MDPI

Article

CD33 rs2455069 SNP: Correlation with Alzheimer's Disease and Hypothesis of Functional Role

Fabiana Tortora [1,†] , Antonella Rendina [1,†] , Antonella Angiolillo [2] , Alfonso Di Costanzo [2] ,
Francesco Aniello [3] , Aldo Donizetti [1,3,*] , Ferdinando Febbraio [1,‡] and Emilia Vitale [1,*,‡]

[1] Institute of Biochemistry and Cell Biology, National Research Council (CNR), Via Pietro Castellino 111, 80131 Naples, Italy; fabiana.tortora@ibbc.cnr.it (F.T.); antonella.rendina@ibbc.cnr.it (A.R.); ferdinando.febbraio@cnr.it (F.F.)

[2] Centre for Research and Training in Medicine of Aging, Department of Medicine and Health Science "V. Tiberio", University of Molise, 86100 Campobasso, Italy; angiolillo@unimol.it (A.A.); alfonso.dicostanzo@unimol.it (A.D.C.)

[3] Department of Biology, University of Naples Federico II, 80126 Naples, Italy; faniello@unina.it

[*] Correspondence: aldo.donizetti@unina.it (A.D.); emilia.vitale@cnr.it (E.V.)

[†] These authors contributed equally to this work.

[‡] These authors contributed equally to this work.

Abstract: The *CD33* gene encodes for a member of the sialic-acid-binding immunoglobulin-type lectin (Siglec) family, and is one of the top-ranked Alzheimer's disease (AD) risk genes identified by genome-wide association studies (GWAS). Many *CD33* polymorphisms are associated with an increased risk of AD, but the function and potential mechanism of many CD33 single-nucleotide polymorphisms (SNPs) in promoting AD have yet to be elucidated. We recently identified the *CD33* SNP rs2455069-A>G (R69G) in a familial form of dementia. Here, we demonstrate an association between the G allele of the rs2455069 gene variant and the presence of AD in a cohort of 195 patients from southern Italy. We carried out in silico analysis of the 3D structures of *CD33* carrying the identified SNP to provide insights into its functional effect. Structural models of the CD33 variant carrying the R69G amino acid change were compared to the CD33 wild type, and used for the docking analysis using sialic acid as the ligand. Our analysis demonstrated that the CD33-R69G variant may bind sialic acid at additional binding sites compared to the wild type, thus potentially increasing its affinity/specificity for this molecule. Our results led to a new hypothesis of rs2455069-A>G SNP as a risk factor for AD, suggesting that a long-term cumulative effect of the CD33-R69G variant results from the binding of sialic acid, acting as an enhancer of the CD33 inhibitory effects on amyloid plaque degradation.

Keywords: CD33; Alzheimer's disease; sialic acid; phagocytosis

1. Introduction

According to recent reports, it is estimated that the number of cases of dementia in the world is expected to rise to 152 million by 2050 as the population ages [1,2], with AD being the most common form of dementia currently [3,4]. Most AD cases are sporadic, with late-onset AD (LOAD) occurring at the age of 65 years or older. Approximately 5% of cases are classified as early-onset AD (EOAD), as reported in a systematic review and meta-analysis [5], where the age of onset is before 65 years and a percentage is considered as familial [6,7]. Except for a small number of early-onset cases who are afflicted due to highly penetrant single-gene mutations, AD, especially in the late-onset form, is genetically heterogeneous, with a polygenic or oligogenic risk inheritance [7]. Genome-wide association studies (GWAS) identified several genetic loci associated with increased susceptibility to LOAD affecting lipid metabolism, tau binding proteins, amyloid precursor protein (APP) metabolism [8], protein catabolism, immune cells, and microglia [9]. Of these, microglial

cells are a key participant in LOAD pathogenesis, with many susceptibility loci, including *CD33*, playing a role [10–15]. CD33 belongs to a subfamily of the Ig superfamily comprising a single N-terminal V-set Ig domain that binds sialic-acid-containing glycans through a conserved positively charged arginine that interacts with the negatively charged C-1 carboxyl group on the sialic acid [16]. CD33 may be activated continuously by sialic-acid-containing glycoproteins and glycolipids, which are structural elements of amyloid plaques in the brains of Alzheimer's disease patients, resulting in an inhibition of microglia-mediated immune activation [17,18]. CD33 can affect microglial phagocytosis through crosstalk with the transmembrane receptor TREM2. Knockout of CD33 resulted in improved cognition and decreased amyloid β pathology in 5xFAD mice, whereas these effects were completely abrogated by additional knockout of TREM2 [19]. TREM2, through an interaction with DAP12/TYROBP, stimulates the tyrosine–protein kinase SYK, which in turn triggers the activation of microglial cell phagocytosis [20,21]. This activation could be counteracted by sialic-acid-activated CD33, which stimulates SHP1/SHP2 tyrosine phosphatases, thereby inhibiting microglial phagocytosis [13,21,22].

In accordance with the effect of *CD33* on the inhibition of microglial phagocytosis, primary microglial cells from *CD33* knockout mice [12] or deletion of *CD33* in human macrophages and microglia resulted in an increased phagocytosis of amyloid β1-42 [23]. This was further supported by targeting *CD33* in a mouse model of AD using an adeno-associated virus (AAV) vector encoding a *CD33* microRNA (miRCD33), which demonstrated a reduction in amyloid β40 and β42 levels in brain extracts [24].

Different *CD33* polymorphisms have been positively or negatively correlated with AD susceptibility. These SNPs can affect the expression level, the structure, and the function of CD33, leading to changes in microglia-mediated clearance of amyloid β, likely contributing to the accumulation of senile plaques in the brain [12,25–27]. The CD33 variant rs3865444(A) negatively correlates with AD, has a reduced *CD33* expression [12], and is coinherited with rs12459419(T), which modulates the splicing efficiency of the IgV domain exon involved in sialic acid binding [13,28].

We recently identified the SNP rs2455069 in exon 2 of the human *CD33* gene on chromosome 19, which is involved in an unusual familial form of dementia [29]. This SNP has been previously identified in a block of eight correlated SNPs in *CD33* significantly associated with cognitive decline in univariate analysis in the all-female cohort, with highly significant effects on hyaluronan synthase-1 (HAS1) expression in the temporal cortex [30]. The presence of the minor allele at rs2455069 leads to a change from an arginine to a glycine in the position 69 (R69G). Here, we demonstrate a positive correlation between the *CD33* SNP rs2455069 and Alzheimer's disease in a cohort of patients from a region of southern Italy, and secondly, that the R69G amino acid change affects the CD33 structure in a way that alters its binding affinity for sialic acid residues. We propose that the mechanism of the long-term cumulative effects of the CD33-R69G variant are mediated through an increased binding of sialic acid, which then acts as a more effective enhancer of the CD33 inhibitory effects on amyloid plaque degradation.

