

1 G3

2

3

4 **Elucidating the molecular determinants of A β aggregation with deep mutational
5 scanning**

6

7

8 Vanessa E. Gray¹, Katherine Sitko¹, Floriane Z. Ngako Kameni¹, Miriam Williamson¹,
9 Jason J. Stephany¹, Nicholas Hasle¹, and Douglas M. Fowler^{1,2,3}

10

11 ¹Department of Genome Sciences, University of Washington, Seattle, WA, USA

12 ²Department of Bioengineering, University of Washington, Seattle, WA, USA

13 ³Genetic Networks Program, CIFAR, Toronto, ON, Canada

14

15

16 Correspondence to:

17 Douglas M. Fowler
18 University of Washington
19 William H. Foege Hall, S041
20 3720 15th Ave NE, Seattle, WA 98195
21 dfowler@uw.edu

22

23

24

25 **Abstract**

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

42

43

44 **Introduction**

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

60 Despite the importance of A β aggregation in Alzheimer's disease etiology, our
61 understanding of the sequence determinants of aggregation is sparse and largely
62 derived from *in vitro* studies. In the past decade, several assays based on the budding
63 yeast *S. cerevisiae* have been used to study protein aggregation (Bagriantsev and
64 Liebman 2006; Haar *et al.* 2007; Caine *et al.* 2007; Morell *et al.* 2011; D'Angelo *et al.*
65 2013). Notably, a growth-based assay that separates toxicity from aggregation offers a

66 way to investigate how changes in A β sequence impact aggregation propensity (Morell
67 *et al.* 2011) (**Figure 1A**). In this assay, A β is cytoplasmically localized to eliminate its
68 aggregation-associated toxicity (Treusch *et al.* 2011; D'Angelo *et al.* 2013). To link A β
69 aggregation to yeast growth, A β is fused to an essential protein, dihydrofolate reductase
70 (DHFR) via a short peptide linker. The result is that DHFR activity depends on the
71 solubility of A β . Thus, upon treatment with the competitive DHFR inhibitor methotrexate,
72 yeast expressing soluble A β variants grow rapidly, whereas yeast expressing
73 aggregating A β variants grow slowly.

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

80 Thus, to fully understand the molecular determinants of A β_{42} aggregation in a cell-based
81 system, we combined the yeast growth-based aggregation assay with deep mutational
82 scanning (Araya and Fowler 2011; Fowler and Fields 2014; Fowler *et al.* 2014) to
83 measure the effect of 791 of the possible 798 single amino acid substitution on the
84 aggregation propensity of A β_{42} . We used high-throughput DNA sequencing to track the
85 frequency of each A β_{42} variant during the selection, enabling us to assign a solubility
86 score to every variant. We present the first large-scale, cell-based mutational analysis of
87 A β , illuminating the physicochemical properties of amino acids that abrogate, promote

88 or do not affect A β aggregation. Of 791 single amino acid A β variants we evaluated,
89 ~75% maintained or increased aggregation. In addition, we identified 11 positions at
90 which substitutions, particularly to hydrophilic and charged amino acids, disrupted A β
91 aggregation. These critical positions were similar but not identical to critical positions
92 identified in previous A β mutagenesis studies. Finally, we analyzed our large-scale
93 mutagenesis data in the context of different A β aggregate structural models, finding that
94 some structures were plausible whereas others were not.

95

96 METHODS

97 Library construction

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

103 A separate PCR reaction was performed for each codon. These reactions contained 40
104 ng template (p416GAL1-A β -DHFR) and 10 μ M forward and reverse primers (IDT,
105 custom oligos) in a total reaction volume of 30 μ L. The following cycling conditions were
106 used: 95C 3min, 8x [98C 20 sec, 60C 15 sec, 72C 9 min], 72C 9 min. After PCR, 7.5
107 μ L of each product was run on a 1.5% agarose gel for 30 min at 100V to check for a
108 single product. The remaining 22.5 μ L aliquots of product were each digested for an

109 hour at 37C with 0.6 µL of DpnI (NEB, R0176S). After digestion, 4 µL of each linear
110 product was transformed into a 50 µL of TOP10F Chemically Competent *E. coli*
111 (ThermoFisher, C303003) according to manufacturer's instructions, with the following
112 modifications: the protocol was done in a 96 well plate, and cells were recovered in a
113 total volume of 200 µL SOC. After recovery, cells were transferred to a deep well plate
114 with 1.6-1.8 mL of ampicillin LB and shaken overnight. To estimate colony count, 50 µL
115 of culture was plated on an LB + ampicillin agar plate. Deep well plates and agar plates
116 were incubated at 37C overnight. After incubation, all 42 deep well plate cultures were
117 combined and subject to Midiprep (Sigma, NA0200).

118 *4.3.2 Plasmids, yeast strains and growth conditions*

119 To create a galactose-regulated A β -DHFR expression system, we directionally cloned
120 DHFR into p416 (URA3, GAL1 promoter, CEN) (Mumberg, Müller, & Funk, 1994) using
121 Blp and Spel and then cloned the human A β_{42} coding sequence into the Spel and
122 HindIII of the same vector. A β -GFP variants were cloned using the same scheme. All
123 A β variants were cloned into p416 and transformed in W303 strain (MAT α /MAT α {leu2-
124 3,112 trp1-1 can1- 100 ura3-1 ade2-1 his3-11,15} [phi+]). Cells were grown at 30C in
125 synthetic complete (SC) media lacking uracil and supplemented with 2% glucose.

