Supplementary Information for

Predicting Antibody Complementarity Determining Region Structures without Classification

Yoonjoo Choi and Charlotte M. Deane

Department of Statistics
Oxford University
1 South Parks Road
Oxford, OX1 3TG
UK

Fig. S1 The CDR-H3 loop in 1NSN has sequence TRGNGD. The asparagine residue 115 does not make any contact ("0"). The C_α atom in the glycine 116 is in an external contact ("1") with leucine 52 in the light chain. The aspartic acid residue 117 makes an internal contact ("2") with valine 2 in the heavy chain. The residue numbers here are assigned according to the IMGT numbering.
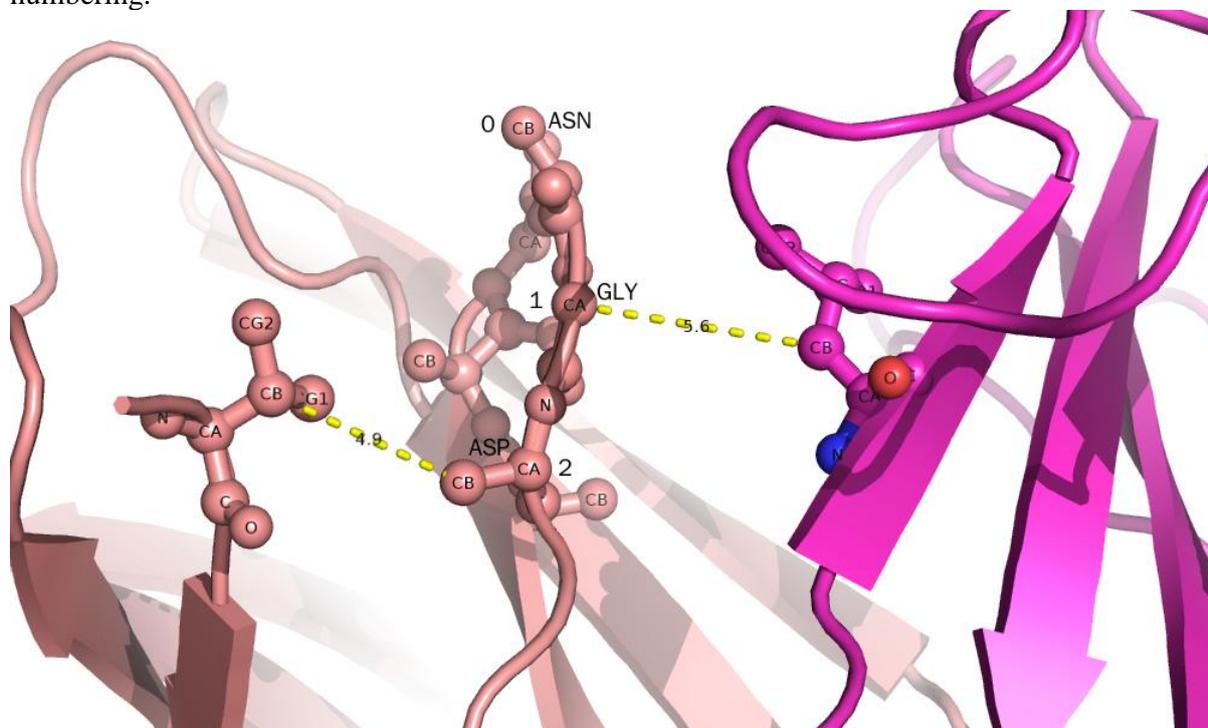
