Conformational analysis of antibody CDR loops upon binding

Chu'nan Liu, Oliver E.C. Hood, Lilian Denzler, and Andrew C.R. Martin
Structural and Molecular Biology, Division of Biosciences
University College London
Gower Street, London WC1E 6BT

November 17, 2023

Abstract

Antibodies have increasingly been developed as drugs with over 100 now licenced in the US or EU. During development, it is often necessary to increase or reduce the affinity of an antibody and rational attempts to do so rely on having a structure of the antibody-antigen complex often obtained by modelling.

The antigen binding site consists primarily of six loops known as complementarity-determining regions (CDRs), and an open question has been whether these loops are flexible — in other words, does the unbound conformation reflect the bound conformation. Existing surveys of antibody-antigen complex structures have only looked at CDR flexibility in case studies or small-scale surveys. With an increasing number of antibody structures deposited in the Protein Data Bank, a large-scale survey of CDR conformational change during binding is now possible. To this end, we built a dataset, AbAgDb, that currently includes 177 antibodies with high-quality CDRs, each of which has at least one bound and one unbound structure. We analysed the conformational change of the C_α backbone of each CDR upon binding and found that, in most cases, the CDRs (other than CDR-H3)

show minimal movement, while 70.6% and 87% of CDR-H3s showed global Cα RMSD $\leq$1.0Å and $\leq$2.0Å, respectively. We also compared bound CDR conformations with the conformational space of unbound CDRs and found most of the bound conformations are included in the unbound conformational space. In future, our results will contribute to developing insights into antibodies and new methods for modelling and docking.

1 Introduction

Antibodies are increasingly used as drugs owing to high affinity and specificity and the ability to bind targets that are undruggable with small molecule drugs. At the time of writing, there are 124 antibody-based drugs approved in the United States or European Union with 19 novel antibody therapeutics having been approved since January 2022 and 19 currently in review (Antibody Society, *Antibody therapeutics approved or in regulatory review in the EU or US*, `https://www.antibodysociety.org/resources/approved-antibodies/`, 24 July 2023). Antibody-based drug development relies largely on time- and cost-intensive experimental approaches, which potentially can benefit substantially from computational methods such as structure- and machine learning-based design[1, 2, 3]. An important step in structure-based design is to identify antibody-antigen interacting sites and obtain the structure of the complex[2]. This would allow for further engineering of the binding sites to obtain antibodies with desirable binding affinities (increased or decreased), an increase in affinity through rational design based on a modelled antibody having been first achieved by Roberts *et al.* in 1989[4]

Antigen binding sites are the regions of the antibody surface that bind to their cognate antigens. They consist, primarily, of six complementarity-determining regions (CDRs), or 'hypervariable loops', three from the heavy chain and three from the light chain[5]. Previous surveys of CDR loop structures showed that, with the exception of CDR-H3, the main-chain conformation of the other five loops can be grouped into 'canonical structures' which can be identified by sequence templates[6, 7, 8]. However, the ques-

tion of whether the canonical structures, or the conformation of CDR-H3, are retained upon binding, has not been considered explicitly, and the complexed/uncomplexed state has generally been ignored in existing studies.

There are three models describing ways in which protein-protein (including antibody-antigen) interactions can occur: (1) The 'lock-and-key' model states that there is little conformational change upon binding. (2) The 'induced-fit' model suggests that the bound conformation at the interface (of one or both partners) is induced by binding with the interface of the unbound structure(s) having a distinct and different conformation[9, 10]. This will incur an enthalpic penalty as the conformation of one (or both) structures will have to move away from the energy minimum seen in the unbound conformation. Thus some of the energy gained from binding is 'wasted' in stressing the conformation of one or both proteins. (3) The 'Conformational-selection' model[11] suggests that one, or both, structures are mobile and that structural studies have 'frozen out' a single conformation of the free antibody that happens to be different from that present in the complex. However, this will incur an entropic penalty unless both proteins are able to move in concert in the complex. Recent surveys of general protein-protein interactions have suggested combinations of models, including conformational-selection and induced-fit[9].

In the case of antibodies which undergo a rapid evolutionary process to optimise binding through somatic hypermutation, it would be reasonable to expect that germline antibodies (which need to bind a range of antigens without a need for high affinity), may fit the induced-fit or conformational-selection models, with affinity maturation leading to higher affinity through a lock-and-key interaction. Indeed, this has been supported by observations of multiple pre-existing conformations of the same antibody primarily in germline antibodies[12], but less frequently in mature antibodies[13].

To aid in developing new computational methods for antibody-antigen complex prediction and for understanding antibody-antigen interactions, we built a database, AbAgDb (built upon AbDb[14]), that includes both unbound and bound conformers for each antibody. The current version contains 177 groups of antibody structures with those in the same group having the

same sequence and at least one unbound and one bound conformation. We examined CDR loop flexibility by analysing conformational change between unbound and bound conformer pairs for each CDR. We also analysed their binding mode by comparing bound conformations against the unbound CDR conformational space represented by canonical structures. CDR canonical structure clusters were derived by employing a similar approach to previous studies[7, 8], but using only unbound structures and used 1,091 CDRs from quality-filtered unbound antibodies obtained from AbDb.

2 Materials and Methods

Because there may be multiple structures of the same antibody (both free and with the same or different antigens) we define the term 'antibody' to mean an antibody with a distinct sequence present in any such set, while we define the term 'entry' to refer to each individual structure present in AbDb for each antibody.

2.1 Antibodies with both unbound and bound conformations

All files used in this work were collected from the latest release (manuscript in preparation) of AbDb[14] in which file names are formatted as the four-character PDB code, an underscore, an integer index (to distinguish antibody entries — i.e. multiple structures within a PDB file), followed by optional characters indicating the antigen type: protein and peptide (**P**), hapten (**H**), nucleic acid (**N**). An empty antigen type character indicates an unbound antibody. To non-redundantize antibodies in AbDb, sequences of all antibody structures (12,205 entries) are collected, split by chain and merged into a single FASTA file containing 21,536 chains, used as input to CD-HIT[15] and clustered at a sequence identity of 100%. This way, each heavy or light chain is assigned to a cluster and each conventional antibody (V_H + V_L) can be represented by a pair of cluster IDs (single-chain antibodies are represented by a single cluster ID). Antibodies with the same cluster-ID or ID

pair were grouped together as they have the same sequence. This led to 3,320 unique V_H/V_L antibodies (9,622 entries) and 836 unique single-chain antibodies (2,292 entries). These were then filtered to remove any problematic antibodies that could not be numbered automatically and only those having both bound and unbound structures were retained, leading to 3,040 entries representing 559 antibodies.

2.2 CDR structure quality filtering

Quality filtering started with the 3,040 entries collected in the last step, numbered using the Martin scheme (a refinement of Chothia numbering in which the position of framework insertions and deletions is also structurally correct[16]). We adapted the filtering procedure from North *et al.*[8] to create the pipeline shown in Figure 1.

The pipeline retains only entries that represent Fv structures (with both V_H and V_L domains) having a resolution of at least 2.8Å and which are of high quality; entries with missing residues, large B-factors and non-proline residues having a cis-peptide bond in any of the CDRs are eliminated.

2.3 CDR loop conformation analysis upon binding

The conformations of the CDRs themselves may change on binding and this can be evaluated by calculating a 'local' Cα RMSD by comparison of the CDR loops in the bound and unbound structures. Alternatively, a CDR may move with respect to the supporting framework, which we refer to as loop 'flapping'. This can be evaluated by calculating a 'global' RMSD where the supporting framework is fitted and the Cα RMSD is calculated over the CDR and comparing this with the local Cα RMSD. While the global RMSD will be affected by *both* any local conformational change and by loop flapping, a high *global* Cα RMSD with a low *local* Cα RMSD will indicate significant loop flapping. When calculating global RMSD, fitting is performed only on the V_H framework for the heavy-chain CDRs and on the V_L framework for the light-chain CDRs. This is to avoid shifts in CDR positions resulting from changes in the V_H/V_L packing angle which could result from antigen binding[17].

