Antibody Recognition of Disordered Antigens

Graphical Abstract

Highlights

- Disordered antigens are bona fide targets of antibody recognition
- Disordered epitopes are smaller, but more efficient, than ordered epitopes
- Recognition of disordered epitopes is more sensitive to epitope sequence variation
- Distribution of hotspots in the disordered antigen-antibody interface is asymmetric

Authors

Christopher A. MacRaidl, Jack S. Richards, Robin F. Anders, Raymond S. Norton

Correspondence

chris.macrailld@monash.edu

In Brief

MacRaidl et al. examine the molecular basis for the antibody recognition of disordered antigens, finding that disordered proteins are frequent targets of high-affinity antibodies. Structural details of these interactions are revealed, shedding light on the interplay between conformational disorder and the specificity of molecular recognition.
Antibody Recognition of Disordered Antigens

Christopher A. MacRaild,1,* Jack S. Richards,2 Robin F. Anders,3 and Raymond S. Norton1

1Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, VIC 3052, Australia
2Centre for Biomedical Research, Burnet Institute, Melbourne, VIC 3004, Australia
3Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC 3086, Australia

*Correspondence: chris.macraild@monash.edu
http://dx.doi.org/10.1016/j.str.2015.10.028

SUMMARY

Disordered proteins are important antigens in a range of infectious diseases. Little is known, however, about the molecular details of recognition of disordered antigens by their cognate antibodies. Using a large dataset of protein antigens, we show that disordered epitopes are as likely to be recognized by antibodies as ordered epitopes. Moreover, the affinity with which antigens are recognized is, unexpectedly, only weakly dependent on the degree of disorder within the epitope. Structurally defined complexes of ordered and disordered protein antigens with their cognate antibodies reveal that disordered epitopes are smaller than their ordered counterparts, but are more efficient in their interactions with antibody. Our results demonstrate that disordered antigens are bona fide targets of antibody recognition, and that recognition of disordered epitopes is particularly sensitive to epitope variation, a finding with implications for the effects of disorder on the specificity of molecular recognition more generally.

INTRODUCTION

There is an increasing recognition that many proteins naturally lack a defined folded state, and that their function depends instead on conformational disorder (Dunker et al., 2002; Wright and Dyson, 2015). These intrinsically disordered proteins are widespread in nature and are abundant in a range of pathogenic organisms. Several parasite species have an unusually high proportion of disordered proteins (Feng et al., 2006), and although the extent of disorder in viral proteomes is highly variable, some are predicted to be extensively disordered (Xue et al., 2014). Nonetheless, the implications of protein disorder for immune recognition by B cells and antibodies have received remarkably little attention.

On the one hand, it has been suggested that intrinsically disordered proteins will elicit weak immune responses or even be completely non-immunogenic (Dunker et al., 2002). This argument is based in part on the fact that disordered proteins often adopt relatively well-defined conformations when bound to partner proteins or to antibodies, and that this is achieved by a process of coupled folding and binding, which places both kinetic and thermodynamic constraints on the interaction (Uversky, 2013). As a result, the interactions of disordered proteins are often of relatively low affinity, despite maintaining high specificity. These properties are ideally suited to mediators of signals that must be switched on or off rapidly, and, accordingly, disorder is particularly common in proteins involved in signal transduction (Wright and Dyson, 2015). They are quite distinct, however, from the properties of typical antibody-antigen interactions, which are selected for high affinity. Consistent with the argument that disordered proteins may be poor immunogens, it has been observed that functionally important sites on protein antigens are often highly flexible, or are surrounded by flexible loops (Colman, 1997; Kwong et al., 2002; MacRaild et al., 2011); this flexibility is proposed to serve as a means of immune evasion, with “conformational masking” being mediated by the entropic cost of inducing order in the otherwise flexible antigen (Kwong et al., 2002).

In sharp contrast, however, disordered antigens can be immunodominant. In some instances these antigens, despite their immunodominance, fail to contribute to immune protection, and thus are believed to function as a “smokescreen”, diverting the immune system from targets with greater protective potential (Kemp et al., 1987). Nonetheless, numerous B-cell epitopes have been characterized in disordered proteins, and many of these appear to contribute to functional immune responses and therefore represent potential vaccine candidates (Adda et al., 2012; Foquet et al., 2014; Foucault et al., 2010; Olugbile et al., 2009; Raj et al., 2014; Yagi et al., 2014). For example, the protective effects of the most advanced malaria vaccine, RTS,S, appear to be mediated by antibodies to the disordered repeats of the circumsporozoite protein (Dyson et al., 1990; Foquet et al., 2014).

To explore the implications of disorder for antibody recognition, we have examined the immune epitope database (IEDB) (Vita et al., 2015) for the presence of antigens that are predicted to be disordered. We find that disordered antigens are no less likely than ordered antigens to be recognized by antibodies, and that the affinity of the antibody-antigen interaction is only weakly dependent on epitope disorder. Using a dataset of structurally defined antibody-antigen complexes, we reveal structural features that contribute to the unexpectedly high affinity of the interactions between disordered antigens and their cognate antibodies.

