ABangle: characterising the VH–VL orientation in antibodies

J.Dunbar1, A.Fuchs2, J.Shi3 and C.M.Deane1,4

1Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. 2pRED Informatics, Roche Diagnostics GmbH, 82377 Penzberg, Germany and 3Informatics, UCB Pharma, 216 Bath Road, Slough SL1 4EN, UK

4To whom correspondence should be addressed.
E-mail: deane@stats.ox.ac.uk

Received March 7, 2013; revised April 12, 2013; accepted April 23, 2013

Edited by Anthony Rees

The binding site of an antibody is formed between the two variable domains, VH and VL, of its antigen binding fragment (Fab). Understanding how VH and VL orientate with respect to one another is important both for studying the mechanisms of antigen specificity and affinity and improving antibody modelling, docking and engineering. Different VH–VL orientations are commonly described using relative measures such as root-mean-square deviation. Recently, the orientation has also been characterised using the absolute measure of a VH–VL packing angle. However, a single angle cannot fully describe all modes of orientation. Here, we present a method which fully characterises VH–VL orientation in a consistent and absolute sense using five angles (HL, HC1, LC1, HC2 and LC2) and a distance (dc). Additionally, we provide a computational tool, ABangle, to allow the VH–VL orientation for any antibody to be automatically calculated and compared with all other known structures. We compare previous studies and show how the modes of orientation being identified relate to movements of different angles. Thus, we are able to explain why different studies identify different structural clusters and different residues as important. Given this result, we then identify those positions and their residue identities which influence each of the angular measures of orientation. Finally, by analysing VH–VL orientation in bound and unbound forms, we find that antibodies specific for protein antigens are significantly more flexible in their unbound form than antibodies specific for hapten antigens. ABangle is freely available at http://opig.stats.ox.ac.uk/webapps/abangle.

Keywords: antibody modelling/antibody structure/computational tools/domain orientation/protein engineering

An effective immune system must have the ability to specifically recognise an almost limitless number of potentially pathogenic molecules. This range in specificity is provided by diversity in the chemical composition and shape of the antibody’s antigen binding site (Amzel and Poljak, 1979). Most of this diversity is thought to be achieved through variation in the sequence and structure of six loops, known as the complementarity determining regions (CDRs) (Wu and Kabat, 1970). Three CDRs are located on each of the variable domains: L1, L2 and L3 on VL; H1, H2 and H3 on VH (Chothia et al., 1989). The two domains, VH and VL, associate non-covalently to bring the CDRs in proximity and form the antigen binding site between them (Chothia et al., 1985; Vargas-Madrazo and Paz-Garcia, 2003).

The geometry of the binding site is further modulated by how the VH and VL domains orientate with respect to one another (Colman et al., 1987; Colman, 1988; Foote and Winter, 1992). This variation has been proposed as an additional mechanism to increase the repertoire of antibody specificity (Davies and Metzger, 1983; Chothia et al., 1985; Stanfield et al., 1993; Khalifa et al., 2000; Vargas-Madrazo and Paz-Garcia, 2003). Several studies have also observed that mutations to residues at framework positions (i.e. those not in the CDRs of the Fv) in the VH–VL interface can act to change antigen affinity (Riechmann et al., 1988; Foote and Winter, 1992; Banfield et al., 1997). These positions are distant from the binding site and unable to make direct contact with the antigen. Therefore, their effect on antigen affinity must be due to a structural change of the binding site geometry: modifying the VH–VL orientation.

In general, current computational tools for modelling antibody structures take a conservative approach when predicting the VH–VL pose (Whiteleg and Rees, 2000; Marcatili et al., 2008; Almagro et al., 2011). Most methods copy the orientation from a known Fv with a high sequence similarity. Perhaps the most complex algorithm, Rosetta Antibody (Sircar et al., 2009; Sivasubramanian et al., 2009), attempts to optimise the pose using an orientation sampling protocol. Recently, attempts have been made to systematically identify positions within the Fv which influence inter-domain orientation (Abhinandan and Martin, 2010; Chailyan et al., 2011). To approach this problem, a consistent description of the VH–VL orientation is required.

