FEATURED ARTICLE

INVESTIGATION

G3 Genes | Genomes | Genetics

Elucidating the Molecular Determinants of Aβ Aggregation with Deep Mutational Scanning

Vanessa E. Gray,* Katherine Sitko,* Floriane Z. Ngako Kameni,* Miriam Williamson,* Jason J. Stephany,* Nicholas Hasle,* and Douglas M. Fowler*,†,‡

*Department of Genome Sciences, †Department of Bioengineering, University of Washington, Seattle, WA, and ‡Genetic Networks Program, CIFAR, Toronto, ON, Canada

ORCID IDs: 0000-0001-8007-4891 (V.E.G.); 0000-0001-7614-1713 (D.M.F.)

ABSTRACT Despite the importance of Aβ aggregation in Alzheimer's disease etiology, our understanding of the sequence determinants of aggregation is sparse and largely derived from *in vitro* studies. For example, *in vitro* proline and alanine scanning mutagenesis of $A\beta_{40}$ proposed core regions important for aggregation. However, we lack even this limited mutagenesis data for the more disease-relevant $A\beta_{42}$. Thus, to better understand the molecular determinants of $A\beta_{42}$ aggregation in a cell-based system, we combined a yeast DHFR aggregation assay with deep mutational scanning. We measured the effect of 791 of the 798 possible single amino acid substitutions on the aggregation propensity of $A\beta_{42}$. We found that ~75% of substitutions, largely to hydrophobic residues, maintained or increased aggregation. We identified 11 positions at which substitutions, particularly to hydrophilic and charged amino acids, disrupted Aβ aggregation. These critical positions were similar but not identical to critical positions identified in previous Aβ mutagenesis studies. Finally, we analyzed our large-scale mutagenesis data in the context of different Aβ aggregate structural models, finding that the mutagenesis data agreed best with models derived from fibrils seeded using brain-derived Aβ aggregates.

KEYWORDS
Amyloid
Amyloid beta
Deep mutational scanning
Protein aggregation

Protein aggregation affects all known organisms from bacteria to humans and is implicated in a number of human diseases. Decades of genetic, biochemical and epidemiological work suggests that aggregation of the amyloid β (Aβ) peptide is related to the incurable neurodegeneration associated with Alzheimer's disease (Hardy and Selkoe 2002; Lesné *et al.* 2008; Bertram and Tanzi 2008; Shankar *et al.* 2009; Masters and Selkoe 2012; Hardy 2017). Aβ peptide is generated by post-translational cleavage of the transmembrane amyloid β precursor protein at variable positions to produce peptides that range from 38 to 43 amino acids in length. The most aggregation-prone form of Aβ is 42 amino acids long ($A\beta_{42}$), though $A\beta_{40}$ is present at higher concentrations in human cerebrospinal fluid (Jarrett *et al.* 1993; Iwatsubo *et al.* 1994; Dahlgren *et al.* 2002). The aggregation of Aβ begins with a shift in equilibrium from soluble monomers to oligomers, and these oligomers may nucleate amyloidogenesis (Matsumura *et al.* 2011; Barz *et al.* 2018). In Alzheimer's disease, Aβ fibrils accumulate in the extracellular space forming the major component of amyloid plaques, a defining feature of the disease.

Despite the importance of Aβ aggregation in Alzheimer's disease etiology, our understanding of the sequence determinants of aggregation is sparse and largely derived from *in vitro* studies. In the past decade, several assays based on the budding yeast *S. cerevisiae* have been used to study protein aggregation (Bagriantsev and Liebman 2006; Haar *et al.* 2007; Caine *et al.* 2007; Morell *et al.* 2011; D'Angelo *et al.* 2013). Notably, a growth-based assay that separates toxicity from aggregation offers a way to investigate how changes in Aβ sequence impact aggregation propensity (Morell *et al.* 2011) (Figure 1A). In this assay, Aβ is cytoplasmically localized to eliminate its aggregation-associated toxicity (Treusch *et al.* 2011; D'Angelo *et al.* 2013). To link Aβ aggregation to yeast growth, Aβ is fused to an essential protein, dihydrofolate reductase (DHFR) via a short peptide linker. The result is that DHFR activity depends on the solubility of Aβ. Thus, upon

doi: https://doi.org/10.1534/g3.119.400535
Manuscript received July 10, 2019; accepted for publication September 8, 2019; published Early Online September 26, 2019.

Supplemental material available at FigShare: https://doi.org/10.6084/m9.figshare.8330297.
[1]Corresponding author: University of Washington, William H. Foege Hall, S041, 3720 15th Ave NE, Seattle, WA 98195, E-mail: dfowler@uw.edu

treatment with the competitive DHFR inhibitor methotrexate, yeast expressing soluble Aβ variants grow rapidly, whereas yeast expressing aggregating Aβ variants grow slowly.

Mutagenesis can elucidate the role of individual residues in protein aggregation. For example, *in vitro* proline (Williams *et al.* 2004) and alanine (Williams, Shivaprasad, and Wetzel 2006) scanning mutagenesis of $A\beta_{40}$ revealed core regions important for aggregation. However, we lack even this limited mutagenesis data for the more disease-relevant $A\beta_{42}$ and, so far, the majority of mutagenesis studies have been performed *in vitro*.

Thus, to fully understand the molecular determinants of $A\beta_{42}$ aggregation in a cell-based system, we combined the yeast growth-based aggregation assay with deep mutational scanning (Araya and Fowler 2011; Fowler and Fields 2014; Fowler *et al.* 2014) to measure the effect of 791 of the possible 798 single amino acid substitution on the aggregation propensity of $A\beta_{42}$. We used high-throughput DNA sequencing to track the frequency of each $A\beta_{42}$ variant during the selection, enabling us to assign a solubility score to every variant. We present the first large-scale, cell-based mutational analysis of Aβ, illuminating the physicochemical properties of amino acids that abrogate, promote or do not affect Aβ aggregation. Of 791 single amino acid Aβ variants we evaluated, ~75% maintained or increased aggregation. In addition, we identified 11 positions at which substitutions, particularly to hydrophilic and charged amino acids, disrupted Aβ aggregation. These critical positions were similar but not identical to critical positions identified in previous Aβ mutagenesis studies. Finally, we analyzed our large-scale mutagenesis data in the context of different Aβ aggregate structural models, finding that some structures were plausible whereas others were not.