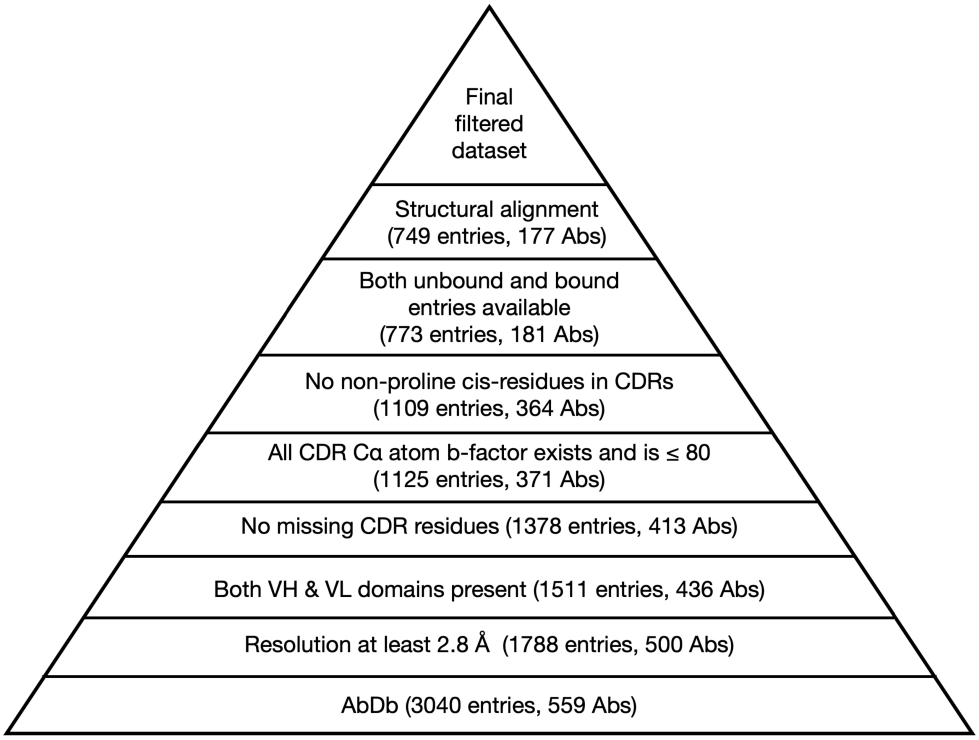

Figure 1: **Filtering AbDb files**. Starting from the bottom, we eliminate structures with resolution worse than 2.8Å and retain antibodies (Abs) that have both heavy and light variable domains, then eliminate files with missing residues in any of the 6 CDRs, or C_α atom B-factor is missing (i.e. 0) or > 80, or a non-proline cis-residue is present in an unbound antibody, leading to 364 Abs with 1109 entries. We then retained antibodies with both unbound and bound structures (181 Abs with 773 entries) and performed global and local fitting. Finally, we eliminated unbound/bound structure pairs whose framework region showed $\geq$1.0Å global Cα RMSD to minimise the impact of the framework region on CDR conformational change and followed by re-checking that both unbound and bound structures are available for an antibody, which led to the elimination of four antibodies. This led to a final set of 749 entries representing 177 antibodies. See Supplementary File 'Supp1-unbound-and-bound-abs.csv' for the initial dataset of entries with both bound unbound structures from AbDb. See Supplementary File 'Supp2-antibody-filtering.csv' for information on entries retained and rejected at each step.

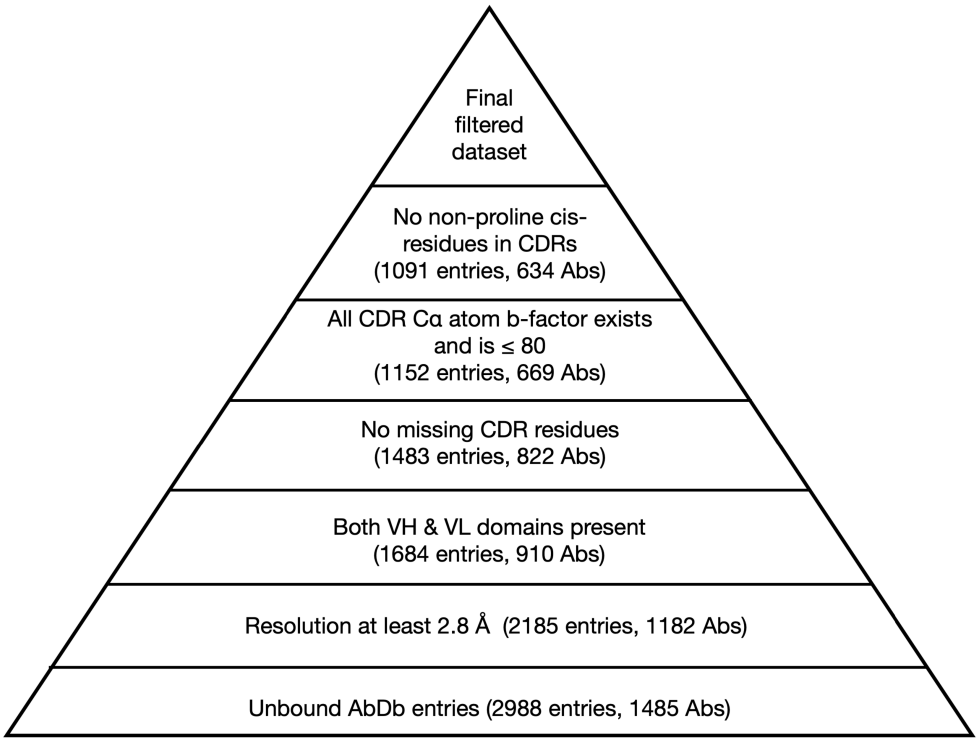

Figure 2: **Filtering AbDb unbound structures**. Filtering steps for unbound antibody structures using the same protocol as in Figure 1 with the number of entries and antibodies retained at each step. See Supplementary File 'Supp3-unbound-filtering.csv' for information on entries retained and rejected at each step.

CDRs were defined using the AbM (Martin) loop definition[18, 19]: CDR-L1 (L24–L34), CDR-L2 (L50–L56), CDR-L3 (L89–L97), CDR-H1 (H26–H35), CDR-H2 (H50–H58), CDR-H3 (H95–H102) using Chothia or Martin numbering[16]. Structure fitting and RMSD calculation was performed using ProFit (an implementation of the McLachlan fitting algorithm[20] available at `http://www.bioinf.org.uk/software/profit/`). The fitted framework region constitutes non-CDR residues, but excludes the N-terminal two residues (H1, H2, L1, L2) and the C-terminal six residues (H109–H113, L106–L110) owing to high flexibility that can lead to fitting errors and sometimes leads to missing residues in x-ray crystal structures.

2.4 CDR canonical structure clustering

Unbound CDR structures were clustered using an updated procedure based on the work of Martin and Thornton[7] and of North *et al.*[8]. We collected *all* unbound antibodies (numbered according to the Martin scheme[16]) from AbDb with both heavy and light variable domains (V_H/V_L) and filtered them using the same quality criteria described in Figure 1 with the exception of the requirement for having both bound and unbound structures. This led to a set of 1,091 unbound entries (Figure 2).

CDR loops were grouped based on CDR type (i.e. CDR-L1, CDR-L2, CDR-L3, CDR-H1, CDR-H2 and CDR-H3), and each group was further partitioned according to loop length and the position of any cis-proline residues. We refer to such groups as CDR '**L**ength and **R**esidue **C**onfiguration' (LRC) groups. For example, the LRC group 'L3-9-cis95' denotes a group of CDR-L3 loops composed of 9 residues with a cis-proline at position L95.