When making a comparison between any two protein structures, it is common to use distance-based metrics such as the root-mean-square deviation (RMSD) of equivalent atoms. This measure has been used by several studies to quantify the relative changes in the VH–VL orientation for specific pairs of structures (e.g. Li et al., 2000; Narayanan et al., 2009; Sela-Culang et al., 2012). In other cases, the angle required to rotate between the two conformations has been reported (Colman et al., 1987; Stanfield et al., 1993; Banfield et al., 1997; Teplyakov et al., 2011). While these relative measures do quantify changes in domain orientation, they do not consistently report on how the pose changes between antibody...
structures. For example, while two pairs of structures may both differ by 2.5 Å RMSD, one is unable to tell whether the pairs differ in the same way. Additionally, there is no direct way to tell how the pairs relate to one another. Similarly, a rotation angle may be reported for each pair (Stanfield et al., 1993; Banfield et al., 1997; Teplyakov et al., 2011). However, the angle can be reported along an arbitrary axis such that its direction is unclear.

Although the CDRs of the antibody Fv are highly variable, the structure of the framework regions of the VH and the VL domains are relatively conserved (Chothia et al., 1989). In both VH and VL, the framework has β-sandwich architecture. This conservation can be utilised to define the VH–VL orientation in an absolute sense. In doing so, the structural space of the antibody Fv can be quantified.

Abhinandan and Martin (2010) defined such an absolute measure with the VH–VL packing angle. This was a torsion angle measured between a vector fitted through the Cα coordinates of conserved positions in the interface β-sheet of the VH domain and a similar vector on the VL domain. Their packing angle varied from −60.8° to −31.0° in known molecules and allowed each antibody to be placed on an absolute scale of structural space. With this definition of VH–VL orientation Abhinandan and Martin were able to select positions that are influential in determining pose: L38, L40, L41, L44, L46, L87, H33, H42, H45, H60, H62, H91 and H105 under the Chothia numbering scheme (Chothia and Lesk, 1987).

A different approach was taken by Chailyan et al. (2011). They focused on identifying different types of VH–VL interface within antibody structures by using a relative measure, global distance test – high accuracy (Zemla, 2003; Read and Chavali, 2007). This measure calculates the structural similarity between the two structures by assessing the fraction of corresponding positions that can be superimposed within different distance thresholds. Analysis of the VH–VL interface structural similarity of 101 Fv regions identified essentially two clusters of antibody structures (A and B). A similar clustering was also identified by Sivasubramanian et al. (2009) using an RMSD measure. Given this clustering, a difference in the orientation of the variable domains and binding site geometry would be expected to be observed. Indeed, Chailyan et al. calculated that structures in cluster B had a significantly smaller binding site area than those in A. Structures in B were also found to be specific for smaller antigens. However, no equivalent clustering is observed in Abhinandan and Martin’s packing angle measure of VH–VL orientation. Furthermore, the eight positions that Chailyan et al. found best discriminated between clusters (L8, L28, L36, L41, L42, L43, L44 and L66) are in agreement with only two of Abhinandan and Martin’s positions, L41 and L44. As highlighted in a recent review (Kuroda et al., 2012), this inconsistency may be due to the inability of a single torsion angle to capture all modes of orientation variation between the domains.

Here, we describe a method for fully characterising the VH–VL orientation in an absolute sense. This allows us to make a comparison of the VH–VL pose of a single Fv region, to that of all other known structures. This method has been implemented in the computational tool, ABangle, available at http://opig.stats.ox.ac.uk/webapps/abangle. To demonstrate its use, we compare how Chailyan et al.’s clusters differ in orientation. We resolve the apparent discrepancy in the positions assigned to be important for Abhinandan and Martin’s packing angle and Chailyan et al.’s clusters. Additionally, we use ABangle to find those positions and their residue identities that are most influential for determining VH–VL orientation. Finally, a case study analysis of the conservation in orientation in sequence-identical Fv structures suggests that those that bind to hapten antigens are more rigid than those that bind to larger protein antigens.
Each of these positions:

\[
S_i = \frac{\sum_{j=1, j \neq i}^R \text{RMSD}_{ij}}{\text{RMSD}_{ij}}
\]

\[
\text{RMSD}_{ij} = \sqrt{\frac{1}{N} \sum_{n=1}^N (d_{n,ij} - \overline{d_{ij}})^2}
\]

\[
\overline{d_{ij}} = \frac{\sum_{n=1}^N d_{n,ij}}{N}
\]

where \( R \) is the number of selected positions in the domain and \( N \) is the number of structures in the non-redundant set. The Euclidean distance between Cα coordinates of the \( i \)th and \( j \)th positions in the \( n \)th structure is denoted as \( d_{n,ij} \).

