Epitomic analysis of the specificity of conformation-dependent, anti-aß amyloid monoclonal antibodies.

Jorge Mauricio Reyes-Ruiz[1], Rie Nakajima[2], Ibtisam Baghallab[3], Luki Goldschmidt[4], Taha Kumosani[3], Philip L. Felgner[2], and Charles Glabe[1*]

[1]Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA 29697-3900

[2]Vaccine Research and Development Center, Department of Physiology and Biophysics, University of California Irvine, Irvine, California, USA.

[3]Biochemistry Department, Faculty of Science, King Abdulaziz University, Jeddah - Saudi Arabia

[4]Institute for Protein Design, Department of Biochemistry, University of Washington**,** NanoES 396**,** Seattle, WA 98195-7370

[*]Corresponding author: Charles Glabe

Email: cglabe@uci.edu

Running title: Epitomic analysis of anti-aß antibodies

Keywords: Bioinformatics, Alzheimer Disease, Amyloid antibodies, Discontinuous epitopes, Epitope mapping, Phage display, Immunotherapy

Abstract

Antibodies against Aß amyloid are indispensable research tools and potential therapeutics for Alzheimer's Disease, but display unusual properties, such as specificity for aggregated forms of the peptide, ability to distinguish several polymorphic aggregate structures and ability to recognize generic aggregation-related epitopes formed by unrelated amyloid sequences. Understanding the mechanisms underlying these unusual properties of anti-amyloid antibodies and the structures of their corresponding epitopes is crucial for the understanding why they display different therapeutic activities and for the development of more effective therapeutic agents. Here we employed a novel "epitomic" approach (1) to map the fine structure of the epitopes of 28 monoclonal antibodies against amyloid-beta using immunoselection of random sequences from a phage display, deep sequencing and pattern analysis to define the critical sequence elements recognized by the antibodies. Although most of the antibodies map to linear epitopes in the amino terminal 1-14 residues of Aß, the antibodies display differences in the target sequence residues that are critical for binding and in their individual preferences for non-target residues, indicating that the antibodies bind to alternative conformations of the sequence by different mechanisms. Epitomic analysis identifies more non-overlapping sequence Aß segments than peptide array approaches that may constitute the conformational epitopes that underly the aggregation specificity of antibodies. Aggregation specific antibodies also recognize sequences that display a significantly higher predicted propensity for forming amyloid than antibodies that recognize Aß monomer, indicating that the ability of the random sequences to aggregate into amyloid structures is a critical element of their binding mechanism.

Introduction

Amyloids are intermolecularly hydrogen bonded ß-sheet aggregates that have a regularly repeating lattice structure (2,3). Many proteins are able to form amyloids upon unfolding or misfolding and are frequently associated with neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) and prion diseases (4). Alzheimer's disease contains two canonical amyloids: Aß amyloid, derived from APP and tau amyloid and in approximately 30% of AD, cortical Lewy bodies are also present containing α-synuclein amyloid. Conformational-dependent monoclonal antibodies against amyloid beta are an invaluable tool for research in the role of amyloids in Alzheimer's disease because they recognize epitopes that are differentially displayed on distinct structural polymorphs or folded states of the peptide, providing insight into the role of polymorphisms in the pathogenic spectrum of the disease (5). Many of the antibodies raised against Aß amyloid specifically bind to aggregated oligomeric or fibrillar forms of the peptide and do not bind to monomer or the amyloid precursor protein (6-9). Moreover, many of these aggregation state-specific antibodies recognize aggregates formed from unrelated sequences, indicating that amyloids display generic epitopes as a consequence of their structure. Understanding the mechanisms underlying the unusual specificities of anti-amyloid antibodies and the structures of their corresponding epitopes is crucial for the understanding of the immune response to amyloid and for the development of effective therapeutic agents.

We have previously reported polyclonal rabbit sera that distinguish two classes of amyloid aggregates. A11 sera was raised against oligomeric Aß mimics made of Aß42 coupled via the carboxyl terminus to colloidal gold that recognizes prefibrillar oligomers, but not fibrils or monomer (6). Subsequently we reported OC serum that was raised against Aß42 fibrils that is specific for fibrils and fibrillar oligomers but not prefibrillar oligomers or monomer (10). A11 and OC serum recognize prefibrillar oligomers and fibrils respectively from several other amyloid forming proteins, including α-synuclein and islet amyloid polypeptide (IAPP), indicating that some of the antibodies in the sera recognize generic amyloid aggregates that do not depend on the precise peptide sequence. We also raised rabbit antiserum against Aß42 annular protofibrils, which are pore-like structures, but this serum labels the same type of prefibrillar oligomers as A11, indicating that these sera have related specificities (11). We cloned 6 monoclonals from A11 serum-producing rabbits (8), 2 monoclonals from annular protofibril

vaccinated rabbits (Glabe, unpublished) and 23 monoclonals from OC serum-producing rabbits (9). For the OC monoclonals (mOC), the epitope was mapped using the PepSpots method of overlapping 10mers that vary by a single amino acid from position -5 to 45. Nineteen of the 23 mOC monoclonals gave a pattern of reactivity that mapped to either a linear or discontinuous epitope in the amino terminal 2/3 of the molecule (9) while 16 map to the amino terminal residues from 1-11 (9). Although these antibodies bind to the same region of Aß, they bind selectively to alternative fibrils structures of Aß, indicating that this region is conformationally polymorphic (12). Here we characterize the specificity of 28 of these monoclonal antibodies in fine detail using the epitomic analysis of the binding specificity (1) by combining random sequence phage display, Illumina next generation sequencing and bioinformatics. The results indicate that there is a broad diversity of preferred epitopes of the antibodies and indicate that the sequences immunoselected by the aggregation-specific antibodies are predicted to have a higher amyloid-forming propensity than the antibodies that bind monomer. This diversity of epitope specificities of antibodies that bind to Aß may underly the differences in therapeutic effectiveness of anti-amyloid monoclonal antibodies in clinical trials.

Results

Library randomness, bias and non-specific binding of phage.

In order to assess the randomness and amino acid bias in the library as purchased from the manufacturer, DNA from an aliquot of the library was extracted, the random sequences amplified by PCR and deep sequenced, yielding 19,434 sequences of which 97.4% were unique sequences (Appendix file 1, LibrarySeqOnly, column A, Original unamplified library). We also amplified the library 2 successive rounds and examined the effect on the number of unique sequences observed. (Appendix file 1, LibrarySeqOnly, columns B, C, Amplified round 1 and round 2). In the first round of amplification, 93.4% of the sequences were unique while in the second round of amplification 93.9% were unique. To further compare the randomness of the library, we compared the unique sequences between the original library and the first amplification and found that 2.1% of the unique sequences were observed in both libraries. This is the same as the number of unique sequences that were observed twice in the same original unamplified library (2.1%). These results indicate that while some sequences may have amplified preferentially, the vast majority of sequences in the library remain random.

The library has an inherent bias in terms of the frequency of nucleotides observed as shown in Appendix Table I. Thymidine is over-represented, while adenine is the lowest abundance. We also calculated the frequency of amino acid residues observed in the random library as shown in Appendix Table II. Not surprisingly, leucine and serine, amino acids that have 6 codons are abundantly encountered, although arginine which also has 6 codons is encountered with approximately half the frequency. Threonine with 4 codons is also abundant, while Proline is surprisingly abundant with only 2 codons. Cysteine is the lowest abundance amino acid observed. These results are very similar to the manufacturer's observed amino acid frequency published in their data sheet for the library. Stop codons are observed with a low frequency of 0.45%, presumably because stop codons halt the expression of the pIII protein that the random sequence is fused to that is required for infectivity. It is not clear why any stop codons are observed if pIII is absolutely required for infectious activity, but it is possible that this may arise from the initial transfection of the phagemid where a sequence containing a stop codon is co-transfected into the same bacterium and packaged into phage.

