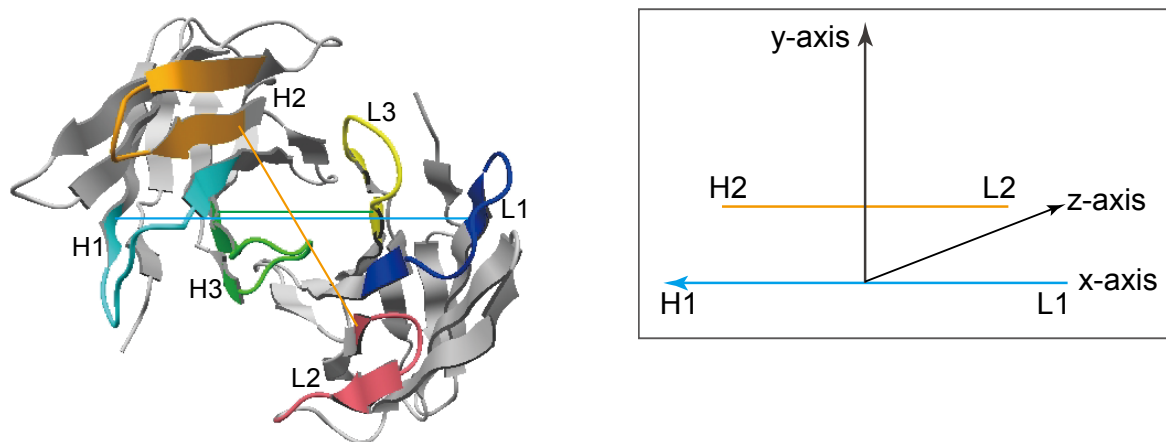


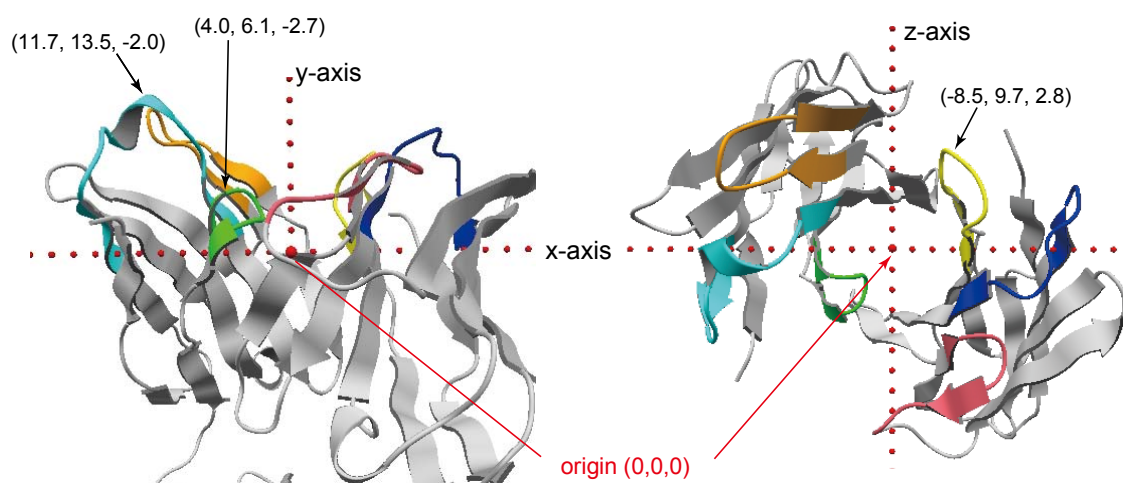
Supplemental Data

Figure S1 A new coordinate system



S1A) The determination of three axes in a new coordinate system.

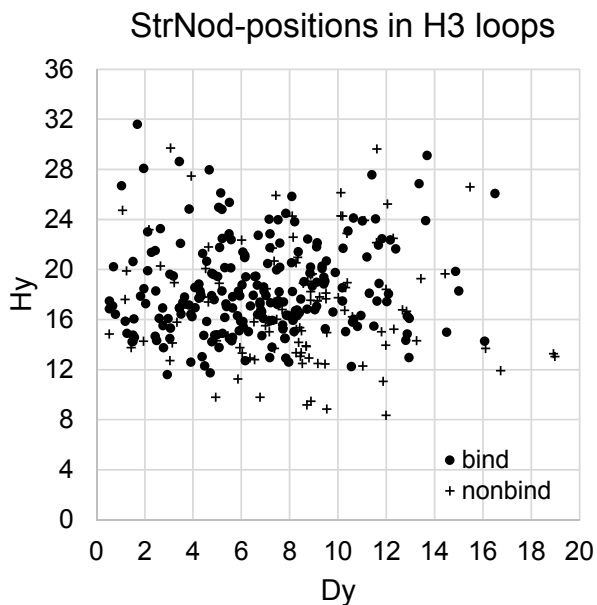
The line connecting between $C\alpha$ atoms in the residues of position 1 in H1 and L1 loops (H1-L1 line colored cyan), that between H2 and L2 loops (H2-L2 line colored orange), and that between H3 and L3 loops (H3-L3 line colored green) are considered, as shown in the left figure. H1-L1 line is almost overlapped with H3-L3 line. H2-L2 line is located about 6 Å above H1-L1 line, tilting around 60 degrees. H1-L1 line is defined as the x-axis in the new coordinate system. The line perpendicular to H1-L1 line and through H2-L2 line is defined as the y-axis, which lies on the shortest line between H1-L1 and H2-L2 lines. The exterior product between the x- and y-axes is used as the z-axis. The CDR loops are colored cyan, orange, green, blue, pink and yellow for H1, H2, H3, L1, L2 and L3, respectively. All the figures were drawn by the interactive molecular viewer, jV¹.