RESULTS

Disordered Proteins Are Common Targets of Antibody Responses

Although there is increasing awareness of the importance of disordered antigens in the immune response to a number of
Disorder Is No Barrier to the Development of an Antibody Response

(A) Disordered epitopes (red) are abundant in a dataset extracted from the IEDB.

(B) The average fraction of IEDB B-cell assays that are positive for antibody binding is higher for assays performed on epitopes that are predicted by IUPred to be disordered (red), than those predicted to be ordered (blue) or to contain a mix of both ordered and disordered residues (purple). Errors on the mean are estimated by bootstrap resampling from the set of antigens. See also Dataset S1.

(C) The fraction of IEDB B-cell assays that are positive (black) and the average relative SASA for residues in a non-redundant subset of the PDB (green), plotted as a function of IUPred score.

(D) Antibody binding affinities reported in the IEDB for ordered (blue), mixed (purple), and disordered (red) epitopes. Boxplots show the median (horizontal line) and interquartile range (box), with whiskers extending to 1.5 times the interquartile range, and any outlying points plotted. See also Dataset S1.

To explore this further, we sorted the residues in our antigen dataset according to their IUPred score, and assessed the relationship between IUPred score and the frequency with which B-cell assays returned positive results. We find that this proxy for antigenicity increases with increasing IUPred score over the majority of that scale (Figure 1C, black line). In other words, as the predicted disorder of residues in this dataset increases, so does the likelihood that B-cell assays involving those residues will return a positive result for antibody binding. Given that accessibility and IUPred score are expected to be correlated, and taking into account the fact that the antigenicity of residues in structured antigens is strongly dependent on the accessibility of these residues (Novotny et al., 1986), this effect may mask any negative impact of disorder on antigenicity. To test for this, we examined a non-redundant set of protein structures from the PDB (Berman et al., 2002) and assessed the relationship between relative solvent-accessible surface area (SASA) and IUPred score for each residue in this dataset. We find that the relationship between IUPred score and the average relative SASA tracks the relationship between IUPred score and antigenicity in the IEDB dataset, except at very small IUPred scores (Figure 1C). Both antigenicity (as measured by the frequency of positive B-cell assays) and solvent exposure increase gradually with IUPred score over the range 0.02–0.8, before increasing more sharply at higher IUPred scores. This result confirms the importance of solvent exposure as a key determinant of antigenicity (Novotny et al., 1986), and suggests that the dependence of antigenicity on IUPred score may largely reflect the increase in accessibility that accompanies increasing disorder. More importantly, it argues strongly against the suggestion that disordered epitopes are markedly less likely to be targets of an antibody response.

Antibody Affinity Is Only Weakly Influenced by Epitope Disorder

It has been suggested that antibodies are unlikely to bind disordered antigens with high affinity (Dunker et al., 2002), owing to the entropic cost associated with the disorder-order transition that accompanies binding (Dyson and Wright, 2002; Kwong et al., 2002). In support of this argument, it is claimed that disordered proteins and peptides bind with lower affinity to their binding partners. For example, there is a 25-fold difference in median affinity between the protein-protein and protein-peptide complexes in the dataset of Chen et al. (2013), while a 2 kcal/mol difference in average binding free energy (equivalent to an approximately 30-fold affinity difference) was observed in a dataset of disordered and ordered protein-protein interactions (Huang and Liu, 2013). However, it is unclear to what extent we can generalize from these datasets (Pancsa and Fuxreiter, 2012). To examine the magnitude of this effect for antibody recognition, we have identified those epitopes in our dataset for which affinity measurements reported for antibodies binding to 284 distinct epitopes, the median affinity of antibodies for disordered epitopes is only 5-fold lower than that for ordered epitopes (Figure 1D). We observe only a weak correlation between affinity and the average IUPred score for each epitope (Speaman’s
1.0–1.5 kcal/mol/residue (Baxa et al., 2014), while somewhat
lost due to folding of the protein backbone put that value at
Recent state-of-the-art computational estimates of the entropy
the magnitude of the entropic cost that is proposed to explain it.

Disordered Epitopes Are Small, Continuous Epitopes
(A) The number of contact residues comprising both the antibody paratope (left) and the antigen epitope (right) is smaller for disordered epitopes (red) than for ordered epitopes (blue).

(B) Antibody-antigen interactions involving ordered epitopes (blue) bury more surface area on the antibody (left) and the antigen (right) than do the interactions of disordered epitopes (red).

(C) The distribution of epitope span, the minimum number of contiguous residues that contain all of the epitope, for ordered epitopes (blue) is bimodal, while disordered epitopes (red) comprise only short, linear epitopes. The region of the distribution containing all disordered epitopes is shown at higher resolution (inset).