Fig. S2 A schematic view of idealised docking and CDR prediction on the Bound-Free set. The antigen from the native antigen-bound structure is docked using ZDOCK to the antigen-free structure. Next the CDRs are predicted on the antigen-free framework in the presence of this approximately docked antigen. These predicted CDRs are then compared to those found in the native antigen bound structure.
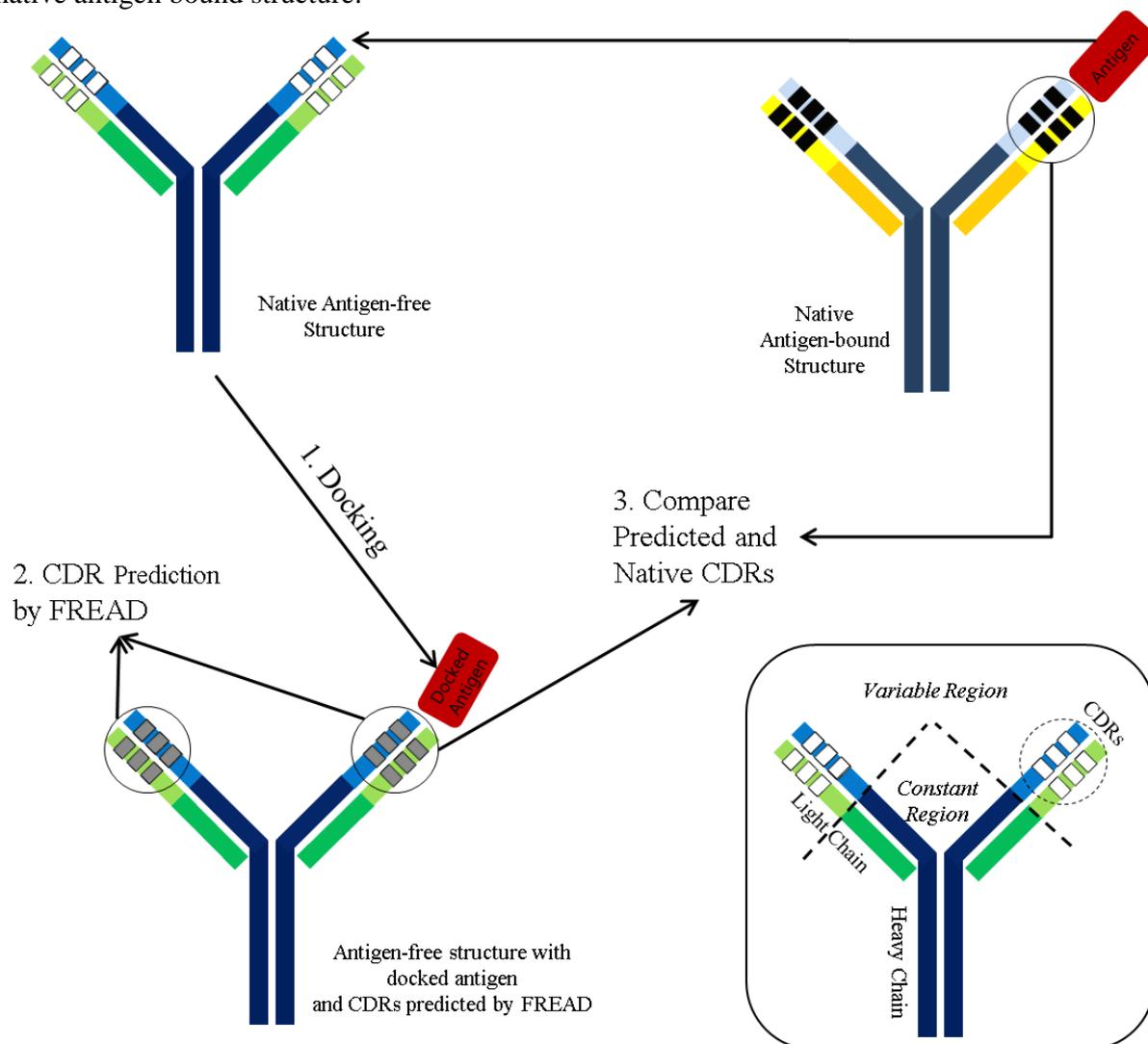
Fig. S3 Structural difference and its relationship to contact profiles. Chain B (purple) and chain H (cyan) in 1XF2 share 100% sequence identity (VRGGYRPYYAMDY for CDR-H3), but different structural conformations due to the antigen binding. Structural distortions occur on the residue pairs in different contacts (2222100111212 and 2222012111212 for chain B and H respectively).
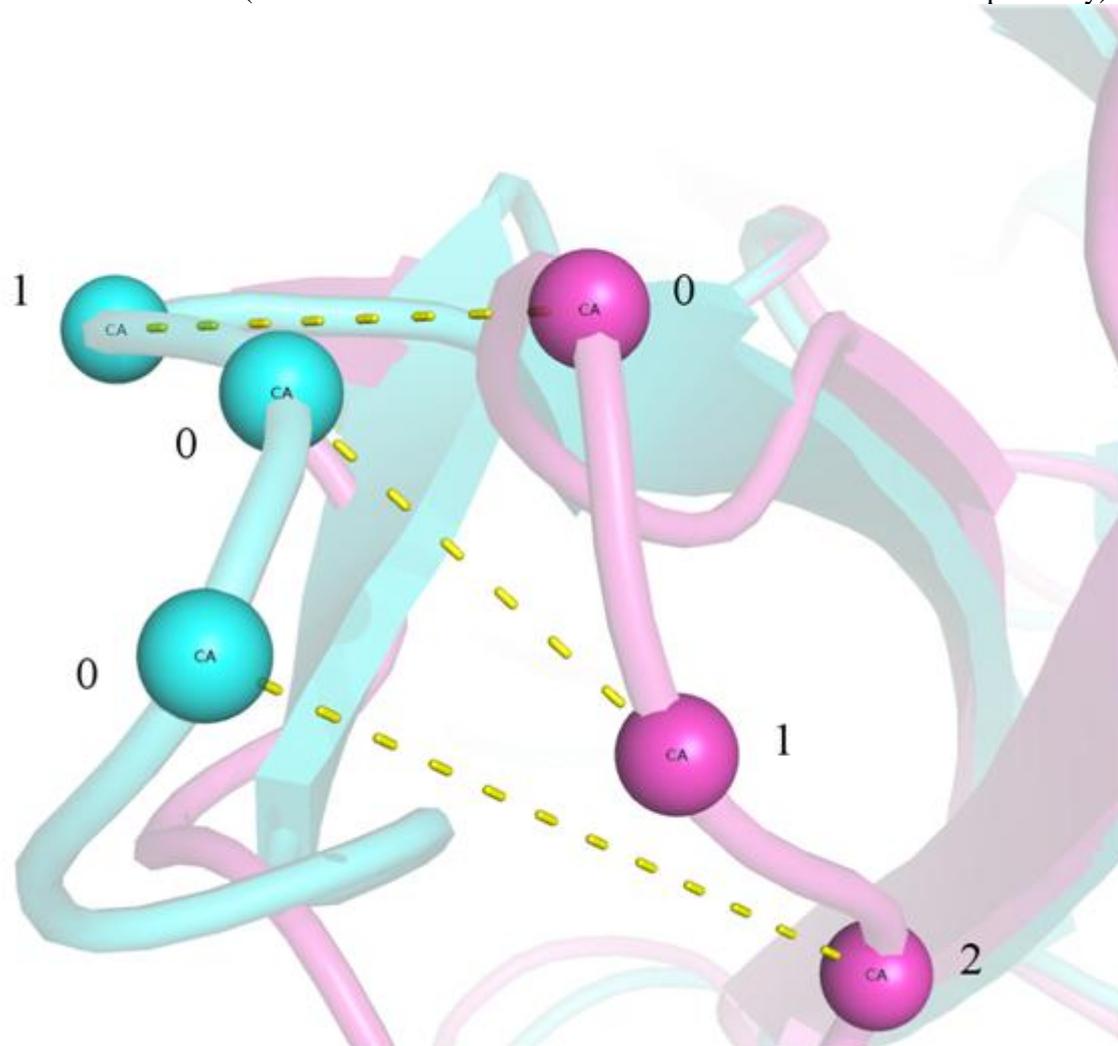