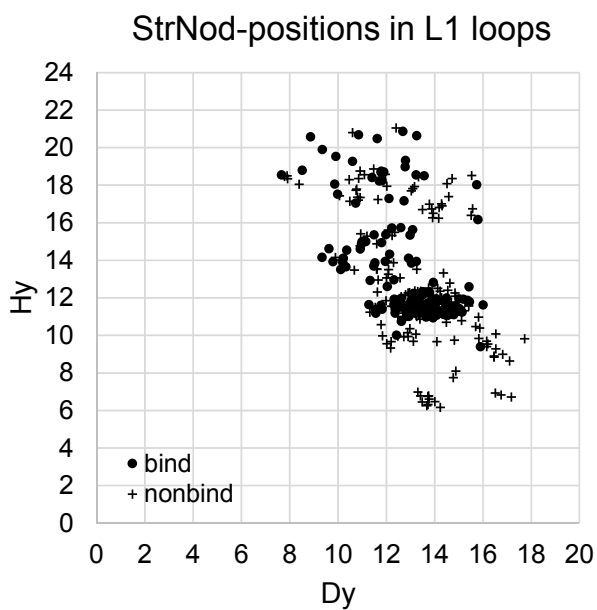
S1B) The standard view of an antibody (PDBID 2vxt) in the new coordinate system.

The origin of the new coordinate system and the coordinates of the highest $C\alpha$ atoms in H1 (cyan), H3 (green) and L3 (yellow) loops are shown. The CDR loops are colored in the same manner as in S1A.

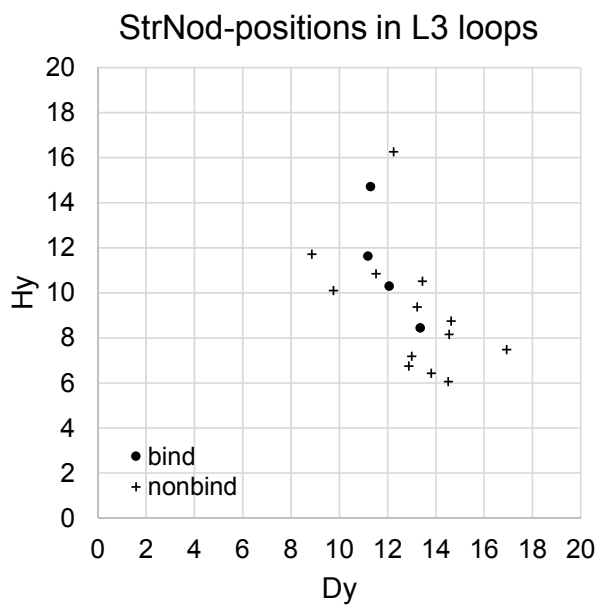
Figure S2 Antigen-binding properties in StrNod-positions (Related to Figure 1)



S2A)



S2B)

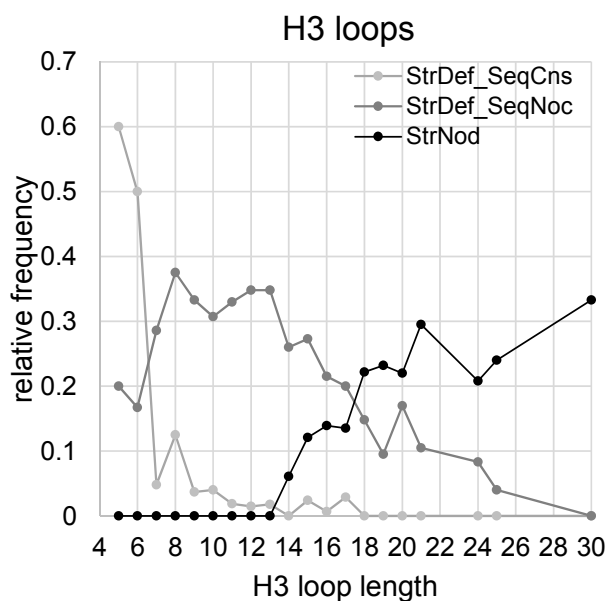


S2C)