To give context to this result, we find it instructive to consider the magnitude of the entropic cost that is proposed to explain it. Recent state-of-the-art computational estimates of the entropy lost due to folding of the protein backbone put that value at 1.0–1.5 kcal/mol/residue (Baxa et al., 2014), while somewhat higher estimates of the conformational entropy loss that accompanies disordered protein interactions have been inferred from changes in nuclear magnetic resonance relaxation rates and from calorimetric measurements (Bracken et al., 1999). In contrast, the affinity difference observed here equates to a free energy difference of only 0.1 kcal/mol/residue over a typical nine-residue disordered epitope (see below, Figure 2). Alternatively, we may compare the 5-fold difference in affinity seen here with the 104-fold loss of affinity observed when one partner of a high-affinity protein-protein interaction becomes disordered (Papadakos et al., 2015). Thus, it is clear that antibodies recognizing disordered epitopes are able to compensate very effectively for the entropy loss that accompanies binding.

Disordered Epitopes Are Smaller than Ordered Epitopes
Studies of the interactions between disordered proteins and their binding partners have revealed that these interactions typically involve a larger interface surface area than do the interactions between ordered proteins (Meszaros et al., 2007; Wong et al., 2013). This may be a mechanism to compensate for the entropic cost of binding a disordered protein, as the larger interface offers greater opportunity for favorable enthalpic or solvent-driven contributions to binding. To examine whether this also applies to the interactions of antibodies with disordered epitopes, we assembled a dataset of 872 structures of antibodies in complex with protein antigens, and predicted epitope disorder using IUPred, as before. A total of 69 of these complexes involved epitopes in which all residues are predicted to be disordered, and the remaining epitopes are classed as ordered. We compared the sizes of the interfaces made by antibodies with ordered and disordered epitopes. We find that disordered epitopes are in fact substantially smaller than ordered epitopes, in terms of both the number of residues in contact with antibody (Figure 2A) and the buried surface area (Figure 2B). Likewise, paratopes cognate to disordered epitopes are smaller in area and in number of residues than is the case for paratopes to ordered epitopes. A typical disordered epitope comprises nine residues and makes contact with 21 paratope residues on the antibody, burying 695 and 512 Å2 on the antigen and antibody, respectively. In contrast, the median ordered epitope comprises 21 residues in contact with a 24-residue paratope and buries 847 and 788 Å2, respectively. Thus, in contrast to the interfaces between disordered proteins and their physiological binding partners, the interfaces between disordered epitopes and their cognate antibodies are small in area and involve relatively few residues. In this respect, disordered epitopes appear to be more similar to the peptide-protein interactions characterized by London et al. (2010), in which average interface surface areas of 512 ± 177 Å2 were observed.

We also examined the epitope span, the minimum number of contiguous residues that encompass all of the epitope, as a measure of the size of the structurally intact epitope (Figure 2C). Ordered epitopes show a bimodal distribution of epitope span; a small group of epitopes with spans <20 residues corresponding to simple linear epitopes, and a larger group with much longer spans corresponding to conformational epitopes. Disordered epitopes in this structural dataset are exclusively in the linear epitope class.
Antigen Disorder Does Not Affect Affinity Maturation of Cognate Antibodies

It has been suggested that antigen disorder may have implications for affinity maturation (Anders, 1986; MacRaid et al., 2015; Penkett et al., 1998), the process by which B-cell receptors develop affinity and specificity for a specific antigen through somatic mutation as they mature into antibodies. To test this, we compared the sequences of the murine antibodies in our structural dataset with those of germline V-gene sequences using IgBlast (Ye et al., 2013), and used the number of inferred V-gene somatic mutations as a proxy for the extent of affinity maturation. We find no evidence for an effect of antigen disorder on the number of somatic mutations (Figure 3), indicating that, at least by this measure, affinity maturation is not compromised by antigen disorder.

Disordered Epitopes Are Enveloped by Concave Paratopes

Figure 2B suggests that, on average, a larger surface area is buried on disordered antigens than is buried on their cognate antibodies, implying a tendency for the paratopes that recognize disordered antigens to be concave. Previous work has demonstrated that the topology of antibody paratopes depends to some extent on the nature of the antigen (MacCallum et al., 1996). We therefore examined the paratope geometry of the antibodies in our structural database. The ratio of buried surface area on the antibody to buried surface area on the antigen is significantly lower for disordered epitopes than it is for ordered epitopes in our dataset (Figure 4A). For ordered epitopes the median value of this ratio is 0.93, equivalent to an effective radius of curvature of 39 Å, and for more than 25% of ordered epitopes this ratio is >1, indicating a convex paratope. In contrast, this ratio is <1 for all disordered epitope complexes in the dataset, suggesting that paratopes cognate to disordered epitopes may always be concave, and the median value for disordered epitopes is 0.74, equivalent to an effective radius of curvature of 9 Å. Paratope topology is determined, at least in part, by variations in complementarity-determining region (CDR) length (Collis et al., 2003), and accordingly we find that the average length of CDRs H2, H3, and L1 differ significantly between antibodies recognizing ordered and disordered antigens (Figure 4B). In antibodies to disordered antigens, the central CDR H3 is shorter, and the flanking CDRs H2 and L1 are longer, consistent with the concave topology inferred from comparison of buried surface areas.