Fig. S4 The results of the FREAD variants on the Bound-Free set. The numbers on each bar are standard deviations (Full details of these results are given in Table S9). "Best" refers to the average RMSD of the lowest RMSD structures FREAD produced.
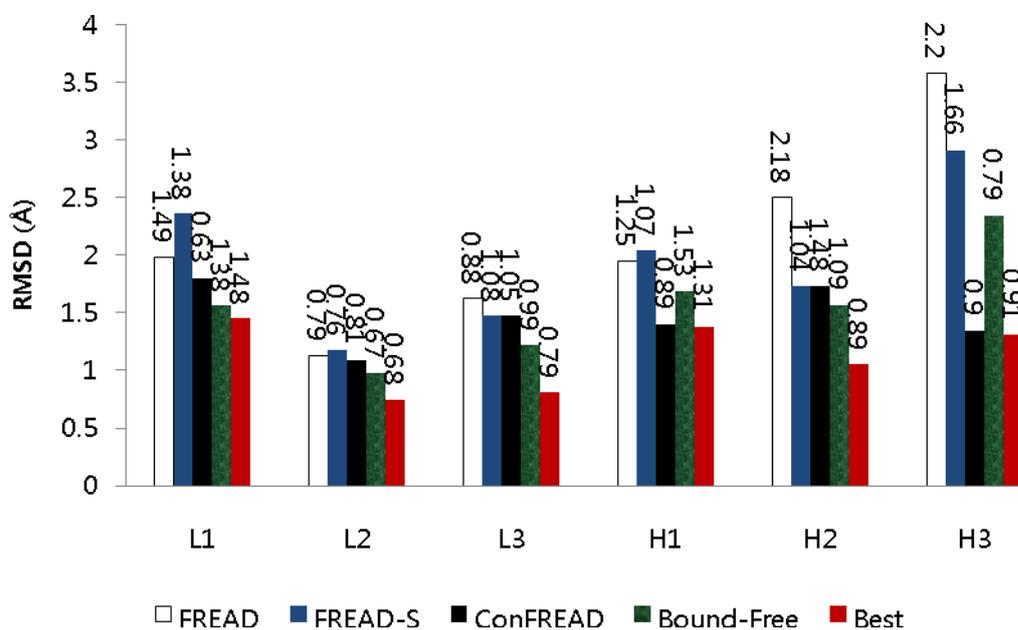
Fig. S5 Contact profile composition of CDRs in the native set. The y-axis gives the percentage of residues involved in the four 'contact' types (from top to bottom: no contact, internal contact only, external contact only, and both external and internal contact). The first bars are sums of antigen bound structures and unbound structures. The second bars are contact type frequencies of antibody structures where antigens are not present. The third bars are from antigen bound structures.