Non-specific binding may have a significant deleterious effect on the analysis of the patterns observed that bind to the paratope of the antigen combining site of the antibody because they do not fit any specific pattern and therefore lower the minimum percentage of the total sequences needed to match an epitope pattern (see C% below). We found that there are two types of non-specific binding sequences. The first type are random sequences that bind specifically to the protein A magnetic beads and the constant regions of the antibodies. The patterns of key amino acid residues for antibody binding that arise from these

sequences are found in all replicates of the control beads without any antibody and also found in all replicates containing antibody of the same isotype regardless of antibody's specificity. These sequences were identified and removed from the sequence files in the Bash processing script used to extract the peptide sequences from the Illumina data files. The second type of non-specific binding sequences arise from the binding of phage to the beads or antibody through an interaction other than with the random sequence fused to pIII. These sequences display the same frequency distribution as the unselected random library in terms of the presence of a high percentage of unique sequences observed only once in the sequencing run.

The purpose of repeating the immunoselection step is to increase the proportion of specific sequences, so we investigated the effect of immunoselection followed by amplification of the phage at each of the 3 steps. Using antibody mOC1 as an example, in the first immunoselection step prior to amplification 73.2% of the sequences occur a single time. At the second immunoselection prior to amplification, 98.7% are single. In the third immunoselection, 74.0% of the sequences are single reads, indicating that a substantial amount of non-specific random sequences remains after repeated immunoselection.

Epitomic analysis of monoclonal antibody specificity

We performed 3 immunoselection or panning steps for each antibody amplifying the eluted phage after each step and sequenced each round of panning before and after amplification. We analyzed all 6 of these sequencing groups and found that the patterns of specific binding patterns were largely the same for each antibody, although the number of specific sequences increases with successive pannings (data not shown). We chose to present the data for the unamplified samples after 3 rounds of pannings because they have been subjected to 3 rounds of immunoselection and only 2 rounds of amplification (Appendix file 2 Immunoselected sequences). A total of 6,282,154 sequence reads were obtained from the 28 monoclonals including the unamplified and amplified steps of the three pannings. This includes 22 of the mOC series monoclonals and 6 of the mA11 monoclonals. Two of the mA11 monoclonals were not analyzed because the hybridomas producing them were lost (mA11-201 and mA11-121) and the two monoclonals raised against Aß42 annular protofibrils (mA11-09 and mA11-89) were included with the mA11 monoclonals because their properties overlap in terms of binding to ß-barrel structures like annular protofibrils and hemolysins (13,14). Once the sequences were processed by removing the non-specific sequences and duplicates, we found a total of 1,713,634 unique peptide sequences. For each unique sequence found, we also counted the number of times it was observed. For the third unamplified step used in the analysis, the range was from 223 unique sequences for mOC51 to 53,621 for mOC98. The average number of unique sequences per monoclonal in the third unamplified panning was 10,959. The unique sequences were sorted by the descending number of times their sequence was observed as a surrogate for relative binding activity.

The sequences were analyzed using the PRATT 2.1 amino acid sequence pattern recognition algorithm that is a method for the identification of patterns in a set of unaligned protein sequences (15). Pratt returns patterns expressed in Prosite notation that include the sequence pattern, the number of sequences that match the pattern ("hits") and a "fitness" score, which is a non-statistical parameter that reflects the non-randomness or complexity of the pattern. Pratt has several variable parameters that significantly influence the patterns found. One of the most important parameters is C% (or CM), the minimum number of unique sequences to match in a pattern. If C% is set too high, no patterns are found or only low fitness patterns are found. If C% is set too low, then many patterns are observed that are sub-patterns of a common parent pattern. As C% values decline, the length of the patterns and their fitness increases, depending on the specificity of the antibody. We illustrated the effect of C% values again using mOC1 (Table I). As the value of C% drops, the number of sequences satisfying the pattern ("hits") drops and the fitness increases. The flexibility number (FN) specifies the maximum length or flexibility of arbitrary sequence spacers or "wild card" regions that can be any amino acid. We chose a FN value of 0 so that only fixed length patterns where X is one residue were evaluated. While Pratt returns the same patterns with larger

values of FN, the program takes longer to run and Pratt also finds more low fitness patterns because of the increased flexibility. Pattern length (PL) is the maximum pattern length allowed and it must be set long enough to accommodate the longest epitope encountered. If it is set too low, Pratt will find multiple lower fitness sub-patterns that segments of a larger pattern. Longer values do not yield any more or different patterns, so we used a value of 12 because the maximum sequence length is 12. For most of the analysis we used C% = 10; FN = 0 and PL = 12. If using these parameters did not find any patterns with fitness values above those observed for the random sequence library alone (fitness > 12.50), then we adjusted the C% value down until higher fitness patterns were obtained. Non-specific binding has a significant effect on C%, because the non-specific sequences do not fit a specific pattern, lowering the threshold number of C% required to observe the specific patterns. The non-specific sequences give the same low fitness patterns of 3-4 residues of the most abundant amino acids (S, L and P) that the unselected library does, with a fitness score of 12.5 for C% = 1 and 16.68 for C% = 0.1 (Table I).

We first analyzed the effect of non-specific sequences by comparing the patterns found in the single copy sequences that contain the vast majority of non-specific sequences to the patterns obtained from the top 100 most frequently encountered sequences that bind in a sequence specific fashion. Surprisingly there is very little difference in the patterns found, although the exact pattern and fitness score can vary slightly comparable to that caused by changing the C% value. Because there was little difference for all of the antibodies between the highly observed sequences and the total sequences, we show the data obtained from the total sequences here (Table II). Only the top 3 patterns with a fitness score above 12.5 are shown in Table II. If more than one pattern had the same fitness score, the patterns with the highest number of hits are shown.

Half of the antibodies (mA11-09, mOC15, mOC29, mOC31, mOC41, mOC51, mA11-55, mA11-89, mOC98, mOC107, mOC108, mA11-118, mA11-204 and mA11-205) displayed no specific pattern at C% = 10, returning low fitness value pairs and triplets of amino acids that are similar in composition and fitness to the patterns observed for the random library sequences alone that consist of the most abundant amino acids, such as leucine, proline, threonine and serine, indicating the patterns are dominated by non-specific sequences (Appendix Table II). All of these antibodies display specific patterns at lower values of C% as listed in Table II. All of the mOC series antibodies and 5 of the 6 A11 antibodies (mA11-09, 55, 118, 204 and 205) display patterns related to the Aß sequence while mA11-89 displays a pattern unrelated to Aß. Individual peptides belonging to the mA11-89 patterns give multiple high identity hits in the non-redundant protein data base (data not shown). It is not clear why this antibody binds to Aß oligomers and oligomers from other amyloid sequences, although the sequence pattern observed may form a mimotope that is common to the ß-barrel oligomers that mA11-89 binds to.

Although most of the antibodies bind to the amino terminal third of Aß, each antibody displays differences in its binding preference in terms of the precise location on the target sequence, the identity of target amino acids required for binding, the location and composition of wild card positions where any amino acid or a subset of alternative residues are permitted and locations where non-target amino acids are preferred over target residues. Non-target amino acids such as a S or P at sequence position 1 or 2 are preferred or tolerated in several antibodies (S or P: mOC16, 24, 31, 51, 87 and mA11-118). Many of the antibodies that map to a region of Aß that contain an F or Y residue can accommodate either amino acid or sometimes W, while mOC116 only recognizes sequences containing the target F at position 20. Inspection of the actual sequences immunoselected by the antibodies indicate that mOC108 has a strong positional preference for sequences that begin at LVF or IVF, indicating that it prefers a free amino terminus. Sequences beginning at [IL]VF occur at approximately 93% of the total sequences recognized by mOC108.

Twenty-six of the monoclonals clearly recognize an Aß related epitope that is located in the amino-terminal (1-15) or central (16-30) thirds of the molecule or both (Fig. 1). Ten antibodies (mOC1, mA11-09, mOC15, mOC22, mOC23, mOC64, mOC98, mOC107, mA11-118 and mA11-204) all

have epitopes centered on residues 1-5 DAEFR. Ten antibodies (mOC3, mOC16, mOC24, mOC29, mOC31, mOC41, mOC51, mOC87, mOC88 and mA11205) prefer residues 4-7 FRH, while mOC antibodies 78 and 86 bind to a site centered on residues 8-10 (SGY in Aß). Four monoclonal antibodies bind to epitopes in the central third of Aß. mOC108 binds to 17-19 (LVF), mOC9 and mA11-55 prefer the residues LVFFA (residues 17-21) and mOC 116 binds to residues 20-23 (FAED). One antibody, mOC76 gives a pattern, G-[IV]-[IV]-I-x-S, that is related to carboxyl terminal sequence residues 38-43 of Aß42, even though our previous epitope mapping study indicated that this antibody maps to residues 5-10, RHDSGY using the "PepSpots" overlapping 10mer epitope mapping method (9).