Consistent with this strong tendency for concave paratope topology, we observe that the average surface area buried on individual paratope and epitope residues is dependent on the nature of the epitope (Figure 4C). Disordered epitopes bury, on average, 24.1 Å² of antibody surface per paratope residue, significantly less than the 32.8 Å² per paratope residue buried by ordered epitopes. In contrast, the area buried per residue in the antigen shows the opposite trend, with disordered epitopes burying 75.6 Å² per residue, much more than the 44.9 Å² per residue buried by ordered epitopes. Together, these data suggest a more intimate interaction between disordered antigens and their antibodies, in which a smaller epitope is more completely buried within the concave paratope. Consistent with this interpretation, the interface shape complementarity between disordered epitopes and their cognate paratope is higher than it is for ordered epitopes (Figure 4D). Similarly, peptide-protein interfaces appear better packed than do protein-protein interfaces (London et al., 2010), as measured by the RosettaHoles packing score.

Structural Basis for the Unexpectedly High Affinity of Antibodies to Disordered Epitopes

The data in Figures 2 and 4 show that disordered epitopes are small, but have much more extensive contacts per epitope residue and better shape complementarity with their cognate antibody than do ordered epitopes. As proposed previously for peptide-protein interactions (London et al., 2010), this suggests a possible explanation for the unexpected efficiency with which these antibodies compensate for the large entropic cost of binding to a disordered epitope. By limiting the number of residues involved in the epitope this entropic cost is minimized, while by maximizing the extent and complementarity of the interaction the compensating enthalpic component of the binding free energy is optimized.

To explore this further, we cross-referenced the antibody-antigen binding affinities reported in the IEDB with our structural dataset by matching antibody names and epitope identifiers in the respective IEDB records. In this way, we assigned 357 experimentally determined $K_D$ values to the corresponding antibody-antigen complex structures. We confirmed that the affinities in this structural dataset are similar to those observed for the sequence-based antigen dataset (compare Figures 1D and 5A). We also compared the free energy of binding as a fraction of the buried surface area on the antibody and the antigen (Figure 5B) and of the number of antibody and antigen contact residues (Figure 5C). By both measures, the interactions of disordered epitopes are more efficient: they gain almost 30%
more binding free energy for each unit area of antibody buried, and gain almost twice the free energy per antigen residue in contact with antibody. There is a modest but highly significant correlation between binding affinity and total interface surface area for ordered antigens ($\rho = 0.29, p = 9 \times 10^{-8}$), consistent with an extensive body of literature on diverse classes of protein interactions (Chen et al., 2013; Lo Conte et al., 1999; Myslinski et al., 2011; Spolar et al., 1989). Remarkably, that correlation appears to be absent for disordered antigens ($\rho = 0.005$), although this difference in correlation does not reach formal significance over the relatively small number of disordered antigens for which affinities are available ($n = 26, p = 0.08$ by permutation test). We also observe a significant correlation between interface shape complementarity and the binding free energy per unit total buried surface area ($\rho = 0.40, p = 1 \times 10^{-3}$ over the full dataset), suggesting that shape complementarity may be an important mechanism by which antibodies to disordered epitopes achieve their binding efficiency. Strikingly, however, there is no correlation between shape complementarity and overall affinity ($\rho = 0.2 \times 10^{-4}$).

Hydrogen bonds and salt bridges are known to be enriched in the interfaces of disordered proteins with ordered binding partners (Wong et al., 2013), and in peptide-protein interactions (London et al., 2010). Accordingly, we examined the abundance of these polar contacts in the current structural dataset. In contrast to these previous results in related systems, we find fewer intermolecular polar contacts, in absolute terms, in antibody-antigen complexes involving disordered epitopes than in complexes involving ordered antigens (Table 1). Nonetheless, because of the smaller size of the complexes of disordered antigens, the density of these contacts is higher; residues within disordered epitopes are more than 50% more likely to be
than in ordered epitopes. These data suggest that epitope polymorphisms are more likely to modulate antibody recognition in disordered epitopes with disordered epitope residues almost twice as likely as ordered epitope residues to be hot spots (Figure 6C, right).

\[ \text{Figure 5. Antibodies Bind Disordered Epitopes with Greater Efficiency} \]

(A) The affinity of antibodies in our structural dataset differs only slightly between ordered (blue) and disordered (red) epitopes. (B) Antibodies to disordered epitopes (red) obtain more binding free energy per unit of buried surface area on the antibody (left) than do antibodies to ordered epitopes (blue). No such difference is seen in binding free energy per unit of buried antigen surface (right). (C) Antibodies to disordered epitopes (red) gain less binding free energy per contact residue in the antibody (left) than do antibodies to ordered epitopes (blue), but more energy per contact residue in the antigen (right).