METHODS

Library construction

The library was cloned using *in vivo assembly* (García-Nafría *et al.* 2016). First, a forward primer containing a 5′ homology region, an NNK codon, and a 3′ extension region was designed for each codon in $A\beta_{42}$ (Table S1). The homology and extension regions were at least 15 nucleotides in length and had melting temperatures greater than 55C. Reverse primers were the reverse complement of the 5′ homology region.

A separate PCR reaction was performed for each codon. These reactions contained 40 ng template (p416GAL1-Aβ-DHFR) and 10 μM forward and reverse primers (IDT, custom oligos) in a total reaction volume of 30 μL. The following cycling conditions were used: 95C 3min, 8x [98C 20 sec, 60C 15 sec, 72C 9 min], 72C 9 min. After PCR, 7.5 μL of each product was run on a 1.5% agarose gel for 30 min at 100V to check for a single product. The remaining 22.5 μL aliquots of product were each digested for an hour at 37C with 0.6 μL of DpnI (NEB, R0176S). After digestion, 4 μL of each linear product was transformed into a 50 μL of TOP10F Chemically Competent *E. coli* (ThermoFisher, C303003) according to manufacturer's instructions, with the following modifications: the protocol was done in a 96 well plate, and cells were recovered in a total volume of 200 μL SOC. After recovery, cells were transferred to a deep well plate with 1.6-1.8 mL of ampicillin LB and shaken overnight. To estimate colony count, 50 μL of culture was plated on an LB + ampicillin agar plate. Deep well plates and agar plates were incubated at 37C overnight. After incubation, all 42 deep well plate cultures were combined and subject to Midiprep (Sigma, NA0200).

4.3.2 Plasmids, yeast strains and growth conditions

To create a galactose-regulated Aβ-DHFR expression system, we directionally cloned DHFR into p416 (URA3, GAL1 promoter, CEN) using Blp and SpeI and then cloned the human $A\beta_{42}$ coding sequence into the SpeI and HindIII of the same vector. Aβ-GFP variants were cloned using the same scheme. All Aβ variants were cloned into p416 and transformed in W303 strain (MATa/MATα {leu2-3,112 trp1-1 can1- 100 ura3-1 ade2-1 his3-11,15} [phi+]). Cells were grown at 30C in synthetic complete (SC) media lacking uracil and supplemented with 2% glucose.

Methotrexate selection assay

Transformed yeast were inoculated into 5 mL (low-throughput) or 300 mL (co-culture and deep mutational scan) of C-Ura, 2% glucose media, grown in a rotating/shaking, 30C incubator overnight and then transferred to 5 mL or 300 mL 2% raffinose media to remove the glucose repression acting on the *gal1* promoter. After two hours in 2% raffinose, yeast were back-diluted to an OD of 0.01 into 5 mL or 300 mL 2% galactose to induce $A\beta_{42}$-DHFR expression in the presence or absence of 80 μM methotrexate (TCI America, M-1664) and 1 mM sulfanilamide (Sigma, S-9251). In 5 mL experiments, yeast growth was measured over 48h using a spectrophotometer that detects 660 nm wavelengths. The following equation was used to calculate doubling times from two time points: $(\mathrm{Log}_{10}(OD_{T2}/OD_{T1})/\mathrm{Log}_{10}(2))/\Delta T$, where OD represents the optical density at 600nm at a time point (T). For co-culture experiments, yeast with aggregating and nonaggregating variants were inoculated at equal densities in 300 mL. Ten OD units of yeast were collected from 300 mL cultures every 12h, spun down, concentrated and stored in -80C. At the end of the experiment, frozen yeast were thawed and then their plasmids were extracted using a DNA Clean and Concentrator kit (Zymo Research, D4013). Extracted plasmids were prepped and sequenced using Sanger sequencing. For the deep mutational scan, 300 mL cultures were sampled at the following timepoints: input, 28h (OD ≈ 1.0), 31.5h (OD ≈ 2.0), 35h (OD ≈ 3.0), 38h (OD ≈ 4.5), and 40h (OD ≈ 6.0). Cultures were spun down, concentrated and stored in -80C. Plasmids were extracted from yeast with Yeast Plasmid Miniprep 1 kit (Zymo Research, D-2001). Library fragments were amplified in 17 PCR cycles using primers specific to DNA sequences that flank Aβ-DHFR in p416, and sequenced by an Illumina NextSeq sequencer using paired-end reads (Table S1).

Variant effect analysis

Enrich2 was used to calculate solubility scores for each Aβ variant from sequencing fastq files (Rubin *et al.* 2017). The Enrich2 pipeline calculates a variant's score in three steps. First, a variant's normalized frequency ratios are tabulated for each timepoint by dividing the frequency of a variant's sequencing reads by all mapped reads and normalizing by the wild-type frequency ratio. Sequencing reads were stringently filtered for quality; we require each base have a Phred score greater than 20 and no uncalled bases. Second, a weighted linear least squares regression line is fit to the normalized frequency ratios across time points. Third, the slope of the regression line is calculated, averaged across the three replicates and $\log_2$ transformed. This averaged $\log_2$ slope reflects a variant's aggregation propensity. Solubility scores below 0 denote variants that are more aggregation-prone than wild-type, whereas scores above 0 indicate that a variant has increased solubility compared to wild-type.