126 *Methotrexate selection assay*

127 Transformed yeast were inoculated into 5 mL (low-throughput) or 300 mL (co-culture
128 and deep mutational scan) of C-Ura, 2% glucose media, grown in a rotating/shaking,
129 30C incubator overnight and then transferred to 5 mL or 300 mL 2% raffinose media to

130 remove the glucose repression acting on the *gal1* promoter. After two hours in 2%
131 raffinose, yeast were back-diluted to an OD of 0.01 into 5 mL or 300 mL 2% galactose
132 to induce A β ₄₂-DHFR expression in the presence or absence of 80 μ M methotrexate
133 (TCI America, M-1664) and 1 mM sulfanilamide (Sigma, S-9251). In 5 mL experiments,
134 yeast growth was measured over 48h using a spectrophotometer that detects 660 nm
135 wavelengths. The following equation was used to calculate doubling times from two time
136 points: $(\text{Log}_{10}(\text{OD}_{T_2}/\text{OD}_{T_1})/\text{Log}_{10}(2))/\Delta T$, where OD represents the optical density at
137 600nm at a time point (T). For co-culture experiments, yeast with aggregating and
138 nonaggregating variants were inoculated at equal densities in 300 mL. Ten OD units of
139 yeast were collected from 300 mL cultures every 12h, spun down, concentrated and
140 stored in -80C. At the end of the experiment, frozen yeast were thawed and then their
141 plasmids were extracted using a DNA Clean and Concentrator kit (Zymo Research,
142 D4013). Extracted plasmids were prepped and sequenced using Sanger sequencing.
143 For the deep mutational scan, 300 mL cultures were sampled at the following
144 timepoints: input, 28h (OD H 1.0), 31.5h (OD H 2.0), 35h (OD H 3.0), 38h (OD H 4.5),
145 and 40h (OD H 6.0). Cultures were spun down, concentrated and stored in -80C.
146 Plasmids were extracted from yeast with Yeast Plasmid Miniprep 1 kit (Zymo Research,
147 D-2001). Library fragments were amplified in 17 PCR cycles using primers specific to
148 DNA sequences that flank A β -DHFR in p416, and sequenced by an Illumina NextSeq
149 sequencer using paired-end reads (**Table S1**).
150 *Variant effect analysis*

151 Enrich2 was used to calculate solubility scores for each A β variant from sequencing
152 fastq files (Rubin *et al.* 2017). The Enrich2 pipeline calculates a variant's score in three
153 steps. First, a variant's normalized frequency ratios are tabulated for each timepoint by
154 dividing the frequency of a variant's sequencing reads by all mapped reads and
155 normalizing by the wild-type frequency ratio. Sequencing reads were stringently filtered
156 for quality; we require each base have a Phred score greater than 20 and no uncalled
157 bases. Second, a weighted linear least squares regression line is fit to the normalized
158 frequency ratios across time points. Third, the slope of the regression line is calculated,
159 averaged across the three replicates and log₂ transformed. This averaged log₂ slope
160 reflects a variant's aggregation propensity. Solubility scores below 0 denote variants
161 that are more aggregation-prone than wild-type, whereas scores above 0 indicate that a
162 variant has increased solubility compared to wild-type.

163 *Classifying A β variants using synonymous mutations*

164 Variant classifications (*i.e.*, WT-like, more aggregation-prone, more soluble) were
165 assigned using the distribution of 39 synonymous mutations from the deep mutational
166 scan. We define WT-like as any variant with a score within ± 2 standard deviations of the
167 synonymous variant mean [-0.26,0.39]. A variant is more-aggregation prone than
168 wildtype if its score is greater than 0.39 or more soluble if its score is lower than -0.26.

169 *Data and code availability*

170 Raw sequencing data will be made available upon publication in the NCBI GEO
171 database. Code is available at <https://github.com/FowlerLab/amyloidBeta2019>.

172 **Results**

173 First, we verified that the DHFR-based yeast aggregation assay could differentiate
174 between aggregating wild type A β (A β_{WT}) and a nonaggregating (A β_{F19D}) variant (Morell
175 *et al.* 2011). As expected, in a mixed culture treated with methotrexate, A β_{F19D}
176 outcompeted A β_{WT} (**Figure 1B**). We used fluorescence microscopy of A β -GFP fusions
177 to confirm that ~30-70% of yeast expressing A β_{WT} -GFP had cytoplasmic punctae
178 compared to ~0-20% of cells expressing A β_{F19D} -GFP across five fields of view (**Figure**
179 **1C-D**). Thus, we concluded the assay could be used in a deep mutational scan to
180 measure the aggregation propensity of variants of A β .

181

182 Using this assay, we conducted a deep mutational scan of A β that yielded solubility
183 scores for 791 single amino acid variants, representing 99.1% of the possible single
184 variants. Solubility scores were calculated by taking the weighted least squares slope of
185 each variant's frequency ratios across six time points. (**see Methods**). The slopes from
186 each replicate were well correlated (Pearson's R 0.78 to 0.92; **Figure 2A**, **Figure S1A**).
187 Replicate slopes were averaged and log₂ transformed to produce final solubility scores
188 such that wild-type had a solubility score of zero (**Table S2**). Positive solubility scores
189 indicated less aggregation and negative scores indicated increased aggregation.

190

191 Solubility scores ranged from -2.38 (most aggregating) to 1.45 (most soluble). The
192 mean (median) solubility score for all variants was 0.09 (0.08), which was similar to the
193 solubility scores of the 39 synonymous variants in our library (mean: 0.06; median:
194 0.08). Because we did not expect synonymous variants to affect aggregation

195 propensity, we used their distribution of scores to identify WT-like variants (**Figure 2B**).
196 In total, we found that 344 (43.4%) of A β variant scores were within two standard
197 deviations of the synonymous score mean and thus had WT-like effects (WT-like range:
198 [-0.26,0.39]). Additionally, we found 246 (31.1%) variants to be more aggregation-prone
199 than A β_{WT} and 201 (25.4%) variants to be more soluble. Therefore, ~75% of A β variants
200 maintained or increased the peptide's propensity to aggregate in yeast cells.