Boxplots show the median (horizontal line) and interquartile range (box), with whiskers extending to 1.5 times the interquartile range, and any outlying points plotted.

Involved in a salt bridge and twice as likely to be involved in a hydrogen bond with the antibody compared with residues in ordered antigens.

**Implications for the Specificity of Antibody Recognition by Disordered Antigens**

The preceding results establish that disordered epitopes are intimately enveloped by their cognate antibody, making more efficient use of a limited interface to generate high binding energies. This may have implications for the specificity of the interactions between disordered antigens and their cognate antibodies. In particular, we predict that the more extensive contacts made by each residue of a disordered epitope will make antibody recognition of disordered antigens more sensitive to sequence variation within the epitope. To test this hypothesis, we have performed computational alanine scans on each of the structures in our dataset to estimate the consequences of mutating each interface residue for the affinity of the complex. The consequences of mutations of paratope residues are relatively insensitive to the degree of order in the cognate epitope (Figure 6A). In contrast, mutations within disordered epitopes are much more frequently associated with a significant unfavorable change in binding affinity (i.e. a large positive value of $\Delta \Delta G^\circ$, Figure 6B).

It is widely recognized that a few key residues make dominant contributions to the binding energy of protein-protein interactions (Bogan and Thorn, 1998; Clackson and Wells, 1995). These “hot-spot” residues are thus the key determinants of binding affinity and specificity. We find that the paratopes of antibodies to disordered epitopes are slightly depleted in hot-spot residues (defined as residues with a predicted $\Delta \Delta G^\circ > 1.5$ kcal/mol, relative to those of antibodies recognizing ordered epitopes (Figure 6C, left). In contrast, disordered epitopes are heavily enriched in hot-spot residues, with disordered epitope residues almost twice as likely as ordered epitope residues to be hot spots (Figure 6C, right). These data suggest that epitope polymorphisms are more likely to modulate antibody recognition in disordered epitopes than in ordered epitopes.

**DISCUSSION**

Disordered antigens are important targets of natural immune responses to a number of pathogens, and are increasingly attracting attention as potential vaccine components (Adda et al., 2012; Foquet et al., 2014; Foucault et al., 2010; Olugbile et al., 2009; Raj et al., 2014; Yagi et al., 2014). In spite of this, little is known about the implications of conformational disorder for the development of an effective immune response. Indeed, it is widely speculated that conformational disorder may impede the development of a high-affinity and specific antibody response (Anders, 1986; Dunker et al., 2002; Kemp et al., 1987; Kwong et al., 2002; MacRaid et al., 2011, 2015). The results presented here suggest otherwise: disordered proteins appear to be more antigenic than ordered proteins, reflecting the greater exposure of disordered regions to solvent (and hence to recognition by antibodies) (Figures 1B and 1C). Likewise, the difference in the affinities of the interactions made by antibodies with ordered and disordered epitopes is very much smaller than might be expected, given the large, unfavorable entropic contribution arising from the disorder-order transition that necessarily accompanies binding (Figure 1D). Indeed, antibodies bind disordered epitopes with higher efficiency than ordered epitopes, in the sense that they achieve more binding free energy per unit of interface surface area, or per epitope residue (Figure 5). This efficiency is presumably driven by the need to minimize the extent of disordered epitopes, and thus minimize the entropic cost of inducing order.

Our analysis of the structures of disordered epitopes in complex with antibodies reveals features of these interactions that may contribute to this increase in binding efficiency. We find that disordered epitopes tend to bind deep within concave paratopes, burying a much larger proportion of each interacting residue than do ordered epitopes (Figure 4). Moreover, they show greater shape complementarity with their cognate antibody, and possess a greater density of intermolecular polar contacts (Figure 5 and Table 1). This may reflect the flexibility of disordered proteins and their consequent ability to better adapt to the shape of their binding partner. Indeed, this flexibility permits some disordered epitopes to adopt distinct conformations...
bound to different antibodies (Chu et al., 2014; Deng et al., 2014). Additional factors outside the scope of the current datasets may also be at play here; for example, we are unable to assess the extent to which residual structure exists within otherwise disordered epitopes, which may reduce the entropic cost of binding (Borcherds et al., 2014). Likewise, the extent to which transient or “fuzzy” interactions modulate the affinities characterized here is unknown (Fuxreiter and Tompa, 2012), although in at least one instance fuzzy interactions do appear to modulate the affinity and specificity with which an antibody recognizes its disordered antigen (Morales et al., 2015).