Our previous studies indicated that 6 of the antibodies (mOC1, 24, 31, 51, 78 and 98) recognize a discontinuous epitope because they react with more than one segment of the Aß sequence using the overlapping peptide PepSpots method (16). The epitomic analysis of the antibodies also provides support for the recognition of discontinuous epitopes consisting of non-overlapping sequence segments of Aß. For example, antibody mOC98 binds to both residues 2-5 and residues 8-11 as the top two patterns observed (Table II). In consideration of the possibility that other Aß segments of discontinuous epitopes may be lower ranked patterns, we ran the PRATT analysis at lower values of C% (1%) and output all of the patterns with a fitness score above 12.5. We then searched for patterns containing other segments of Aß. This analysis identified 16 antibodies that recognize more than 1 non-overlapping segment of Aß as shown in Table III. While the epitomic analysis confirms the discontinuous epitopes identified by the PepSpots assay for antibodies mOC3, 51, 78 and 98, it fails to confirm a second binding segment for mOC1, and 24. In addition, it identifies 12 additional antibodies with discontinuous epitopes that were not observed using the PepSpots membranes (mOC9, 29, 31, 41, 64, 76, 86, 87, 98, and mA11-55, 118 and 204). The resulting map of antibody binding sites on Aß including the major epitopes for 27 of the 28 antibodies and the minor epitopes for the antibodies that recognize more than 1 segment of Aß is shown on Fig. 1. Only two segments of the Aß sequence are not recognized by any of the antibodies, residues 13-16 (HHQK) and residues 28-35 (KGAIIGLM).

In order to confirm the binding sites for the antibodies on the Aß target sequence and attempt to resolve the conflicts between the random sequence epitomic approach and the PepSpots method, we repeated the peptide array using synthesized overlapping 10mers of the Aß sequence from the -4 position to residue 45 and arraying the sequences on glass slides and probing the array with the antibodies. The data is shown in Figure 2 and results of the PepSpots assay, the Epitomic analysis (Table III) and the microarray data are summarized in Table IV. The microarray data indicates that only 5 of the 16 antibodies identified by epitomic analysis as binding discontinuous epitopes are confirmed by the two peptide array assays (mOC31, mOC51, mOC78, mOC98 and mA11-55). Even though all 3 approaches agree, the actual segments are different among the three different assays. For example, mOC51 binds to residues 3-7 and 19-25 according to PepSpots, residues 6-11 and 17-21 according to Epitomic analysis and 13-21 and 28-37 by the microarray data. The two peptide arrays don't agree with each other for several of the antibodies. The epitomic analysis finds more discontinuous epitopes than the peptide array approach and this may be due to the fact that in random sequence selection, both segments may be present simultaneously to create a stronger binding pair than the individual segments by themselves. The microarray results also confirm that mOC76 binds to the carboxyl-terminal sequences of Aß, binding most strongly to the peptide ending at residue 45 (GVVIATVI) and fails to confirm the data from the spots assay that indicate binding at the amino terminus.

Most of the antibodies are specific for aggregated forms of Aß and do not recognize monomer or the amyloid precursor protein, so it is not obvious from the epitope fine structures why the sequences displayed by the phage bind. To investigate this issue further, we analyzed the specific sequences immunoselected for the propensity to form amyloid by three different algorithms, AGGRESCAN (17), Waltz (18) and the 3D profile method (19). The first two methods are sequence-based while the last method is structure based. The three methods give

a general agreement on the amyloid forming propensity, but the Waltz algorithm provides higher estimates and the 3D profile methods gives lower estimates of amyloid formation. We analyzed the top 100 sequences for each antibody and 100 unselected random sequences for amyloid formation propensity and the results are shown in Table V. Although there is considerable variation in the amyloid-forming propensity for some of the antibodies, the antibodies that are specific for amyloid aggregates are associated with sequences that have a significantly higher propensity for forming amyloid than the antibodies that recognize monomer ($p < 0.002$). Since there are about 5 copies of the pIII protein all located on the head side of the filamentous phage, it is conceivable that they could form amyloid-like intermolecular aggregates on the surface of the phage. Some of the antibodies that recognize monomer have sequences with very low amyloid forming propensity that is lower than the average amyloid forming propensity for random 12mers from the unselected library and lower than that for the proteins, APP, Tau, Synuclein and BSA. This preference for low amyloid forming propensity may be a reflection of the preference of these antibodies for unstructured regions of peptide. We have previously reported that some of these antibodies bind to amyloid plaques and intraneuronal amyloid in human and transgenic mouse brain while others do not (9,20). In addition, mOC31 specifically recognizes vascular amyloid (9,21). We also found that two of the antibodies, mOC22 and mOC23 differentially recognize distinct regions of cored plaques (Fig. 3). mOC22 recognizes the entire cored plaque while mOC23 preferentially binds to the rim of the plaque, indicating that even though both antibodies bind to the same amino terminal region of Aß, their distinct epitomic specificities are meaningful in their ability to identify inhomogeneities in plaque structure. Diffuse plaques stain uniformly with both antibodies indicating that different types of plaques display different sets of epitopes (Fig. 3). This provides further evidence for the utility of these antibodies in distinguishing different types of amyloid in human brain.

Discussion

A standard approach for epitope mapping is the SPOTS assay, where a series of overlapping peptides are synthesized and arrayed and used to determine the region of protein sequence to which an antibody binds. While this is a reasonably inexpensive and facile method, it lacks fine details of the interactions and fails to identify many discontinuous conformational epitopes, such as those that are found on amyloid aggregates. Phage display has also been used to identify epitopes, but the classical approach of sequencing positive individual clones limits the number of samples one can realistically process at once (22). Here we combine random sequence phage display and next generation sequencing to identify the fine details of monoclonal antibody recognition of Aß and its aggregates. Thousands of immunoselected sequences were analyzed using the Pratt 2.1 pattern recognition algorithm for patterns common to specifically bound phage that define the preferred epitope sequences recognized by the antibody. By coupling immunoselection of random sequences with deep sequencing, the results provide unprecedented detail into the nature of the epitope and its interaction with the antibody. Not only do the results identify the residues of the target sequence that constitute the binding site, but they also provide details about non-target residues that are allowed or preferred for binding. Epitomic analysis also identifies non-overlapping segments that may constitute conformational epitopes for antibodies that fail to bind to these sequences using peptide array methods. Identifying conformational epitopes for monoclonal antibodies is an important and challenging topic for immunology and epitomic analysis may be a rapid and facile means of investigating conformational specificity. The availability of multiple peptide sequences simultaneously for antibody binding may allow the identification of binding segments where each individual sequence binds too weakly to allow identification using single linear peptide segment approaches, like peptide arrays. In addition, some antibodies appear to bind one segment more tightly that the other so they may be misidentified as having a single linear epitope by peptide arrays when they actually recognize a discontinuous epitope. Reexamination of these antibodies using epitomic analysis may be warranted. These details may be very useful for understanding the binding mechanism and modeling the binding interaction in the absence of detailed structural information of the antibody-antigen complexes. This insight could

also be very useful for predicting antibody cross reactivity.