CDR loops were then converted to vectors of sine and cosine values of dihedral angles (ϕ and ψ) of each residue. Each LRC group was converted to a matrix of shape $n \times 4L$ where n denotes the number of loops, and L denotes the loop length. For example, a loop of length 9 (e.g. group 'L3-9-cis95') is converted to a 36-dimensional vector, and a set of n loops would be represented as $n \times 36$ matrix:

$$\begin{bmatrix} \sin\phi_1^1,\ \cos\phi_1^1,\ \sin\psi_1^1,\ \cos\psi_1^1,\ \ldots,\ \sin\phi_9^1,\ \cos\phi_9^1,\ \sin\psi_9^1,\ \cos\psi_9^1 \\ \ldots \\ \sin\phi_1^n,\ \cos\phi_1^n,\ \sin\psi_1^n,\ \cos\psi_1^n,\ \ldots,\ \sin\phi_9^n,\ \cos\phi_9^n,\ \sin\psi_9^n,\ \cos\psi_9^n \end{bmatrix}$$

Each matrix was then clustered using the Affinity Propagation (AP) method[21]. AP clustering is a message-passing-based method which has the advantage over other clustering methods of taking all data points into consideration for deciding cluster representatives. Each data point in this case is the $4L$-element vector description of a loop as described above. The resulting clusters are called '**AP clusters**'. The distance between a pair of loops of the same length is calculated as the squared Euclidean distance.

For example, the distance between a pair of loops of the same length L is calculated as:

$$f(a,b) = (\sin a - \sin b)^2 + (\cos a - \cos b)^2 \tag{1}$$

$$D(i,j) = \sum_{r=1}^{L} f(\phi_r^i,\ \phi_r^j) + f(\psi_r^i,\ \psi_r^j) \tag{2}$$

where i and j denote the indices of two loop conformations of interest, r denotes a residue index, and L is the loop length. The similarity between two data points ($S(i,j)$) is the negative squared Euclidean distance (Equation 3, below). The self-similarity S_{self}, which affects the final number of clusters (as described by North *et al.*[8]) is set to the mean of similarities between all non-self pairs of CDR loops within an LRC group i.e.

$$S(i,\ j) = -D(i,j) \tag{3}$$

$$S_{self} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} S(i,\ j) \tag{4}$$

After clustering in torsional space, to decide whether a pair of AP clusters are similar in Cartesian space, we compared all possible pairs of cluster exemplars using the same criteria described by Martin and Thornton[7]. As explained by Martin and Thornton, a difference in backbone torsion angles may correspond to a much smaller movement in Cartesian space. A pair of AP clusters are merged if their exemplar CDR structures meet all three conditions: after fitting CDRs (on C_α atoms), the Cα RMSD between the exemplars <1.0Å, the maximum distance between C_α atoms at equivalent positions <1.5Å, and the maximum distance between C_β atoms at equivalent positions <1.9Å. We refer to this merging criteria as 'CartesianCriteria'. The final merged AP clusters are called '**Canonical clusters**'.

We ensured that this new clustering protocol was consistent with the Chothia canonical classes (i.e. none of the clusters used here contained more than one Chothia canonical class). The method used to assign the structures

described by Martin and Thornton[7] to clusters, together with a comparison of the clustering used here with that work (including the Chothia canonical assignments) is given in Supplementary File 'ClusterComparison.pdf' (Tables S1–S5).

2.5 Comparison of bound CDR loop conformations with unbound conformational space

The procedure to compare a bound CDR conformation with the unbound CDR conformational space is illustrated in Figure 3 and uses the following approach:

1. The conformational space of a CDR of a given length within an LRC group is represented as a set of AP clusters $\mathbf{A} = \{a_1, a_2, \ldots, a_i\}$ and a set of Canonical clusters $\mathbf{C} = \{c_1, c_2, \ldots, c_k\}$. As a result of post-cluster Cartesian merging, one canonical cluster may contain multiple AP clusters, and consequently, each AP cluster can be mapped to a single Canonical cluster.

2. A single CDR conformation is denoted as a_i^j where i denotes the AP cluster and j denotes the conformation within that cluster. The representative (or 'exemplar') of an AP cluster a_i is denoted as a_i^e. As explained above, each a_i^j is represented as a vector of ϕ and ψ sine and cosine values giving a vector size of $4 \times L$ where L is the loop length.

3. The radius of an AP cluster $r(a_i)$ is calculated as

$$r(a_i) = \max\left(D(a_i^e, a_i^j)\right)$$

   where $D()$ is defined in Equation 2 above. In other words, this is the maximum torsional distance between the AP cluster exemplar and any of its members.

4. The query bound conformation is transformed into a trigonometric vector as described previously, denoted as x.

5. For the unbound conformation of the same antibody, we identify its AP cluster a_u and Canonical cluster c_u. In the example in Figure 3, representing the LRC group 'H2-10-allT', the unbound conformation belongs to AP cluster a_u in Canonical cluster c_u, which here is c_1.

6. We then locate the closest AP cluster to the bound conformation x (i.e. the AP cluster having the minimum value of $D(x, a_i^e)$) denoted as a_B,

7. If $D(x, a_B) \leq r(a_i)$ (i.e. the conformation falls within the radius of the cluster), then x is a member of AP cluster a_B and the associated Canonical cluster, c_B, is identified.

8. If $D(x, a_B) > r(a_i)$ (i.e. the conformation falls outside the radius of the cluster), then x is not a member of an existing AP cluster, but if it passes the 'CartesianCriteria' (defined above), then it will be a member of the Canonical cluster c_B of which a_B is a member. If it does not pass the CartesianCriteria, then conformation x is a novel conformation not observed in the unbound structures.

Comparing the AP cluster and Canonical cluster labels of such unbound/bound conformation pairs i.e. comparing a_u with a_B and c_u with c_B, we can define four types of conformational change: 'Identical AP cluster', 'AP-cluster shift', 'Canonical cluster shift', and 'Non-canonical conformation' as described in Table 1 and Figure 3.

3 Results

3.1 Dataset of antibodies with unbound and bound conformers

As stated above, we use the term 'antibody' to refer to any set of bound or unbound structures having the same sequence and 'entries' to refer to the individual structures. As described in the Materials and Methods, Figure 1 shows a schematic of the filtering procedure with the number of AbDb entries or antibodies retained at each stage indicated. After filtering, we identified

Table 1: Types of conformational change upon binding

Type name	**Definition**
Identical AP cluster	The bound conformation belongs to the same AP cluster as that of its unbound conformation. i.e. there is negligible conformational change (Figure 3(1) $a_B = a_u, c_B = c_u$).
AP-cluster shift	The bound conformation belongs to a different AP cluster from that of its unbound conformation, but is within the same Canonical cluster as the unbound conformation. i.e. there is a larger conformational change in torsional space, but would be placed in the same Canonical class (Figure 3(2), $a_B \neq a_u, c_B = c_u$).
Canonical-cluster shift	The bound conformation is different from the unbound conformation but matches a different Canonical cluster observed in another antibody (Figure 3(3), $a_B \neq a_u, c_B \neq c_u, c_B \in C$).
Non-canonical conformation	The bound conformation is different from the unbound conformation and is not seen in any other unbound the CDR unbound antibodies (Figure 3(4), $a_B \neq a_u, c_B \neq c_u, c_B \notin C$).