These caveats notwithstanding, similar observations have been made for the interactions that peptides and disordered proteins make with other binding partners (London et al., 2010; Meszaros et al., 2007; Wong et al., 2013), suggesting that the strategies of concave binding sites, high surface complementarity, and dense polar contacts may contribute to the interactions of disordered proteins more generally. They are, however, particularly pronounced in the current dataset, for at least two reasons. First, B cells mature under strong selective pressure for affinity to antigen, meaning that the antibody-antigen interactions studied here are selected for optimal affinity to an extent that may not apply to the interactions of disordered proteins in other contexts (Zhou, 2012). Second, whereas the interfaces disordered proteins make with antibodies are small (Figure 2), limited by the extent of both the disordered epitope and antibody paratope generally, the interfaces that disordered proteins make with other protein binding partners are often much larger (Wong et al., 2013), limiting the need for the extreme spatial efficiency seen here.

Very recently, we showed that the most flexible regions of the highly disordered malaria antigen MSP2 were less immunogenic than slightly more constrained regions (MacRaid et al., 2015). Based on this we hypothesized that, in the context of disordered antigens, intermediate levels of flexibility may be optimal for immunogenicity. Although we find no clear support for this hypothesis in the current analysis, direct comparison with our data on MSP2 is challenging because MSP2 has a uniformly high IUPred score (>85% of residues score >0.8), leading to weak correlation between IUPred score and the experimental measures of conformational flexibility analyzed in our prior work.

The implications of conformational disorder for the specificity of protein interactions remain a matter of active debate (Haynes et al., 2006; Huang and Liu, 2013; Schnell et al., 2007; Tokuriki and Tawfik, 2009a; Zhou, 2012). We have shown that individual residues of disordered epitopes are more likely to be energetically important for recognition by antibody than their counterparts in ordered epitopes. Very similar observations have been made for other interactions involving disordered proteins and peptides (Chen et al., 2013; Wong et al., 2013), suggesting that, in general, interactions with disordered proteins may be more sensitive to mutation in the disordered partner, and in that sense more specific than the interactions of ordered proteins. One striking observation from the data presented here is the asymmetry seen between antibody and antigen in measures of binding efficiency (Figure 5) and the distribution of hot-spot residues (Figure 6). While disordered epitopes contribute more binding energy per residue than do ordered epitopes, the opposite trend holds (albeit weakly) for their cognate antibodies; similarly, hot-spot residues are enriched within disordered epitopes but depleted within their cognate paratopes when compared with their ordered counterparts. Again, there is evidence for such asymmetry in the interactions of disordered proteins more broadly (Meszaros et al., 2007; Wong et al., 2013). This implies that, while the recognition of disordered proteins may be particularly sensitive to sequence variation in the disordered binding partner, it will be relatively insensitive to variation in the ordered binding partner. In other words, the effects of disorder on specificity may themselves be asymmetric, being dependent

### Table 1. Disordered Epitopes Are Enriched in Polar Contacts with Antibody

<table>
<thead>
<tr>
<th></th>
<th>Ordered</th>
<th>Disordered</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>H bonds</td>
<td>9.1 ± 0.2</td>
<td>7.8 ± 0.4</td>
<td>0.007</td>
</tr>
<tr>
<td>Per 100 Å²</td>
<td>1.11</td>
<td>1.32</td>
<td>0.002</td>
</tr>
<tr>
<td>Per Ag residue</td>
<td>0.52</td>
<td>1.0</td>
<td>6 × 10⁻¹⁶</td>
</tr>
<tr>
<td>Salt bridges</td>
<td>1.73 ± 0.05</td>
<td>1.55 ± 0.2</td>
<td>NS</td>
</tr>
<tr>
<td>Per 100 Å²</td>
<td>0.22</td>
<td>0.26</td>
<td>NS</td>
</tr>
<tr>
<td>Per Ag residue</td>
<td>0.11</td>
<td>0.18</td>
<td>0.002</td>
</tr>
</tbody>
</table>

NS, not significant.

![Figure 6. Antibody Recognition of Disordered Epitopes Is More Sensitive to Mutation](image-url)
on which side of the protein complex variation occurs. It is not clear that this asymmetry has been recognized in previous attempts to address the interplay of conformational disorder and specificity (Huang and Liu, 2013). It is striking, also, to contrast the clear differences in the distribution of stability effects of mutations observed here (Figures 6A and 6B) to the uniform distribution of the effects of mutations on the stability of globular proteins (Tokuriki et al., 2007). The distribution of mutation effects has implications for protein evolvability (Tokuriki and Tawfik, 2009b); it is likely, therefore, that these differences contribute to the distinct patterns of sequence evolution seen in the interaction motifs of disordered proteins (Brown et al., 2011).

Disordered proteins and regions of low sequence complexity are particularly abundant in a range of pathogenic species (Feng et al., 2006; Pizzi and Frontali, 2001; Xue et al., 2014), and the consequences of this disorder for antigenic responses to these pathogens have been the subject of extensive speculation (Kemp et al., 1987; Ridley, 1991; Zilversmit et al., 2010). The results presented here demonstrate that disordered epitopes are bona fide targets of high-affinity and specific antibody responses. They imply that the entropic cost of binding to a disordered epitope is in itself no barrier to the development of a high-affinity and specific antibody response, in contrast to some suggestions (Kwong et al., 2002; MacRaid et al., 2015; Penkett et al., 1998). Indeed, they suggest that antibody recognition of disordered antigens is likely to be particularly sensitive to polymorphism in the antigen sequence. Disordered regions in protein antigens are also frequently more polymorphic than ordered regions; our findings suggest that the effects of some of these polymorphisms on antibody recognition may be particularly acute. It is possible, therefore, that the abundance of disordered proteins in some pathogens is, to an extent, due to the tolerance of these proteins for sequence variation that will be particularly effective in enabling the evasion of recognition by antibodies.