2. Results

2.1. Genetic Analysis

Based on the evidence of the possible involvement in an unusual form of dementia of the SNP rs2455069-A>G in exon 2 of *CD33* gene [29], leading to a change from arginine to glycine in position 69, we widened the analysis of this SNP to a cohort of AD patients from southern Italy. We found that the genotype frequency for the rs2455069 polymorphism in the control group was 11.1% ($n = 15$) for the GG, 53.0% ($n = 77$) for the AG, and 31.8% ($n = 43$) for the AA genotypes. In AD patients, the frequency of GG, AG, and AA genotypes were 17.9% ($n = 35$), 65.1% ($n = 127$), and 16.9% ($n = 33$), respectively. While genotypes in the LOAD sample were not in Hardy–Weinberg equilibrium (HWE, $p = 0.00013$), the genotype frequencies in controls did not deviate significantly from the Hardy–Weinberg equilibrium (HWE) ($p = 0.08301$). The G allele frequency was 39.6% in controls and 50.5%

in AD patients (p = 0.00582). A significant association between AD and the GG or AG genotype was found (Table 1), particular in the dominant model (AG + GG vs. AA), leading to the hypothesis that the R69G change may have had an impact on the increased risk of LOAD in the analyzed cohort.

Table 1. Association between the genetic polymorphism CD33-rs2455069 and the risk of AD.

CD33 Genotype	OR [a] (95% C.I. [b])	Chi2	p
AA	Ref		
AG	2.149 (1.259–3.668)	8.02	p = 0.00463
GG	3.040 (1.428–6.476)	8.58	p = 0.00341
AG + GG	2.294 (1.363–3.862)	10.03	p = 0.00154
A	Ref		
G	1.555 (1.135–2.130)	7.61	p = 0.00582

[a] OR = odds ratio; [b] C.I. = confidence interval.

2.2. In Silico Structural Analysis

To evaluate the impact of the R69G change on the CD33 protein function, we carried out a series of in silico analyses. Six human CD33 partial structures (Siglec-3) were available in the RCSB Protein Data Bank [31] (Supplementary Table S1, Supplementary Figure S1). We prepared the monomeric form of each CD33 structure, summarized in Supplementary Table S1, using the web-based platform CHARMM-GUI. To reduce possible artifacts due to binding with mimetic substrates, we used a protein model based on the CD33 structure PDB-ID 5IHB, representing the protein region interacting with NAG as a nonspecific ligand. We compared the two structures obtained after CHARMM minimization carrying either arginine or glycine at position 69 of the CD33 protein. We did not observe significant changes in the protein backbone after basic energy minimization, but small rearrangements, both in the α-helix containing the glycine residue (red arrow in Figure 1A) and in a β-sheet of the repeated structural type C2 Ig domain (black arrow in Figure 1A) were evident. Most importantly, few differences were observed in the structures of sidechains, which showed small changes in the connecting region between the two domains (Figure 1B).
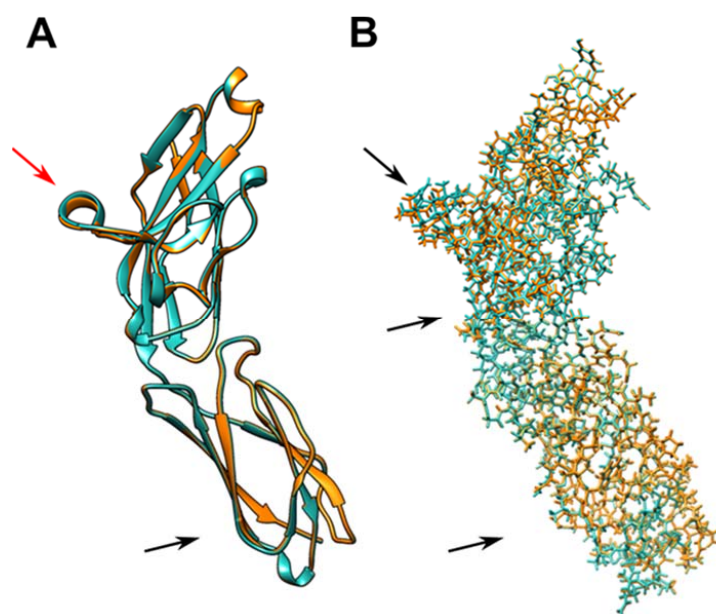

Figure 1. Structural representation of CD33 variants. (**A**) Ribbon representation of the backbone of superimposed CD33 monomers (PDB ID 5IHB) with Arg69 (cyan) and Gly69 (orange) after minimization

by CHARMM. (**B**) Stick representation of backbone and sidechains of structures in (**A**). The arrows indicate the area with major changes in the structure of sidechains after basic energy minimization.

