chain guanidine group of R158 makes a polar contact with the backbone carbonyl oxygen atom of R180. However, after the 5 ns MD simulation, this polar contact was broken and the side chain conformation of R158 is flipped towards the E165 with water molecules surrounding the guanidine group of R158 thus preventing it from making any contacts with R180. Due to this side chain flip, the helix tilt was generated causing a 5 Å structural deviation as discussed above. The initial conformation of the R158 side chain is stabilized by the polar contacts with E165, R180 and one water molecule. After 5 ns MD simulation, the polar contact with E165 was still maintained but the polar contact with R180 was replaced with multiple polar contacts with four water molecules.

It is interesting to see that the polar contact between R158 and R180 did not exist in the original X-ray crystal structure but was seen in the homology model after preparation for the MD simulation. Part of the reason for this difference is due to the absence of hydrogens in the crystal structure. During the system builder step of the MD simulation protocol, plenty of water molecules were added to the system along with the physiological salt concentration and other ions to neutralize the overall system. Additionally before the production run of the MD simulation, multiple equilibration runs were performed due to which the polar contact was established between R158 and R180. Thus, this polar contact was visible in the frame-0 of the MD simulation which was taken as the model before MD that was superposed onto the final model taken after the completion of the 5 ns MD simulations.

Conclusion and Future directions:

The current study sheds light on the structural deviations in ApoE4 compared to ApoE2 and ApoE3 by using a short MD simulation as basis. Initially we hypothesized that R112 and R158 may play a role in the structural flexibility/deviations in the ApoE4 protein compared to ApoE2 & ApoE3 proteins.

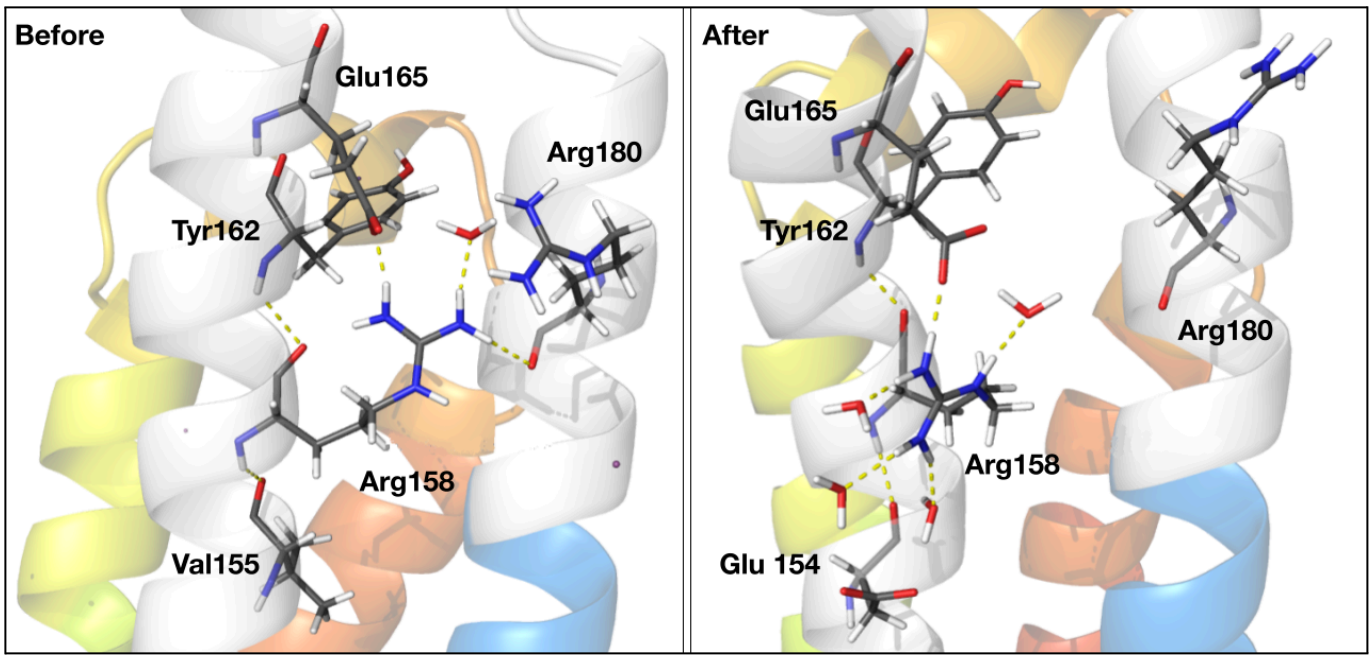


Figure 4. R158-R180 polar contact before & after 5 ns MD.