EXPERIMENTAL PROCEDURES

Linear Peptide Epitopes from the Immune Epitope Database

From the IEDB, we extracted antibody binding assays involving linear peptide epitopes. From these assays we retained those involving antigens for which five or more assay records are available and for which we could determine the complete antigen sequence, giving a final dataset of 105,235 antibody binding assays covering 54,293 unique epitopes on 1,946 antigens. Affinity measurements are reported for 666 binding assays covering 284 of these epitopes (Dataset S1). IUPred scores (Dosztanyi et al., 2005) were calculated over the full sequence of each antigen, and epitopes were classified as “disordered” if the IUPred score was $>0.5$ for all residues within the epitope and for five flanking residues on each side, “ordered” if the IUPred score was $<0.5$ for all such residues, and “mixed” otherwise. The relationship between IUPred score and solvent exposure was estimated from a non-redundant selection of chains extracted from the PDB (Berman et al., 2002), using IUPred and DSSP (Kabsch and Sander, 1983) as described in Supplemental Experimental Procedures.

Structural Analysis of Disordered Antigen-Antibody Interactions

We extracted 872 structures from the PDB based on IEDB records describing three-dimensional structures of antibody complexes with protein or peptide antigens. Binding affinities for each antibody-antigen complex were extracted from the IEDB by cross-referencing the IEDB epitope ID and antibody name from each three-dimensional structure record with records reporting quantitative binding affinities. Where multiple affinities were reported for a single complex, we used the mean of the $pK_a$ values (Dataset S2). SASA and secondary structures were determined using DSSP, shape complementarity of the antibody-antigen interface was assessed with the $S_v$ statistic (Lawrence and Colman, 1993), and hydrogen bonds between antigen and antibody were identified using HBPlus (McDonald and Thornton, 1994) using default parameters. Further details of the analysis are presented in Supplemental Experimental Procedures.

For the inference of somatic V-gene mutations, the antibody sequences as reported in the PDB SEQRES field were compared with the appropriate IMGT germline V-gene reference dataset using IgBlast (Ye et al., 2013). Significant species-specific differences in the number of V-gene mutations were observed, so the dataset was segregated by species. However, there were insufficient sequences available for a robust analysis for species other than Mus musculus, so only murine antibodies are included in the final analysis.

Computational alanine scans were performed using FoldX 3.0 beta 6.1 (Schymkowitz et al., 2005) The FoldX RepairPDB command was run first, to optimize the complex conformation under the FoldX force field, followed by the complex_alascan command. All parameters were maintained at default values except $<VdWDesign>$, which was set to 0 for the complex_alascan procedure. We define hot-spot residues as those with predicted $\Delta\Delta G^\circ$ $> 1.5$ kcal/mol.

Statistical Analysis

Comparisons between classes were performed using the Wilcoxon rank-sum test, except for the comparisons between CDR lengths, which used bootstrap resampling to test for differences in the mean.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two datasets, and one figure and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.10.028.

AUTHOR CONTRIBUTIONS

C.A.M. conceived the study, collated and analyzed the data, and drafted the manuscript. J.S.R., R.F.A., and R.S.N. contributed to data analysis and interpretation, and revised the manuscript.

ACKNOWLEDGMENTS

This work was supported in part by an Australian National Health and Medical Research Council (NHMRC) project grant (1042520). R.S.N. and J.S.R. acknowledge fellowship support from the NHMRC.

Received: August 4, 2015
Revised: October 12, 2015
Accepted: October 28, 2015
Published: December 17, 2015

REFERENCES


Supplemental Information

Antibody Recognition of Disordered Antigens

Christopher A. MacRaild, Jack S. Richards, Robin F. Anders, and Raymond S. Norton
Supplemental Information

**Antibody recognition of disordered antigens**

Christopher A. MacRaild\(^1\), Jack S. Richards\(^2\), Robin F. Anders\(^3\) and Raymond S. Norton\(^1\)

\(^1\)Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville 3052, Australia. \(^2\)Centre for Biomedical Research, Burnet Institute, Melbourne, Victoria 3004, Australia and \(^3\)Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia.