The results show that all of the antibodies display distinct preferences for epitopes consisting of a mixture of target and non-target residues and positions where a number of different amino acids are accommodated. This suggests that the regions of Aß containing these epitopes are able to adopt a number of different structures such that the epitopes are displayed or not in a combinatorial fashion. This provides an explanation for why the antibodies bind to different Aß fibril structures (12), and bind to different types of amyloids *in vitro* and *in vivo* (9,20,21,23). Amyloids are intermolecularly hydrogen bonded cross-ß structures that are structurally polymorphic (reviewed in (24,25)), and the fact that Aß displays a number of different structural polymorphisms is consistent with the fact that the immune response gives rise to a large number of different antibodies that have different binding modes for the different structures of the same peptide segments of Aß. The fact that so many of the antibodies that bind to the amino terminal segment of Aß are specific for aggregated Aß and recognize different structural variants indicates that the commonly held view that antibodies that bind to a linear segment of a peptide sequence are "non-conformational" is an over simplification at least for anti-amyloid antibodies. A more accurate view may be whether it can adopt more than one structure. An implication of the conformational polymorphisms of the Aß peptide is that a peptide sequence of Abeta does not predict what type of antibody will be produced in response to immunization and whether it will be therapeutically effective because it may give rise to different immune responses in different people. Mimotopes such as the ones identified here may give a more predictable immune response because it may adopt a single predominant structure and give rise to a more predictable immune response.
We hoped to also gain some insight into some of the peculiar properties of anti-amyloid antibodies, such as why many of them are specific for aggregates, why they recognize generic epitopes that do not depend on a specific amino acid sequence and why the A11 and OC series antibodies recognize mutually exclusive aggregation-specific epitopes. A11 binds known repeating antiparallel beta structures, such as antiparallel Aß42 prefibrillar oligomers (26), hemolysin pores (14) and beta cylindrins (27) and antibodies in A11 serum also bind to α-sheet structures (28). OC binds to known parallel, in-register structures, such as Aß (12), α-synuclein and IAPP fibrils (7). In antiparallel ß sheets the strands alternate in opposite orientations, so it seems more likely that antibodies that are specific for antiparallel ß sheets would bind to discontinuous epitopes. Indeed, half of the A11 antibodies appear to bind to more than one Aß segment and the combinatorial nature of the discontinuous segments provides and additional layer of specificity to the antibodies with these epitopes. However, approximately half of the mOC series monoclonals also display discontinuous epitopes, so this feature alone does not explain specificity of the A11 and OC antibodies. If the OC antibodies bind to parallel ß sheets, then it would seem to indicate that these antibodies must bind to the ends of the beta sheet hair pin which is the only part of the fibril structure where both segments would be exposed for antibody binding.

The majority of the antibodies are specific for aggregation-related epitopes and do not bind Aß monomer or APP. The data indicate that these sequences that bind to these antibodies have a significantly higher predicted propensity to aggregate than the random sequences as a whole or the sequences preferred by the antibodies that bind to Aß monomer. This suggests that the random sequence displayed by the phage pIII protein may be aggregated even though there are a relatively low number of copies of the pIII protein. It is interesting that many of the antibodies that bind monomeric Aß show a lower than average propensity to aggregate. It isn't clear whether this means that they are more likely to be unstructured which facilitates the induced fit in the antigen combining site or whether they adopt a common non-aggregating structure.

We also found that the vast majority of the antibodies display a unique pattern of amino acid residues that is specific to the antibody and is most likely determined by the unique structure of the complementary paratope. This could be very useful to serve as a fingerprint to identify the antibody in complex mixtures of antibodies, such as in human blood or CSF. This would allow the identification

of antibodies that could be protective for disease or serve as biomarkers of any immune related disease, including infectious disease, autoimmune disease and cancer. For AD, it could enable the identification of disease subtypes and precision medicine approaches to AD immunotherapy.

Experimental procedures

Phage display

Twenty eight rabbit monoclonal antibodies consisting of 5 antibodies derived from A11 serum and 23 derived from OC serum were produced and characterized as previously described (8,9). One µg of each antibody was used to pan a dodecapeptide library with a complexity of 10^9 individual clones expressed in the coat protein pIII of the M13 phage according to the manufacturer's recommended protocol (PhD Phage display library, NE BioLabs). Phage library without adding antibody was used to control for sequences binding only to the beads (Dynabeads Protein A, 10002D obtained from Novex Life Technologies). Three rounds of panning and amplification were carried out. Phage obtained at each step were collected and stored for analysis.

Phage DNA isolation and library preparation

Phage DNA was isolated using a standard phenol:chloroform method (29). Quality was assessed by visualization in a 1% agarose gel, and its concentration measured by spectrophotometry. One hundred ng of phage DNA were used as template for PCR amplification for the Next Generation Sequencing step. The phage DNA amplicons were barcoded, pooled and a 10 nM library was sequenced commercially on an Illumina MiSeq platform. (Laragen Inc, Culver City, California, USA).

Data analysis

The Illumina sequencing data was processed by a Bash script that extracts the DNA sequence coding for the dodecapeptides and translates them to protein and counts how many times each unique sequence was found in the sequencing file (frequency) and removes common background sequences due to unspecific binding to the protein A beads. The peptide sequences were sorted by frequency and converted to FASTA format as previously described (1). The FASTA sequences were written with an identifier label that contains the antibody name, the unique sequence number and the number of times the sequence was observed in the following pattern: >(antibody)*(sequence #)_(frequency) to make machine counting of the frequency easier. The sequences were analyzed to determine the amino acid sequence patterns they contain using Pratt 2.1 (15) that was edited and recompiled to accommodate up to 200,000 sequences (1).

Array analysis

For microarray fabrication, biotinylated amyloid peptides spanning the amyloid beta sequence (-5 to 45) were pre-incubated with NeutrAvidin (ThermoFisher Scientific, MA, USA) in phosphate-buffered saline (PBS) at molar ratio of 4:1 for 1 hr at room temperature with gentle agitation. Final concentration of peptides in the complex was 0.5 mg/ml. Tween 20 (T-PBS) and glycerol was added to the peptide complex solution at final concentration of 0.001%, and then the solution was spotted onto nitrocellulose coated glass AVID slides (Grace Bio-Labs, Inc., OR, USA) using an Omni Grid 100 contact microarray printer (Genomic Solutions). One nanoliter of peptide solution was delivered onto the membrane, corresponding approximately 0.5 ng of peptide per spot. Slides were stored in desiccator until use. Monoclonal antibodies were diluted in protein array blocking buffer (GVS, Sanford, ME, USA) to a final concentration of 5 ng/ml. Concurrently, arrays were rehydrated in blocking buffer for 30 min. Blocking buffer was removed, and arrays were probed with samples using sealed chambers to avoid cross-contamination between the pads. Arrays were incubated overnight at 4°C with gentle agitation. Arrays were washed three times with washing buffer, Tris-buffered saline (TBS) containing 0.05% Tween 20 (T-TBS), and bound antibodies were detected by Cy5-conjugated goat anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA, USA), diluted 1:200 in blocking buffer. After 1h at room temperature (RT) incubation, arrays were then washed three times with T-TBS and once with water. They were air dried by centrifugation at

500 × g for 10 min. Images were acquired using Perkin Elmer ScanArray Express HT confocal laser scanner at a wavelength of 670 nm and signal intensities were quantified using ProScanArray Express software (Perkin Elmer, Waltham, MA). All signal intensities were corrected for spot-specific background.

Fluorescence microscopy

Alzheimer human brain sample from frontal cortex was sectioned into 40 micrometer sections using a Vibratome. Antibody mOC22 was biotinylated with NHS-biotin (ThermoFisher EZ-Link Catalog # 20217) according to the manufacturer's instructions. The sections were washed in Tris-buffered saline (TBS) three times for five minutes each. The sections were incubated overnight at room temperature in mOC23 at a concentration of 0.1 μg/ml in TBS containing 1% BSA and then washed 3 times in TBS. The sections were incubated with goat anti rabbit IgG labeled with AlexaFluor 647 (A32728, Thermo Fisher Scientific), 1 μg/ml in TBS with 1% goat serum for 1 hr. Sections were washed 3 times in TBS for 5 minutes and incubated in TBS containing 1% rabbit serum for 30 minutes. The sections were then incubated overnight in 0.1 μg/ml biotnylated mOC22 and then washed 3 times in TBS, incubated in TBS-1% BSA and incubated with 1 μg/ml AlexaFluor 488 streptavidin (ThermoFisher S21374)for 1 hr, then washed 3 times with TBS and mounted on a slide and imaged with a Zeiss LSM 700 laser confocal microscope.

Acknowledgments
Authors wish to thank Dr. Jeffrey Glabe for help with modifying and recompiling PRATT, Mr. Matt Gargus, Mr. Noe Neira Garza and Ms. Dania Alvarez for helping to write the script to extract the peptide sequences from the sequencing files.

Funding and additional information
This work was supported by NIH grants AG056507 and AG061857 (CG), and the Cure Alzheimer Fund (CG).

Conflict of interest: The authors declare no conflicts of interest.