Positions that are less conserved have a higher structural variation score \( (S_i) \). In order to give an estimate of how many positions should be included in the VH coreset, the position with the highest \( S_i \) was removed and the scores recomputed. The process was repeated until all positions were removed, noting the score at removal and reducing \( R \) by 1 upon each iteration. The point where the score decreased approximately linearly with \( R \) was used to choose a cut-off for the number of positions to include in the coreset (Supplementary Fig. S1).

Thirty-five positions were chosen for the VH coreset. An analogous procedure was performed for the light variable domain to form the VL coreset, again containing 35 positions (Supplementary Table SI). With this number of positions remaining, the value of \( S_i \) for the least conserved among the 35 positions was approximately the same for both the VH and the VL procedures.

Figure 1a shows the location of the 35 positions used for each coreset. As might be expected, these positions are predominantly located on the β-strands of the framework and form the core of each domain.

**Defining frames of reference and consensus structures**

To measure the VH–VL orientation, a definition of the location of each domain must be made. These locations are described using frames of reference. One method to define frames of reference is to use conserved features on each individual domain. Abhinandan and Martin achieved this by fitting vectors through conserved positions in the β-sheets of the VH–VL interface. This description of the domain position is sensitive to the relative spacing of a small number of points. The effect of this when measuring a torsion angle is small. However, measurement of other modes of rigid-body orientation would be affected to a greater extent.

To minimise the effect of local deformations, we used the coreset positions to register frames of reference onto the domains. In doing so, we capture the rigid-body locations of the whole of each domain and not just those positions at the interface.

The VH frame of reference was created as follows. The VH domains in the non-redundant dataset were clustered using CDHIT, applying a sequence identity cut-off of 80% over framework positions in the domain. One structure was randomly chosen from each of the 30 largest clusters.

This set of domains was aligned over the VH coreset positions using Mammoth-mult (Lupyan et al., 2005). From this alignment we extracted the Cα coordinates corresponding to
the same eight structurally conserved positions in the β-sheet interface as identified by Abhinandan and Martin (H36, H37, H38, H39, H89, H90, H91 and H92). Through the resulting 240 coordinates we fit a plane. This was done by taking the first two components of a principal components analysis. The plane is the frame of reference for the VH domain. A consensus VH structure was also created by taking the mean Cα coordinate for each of the VH coreset positions in the 30 aligned structures.

To register the reference frame plane onto an individual VH structure, we perform a superposition of the consensus structure to the coreset positions of the real VH domain, using TM-align (Zhang and Skolnick, 2005). The resulting transformation matrix can then be used to map the plane onto the structure of the individual VH domain.

An analogous procedure was performed to create the reference frame plane and the consensus structure for the VL domain. Here, positions L35, L36, L37, L38, L85, L86, L87 and L88 were used to fit the plane. Figure 1b shows the VH and VL reference frame planes and consensus structures when they have been mapped to the structure 1B4J_HL.

Choosing an axis to measure VH–VL orientation about

The procedure described above allows us to map the two reference frame planes onto any Fv structure. We can therefore think of measuring VH–VL orientation as equivalent to measuring the orientation between the two planes. To do this fully and in an absolute sense requires at least six parameters: a distance, a torsion angle and four bend angles. These parameters must be measured about a consistently defined vector that connects the planes. We used the vector that had the most conserved length in the structures in our non-redundant set defined as described later. This choice was made in order to maximise the variation in orientation, which may be described using angular measures. Thus, it effectively acts as the pivot axis of VH–VL orientation. We call this vector C.