　　These predictions led us to hypothesize that the effect of the arginine-to-glycine change at position 69 could predominantly affect the reorganization of polar sidechains, and only locally, the rearrangement of the backbone. In agreement with these observations, the analysis carried out on 200 structural models of CD33 (PDB ID 5IHB), 100 with the Arg69 residue and 100 with the Gly69 residue, indicated the presence of few changes in the boundary spatial coordinates (min and max) of the entire CD33 structure (Figure 2). These differences were hypothesized to be due to the increased length of the sidechain consisting of a three-carbon aliphatic straight chain ending in a guanidino group as a consequence of the change from a glycine to an arginine residue in position 69.
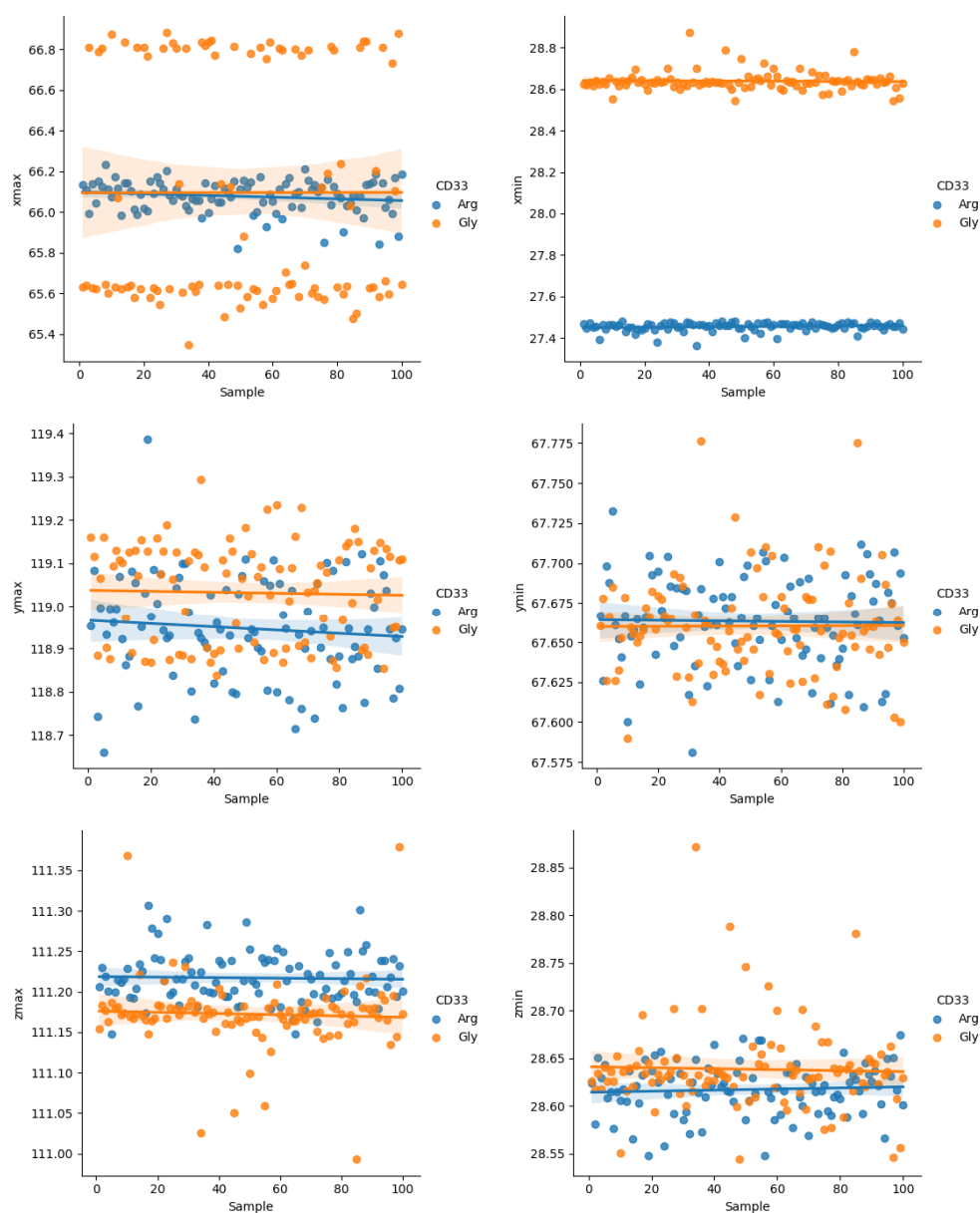

Figure 2. Variations of the maximum and minimum dimensional values in the XYZ axes of the two CD33 variants (PDB ID 5IHB). The variations were obtained after minimization by CHARMM. Blue dots refer to the structures with the arginine residue, whereas the orange dots refer to the structures with a glycine residue in position 69.

2.3. Docking Analysis

The predicted binding sites for sialic acid in the N-terminal domain (aa19-135) of the monomeric form of CD33 (PDB ID 5IHB) carrying either an arginine or a glycine residue in position 69 are reported in Figure 3. The affinity of each binding site toward the ligand is expressed as the binding energy (kcal/mol), with lower values indicating a greater bond affinity. In the Arg69 variant, the values were −4.8 kcal/mol near Arg119 and −5.3 kcal/mol near Arg91, while the values were slightly increased in the form with Gly69, with values of −5.2 kcal/mol near Arg119 and −5.7 kcal/mol near Arg91 (Figure 3).
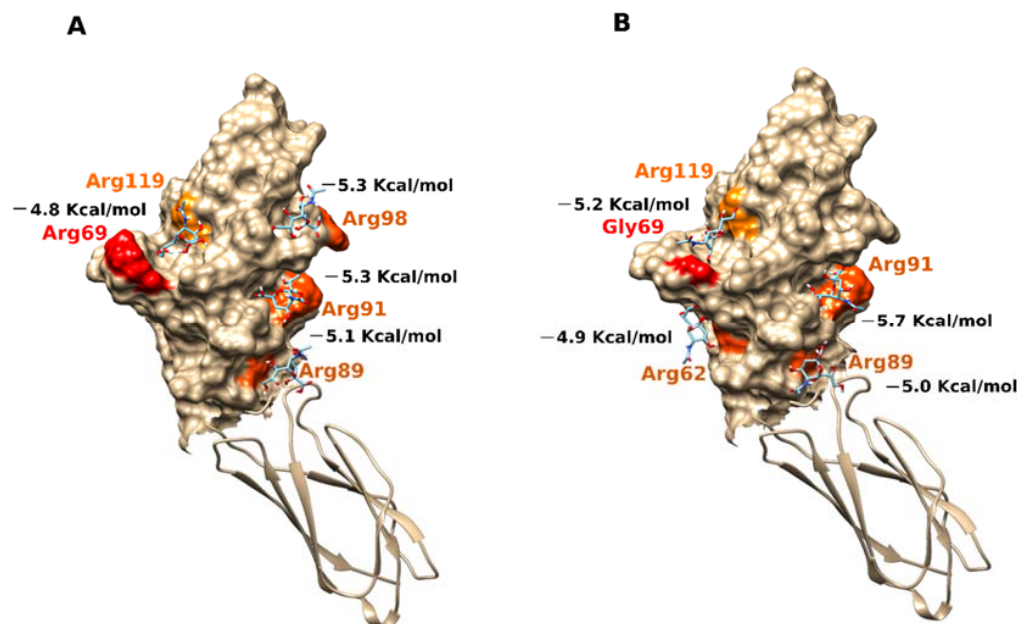

Figure 3. Sialic acid interactions with the two SNPs of CD33 structures. (**A**) Surface representation of functional domain of the CD33 predominant form with Arg in position 69, colored in red (PDB ID 5IHB). (**B**) Surface representation of functional domain of the CD33 variant form with Gly in position 69, colored in red (PDB ID 5IHB). The areas in orange indicate the binding sites of sialic acid, in stick representation, in correspondence with the positively charged arginine residues with which the ligand establishes saline bridges. The values of the binding energy in kcal/mol are reported for each binding site of sialic acid.