201

202 To verify that our deep mutational scan accurately measured variant effects on
203 aggregation, we tested six A β variants, G38F, K16V, A42V, F19Y, L17S and L34R, that
204 spanned the solubility score range in a low-throughput validation assay. The growth rate
205 of methotrexate-treated yeast expressing each A β variant was measured and compared
206 to the aggregation propensity scores (**Figure 2C, S1B**). We found that low-throughput
207 assay results strongly correlated with the solubility scores derived from deep mutational
208 scanning ($R^2 = 0.98$). Thus, our deep mutational scan reliably measured A β variant
209 aggregation propensity in the yeast assay.

210

211 To explore the effects of each amino acid substitution on A β aggregation, we created an
212 A β sequence-aggregation map (**Figure 2D**). Substitutions to aspartic acid and proline
213 were most associated with A β solubility, as evinced by their median scores of 0.64 and
214 0.56, respectively (**Figure S2A**). Conversely, the most aggregation-associated
215 substitutions were hydrophobic tryptophan and phenylalanine, with scores of -0.60 and -
216 0.51, respectively. Moreover, hierarchical clustering of all 791 solubility scores by amino

217 acid revealed that hydrophobic substitutions, except alanine, clustered together and
218 were associated with greater aggregation than other classes of substitutions.

219

220 Next, we characterized each position in A β based on its mutational profile. Hierarchical
221 clustering of variant solubility scores by position identified six distinct clusters (**Figures**
222 **2E-F; S2B-C**). In cluster 1, comprising positions 17-20, 31-32, 34-36, 39 and 41,
223 substitutions tended to decrease A β aggregation compared to substitutions in other
224 clusters (cluster 1 mean solubility scores = 0.64, all other clusters = -0.28; **Figure S2D**).
225 In cluster 1, even substitutions to hydrophobic amino acids slightly decreased
226 aggregation (mean solubility score = 0.17). The effects of substitutions in cluster 2 were
227 similar to but less extreme than in cluster 1. Both clusters 1 and 2 are largely
228 comprised of hydrophobic positions in the wild type A β sequence. Indeed, 80% of A β
229 positions with hydrophobic wild type residues are in clusters 1 and 2. In stark contrast,
230 within clusters 4, 5 and 6, hydrophobic substitutions generally increase protein
231 aggregation (all mean: -0.15, -0.12 and -0.45; hydrophobic means: -0.29, -0.65, and -
232 1.04). Cluster 3 contains only two positions, 37 and 38. Here, every substitution except
233 proline increased aggregation (all mean: -0.99, hydrophobic mean: -1.56). Given that
234 cluster 1 is characterized by hydrophobic positions where hydrophilic substitutions
235 profoundly decreased aggregation, we suggest that this cluster defined buried β -strands
236 in the A β sequence.

237

238 Next, we compared our solubility scores to previous alanine and proline scans which
239 reported A β_{40} fibril thermodynamic stability *in vitro* ($\Delta\Delta G$). $\Delta\Delta G$ values were determined

240 by measuring variant A β monomer concentration remaining in solution after fibril
241 formation reached equilibrium (Williams *et al.* 2004; Williams, Shivaprasad, and Wetzel
242 2006a). We found that the effects of proline substitution in our assay were correlated
243 with proline $\Delta\Delta G$ values ($R^2 = 0.40$), while the effects of alanine substitutions in our
244 assay were less correlated with alanine $\Delta\Delta G$ values ($R^2 = 0.17$; **Figure 3A**). In our
245 alanine and proline comparisons, we found the greatest correlation at positions 17-20
246 and 31-32, where substitutions decreased aggregation in all studies (**Figure S3**). The
247 most notable disagreement between studies was for alanine substitutions at positions
248 37 and 38. In our assay, alanine substitutions caused a profound increase in
249 aggregation, whereas the *in vitro* alanine scan showed the opposite effect.
250

251 We also compared our buried β -strand positions from cluster 1 to β -stands proposed
252 based on the *in vitro* alanine and proline scans, finding some concordance (**Figure 3B**).
253 The single amino acid scans identify three regions that disrupt fibril elongation
254 thermodynamics when mutated. The regions include positions 15-21, 24-28, and 31-36
255 for the proline scan and positions 18-21, 25-26, and 32-33 for the combined alanine
256 and proline scans (Williams *et al.* 2004; Williams, Shivaprasad, and Wetzel 2006b).
257 Given the generally highly disruptive nature of proline substitutions (Gray *et al.* 2017), it
258 is not surprising that the proline scan would nominate many positions. Our deep
259 mutational scan, on the other hand, does not reveal a central β -strand or strong
260 decrease in aggregation with alanine or proline substitution from positions 24-28. We
261 speculate that this difference is due either to the distinct experimental approaches used
262 or to the different A β species (A β_{40} vs. A β_{42}).

263

264 **Discussion**

265 We used deep mutational scanning to characterize 791 A β variants in a yeast-based
266 aggregation assay. Proline and aspartic acid substitutions were most disruptive of A β
267 aggregation, while tryptophan and phenylalanine increased aggregation most.
268 Additionally, we used unsupervised clustering to determine the regions of A β most
269 important for aggregation. We conclude that these regions are most likely to form buried
270 β -stands, which are necessary for aggregation and sensitive to amino acid substitutions
271 (Jahn *et al.* 2010; Abrusán and Marsh 2016). These include positions 17-20, 31-32, 34-
272 35, 39 and 41. While other positions could also form β -stands, the positions in cluster 1
273 are most likely to form the buried cores of A β aggregates in our cell-based assay.