To identify C, the reference frame planes were registered onto each of the structures in our non-redundant set and a mesh placed on each plane (Fig. 1b). Each structure, therefore, had equivalent mesh points and thus equivalent VH–VL mesh point pairs. The Euclidean distance was measured for each pair of mesh points in each structure. The pair of points with the minimum variance in their separation distance was identified. Figure 1c shows where these two points are located on the planes. The vector which joins these points is defined as C. As we can identify where these points are located on each individual domain, we can also map C onto every Fv structure and define a coordinate system about it in a consistent manner.

Defining a coordinate system and measures for VH–VL orientation

The coordinate system is fully defined using vectors, which lie in each plane and are centred on the points corresponding to C. H1 is the vector running parallel to the first principal component of the VH plane, while H2 runs parallel to the second principal component. They therefore lie approximately parallel and perpendicular to the strands in the VH β-sheet interface, respectively. L1 and L2 are similarly defined on the VL domain. Figure 1d shows the coordinate system defined on an Fv region.

To describe the VH–VL orientation we use six measures, a distance and five angles. These are defined in the coordinate system as follows:

- The length of C, dC
- The torsion angle, HL, from H1 to L1 measured about C.
- The bend angle, HC1, between H1 and C.
- The bend angle, HC2, between H2 and C.
- The bend angle, LC1, between L1 and C.
- The bend angle, LC2, between L2 and C.

The HL angle is a torsion angle between the two domains and it is similar to Abhinandan and Martin’s packing angle. The HC1 and LC1 bend angles are equivalent to tilting-like variations of one domain with respect to the other. The HC2 and LC2 bend angles describe twisting-like variations of one domain to the other.

These measures provide a method of placing the orientation of individual structures onto an absolute scale. By describing the pose of the VH and VL domains for all known Fv structures, one can compare both individual and groups of structures in a consistent manner.

Identification of Chailyan et al.’s interface types within the non-redundant set

Chailyan et al. have previously identified two types of VH–VL interface by clustering a curated set of antibody structures. They denoted these as clusters A and B. To investigate if and how these sets of structures differ in our orientation measures, we assigned each of the structures in our non-redundant set to one of these two clusters.

The residue identity at position L44 was found to be most informative of cluster membership by Chailyan et al. Structures in cluster A all have proline at L44. Three hundred and two of our 351 structures have this residue identity.

Structures in cluster B can have either phenylalanine, valine or isoleucine at L44. In Chailyan et al.’s dataset, these residues occurred in the ratio 24 : 5 : 2 and therefore cluster B was largely characterised as having phenylalanine at L44. In our non-redundant dataset we identified structures in cluster B with the residues phenylalanine, valine and isoleucine in the ratio 19 : 16 : 10. This difference in proportion of residues led us to stratify the structures in cluster B by the amino-acid type at L44 (phe-L44, val-L44 and ile-L44). Four structures in our dataset were not placed in either cluster. Three of these had asparagine at L44 and one had leucine at this position.

Random forest regression

Our method allows for each structure to be placed on an absolute scale of orientation. Previous studies have identified positions that are thought to be influential for determining VH–VL orientation (e.g. Abhinandan and Martin, 2010; Chailyan et al., 2011). Here, we wish to find both the positions that are important and their residue identity for which they become so. For instance, position L44 can be either proline, phenylalanine, valine, leucine, asparagine or isoleucine. Although the presence of phenylalanine at this position may be influential for the HC1 and LC1 angles, it may not influence the HL torsion angle. Or conversely, the presence of valine at L44 may be informative for HL but not for the HC1 or LC1 angles.

We performed a regression on the distributions of our angles using a random forest algorithm (Breiman, 2001). This
algorithm builds an ensemble, or forest, of decision trees based on input variables in order to predict a response variable. Each decision tree is built on a subset of the data to prevent over-fitting. Those input variables that are most informative of the response variable are identified by randomly permuting each of them in turn and assessing the reduction in prediction performance. Here, we use conditional inference trees as they have been shown to allow for input variable importance assignment that is unbiased by their entropy (Strobl et al., 2007).