Table S1 A full list of the Native Set. The first column is the PDB code, the second and the third columns are the chain identifiers of the light and heavy chains.

| \<The Native Set with antigens\>											
Code	V_L	V_H	Code	V_L	V_H	Code	V_L	V_H	Code	V_L	V_H
1NDM	A	B	3DSF	L	H	1EAP	A	B	1MEX	L	H
1RIU	L	H	1I9R	Y	K	2MCP	L	H	1V7N	N	J
1AI1	L	H	2Z92	B	A	1YUH	A	H	1F3D	J	K
1WEJ	L	H	1A4K	A	H	2NYY	C	D	1RZJ	L	H
1OB1	D	E	1NC4	A	D	1CFS	A	B	1IND	L	H
1MPA	L	H	1T2Q	L	H	2CMR	L	H	1FIG	L	H
3CFB	L	B	2IFF	L	H	1CE1	L	H	1TZG	M	H
1C5C	L	H	1IFH	L	H	1NAK	L	I	1GGI	M	J
1SM3	L	H	2R56	L	H	35C8	L	H	2HMI	C	D
2W65	D	A	1E4X	L	H	2IPU	K	G	2HRP	M	H
1KCR	L	H	1NCW	L	H	1FJ1	A	D	1YEE	L	H
1OSP	L	H	1FNS	L	H	4FAB	L	H	1RJL	A	B
1N0X	M	H	1NSN	L	H	2HKF	L	H	1ETZ	A	H
1IKF	L	H	2AJX	L	H	3BZ4	G	D	1OP5	K	M
\<The Native Set without antigens\>											
Code	V_L	V_H	Code	V_L	V_H	Code	V_L	V_H	Code	V_L	V_H
1E6O	L	H	1RMF	L	H	2Q76	A	D	1IGI	L	H
1MCO	L	H	1BM3	L	H	1NLB	L	H	1MJU	L	H
3FZU	L	C	1AQK	L	H	1BZ7	A	B	1ADQ	L	H
1L7T	L	H	1HKL	L	H	1U6A	L	H	1UM5	L	H
3CFE	A	B	1UWG	X	Y	1PKQ	F	B	12E8	L	H
1ZA6	G	D	1MAM	L	H	1Q9O	C	B	1DFB	L	H
2FJF	M	P	1IAI	L	H	1A5F	L	H	1IT9	L	H
1NLD	L	H	1NGZ	A	B	3EYQ	C	D	1P7K	A	H
1PG7	L	Z	3EO9	L	H	1IGY	C	B	1F4Y	L	H
2PR4	L	H	2W9D	L	H	1MFE	L	H	2IG2	L	H
1F11	A	D									

Table S2 A full list of the RA set. The first column is the PDB code, the second and the third columns are the chain identifiers of the light and heavy chains.