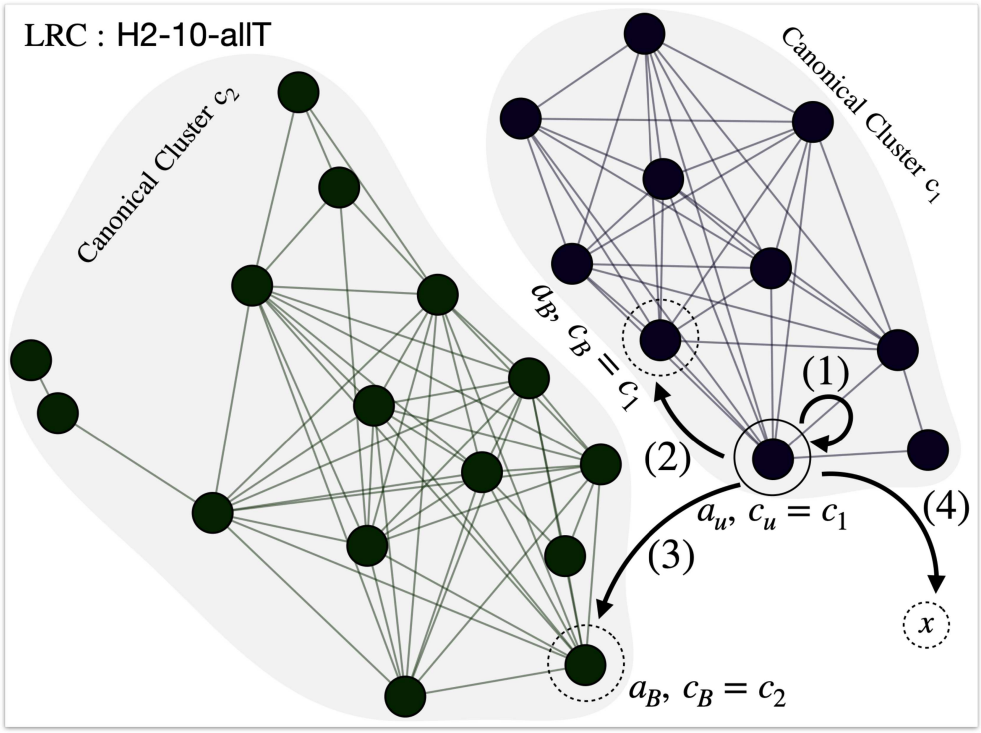

Figure 3: **Conformational change types**. The two major Canonical clusters (i.e. subgraphs) of the LRC group 'H2-10-allT' are shown and are denoted as c_1 and c_2. Each Canonical cluster consists of AP clusters (i.e. nodes). The AP cluster of the unbound conformer is labelled as a_u, and its Canonical cluster is denoted as c_u, in this case, c_1. The AP cluster and Canonical cluster of a bound conformation x are denoted as a_B and c_B on the graph. Comparing a_B, c_B with a_u, c_u, we can define four types of CDR conformational change upon binding: **(1) Identical AP cluster**: the bound conformer is merged with the same AP cluster as the unbound. **(2) AP-cluster shift**: the bound conformer is merged with a different AP cluster in the same Canonical cluster as the unbound. **(3) Canonical-cluster shift**: the bound conformer is merged with an AP cluster in a Canonical cluster different from the unbound; **(4) Non-canonical conformation**: x is not merged with any AP cluster or Canonical cluster. See Table 1.

177 antibodies that each had at least one bound and one unbound structure from a total of 749 AbDb entries (369 unbound and 380 bound).

3.2 CDR loop movement upon binding

Global and local fitting were performed on all possible unbound/bound pairs of entries for each antibody, and the distribution of conformational change, as represented by the median of the Cα RMSD for those pairs, is shown in Figure 4. For example, the mouse anti-hen egg white lysozyme antibody HyHEL-63 (PDB: 1dqq) has three associated unbound entries (1dqq_0, 1dqq_1, 1dqm_0) and three bound entries (1nbz_0P, 1dqj_0P, 1nby_0P). Thus, in this example, nine Cα RMSD values are obtained for each CDR, from which the medians are calculated and used to plot the distribution.

As expected, the Cα RMSD from local fitting (representing shape change within a loop) is lower than from global fitting. As shown in Figure 4, when globally fitted, the third quartiles of non-CDR-H3 loops are close to 0.5Å and over 70% of non-CDR-H3 loops (except CDR-H2, 68%) showed a global Cα RMSD of under 0.5Å (Table 2(1)). This suggests these loops generally show no significant backbone movement upon binding, since a Cα RMSD of up to 0.5Å is frequently seen between multiple structures of the same protein crystallized in different conditions.

Meanwhile, when locally fitted, the third quartiles for non-CDR-H3 loops dropped to 0.25Å (Figure 4), and over 90% of non-CDR-H3 loops showed a local fitting Cα RMSD below 0.5Å (Table 2(2)). Comparing global and local fitting, we observed the average percentage of antibodies with a CDR having Cα RMSD below 0.5Å increased from 68.2% (average of percentages in Table 2(1), column 1) to 90.7% (average of percentages in Table 2(2), column 1) and those within the range of 0.5Å and 1.0Å dropped from 24.2% to 5.3% (average of column 2 percentages in Table 2(1) and Table 2(2) respectively) This clearly demonstrates that part of the global Cα RMSD is caused by a small degree of loop 'flapping'. We also calculated the difference between global and local Cα RMSD for each antibody as an indicator of the amount of loop flapping (Table 2(3)). Generally, we observed an average of 70.5%

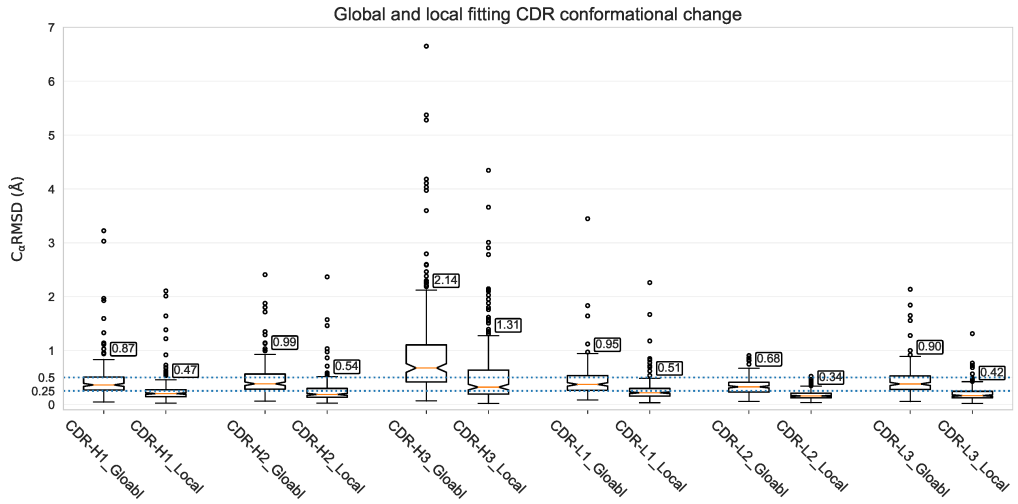

Figure 4: **CDR conformational change distribution from global and local fitting**. Distributions of conformational change (measured as Cα RMSD) from global fitting (superposing on framework region with suffix "_FR") and local fitting (superposing on CDR region with suffix "_CDR") are plotted. Each box represents the first quartile, median and the third quartile with the whiskers represent the lower and upper fence. Outliers are shown as circles using upper fence ($Q3 + 1.5 \times IQR$ meaning 3rd quartile plus one and half inter-quartile range), and the upper fence values for each box plot are labelled. To assist comparison, Cα RMSD at 0.25Å and 0.5Å are plotted as dashed lines.

Table 2: Numbers of antibodies in different RMSD ranges.