Classifying Aβ variants using synonymous mutations

Variant classifications (*i.e.*, WT-like, more aggregation-prone, more soluble) were assigned using the distribution of 39 synonymous mutations from the deep mutational scan. We define WT-like as any variant

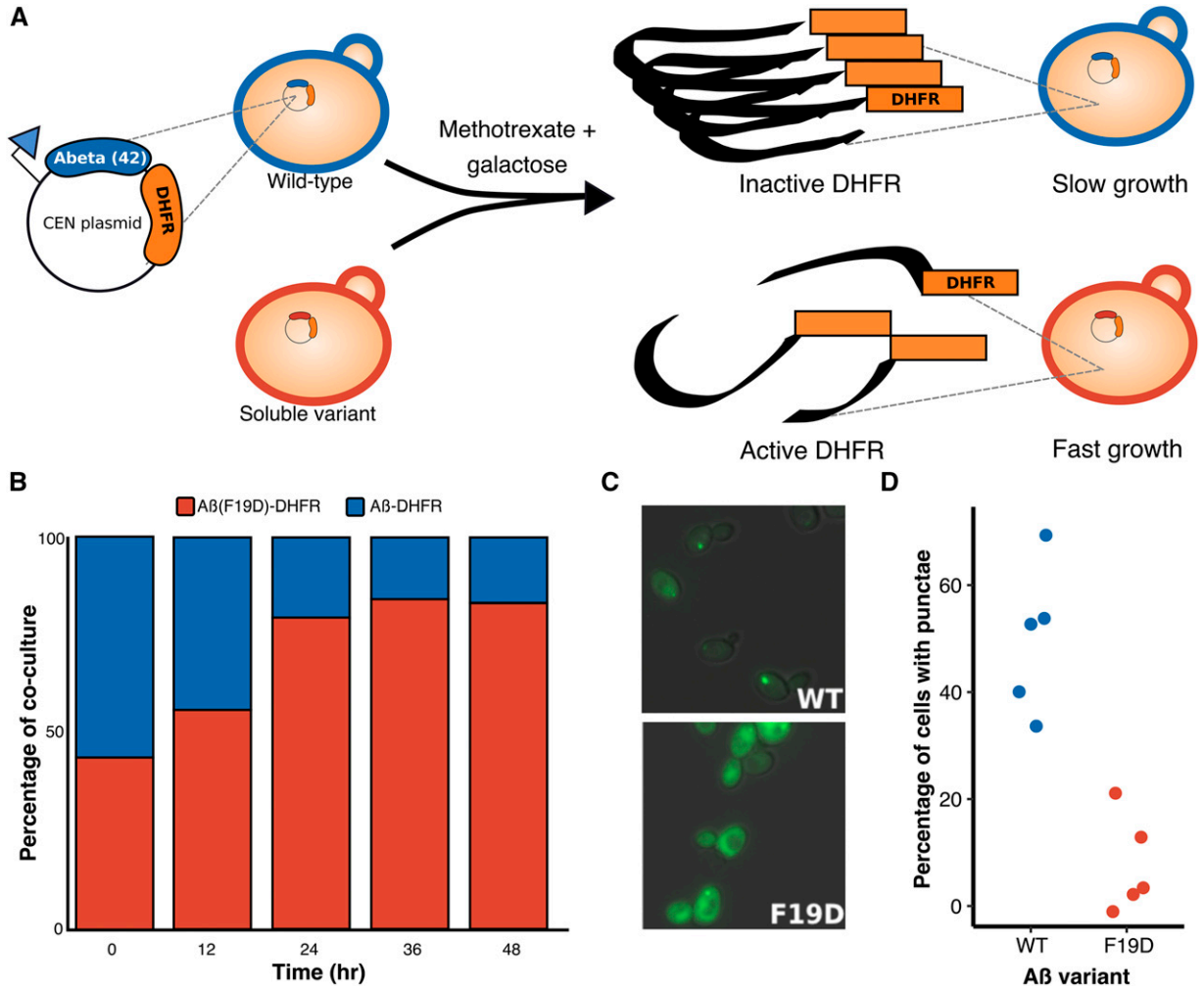

Figure 1 A-D. A yeast-based aggregation assay distinguishes between soluble and aggregation-prone variants of Aβ. A schematic of the assay shows plasmid-based expression of Aβ-DHFR and a nonaggregating variant of Aβ fused to DHFR, which lead to slow and fast yeast growth in the presence of methotrexate, respectively (A). A stacked bar graph shows the percentage of Aβ-DHFR and $A\beta^{19}_{FD}$-DHFR in co-culture (y-axis) every 12 hr for 48 hr (x-axis; B). Fluorescence light microscopy shows the aggregation patterns of Aβ-GFP (WT) and $A\beta^{19}_{FD}$-GFP (^{19}FD) 16h after induction of expression (C). A bar graph shows the percentage of yeast cells with punctae (y-axis) in five fluorescence microscopy images of Aβ-DHFR (WT) or $A\beta^{19}_{FD}$-DHFR (^{19}FD; x-axis; D).

with a score within ± 2 SD of the synonymous variant mean [-0.26,0.39]. A variant is more-aggregation prone than wildtype if its score is greater than 0.39 or more soluble if its score is lower than -0.26.

Data and code availability

Raw sequencing data is available in the NCBI GEO database (accession number GSE139122). Code and variant scores are available at https://github.com/FowlerLab/amyloidBeta2019. Supplemental material available at FigShare: https://doi.org/10.6084/m9.figshare.8330297.

RESULTS

First, we verified that the DHFR-based yeast aggregation assay could differentiate between aggregating wild type Aβ ($A\beta_{WT}$) and a nonaggregating ($A\beta^{19}_{FD}$) variant (Morell *et al.* 2011). As expected, in a mixed culture treated with methotrexate, $A\beta^{19}_{FD}$ outcompeted $A\beta_{WT}$ (Figure 1B). We used fluorescence microscopy of Aβ-GFP fusions to confirm that ~30–70% of yeast expressing $A\beta_{WT}$-GFP had cytoplasmic punctae compared to ~0–20% of cells expressing $A\beta^{19}_{FD}$-GFP across five fields of view (Figure 1C-D). Thus, we concluded the assay could be used in a deep mutational scan to measure the aggregation propensity of variants of Aβ.

Using this assay, we conducted a deep mutational scan of Aβ that yielded solubility scores for 791 single amino acid variants, representing 99.1% of the possible single variants. Solubility scores were calculated by taking the weighted least squares slope of each variant's frequency ratios across six time points. (**see Methods**). The slopes from each replicate were well correlated (Pearson's R 0.78 to 0.92; Figure 2A, Figure S1A). Replicate slopes were averaged and $\log_2$ transformed to produce final solubility scores such that wild-type had a solubility score of zero (Table S2). Positive solubility scores indicated less aggregation and negative scores indicated increased aggregation.