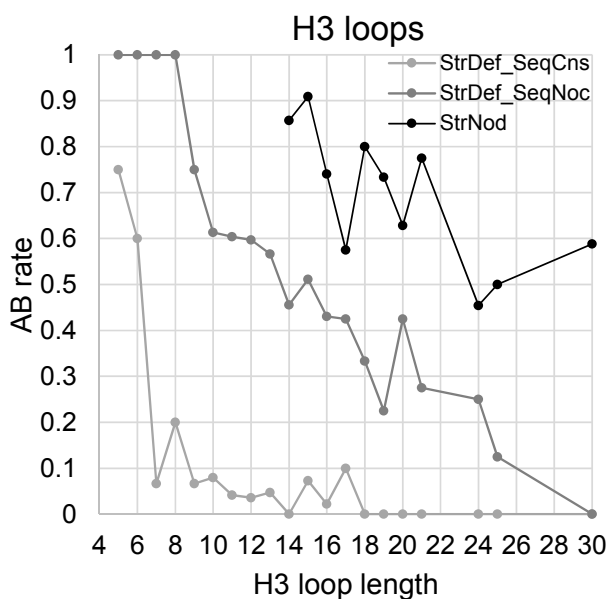
S2) The relationship between Hy and Dy values of the residues belonging to StrNod-positions.

The relationship in H3, L1 or L3 loop is shown in S2A, S2B or S2C, respectively. Filled circles or plus signs indicate antigen-binding or non antigen-binding residues, respectively.

Figure S3 The effects of H3 loop lengths on antigen binding (Related to Figure 2)



S3A)



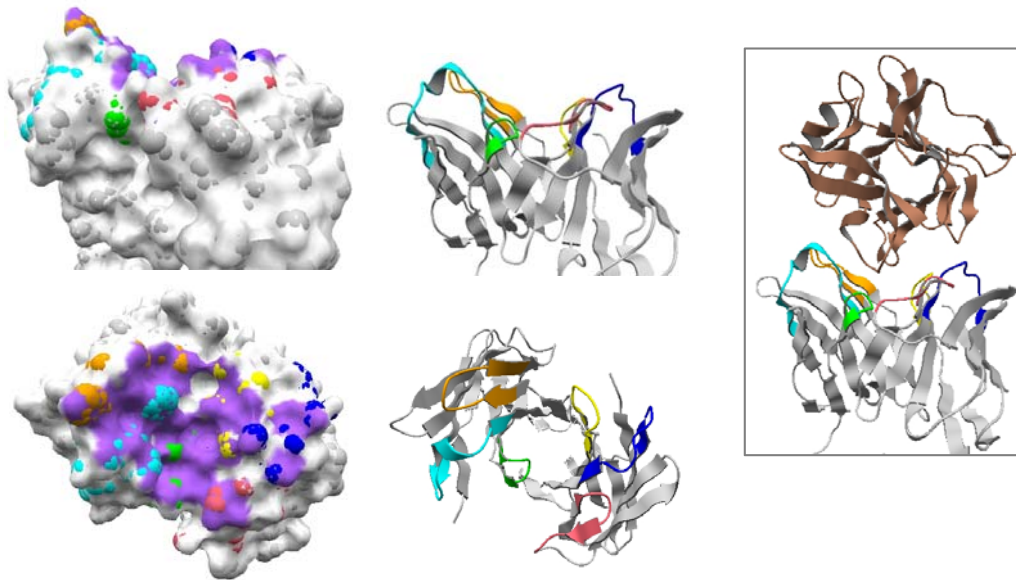
S3B)

S3A) The fraction of antigen-binding residues as a function of the H3 loop length.

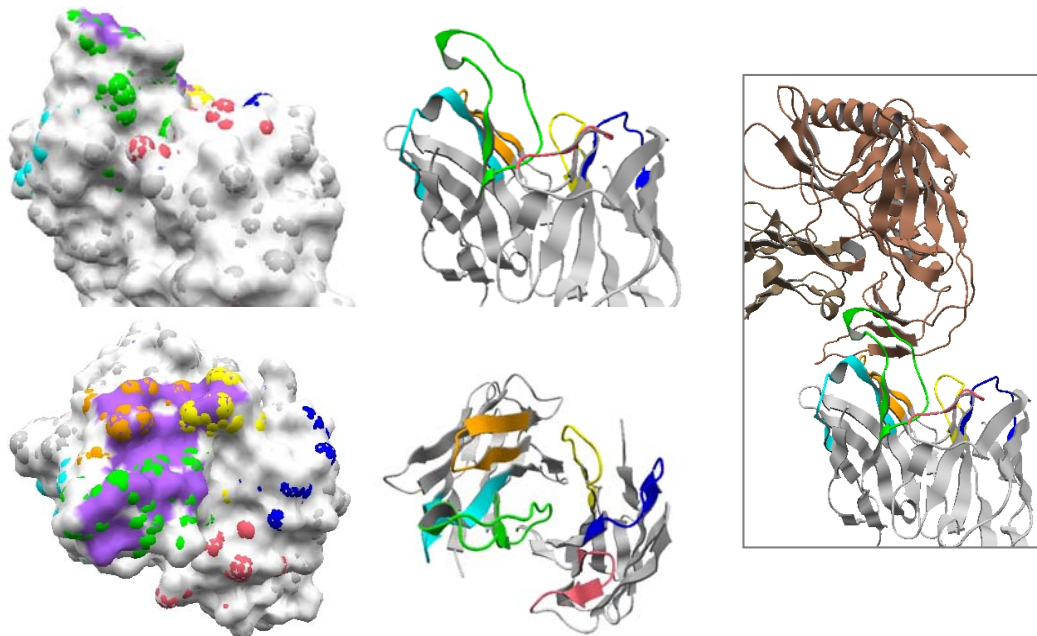
The fraction in StrDef_SeqCns-, StrDef_SeqNoc- or StrNod-position is shown by light-grey, grey or black line, respectively. In short H3 loops, StrDef_SeqNoc-positions mainly contribute to antigen binding, while in long H3, StrNod-positions are main contributors to antigen binding. In addition, in short H3 loops (five and six residue lengths), StrDef_SeqCns-positions are also involved in antigen binding.