Code	V_L	V_H	Code	V_L	V_H	Code	V_L	V_H	Code	V_L	V_H
2DDQ	L	H	1QBL	L	H	1CLZ	L	H	2FJH	L	H
1DQQ	A	B	1VFA	A	B	2CJU	L	H	2G5B	A	B
1Z3G	L	H	1IQD	A	B	1FOR	L	H	2H2P	D	C
2AI0	L	H	1K4C	B	A	1KB5	L	H	1FBI	L	H
1TET	L	H	1JHL	L	H	2AJU	L	H	1BJ1	L	H
1BQL	L	H	2AEP	L	H	1ZTX	L	H	1WC7	L	H
1CGS	L	H	2FBJ	L	H	1MCP	L	H	1ZAN	L	H
1MLB	A	B	1IGT	A	B	1NCA	L	H	2AJ3	A	B
2C1P	L	H	2FD6	L	H	2FJG	L	H	2DQU	L	H
2BDN	L	H	2ADF	L	H	2H1P	L	H	1F58	L	H
1FGN	L	H	2JEL	L	H	1FPT	L	H	1HZH	L	H
1JPT	L	H	1YNT	A	B	2ADG	A	B	1G9M	L	H
1A6T	A	B	1DBA	L	H	1IGM	L	H	2B4C	L	H
1KEM	L	H	2B2X	L	H						

Table S3 The detailed results of CDR prediction on the Native Set using DB-I.

CDR		Mean	STD	Median	Coverage
L1	FREAD	0.81	0.75	0.60	100.0%
	FREAD-S	1.25	1.05	0.97	100.0%
	ConFREAD	0.86	0.80	0.69	91.8%
	Best	0.39	0.39	0.69	100.0%
L2	FREAD	0.42	0.34	0.35	100.0%
	FREAD-S	0.59	0.38	0.56	100.0%
	ConFREAD	0.56	0.36	0.47	99.0%
	Best	0.22	0.22	0.47	100.0%
L3	FREAD	0.96	0.88	0.75	100.0%
	FREAD-S	0.80	0.62	0.64	100.0%
	ConFREAD	0.69	0.52	0.53	92.8%
	Best	0.37	0.33	0.53	100.0%
H1	FREAD	0.99	0.88	0.77	100.0%
	FREAD-S	1.13	0.96	0.80	100.0%
	ConFREAD	0.84	0.67	0.65	87.6%
	Best	0.56	0.74	0.65	100.0%
H2	FREAD	0.88	0.88	0.61	100.0%
	FREAD-S	0.96	0.87	0.67	100.0%
	ConFREAD	0.94	0.91	0.62	88.7%
	Best	0.46	0.56	0.62	100.0%
H3	FREAD	2.26	2.09	1.99	100.0%
	FREAD-S	1.38	1.56	0.71	100.0%
	ConFREAD	1.24	1.14	0.69	67.0%
	Best	0.91	0.97	0.69	100.0%

Table S4 The detailed results of CDR prediction on the Native Set using DB-E.

CDR		Mean	STD	Median	Coverage
L1	FREAD	1.61	1.99	0.87	100.0%
	FREAD-S	1.59	1.98	0.86	100.0%
	ConFREAD	0.94	0.77	0.78	34.0%
	Best	1.32	1.99	0.78	100.0%
L2	FREAD	0.50	0.45	0.38	100.0%
	FREAD-S	0.48	0.33	0.39	100.0%
	ConFREAD	0.51	0.31	0.42	92.8%
	Best	0.36	0.33	0.42	100.0%
L3	FREAD	1.49	1.43	1.11	100.0%
	FREAD-S	1.34	1.38	0.95	100.0%
	ConFREAD	0.84	0.41	0.76	58.8%
	Best	0.69	0.51	0.76	100.0%
H1	FREAD	1.49	1.35	1.01	100.0%
	FREAD-S	1.55	1.49	1.02	100.0%
	ConFREAD	0.96	0.58	0.77	59.8%
	Best	1.03	1.26	0.77	100.0%
H2	FREAD	1.60	1.11	1.34	100.0%
	FREAD-S	2.04	1.47	1.62	100.0%
	ConFREAD	1.33	0.68	1.29	42.3%
	Best	1.15	0.86	1.29	100.0%
H3	FREAD	3.81	1.99	3.46	88.7%
	FREAD-S	3.48	2.21	3.13	88.7%
	ConFREAD	0.97	0.66	0.61	6.2%
	Best	2.97	1.90	0.61	88.7%