References

1. Baghallab, I., Reyes-Ruiz, J. M., Abulnaja, K., Huwait, E., and Glabe, C. (2018) Epitomic Characterization of the Specificity of the Anti-Amyloid Abeta Monoclonal Antibodies 6E10 and 4G8. *J Alzheimers Dis* **66**, 1235-1244
2. Eisenberg, D. S., and Sawaya, M. R. (2017) Structural Studies of Amyloid Proteins at the Molecular Level. *Annu Rev Biochem* **86**, 69-95
3. Glabe, C. G. (2008) Structural classification of toxic amyloid oligomers. *J Biol Chem* **283**, 29639-29643
4. Dobson, C. M. (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**, 329-332
5. Glabe, C. G. (2004) Conformation-dependent antibodies target diseases of protein misfolding. *Trends Biochem Sci* **29**, 542-547
6. Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486-489
7. Kayed, R., Head, E., Sarsoza, F., Saing, T., Cotman, C. W., Necula, M., Margol, L., Wu, J., Breydo, L., Thompson, J. L., Rasool, S., Gurlo, T., Butler, P., and Glabe, C. G. (2007) Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. *Mol Neurodegener* **2**, 18
8. Kayed, R., Canto, I., Breydo, L., Rasool, S., Lukacsovich, T., Wu, J., Albay, R., 3rd, Pensalfini, A., Yeung, S., Head, E., Marsh, J. L., and Glabe, C. (2010) Conformation dependent monoclonal antibodies distinguish different replicating strains or conformers of prefibrillar Abeta oligomers. *Mol Neurodegener* **5**, 57
9. Hatami, A., Albay, R., 3rd, Monjazeb, S., Milton, S., and Glabe, C. (2014) Monoclonal antibodies against Abeta42 fibrils distinguish multiple aggregation state polymorphisms in vitro and in Alzheimer disease brain. *J Biol Chem* **289**, 32131-32143
10. Kayed, R., Head, E., Sarsoza, F., Saing, T., Cotman, C. W., Necula, M., Margol, L., Wu, J., Breydo, L., Thompson, J. L., Rasool, S., Gurlo, T., Butler, P., and Glabe, C. G. (2007) Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. *Mol Neurodegener* **2**, 18
11. Lasagna-Reeves, C. A., Glabe, C. G., and Kayed, R. (2011) Amyloid-beta annular protofibrils evade fibrillar fate in Alzheimer disease brain. *J Biol Chem* **286**, 22122-22130
12. Walti, M. A., Ravotti, F., Arai, H., Glabe, C. G., Wall, J. S., Bockmann, A., Guntert, P., Meier, B. H., and Riek, R. (2016) Atomic-resolution structure of a disease-relevant Abeta(1-42) amyloid fibril. *Proc Natl Acad Sci U S A*

13. Kayed, R., Pensalfini, A., Margol, L., Sokolov, Y., Sarsoza, F., Head, E., Hall, J., and Glabe, C. (2009) Annular protofibrils are a structurally and functionally distinct type of amyloid oligomer. *J Biol Chem* **284**, 4230-4237
14. Yoshiike, Y., Kayed, R., Milton, S. C., Takashima, A., and Glabe, C. G. (2007) Pore-forming proteins share structural and functional homology with amyloid oligomers. *Neuromolecular Med* **9**, 270-275
15. Jonassen, I., Collins, J. F., and Higgins, D. G. (1995) Finding flexible patterns in unaligned protein sequences. *Protein Sci* **4**, 1587-1595
16. Reineke, U., and Sabat, R. (2009) Antibody epitope mapping using SPOT peptide arrays. *Methods Mol Biol* **524**, 145-167
17. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X., and Ventura, S. (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65
18. Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., Morris, K. L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J. W., and Rousseau, F. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* **7**, 237-242
19. Goldschmidt, L., Teng, P. K., Riek, R., and Eisenberg, D. (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci U S A* **107**, 3487-3492
20. Pensalfini, A., Albay, R., 3rd, Rasool, S., Wu, J. W., Hatami, A., Arai, H., Margol, L., Milton, S., Poon, W. W., Corrada, M. M., Kawas, C. H., and Glabe, C. G. (2014) Intracellular amyloid and the neuronal origin of Alzheimer neuritic plaques. *Neurobiol Dis* **71C**, 53-61
21. McLean, D., Cooke, M. J., Albay, R., 3rd, Glabe, C., and Shoichet, M. S. (2013) Positron Emission Tomography Imaging of Fibrillar Parenchymal and Vascular Amyloid-beta in TgCRND8 Mice. *ACS Chem Neurosci* **4**, 613-623
22. Abbott, W. M., Damschroder, M. M., and Lowe, D. C. (2014) Current approaches to fine mapping of antigen-antibody interactions. *Immunology* **142**, 526-535
23. Hatami, A., Monjazeb, S., Milton, S., and Glabe, C. G. (2017) Familial Alzheimer's Disease Mutations within the Amyloid Precursor Protein Alter the Aggregation and Conformation of the Amyloid-beta Peptide. *J Biol Chem* **292**, 3172-3185
24. Gallardo, R., Ranson, N. A., and Radford, S. E. (2019) Amyloid structures: much more than just a cross-beta fold. *Curr Opin Struct Biol* **60**, 7-16
25. Fandrich, M., Nystrom, S., Nilsson, K. P. R., Bockmann, A., LeVine, H., 3rd, and Hammarstrom, P. (2018) Amyloid fibril polymorphism: a challenge for molecular imaging and therapy. *J Intern Med* **283**, 218-237
26. Cerf, E., Sarroukh, R., Tamamizu-Kato, S., Breydo, L., Derclaye, S., Dufrene, Y. F., Narayanaswami, V., Goormaghtigh, E., Ruysschaert, J. M., and Raussens, V. (2009) Antiparallel beta-sheet: a signature structure of the oligomeric amyloid beta-peptide. *Biochem J* **421**, 415-423
27. Laganowsky, A., Liu, C., Sawaya, M. R., Whitelegge, J. P., Park, J., Zhao, M., Pensalfini, A., Soriaga, A. B., Landau, M., Teng, P. K., Cascio, D., Glabe, C., and Eisenberg, D. (2012) Atomic view of a toxic amyloid small oligomer. *Science* **335**, 1228-1231
28. Shea, D., Hsu, C. C., Bi, T. M., Paranjapye, N., Childers, M. C., Cochran, J., Tomberlin, C. P., Wang, L., Paris, D., Zonderman, J., Varani, G., Link, C. D., Mullan, M., and Daggett, V. (2019) alpha-Sheet secondary structure in amyloid beta-peptide drives aggregation and toxicity in Alzheimer's disease. *Proc Natl Acad Sci U S A* **116**, 8895-8900
29. Sambrook, J., and Russell, D. W. (2006) Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc* **2006**

Figures and Tables

Figure 1.

Figure 1. Map of antibody binding sites on Aß. The binding sites identified in Tables III and IV are shown as color coded bars. The major epitope is shown by bars above the Aß sequence and the minor epitope for antibodies that bind to more than one Aß segment is shown by the bars below the Aß sequence. No binding site is indicated for mA11-89 because it does not bind to Aß.

Figure 2

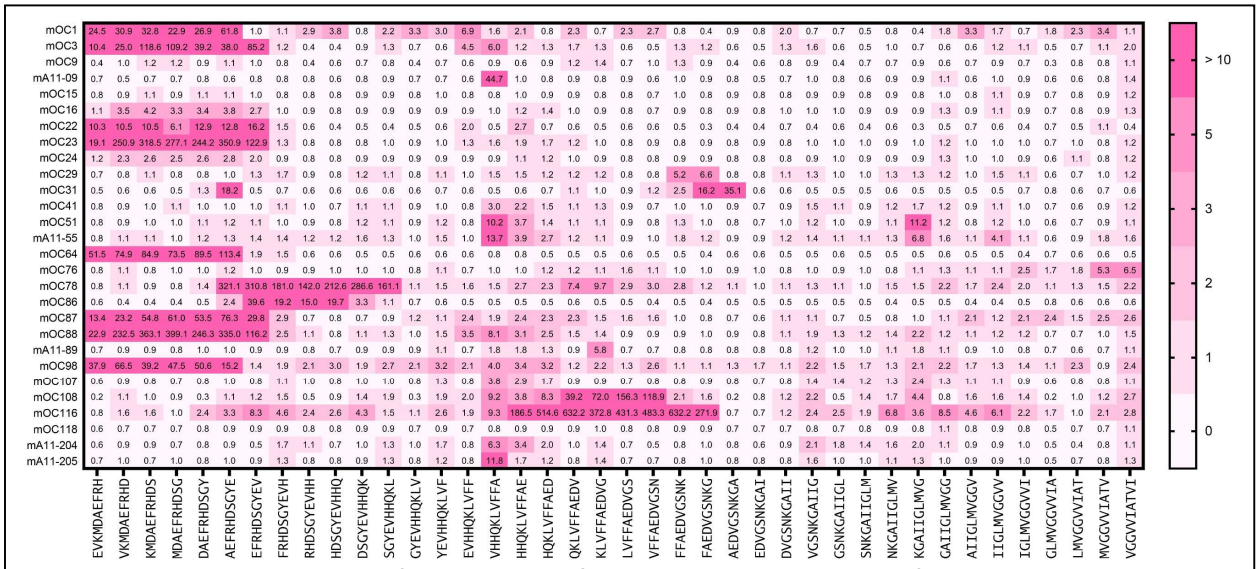

Figure 2. Microarray analysis of overlapping Aß peptide 10mer sequences from -4 to 45 as indicated along the bottom of the figure. The spot intensity is indicated by pink to red boxes containing the fold increase above background of the binding.