274

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

284

285 Given that we collected large-scale mutagenesis data in a cell-based system, we
286 examined how our results compared to structural models of A β fibrils. Some models
287 such as 1IYT (Crescenzi *et al.* 2002) and 2NAO (Wälti *et al.* 2016), showed very little to
288 no overlap with either our proposed buried β -strands or those proposed by Williams *et*
289 *al.* (2004, 2006) (**Figure 3C**). Other models contained three β -strand regions
290 reminiscent of those suggested by Williams *et al.* (2004, 2006): 2MXU (Xiao *et al.*
291 2015), 5KK3 (Colvin *et al.* 2016), and 5OQV (Gremer *et al.* 2017). Yet other models
292 propose β -strand patterns more similar to ours. These include 2BEG (Lührs *et al.* 2005),
293 2LNQ (Gremer *et al.* 2017), 2LMP and 2LMN (Lu *et al.* 2013). Since our β -strand
294 patterns were derived from data gathered in a cell-based assay, we hypothesized that
295 they would be most consistent with structural models based on *in vivo*-derived fibrils.
296 Indeed, the 2LMP and 2LMN models were based on fibrils seeded from plaques
297 isolated from the brains of individuals afflicted by Alzheimer's disease. Moreover, every
298 model besides 2LMP and 2LMN was constructed using NMR or cryo-EM data from
299 laboratory grown fibrils. These models are less concordant with our cell-based
300 mutational data, which suggests that there are important structural differences between
301 *in vitro* and *in vivo* derived fibrils.

302

303 Two major differences exist between the experimental conditions used by Williams *et al.*
304 (2004, 2006) and in our work, and may explain the difference in β -strands proposed in
305 our respective *in vitro*- and *in vivo*-derived models. First, Williams *et al.* (2004, 2006)
306 incubate A β in the absence of any other proteins, while our yeast-based system
307 provides key players that affect protein aggregation, such as chaperone proteins and

308 molecular crowding. Second, Williams *et al.* (2004, 2006) incubate A β peptides at 37C,
309 whereas our yeast-based experiments required a lower temperature of 30C. This
310 temperature difference may yield differences in folding kinetics. Further experiments are
311 required to determine the contribution of these experimental differences to β -strand
312 formation in A β .

313

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

320

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

330

331 **Acknowledgements**

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

341

342 **References**

- 343 Abrusán, G., and J. A. Marsh, 2016 Alpha Helices Are More Robust to Mutations than
344 Beta Strands. PLoS Comput Biol 12: 1–16.
- 345 Antzutkin, O. N., R. D. Leapman, J. J. Balbach, and R. Tycko, 2002 Supramolecular
346 structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy
347 and solid-state nuclear magnetic resonance. Biochemistry 41: 15436–15450.
- 348 Araya, C. L., and D. M. Fowler, 2011 Deep mutational scanning: assessing protein
349 function on a massive scale. Trends Biotechnol. 29: 435–442.
- 350 Bagriantsev, S., and S. Liebman, 2006 Modulation of Abeta42 low-n oligomerization
351 using a novel yeast reporter system. BMC Biol. 4: 1–12.
- 352 Bajaj, K., P. C. Dewan, P. Chakrabarti, D. Goswami, B. Barua *et al.*, 2008 Structural
353 correlates of the temperature sensitive phenotype derived from saturation
354 mutagenesis studies of CcdB. Biochemistry 47: 12964–12973.
- 355 Barz, B., Q. Liao, and B. Strodel, 2018 Pathways of Amyloid- β Aggregation Depend on
356 Oligomer Shape. J. Am. Chem. Soc. 140: 319–327.
- 357 Benzinger, T. L., D. M. Gregory, T. S. Burkoth, H. Miller-Auer, D. G. Lynn *et al.*, 1998
358 Propagating structure of Alzheimer's beta-amyloid(10-35) is parallel beta-sheet with