Table S5 The detailed results of CDR prediction on the RA-Native set.

CDR		Mean	STD	Median	Coverage
L1	FREAD	1.09	0.73	0.90	100.0%
	FREAD-S	0.94	0.46	0.86	100.0%
	ConFREAD	0.90	0.51	0.73	90.7%
	RosettaAntibody	0.83	0.36	0.78	100.0%
L2	FREAD	0.59	0.31	0.51	100.0%
	FREAD-S	0.75	0.42	0.68	100.0%
	ConFREAD	0.72	0.38	0.68	100.0%
	RosettaAntibody	0.58	0.25	0.54	100.0%
L3	FREAD	1.00	0.58	0.93	100.0%
	FREAD-S	1.01	0.80	0.80	100.0%
	ConFREAD	0.80	0.67	0.61	94.4%
	RosettaAntibody	0.91	0.53	0.83	100.0%
H1	FREAD	0.88	0.41	0.82	100.0%
	FREAD-S	1.20	0.72	1.13	100.0%
	ConFREAD	0.84	0.59	0.69	96.3%
	RosettaAntibody	0.83	0.41	0.84	100.0%
H2	FREAD	0.90	0.62	0.73	100.0%
	FREAD-S	1.23	0.93	1.01	100.0%
	ConFREAD	0.83	0.72	0.69	81.5%
	RosettaAntibody	1.28	1.49	0.93	100.0%
H3	FREAD	1.85	1.62	1.78	100.0%
	FREAD-S	1.53	1.54	0.91	100.0%
	ConFREAD	1.38	1.24	0.84	68.5%
	RosettaAntibody	2.15	1.33	1.80	100.0%

Table S6 The detailed results of CDR prediction on the RA-Model set

CDR		Mean	STD	Median	Coverage
L1	FREAD	0.75	0.51	0.70	96.3%
	FREAD-S	0.95	0.57	0.86	96.3%
	ConFREAD	0.99	0.58	0.83	94.4%
	RosettaAntibody	0.82	0.35	0.76	100.0%
L2	FREAD	0.68	0.42	0.61	96.3%
	FREAD-S	0.86	0.43	0.85	96.3%
	ConFREAD	0.65	0.24	0.61	96.3%
	RosettaAntibody	0.59	0.25	0.54	100.0%
L3	FREAD	0.99	0.90	0.79	98.1%
	FREAD-S	1.13	0.94	0.83	98.1%
	ConFREAD	0.98	0.60	0.84	96.3%
	RosettaAntibody	0.90	0.53	0.83	100.0%
H1	FREAD	1.17	1.01	0.82	100.0%
	FREAD-S	1.49	0.96	1.16	100.0%
	ConFREAD	1.08	0.60	0.86	94.4%
	RosettaAntibody	0.82	0.41	0.82	100.0%
H2	FREAD	1.61	1.70	0.95	100.0%
	FREAD-S	1.86	1.75	1.41	100.0%
	ConFREAD	1.23	0.73	1.23	75.9%
	RosettaAntibody	1.03	0.55	0.91	100.0%
H3	FREAD	3.12	1.88	3.12	98.1%
	FREAD-S	2.07	1.40	1.68	98.1%
	ConFREAD	1.98	1.44	1.31	55.6%
	RosettaAntibody	2.96	2.08	2.65	100.0%

Table S7 A full list of the results (global loop RMSD) of the CDR-H3 loops of RA-Native set.