Figure 3

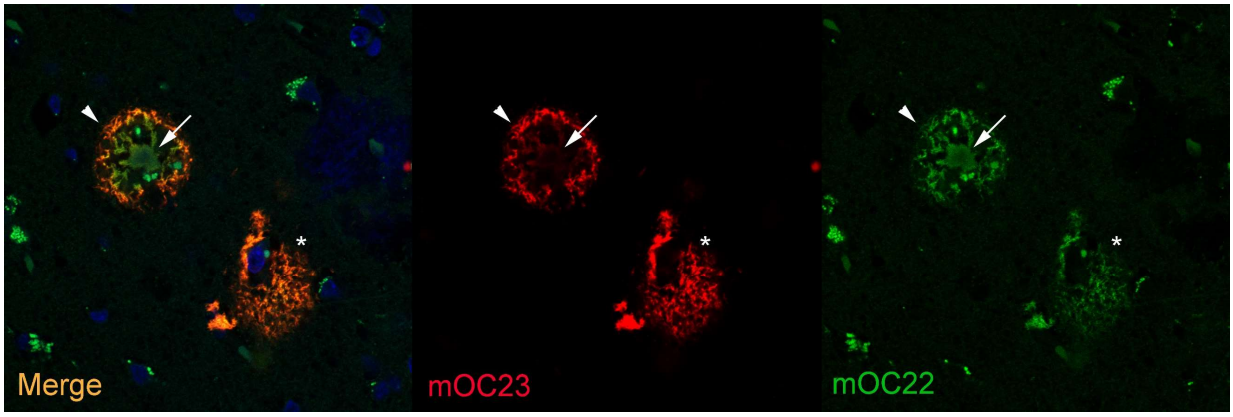

Figure 3. Differential staining of plaques in human AD brain by mOC22 (green) and mOC23 (red). mOC22 stains the entire cored plaque (arrow) while mOC23 only stains the rim (arrow head). A diffuse plaque is stained by both antibodies (asterix). The cored plaque is approximately 50 micrometers in diameter.

Table I
Effect of C% on patterns

mOC1 C% = 75%

	fitness	hits	Pattern
1:	8.34	1148	E-x-R
2:	8.34	1192	A-E

mOC1 C% = 50%

1:	12.51	895	A-E-x-R
2:	11.52	744	A-E-[FY]
3:	11.09	837	E-[FWY]-R
4:	10.93	726	D-[AGS]-E

mOC1 C% = 20%

1:	16.68	351	A-E-F-R
2:	16.68	364	D-A-E-x-R
3:	15.77	321	D-A-E-[FL]
4:	15.68	295	S-[AS]-E-x-R

mOC1 C% = 10%

1:	19.86	168	S-A-E-[FY]-R
2:	19.85	161	S-[AS]-E-F-R
3:	19.55	155	D-A-E-[FLM]-R
4:	19.27	146	D-[AGS]-E-F-R

Unselected random sequences C% = 1

A	1:	12.51	178(178)	P-P-S
B	2:	12.51	180(179)	P-S-x-S
C	3:	12.51	194(193)	P-S-x-P
D	4:	12.51	193(193)	T-L-S

Unselected random sequences C% = 0.1

A	1:	16.68	18(18)	S-P-x(4)-S-P
B	2:	16.68	29(29)	S-L-L-P
C	3:	16.68	20(20)	S-x(3)-S-x-L-P
D	4:	16.68	18(18)	S-S-x(3)-S-x-L

Table II Major epitope patterns observed for 28 monoclonal antibodies.

mOC1 fitness hits Sequence 16.68 1296 D-A-E-x-R 16.68 1278 A-E-F-R 16.68 855 D-A-E-F	mOC3 fitness hits Sequence 16.68 702 E-F-R-H 15.69 676 P-[EQ]-x-R-H 12.51 1505 E-x-R-H	mOC9 C% = 0.5 fitness hits Sequence 17.15 50 L-[ANSTV]-F-[FWY]-A 16.71 51 L-[ASTV]-[FW]-[FWY]-A 16.68 56 L-x-F-F-A	mA11-09 C% = 1 fitness hits Sequence 15.69 111 A-E-[FY]-R 15.13 93 D-A-A-[ADS]* 15.10 92 D-[AGS]-E-x-R
mOC15 C% = 1 fitness hits Sequence 15.75 50 A-E-F-[KR] 15.69 60 A-E-[FY]-R 15.69 50 A-[EQ]-F-R	mOC16 fitness hits Sequence 18.97 429 P-E-[FHWY]-R-H 16.68 598 E-F-R-H 15.26 428 P-E-[FWY]-x-H	mOC22 fitness hits Sequence 19.86 817 D-A-E-[FY]-R 18.54 752 D-[AGSTV]-E-F-R 16.68 113 A-E-F-R	mOC23 fitness hits Sequence 14.8 791 A-E-[FHWY]-R 12.51 898 A-E-x-R 12.51 792 A-x-F-R
mOC24 fitness hits Sequence 20.85 482 P-E-F-R-H 19.43 448 S-x-E-[FWY]-R-H 19.29 468 S-[APS]-E-x-R-H	mOC29 C% = 0.5 fitness hits Sequence 16.68 62 A-E-F-R 14.8 53 E-[FHWY]-R-H 12.51 91 A-E-x-R	mOC31 C% = 0.2 fitness hits Sequence 18.97 20 P-E-[FHWY]-R-H 16.68 26 E-F-R-H 16.68 22 P-E-x-R-H	mOC41 C% = 1 fitness hits Sequence 16.68 24 A-E-F-R 15.26 24 E-[FHW]-R-H 15.09 25 D-A-[DER]-x-R
mOC51 C% = 2 fitness hits Sequence 20.85 7 A-E-F-R-H 20.85 7 P-E-F-R-H 20.85 5 S-P-E-x-R-H	mA11-55 C% = 2 fitness hits Sequence 18.56 54 L-x-F-F-A-[ADEST] 17.87 52 L-[ATV]-F-[FWY]-A 16.74 50 L-[GST]-[ADEGST]-R-S	mOC64 C% = 1 fitness hits Sequence 20.11 159 A-E-[FW]-x(2)-G-G 19.31 174 G-A-E-W-[NRS] 19.31 162 A-E-W-[NRS]-S	mOC76 fitness hits Sequence 1: 15.7 530 G-[IV]-V-I 2: 15.7 579 G-I-[IV]-I 3: 15.1 481 G-[ILV]-x-I-x-S
mOC78 fitness hits Sequence 16.68 1079 S-G-Y-E 15.11 965 S-G-F-[EQT] 14.38 1022 H-[AGLPV]-S-G	mOC86 fitness hits Sequence 14.71 1213 H-[ALPV]-S-G 12.51 1344 S-G-x-E 12.51 1210 S-G-Y	mOC87 fitness hits Sequence 16.68 780 E-F-R-H 15.69 740 P-[EQ]-x-R-H 13.81 683 P-[EQ]-[FHWY]-R	mOC88 C% = 5 fitness hits Sequence 15.11 345 D-x-R-H-[DST] 14.38 331 G-[DENST]-x-R-H 12.51 329 F-R-H
mA11-89 C% = 1 fitness hits Sequence 32 105 H-L-[AD]-C-A-[ADGV]-C-R 30 108 H-L-[AD]-C-A-x-C-R-[FLM] 30 104 L-D-C-A-[ADV]-C-R-[LMV]	mOC98 fitness hits Sequence 16.68 576 A-E-F-R 15.69 533 S-G-[FY]-E 15.68 538 D-[AS]-E-x-R	mOC107 C% = 1 fitness hits Sequence 15.72 182 A-E-F-[RS] 15.69 232 A-E-[FY]-R 15.68 186 D-[AS]-E-x-R	mOC108 C% = 1 fitness hits Sequence 16.68 168 I-V-F-P * 16.68 113 L-V-F-P * 15.73 119 V-F-P-[AP] *
mOC116 fitness hits Sequence 16.68 1253 F-A-E-D 12.51 2158 F-A-E 12.51 1387 F-x-E-D	mA11-118 C% = 0.1 fitness hits Sequence 15.94 517 S-A-E-[FW] 15.94 448 S-x-A-E-[FW] 15.94 410 A-E-[FW]-S	mA11-204 C% = 1 fitness hits Sequence 15.69 305 A-E-[FY]-R 13.93 266 A-[DEGNQST]-F-R 12.51 335 E-F-R	mA11-205 C% = 0.2 fitness hits Sequence 16.68 102 A-E-F-R 16.68 101 E-F-R-H 15.17 102 V-V-[ILV]-x-S