Solubility scores ranged from -2.38 (most aggregating) to 1.45 (most soluble). The mean (median) solubility score for all variants was 0.09 (0.08), which was similar to the solubility scores of the 39 synonymous variants in our library (mean: 0.06; median: 0.08). Because we did not expect synonymous variants to affect aggregation propensity, we used their distribution of scores to identify WT-like variants (Figure 2B). In total, we found that 344 (43.4%) of Aβ variant scores were within two

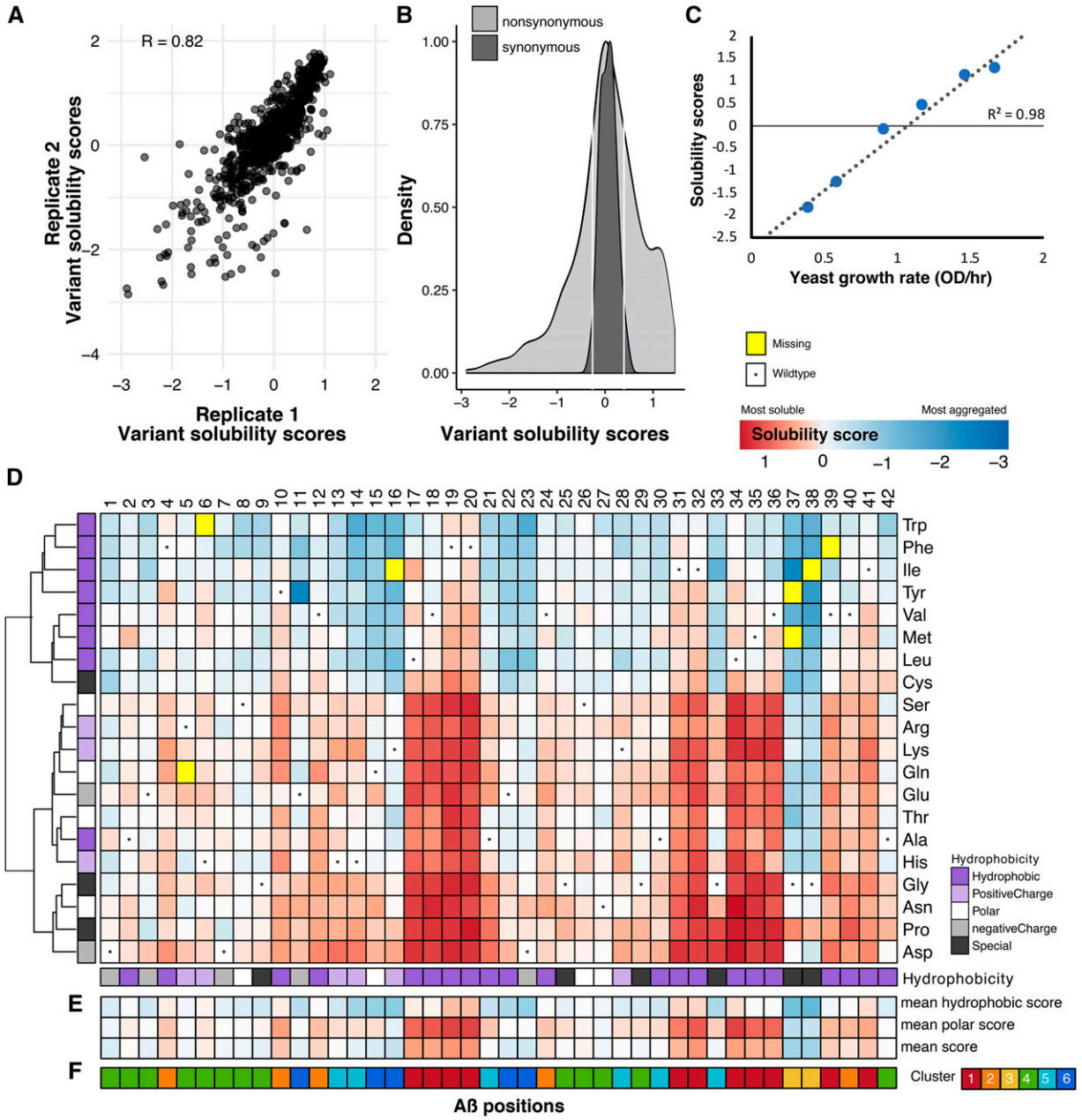

Figure 2 A-F. Solubility scores for 791 Aβ variants. Solubility scores reliably measure the effects of Aβ sequence on aggregation propensity. A scatter plot shows the correlation between two of three biological replicates that were averaged to yield final solubility scores (A; Figure S1A). The distribution of solubility scores (x-axis) of synonymous variants was used to determine cutoffs that define variants that are wild-type-like or more/less aggregation-prone than wild-type. The density plot shows distributions of nonsynonymous (light gray) and synonymous (dark gray) variants and the white lines show the lower (-0.26) and upper (0.39) bounds for wild-type-like variants (B). The scatterplot shows the correlation between our solubility scores (y-axis) and a low-throughput yeast growth assay that measured yeast growth rate as a proxy for Aβ solubility (C; Figure S1B). The heatmap shows the effect of 791 Aβ variants on solubility with Aβ positions on the x-axis and mutant amino acids on the y-axis. A variant's color denotes its solubility: red is most soluble, white is wild-type-like and, dark blue is most aggregated, whereas yellow variants are missing from our variant library and dots denote the wild-type amino acid at a given position. The annotation tracks on the x- and y-axes display the hydrophobicity of each wild-type and mutant amino acid, respectively. The heatmap's y-axis has been re-ordered using hierarchical clustering on the solubility score vectors (D). For each position, the mean solubility score at each position is depicted using the same color scheme as the main heatmap. Additionally, the mean solubility scores for all hydrophobic and polar substitutions are shown (E; Figure S2A). Heirarchical clustering on the x-axis yielded 6 distinct clusters: 1 (red), 2 (orange), 3 (yellow), 4 (green), 5 (light blue), and 6 (dark blue; F; Figure S2B-C).

standard deviations of the synonymous score mean and thus had WT-like effects (WT-like range: [-0.26,0.39]). Additionally, we found 246 (31.1%) variants to be more aggregation-prone than $A\beta_{WT}$ and 201 (25.4%) variants to be more soluble. Therefore, ~75% of Aβ variants maintained or increased the peptide's propensity to aggregate in yeast cells.