However, our MD simulations revealed that major structural deviations in ApoE4 are caused due to the broken polar contact between R158 and R180. In future, longer MD simulations will be performed up to 100 ns to check whether this finding still supports our current finding. Additionally, full length ApoE proteins will be simulated in future.

The structural flexibility that we identified in this study is probably contributing to the major structural changes in ApoE4 due to which, it is unable to bind to the ApoE receptor which in turn might affect the lipid transportation in the brain and cause an imbalance in the lipid:protein ratio. Due to this imbalance, one can predict that the protein content in the cytoplasm may increase and alter the solubility properties of the beta amyloid protein leading to the formation of plaques. Furthermore, similar MD simulations should be performed in future using the full length ApoE proteins including the membrane to cross check whether our current findings match with them. Our predictions in this conclusion need further experimental validation in the future.

Acknowledgements: We thank The Yedidi Institute of Discovery and Education, Toronto for scientific collaborations and funding.

Conflict of interest: The authors declare no conflict of interest in this study.

Author contributions: A.Y., A.D. and S.B. performed MSA, Homology modeling, MD simulations. A.Y. and A.D. prepared the post simulation analysis reports. S.G. supervised A.Y., A.D. and S.B. R.S.Y. is the principal investigator who designed the project, trained A.Y., A.D. and S.B., secured required material for the project, provided the laboratory space and facilities needed. R.S.Y. wrote and edited the manuscript.

References

1. Kumar A, Sidhu J, Lui F, Tsao JW. Alzheimer Disease. In: StatPearls. Treasure Island (FL): StatPearls Publishing; February 12, 2024.
2. Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules. 2020;25(24):5789. Published 2020 Dec 8. doi:10.3390/molecules25245789.
3. Maitre, M., Jeltsch-David, H., Okechukwu, N.G. et al. Myelin in Alzheimer's disease: culprit or bystander?. acta neuropathol commun 11, 56 (2023). https://doi.org/10.1186/s40478-023-01554-5.
4. Chartier-Harlin MC, et al.Harold D, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 2009;41:1088–93.
5. Lambert JC, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 2009;41:1094–9.
6. Apolipoprotein E, epsilon 4 allele as a major risk factor for sporadic early and late-onset forms of Alzheimer's disease: analysis of the 19q13.2 chromosomal region. Hum Mol Genet. 1994;3:569–74.
7. ApoE genotype is a risk factor in nonpresenilin early-onset Alzheimer's disease families. Am J Med Genet 1998;81:117.