The 3 highest fitness Pratt patterns are shown. Positions denoted by x can be any amino acid. Positions denoted by square brackets denote the preferred amino acids at wild card positions. C% = 10 was used unless a value for C% is shown. * denotes patterns where a strong position effect is observed with the pattern beginning at the first residue of the sequence.

Table III Antibodies that recognize more than 1 non-overlapping segment of Aß

mOC3 C% = 1 Rank Fitness Hits Sequence 1 16.68 702 E-F-R-H 103 20.85 65 F-F-A-E-D 250 16.68 109 F-F-A-E	mOC9 C% = 1 Rank Fitness Hits Sequence 8 26.23 5 L-[TV]-F-F-A-E-[DQRS] 85 20.85 10 D-A-E-F-R 86 20.85 6 D-A-E-Y-R	mOC29 C% = 1 Rank Fitness Hits Sequence 1 19.43 37 D-A-E-[FWY]-R 5 15.69 46 L-x-F-[FY]-A 8 15.68 35 L-x-F-F-[AS]	mOC31 C% = 0.1 Rank Fitness Hits Sequence 1 22.05 10 S-[APS]-E-[FWY]-R-H 331: 15.11 10 V-G-G-[ATV] 356: 15.10 11 A-x-D-[ALV]-G
mOC41 C% = 1 Rank Fitness Hits Sequence 1 16.68 25 A-E-F-R 17 14.80 25 S-G-[FHWY]-E 27 14.15 25 S-G-[FY]-[AET]	mOC51 C% = 1 Rank Fitness Hits Sequence 25 28.19 2 A-E-F-R-H-D-[SV] 46 20.85 3 E-F-R-H-D 50 20.85 3 S-G-Y-E-V	mA11-55 C% = 1 Rank Fitness Hits Sequence 202 19.85 28 L-[TV]-F-F-A 290 15.10 23 H-x-S-G-[FHY] 310 12.51 23 S-G-x-E	mOC64 c% = 1 Rank Fitness Hits Sequence 55 16.68 201 A-E-F-x-H 59 16.68 350 A-E-F-R 378 12.51 177 S-G-x-E
mOC76 c% = 1 Rank Fitness Hits Sequence 7 24.07 48 G-[IV]-V-I-A-S 194 15.69 55 E-[FY]-R-H 303 12.51 77 E-x-R-H	mOC78 C% = 1 Rank Fitness Hits Sequence 8 20.85 136 S-G-Y-E-V 150 16.68 1080 S-G-Y-E 437 12.51 102 V-G-S	mOC86 Rank Fitness Hits Sequence 16 19.86 178 H-x-S-G-[FY]-E 21 19.86 146 S-G-Y-E-[LV] 402 10.93 129 V-G-[AGS]	mOC87 Rank Fitness Hits Sequence 21 20.85 185 A-E-F-R-H 57 19.86 75 E-[FY]-R-H-D 357 12.51 71 S-G-x-E
mOC98 C% = 1 Rank Fitness Hits Sequence 1 16.68 626 A-E-F-R 2 15.69 557 S-G-[FY]-E 4 15.10 550 H-x-S-G-[FHY]	mOC108 C% = 1 Rank Fitness Hits Sequence 1 16.68 113 L-V-F-P 2 16.68 168 I-V-F-P 66 11.52 133 S-G-[FY]	mA11-118 C% = 1 Rank Fitness Hits Sequence 20 15.94 229 A-E-[FW]-x(3)-S 114 12.51 182 L-x(3)-A-E 204 11.53 183 F-A-[ES]	mA11-204 c% = 0.3 Rank Fitness Hits Sequence 1 19.86 100 D-A-E-[FY]-R 76 12.51 105 V-G-S 381 12.51 90 G-S-N

Table IV Summary of discontinuous epitopes

	PepSpots	Epitomics	Array
mOC1	+	-	+
mOC3	-	+	+
mOC9	-	-	-
mA11-09	-	+	-
mOC15	-	-	-
mOC16	-	-	-
mOC22	-	-	-
mOC23	-	-	-
mOC24	+	-	-
mOC29	-	+	-
mOC31	+	+	+
mOC41	-	+	-
mOC51	+	+	+
mA11-55	+	+	+
mOC64	-	+	-
mOC76	-	+	-
mOC78	+	+	+
mOC86	-	+	-
mOC87	-	+	-
mOC88	-	-	+
mA11-89	-	-	-
mOC98	+	+	+
mOC107	-	-	-
mOC108	-	+	-
mOC116	-	-	+
mA11-118	-	+	-
mA11-204	-	+	-
mA11-205	-	-	-

+ indicates a discontinuous epitope

Table V Aggregation propensity of immunoselected peptides.

Name	Waltz % hits	AggScan % hits	3D % hits	Average
RandomSeq	32	28		
BSA	41.5	31.6		
APP771	35.8	26.5		
Synuclein	25.7	51.4		
mOC1	3	12	11	8.7
mOC3	29	13	5	15.7
mA11-09	52	36	15	34.3
mOC9	63	50	23	45.3
mOC15	42	36	15	31.0
mOC16	23	2	7	10.7
mOC22	33	20	11	21.3
mOC23	25	13	3	13.7
mOC24	16	5	3	8.0
mOC29	58	47	28	44.3
mOC31	38	29	19	28.7
mOC41	54	47	32	44.3
mOC51	51	32	9	30.7
mA11-55	46	34	23	34.3
mOC64	58	22	11	30.3
mOC76	69	80	58	69.0
mOC78	54	34	25	37.7
mOC86	51	44	29	41.3
mOC87	25	5	5	11.7
mOC88	20	9	4	11.0
mA11-89	66	72	42	60.0
mOC98	22	13	9	14.7
mOC107	44	40	17	33.7
mOC108	51	40	9	33.3
mOC116	65	43	9	39.0
mA11-118	36	33	16	28.3
mA11-204	46	36	16	32.7
mA11-205	51	43	20	38.0
4G8	53	79	21	51.0
6E10	23	7	4	11.3

Supplemental Information

Appendix I

Appendix Table I
Nucleotide Frequency

Nucleotide	Percent
G	26.39
A	18.15
T	34.24
C	21.22

Appendix Table II Amino Acid Frequency

Amino Acid	Frequency	Amino Acid	Frequency
S	.109	D	.039
P	.103	Y	.038
L	.092	Q	.036
T	.092	I	.035
A	.069	F	.031
H	.052	M	.029
R	.052	E	.027
V	.050	K	.025
N	.048	W	.019
G	.042	C	.010