To verify that our deep mutational scan accurately measured variant effects on aggregation, we tested six Aβ variants, G38F, K16V, A42V, [19]

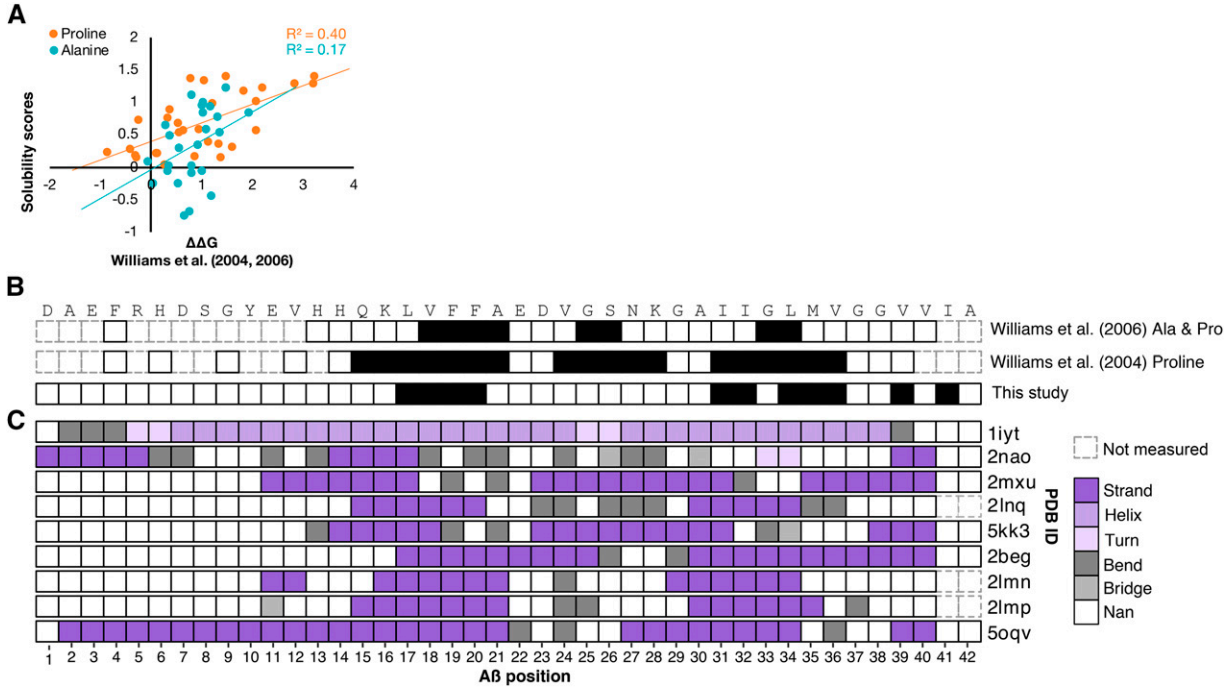

Figure 3 Comparison of yeast cell-based solubility scores to *in vitro* aggregation measurements and Aβ structural models. The scatterplot shows the correlation between our solubility scores (y-axis) and two single amino acid scans that measured the effect of proline (orange) or alanine (teal) variants on the thermodynamic stability of aggregates, relative to wild type (ΔΔG) (A; Figure S3). The first two tracks show unmeasured mutations (dashed gray) and the Aβ buried β-strand positions (black) suggested by proline scanning alone, or by proline and alanine scanning together (Williams *et al.* 2004; Williams, Shivaprasad, and Wetzel 2006). The third track shows positions with the greatest increase in solubility when mutated in our large-scale mutagenesis study, found in cluster 1 (B). The next nine tracks show the secondary structure of nine models of Aβ aggregate structure for each Aβ position (x-axis; C). The Aβ wild-type sequence is shown at the top.

FY, L17S and L34R, that spanned the solubility score range in a low-throughput validation assay. The growth rate of methotrexate-treated yeast expressing each Aβ variant was measured and compared to the aggregation propensity scores (Figure 2C, S1B). We found that low-throughput assay results strongly correlated with the solubility scores derived from deep mutational scanning (R^2 = 0.98). Thus, our deep mutational scan reliably measured Aβ variant aggregation propensity in the yeast assay.

To explore the effects of each amino acid substitution on Aβ aggregation, we created an Aβ sequence-aggregation map (Figure 2D). Substitutions to aspartic acid and proline were most associated with Aβ solubility, as evinced by their median scores of 0.64 and 0.56, respectively (Figure S2A). Conversely, the most aggregation-associated substitutions were hydrophobic tryptophan and phenylalanine, with scores of -0.60 and -0.51, respectively. Moreover, hierarchical clustering of all 791 solubility scores by amino acid revealed that hydrophobic substitutions, except alanine, clustered together and were associated with greater aggregation than other classes of substitutions.

Next, we characterized each position in Aβ based on its mutational profile. Hierarchical clustering of variant solubility scores by position identified six distinct clusters (Figures 2E-F; S2B-C). In cluster 1, comprising positions 17-20, 31-32, 34-36, 39 and 41, substitutions tended to decrease Aβ aggregation compared to substitutions in other clusters (cluster 1 mean solubility scores = 0.64, all other clusters = -0.28; Figure S2D). In cluster 1, even substitutions to hydrophobic amino acids slightly decreased aggregation (mean solubility score = 0.17). The effects of substitutions in cluster 2 were similar to but less extreme than in cluster 1. Both clusters 1 and 2 are largely comprised of hydrophobic positions in the wild type Aβ sequence. Indeed, 80% of Aβ positions with hydrophobic wild type residues are in clusters 1 and 2. In stark contrast, within clusters 4, 5 and 6, hydrophobic substitutions generally increase protein aggregation (all mean: -0.15, -0.12 and -0.45; hydrophobic means: -0.29, -0.65, and -1.04). Cluster 3 contains only two positions, 37 and 38. Here, every substitution except proline increased aggregation (all mean: -0.99, hydrophobic mean: -1.56). Given that cluster 1 is characterized by hydrophobic positions where hydrophilic substitutions profoundly decreased aggregation, we suggest that this cluster defined buried β-strands in the Aβ sequence.