(1) Global fitting Cα RMSD (Å)						
CDR	≤ 0.5	$(0.5, 1.0]$	$(1.0, 2.0]$	$(2.0, 3.0]$	$(3.0, 4.0]$	> 4.0
H1	131 (74%)	36 (20%)	8 (5%)	0	2 (1%)	0
H2	120 (68%)	46 (26%)	9 (5%)	1 (1%)	0	1 (1%)
H3	60 (34%)	65 (37%)	29 (16%)	15 (8%)	2 (1%)	6 (3%)
L1	127 (72%)	46 (26%)	3 (2%)	0	1 (1%)	0
L2	155 (88%)	22 (12%)	0	0	0	0
L3	130 (73%)	42 (24%)	4 (2%)	1 (1%)	0	0

(2) Local fitting Cα RMSD (Å)						
CDR	≤ 0.5	$(0.5, 1.0]$	$(1.0, 2.0]$	$(2.0, 3.0]$	$(3.0, 4.0]$	> 4.0
H1	163 (92%)	9 (5%)	3 (2%)	2 (1%)	0	0
H2	166 (94%)	7 (4%)	3 (2%)	1 (1%)	0	0
H3	120 (68%)	28 (16%)	20 (11%)	6 (3%)	2 (1%)	1 (1%)
L1	167 (94%)	7 (4%)	2 (1%)	1 (1%)	0	0
L2	176 (99%)	1 (1%)	0	0	0	0
L3	172 (97%)	4 (2%)	1 (1%)	0	0	0

(3) Difference between global and local fitting Cα RMSD (Å)						
CDR	≤ 0.25	$(0.25, 0.5]$	$(0.5, 1.0]$	$(1.0, 2.0]$	$(2.0, 3.0]$	> 3.0
H1	138 (78%)	32 (18%)	6 (3%)	1 (1%)	0	0
H2	122 (69%)	44 (25%)	9 (5%)	1 (1%)	0	1 (1%)
H3	72 (41%)	57 (32%)	33 (19%)	12 (7%)	3 (2%)	0
L1	141 (80%)	33 (19%)	2 (1%)	1 (1%)	0	0
L2	153 (86%)	22 (12%)	2 (1%)	0	0	0
L3	122 (69%)	44 (25%)	8 (5%)	3 (2%)	0	0

The percentage of each count relative to the entire dataset (177 antibodies) is given in parentheses. The difference between global and local fitting Cα RMSD, which reflects loop flapping, is calculated as the absolute value of the difference between the global and local Cα RMSD of each pair of unbound/bound entries.

	≤1.0Å		≤2.0Å	
	Non-H3	H3	Non-H3	H3
Global	96.7%	70.6%	99.3%	87.0%
Local	98.5%	83.6%	99.5%	94.9%

Table 3: Summary of global and local C_α fitting for non-CDR-H3 loops and CDR-H3 loops. The percentage of loops having Cα RMSD values ≤1.0Å and ≤2.0Å are shown.

of CDRs in antibodies that showed a difference of up to 0.25Å (average of percentages in Table 2(3), column 1; i.e. no loop flapping) and 21.8% of antibodies between 0.25Å and 0.5Å (average of percentages in Table 2(3), column 2; i.e. minimal flapping).

The exception is CDR-H3 with a boxplot upper fence value (see legend to Figure 4) of 1.31Å from local fitting (Figure 4). However, this is still lower than the upper fence value of 2.14Å from global fitting. The percentage of antibodies showing a Cα RMSD below 0.5Å increases from 34% for global fitting to 68% for local fitting (CDR-H3 in column 1 of Table 2(1) compared with Table 2(2)). Thus, CDR-H3 more frequently shows larger scale flapping movements than the other CDRs. 19% of CDR-H3 loops showed a Cα RMSD difference (local *vs.* global) between 0.5Å and 1.0Å, whereas this value was ≤5% for non-CDR-H3 loops (CDR-H3 in Table 2(3), column 3). Thus loop 'flapping' is more common in CDR-H3 upon binding than in non-CDR-H3 loops.

Table 3 summarises these findings for CDR-H3 and non-CDR-H3 loops at cutoffs of ≤1.0Å and ≤2.0Å. 96.7% of non-CDR-H3 loops show a global fit with a Cα RMSD of ≤1.0Å, while 99.3% show a global fit of ≤2.0Å. The local fitting values rise to 98.5% and 99.5%, respectively. This suggests that non-CDR-H3 loops rarely change conformation on binding. Further, the fact that the percentage of non-CDR-H3 loops with local and global Cα RMSD ≤2.0Å is virtually unchanged suggests that while some loop flapping occurs, it is only a small effect (mostly ≤1.0Å).

On the other hand, 70.6% of CDR-H3 loops show a global fit with a Cα RMSD of ≤1.0Å, while 87.0% show a global fit of ≤2.0Å. The local

fitting values rise to 83.6% and 94.9% respectively. This suggests that while changes in CDR-H3 conformation on binding are still uncommon, they are much more common than for the non-CDR-H3 loops. The fact that ∼8% and ∼13% more of the CDR-H3 loops have local Cα RMSD of ≤2.0Å and ≤1.0Å respectively suggests both that loop flapping is much more common in CDR-H3 than it is in the other CDRs and that the degree of flapping is greater than with the other CDRs.

CDR conformational change from global fitting was also plotted against loop length (Figure 5). A single loop length group dominates CDR-H1, CDR-H2, CDR-L2, and CDR-L3. CDR-L1 has two major groups — 11 and 16 residues. In contrast, CDR-H3 has diverse loop lengths, with the majority between 7 and 16 residues. For CDR-H3 loops, little correlation between conformational change and loop length was observed (Spearman rank correlation coefficient between global Cα RMSD and loop length is 0.13; p-value of 0.08). However, we do observe a larger conformational change when the loop becomes longer for ten antibodies with CDR-H3 loop length ≥ 17 residues: the CDR-H3 global Cα RMSD from such antibodies ranges between 0.93Å and 6.65Å, see Figure 5(H3). Though it appears that the longer loops might imply a larger conformational change upon binding, this may be a result of the limited number of antibodies with such long CDR-H3 loops (only ten antibodies have a CDR-H3 loop longer than 16 residues which only accounts for 7% of entries).

3.3 CDR conformational clustering

The LRC distribution of each CDR is shown in Figure 6 sorted by group size. Both CDR-L2 and CDR-H1 are dominated by a single group. 'L2-7-allT' that accounts for 99% of entries for CDR-L2 while 'H1-10-allT' accounts for 89% of entries for CDR-H1 where the second biggest group ('H1-11-allT') only represents 5%. CDR-H2 is dominated by 'H2-10-allT' accounting for 68% of entries followed by two smaller groups 'H2-9-allT' (25%) and 'H2-12-allT' (6%). CDR-L1 is dominated by 'L1-11-allT' (46%) followed by 'L1-16-allT' (14%) and six smaller groups each of which accounts for less than 8% of