We found similar results when we performed docking analyses on the other two structures: the N-terminal domain (aa19-135) (PDB ID 5J06, 5J0B), which carried a glycine residue in position 69; and the repeated structural type C2 Ig domain (aa145-228). In the PDB-ID 5J06 structure, the values were −4.7 kcal/mol in Arg119 proximity and −5.2 kcal/mol in Arg91 proximity; while in the PDB-ID 5J0B structure, the predicted binding affinity values were −4.7 kcal/mol in Arg119 proximity and −5.3 kcal/mol in Arg91 proximity (Supplementary Figure S2). We confirmed these results through the docking analyses carried out on the other three CD33 structures having only the N-terminal domain (PDB ID 6D48, 6D49, and 6D4A). In fact, we always observed a greater binding affinity in the Arg91 residue part than in the Arg119 part of the functional domain (Supplementary Figure S3), suggesting a better chance of binding at this site. This result was in contrast with the evidence from the binding of a sialic acid mimetic ligand FVP in the proximity of Arg119, as found by others [32]. However, by analyzing the asymmetric unit assembly of each resolved structure in that study, we observed the formation of dimers and tetramers involving interfaces near Arg91, which likely prevented the binding of ligands near this residue. Considering that the most reactive side of the D2:IgV domain is represented by the positively charged patch comprising arginine residues 89 and 91, the oligomerization trend involving this surface could be explained easily. Notwithstanding the site of binding,

our results suggested a general increase in the binding affinity and specificity of the glycine variant of CD33 toward ligands.

2.4. Virtual Screening Analysis

Plotting the carbon 1 atomic coordinate of the best sialic acid poses against the affinity values (kcal/mol) predicted by the docking analysis, we obtained a conceptual understanding of the ligand distributions on the protein surface with respect to the arginine or glycine variants in position 69 of the CD33 protein. The results highlighted a noteworthy increase in the affinity of several binding sites for sialic acid in the glycine 69 variant of CD33 (Supplementary Figure S4). Indeed, the analysis of the distributions of the best ligands on the D2:IgV domain of CD33, obtained by plotting the XYZ coordinates of the C1 atom of sialic acid structures (Supplementary Figure S5), showed binding in specific sites of CD33 receptors for both variants with arginine and glycine residues in position 69. When comparing the position of arginine residues in the D2:IgV domain of the PDB-ID 5IHB structure with the best ligand poses by overlaying the YZ coordinates of the Cζ atom of arginine sidechains with the coordinates of C1 atom of sialic acid poses (Supplementary Figure S6), we observed a number of sialic-acid-binding molecules close to arginine residues 91, 98, and 119. However, the C1 atom of sialic acid near the approximate Z coordinate 70,000, included in the range of Y coordinates 85,000–95,000 and 110,000–115,000, were instead close to the arginine 62 and 89 residues in the 3D representation, respectively. It was apparent that ligands basically spread out on the surface of the Arg69 structure of CD33, with most of the sialic acid binding near arginine residues 91, 119, 89, 98, and 62, ordered by the number of bonded ligands (Supplementary Figure S6A). In contrast, an increased affinity and number of sialic acid molecules was predicted near the position of arginine 91 in the Gly69 variant of CD33 (Supplementary Figure S6B).

These data suggested the possibility of the presence of more than one binding site for sialic acid derivatives with a similar specificity. In fact, it was not surprising that the more reactive domain near arginine 91 was involved in the binding at the interfaces between the CD33 structures (single and double domains) in the crystallographic unit cell, preventing the recognition of this other sialic-acid-binding site in the crystallographic structures resolved in presence of ligands [32].

The 3D representations in Figure 4 illustrate the best orientations of sialic acid molecules in the CD33 binding sites obtained by virtual screening analysis. In the CD33 structure presenting an arginine residue in position 69, we observed the prevalent distribution of the ligands in the two already identified binding sites involving arginine 91 and 119 (Figure 4A). However, two additional binding sites were predicted near arginine 62 and 89. The number of ligands in these two secondary sites increased in the representation of the second-best poses (Figure 4B), having a lower affinity with respect to the previous ones. In contrast, the best poses of sialic acid molecules were located exclusively near arginine 91 and 119 in the Gly69 variant of CD33 (Figure 4C), and only a few ligand poses were predicted near arginine 62 and 89 (Figure 4D). In addition, when observing the manner the sialic acid binds to the sites on the CD33 surface, we can point out that the orientation of sialic acid molecules was very often preferentially maintained in the binding sites of the Gly69 variant of CD33 (Figure 4). Thus, in addition to the increase in affinity, it was likely that the Gly69 variant of CD33 bound to the sialic acid with a higher specificity than the Arg69 variant. These results supported the idea that the balance of the binding equilibrium constant for CD33–sialic acid interaction changed slightly in a positive direction in the Gly69 variants. This suggested a more efficient binding of sialic acid at a lower concentration by this CD33 variant.
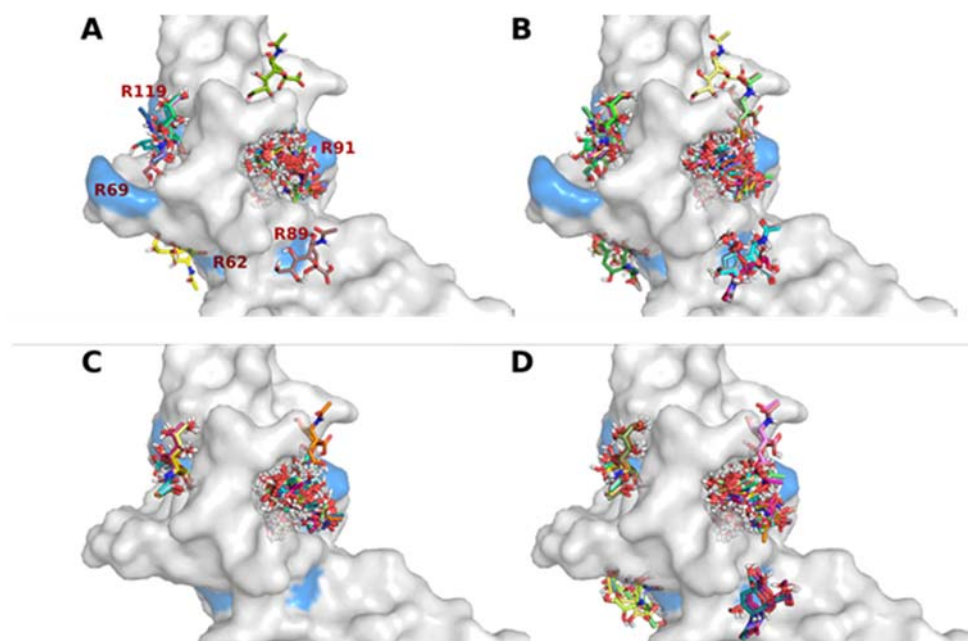
Figure 4. Detail of the binding sites in the D2:IgV domain for sialic acid molecules, as derived from docking analyses. Surface representation, in light gray, of the N-terminal D2:IgV functional domain of arginine 69 models of CD33, in which the docked poses of sialic acid molecules with the best (**A**) and the second best (**B**) ligand mode rmsd are shown in stick figures. Surface representation, in light gray, of the N-terminal D2:IgV functional domain of glycine 69 models of CD33, in which the docked poses of sialic acid molecules with the best (**C**) and the second best (**D**) ligand mode rmsd are shown in stick figures. The surfaces of arginine residues 62, 69, 89, 91, and 119 are presented in light blue.