Next, we compared our solubility scores to previous alanine and proline scans which reported $A\beta_{40}$ fibril thermodynamic stability *in vitro* (ΔΔG). ΔΔG values were determined by measuring variant Aβ monomer concentration remaining in solution after fibril formation reached equilibrium (Williams *et al.* 2004; Williams, Shivaprasad, and Wetzel 2006). We found that the effects of proline substitution in our assay were correlated with proline ΔΔG values (R^2 = 0.40), while the effects of alanine substitutions in our assay were less correlated with alanine ΔΔG values (R^2 = 0.17; Figure 3A). In our alanine and proline comparisons, we found the greatest correlation at positions 17-20 and 31-32, where substitutions decreased aggregation in all studies (Figure S3). The most notable disagreement between studies was for alanine substitutions at positions 37 and 38. In our assay, alanine substitutions caused a profound increase in aggregation, whereas the *in vitro* alanine scan showed the opposite effect.

We also compared our buried β-stand positions from cluster 1 to β-stands proposed based on the *in vitro* alanine and proline scans, finding some concordance (Figure 3B). The single amino acid scans identify three regions that disrupt fibril elongation thermodynamics when mutated. The regions include positions 15-21, 24-28, and

31-36 for the proline scan and positions 18–21, 25–26, and 32–33 for the combined alanine and proline scans (Williams *et al.* 2004; Williams, Shivaprasad, and Wetzel 2006). Given the generally highly disruptive nature of proline substitutions (Gray *et al.* 2017), it is not surprising that the proline scan would nominate many positions. Our deep mutational scan, on the other hand, does not reveal a central β-strand or strong decrease in aggregation with alanine or proline substitution from positions 24-28. We speculate that this difference is due either to the distinct experimental approaches used or to the different Aβ species ($Aβ_{40}$ *vs.* $Aβ_{42}$).

DISCUSSION

We used deep mutational scanning to characterize 791 Aβ variants in a yeast-based aggregation assay. Proline and aspartic acid substitutions were most disruptive of Aβ aggregation, while tryptophan and phenylalanine increased aggregation most. Additionally, we used unsupervised clustering to determine the regions of Aβ most important for aggregation. We conclude that these regions are most likely to form buried β-stands, which are necessary for aggregation and sensitive to amino acid substitutions (Jahn *et al.* 2010; Abrusán and Marsh 2016). These include positions 17-20, 31-32, 34-35, 39 and 41. While other positions could also form β-stands, the positions in cluster 1 are most likely to form the buried cores of Aβ aggregates in our cell-based assay.

Due to the noncrystalline nature of Aβ fibrils, traditional techniques such as X-ray crystallography and solution-state NMR cannot be used to solve Aβ's aggregate structure. Instead, structural models have been developed by amassing constraints, such as the direction and register of β-sheets. For example, solid-state nuclear magnetic resonance studies suggest that Aβ fibrils are parallel, in register β-sheets (Benzinger *et al.* 1998; Gregory *et al.* 1998; Antzutkin *et al.* 2002; Tycko 2011). Many of these structural models are problematic because they are generated from constraints derived from *in vitro* experimental data, which may not be representative of *in vivo* conditions.

Given that we collected large-scale mutagenesis data in a cell-based system, we examined how our results compared to structural models of Aβ fibrils. Some models such as 1IYT (Crescenzi *et al.* 2002) and 2NAO (Wälti *et al.* 2016), showed very little to no overlap with either our proposed buried β-strands or those proposed by Williams *et al.* (2004, 2006) (Figure 3C). Other models contained three β-strand regions reminiscent of those suggested by Williams *et al.* (2004, 2006): 2MXU (Xiao *et al.* 2015), 5KK3 (Colvin *et al.* 2016), and 5OQV (Gremer *et al.* 2017). Yet other models propose β-strand patterns more similar to ours. These include 2BEG (Lührs *et al.* 2005), 2LNQ (Gremer *et al.* 2017), 2LMP and 2LMN (Lu *et al.* 2013). Since our β-strand patterns were derived from data gathered in a cell-based assay, we hypothesized that they would be most consistent with structural models based on *in vivo*-derived fibrils. Indeed, the 2LMP and 2LMN models were based on fibrils seeded from plaques isolated from the brains of individuals afflicted by Alzheimer's disease. Moreover, every model besides 2LMP and 2LMN was constructed using NMR or cryo-EM data from laboratory grown fibrils. These models are less concordant with our cell-based mutational data, which suggests that there are important structural differences between *in vitro* and *in vivo* derived fibrils.

Two major differences exist between the experimental conditions used by Williams *et al.* (2004, 2006) and in our work, and may explain the difference in β-strands proposed in our respective *in vitro*- and *in vivo-derived* models. First, Williams *et al.* (2004, 2006) incubate Aβ in the absence of any other proteins, while our yeast-based system provides key players that affect protein aggregation, such as chaperone proteins and molecular crowding. Second, Williams *et al.* (2004, 2006) incubate Aβ peptides at 37C, whereas our yeast-based experiments required a lower temperature of 30C. This temperature difference may yield differences in folding kinetics. Further experiments are required to determine the contribution of these experimental differences to β-strand formation in Aβ.

Deep mutational scanning data could contribute to the investigation of Aβ fibril structure beyond the analysis of existing models we present. For example, others have used site-saturation mutagenesis and deep mutational scanning data to evaluate proposed structural models (Bajaj *et al.* 2008; Khare *et al.* 2019). Additionally, deep mutational scanning data have now been used to generate distance constraints for the prediction of tertiary protein structure (Schmiedel and Lehner 2018; Rollins *et al.* 2018).