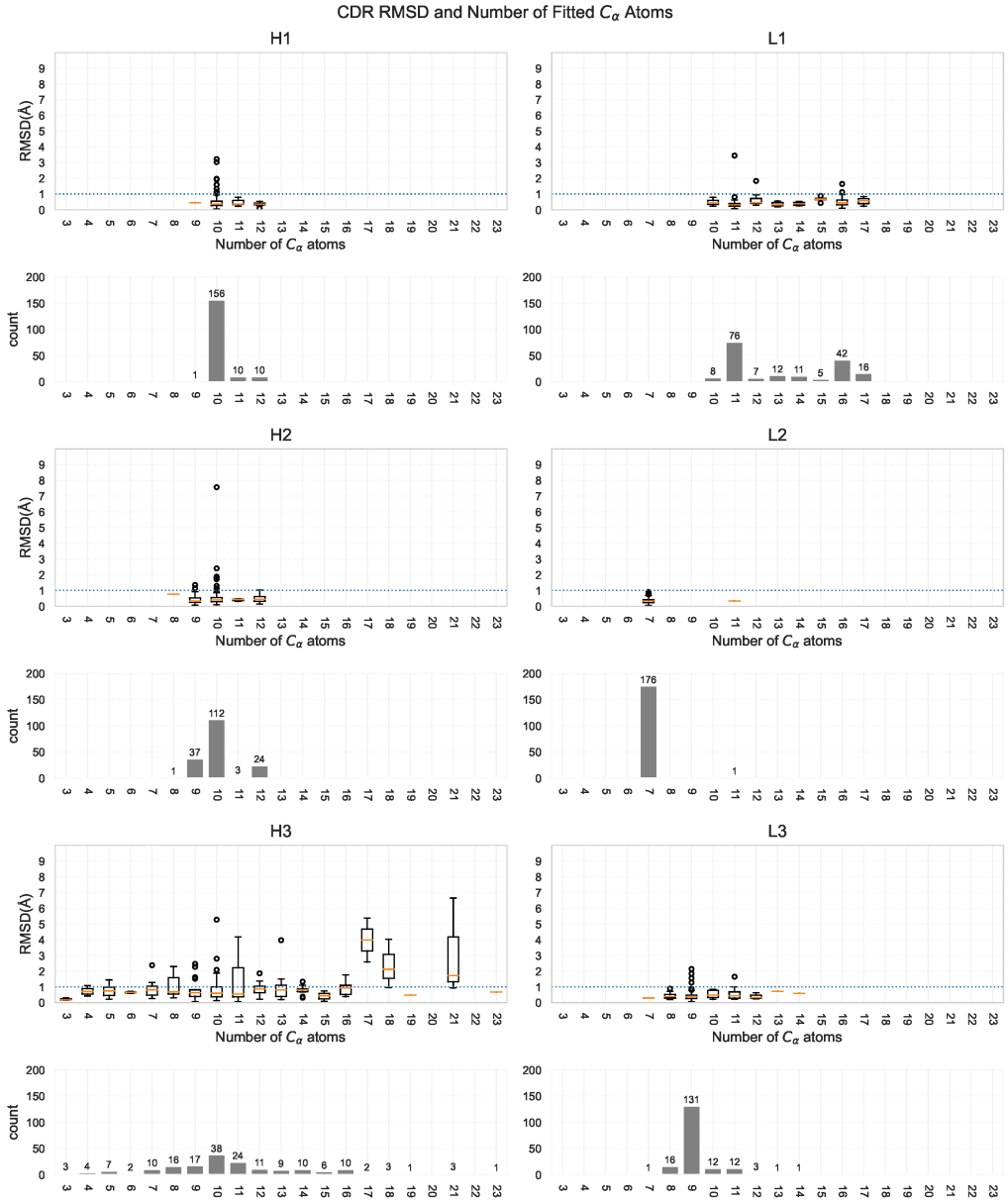

Figure 5: **CDR loop movement upon complexation against loop length**. The global Cα RMSD of each CDR upon binding (Cα RMSD) versus loop length (number of residues) are plotted as boxplots, with outliers (exceeding upper fence values $Q3+1.5\times IQR$) shown as circles. A horizontal dashed line is drawn at 1.0Å Cα RMSD on each box plot. The number of CDR loops of each loop length are also plotted as histograms.

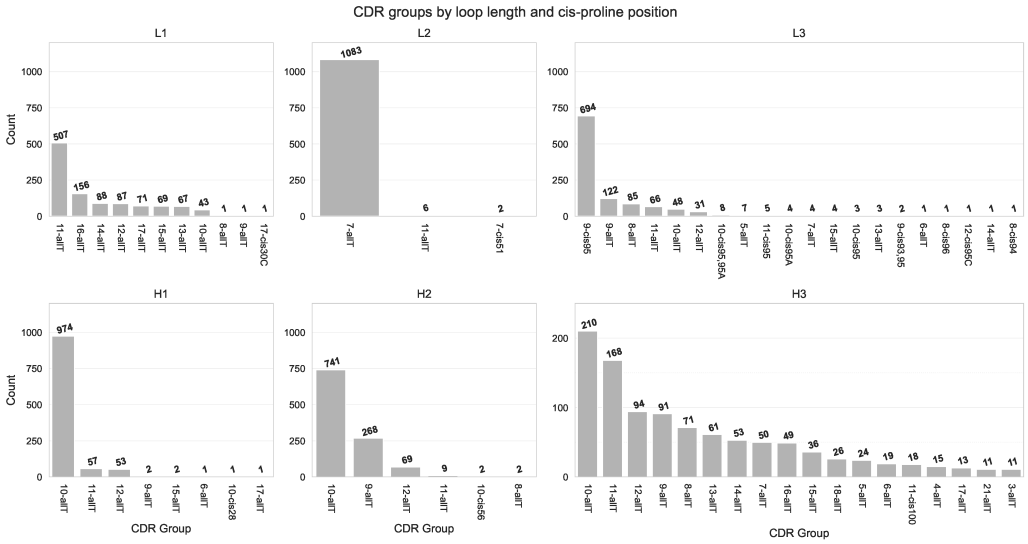

Figure 6: **LRC groups**. Each subplot shows the number of entries in each LRC group. For CDR-H3, only groups with more than 10 entries are shown.

entries. Similarly, CDR-L3 has a single dominant group ('L3-9-cis95', 64%) followed by 'L3-9-allT' (11%) and four smaller groups each representing up to 8%, with the rest being much less well populated.

The same descriptor was applied to CDR-H3 which consists of numerous small groups — the two most common LRC groups ('H3-10-allT' and 'H3-11-allT') account for 19% and 15% of entries respectively, while six groups each represents 5–9% of entries. The rest of the CDR-H3 LRC groups are much less common.

To derive a representation of the unbound CDR conformational space, we performed CDR torsional clustering within each LRC group to generate 'AP clusters'. As an example, Figure 7 shows the clustering results for the largest LRC groups of each CDR. Groups including 'L2-7-allT', 'H1-10-allT', 'L1-11-allT', 'L1-16-allT' and 'H2-9-allT' comprise a leading conformational cluster with a few smaller clusters. Groups including 'H2-10-allT' and 'L3-9-allT' are composed of two leading conformational clusters, and 'H2-10-allT' has an extra small cluster. Group 'L3-9-cis95' is dominated by a single conformational cluster. After torsional clustering, we performed Cartesian cluster merging to replicate the Chothia canonical clusters as described by

Martin and Thornton[7].

Although CDR-H3 does not follow the canonical class rules adopted by the other CDRs, we clustered the observed LRC groups for CDR-H3 in the same way, forming AP (torsional) clusters and then 'canonical' clusters by Cartesian cluster merging.

3.4 CDR conformational change types

The numbers of antibodies of each conformational change type (as described in Table 1) are summarized in Table 4. For all CDRs except CDR-H3, 98–100% of bound conformations are observed in unbound antibodies (Table 4 column 'Sum(NR)'). While some degree of conformational change at the torsional level is observed frequently (40–74% of the time; Table 4 column 'AP ClusterShift'), significant change is rare (1–3% change canonical cluster; 0–2% to a conformation not seen as part of a canonical cluster in unbound antibodies). In contrast, for CDR-H3, only 87% of bound conformations can be found in the unbound conformational space. While it is still the case that, in general, CDR-H3 does not significantly change conformation on binding (78% of cases do not change AP cluster); when this does occur, there is a higher chance of a larger change, either to a conformation seen in another antibody (12% 'Canonical cluster shift') or to a conformation not seen in other antibodies (11% 'Non-canonical conformation').

In addition, we plotted the density distribution of local Cα RMSD for antibodies of each conformational change type (Figure 8). Generally, the conformational change for CDRs of 'Identical-AP' and 'AP-cluster shift' conformational change type is insignificant (around 0.5Å), whereas those of 'Canonical cluster shift' and 'Non-canonical conformation' types are larger and more wide-ranging. Examples of unbound/bound pairs for each conformational change type are provided in Figure 9, and the loop 'flapping' effect is evident in Figure 9d and 9e where the local Cα RMSD is small and much lower than the global Cα RMSD.