In order to find both positions and residues that are important for the orientation angles, we created binary input variables for the regression algorithm. For instance, the residue at position L87 can be either phenylalanine, tyrosine, isoleucine or histidine. We therefore create four variables L87F, L87Y, L87I and L87H. These are 1 for a structure when the corresponding residue is present at L87 and 0 otherwise. The L87F binary variable therefore differentiates between structures that have phenylalanine at position L87 or any other residue instead.

Binary variables were created for each position that is in the VH–VL interface and therefore able to directly mediate inter-domain orientation. These positions were identified by examining the change in solvent accessible surface area (SASA) when the domains are taken individually and in complex with one another. SASA was calculated using JOY (Mizuguchi et al., 1998). Any position that had a change in SASA >15% in more than 5% of the structures in the non-redundant set was defined as an interface position. We found 64 such positions, 30 from the light chain and 34 from the heavy chain.

As positions that are highly conserved in their identity are unlikely informative about variation in VH–VL orientation, those variables that were 1 for 80% or more structures were discarded. Similarly, residues which appear at positions infrequently are unable to provide statistically significant information about the pose. Therefore those variables which accounted for <2% of structures were also discarded. We also combined those variables which were deemed to be highly correlated. The Jaccard distance (Jaccard, 1908) was used as a measure of variable dissimilarity, with those variables of a score of <0.4 being combined. For instance, the variables L42G and L43T were combined to form the variable L42G/L43T. This has the value of 1 when a structure has either a glycine at L42 or a threonine at position L43. Combinations only occur between variables relating to different positions.

This resulted in a total of 349 input variables. For each of the angular measures (HL, HC1, HC2, LC1 and LC1), 50 conditional inference forests were built using the R package ‘party’ (Strobl et al., 2009). From these, the mean importance measure was extracted and the variables ranked in accordance to it.

Results

Distributions of the measures

Our VH–VL orientation measures were calculated for all of the Fv regions in our dataset. The distribution of each measure is shown for the non-redundant set in Fig. 2a.

As described in the Materials and methods section, the vector C was chosen to have the most conserved length over the non-redundant set of structures. The distance, dc, is this length. It has a mean value of 16.2 Å and a standard deviation of only 0.3 Å.

The HL torsion angle has the largest range. It varies from $-72.2^\circ$ to $-45.14^\circ$. The angle with the smallest range is HC1, which varies from 64.8$^\circ$ to 77.4$^\circ$. However, variation in each angle cannot be compared on the same scale. For instance, a $1^\circ$ change in the HC1 angle is not equivalent to a $1^\circ$ change in the HC2 angle. They describe different directions of movement and therefore affect the physical coordinates of the domains by different amounts.

We compared our absolute measures with the relative measure of orientation RMSD, as defined by Narayanan et al. (2009). This was carried out by calculating the change in our absolute measures and the orientation difference as measured by RMSD, between every pair of structures in the non-redundant set. Our measure that was most correlated with RMSD was the HC2 angle (Spearman’s $\rho = 0.54$). However, for a given RMSD there can be a range of angle changes. This range of angle differences increases with RMSD. For instance, pairs of structures with an RMSD of 0.5 ± 0.1 Å range in HL angle difference from 0.0$^\circ$ to 4.8$^\circ$, while those with an RMSD of 3 ± 0.1 Å range from 0.0$^\circ$ to 19.6$^\circ$.

When plotting the angles against each other no direct correlation is observed between any of the measures. This suggests that the orientation of the variable domains does not vary from one structure to another about a single axis, i.e. one cannot define a single torsion angle to adequately define VH–VL orientation.

Comparing the orientation of interface-type clusters

In order to compare with Chailyan et al.’s work, structures in the non-redundant set were stratified into four subsets in accordance to the residue present at position L44 (see the section Identification of Chailyan et al.’s interface types within the non-redundant set). Figure 2 shows how these subsets of structures differ in their orientation measures.