In summary, we used deep mutational scanning to elucidate the effects of amino acid substitutions on Aβ aggregation in a cell-based model. We used these large-scale mutagenesis data to propose positions critical for Aβ aggregation. Our results conflict with some previous *in vitro* reports of the effects of substitutions on Aβ aggregation and with some models of Aβ fibril structure. This outcome highlights the difficulties of studying protein aggregation and emphasizes the potential utility of *in vivo* or cell-based models. We suggest that deep mutational scanning of other aggregation-prone proteins such as α-synuclein or transthyretin could help reveal the relationship between sequence, structure and aggregation.

ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences (grant R01GM109110 to D.M.F.) and the Alzheimer's Association. D.M.F. is a CIFAR Azrieli Global Scholar. V.E.G. was supported by a National Science Foundation Graduate Research Fellowship and a NHGRI Ruth L. Kirschstein National Research Service Award (T32HG000035). N.H. was supported by a NCI Ruth L. Kirschstein National Research Service Award (F30CA236335-01). D.M.F. conceived of the project. D.M.F., V.E.G. and K.S. designed experiments. N.H designed the Aβ library. V.E.G., K.S., F.N.K., J.S. executed experiments. V.E.G., M.W. and F.N.K. analyzed data. V.E.G. and D.M.F. wrote the manuscript.

LITERATURE CITED

Abrusán, G., and J. A. Marsh, 2016 Alpha Helices Are More Robust to Mutations than Beta Strands. PLOS Comput. Biol. 12: e1005242. https://doi.org/10.1371/journal.pcbi.1005242

Antzutkin, O. N., R. D. Leapman, J. J. Balbach, and R. Tycko, 2002 Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. Biochemistry 41: 15436–15450. https://doi.org/10.1021/bi0204185

Araya, C. L., and D. M. Fowler, 2011 Deep mutational scanning: assessing protein function on a massive scale. Trends Biotechnol. 29: 435–442. https://doi.org/10.1016/j.tibtech.2011.04.003

Bagriantsev, S., and S. Liebman, 2006 Modulation of Abeta42 low-n oligomerization using a novel yeast reporter system. BMC Biol. 4: 32. https://doi.org/10.1186/1741-7007-4-32

Bajaj, K., P. C. Dewan, P. Chakrabarti, D. Goswami, B. Barua *et al.*, 2008 Structural correlates of the temperature sensitive phenotype derived from saturation mutagenesis studies of CcdB. Biochemistry 47: 12964–12973. https://doi.org/10.1021/bi8014345

Barz, B., Q. Liao, and B. Strodel, 2018 Pathways of Amyloid-β Aggregation Depend on Oligomer Shape. J. Am. Chem. Soc. 140: 319–327. https://doi.org/10.1021/jacs.7b10343

Benzinger, T. L., D. M. Gregory, T. S. Burkoth, H. Miller-Auer, D. G. Lynn *et al.*, 1998 Propagating structure of Alzheimer's beta-amyloid(10–35) is parallel beta-sheet with residues in exact register. Proc. Natl. Acad. Sci. USA 95: 13407–13412. https://doi.org/10.1073/pnas.95.23.13407

Bertram, L., and R. E. Tanzi, 2008 Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat. Rev. Neurosci. 9: 768–778. https://doi.org/10.1038/nrn2494

Caine, J., S. Sankovich, H. Antony, L. Waddington, P. Macreadie *et al.*, 2007 Alzheimer's Abeta fused to green fluorescent protein induces growth stress and a heat shock response. FEMS Yeast Res. 7: 1230–1236. https://doi.org/10.1111/j.1567-1364.2007.00285.x

Colvin, M. T., R. Silvers, Q. Z. Ni, T. V. Can, I. Sergeyev *et al.*, 2016 Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. J. Am. Chem. Soc. 138: 9663–9674. https://doi.org/10.1021/jacs.6b05129

Crescenzi, O., S. Tomaselli, R. Guerrini, S. Salvadori, A. M. D'Ursi *et al.*, 2002 Solution structure of the Alzheimer amyloid beta-peptide (1–42) in an apolar microenvironment. Similarity with a virus fusion domain. Eur. J. Biochem. 269: 5642–5648. https://doi.org/10.1046/j.1432-1033.2002.03271.x

D'Angelo, F., H. Vignaud, J. Di Martino, B. Salin, A. Devin *et al.*, 2013 A yeast model for amyloid-β aggregation exemplifies the role of membrane trafficking and PICALM in cytotoxicity. Dis. Model. Mech. 6: 206–216. https://doi.org/10.1242/dmm.010108

Dahlgren, K. N., A. M. Manelli, W. B. Stine, L. K. Baker, G. A. Krafft *et al.*, 2002 Oligomeric and fibrillar species of amyloid-beta peptides differentially affect neuronal viability. J. Biol. Chem. 277: 32046–32053. https://doi.org/10.1074/jbc.M201750200

Fowler, D. M., and S. Fields, 2014 Deep mutational scanning: a new style of protein science. Nat. Methods 11: 801–807. https://doi.org/10.1038/nmeth.3027

Fowler, D. M., J. J. Stephany, and S. Fields, 2014 Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat. Protoc. 9: 2267–2284. https://doi.org/10.1038/nprot.2014.153

García-Nafría, J., J. F. Watson, and I. H. Greger, 2016 IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Sci. Rep. 6: 27459. https://doi.org/10.1038/srep27459

Gray, V. E., R. J. Hause, and D. M. Fowler, 2017 Analysis of Large-Scale Mutagenesis Data To Assess the Impact of Single Amino Acid Substitutions. Genetics 207: 53–61. https://doi.org/10.1534/genetics.117.300064

Gregory, D. M., T. L. Benzinger, T. S. Burkoth, H. Miller-Auer, D. G. Lynn *et al.*, 1998 Dipolar recoupling NMR of biomolecular self-assemblies: determining inter- and intrastrand distances in fibrilized Alzheimer's beta-amyloid peptide. Solid State Nucl. Magn. Reson. 13: 149–166. https://doi.org/10.1016/S0926-2040(98)00086-1

Gremer, L., D. Schölzel, C. Schenk, E. Reinartz, J. Labahn *et al.*, 2017 Fibril structure of amyloid-β(1–42) by cryo-electron microscopy. Science 358: 116–119. https://doi.org/10.1126/science.aao2825

Haar, von der, T., L. J. P. Wright, J. Zenthon, and M. F. Tuite, 2007 Development of a novel yeast cell-based system for studying the aggregation of Alzheimer's disease-associated Abeta peptides in vivo. Neurodegener Dis 4: 136–147.