- 359 residues in exact register. *Proceedings of the National Academy of Sciences* 95:
360 13407–13412.
- 361 Bertram, L., and R. E. Tanzi, 2008 Thirty years of Alzheimer's disease genetics: the
362 implications of systematic meta-analyses. *Nat. Rev. Neurosci.* 9: 768–778.
- 363 Caine, J., S. Sankovich, H. Antony, L. Waddington, P. Macreadie *et al.*, 2007
364 Alzheimer's Abeta fused to green fluorescent protein induces growth stress and a
365 heat shock response. *FEMS Yeast Res.* 7: 1230–1236.
- 366 Colvin, M. T., R. Silvers, Q. Z. Ni, T. V. Can, I. Sergeev *et al.*, 2016 Atomic Resolution
367 Structure of Monomeric A β 42 Amyloid Fibrils. *J. Am. Chem. Soc.* 138: 9663–
368 9674.
- 369 Crescenzi, O., S. Tomaselli, R. Guerrini, S. Salvadori, A. M. D'Ursi *et al.*, 2002 Solution
370 structure of the Alzheimer amyloid beta-peptide (1-42) in an apolar
371 microenvironment. Similarity with a virus fusion domain. *Eur. J. Biochem.* 269:
372 5642–5648.
- 373 D'Angelo, F., H. Vignaud, J. Di Martino, B. Salin, A. Devin *et al.*, 2013 A yeast model for
374 amyloid- β aggregation exemplifies the role of membrane trafficking and PICALM in
375 cytotoxicity. *Dis Model Mech* 6: 206–216.
- 376 Dahlgren, K. N., A. M. Manelli, W. B. Stine, L. K. Baker, G. A. Krafft *et al.*, 2002
377 Oligomeric and fibrillar species of amyloid-beta peptides differentially affect neuronal
378 viability. *J. Biol. Chem.* 277: 32046–32053.
- 379 Fowler, D. M., and S. Fields, 2014 Deep mutational scanning: a new style of protein
380 science. *Nat. Methods* 11: 801–807.
- 381 Fowler, D. M., J. J. Stephany, and S. Fields, 2014 Measuring the activity of protein
382 variants on a large scale using deep mutational scanning. *Nat Protoc* 9: 2267–2284.
- 383 García-Nafría, J., J. F. Watson, and I. H. Greger, 2016 IVA cloning: A single-tube
384 universal cloning system exploiting bacterial In Vivo Assembly. *Sci Rep* 6: 1–12.
- 385 Gray, V. E., R. J. Hause, and D. M. Fowler, 2017 Analysis of Large-Scale Mutagenesis
386 Data To Assess the Impact of Single Amino Acid Substitutions. *Genetics* 207: 53–
387 61.
- 388 Gregory, D. M., T. L. Benzinger, T. S. Burkoth, H. Miller-Auer, D. G. Lynn *et al.*, 1998
389 Dipolar recoupling NMR of biomolecular self-assemblies: determining inter- and
390 intrastrand distances in fibrilized Alzheimer's beta-amyloid peptide. *Solid State Nucl
391 Magn Reson* 13: 149–166.
- 392 Gremer, L., D. Schölzel, C. Schenk, E. Reinartz, J. Labahn *et al.*, 2017 Fibril structure of
393 amyloid- β (1-42) by cryo-electron microscopy. *Science* 358: 116–119.

- 394 Haar, von der, T., L. Jossé, P. Wright, J. Zenthon, and M. F. Tuite, 2007 Development
395 of a novel yeast cell-based system for studying the aggregation of Alzheimer's
396 disease-associated Abeta peptides in vivo. *Neurodegener Dis* 4: 136–147.
- 397 Hardy, J., 2017 The discovery of Alzheimer-causing mutations in the APP gene and the
398 formulation of the "amyloid cascade hypothesis". *FEBS J.* 284: 1040–1044.
- 399 Hardy, J., and D. J. Selkoe, 2002 The amyloid hypothesis of Alzheimer's disease:
400 progress and problems on the road to therapeutics. *Science* 297: 353–356.
- 401 Iwatsubo, T., A. Odaka, N. Suzuki, H. Mizusawa, N. Nukina *et al.*, 1994 Visualization of
402 A beta 42(43) and A beta 40 in senile plaques with end-specific A beta monoclonals:
403 evidence that an initially deposited species is A beta 42(43). *Neuron* 13: 45–53.
- 404 Jahn, T. R., O. S. Makin, K. L. Morris, K. E. Marshall, P. Tian *et al.*, 2010 The common
405 architecture of cross-beta amyloid. *J. Mol. Biol.* 395: 717–727.
- 406 Jarrett, J. T., E. P. Berger, and P. T. Lansbury, 1993 The carboxy terminus of the beta
407 amyloid protein is critical for the seeding of amyloid formation: implications for the
408 pathogenesis of Alzheimer's disease. *Biochemistry* 32: 4693–4697.
- 409 Khare, S., M. Bhasin, A. Sahoo, and R. Varadarajan, 2019 Protein model discrimination
410 attempts using mutational sensitivity, predicted secondary structure, and model
411 quality information. *Proteins* 87: 326–336.
- 412 Lesné, S., L. Kotilinek, and K. H. Ashe, 2008 Plaque-bearing mice with reduced levels
413 of oligomeric amyloid-beta assemblies have intact memory function. *Neuroscience*
414 151: 745–749.
- 415 Lu, J.-X., W. Qiang, W.-M. Yau, C. D. Schwieters, S. C. Meredith *et al.*, 2013 Molecular
416 structure of β-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* 154: 1257–
417 1268.
- 418 Lührs, T., C. Ritter, M. Adrian, D. Riek-Lohr, B. Bohrmann *et al.*, 2005 3D structure of
419 Alzheimer's amyloid-beta(1-42) fibrils. *Proceedings of the National Academy of
420 Sciences* 102: 17342–17347.
- 421 Masters, C. L., and D. J. Selkoe, 2012 Biochemistry of amyloid β-protein and amyloid
422 deposits in Alzheimer disease. *Cold Spring Harb Perspect Med* 2: a006262–
423 a006262.
- 424 Matsumura, S., K. Shinoda, M. Yamada, S. Yokojima, M. Inoue *et al.*, 2011 Two distinct
425 amyloid beta-protein (Abeta) assembly pathways leading to oligomers and fibrils
426 identified by combined fluorescence correlation spectroscopy, morphology, and
427 toxicity analyses. *J. Biol. Chem.* 286: 11555–11562.
- 428 Morell, M., N. S. de Groot, J. Vendrell, F. X. Avilés, and S. Ventura, 2011 Linking
429 amyloid protein aggregation and yeast survival. *Mol Biosyst* 7: 1121–1128.