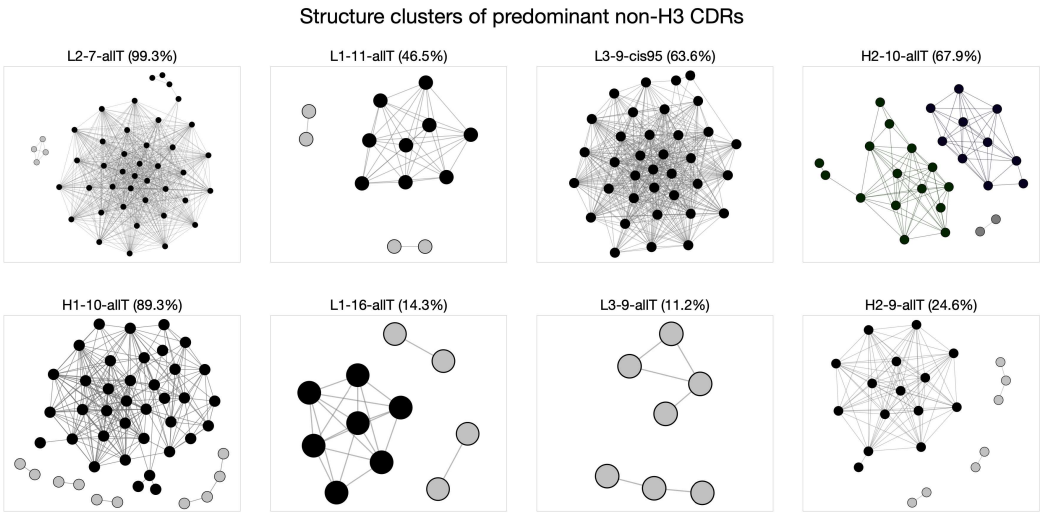

Figure 7: **Structure clusters of predominant non-CDR-H3 CDR LRC groups**. Subplot titles are CDR LRC group names, and the percentage given in parenthesis denotes the ratio of the group size (number of AbDb entries) to the entire set (1091 entries). Each node represents an AP cluster which consists of a set of similar CDR structures and from which a representative structure (also called an exemplar structure) was identified. The edges between pairs of nodes indicate the exemplar structures of both nodes are similar following our criteria under Cartesian space and thus belong to the same Canonical cluster. Nodes, directly or indirectly connected, are given the same colour. The major groups are coloured in black and smaller ones in grey. Note we use edges to indicate connectivity only, which means the distance between a pair of nodes is trivial in this case. The placing of nodes in the figure is purely illustrative.

Table 4: Counts of antibodies of each conformational change type

CDR	Unbound conformational space				
	Identical-AP	**AP-cluster shift**	**Canonical-cluster shift**	**Sum (NR)**	**Non-canonical conformation**
H1	127 (72%)	95 (54%)	3 (2%)	174 (98%)	3 (2%)
H2	123 (69%)	89 (50%)	3 (2%)	174 (98%)	4 (2%)
H3	138 (78%)	6 (3%)	21 (12%)	154 (87%)	19 (11%)
L1	144 (81%)	70 (40%)	5 (3%)	175 (99%)	3 (2%)
L2	87 (49%)	131 (74%)	1 (1%)	177 (100%)	0 (0%)
L3	114 (64%)	91 (51%)	2 (1%)	175 (99%)	3 (2%)

Because one antibody can have more than one unbound or/and bound entries, it can fall into multiple conformational change types and therefore the total number of cases from the four types can exceed the number of antibodies in the entire set (177 antibodies). **Sum (NR)** is the sum of non-redundant antibodies whose bound conformation can be found in the unbound conformational space ('Identical-AP', 'AP-cluster shift', 'Canonical-cluster shift').

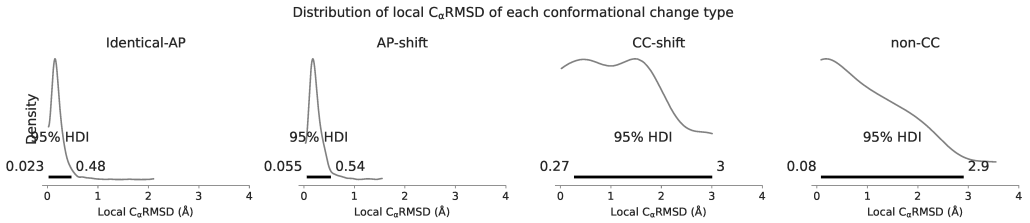

Figure 8: **Local Cα RMSD of antibodies of each conformational change type**. Each subplot is a Kernel Density Estimation of the local Cα RMSD of unbound/bound CDR conformation pairs found in each conformational change type. The number in each subplot parenthesis indicates the number of antibodies.

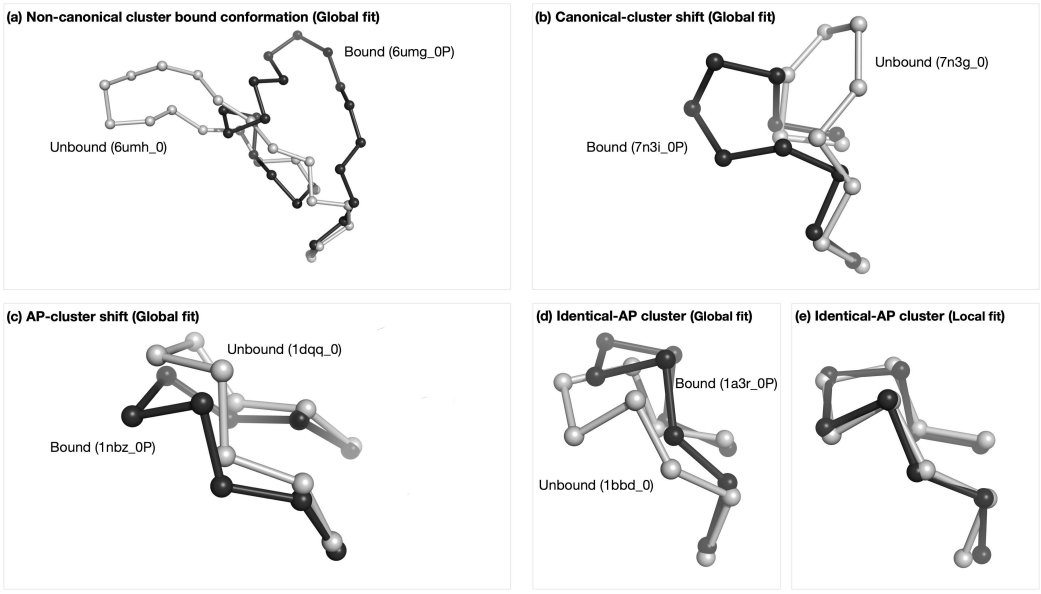

Figure 9: **Example of each conformational change type**. The figure shows one locally fitted CDR-H3 unbound/bound pair for each conformational change type. The bound CDR-H3 loop is coloured black and the unbound loop is grey. **(a) Non-canonical cluster conformation**: unbound (6umh_0) and bound (6umg_0P), global Cα RMSD of 8.41Å; **(b) Canonical-cluster shift**: unbound (7n3g_0) and bound (7n3i_0P), global Cα RMSD of 5.28Å; **(c) AP-cluster shift**: unbound (1kcv_0) and bound (1kcs_0P), global Cα RMSD of 2.38Å; **(d)** and **(e) Identical-AP cluster** — (d) is locally fitted and **(e)** is globally fitted to show the loop 'flapping' effect. **(d)** global Cα RMSD 2.27Å, **(e)** global Cα RMSD 0.54Å.

4 Discussion

There is a common belief that antibody CDRs are flexible and likely to change conformation on binding. However, from a thermodynamic perspective, given all other things being equal, a rigid lock-and-key interaction will result in optimal affinity with no loss of enthalpy or entropy. Consequently, in this work, we provide a survey of CDR conformational change upon binding by directly comparing the unbound and bound conformers of the same antibody.

We implemented a filtering pipeline to pool high-quality antibody structures from AbDb[14] and built a dataset (AbAgDb) which consists of 177 antibodies with bound and unbound structures. Currently, we limit the type of antibody to those with conventional Fvs containing both V_H and V_L domains. Compared with a previously published dataset for antibody-antigen cases[22], our dataset has expanded the number of antigen types and examples. We believe that maintaining this dataset is beneficial for the development of new computational tools for antibody-related tasks, such as epitope prediction and antibody-antigen complex prediction. As reviewed in a recent survey[2], one of the major challenges in developing computational tools for antibody development is data completeness.