Hardy, J., 2017 The discovery of Alzheimer-causing mutations in the APP gene and the formulation of the "amyloid cascade hypothesis". FEBS J. 284: 1040–1044. https://doi.org/10.1111/febs.14004

Hardy, J., and D. J. Selkoe, 2002 The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297: 353–356. https://doi.org/10.1126/science.1072994

Iwatsubo, T., A. Odaka, N. Suzuki, H. Mizusawa, N. Nukina *et al.*, 1994 Visualization of A beta 42(43) and A beta 40 in senile plaques with end-specific A beta monoclonals: evidence that an initially deposited species is A beta 42(43). Neuron 13: 45–53. https://doi.org/10.1016/0896-6273(94)90458-8

Jahn, T. R., O. S. Makin, K. L. Morris, K. E. Marshall, P. Tian *et al.*, 2010 The common architecture of cross-beta amyloid. J. Mol. Biol. 395: 717–727. https://doi.org/10.1016/j.jmb.2009.09.039

Jarrett, J. T., E. P. Berger, and P. T. Lansbury, 1993 The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. Biochemistry 32: 4693–4697. https://doi.org/10.1021/bi00069a001

Khare, S., M. Bhasin, A. Sahoo, and R. Varadarajan, 2019 Protein model discrimination attempts using mutational sensitivity, predicted secondary structure, and model quality information. Proteins 87: 326–336. https://doi.org/10.1002/prot.25654

Lesné, S., L. Kotilinek, and K. H. Ashe, 2008 Plaque-bearing mice with reduced levels of oligomeric amyloid-beta assemblies have intact memory function. Neuroscience 151: 745–749. https://doi.org/10.1016/j.neuroscience.2007.10.054

Lu, J.-X., W. Qiang, W.-M. Yau, C. D. Schwieters, S. C. Meredith *et al.*, 2013 Molecular structure of β-amyloid fibrils in Alzheimer's disease brain tissue. Cell 154: 1257–1268. https://doi.org/10.1016/j.cell.2013.08.035

Lührs, T., C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann *et al.*, 2005 3D structure of Alzheimer's amyloid-beta(1–42) fibrils. Proc. Natl. Acad. Sci. USA 102: 17342–17347. https://doi.org/10.1073/pnas.0506723102

Masters, C. L., and D. J. Selkoe, 2012 Biochemistry of amyloid β-protein and amyloid deposits in Alzheimer disease. Cold Spring Harb. Perspect. Med. 2: a006262. https://doi.org/10.1101/cshperspect.a006262

Matsumura, S., K. Shinoda, M. Yamada, S. Yokojima, M. Inoue *et al.*, 2011 Two distinct amyloid beta-protein (Abeta) assembly pathways leading to oligomers and fibrils identified by combined fluorescence correlation spectroscopy, morphology, and toxicity analyses. J. Biol. Chem. 286: 11555–11562. https://doi.org/10.1074/jbc.M110.181313

Morell, M., N. S. de Groot, J. Vendrell, F. X. Avilés, and S. Ventura, 2011 Linking amyloid protein aggregation and yeast survival. Mol. Biosyst. 7: 1121–1128. https://doi.org/10.1039/c0mb00297f

Rollins, N. J., K. P. Brock, F. J. Poelwijk, M. A. Stiffler, N. P. Gauthier *et al.*, 2019 Inferring protein 3D structure from deep mutation scans. Nat. Genet. 51: 1170–1176. https://doi.org/10.1038/s41588-019-0432-9

Rubin, A. F., H. Gelman, N. Lucas, S. M. Bajjalieh, A. T. Papenfuss *et al.*, 2017 A statistical framework for analyzing deep mutational scanning data. Genome Biol. 18: 150, 19: 17. https://doi.org/10.1186/s13059-017-1272-5

Schmiedel, J., and B. Lehner, 2019 Determining protein structures using genetics. Nat. Genet. 51: 1177–1186. https://doi.org/10.1038/s41588-019-0431-x

Shankar, G. M., M. A. Leissring, A. Adame, X. Sun, E. Spooner *et al.*, 2009 Biochemical and immunohistochemical analysis of an Alzheimer's disease mouse model reveals the presence of multiple cerebral Abeta assembly forms throughout life. Neurobiol. Dis. 36: 293–302. https://doi.org/10.1016/j.nbd.2009.07.021

Treusch, S., S. Hamamichi, J. L. Goodman, K. E. S. Matlack, C. Y. Chung *et al.*, 2011 Functional links between Aβ toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. Science 334: 1241–1245. https://doi.org/10.1126/science.1213210

Tycko, R., 2011 Solid-state NMR studies of amyloid fibril structure. Annu. Rev. Phys. Chem. 62: 279–299. https://doi.org/10.1146/annurev-physchem-032210-103539

Wälti, M. A., F. Ravotti, H. Arai, C. G. Glabe, J. S. Wall *et al.*, 2016 Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. Proc. Natl. Acad. Sci. USA 113: E4976–E4979. https://doi.org/10.1073/pnas.1600749113

Williams, A. D., E. Portelius, I. Kheterpal, J.-T. Guo, K. D. Cook *et al.*, 2004 Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. J. Mol. Biol. 335: 833–842. https://doi.org/10.1016/j.jmb.2003.11.008

Williams, A. D., S. Shivaprasad, and R. Wetzel, 2006 Alanine scanning mutagenesis of Abeta(1–40) amyloid fibril stability. J. Mol. Biol. 357: 1283–1294. https://doi.org/10.1016/j.jmb.2006.01.041

Xiao, Y., B. Ma, D. McElheny, S. Parthasarathy, F. Long *et al.*, 2015 Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. Nat. Struct. Mol. Biol. 22: 499–505. https://doi.org/10.1038/nsmb.2991

Communicating editor: B. Andrews