S3B) The fraction of AB rates as a function of the H3 loop length.

The fraction in StrDef_SeqCns-, StrDef_SeqNoc- or StrNod-position is shown by light-grey, grey or black line, respectively. The antigen-binding (AB) rates in StrDef_SeqNoc- and StrNod-positions decrease with the increase of the H3 loop length, where the number of StrDef_SeqNoc-positions remains constant, while that of StrNod-positions increases. It indicates that StrDef_SeqNoc-positions in long H3 loops are no longer main contributors to antigen binding.



1) 2vxt



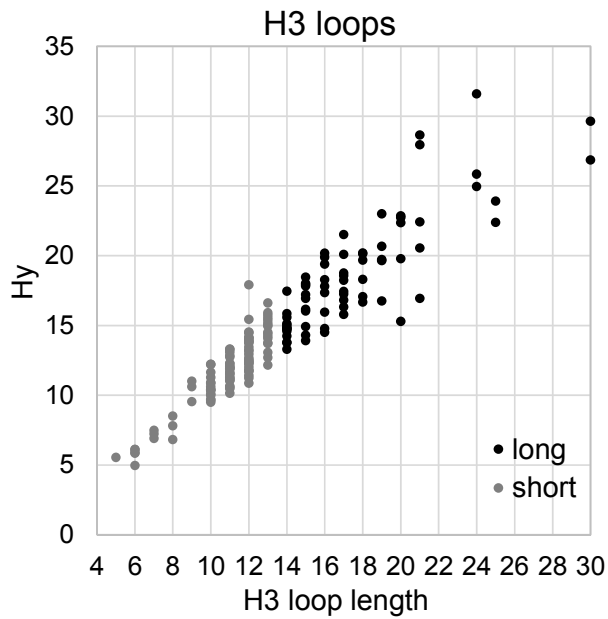
2) 1g9m

S3C) The concave and convex shapes of the CDR surfaces.

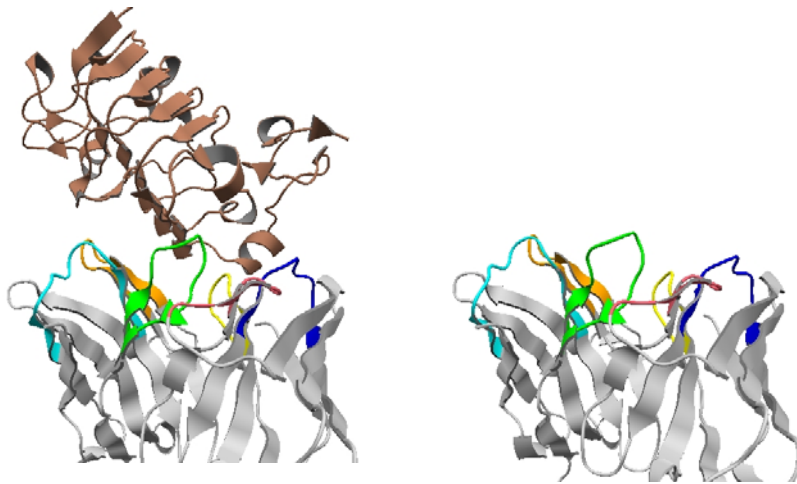
The molecular surfaces and structures of antibody variable fragments in 2vxt and 1g9m are shown. The right figures show the antibody-antigen complex structures. The molecular surfaces were calculated by the program MSP², and all the figures were drawn by the interactive molecular viewer, jV¹. The antigen-binding regions on the molecular surfaces are colored purple, and the CDR loops are colored in the same manner as in S1A. The antigen molecules are colored brown.