3. Discussion

Although the association of CD33 with Alzheimer's diseases has not been thoroughly elucidated, literature data supports the hypothesis that through the modulation of immune cell functions (such as phagocytosis, cytokine release, apoptosis, etc.), CD33 could be implicated in the suppression of amyloid fragment uptake [12,18,23,25–27]. Here, we demonstrated a correlation between an SNP variant of the *CD33* gene (SNP rs2455069-A>G) and Alzheimer's disease risk, and proposed a new hypothesis for its functional role. Using different in silico approaches, we demonstrated that the binding site for sialic acid could involve additional regions of the protein to the previously documented Arg119 residue. This residue was demonstrated to be essential for the binding of the sialic acid mimetic [32]. Indeed, it was assumed that the binding site for sialic acid was located at the D2:IgV domain of the CD33 receptor, and contained the arginine 119 residue, which was positively charged at physiological pH. Because sialic acid molecules are typically terminal residues on glycoconjugates, the positively charged arginine forms salt bridges with the sialic acid portion of sugar residues, enabling stable interactions [18,33–35]. Due to the presence of FVP (a subtype-selective sialic acid mimetic ligand) in the 6D49 and 6D4A structures (recently also demonstrated in the 7AW6 structure), the putative binding site was located near the Arg69 and Arg119 positions (Supplementary Figure S7). In agreement with those results, we confirmed the binding of sialic acid to Arg119, but also demonstrated the possibility that the ligand may bind to other regions of the protein with higher affinity that include Arg91. The presence of an additional binding site was recently reported for another Siglec family member [36], pointing out the need to further explore the binding properties of CD33 toward sialic-acid-containing molecules. A mutation of arginine 119 to alanine did not affect phagocytosis of U937 cells [27], raising a question on the role of this residue in sialic-acid-containing glycan binding, and supporting a hypothesis of additional binding sites for sialic acid. Our *in-silico* analyses indeed supported the possibility of additional

binding sites for sialic acid. In addition, our studies also supported a hypothesis that the CD33-R69G variant may bind sialic acid with greater specificity and affinity, and results in a more efficient binding of the variant to sialic acid, at lower concentrations than that of the wild type. Interestingly, the dependence of binding on sialic acid concentrations was consistent with observations that sialic acid content increases in old-age females [37], who have an increased risk of LOAD compared to males of the same age [38]. In the light of these observations, new perspectives have opened up in the study of CD33, with novel and interesting understandings of the interactions leading to late-onset Alzheimer's disease. Several polymorphisms in the CD33 gene could affect the structure of this receptor, and in turn, the binding affinity for sialic-acid-containing gangliosides. We hypothesized that the microglial phagocytosis, and in turn, the efficient removal of neurotoxic protein aggregates, might also depend on a variety of factors, including dietary habits, sex-specific accumulation of sialic acid, and aging, among others, affecting ganglioside expression in the brain. This may establish a complex inter-relationship between genetic risk factors and additional predictors in the progression of LOAD. Overall, our results led us to hypothesize that the rs2455069-A>G SNP is a risk factor for LOAD, and that the mechanism of the long-term cumulative effects of the CD33-R69G variant are mediated through an increased binding of sialic acid, which then acts as a more effective enhancer of the CD33 inhibitory effects on amyloid plaque degradation (Figure 5).
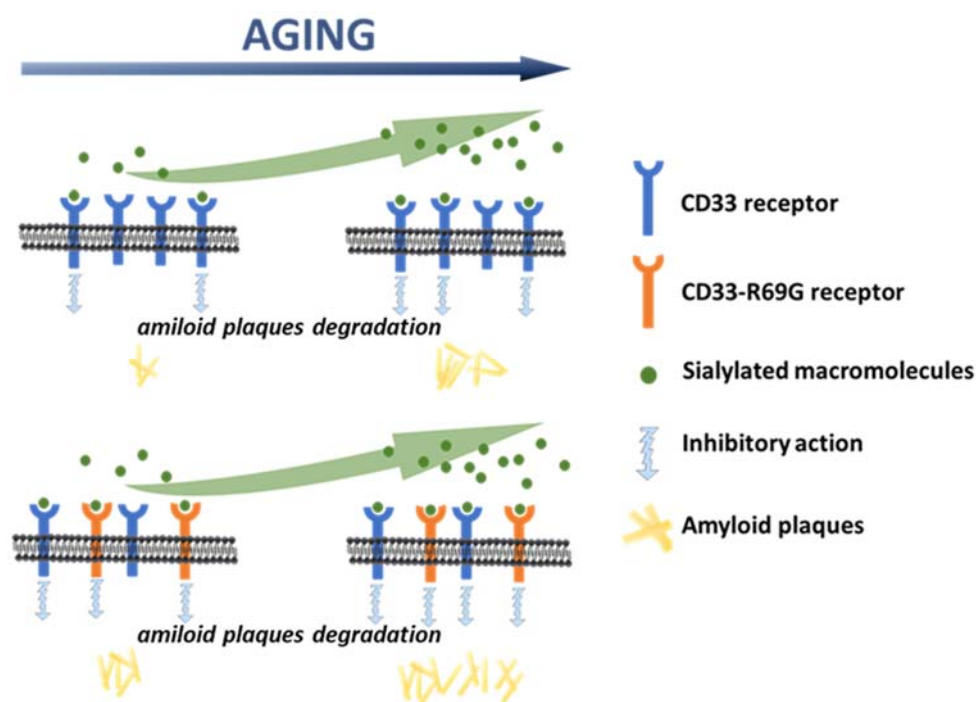

Figure 5. Model describing the plaque accumulation during aging. The increased effect of the CD33-R69G variant in inhibiting the microglial phagocytosis results in a lower extent of clearance of amyloid plaques, supporting late-onset AD.