8. Houlden H., Farrer La, *et al*. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: A meta-analysis. JAMA. 1997;278:1349–1356.
9. Corder EH, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993;261:921–3.
10. Farrer La *et al*. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: A meta-analysis. JAMA. 1997;278:1349–1356.
11. Willam Rebeck G, Reiter JS, Strickland DK, Hyman BT. Apolipoprotein E in sporadic Alzheimer's disease: Allelic variation and receptor interactions. Neuron. 1993;11:575–580.
12. Farlow MR, et al. Impact of APOE in mild cognitive impairment. Neurology. 2004;63:1898.
13. Whitehair DC, et al. Influence of apolipoprotein E ε4 on rates of cognitive and functional decline in mild cognitive impairment. Alzheimer's dement. 2010;6:412–419.
14. Cosentino S, et al. APOE ε4 allele predicts faster cognitive decline in mild Alzheimer disease. Neurology. 2008;70:1842–1849.
15. Kok E, et al. Apolipoprotein E–dependent accumulation of Alzheimer disease–related lesions begins in middle age. Ann Neurol. 2009;65:650–657.
16. Castellano JM, et al. Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. Sci Transl Med. 2011;3:89ra57.
17. Morris JC, et al. APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. Ann Neurol. 2010;67:122–131.
18. Piela R, Rodriguez BL, Launer LJ. Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies: The Honolulu-Asia Aging Study. Diabetes. 2002;51:1256–62.
19. Matsuzaki T, et al. Insulin resistance is associated with the pathology of Alzheimer disease: the Hisayama study. Neurology. 2010;75:764–70.
20. Irie F, et al. Enhanced risk for Alzheimer disease in persons with type 2 diabetes and APOE epsilon4: the Cardiovascular Health Study Cognition Study. Arch Neurol. 2008;65:89–93.
21. Huang Y. Aβ-independent roles of apolipoprotein E4 in the pathogenesis of Alzheimer's disease. Trends Mol Med. 2010;16:287–294.
22. Brecht WJ, et al. Neuron-specific apolipoprotein E4 proteolysis is associated with increased Tau phosphorylation in brains of transgenic mice. J Neurosci. 2004;24:2527–2534.
23. Mahley RW, Weisgraber KH, Huang Y. Apolipoprotein E4: A causative factor and therapeutic target in neuropathology, including Alzheimer's disease. Proc Natl Acad Sci USA. 2006;103:5644–5651.
24. Mauch DH, et al. CNS synaptogenesis promoted by glia-derived cholesterol. Science. 2001;294:1354–1357.
25. Pfrieger FW. Cholesterol homeostasis and function in neurons of the central nervous system. Cell Mol Life Sci. 2003;60:1158–1171.
26. Dickson DW, Heckman MG, Murray ME, Soto AI, Walton RL, Diehl NN, van Gerpen JA, Uitti RJ, Wszolek ZK, Ertekin-Taner N, et al. APOE ε4 is associated with severity of Lewy body pathology independent of Alzheimer pathology. Neurology. 2018;91:e1182–95.
27. Outeiro TF, Koss DJ, Erskine D, Walker L, Kurzawa-Akanbi M, Burn D, Donaghy P, Morris C, Taylor JP, Thomas A, et al. Dementia with Lewy bodies: an update and outlook. Mol neurodegeneration. 2019;14:5. 28.van Steenoven I, Koel-Simmelink MJA, Vergouw LJM, Tijms BM, Piersma SR, Pham TV, Bridel C, Ferri GL, Cocco C, Noli B, et al. Identification of novel cerebrospinal fluid biomarker candidates for dementia with Lewy bodies: a proteomic approach. Mol neurodegeneration. 2020;15:
28. van Steenoven I, Koel-Simmelink MJA, Vergouw LJM, Tijms BM, Piersma SR, Pham TV, Bridel C, Ferri GL, Cocco C, Noli B, et al. Identification of novel cerebrospinal fluid biomarker candidates for dementia with Lewy bodies: a proteomic approach. Mol neurodegeneration. 2020;15
29. Silva MVF, Loures CMG, Alves LCV, de Souza LC, Borges KBG, Carvalho MDG. Alzheimer's disease: risk factors and potentially protective measures. *J Biomed Sci*. 2019;26(1):33. Published 2019 May 9. doi:10.1186/s12929-019-0524-y.

Full figure legends:

Figure 1. Multiple sequence alignment of ApoE2, ApoE3 and ApoE4 using CLUSTAL Omega. Fasta sequences of ApoE2, ApoE3 and ApoE4 proteins were taken from the PDB IDs 1NFO, 1OR3 and 8AX8, respectively. A total of 162 amino acids showed proper alignment indicated by the asterisks. Positions 112 and 158 showed mismatch as expected.

Figure 2. The overall RMSDs and RMSFs for ApoE isoforms. The top and bottom panels display the overall RMSF profiles and RMSD profiles, respectively of ApoE2 (left panel), ApoE3 (middle panel) and ApoE4 (right panel). All three proteins show three major fluctuation areas, FI, FII and FIII of which the FIII of ApoE4 is highest compared to the rest. The three RMSD curves for ApoE3 are parallel to each other but they overlap partially for ApoE2 and ApoE4.

Figure 3. ApoE4 helix tilt (5 Å) after 5 ns MD simulation. The homology model of ApoE4 before and after 5 ns MD simulations were superposed for structural analysis. A helix tilt of 5 Å was identified that continued into the following loop. This helix tilt is clearly visible both in the side view and top view.

Figure 4. R158-R180 polar contact before & after 5 ns MD. R158 shows multiple polar contacts within the helix and also with R180 from the neighboring helix before simulation. The polar contact between R158 and R180 was broken during the MD simulation thus creating the helix tilt.