In 2vxt whose H3 loop consists of six residues, the center of CDR is concave because of the short H3 loop, therefore, the antigen can bind with all the CDR loops (the contributing rate of H3 to antigen binding: 0.19). On the other hand, in 1g9m whose H3 loop consists of 21 residues, the center of CDR is convex due to the long H3 loop. Therefore, the antigen mainly binds to H3 loop, of which contributing rate to antigen binding is 0.54.

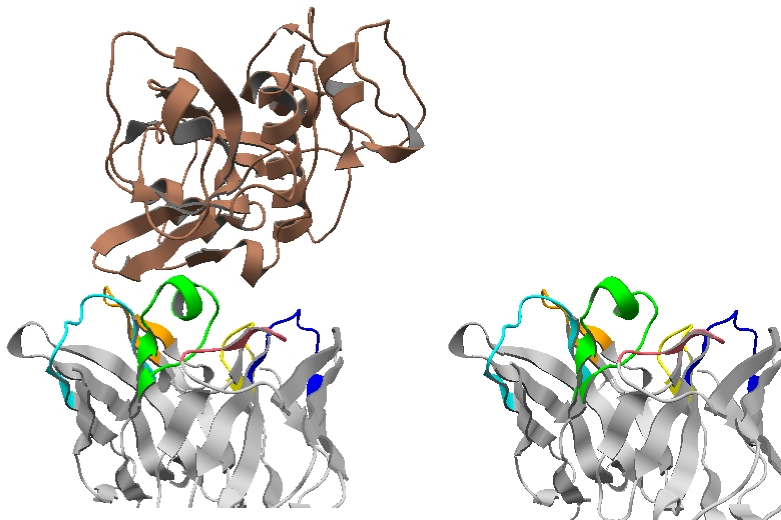
Figure S4 Diverse conformations of long H3 loops and their effects on antigen binding (Related to Figure 3)



S4A) The distribution of the largest Hy value in each H3 loop, as a function of the H3 loop length. Filled circles colored grey or black show the data of short (the length shorter than 14 residues) or long (longer than or equal to 14 residues) H3 loops, respectively.



1) 3p0y (bound) – 3p0v (unbound)

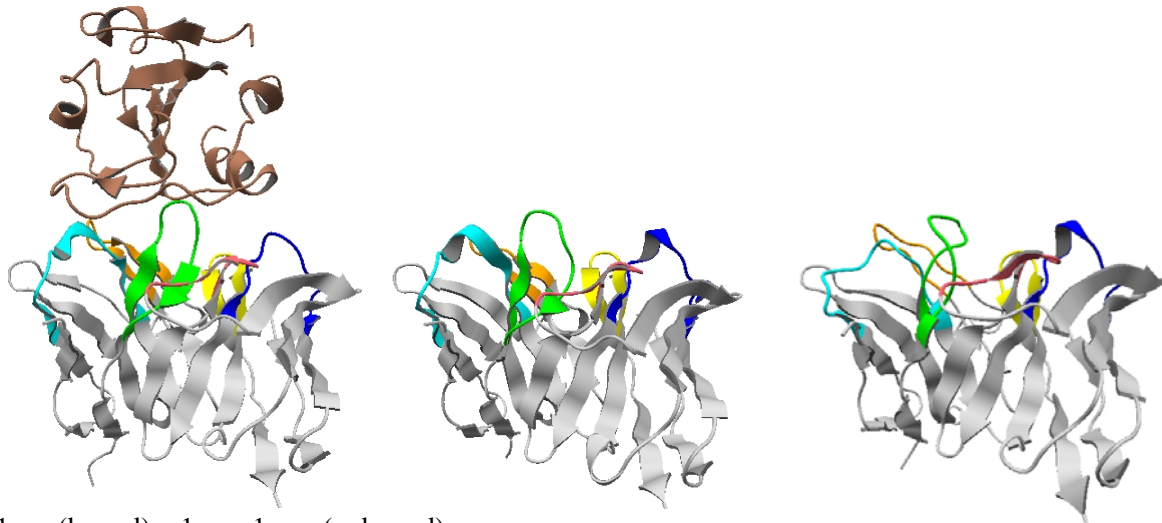


2) 4pp2 (bound) – 4poz (unbound)

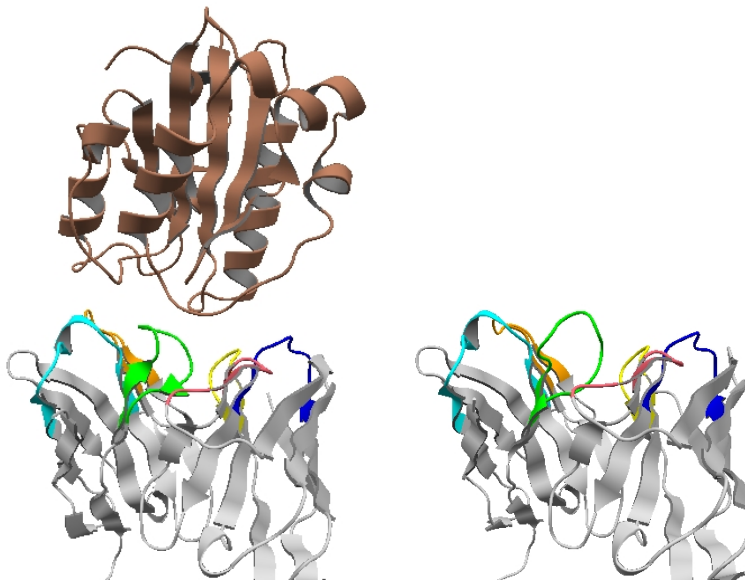
S4B) The similar or distinct H3 loop conformations in antigen-bound and unbound antibodies.