Most striking is the location of the phe-L44 subset of structures on the LC1 and HC1 bend angle distributions. The structures in Chailyan et al.’s cluster B were predominantly from this subset. Those structures in the pro-L44 subset best represent Chailyan et al.’s cluster A. A Kolmogorov–Smirnov (K–S) test (Massey, 1951) comparing the distribution of HC1 angles for the phe-L44 subset with that for the pro-L44 subset showed that the former is significantly more acute ($P = 4.0 \times 10^{-12}$). The same is true for the LC1 angle ($P = 3.6 \times 10^{-10}$). These differences correspond to a tilting of the variable domains towards each other at the binding site in cluster B structures relative to those in cluster A. It is also indicative of the significantly smaller binding site area in cluster B structures than in cluster A structures observed by Chailyan et al.

We therefore propose that the difference in orientation that Chailyan et al. describe with their clusters is in the HC1 and LC1 angles. This is not the same mode of orientation variation described by Abhinandan and Martin’s packing angle. In fact, a change in Abhinandan and Martin’s torsion angle is approximately perpendicular to a change described by the HC1 or LC1 angles. Therefore, the apparently inconsistent sets of positions that these studies propose as influential for VH–VL orientation may be due to identifying positions that are important for different directions of pose.
However, Chailyan et al.'s cluster B also contained a small number of structures that did not have phenylalanine at L44. Instead, valine or isoleucine was present. We find that structures in our ile-L44 subset do not have significantly different preferences for orientation in any measure than structures with a residue other than isoleucine at position L44. In contrast, those structures in the val-L44 subset have significantly different HL, HC2 and LC2 angles to structures with a residue other than valine at position L44 (K–S test $P$ values $5.7 \times 10^{-5}$, $9.1 \times 10^{-4}$ and $2.0 \times 10^{-6}$, respectively). This is not the same mode of orientation differentiation that is found in phe-L44 structures and is more similar to the mode of orientation variation described by Abhinandan and Martin’s packing angle.

These results indicate that different residues at the same position may influence the VH–VL orientation in different directions. Residues at position L44 can discriminate between structures that have preferences for either the HL torsion angle or the HC1 and LC1 angles. This may explain why L44 was one of only two positions that both Chailyan et al. and Abhinandan and Martin assigned high importance in their descriptions of orientation.

Those structures with phenylalanine at L44 have light chains predominantly from the mouse IGLV1 subgroup. Those with valine at the same position instead are predominantly from the mouse IGLK10 subgroup. As these subsets were found to have distinct orientations, we investigated the

---

**Fig. 2.** Histograms showing the distribution of each of our VH–VL orientation measures for the non-redundant set of structures. Each antibody variable domain can be placed at a position in this structural space. The location on each distribution of structures with position L44 occupied by (a) proline or phenyalanine or (b) valine or isoleucine are shown in each measure. Each line represents the Gaussian density estimation for the relevant distribution.
effect of heavy and light subgroup pairings on the orientation measures. Further dependence on the subgroup type was not found. Similarly, a recent study compared Abhinandan and Martin’s packing angle with VH–VL subgroup pairing and found no particular preference for subgroup pairs (Jayaram et al., 2012). Therefore, we moved to consider the residue identity at individual positions for their influence on the orientation measures.

**Important positions and residues for determining VH–VL orientation**

Table I lists the top 10 positions and residues identified by the random forest algorithm as being important in determining each of our angular measures of VH–VL orientation. For instance, L87F is the highest scoring position and residue for the HL angle. Therefore, to model the structure of an antibody that has phenylalanine at position L87, one should only use template structures of antibodies that share this property in order to better predict the VH–VL orientation with respect to the HL angle. However, as this position does not score highly for the other measures, not using this information will not affect the prediction of the VH–VL orientation in the other angles.

Three of the positions that Abhinandan and Martin found to be influential for their packing angle (L44, L87 and H62) are also identified by our method to be influential for our similar torsion angle, HL. A further three positions, L38, L41 and L46, that Abhinandan and Martin found to be influential are identified as being important for at least one of our other measures.

Five of the eight positions that Chailyan et al., proposed as influential for VH–VL orientation also score highly for the HC1 or LC1 measures. As shown in the previous section, these angles best discriminate between the authors’ two clusters of structures. The remaining three positions were not interface positions and therefore were not included in our analysis.