\* To whom correspondence should be addressed: Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville 3052. AUSTRALIA. Phone: (+61 3) 9903 9101. FAX: (+61 3) 9903 9582 E-mail: chris.macraild@monash.edu
**Supplemental Data**

**Figure S1** Related to Figure 1. Comparison of the epitope classifications from IUPred and DISOPRED. A. Histogram of epitopes classified as ordered (blue), mixed (purple), and disordered (red) by DISOPRED (light shades) and IUPred (dark shades). B. Histogram of the joint IUPred/DISOPRED classification. The population of each bin as a percentage of the corresponding IUPred class is shown above each bar. C. The average fraction of IEDB B-cell assays that are positive for antibody binding is higher for assays performed on epitopes that are predicted by DISOPRED (light shades) or IUPred (dark shades) to be disordered (red), than those predicted to be ordered (blue) or to contain a mix of both ordered and disordered residues (purple).
**Dataset S1** Related to Figure 1. Spreadsheet detailing the linear peptide epitopes for which affinity data are available. Includes IEDB epitope ID, average IUPred score for the epitope and flanking residues, disorder classification (ordered, disordered or mixed), and the average of the reported pKₐ for each epitope.

**Dataset S2** Spreadsheet detailing protein antigen-antibody complexes. Includes IEDB Assay ID, disorder classification, IEDB epitope ID, Interface surface area, the average of the reported pKₐ for each complex, PDB ID, and the IUPred scores for each epitope residue.
Supplemental Experimental Procedures

Linear peptide epitopes from the immune epitope database. The IEDB catalogues epitope-specific experimental assays extracted from the primary literature. It includes data from immune responses developed in any vertebrate species, in the context of both natural and experimental exposure to antigen (Vita et al., 2015). Included in the database are B-cell or antibody binding assays, in which antibody recognition of a defined epitope is tested. We extracted 139,790 of these antibody-binding assays which involved linear peptide epitopes ranging in length from 5 to 25 residues extracted from 4522 protein antigens. We filtered these assays to retain those involving antigens for which 5 or more assay points are available and for which we could determine the complete antigen sequence, giving a final dataset of 105,235 antibody binding assays covering 54,293 unique epitopes on 1946 antigens. Affinity measurements are reported in the IEDB for 666 of these binding assays, covering 284 epitopes, (Supplemental Dataset 1). In order to identify disordered epitopes within this dataset, IUPred (Dosztanyi et al., 2005) and DISOPRED3 (Jones and Cozzetto, 2015) scores were calculated over the complete sequence of each antigen using the default parameters for each predictor. Epitopes were classified as disordered if the score from the relevant predictor was ≥ 0.5 for all residues within the epitope and for five flanking residues on either side, as ordered if the relevant score was < 0.5 for all such residues, and mixed otherwise. Comparison of the results from the IUPred and DISOPRED classifications suggested that the two performed similarly, except that DISOPRED predicts a much larger number of mixed epitopes (Fig S1). This is consistent with the known sensitivity of DISOPRED to short regions of disorder, in contrast to the default behaviour of IUPred which is tuned to the prediction of longer disordered regions (Dosztanyi et al., 2005; Jones and Cozzetto, 2015). In order to focus our analysis on epitopes within regions of extensive disorder, rather than on epitopes located in flexible loops within otherwise ordered antigens, we chose to use the IUPred classifications for all further analysis.

The relationship between IUPred score and solvent exposure was estimated from a non-redundant selection of the PDB (Berman et al., 2002), filtered at the level of 30% sequence identity using the search options of the PDB website. Each chain in the resulting set of structures was analysed in isolation, with the relative solvent-accessible surface area (rSASA), the ratio of the observed SASA for each residue to that expected for that residue type in the context of a short disordered peptide, determined using DSSP (Kabsch and Sander, 1983). Residues were then binned according to IUPred score, and the average rSASA for each bin calculated.

Structural analysis of disordered antigen-antibody interactions. We extracted 872 structures from the PDB based on IEDB records describing three-dimensional structures of antibody complexes with protein or peptide antigens (Supplemental Dataset 2). The protein chains representing the antibody and antigen were identified based on the IEDB annotations, and SASA and secondary structures were determined using DSSP. SASA was calculated separately for the isolated antibody and antigen and for the complex. We define the epitope as comprising all residues of the antigen that have a smaller SASA in the complex than in the free state. Similarly, the paratope comprises all antibody residues for which SASA is smaller in the complex. Epitopes were characterised for disorder using IUPred score as described above. As ordered and mixed epitopes have similar properties (Fig 1), we consider these classes together in the structural analysis. Shape complementarity of the antibody-antigen interface was assessed with the Sₚ statistic (Lawrence and Colman, 1993) as implemented in the CCP4 suite. Hydrogen bonds between antigen and antibody were identified using HBPlus (McDonald and Thornton, 1994) with all parameters set to their default values. CDRs were identified according to the definitions of Kabat (Kabat et al., 1983), using AbNum (Abhinandan and Martin, 2008).
Binding affinities for each antibody-antigen complex were extracted from the IEDB by cross-referencing the IEDB epitope ID and antibody name from each three-dimensional structure record with records reporting quantitative binding affinities. Where multiple affinities were reported for a single complex, we used the mean of the reported pK_d values.

**Supplemental References**