The CDR loops are colored in the same manner as in S1A. The antigen molecules are colored brown.

1, 2) The structures of two pairs of antigen-bound and unbound antibodies, the H3 loops of which have similar conformations. The upper figures show the pair of antigen-bound (3p0y) and unbound (3p0v) antibodies, where the H3 loops consist of 14 residues and have broad conformations. On the other hand, the lower figures show the pair of antigen-bound (4pp2) and unbound (4poz) antibodies, of which the H3 loops consist of 16 residues and form helix conformations.



3) 1oaz (bound) – 1oag, 1ocw (unbound)

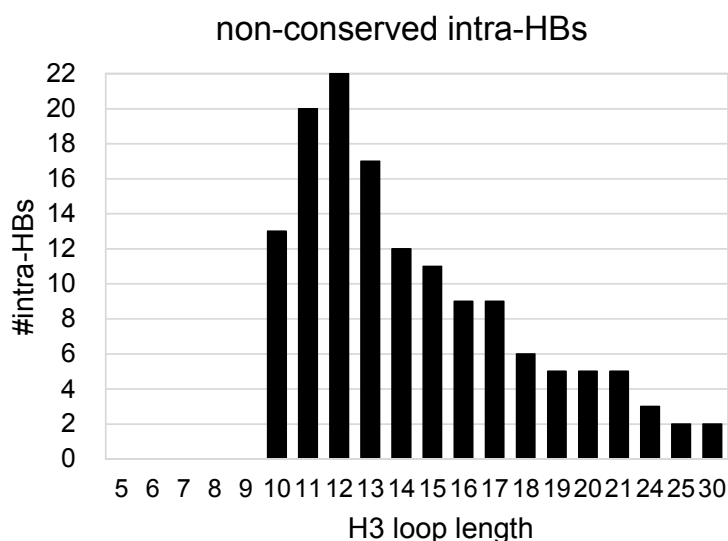


4) 3hi6 (bound) - 3hi5 (unbound)

3, 4) The structures of two pairs of antigen-bound and unbound antibodies, the H3 loops of which consist of 13 residues and have distinct conformations. The upper figures show the pair of antigen-bound (1oaz) and unbound (1oag, 1ocw) antibody structures, where the H3 loops have straight, broad and twist conformations, respectively. These structures have been determined by the same authors ³. The authors have determined two and four antigen unbound and bound structures of this antibody and shown that the two unbound structures have different conformations of H3 and L3 loops (1oag and 1ocw), and three of the four bound structures have the similar conformations to that in one of the unbound structures. The remaining bound structure (1oaz) has the different conformation of H3 and L3 loops. Thus, the structural analysis revealed the “preexisting conformational diversity” of this protein.

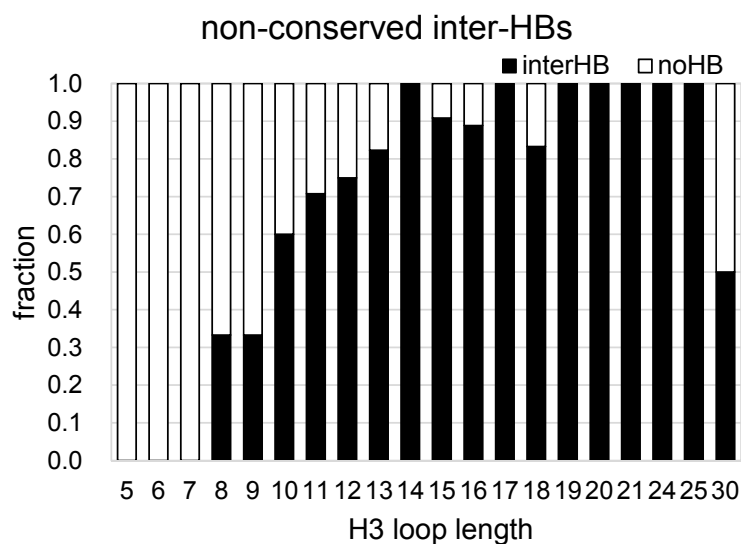
On the other hand, the lower figures show the pair of antigen-bound (3hi6) and unbound (3hi5) antibody structures, of which the H3 loops form bent and broad conformations, respectively. In the bound

structure, a divalent ion (Mn) is located in a metal ion-dependent adhesion site of the antigen, and it binds to D101 in H3 loop, which may lead to the conformational change in the H3 loop⁴.