The HC2 measure is found to have a strong dependence on heavy chain variables and especially on the position H62. Examination of those structures which have lysine at H62 and those with serine at H62 finds that they have different preferences for the HC2 angle. Those with lysine have significantly smaller HC2 angles than those with serine (K–S test, $P$ value = 2.0 $\times$ 10$^{-14}$). Further investigation of the relationship between HC2 and the residue at position H62 revealed that the size of the residue present is generally inversely related to the size of the angle. For instance, those structures with aspartic acid at H62 have small angles while those with alanine have large ones. The size of the amino-acid affects the packing at the domain interface and therefore the VH–VL orientation as measured by HC2.

**Location of important positions on the VH–VL interface**

Figure 3 shows the location on the variable domains of the positions we have identified to be influential for each angle. The HC1 and LC1 measures describe a tilting-like motion of one domain towards the other. The positions we identify as important for these angles tend to be in the core of the interface and predominantly on the VL domain. Whereas LC2 and HC2 describe a twisting-like motion of one domain with respect to the other. In this case, the positions tend to be on the periphery of the inter-domain interface (predominantly on the VL domain for LC2 and VH for HC2). The positions that are important for the HL torsion angle tend to also be important for either the HC2 or the LC2 measures (i.e. sites on the periphery of the interface).

**Variation in orientation between sequence-identical structures is dependent on antigen type**

In this section, we consider if our VH–VL orientation measures are informative with respect to antibody–antigen binding. To test the conservation of the VH–VL orientation in antibodies, sequence-identical structures with the same bound-state were identified in the full dataset. In the set of Fvs with antigens bound, 205 sequences were identified that had two or more structures. The difference in the VH–VL orientation angle was calculated for each pair of sequence-identical structures and the mean difference calculated for each sequence case. Similarly, in the set of Fvs with no antigen bound, 45 sequences were identified that had two or more structures and the mean difference in angles was calculated for each.

The variation was found to be different between the bound and unbound sets of structures in only the HL angle. The bend angles are more likely to be conserved in sequence-identical structures as large differences would imply a loss of contacts at the VH–VL interface. However, larger changes may occur in a torsion angle motion and still maintain the VH–VL contact surface.

Figure 4 shows the distributions of the HL angle variation for unbound structures of sequence-identical antibodies and

<table>
<thead>
<tr>
<th>Table I.</th>
<th>$X$ represents the variable L36V$/L38E$/L42H$/L43L?$/L44F$/L45T$/L46G$/L49G/L95H</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Angle</strong></td>
<td><strong>Top 10 important input variables</strong></td>
</tr>
<tr>
<td><strong>HC1</strong></td>
<td>L38$,L56P$, L41D$, L89A, L97V, L94N, L34H, L34N, L96W, L100A</td>
</tr>
</tbody>
</table>

$^a$Denotes those positions also found to be influential by Chailyan et al.

$^b$Denotes positions also found to be influential by Abhinandan and Martin.

Fig. 3. The location of positions that are found to be influential for the angular measures. Those positions which are influential for more than one measure are coloured in the order of priority: LC1, HC1, LC2, HC2 and HL. Positions that are deemed to be influential for the HC2 and LC2 measures are located on the periphery of the VH–VL interface, while those for the LC1 and HC1 measures pack into the centre of the interface.
bound structures of sequence-identical structures. The distribution of HL angle differences for those structures which are <90% sequence identical is also shown in order to demonstrate the background variation. We find that the bound structures of sequence-identical antibodies have a more conserved HL angle than unbound structures. Although not a direct indication of the dynamics of the molecule, this result may reflect the effect of complex formation reducing the structural space available to the antibody. If this is true, then the degree to which the structure is stabilised is likely to be dependent on the size of the antigen it binds.

To test whether there is any dependence of orientation variation on antigen size, we stratified the sets of sequence-identical structures into three types of antibodies: hapten binding, peptide (or carbohydrate) binding and protein binding. There were 99 protein binding, 38 peptide binding and 68 hapten-binding sets of bound structures and 15 protein binding, 11 peptide binding and 19 hapten-binding sets of unbound structures. Figure 5 shows the distributions in HL angle variation for the bound and unbound sets stratified in this way. For protein binders, we found that the variation in unbound structures was significantly larger than the variation in bound structures (P value = 0.0014). However, neither the peptide nor the hapten distributions were significantly different between unbound and bound forms. The variation in angle in unbound structures is also significantly larger for protein binders than for hapten binders (P value = 0.0076), while the variation in bound structures is not significantly different.