- 430 Rollins, N. J., K. P. Brock, F. J. Poelwijk, M. A. Stiffler, N. P. Gauthier *et al.*, 2018 3D
431 protein structure from genetic epistasis experiments. 1–17.
- 432 Rubin, A. F., H. Gelman, N. Lucas, S. M. Bajjalieh, A. T. Papenfuss *et al.*, 2017 A
433 statistical framework for analyzing deep mutational scanning data. *Genome Biol.* 18:
434 1–15.
- 435 Schmiedel, J., and B. Lehner, 2018 Determining protein structures using genetics. 1–
436 49.
- 437 Shankar, G. M., M. A. Leissring, A. Adame, X. Sun, E. Spooner *et al.*, 2009 Biochemical
438 and immunohistochemical analysis of an Alzheimer's disease mouse model reveals
439 the presence of multiple cerebral Abeta assembly forms throughout life. *Neurobiol.*
440 *Dis.* 36: 293–302.
- 441 Treusch, S., S. Hamamichi, J. L. Goodman, K. E. S. Matlack, C. Y. Chung *et al.*, 2011
442 Functional links between A β toxicity, endocytic trafficking, and Alzheimer's disease
443 risk factors in yeast. *Science* 334: 1241–1245.
- 444 Tycko, R., 2011 Solid-state NMR studies of amyloid fibril structure. *Annu Rev Phys*
445 *Chem* 62: 279–299.
- 446 Wälti, M. A., F. Ravotti, H. Arai, C. G. Glabe, J. S. Wall *et al.*, 2016 Atomic-resolution
447 structure of a disease-relevant A β (1-42) amyloid fibril. *Proc. Natl. Acad. Sci. U.S.A.*
448 113: 1–9.
- 449 Williams, A. D., E. Portelius, I. Kheterpal, J.-T. Guo, K. D. Cook *et al.*, 2004 Mapping
450 abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol.*
451 *Biol.* 335: 833–842.
- 452 Williams, A. D., S. Shivaprasad, and R. Wetzel, 2006a Alanine scanning mutagenesis of
453 Abeta(1-40) amyloid fibril stability. *J. Mol. Biol.* 357: 1283–1294.
- 454 Williams, A. D., S. Shivaprasad, and R. Wetzel, 2006b Alanine Scanning Mutagenesis
455 of A β (1-40) Amyloid Fibril Stability. *J. Mol. Biol.* 357: 1283–1294.
- 456 Xiao, Y., B. Ma, D. McElheny, S. Parthasarathy, F. Long *et al.*, 2015 A β (1-42) fibril
457 structure illuminates self-recognition and replication of amyloid in Alzheimer's
458 disease. *Nat. Struct. Mol. Biol.* 22: 499–505.
- 459

460 **Figure Legends**

461 **Figure 1A-D. A yeast-based aggregation assay distinguishes between soluble and**
462 **aggregation-prone variants of A β .** A schematic of the assay shows plasmid-based
463 expression of A β -DHFR and a nonaggregating variant of A β fused to DHFR, which lead
464 to slow and fast yeast growth in the presence of methotrexate, respectively (**A**). A
465 stacked bar graph shows the percentage of A β -DHFR and A β_{F19D} -DHFR in co-culture
466 (y-axis) every 12 hours for 48 hours (x-axis; **B**). Fluorescence light microscopy shows
467 the aggregation patterns of A β -GFP (WT) and A β_{F19D} -GFP (F19D) 16h after induction of
468 expression (**C**). A bar graph shows the percentage of yeast cells with punctae (y-axis) in
469 five fluorescence microscopy images of A β -DHFR (WT) or A β_{F19D} -DHFR (F19D; x-axis;
470 **D**).

471

472 **Figure 2A-F. Solubility scores for 791 A β variants.** Solubility scores reliably measure
473 the effects of A β sequence on aggregation propensity. A scatter plot shows the
474 correlation between two of three biological replicates that were averaged to yield final
475 solubility scores (**A**; **Figure S1A**). The distribution of solubility scores (x-axis) of
476 synonymous variants was used to determine cutoffs that define variants that are wild-
477 type-like or more/less aggregation-prone than wild-type. The density plot shows
478 distributions of nonsynonymous (light gray) and synonymous (dark gray) variants and
479 the white lines show the lower (-0.26) and upper (0.39) bounds for wild-type-like
480 variants (**B**). The scatterplot shows the correlation between our solubility scores (y-axis)
481 and a low-throughput yeast growth assay that measured yeast growth rate as a proxy
482 for A β solubility (**C**; **Figure S1B**). The heatmap shows the effect of 791 A β variants on

483 solubility with A β positions on the x-axis and mutant amino acids on the y-axis. A
484 variant's color denotes its solubility: red is most soluble, white is wild-type-like and, dark
485 blue is most aggregated, whereas yellow variants are missing from our variant library
486 and dots denote the wild-type amino acid at a given position. The annotation tracks on
487 the x- and y-axes display the hydrophobicity of each wild-type and mutant amino acid,
488 respectively. The heatmap's y-axis has been re-ordered using hierarchical clustering on
489 the solubility score vectors (**D**). For each position, the mean solubility score at each
490 position is depicted using the same color scheme as the main heatmap. Additionally,
491 the mean solubility scores for all hydrophobic and polar substitutions are shown (**E**;
492 **Figure S2A**). Heirarchical clustering on the x-axis yielded 6 distinct clusters: 1 (red), 2
493 (orange), 3 (yellow), 4 (green), 5 (light blue), and 6 (dark blue; **F**; **Figure S2B-C**).