4. Conclusions

We analyzed the functional role of the SNP rs2455069-A>G, which we found to be associated with Alzheimer's disease (AD) in a cohort of patients from southern Italy. The SNP changed a single amino acid in the sequence of CD33, and resulted in a local change in its protein structure, which we derived by in silico analysis. Our data demonstrated that this small change resulted in a slight increase in the binding affinity and specificity of CD33 toward sialylated molecules, its natural ligands. We propose that this change in binding enhanced the inhibitory effects of CD33 on amyloid plaque degradation, and together with other genetic, epigenetic, and environmental factors, could predispose to the accumulation

of amyloid plaques and the onset of AD pathology. These observations support a new approach to the investigation of protein polymorphisms involved in AD, and propose a need for structural analyses to complements genetic studies.

5. Materials and Methods

5.1. Participant Recruitment and Analysis

The participants in the study (n = 330) were recruited at the Centre for Research and Training in Medicine of Aging (CeRMA) at the University of Molise (Italy). The AD patients (n = 195) fulfilled the National Institute of Aging and Alzheimer's Association (NIA-AA) diagnostic criteria for "probable AD with documented decline" [39]. They scored <24 on the Mini Mental State Examination (MMSE) and >0.5 on the Clinical Dementia Rating (CDR). To rule out other potential causes of cognitive impairment, all patients underwent blood tests (including full blood count, erythrocyte sedimentation rate, blood urea nitrogen and electrolytes, thyroid function, vitamin B12, and folate) and brain imaging. One hundred and thirty-five sex/age-matched cognitively healthy subjects (HS) were recruited as a control group. DNA from all participants was analyzed for the presence of the rs2455069 SNP by high-resolution melting analysis (HRM). Genomic DNA was extracted from whole blood with a QIAamp DNA Blood Mini Kit (QIAGEN). PCR amplifications were performed on a SureCycler 8800 Thermal Cycler (Agilent, Santa Clara, CA, USA). The study was conducted in accordance with ethical principles stated in the Declaration of Helsinki, and with approved national and international guidelines for human research. The Institutional Review Board (IRB) of the University of Molise approved the study (IRB Prot. n. 16/2020). Written informed consent was obtained from each participant or caregiver.

5.2. High-Resolution Melting Analysis (HRM)

The high-resolution melting analysis (HRM) was performed as previously described [29] on a LightCycler 480 Instrument II (Roche, Basel, Switzerland) and analyzed using LightCycler® 480 Gene Scanning Software 1.5.1. High-resolution melting Mastermix (Roche) was used. Specific primers used for amplification and sequencing by the Sanger method and the electropherograms were analyzed with Chromas 2.22 software and aligned according to reference sequences present in GenBank (http://www.ncbi.nlm.nih.gov/GenBank/index.html; accessed on 1 September 2019). Templates were amplified in 96-well plates and ~30 ng analyzed in a 20 μL final volume reaction as previously specified [29].

5.3. CD33 Structure Editing and Minimization

Computer simulations were carried out on the 3D crystallographic structures available in the Protein Data Bank [31] with ID numbers 5IHB, 5J06, 5J0B, 6D48, 6D49, 6D4A, and 7AW6. The first four included the N-terminal domain (aa19-135) and the repeated structural domain Ig of type C2 (aa145-228), while the last three structures included only the N-terminal domain. All PDB structures of CD33 (excluding 6D48) incorporated ligands. Characteristics of the seven human CD33 structures (Siglec-3) are summarized in Supplementary Table S1. The PDB files were edited by using Pymol [40]. Any molecules bound in the protein catalytic sites and water molecules bound to the protein during the crystallization process were removed. The CHARMM-GUI web-based platform (http://www.charmm-gui.org; accessed on 15 February 2022) was used to extract a single monomer from each of the CD33 structures contained in the PDB files, and to generate the two single-nucleotide polymorphisms (SNPs) at position 69 for each monomer. All generated structures were optimized offline using the CHARMM program [41] by performing a basic energy minimization with the steepest descent (SD) algorithm of the initial 25–50 steps to remove bad van der Waals contact. A more precise minimization was carried out subsequently with the Newton–Raphson method (ABNR) of 1,000,000 steps to remove potential problems, such as incorrect contacts, collisions, and nonphysical contacts/interactions. When the step-to-step energy change was less than 0.001 kcal/mol, the structure was considered sufficiently minimized.

5.4. Ligand Structure Editing and Minimization

The sialic acid and similar ligand (3′-sialyllactose, 6′-sialyllactose) 3D structures were generated using Avogadro software (http://avogadro.cc; accessed on 15 February 2022), and the structures were optimized through the MMFF94 force field with the steepest descent algorithm until the ΔE reach a value approaching 0 (<10–10).

5.5. Docking Analysis

The minimized monomeric structures of CD33 were used for the docking analysis of sialic acid and similar ligand structures. The docking analysis was carried out using Autodock Vina [42]. The ADT software package was used to determine the grid sizes, visualize the ligand poses, and add hydrogen atoms to the templates [43]. We superimposed the different CD33 structures and generated a single grid box that included all the CD33 structures. The coordinates of the grid box were put in a configuration file (config file in Supplementary Materials) used to perform both the docking and the virtual screening analysis. During the docking procedure, both the proteins and ligands were considered as rigid, and the only torsion that was allowed was for the carboxylic and alcoholic groups of the ligand. A bash shell command was used to pass receptors, ligands, and grid box parameters to Autodock Vina, producing an output file containing the predicted models (ligand poses) and a log file for each analysis. The structures were analyzed and the images were produced by using the PyMOL [40] and the UCSF Chimera 1.13.1 (http://www.cgl.ucsf.edu/chimera; accessed on 15 February 2022) [44] molecular graphics software programs.

5.6. Virtual Screening Analysis

We prepared two datasets of 100 models each of the Arg69 and Gly69 variants of the CD33 structure (PDB ID 5IHB). We performed a basic energy minimization using the CHARMM program [41], as previously described, to make them more relaxed for subsequent analysis. Using an ad hoc script (see Supplementary Materials) to transfer receptors and grid box parameters for automating the docking process, we submitted the two structure datasets (200 structures total) for analysis by Autodock Vina against the sialic acid as ligand, retrieving at least eight poses for each CD33 structure.