These results suggest that the VH–VL orientation for protein-binding antibodies is more flexible than for hapten-binding antibodies. However, upon binding, both types of antibodies are found to have a similar degree of conservation in variable domain pose. The flexibility of a protein leads to a higher entropic cost for binding. Therefore, this result may be due to the fact that larger antigens are able to overcome this cost, while smaller hapten antigens require a more rigid binding partner. This result is physically intuitive and statistically significant. However, the number of data points is small and the analysis would benefit from the availability of more structures, especially in the unbound form.

**ABangle**

We have implemented our method of calculating the VH–VL orientation in antibody structures in the computational tool, ABangle. Given a structure in PDB format, the software fully automates the procedure of recognising all Fv regions present and will calculate and report the orientation measures for each VH–VL pair. This includes calculation for multiple NMR models and an option for automation of orientation calculation for single chain Fvs. The distribution of angles for single chain Fvs is generally similar to that of the standard VH–VL pairs.

ABangle also allows for the analysis of VH–VL orientation. Individual Fv regions found in the PDB can be retrieved using their PDB code and its chain identifiers. Sets of
structures can also be selected by a number of properties including residue identity at a Chothia position, species, heavy or light chain subgroup and CDR loop length. The orientation of these structures can then be visualised in two ways: as plot of the distribution of the orientation measures against the non-redundant set background, e.g. Fig. 2b; or using PyMol (Schrödinger, 2012) by aligning all the structures to either the VH or VL consensus structure.

We provide ABangle at http://opig.stats.ox.ac.uk/webapps/abangle.

Discussion

In this paper, we present a method to fully characterise the VH–VL orientation of antibody structures in an absolute sense. This allows us to investigate not just the relative changes between Fv regions, but how this change relates to variation observed in all structures.

We use our method to explain why two previous studies (Abhinandan and Martin, 2010; Chailyan et al., 2011) identify different framework positions as important for the orientation of the VH and VL domains. We find that the difference between the two clusters identified by Chailyan et al. is related to a change in our LC1 and HC1 angles, while the difference described by Abhinandan and Martin, the VH–VL packing angle, relates instead to a change in our HL torsion angle. Thus, the apparent inconsistency in the positions that these studies find influential for pose is because they have described approximately perpendicular modes of variation.

Our orientation measures have allowed us to investigate which positions and their residue identity affect pose in different directions. We find similar positions to both Abhinandan and Martin and Chailyan et al. in the analogous modes of orientation and identify others that may have a significant influence for different modes of variation. Our measures also offer insight into structural variation between bound and unbound forms of antibodies. We find that the variation in VH–VL orientation in antibodies in their unbound form is dependent on the size of antigen they bind. However, in the bound form, no such dependence is found suggesting a reduction in conformational space available to the antibody.

Our method has been implemented in the computational tool, ABangle. This allows researchers to investigate the structural space of antibodies. We have demonstrated its use for the applications of comparing sets of structures, finding influential positions and investigating the variation of orientation in homologues. It could also be used to compare the orientation of specific antibodies, especially in their unbound and bound forms. The orientation measures allow absolute scales of variation to be quantified. They can therefore be incorporated into Fv modelling protocols as a framework for modulating the VH–VL pose or for model assessment. ABangle’s ability to automatically and rapidly calculate the VH–VL orientation of a number of structures also lends itself to the investigation of the conformational space observed in nuclear magnetic resonance models of the Fv. Similarly, investigations could be made into the structural space of single chain Fvs and the effect that removing the constant region has on the orientation of the variable domains. ABangle opens many possible avenues of research and is available at http://opig.stats.ox.ac.uk/webapps/abangle.

Supplementary data

Supplementary data are available at PEDS online.

Acknowledgements

Terry Baker is thanked for useful discussions concerning the choice of axis about which to measure the VH–VL orientation.

Funding

This work was supported by the Engineering and Physical Sciences Research Council, Roche Diagnostics GmbH and UCB Pharma.

References