494

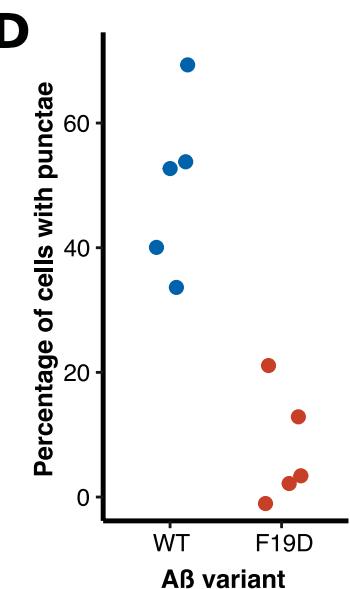
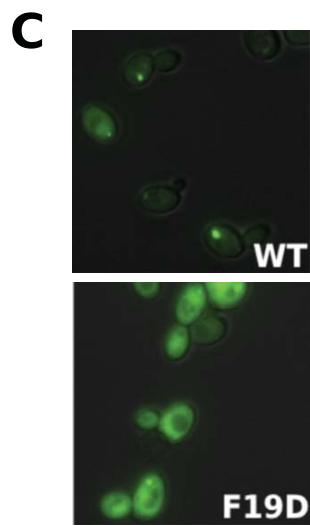
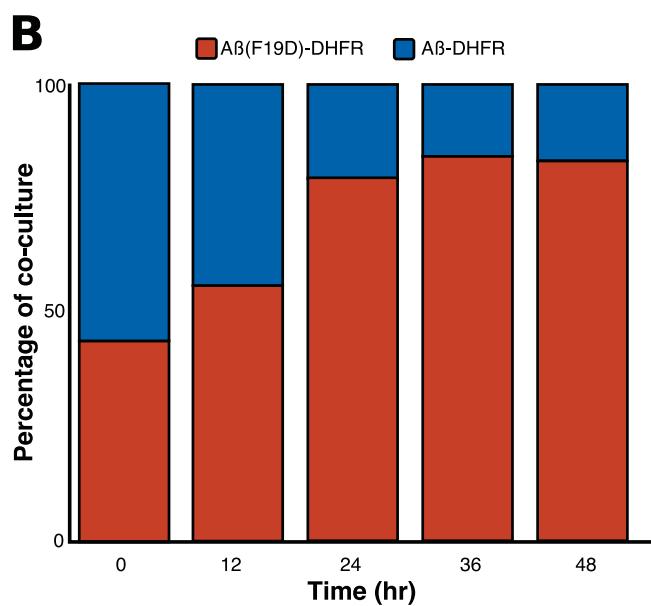
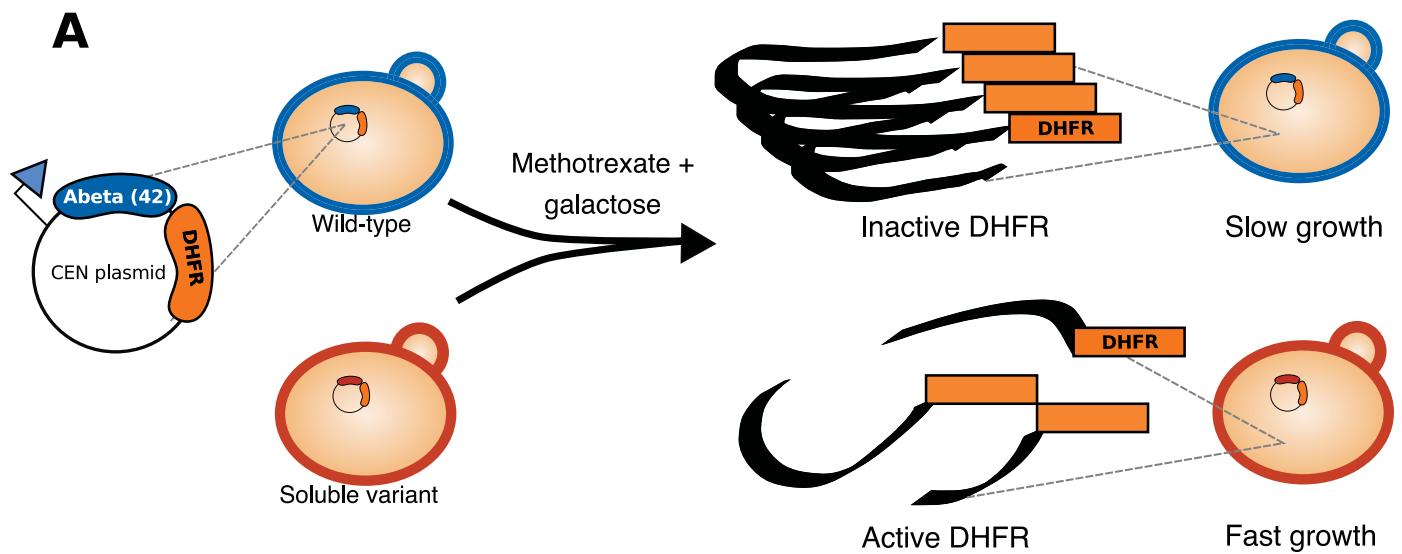
495 **Figure 3. Comparison of yeast cell-based solubility scores to *in vitro* aggregation**
496 **measurements and A β structural models.** The scatterplot shows the correlation
497 between our solubility scores (y-axis) and two single amino acid scans that measured
498 the effect of proline (orange) or alanine (teal) variants on the thermodynamic stability of
499 aggregates, relative to wild type ($\Delta\Delta G$) (**A**; **Figure S3**). The first two tracks show
500 unmeasured mutations (dashed gray) and the A β buried β -strand positions (black)
501 suggested by proline scanning alone, or by proline and alanine scanning together
502 (Williams *et al.* 2004; Williams, Shivaprasad, and Wetzel 2006a). The third track shows
503 positions with the greatest increase in solubility when mutated in our large-scale
504 mutagenesis study, found in cluster 1 (**B**). The next nine tracks show the secondary