Bash script for processing sequencing files.

```
#unzips any file with a .gz at end
for file in /data/users/UserName/UserDirectory/job_trials/*.gz; do
gunzip ${i};done

#Name A11
for file in 1_* 2_* 3_*; do mv "$file" "${file/S*_/A11.fastq}"; done

#Name OC
for file in 21_* 24_* 14_*; do mv "$file" "${file/S*_/OC.fastq}"; done

#Name Lock
for file in 15_* 20_*; do mv "$file" "${file/S*_/Lock.fastq}"; done

#name group Beads
for file in 96_* 97_* 98_*; do mv "$file" "${file/S*_/Bead.fastq}";
done

#Prinseq execute in Perl
for file in *.fastq; do prinseq-lite.pl -fastq ${file} -out_format
3;done

#removes all files EXCEPT files that have good in name from prinseq
output
find . -type f ! -name "*good*" -exec rm -rf {} \;

#search all files that have good and ends with .fastq for TTCTCACTCT
and next 36 basepairs
for file in *good*.fastq; do grep -o -P 'TTCTCACTCT.{0,36}' ${file}>
${file}_hit.txt; done

#removes all files EXCEPT .txt files in directory
find . -type f ! -name "*.txt" -exec rm -rf {} \;

#searches files that end with _hit.txt and removes TTCTCACTCT wherever
its found
for file in *_hit.txt; do sed "s/TTCTCACTCT//g" ${file} >
${file}_final_thirtysix.txt; done

#trims name of files with prinseq_good and ends with .txt and replaces
and renames anything after prinseq_good as dna_sequence
for file in *prinseq_good*.txt; do mv "$file"
"${file/prinseq_good*/dna_sequence}"; done

#for files in specified directory that end with dna_sequence run
python script translate.py and name the output the same with
_translated added
```

```
for file in
/data/users/UserName/UserDirectory/job_trials/*dna_sequence; do python
/data/users/UserName/UserDirectory/translate.py
${file}>${file}_translated; done

#removes all files EXCEPT ones with translated at the end
find . -type f ! -name "*translated" -exec rm -rf {} \;

#
find /data/users/UserName/UserDirectory/job_trials -type f -exec sed -
i 's/*/X/g' {} \;

#
find /data/users/UserName/UserDirectory/job_trials/ -type f -exec sed
-i '/.\{12\}/!d' {} \;

#THIS NEEDS ATTENTION! Bead background files need to be designated
input!!!
for file in *Bead* ; do cat "$file" >> Bead_background.txt && rm
"$file" || break ; done

#Remove/replaces lines that appear in bead background file and removes
from files with translated in name
for file in *translated ; do grep -vxFf *background.txt ${file}>
${file}_bead_filter.txt; done

#These lines provide individual files ranked sorted counted and
labeled
for i in *bead_filter.txt; do
cat ${i} | sort | uniq -c | sort -nr >> ${i}_bead_ranked.txt;
done
for i in *dna_sequence_translated_bead_filter.txt_bead_ranked.txt; do
g=$(basename $i
dna_sequence_translated_bead_filter.txt_bead_ranked.txt);
sed -e "s/[0-9]*//g" ${i} | sed -e "s/ //g" > ${g}_peptides.txt
awk -v g="$g" '{ print ">"g"_peptide_"NR"_"$1$0}' ${i} | awk '{print
$1}' > ${g}_headers.txt;
paste -d '\n' ${g}_headers.txt ${g}_peptides.txt > ${g}_complete.txt
rm ${g}_peptides.txt
rm ${g}_headers.txt
done

#concatenates all A11 files
cat *A11* > A11_all.txt

#concatenates all OC files
cat *OC* > OC_all.txt

#concatenates all Lock files
cat *Lock* > Lock_all.txt
```

```
#aruably the most rigorous lines. This section takes the all.txt files
creates a low counts file for peptides that appear less than 5 and
then also make a file for lines in each file that appear 5 or more
time. This then creates some temporary files to create a .fasta file
with the peptides labeled, ranked, with count in the header.
for i in *all.txt; do
cat ${i} | sort | uniq -c | sort -nr >> ${i}_sorted_ranked.txt;
done
for i in *_sorted_ranked.txt; do
awk '($1 > 4 ) ' ${i}>${i}_five.txt;
done
for i in *_sorted_ranked.txt; do
awk '($1 < 5 ) ' ${i}>${i}_low_counts.txt
done
for i in *_all.txt_sorted_ranked.txt_five.txt; do
g=$(basename $i _all.txt_sorted_ranked.txt_five.txt);
sed -e "s/[0-9]*//g" ${i} | sed -e "s/ //g" > ${g}_peptides.txt
awk -v g="$g" '{ print ">"g"_peptide_"NR"_"$1$0}' ${i} | awk '{print
$1}' > ${g}_headers.txt;
paste -d '\n' ${g}_headers.txt ${g}_peptides.txt > ${g}_final.fasta
rm ${g}_peptides.txt
rm ${g}_headers.txt
done



#continuous

module load enthought_python/7.3.2
module load prinseq-lite/0.20.4
for file in /data/users/UserName/UserDirectory/job_trials/*.gz; do
gunzip ${file};done
for file in 1_* 2_* 3_*; do mv "$file" "${file/S*_/A11.fastq}"; done
for file in 21_* 24_* 14_*; do mv "$file" "${file/S*_/OC.fastq}"; done
for file in 15_* 20_*; do mv "$file" "${file/S*_/Lock.fastq}"; done
for file in 96_* 97_* 98_*; do mv "$file" "${file/S*_/Bead.fastq}";
done
for file in *.fastq; do prinseq-lite.pl -fastq ${file} -out_format
3;done
find . -type f ! -name "*good*" -exec rm -rf {} \;
for file in *good*.fastq; do grep -o -P 'TTCTCACTCT.{0,36}' ${file}>
${file}_hit.txt; done
find . -type f ! -name "*.txt" -exec rm -rf {} \;
for file in *_hit.txt; do sed "s/TTCTCACTCT//g" ${file} >
${file}_final_thirtysix.txt; done
for file in *prinseq_good*.txt; do mv "$file"
"${file/prinseq_good*/dna_sequence}"; done
for file in
/data/users/UserName/UserDirectory/job_trials/*dna_sequence; do python
/data/users/UserName/UserDirectory/translate.py
${file}>${file}_translated; done
find . -type f ! -name "*translated" -exec rm -rf {} \; done
```

```
find /data/users/UserName/UserDirectory/job_trials -type f -exec sed -
i 's/*/X/g' {} \;
find /data/users/UserName/UserDirectory/job_trials/ -type f -exec sed
-i '/.\{12\}/!d' {} \;
for file in *Bead*translated ; do cat "$file" >> Bead_background.txt
&& rm "$file" || break ; done
rm *Bead*good*
for file in *translated ; do grep -vxFf *background.txt ${file}>
${file}_bead_filter.txt; done
for i in *bead_filter.txt; do
cat ${i} | sort | uniq -c | sort -nr >> ${i}_bead_ranked.txt;
done
for i in *dna_sequence_translated_bead_filter.txt_bead_ranked.txt; do
g=$(basename $i
dna_sequence_translated_bead_filter.txt_bead_ranked.txt);
sed -e "s/[0-9]*//g" ${i} | sed -e "s/ //g" > ${g}_peptides.txt
awk -v g="$g" '{ print ">"g"_peptide_"NR"_"$1$0}' ${i} | awk '{print
$1}' > ${g}_headers.txt;
paste -d '\n' ${g}_headers.txt ${g}_peptides.txt > ${g}_complete.txt
rm ${g}_peptides.txt
rm ${g}_headers.txt
done
cat *A11*_bead_filter.txt > A11_all.txt
cat *OC*_bead_filter.txt > OC_all.txt
cat *Lock*_bead_filter.txt > Lock_all.txt
for i in *all.txt; do
cat ${i} | sort | uniq -c | sort -nr >> ${i}_sorted_ranked.txt;
done
for i in *_sorted_ranked.txt; do
awk '($1 > 4 ) ' ${i}>${i}_five.txt;
done
for i in *_sorted_ranked.txt; do
awk '($1 < 5 ) ' ${i}>${i}_low_counts.txt
done
for i in *_all.txt_sorted_ranked.txt_five.txt; do
g=$(basename $i _all.txt_sorted_ranked.txt_five.txt);
sed -e "s/[0-9]*//g" ${i} | sed -e "s/ //g" > ${g}_peptides.txt
awk -v g="$g" '{ print ">"g"_peptide_"NR"_"$1$0}' ${i} | awk '{print
$1}' > ${g}_headers.txt;
paste -d '\n' ${g}_headers.txt ${g}_peptides.txt > ${g}_final.fasta
rm ${g}_peptides.txt
rm ${g}_headers.txt
done
```