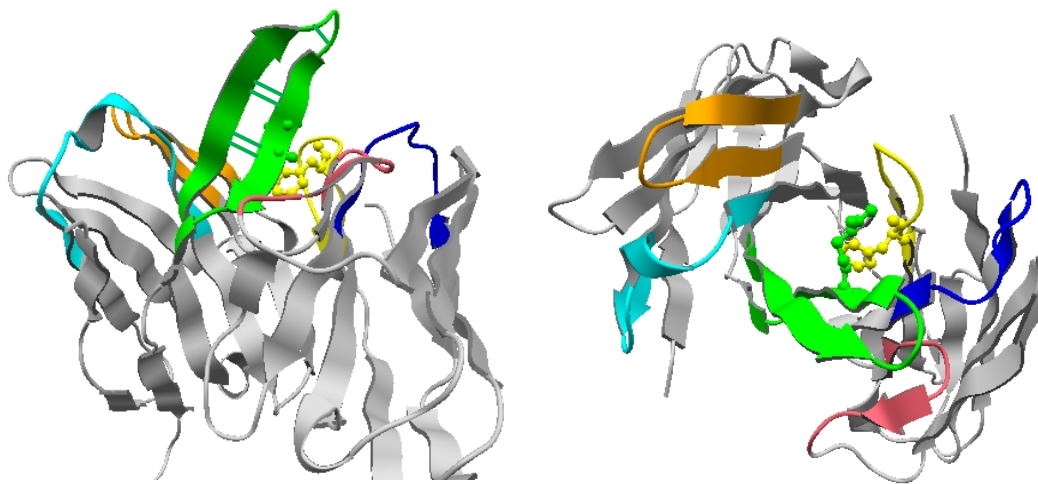
S4C) The average number of non-conserved intra-HBs observed in non-stem regions of H3 loops, as a function of the H3 loop length.

Of 99 short H3 loops (<14 residues), 27 loops do not have non-conserved intra-HBs, while only three loops have no intra-HBs in 72 long H3 loops (>=14 residues).



S4D) The fractions of antibodies with (inter HB, black bar) and without (no HB, white bar) non-conserved inter-CDR loop HBs observed in non-stem regions of H3 loops.

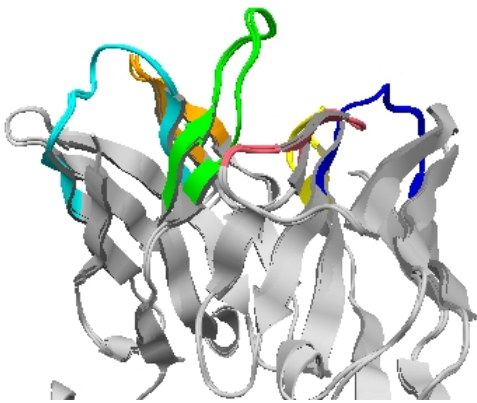
Of 99 short H3 loops (<14 residues), 36 loops do not have non-conserved inter-loop HBs, while only four loops have no inter-loop HBs in 72 long H3 loops (>=14 residues).



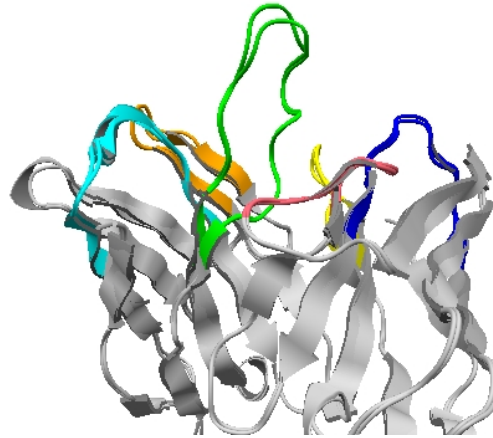
S4E) The straight H3 loop stabilized by regular ladder-like intra-loop HBs.

The structure of an antibody whose H3 loop has a straight conformation is shown (PDBID 3vg9). The ladder-like intra-loop HBs between main-chain atoms are indicated by green lines. The residues involved in non-conserved inter-loop HBs are shown in ball and stick model. The CDR loops are colored in the same manner as in S1A.

In 3vg9, the non-stem region of H3 loop forms intra-loop HBs between main-chain atoms in three pairs of positions, and an inter-loop HB with L3 loop. The ladder-like HBs stabilize the long H3 loop with a straight conformation, while the inter-loop HB contributes to stabilize the H3 loop against the rest of CDR loops.



1) 3rvv (bound) – 3rvu (unbound)



2) 3ru8 (bound) - 1hzh (unbound)

S4F) The identical inter-loop HBs in antigen-bound and unbound antibodies.