We investigated the conformational changes of each CDR loop using global and local fitting while excluding changes resulting from differences in the packing of V_H and V_L domains. In summary, the local C_α conformation of the non-H3 CDRs changes by $\leq$1.0Å in 98.5% of cases and by $\leq$2.0Å in 99.5% of cases indicating that significant conformational changes are rare. In CDR-H3, these percentages drop to 83.6% ($\leq$1.0Å) and 94.9% ($\leq$2.0Å) indicating that conformational change is more common but still unusual. See Table 3.

However, when we look at the *global* Cα RMSD, we find that smaller percentages of all CDRs have Cα RMSD below either 1.0Å or 2.0Å indicating loop flapping. For the non-CDR-H3 loops, the global and local percentages are almost the same when looking at $\leq$2.0Å indicating only a minor flapping effect, but this is much more frequent in CDR-H3 (Table 3).

We went on to cluster unbound CDR conformations in backbone torsion

angles to create 'AP clusters' followed by Cartesian cluster merging to create 'Canonical clusters'. This approach was applied to all six CDRs. We then classified the conformational change on binding into four categories: 'identical AP cluster', 'AP-cluster shift', 'canonical-cluster shift', and 'non-canonical structure' as described in Table 1. In most cases, the bound conformation does not change, at least at the level of a canonical cluster. For the non-CDR-H3 loops, 1–3% change canonical cluster and 0–2% change to a conformation not observed in a canonical cluster in unbound antibodies. For CDR-H3, a change in conformation is more common and more pronounced, with 12% changing to a canonical cluster seen in other unbound antibodies, and 11% changing to a conformation not seen in other unbound antibodies (Table 4).

While these are rare, when we see canonical class shifts, they are all changes to conformations seen in a different antibody. The only exceptions are in three antibodies where three CDR-H3 loops change to a conformation seen in a different entry for the same antibody.

5 Conclusion

There is a common preconception that antibody CDRs are flexible and, in particular, that they are likely to change conformation on binding. In this work, we show that this is rarely the case.

We provide a survey of CDR movement, directly comparing the unbound and bound conformers of the same antibody, both by Cα RMSD and by conformational clustering. Based on our AbAgDb dataset of 177 high-quality antibody structures where both unbound and bound forms are available, we found that significant local conformational change on binding is rare. Only $\sim$1.5% show a local conformational change of $>$1.0Å (Cα RMSD) and $\sim$0.5% show a local conformational change of $>$2.0Å. Conformational change is somewhat more common in CDR-H3, but the vast majority of antibodies still undergo only minimal change in CDR-H3 ($\sim$16.4% show a local conformational change of $>$1.0Å while $\sim$5.1% show a local conformational change of $>$2.0Å).

We also observe a loop 'flapping' effect where there is minimal change in CDR conformation, but the loop 'flaps' about its junction with the framework. This was found always to be a minor effect in non-CDR-H3 loops, but is somewhat more common and larger in CDR-H3.

6 Availability

A snapshot of AbDb (version date: 20220926) as used to build AbAgDb is available at `http://www.abybank.org/abdb/snapshots/abdb_20220926.zip`

ProFit was used for protein structure fitting and can be obtained from `http://www.bioinf.org.uk/software/profit`.

The code for using the CDR conformation classifiers is available at `https://github.com/biochunan/CDRConformationClassification`.

Detailed information on the comparison of the clustering performed here, the Martin and Thornton (1996) clusters and the original canonical classes is provided in Supplementary Files 'ClusterComparison.pdf' and 'SI1-Clusters.csv'.

A CSV file containing the full list of entries indicating which entries remain after each filtering step is provided in Supplementary File 'SI2-unbound_abdbids_and_filtering_step.csv'.

The list of unbound and bound antibody structure pairs in AbAgDb is provided in Supplementary File 'SI3-unbound_and_bound_abs.csv'.

References

[1] Rahmad Akbar, Habib Bashour, Puneet Rawat, Philippe A. Robert, Eva Smorodina, Tudor-Stefan Cotet, Karine Flem-Karlsen, Robert Frank, Brij Bhushan Mehta, Mai Ha Vu, Talip Zengin, Jose Gutierrez-Marcos, Fridtjof Lund-Johansen, Jan Terje Andersen, and Victor Greiff. Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies. *mAbs*, 14(1):2008790, 2022. PMID: 35293269.

[2] Alissa M. Hummer, Brennan Abanades, and Charlotte M. Deane. Advances in computational structure-based antibody design. *Current Opinion in Structural Biology*, 74:102379, 2022.

[3] Andrew C.R. Martin and James Allen. *Bioinformatics Tools for Antibody Engineering*, chapter 5, pages 95–117. John Wiley & Sons, Ltd, 2007.

[4] Sally Roberts, Janet C. Cheetham, and Anthony R. Rees. Generation of an antibody with enhanced affinity and specificity for its antigen by protein engineering. *Nature*, 328:731–734, 1987.

[5] Cyrus Chothia and Arthur M. Lesk. Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology*, 196(4):901–917, 1987.

[6] Bissan Al-Lazikani, Arthur M. Lesk, and Cyrus Chothia. Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology*, 273(4):927–948, 1997.

[7] Andrew C.R. Martin and Janet M. Thornton. Structural families in loops of homologous proteins: Automatic classification, modelling and application to antibodies. *Journal of Molecular Biology*, 263(5):800–815, 1996.

[8] Benjamin North, Andreas Lehmann, and Roland L. Dunbrack. A new clustering of antibody CDR loop conformations. *Journal of Molecular Biology*, 406(2):228–256, 2011.

[9] Peter Csermely, Robin Palotai, and Ruth Nussinov. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences*, 35(10):539–546, 2010.

[10] Daniel E. Koshland Jr. Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences*, 44(2):98–104, 1958.

[11] Jefferson Foote and Cesar Milstein. Conformational isomerism and the diversity of antibodies. *Proceedings of the National Academy of Sciences*, 91(22):10370–10374, 1994.

[12] Gary J. Wedemayer, Phillip A. Patten, Leo H. Wang, Peter G. Schultz, and Raymond C. Stevens. Structural insights into the evolution of an antibody combining site. *Science*, 276(5319):1665–1669, 1997.

[13] Leo C. James, Pietro Roversi, and Dan S. Tawfik. Antibody multispecificity mediated by conformational diversity. *Science*, 299(5611):1362–1367, 2003.

[14] Saba Ferdous and Andrew C.R. Martin. AbDb: antibody structure database—a database of PDB-derived antibody structures. *Database*, 2018, 04 2018. bay040.

[15] Weizhong Li and Adam Godzik. CD-Hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, 05 2006.

[16] Abhinandan K. Raghavan and Andrew C.R. Martin. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. *Molecular Immunology*, 45:3832–3839, 2008.

[17] Abhinandan K. Raghavan and Andrew C.R. Martin. Analysis and prediction of VH/VL packing in antibodies. *Protein Engineering Design and Selection*, 23:689–697, 2010.

[18] Andrew C.R. Martin, Janet C. Cheetham, and Anthony R. Rees. Molecular modeling of antibody combining sites. *Methods Enzymology*, 203:121–153, 1991.

[19] Andrew C.R. Martin, Janet C. Cheetham, and Anthony R. Rees. Modeling antibody hypervariable loops: a combined algorithm. *Proceedings of the National Academy of Sciences*, 86(23):9268–9272, 1989.

[20] Andrew D. McLachlan. Rapid comparison of protein structures. *Acta Crystallographer*, A38:871–873, 1982.

[21] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, 2007.

[22] Johnathan D. Guest, Thom Vreven, Jing Zhou, Iain Moal, Jeliazko R. Jeliazkov, Jeffrey J. Gray, Zhiping Weng, and Brian G. Pierce. An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. *Structure*, 29(6):606–621.e5, 2021.