5.7. Statistical Analysis

Data were analyzed using the SPSS (v. 17.0) statistical software package (SPSS Inc., Chicago, IL, USA). The frequencies of genotypes and allelotypes of the CD33 gene polymorphism were calculated and determined for deviation from the Hardy–Weinberg equilibrium (HWE). The Chi-squared test and odds ratio (OR) were used to evaluate the association of AD risk with different genotypes and alleles. A *p*-value of <0.05 was considered statistically significant.

Author Contributions: F.T. and F.F. performed the molecular modeling and simulation; A.R. performed the in vitro experiments; A.A. and A.D.C. performed the collection of data from patients; A.R., A.D. and E.V. performed the analysis and comparison of experimental data; F.F. performed the analysis of in silico simulations; F.F. and E.V. conceived and designed the research; A.D., F.A., A.D.C., F.F. and E.V. interpreted the results; F.T., A.D., F.F. and E.V. wrote the manuscript; all authors edited the manuscript. All authors have read and agreed to the published version of the manuscript.

References

1. Alzheimer's Disease International. World Alzheimer Report 2019: Attitudes to Dementia. Available online: https://www.alzint.org/u/WorldAlzheimerReport2019.pdf (accessed on 1 September 2019).
2. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **2022**, *7*, e105–e125. [CrossRef]
3. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. *Alzheimer's Dement.* **2019**, *15*, 321–387. [CrossRef]
4. Winblad, B.; Amouyel, P.; Andrieu, S.; Ballard, C.; Brayne, C.; Brodaty, H.; Cedazo-Minguez, A.; Dubois, B.; Edvardsson, D.; Feldman, H.; et al. Defeating Alzheimer's disease and other dementias: A priority for European science and society. *Lancet Neurol.* **2016**, *15*, 455–532. [CrossRef]
5. Zhu, X.-C.; Tan, L.; Wang, H.-F.; Jiang, T.; Cao, L.; Wang, C.; Wang, J.; Tan, C.-C.; Meng, X.-F.; Yu, J.-T. Rate of early onset Alzheimer's disease: A systematic review and meta-analysis. *Ann. Transl. Med.* **2015**, *3*, 38, Erratum in *Ann. Transl. Med.* **2016**, *4*, E4. [CrossRef]
6. Cacace, R.; Sleegers, K.; Van Broeckhoven, C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimer's Dement.* **2016**, *12*, 733–748. [CrossRef]
7. Kamboh, M.I. Genomics and Functional Genomics of Alzheimer's Disease. *Neurotherapeutics* **2021**, 1–21. [CrossRef]
8. Kunkle, B.W.; Grenier-Boley, B.; Sims, R.; Bis, J.C.; Damotte, V.; Naj, A.C.; Boland, A.; Vronskaya, M.; Van Der Lee, S.J.; Amlie-Wolf, A.; et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **2019**, *51*, 414–430, Erratum in *Nat. Genet.* **2019**, *51*, 1423–1424. [CrossRef]
9. Wightman, D.P.; Jansen, I.E.; Savage, J.E.; Shadrin, A.A.; Bahrami, S.; Holland, D.; Rongve, A.; Børte, S.; Winsvold, B.S.; Drange, O.K.; et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **2021**, *53*, 1276–1282, Erratum in *Nat. Genet.* **2021**, *53*, 1722. [CrossRef]
10. Bertram, L.; Lange, C.; Mullin, K.; Parkinson, M.; Hsiao, M.; Hogan, M.F.; Schjeide, B.M.; Hooli, B.; DiVito, J.; Ionita, I.; et al. Genome-wide Association Analysis Reveals Putative Alzheimer's Disease Susceptibility Loci in Addition to APOE. *Am. J. Hum. Genet.* **2008**, *83*, 623–632. [CrossRef]
11. Hollingworth, P.; Harold, D.; Sims, R.; Gerrish, A.; Lambert, J.-C.; Carrasquillo, M.M.; Abraham, R.; Hamshere, M.L.; Pahwa, J.S.; Moskvina, V.; et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* **2011**, *43*, 429–435. [CrossRef]
12. Griciuc, A.; Serrano-Pozo, A.; Parrado, A.R.; Lesinski, A.N.; Asselin, C.N.; Mullin, K.; Hooli, B.; Choi, S.H.; Hyman, B.T.; Tanzi, R.E. Alzheimer's Disease Risk Gene CD33 Inhibits Microglial Uptake of Amyloid Beta. *Neuron* **2013**, *78*, 631–643. [CrossRef]
13. Malik, M.; Simpson, J.F.; Parikh, I.; Wilfred, B.R.; Fardo, D.W.; Nelson, P.T.; Estus, S. CD33 Alzheimer's Risk-Altering Polymorphism, CD33 Expression, and Exon 2 Splicing. *J. Neurosci.* **2013**, *33*, 13320–13325. [CrossRef]
14. Naj, A.C.; Jun, G.; Beecham, G.W.; Wang, L.-S.; Vardarajan, B.N.; Buros, J.; Gallins, P.J.; Buxbaum, J.D.; Jarvik, G.P.; Crane, P.K.; et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* **2011**, *43*, 436–441. [CrossRef]
15. Bao, J.; Wang, X.-J.; Mao, Z.-F. Associations Between Genetic Variants in 19p13 and 19q13 Regions and Susceptibility to Alzheimer Disease: A Meta-Analysis. *Med. Sci. Monit.* **2016**, *22*, 234–243. [CrossRef]
16. Duan, S.; Paulson, J.C. Siglecs as Immune Cell Checkpoints in Disease. *Annu. Rev. Immunol.* **2020**, *38*, 365–395. [CrossRef]
17. Jiang, T.; Yu, J.-T.; Hu, N.; Tan, M.-S.; Zhu, X.-C.; Tan, L. CD33 in Alzheimer's Disease. *Mol. Neurobiol.* **2014**, *49*, 529–535. [CrossRef]
18. Zhao, L. CD33 in Alzheimer's Disease—Biology, Pathogenesis, and Therapeutics: A Mini-Review. *Gerontology* **2019**, *65*, 323–331. [CrossRef]
19. Griciuc, A.; Patel, S.; Federico, A.N.; Choi, S.H.; Innes, B.J.; Oram, M.K.; Cereghetti, G.; McGinty, D.; Anselmo, A.; Sadreyev, R.I.; et al. TREM2 Acts Downstream of CD33 in Modulating Microglial Pathology in Alzheimer's Disease. *Neuron* **2019**, *103*, 820–835.e7. [CrossRef]
20. Ulland, T.K.; Colonna, M. TREM2—A key player in microglial biology and Alzheimer disease. *Nat. Rev. Neurol.* **2018**, *14*, 667–675. [CrossRef]
21. Neumann, H.; Daly, M.J. Variant TREM2 as Risk Factor for Alzheimer's Disease. *N. Engl. J. Med.* **2013**, *368*, 182–184. [CrossRef]