The structures of antigen-bound and unbound antibodies are superimposed by CCP4^{5,6}, where the variable fragments except for CDR loops were overlapped. The CDR loops are colored in the same manner as in S1A.

The left figure shows the pair of antigen-bound (3rvv) and unbound (3rvu) antibody structures, where the H3 loops that consist of 14 residues with broad conformations, form non-stem inter-loop HBs with L2 and L3 at the same position. The right figure shows antigen-bound (3ru8) and unbound (1hzh) antibody structures, of which the H3 loops consist of 20 residues having broad conformations, form non-stem inter-loop HBs with L1 and L3 at the same positions, respectively.

Table S1 Discrimination of probable antigen-binding residues in CDR loops

Definition 1) StrDef_SeqCns-positions: non antigen-binding, 2) StrDef_SeqNoc-positions: antigen-binding, 3) StrNod-positions in H3 and L1: antigen-binding, those in L3: non antigen-binding.

	TP ^{a)}	FP ^{b)}	FN ^{c)}	TN ^{d)}	Total	ACC ^{e)}	MCC ^{f)}	F measure ^{g)}
H1	218	149	217	1,663	2,247	0.84	0.45	0.54
H2	635	535	77	436	1,683	0.64	0.37	0.68
H3	800	689	53	801	2,343	0.68	0.48	0.68
L1	255	403	115	1,250	2,023	0.74	0.37	0.50
L2	132	381	107	748	1,368	0.64	0.17	0.35
L3	464	381	28	666	1,539	0.73	0.54	0.69
Total	2,504	2,538	597	5,564	11,203	0.72	0.44	0.62

a) The number of antigen-binding residues judged as antigen-binding residues correctly.

b) The number of non antigen-binding residues judged as antigen-binding residues incorrectly.

c) The number of antigen-binding residues judged as non antigen-binding residues incorrectly.

d) The number of non antigen-binding residues judged as non antigen-binding residues correctly.

e) The accuracy of the discrimination. $ACC = (TP+TN)/(TP+FP+FN+TN)$

f) The Matthews correlation coefficient.

$$MCC = (TP \times TN - FP \times FN) / \sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))}$$

g) The harmonic mean of precision and recall.

$$F \text{ measure} = 2(\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}), \text{ where } \text{precision} = TP / (TP+FP) \text{ and } \text{recall} = TP / (TP+FN) .$$

Table S2 The numbers of hydrogen bonds between main-chain atoms in positions 4 and N-3 in H3 loops, and the H3 loop conformations (Related to Figure 3)

Observation ^{a)}	Antibody ^{b)}	main-chain HBs in positions 4 and N-3		
		#0 ^{c)}	#1 ^{d)}	#2 ^{e)}
short (length<9) ^{f)}	8	8	0	0
short (length>=9) ^{f)}	83	38	36	9
short (extend) ^{f, g)}	8	8	0	0
broad	22	17	5	0
bent	25	6	19	0
straight	12	0	0	12
helix ^{h)}	5	3	2	0
twisted ^{h)}	4	2	2	0
extend ^{g)}	4	4	0	0
total	171	86	64	21

a) The observed H3 loop conformations.

b) The number of antibodies whose H3 loops are classified into a given class of loop conformations.

c, d, e) The numbers of antibodies whose H3 loops form no, one, and two hydrogen bonds between main-chain atoms in positions 4 and N-3, respectively.

f) The short H3 loops (the length < 14 residues) are divided into three classes, H3 loops with the length shorter than nine residues, those with longer than or equal to nine residues, and those with an extended stem conformation. This is because the formation of the hydrogen bonds between positions 4 and N-3 requires at least nine residues.

g) The “extend” H3 loops have an extended stem conformation and never form the hydrogen bonds between positions 4 and N-3. We did not mention them in the main article.

h) The “helix” and “twisted” conformations are minor H3 loop conformations. Therefore, we did not mention them in the main article.

Table S3 The summary of the key observations for antigen recognition by antibody

	Key observations
1	Structurally definable, sequence conserved (StrDef_SeqCns) positions are infrequently used in antigen binding.
2	The antigen binding typically takes place in the “very high” or “high and centrally located” structurally definable (StrDef) positions, most of which are structurally definable, sequence non-conserved (StrDef_SeqNoc) positions in the heavy chains.
3	The H3 loops utilize different positions for antigen binding depending on the loop length and a cut-off of 14 residues differentiates the short and long H3 loops..
4	The H3 loop lengths affect the contributions of the other CDR loops to antigen binding.
5	The long H3 loops prefer to form non-straight (non- β -ladder) conformations.
6	HB breaks between main-chain atoms in positions 4 and N-3 in H3 loops are an important factor to prevent an extension of the β -ladder, and the numbers of these hydrogen bonds highly correlate with the observed conformations of long H3 loops (straight, bent and broad).
7	The long H3 loops tend to form non-conserved intra- and inter-loop HBs in non-stem regions.
8	The non-conserved inter-loop HBs in long H3 loops are located close to the residues involved in antibody-antigen HBs.