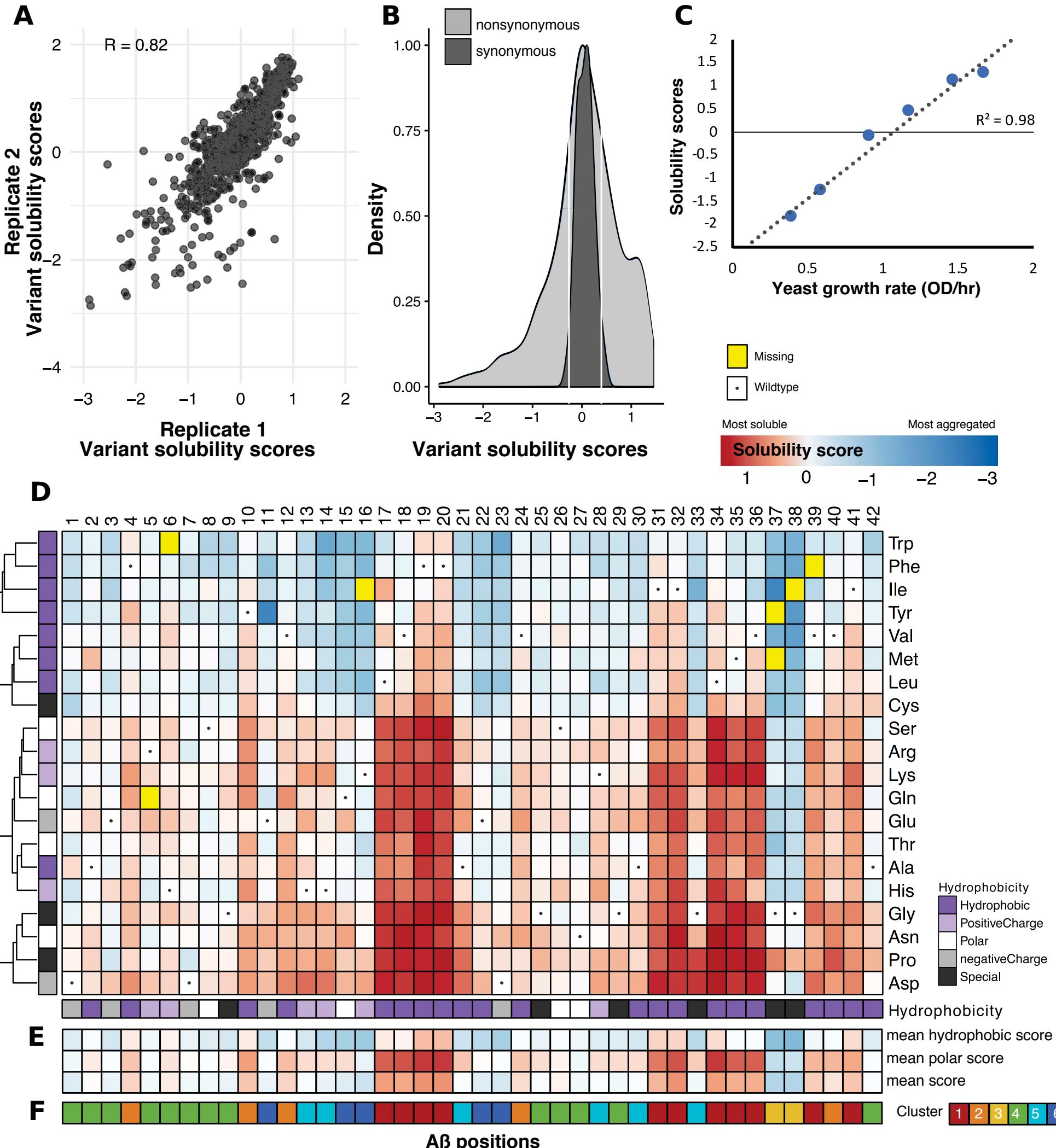
505 structure of nine models of A β aggregate structure for each A β position (x-axis; **C**). The
506 A β wild-type sequence is shown at the top.

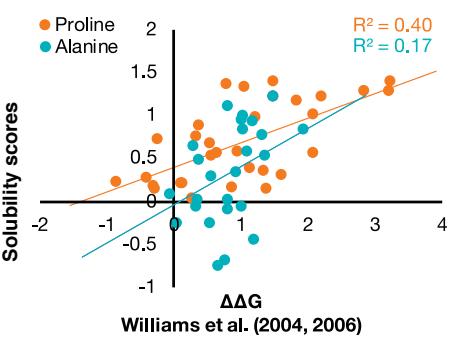
507

508

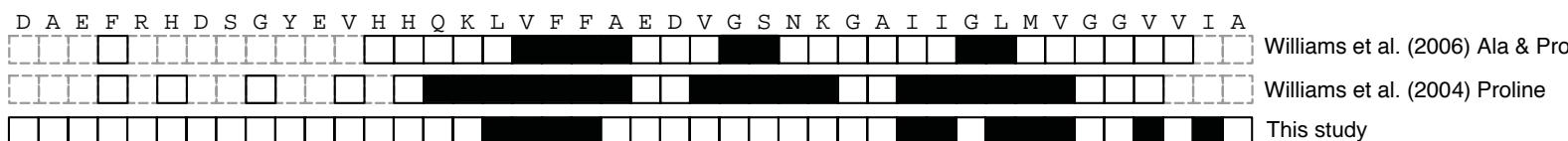
509





A

Williams et al. (2004, 2006)

B**C**