22. Linnartz, B.; Wang, Y.; Neumann, H. Microglial Immunoreceptor Tyrosine-Based Activation and Inhibition Motif Signaling in Neuroinflammation. *Int. J. Alzheimer's Dis.* **2010**, *2010*, 587463. [CrossRef]

23. Wißfeld, J.; Nozaki, I.; Mathews, M.; Raschka, T.; Ebeling, C.; Hornung, V.; Brüstle, O.; Neumann, H. Deletion of Alzheimer's disease-associated CD33 results in an inflammatory human microglia phenotype. *Glia* **2021**, *69*, 1393–1412. [CrossRef]

24. Griciuc, A.; Federico, A.N.; Natasan, J.; Forte, A.M.; McGinty, D.; Nguyen, H.; Volak, A.; LeRoy, S.; Gandhi, S.; Lerner, E.P.; et al. Gene therapy for Alzheimer's disease targeting CD33 reduces amyloid beta accumulation and neuroinflammation. *Hum. Mol. Genet.* **2020**, *29*, 2920–2935. [CrossRef]

25. Efthymiou, A.G.; Goate, A.M. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol. Neurodegener.* **2017**, *12*, 43. [CrossRef]

26. Estus, S.; Shaw, B.C.; Devanney, N.; Katsumata, Y.; Press, E.E.; Fardo, D.W. Evaluation of CD33 as a genetic risk factor for Alzheimer's disease. *Acta Neuropathol.* **2019**, *138*, 187–199. [CrossRef]

27. Bhattacherjee, A.; Jung, J.; Zia, S.; Ho, M.; Eskandari-Sedighi, G.; St. Laurent, C.D.; McCord, K.A.; Bains, A.; Sidhu, G.; Sarkar, S.; et al. The CD33 short isoform is a gain-of-function variant that enhances Aβ1–42 phagocytosis in microglia. *Mol. Neurodegener.* **2021**, *16*, 19. [CrossRef]

28. Raj, T.; Ryan, K.J.; Replogle, J.M.; Chibnik, L.B.; Rosenkrantz, L.; Tang, A.; Rothamel, K.; Stranger, B.E.; Bennett, D.A.; Evans, D.A.; et al. CD33: Increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Hum. Mol. Genet.* **2014**, *23*, 2729–2736. [CrossRef]

29. Rendina, A.; Drongitis, D.; Donizetti, A.; Fucci, L.; Milan, G.; Tripodi, F.; Giustezza, F.; Postiglione, A.; Pappatà, S.; Ferrari, R.; et al. CD33 and SIGLECL1 Immunoglobulin Superfamily Involved in Dementia. *J. Neuropathol. Exp. Neurol.* **2020**, *79*, 891–901. [CrossRef]

30. Nettiksimmons, J.; Tranah, G.; Evans, D.S.; Yokoyama, J.S.; Yaffe, K. Gene-based aggregate SNP associations between candidate AD genes and cognitive decline. *Age* **2016**, *38*, 41. [CrossRef]

31. Burley, S.K.; Berman, H.M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Dalenberg, K.; Duarte, J.M.; Dutta, S.; et al. RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* **2019**, *47*, D464–D474. [CrossRef]

32. Miles, L.A.; Hermans, S.J.; Crespi, G.A.; Gooi, J.; Doughty, L.; Nero, T.L.; Markulić, J.; Ebneth, A.; Wroblowski, B.; Oehlrich, D.; et al. Small Molecule Binding to Alzheimer Risk Factor CD33 Promotes Aβ Phagocytosis. *iScience* **2019**, *19*, 110–118. [CrossRef] [PubMed]

33. Crocker, P.R.; Paulson, J.C.; Varki, A. Siglecs and their roles in the immune system. *Nat. Rev. Immunol.* **2007**, *7*, 255–266. [CrossRef] [PubMed]

34. Avril, T.; Floyd, H.; Lopez, F.; Vivier, E.; Crocker, P. The Membrane-Proximal Immunoreceptor Tyrosine-Based Inhibitory Motif Is Critical for the Inhibitory Signaling Mediated by Siglecs-7 and -9, CD33-Related Siglecs Expressed on Human Monocytes and NK Cells. *J. Immunol.* **2004**, *173*, 6841–6849. [CrossRef] [PubMed]

35. Varki, A.; Angata, T. Siglecs—The major subfamily of I-type lectins. *Glycobiology* **2006**, *16*, 1R–27R. [CrossRef]

36. Yamakawa, N.; Yasuda, Y.; Yoshimura, A.; Goshima, A.; Crocker, P.R.; Vergoten, G.; Nishiura, Y.; Takahashi, T.; Hanashima, S.; Matsumoto, K.; et al. Discovery of a new sialic acid binding region that regulates Siglec-7. *Sci. Rep.* **2020**, *10*, 8647. [CrossRef]

37. Olaru, O.G.; Constantin, G.I.; Pena, C.M. Variation of total serum sialic acid concentration in postmenopausal women. *Exp. Ther. Med.* **2020**, *20*, 2455–2459. [CrossRef]

38. Mosconi, L.; Rahman, A.; Diaz, I.; Wu, X.; Scheyer, O.; Hristov, H.W.; Vallabhajosula, S.; Isaacson, R.S.; De Leon, M.J.; Brinton, R.D. Increased Alzheimer's risk during the menopause transition: A 3-year longitudinal brain imaging study. *PLoS ONE* **2018**, *13*, e0207885. [CrossRef]

39. McKhann, G.M.; Knopman, D.S.; Chertkow, H.; Hyman, B.T.; Jack, C.R., Jr.; Kawas, C.H.; Klunk, W.E.; Koroshetz, W.J.; Manly, J.J.; Mayeux, R.; et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **2011**, *7*, 263–269. [CrossRef]

40. The PyMOL Molecular Graphics System, version 1.8; New York (USA) Schrödinger, LLC: 2015. Available online: Pymol.org/2/support.html (accessed on 15 February 2022).

41. Brooks, B.R.; Brooks, C.L., III; MacKerell, A.D., Jr.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [CrossRef]

42. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]

43. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [CrossRef]

44. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef]