Conformation-Dependent Epitopes Recognized by Prion Protein Antibodies Probed Using Mutational Scanning and Deep Sequencing

Kyle M. Doolan and David W. Colby

Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19716, USA

Correspondence to David W. Colby: Department of Chemical and Biomolecular Engineering, University of Delaware, 150 Academy Street, Newark, DE 19716, USA. colby@udel.edu

http://dx.doi.org/10.1016/j.jmb.2014.10.024

Edited by S. Sidhu

Abstract

Prion diseases are caused by a structural rearrangement of the cellular prion protein, PrP\(^{C}\), into a disease-associated conformation, PrP\(^{Sc}\), which may be distinguished from one another using conformation-specific antibodies. We used mutational scanning by cell-surface display to screen 1341 PrP single point mutants for attenuated interaction with four anti-PrP antibodies, including several with conformational specificity. Single-molecule real-time gene sequencing was used to quantify enrichment of mutants, returning 26,000 high-quality full-length reads for each screened population on average. Relative enrichment of mutants correlated to the magnitude of the change in binding affinity. Mutations that diminished binding of the antibody ICSM18 represented the core of contact residues in the published crystal structure of its complex. A similarly located binding site was identified for D18, comprising discontinuous residues in helix 1 of PrP, brought into close proximity to one another only when the alpha helix is intact. The specificity of these antibodies for the normal form of PrP likely arises from loss of this conformational feature after conversion to the disease-associated form. Intriguingly, 6H4 binding was found to depend on interaction with the same residues, among others, suggesting that its ability to recognize both forms of PrP depends on a structural rearrangement of the antigen. The application of mutational scanning and deep sequencing provides residue-level resolution of positions in the protein–protein interaction interface that are critical for binding, as well as a quantitative measure of the impact of mutations on binding affinity.

© 2014 Elsevier Ltd. All rights reserved.

Introduction

The specificity of protein–protein interactions mediates many biological processes including receptor–ligand binding, protein signaling cascades, cell adhesion, and antibody recognition. Gene based approaches to investigate protein–protein interactions have made use of surface display technologies [1–4] in which proteins are expressed on the surface of a cell, virus, or ribosome, directly linking protein function with its underlying genetic basis. Surface display has been used to identify antibody binding epitopes through the identification of mutations that diminish binding [5–8]. In particular, yeast surface display has enabled the identification of conformational epitopes by taking advantage of yeast's eukaryotic protein folding machinery [6]. Deep sequencing has been used in conjunction with surface display approaches to identify peptide ligand binding sites of the WW domain [9], examine the protein fitness landscape of engineered influenza binding proteins [10], probe complementarity-determining-region biases in the isolation of poliovirus receptor antibodies [11], and explore CH3 domain stability [12]. In these studies, reversible dye-terminator deep sequencing technology was used to obtain a large number of sequences; however, the technology does not generate individual reads longer than 250 base pairs, requiring the construction of complex libraries and use of tandem reads to investigate whole proteins [13,14]. Alternatively, single-molecule real-time (SMRT) deep sequencing [15] permits investigation of open reading frames spanning more than 600 base pairs [16].

Prions cause fatal neurodegenerative diseases that may be transmissible, genetic, or sporadic in
etiology [17]. The underlying molecular basis for the pathogenesis of these diseases is the structural rearrangement of the cellular prion protein, PrP\(^{C}\), into a disease-associated conformation, PrP\(^{Sc}\). Conversion of PrP\(^{C}\) into PrP\(^{Sc}\) occurs via interaction of the two isoforms, resulting in self-propagation of PrP\(^{Sc}\). Structural characterization of recombinant PrP by NMR and X-ray crystallography suggests a structure consistent with that of PrP\(^{C}\) [18,19]. However, atomic-resolution characterization of the disease-associated prion conformation has been hampered by insolubility. Due to these difficulties, epitope accessibility has been used to gain insight into the structure of PrP\(^{Sc}\) [20–25]. While such approaches are critically dependent on accurate epitope definitions, studies on anti-PrP antibody epitopes have raised enigmatic questions [26–28]. For example, while D18, ICSM18, and 6H4 all bind epitopes in helix 1 by peptide technologies [20,29,30], only 6H4 is able to recognize PrP\(^{Sc}\) [22,28,31], although it does so with a lower binding affinity than it has for PrP\(^{C}\) and thus may still be considered conformation dependent. Recombinant prion amyloid fibers are recognized by D18 only after partial denaturation [32]. Recent investigations into the conformational specificity of D18 and 6H4 have provided insight into the roles of PrP secondary structure and disulfide bond formation [33].

We conjectured that yeast surface display and SMRT sequencing would enable characterization of conformational epitopes, which may require contacts at disparate locations along the entire length of PrP. When fully processed, PrP is a 208-amino-acid protein in humans and contains a single disulfide bond. We used yeast surface display and SMRT sequencing to identify epitope residues critical to ligand binding for four anti-PrP antibodies: D18 [30], ICSM18 [34], 6H4 [20], and EP1802Y. Exogenous expression of PrP in yeast produces protein that is post-translationally processed similarly to PrP\(^{C}\) in mammalian cells [35]. A mutant PrP library was constructed, sorted by fluorescence-activated cell sorting (FACS) for mutations diminishing antibody binding, and the responsible mutations were identified by SMRT sequencing. Statistical analysis of the genetic sequences identified critical epitope residues for each antibody and demonstrated detection of secondary-structure-dependent and tertiary-structure-dependent contacts.

Results and Discussion

Library construction and characterization

To display PrP on the surface of yeast, we subcloned the gene encoding mouse PrP for expression as a fusion to the yeast mating protein, Aga2p, with a c-myc tag at the C-terminus to verify full-length expression. Based on this construct, a library of PrP mutants was constructed by error-prone PCR amplification and transformed into yeast to produce a library containing \(8.9 \times 10^5\) total members. We performed SMRT sequencing on the 633-base-pair DNA fragments encoding the library in order to determine the frequency of mutations. Sequencing returned 31,001 high-quality reads containing 27,721 total nucleotide mutations. The sequencing error rate of \(5.90 \times 10^{-5}\) errors per nucleotide was found to be independent of nucleotide position (Supplementary Fig. 1), suggesting a small contribution of sequencing errors to the estimated depth of mutagenesis (<1%). Translation of the genetic sequences to their corresponding protein sequences gives rise to an effective protein mutation rate of 0.593 mutations per protein with a distribution that favors single mutation sequences over those with two or more mutations (Table 1). The mutations were found to be distributed about the entire length of the sequence (Supplementary Fig. 2) and demonstrated nucleotide substitution biases similar to published reports [36] (Supplementary Table 1).

Identification of PrP residues that modulate binding affinity

Genes encoding single-chain variable fragments (scFvs; Supplementary Table 2) of anti-PrP antibodies D18, ICSM18, and 6H4 were synthesized for recombinant expression by yeast as a secreted protein. Secreted scFvs bound native PrP expressed by N2a cells (Supplementary Fig. 3). Anti-PrP scFvs and EP1802Y, a commercially available anti-PrP Mab, were each incubated with the yeast surface display PrP library and binding was analyzed by flow cytometry (Fig. 1a). For each scFv, two major populations were observed: yeast that do not express PrP on their surface either due to plasmid loss or because the PrP contained mutations that made expression unfavorable, and yeast that express PrP with a sequence recognized by the anti-PrP scFvs. Clones appearing within the polygon labeled “Sort” express PrP containing mutations that reduce antibody affinity. This population was isolated by 2–4 successive rounds of FACS at concentrations greater than the wild-type \(K_d\) for each antibody (Supplementary Table 3). The isolated subpopulations exhibited reduced recognition by the scFvs (Fig. 1b). For some scFvs, the most PrP was captured with a lower binding affinity than for PrP\(^{C}\), although it does so with a lower binding affinity than it has for PrP\(^{C}\) and thus may still be considered conformation dependent. Recombinant prion amyloid fibers are recognized by D18 only after partial denaturation [32]. Recent investigations into the conformational specificity of D18 and 6H4 have provided insight into the roles of PrP secondary structure and disulfide bond formation [33].

We conjectured that yeast surface display and SMRT sequencing would enable characterization of conformational epitopes, which may require contacts at disparate locations along the entire length of PrP. When fully processed, PrP is a 208-amino-acid protein in humans and contains a single disulfide bond. We used yeast surface display and SMRT sequencing to identify epitope residues critical to ligand binding for four anti-PrP antibodies: D18 [30], ICSM18 [34], 6H4 [20], and EP1802Y. Exogenous expression of PrP in yeast produces protein that is post-translationally processed similarly to PrP\(^{C}\) in mammalian cells [35]. A mutant PrP library was constructed, sorted by fluorescence-activated cell sorting (FACS) for mutations diminishing antibody binding, and the responsible mutations were identified by SMRT sequencing. Statistical analysis of the genetic sequences identified critical epitope residues for each antibody and demonstrated detection of secondary-structure-dependent and tertiary-structure-dependent contacts.

Results and Discussion

Library construction and characterization

To display PrP on the surface of yeast, we subcloned the gene encoding mouse PrP for expression as a fusion to the yeast mating protein, Aga2p, with a c-myc tag at the C-terminus to verify full-length expression. Based on this construct, a library of PrP mutants was constructed by error-prone PCR amplification and transformed into yeast to produce a library containing \(8.9 \times 10^5\) total members. We performed SMRT sequencing on the 633-base-pair DNA fragments encoding the library in order to determine the frequency of mutations. Sequencing returned 31,001 high-quality reads containing 27,721 total nucleotide mutations. The sequencing error rate of \(5.90 \times 10^{-5}\) errors per nucleotide was found to be independent of nucleotide position (Supplementary Fig. 1), suggesting a small contribution of sequencing errors to the estimated depth of mutagenesis (<1%). Translation of the genetic sequences to their corresponding protein sequences gives rise to an effective protein mutation rate of 0.593 mutations per protein with a distribution that favors single mutation sequences over those with two or more mutations (Table 1). The mutations were found to be distributed about the entire length of the sequence (Supplementary Fig. 2) and demonstrated nucleotide substitution biases similar to published reports [36] (Supplementary Table 1).

Identification of PrP residues that modulate binding affinity

Genes encoding single-chain variable fragments (scFvs; Supplementary Table 2) of anti-PrP antibodies D18, ICSM18, and 6H4 were synthesized for recombinant expression by yeast as a secreted protein. Secreted scFvs bound native PrP expressed by N2a cells (Supplementary Fig. 3). Anti-PrP scFvs and EP1802Y, a commercially available anti-PrP Mab, were each incubated with the yeast surface display PrP library and binding was analyzed by flow cytometry (Fig. 1a). For each scFv, two major populations were observed: yeast that do not express PrP on their surface either due to plasmid loss or because the PrP contained mutations that made expression unfavorable, and yeast that express PrP with a sequence recognized by the anti-PrP scFvs. Clones appearing within the polygon labeled “Sort” express PrP containing mutations that reduce antibody affinity. This population was isolated by 2–4 successive rounds of FACS at concentrations greater than the wild-type \(K_d\) for each antibody (Supplementary Table 3). The isolated subpopulations exhibited reduced recognition by the scFvs (Fig. 1b). For

<table>
<thead>
<tr>
<th>Library size (members)</th>
<th>Number of mutations per member (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>48.5</td>
</tr>
<tr>
<td>1</td>
<td>27.7</td>
</tr>
<tr>
<td>2</td>
<td>12.6</td>
</tr>
<tr>
<td>3</td>
<td>4.7</td>
</tr>
<tr>
<td>≥4</td>
<td>6.51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Library size (members)</th>
<th>Number of mutations per member (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA 8.9 × 10^5</td>
<td>48.5</td>
</tr>
<tr>
<td>Protein 8.9 × 10^5</td>
<td>28.2</td>
</tr>
<tr>
<td>8.0</td>
<td>2.1</td>
</tr>
<tr>
<td>0.98</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Mutational frequency observed for PrP library.
ICSM18-scFv, D18-scFv, and Mab EP1802Y, less than 1% of cells with any binding at the sort concentrations were observed, while for 6H4-scFv, approximately 8% of cells had a binding profile comparable to wild-type PrP. Gel electrophoresis of digested fragments of the plasmids within the sorted populations revealed DNA fragments corresponding to the molecular mass of \( \text{Prnp} \), which was excised for sequencing, as well as a smaller fragment consistent with the insert of the parental vector.

Deep sequencing was performed on the \( \text{Prnp} \) gene fragments of the sorted populations to identify mutations that reduce antibody binding affinity. Sequencing yielded 19,000–30,000 high-quality reads for each sorted population (Table 2). The number of wild-type sequences remaining in the sorted population was used as an experimental measure of the sorting stringency and showed that the population sorted for diminished binding to ICSM18-scFv was the most stringently sorted (4.3% wild type), with other populations showing less stringent sorting (6H4-scFv, 7.6% wild type; Mab EP1802Y, 11.3% wild type; D18-scFv, 35.1% wild type). The number of sequences encoding a single mutation at each amino acid position was determined (Fig. 1c), and their enrichment relative to wild-type sequences was calculated (Fig. 1d). The enriched amino acid positions for both ICSM18-scFv and D18-scFv indicated that mutations of two residues in helix 1 predominantly disrupted binding, with modest enrichment of additional mutations to residues in helix 1 and helix 3 for D18-scFv. 6H4-scFv binding was disrupted by mutating residues in helix 1, previously identified as its epitope [20], as well as several residues in helix 3. Binding of the EP1802Y antibody was disrupted by mutations in a largely linear sequence of residues in helix 3 located at positions 217–226.

Sequences containing two or more mutations generally contained at least one of the deleterious mutations identified in the data containing a single mutation (Supplementary Table 4). Only in the case of ICSM18-scFv, which was the most stringently sorted
population, did the multiple mutations data identify additional substitutions that may impact binding, N151D (238 observations). The error rate of sequencing for the sorted populations was calculated at 3.83 × 10^-5 errors per nucleotide on average, corresponding to a Q-value of 47.9 and suggesting that 12% of apparent mutations actually arose from sequencing errors. However, the number of unique sequences returned exceeded the estimated number of clones in the sorted populations, suggesting that the sequencing error rate may have been higher than calculated. Based on these observations, we estimated the sequencing error to be 2.58 × 10^-4 errors per nucleotide, although we cannot exclude the possibility that these unique sequences arose from sources other than sequencing error. Even this higher estimate of the sequencing error rate is negligible compared to the error associated with chi-squared testing where multiple observations of a specific mutation were necessary to obtain statistical significance. However, if this rate of sequencing error was obtained in the initial library as well, our estimate of the number of true single point mutations would be reduced to 17.85%, providing 940-fold coverage rather than 1143-fold, which is more than sufficient for the current study.

Determination of contact residues from analysis of substitutions

The preceding analysis broadly identified epitope regions; however, the affinity could have been reduced upon mutation for reasons other than the loss of a critical interaction residue, including the introduction of unfavorable interactions and alternate folding of PrP secondary or tertiary structure. Analysis of the substitutions present at each position may help identify which is the case. We calculated the enrichment relative to wild type for all substitutions within the sorted populations based on the number of times a substitution was observed at a particular position in the sorted populations, the number of codons corresponding to the introduced amino acid arising from a single-nucleotide change, and the observed mutation rate. All substitutions that were enriched (p < 10^-5) are identified in Supplementary Tables 5–8.

Residues in contact with the antibody probed are likely to have diminished binding, and thus be enriched in the sorted populations, when mutated to most if not all substitutions present in the initial library. This characteristic was generally observed when examining those residues with the highest positional enrichment (Fig. 2a). In some instances, most available substitutions were enriched at a particular residue, while one or more substitutions available by a single-nucleotide change were not. We observed this for D201 and Y148 in the population sorted for diminished 6H4-scFv binding (Fig. 2b), where substitution of glutamic acid for aspartic acid at position 201 or substitution of phenylalanine for tyrosine at position 148 was not observed to be deleterious to 6H4-scFv binding, presumably due to physicochemical similarity of the side chains. Reduction in expression efficiency may also account for the reduction or absence of particular substitutions, as was the case for the D201A mutation (Supplementary Fig. 4), which was only moderately enriched compared to the majority of the other substitutions at D201.

The analysis of substitutions present also identifies substitutions that are not permissive to antibody recognition due to the introduction of repulsive interactions or changes to secondary or tertiary structure. At position E151 in the ICSM18-scFv sorted population, only E151K diminished binding (Fig. 2c), likely as a result of the net +2 charge change introducing repulsive interactions. All possible substitutions at E151 express sufficiently well to be observed for the 6H4-scFv sorted population, suggesting that the glutamic acid side chain at position 151 does not contribute substantially to affinity in the ICSM18-scFv binding under the conditions tested. Studies have demonstrated persistence of alpha helix 1 in PrP for a
variety of amino acid substitutions in this region \[37,38\],
including specifically the E151K substitution \[37\].
Substitutions to proline and glycine prove to be
exceptions, as these substitutions are known to break
alpha-helical structures based on allowable dihedral
angles \[39\]. We observed that mutations to proline and
glycine at R150 disrupted 6H4-scFv binding (Fig. 2d).
Finally, we also observed that mutations of several
hydrophobic core residues, including M212, reduced
6H4-scFv binding, consistent with the observation that
perturbation of hydrophobic core residues in PrP
disrupts tertiary structure \[40\].

In order to identify only potential contact residues,
we first identified residues demonstrating enrich-
ment in more than half of the substitutions arising
from a single-nucleotide mutation. We then surveyed
the literature on alpha-helical conformation tolerance
for the remaining residues to exclude those that
introduce structural changes. The F197S mutation is
associated with Gerstmann–Sträussler–Scheinker
syndrome \[41\], significantly reduces PrP stability
\[42\], and alters PrP folding \[43\]. For this reason, it
was excluded. Studies into the sensitivity of hydro-
phobic core residues to substitution \[44\] or oxidation
\[40\] revealed that modifications to core residues
produce a protein structure rich in alpha-helical
character but with reduced stability and an altered
orientation of residues in helix 1. We excluded the
previously identified hydrophobic core residues
Y156, V160, Y156, V209, M204, M205, and M212
based on these studies. Finally, we excluded proline
and glycine residues such as P157 and G227 based
on dihedral angle considerations \[39\].

Antibody binding affinity for mutants correlates
with enrichment

In order to experimentally validate our identifica-
tion of substitutions that modulate binding and
examine the relationship between the enrichment

**Fig. 2.** Substitution distributions of selected residues characterize residue interactions with antibody. The enrichment
for each particular substitution is reported for all substitutions arising from a single-nucleotide substitution. (a) Epitope
residues show enrichment of all or near-all substitutions arising from a single-nucleotide substitution consistent with the
side chain of these residues directly contacting paratope residues in the antibody. (b) Some epitope residues show
tolerance of particular substitution, demonstrated by the absence of these substitutions in the sequences obtained from the
sorted populations, as a result of physicochemical similarity of the side chains (c) Positions where only one or a few
substitutions show statistically significant enrichment can be the result of particular substitutions that introduce negative
interactions. (d) Mutations to proline or glycine at many residues were ablative likely by preventing alpha-helical secondary
structure to develop. Mutations to residues critical to protein folding also led to reduction in affinity.
of specific substitutions and contribution to binding affinity, we selected clones from the sorted libraries for further investigation. We first titrated the concentrations of ICSM18-scFv (Fig. 3a) and D18-scFv (Fig. 3b) incubated with wild type, W144S and N152D PrP expressed on the yeast surface in order to determine differences in binding affinity. We found that the W144S mutation resulted in ablation of the interaction with both D18-scFv and ICSM18-scFv with PrP over the concentration range tested, while the N152D again ablated D18-scFv binding over the same concentration range and reduced the affinity of ICSM18-scFv with a dissociation constant increase from 153 ± 67 nM to 3.22 ± 1.28 μM (95% confidence interval). This result is consistent with the observation of these substitutions in the isolated population, though N152D mutations were only observed in sequences with multiple mutations for ICSM18-scFv, highlighting the fact that increasing stringency reduces the number of residues identified by this type of analysis.

We then examined the affinity of 6H4-scFv for five PrP mutants, in addition to wild type, over a range of enrichment values in order to quantitatively examine the relationship between enrichment and affinity (Fig. 3c and d). Highly enriched substitutions such as N152D, D201G, and M212V were found to have the greatest reduction in affinity. The modestly enriched F140S substitution was found to have a small but statistically significant reduction in affinity compared to wild type ($p < 0.05$). The K203E substitution, which was not enriched in the sorted population, did not significantly impact 6H4-scFv binding. The data experimentally demonstrate a positive correlation between enrichment and affinity with a Spearman coefficient of 0.94 ($p = 8.3 \times 10^{-3}$).

Spatial orientation of residues modulating binding affinity

The structural organization of residues enriched in the sorted populations was visualized by mapping onto the structure of mouse PrP [45] (Fig. 4a–c). The most highly enriched residues for both ICSM18-scFv and D18-scFv form a continuous interface on the outward face of helix 1. Those for 6H4-scFv reside along the inner face of helix 1 and at the interface between helix 1 and helix 3 (D201) and also form a continuous binding pocket. In contrast to the other antibodies tested, a linear series of residues that are solvent exposed were enriched for EP1802Y. Mapping of the epitope residues onto the mouse prion protein structure identifies epitopes that are dependent on secondary and tertiary structures and characterizes two different binding pockets (that for ICSM18-scFv/D18-scFv and 6H4-scFv) along alpha helix 1.

The potential that some identified residues modulate binding as the result of the introduction of negative interactions or conformational changes can be observed by weaker colored regions or by highlighting residues buried deep within the hydrophobic core. Identified contact residues are shown in Fig. 4d.

For each antibody tested, the number of contact residues identified by this analysis was inferred to be
dependent upon the stringency of sorting (Supplemental Table 3). More stringent sorting conditions favor the identification of only mutations that dramatically reduce binding affinity, as was the case for ICSM18-scFv, for which two core residues are identified here of the six contact residues found in the crystal structure. Less stringent sorting conditions allow for the detection of mutations with intermediate and low impact on binding affinity. Therefore, the positive identification of contact residues here comes with the caveat that residues not identified may be in contact with the antibody but that their mutation causes a smaller decrease in binding affinity than those isolated.

Comparison of contact residues identified with epitopes determined by established methods

The contact residues identified here by mutagenic library sorting and deep sequencing are generally in agreement with previously reported epitopes for these antibodies using alternative methods and are more precisely defined compared to those identified using peptide arrays (Table 3). The epitopes of D18 and ICSM18 were originally identified by peptide recognition and were determined to include overlapping 25- and 11-residue stretches of PrP that include helix 1 [29,30]; the residues identified as the epitope of both antibodies here are present on the overlapping section of these peptides. The reported co-crystal structure for ICSM18 in complex with human PrP [46] shows a similar set of discontinuous epitope residues composed of externally facing alpha helix 1 residues. In the crystal structure, S142, D143, R150, and E151 were identified by proximity as epitope residues; however, the mutation data suggest that, while these residues may be in close contact with the antibody, their contributions to binding affinity are small, compared to the contributions of the residues identified here, for the ICSM18-scFv/MoPrP interaction. Differences in the antibody format and species variability of PrP in the region of the interface may also play a role providing a distinct epitope.

For 6H4, the mutagenic library sorting and deep sequencing approach identified an epitope composed of several residues in helix 1, previously identified as the region of the epitope by peptide array, as well as D201 located in helix 3 at its interface with helix 1, a contribution that was not detected by peptide array [20].
The availability of a minimal epitope composed of helix 1 only permits detectable binding to linear peptides at high concentrations despite the reduction in affinity due to the absence of D201 by direct mutation to D201 or potentially by mutations to the buried methionine residues on helix 3. The reduction, but not ablation, of PrP binding by 6H4 following the disruption of the disulfide bond has been described [33], consistent with an interaction that depends on tertiary structural organization.

Conclusions

In this study, we identify epitope profiles for the variable domains of three antibodies targeting helix 1 of PrP and one monoclonal antibody targeting the C-terminus of helix 3 with residue-level resolution. Two antibodies, ICSM18 and D18, were found to be dependent on interactions with discontinuous residues along the outer face of helix 1 that form a continuous surface only when the alpha helix is formed (Fig. 4), indicating the mechanism by which these antibodies recognize PrP C but not PrP SC, as conversion of PrP C to PrP SC is accompanied by a loss of alpha-helical secondary structure elements and an increase in beta content [47]. Some of the residues responsible for 6H4 interaction with PrP existing at the helix 1/helix 3 interface are inaccessible in a majority of the NMR structures in the PrP C configuration where some of the residues (E145, N152, and D201) are <15% solvent accessible. This suggests either that these residues become accessible for direct contact as the result of a structural rearrangement or that these residues play a critical role in maintenance of the alpha-helical PrP C structure. Structural dynamics may play a role in allowing access to D201 where the side chain is accessible in 2 of the 20 NMR models. A configuration allowing for surface accessibility of contact residues either is already present in PrP SC or, more likely, may be achieved at a comparable energetic cost, as 6H4 can bind both conformations [31].

Modulation of binding affinity by D201 in helix 3, which is in close proximity to helix 1 in PrP C (Fig. 4), may also explain differences in the strength of interaction of 6H4 with PrP C and PrP SC. This observation is also consistent with the decreased affinity of 6H4 for PrP following denaturation [33].

A variety of established techniques are available for the characterization of antibody epitopes including peptide arrays, alanine scanning, and co-crystallization. Peptide arrays are experimentally simple and inexpensive to employ but offer limited resolution and often cannot identify tertiary structural requirements. Mutational scanning and deep sequencing permits the contributions of each amino acid position to be evaluated in biologically relevant conditions, much like alanine scanning [48]. However, evaluation of many different substitutions at each position may provide information about reactivity across species or naturally occurring genetic variants. We found this to be the case where the N142S mutation was not enriched in the 6H4-scFv sorted population, an attribute that is likely important for the recognition of native human PrP by 6H4 where serine is the wild-type amino acid [20]. We identified the Q222E mutation as ablative in the EP1802Y sorted population, suggesting that the antibody may have reduced or ablated affinity to mink or elk PrP where glutamic acid is the native amino acid. An important consideration for mutational scanning is the possibility that some substitutions may dramatically alter global conformations by the introduction of unfavorable interactions. This concern is minimized but not entirely eliminated in alanine scanning and can be partially identified in mutational scanning by examining the enrichment of all constructed substitutions at each position. Both alanine scanning and mutational scanning approaches are inherently limited to residue-level resolution and are unable to match the atomic resolution of

Table 3. Comparison of anti-PrP antibody epitope residues characterized in this and other studies.

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Form</th>
<th>PrP species</th>
<th>Method</th>
<th>Epitope</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICSM18</td>
<td>Fab</td>
<td>Human</td>
<td>Co-crystallization</td>
<td>142-GSDYEDRYYREN-154</td>
<td>Khalili-Shirazi et al. [29]</td>
</tr>
<tr>
<td>ICSM18</td>
<td>scFv</td>
<td>Mouse</td>
<td>Mutational scanning</td>
<td>153-GSDYEDRYYREN-165</td>
<td>Korth et al. [20]</td>
</tr>
<tr>
<td>D18</td>
<td>Fab</td>
<td>Mouse</td>
<td>Mutational scanning</td>
<td>132-SAMSRPMIHFGNDWEDRYRENMYRP-158</td>
<td>Williamson et al. [30]</td>
</tr>
<tr>
<td>D18</td>
<td>scFv</td>
<td>Mouse</td>
<td>Mutational scanning</td>
<td>132-SAMSRPMIHFGNDWEDRYRENMYRP-158</td>
<td>This study</td>
</tr>
<tr>
<td>6H4</td>
<td>Fab</td>
<td>Bovine</td>
<td>Peptide phage display</td>
<td>141-GNDWEDRTRYREN-155</td>
<td>This study</td>
</tr>
<tr>
<td>6H4</td>
<td>scFv</td>
<td>Mouse</td>
<td>Mutational scanning</td>
<td>220-ESQAYDGRSS-231</td>
<td>This study</td>
</tr>
</tbody>
</table>
co-crystallization structural determination. The residue-level resolution of the method allows residues that do not contribute significantly to binding affinity to be identified, such as E151 for ICSM18-scFv, R150 for D18-scFv, and Y224 for Mab EP1802Y, permitting identification of potential hotspots.

The mutagenic library sorting and SMRT sequencing approach taken here may be applied to the study of other antibody–protein or protein–protein interactions, including conformation-dependent antibodies that recognize other targets [49,50]. The primary method limitation is the requirement that the protein be expressed in the yeast surface display construct; however, owing to yeast eukaryotic protein folding machinery, it is amenable to many types of proteins including high-molecular-weight and multi-domain proteins [51–54]. HiSeq sequencing generates a greater number of reads than SMRT sequencing, an attribute that can be leveraged for more detailed calculations of the protein fitness landscape. However, the short read lengths are most amenable to targeted sequencing approaches that are dependent on a priori identification of areas of interest. Additionally, SMRT sequencing has the ability to identify double mutants regardless of their genetic proximity to one another. The ability to analyze double mutants also enables the possibility of identifying epistasis of genetically distant residues and may be useful for libraries with increased coverage of the double-mutant sequence space. In this work, we obtained full-length sequence coverage using SMRT sequencing and identified conformational contributions that may not have been identified by targeting the established epitopes.

Materials and Methods

Production of anti-PrP scFv

Gene and protein sequences for D18, ICSM18, and 6H4 were obtained from National Center for Biotechnology Information databases (Supplementary Table 2) and gene sequences were commercially synthesized (Genscript USA Inc.) with a C-terminal FLAG tag for detection. Synthesized genes were cloned into yeast secretion vector pTY, linearized by MfeI digestion, and integrated into Ty transposons of the yeast genome using a G-418 selectable marker as previously described [55]. The YVH10 yeast strain, which is a BJ5464 derivative strain containing an additional copy of protein disulfide isomerase inserted in tandem with the exogenous copy, was used [56]. Clones plated onto 300 μg/mL G-418 agar plates were subsequently screened in galactose media to drive scFv secretion under the Gal1-10 promoter to identify highly productive clones [57]. Selected clones were grown in 1 L of galactose media for 72–96 h at 20 °C. Cells were harvested by centrifugation at 3000 g and supernatant was concentrated by ultrafiltration (Millipore UFC701008) using 10-kDa filters. Concentrated scFv concentrations varied from 0.5 to 100 μM as determined by anti-FLAG chemiluminescence and comparison to a FLAG standard (Sigma P7457). Mab EP1802Y was purchased from a commercial source (Abcam ab52604).

Generation of mutant PrP library

The wild-type mouse PrP gene encoding the mature form of the protein (residues 23–231) was obtained by removal of the 3F4 epitope tag from a PrP expression vector by site-directed mutagenesis (Addgene plasmid 1321) [58]. The wild-type PrP gene was cloned into yeast surface display vector pCTCON2 [59] to provide C-terminal fusion to yeast mating protein Aga2 and a tryptophan selection marker. Mutagenic library construction was performed as previously described [60], without the addition of nucleotide analogs and using only a single 30-cycle Thermus aquaticus polymerase amplification. We combined 9.5 μg of the resulting library with 0.5 μg of pCTCON2 plasmid backbone and transformed it into 50 μL of electrocompetent EBY100 strain yeast (~2 × 10^8 yeast). The transformation and homologous recombination led to a library of 8.9 × 10^6 members determined by serial dilution plating onto tryptophan-deficient plates.

Yeast labeling and sorting for ablated antibody binding

The yeast mutant PrP mutant library was grown at 30 °C overnight in tryptophan-deficient glucose media, harvested by centrifugation, resuspended to an optical density of ~1.0 in tryptophan-deficient galactose media, and grown for 20–24 h at 20 °C to induce protein expression. Exogenous expression of PrP in yeast produces protein that is post-translationally processed similarly to PrP\(^\text{\dagger}\) in mammalian cells [53]. Yeast labeling was performed as previously described [59], using near-saturating scFv/antibody and saturating anti-cmyc concentrations [mouse anti-FLAG M2 (Sigma F1804), chicken anti-cmyc (Invitrogen A21281), goat anti-rabbit phycoerythrin (Invitrogen P2771MP), goat anti-mouse-Alexa Fluor 647 (Invitrogen A21235), and goat anti-chicken Alexa Fluor 488 (Invitrogen A11039)]. Labeled yeast were analyzed on an Accuri C6 flow cytometer and sorted on a FACSaria or FACSaria II. Yeast were sorted for 2–4 successive rounds with outgrowth in tryptophan-deficient glucose media and expression in tryptophan-deficient galactose media between each round until the cmyc+/antibody+ population was eliminated.

For clonal analysis, yeast clones were selected from the sorted sub-libraries by dilution onto tryptophan-deficient plates or were constructed by site-directed mutagenesis and gene sequences were determined by Sanger sequencing. Clones were grown, induced for surface-displayed protein expression, and labeled as described above. Titrations for each clone were performed spanning over 4 orders of magnitude in antibody concentration. To determine the normalized fluorescence value, we obtained the minimum and maximum fluorescence values from the wild type with the mean fluorescence of the expressing population was used. A one-parameter model was used to
determine the dissociation constant $K_d$.

\[
\text{normalized fluorescence value} = \frac{[\text{Antibody}]}{[\text{Antibody}] + K_d}
\]

Mutations leading to less than 0.15 normalized fluorescence values throughout all concentrations tested were fit to a horizontal line representing the average normalized fluorescence values.

**Isolation of mutant PrP genes**

Sorted libraries were grown overnight in tryptophan-deficient glucose media and 0.5-mL culture was used for plasmid DNA extraction using the Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research D2004). DNA was further purified and concentrated by isopropanol/ethanol precipitation to remove inhibitors. Purified DNA was used to transform MegaX DH10B T1R Electrocomp Cells (Invitrogen C640003) by electroporation. Transformed cells were passaged into 50 mL Luria–Bertani broth with 100 μg/mL ampicillin for selection and serial dilution onto Luria–Bertani agar plates with 100 μg/mL ampicillin was performed to determine library size. Typical transformations provided a total of 4000–30,000 colony-forming units per library. *Escherichia coli* cultures were passaged once and 100 mL was grown in Luria–Bertani broth with 100 μg/mL ampicillin for DNA extraction using two QIAGEN Plasmid Plus Midi columns (QIAGEN #12943). Bacterially derived plasmid DNA was digested with NheI-HF and SalI-HF restriction enzymes (New England Biolabs R3131S and R3138S) to liberate the 633-base-pair fragment containing the PrP columns (QIAGEN #12943). Bacterially derived plasmid DNA was digested with NheI-HF and SalI-HF restriction enzymes (New England Biolabs R3131S and R3138S) to liberate the 633-base-pair fragment containing the PrP NMR model coordinates of the structured domain of the mouse PrP(121–231) were obtained from the Protein Data Bank (PDB ID: 2L39; model 1) [45] and were loaded into PyMOL [63] for visualization. Surface accessibility was determined using the surface accessibility calculator of the DeepView/Swiss-PdbViewer [64] using a greater than or equal to 15% amino acid surface accessibility criteria.

**SMRT sequencing and read filtration**

End repair was performed on restriction-digested DNA and samples were prepared for SMRT sequencing using the DNA Template Prep Kit 2.0 (Pacific Biosciences). Samples were run in a single SMRT cell with 2 × 45 min processing on PacBio RS II (Pacific Biosciences).

Data were initially processed by using the SMRT Portal RS_ReadsOfInsert module that utilizes the Quiver algorithm to obtain a maximum likelihood consensus sequence. The consensus sequences and accompanying quality values were exported in .fastq format and further processed using a linux script (Supplementary Information). Briefly, a pairwise sequence alignment was performed for each consensus sequence and the wild-type Prnp sequence using the needleall program from the EMBOSS suite [61]. Consensus sequences that produced poor alignments were removed from consideration. From the alignment, both nucleotide and protein mutations were identified. Sequences containing a cysteine mutation or producing a premature stop codon were censored because of the possibility of alternative disulfide bond formation. Ultimately, each sequence provided 19,000–30,000 high-quality reads (Table 2) with an average Q-value of 40.00–55.22 corresponding to error rates from $1.00 \times 10^{-4}$ to $3.00 \times 10^{-6}$ errors per nucleotide.

The algorithm used to calculate the error rate (Quiver) was trained on data providing a template for alignment; we chose to construct de novo assemblies for the sequencing subreads to provide consensus sequences and then used a traditional Needleman–Wunsch algorithm for alignment of the consensus sequences to prevent wild-type bias. We observed a higher appearance of sequences observed only one or a few times, considering that the estimated coverage was approximately 10-fold suggesting that the actual error rate may be as high as $2.58 \times 10^{-6}$ (Q-value of 35.9). Filtering removed 13.6% of sequences (CCS reads), on average, due to low quality, consistent with other similar approaches for SMRT sequencing [62] (see Supplementary Methods for detailed account of data filtering).

**Statistical analysis of mutations by position and substitution**

Single-mutant reads were used to identify epitope residues involved in antibody binding to PrP. Enrichment values were calculated for each position based on the number of times a mutation was observed in the sorted populations, the number of codons corresponding to the introduced amino acid arising from a single-nucleotide change, the nucleotide substitution biases, and the observed sorting stringency. $p$-values were calculated using the chi-squared statistic for each pairwise comparison. The polymerase mutational biases in the library were characterized (Supplementary Table 1). For comparison to traditional epitope identification methods, epitope residues were identified as residues showing statistically significant positional enrichment ($p < 10^{-5}$) and statistically significant enrichment ($p < 10^{-5}$) for more than half of the substitutions arising from a single-nucleotide mutations.

**Mapping of epitope residues onto mouse PrP NMR structure**

The NMR model coordinates of the structured domain of the mouse PrP(121–231) were obtained from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB ID: 2L39; model 1) [45] and were loaded into PyMOL [63] for visualization. Surface accessibility was determined using the surface accessibility calculator of the DeepView/Swiss-PdbViewer [64] using a greater than or equal to 15% amino acid surface accessibility criteria.

**Acknowledgements**

The authors would like to thank Olga Shevchenko for assistance with the SMRT sequencing and Bruce Kingham, Karol Miaskiewicz, and Lynn Opdenaker for helpful discussions. Yeast expression plasmids pCTCON2 and pITY and yeast strains EBY100 and YVH10 were a kind gift from the laboratory of K. Dane
Wittrup.

Funding

This work was supported by the National Institutes of Health (NIH)/National Institute of General Medical Sciences Chemistry–Biology Interface Program (5 T32 GM008550-17) and NIH/National Institute of General Medical Sciences Institutional Development Award Program (8 P20 GM103446-13). Additional support from NIH grant NS064173 and the University of Delaware Center for Bioinformatics and Computational Biology Core Facility was made possible through funding from the NIH (5 P20 RR016472-12) and National Science Foundation (EPS-081425).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.10.024.

Received 13 May 2014; Received in revised form 23 October 2014; Accepted 24 October 2014

Available online 7 November 2014

Keywords: EP1802Y; yeast surface display; epitope mapping; SMRT; discontinuous epitope

Abbreviations used:
SMRT, single-molecule real-time; FACS, fluorescence-activated cell sorting; NIH, National Institutes of Health.

References


Deep Mutational Scanning of Conformational Epitopes


Deep Mutational Scanning of Conformational Epitopes 339

[28] Williamson RA, Peretz D, Pinilla C, Ball H, Bastidas RB, Lieman

[29] Ramachandran GN, Ramakrishnan C, Sasisekharan V.


[42] An
Deep Mutational Scanning of Conformational Epitopes 339


[45] Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagene-