Code	FREAD	FREAD-S	ConFREAD	RosettaAntibody
2DDQ	4.06	3.82	3.82	1.7
1DQQ	0.64	0.37	1.01	0.3
1Z3G	0.59	0.59	0.59	1.8
2AI0	2.50	2.50	0.97	0.1
1TET	0.28	0.28	-	0.8
1BQL	0.32	0.32	0.32	1.5
1CGS	2.10	1.33	1.33	1.6
1MLB	0.73	0.93	0.58	2.9
2C1P	1.84	0.73	0.73	3.6
2BDN	2.55	3.30	-	0.8
1FGN	2.19	0.84	4.64	0.9
1JPT	1.08	1.43	1.24	0.9
1A6T	0.95	1.34	1.34	1.0
1KEM	0.27	0.52	2.46	1.6
1QBL	0.43	0.43	-	1.8
1VFA	0.51	0.23	0.13	2.0
1IQD	2.11	2.11	-	3.7
1K4C	0.12	0.69	0.77	0.6
1JHL	2.13	2.13	1.87	0.9
2AEP	2.29	2.61	-	1.0
2FBJ	3.61	1.60	1.60	1.1
1IGT	1.87	0.89	1.51	1.3
2FD6	0.31	0.33	1.86	1.4
2ADF	2.34	2.43	-	2.0
2JEL	0.95	2.42	-	2.0
1YNT	0.12	0.12	0.12	3.2
1DBA	1.11	0.55	0.60	0.7
2B2X	3.98	0.74	0.84	0.9
1CLZ	1.89	0.60	0.61	1.0
2CJU	2.36	1.26	1.05	1.2
1FOR	2.08	2.94	2.03	1.3
1KB5	3.49	2.29	3.13	1.7
2AJU	0.90	0.93	0.74	2.5
1ZTX	2.02	1.82	3.03	2.8
1MCP	1.46	0.44	0.44	1.1
1NCA	6.03	0.67	0.67	1.6
2FJG	0.50	2.02	0.50	2.6
2H1P	2.03	2.29	-	3.1
1FPT	1.72	3.55	-	3.6

2ADG	1.95	0.32	0.32	3.7
1IGM	1.90	5.51	-	3.1
2FJH	4.09	0.93	-	1.7
2G5B	0.36	0.36	3.63	1.9
2H2P	2.71	0.54	0.54	2.5
1FBI	4.35	0.51	-	5.7
1BJ1	0.34	0.34	0.34	2.0
1WC7	0.67	0.62	0.62	4.1
1ZAN	6.10	6.26	4.62	4.4
2AJ3	0.45	0.66	-	3.3
2DQU	0.48	0.48	0.48	2.7
1F58	0.70	1.48	-	5.3
1HZH	1.42	1.68	-	2.7
1G9M	0.53	0.71	-	2.3
2B4C	7.51	7.66	-	5.9
Mean	1.85	1.53	1.38	2.15
STD	1.62	1.54	1.24	1.33
Median	1.78	0.91	0.84	1.80

Table S8 A full list of the results (global loop RMSD) of the CDR-H3 loops of RA-Model set.

Code	FREAD	FREAD-S	ConFREAD	RosettaAntibody
2DDQ	4.20	3.68	-	1.4
1DQQ	0.71	0.71	1.12	0.3
1Z3G	0.71	0.71	4.62	2.4
2AI0	2.14	2.14	0.79	1.7
1TET	6.58	3.33	-	1.1
1BQL	4.37	4.74	-	1.6
1CGS	1.69	1.77	3.16	1.8
1MLB	4.75	0.74	1.97	1.4
2C1P	1.24	1.75	1.24	4.5
2BDN	4.84	4.84	-	3
1FGN	2.07	1.17	0.63	1.6
1JPT	2.18	2.11	1.12	2.2
1A6T	1.29	1.13	1.13	2
1KEM	3.38	1.60	2.06	2.1
1QBL	3.44	1.05	2.76	3.3
1VFA	1.75	1.19	1.23	1
1IQD	2.22	2.22	-	3.4
1K4C	3.28	0.70	0.54	2.7
1JHL	1.36	2.07	1.60	1.1
2AEP	2.89	2.89	-	1.2
2FBJ	3.74	1.68	-	2.7
1IGT	3.12	1.84	-	1.7
2FD6	3.57	1.08	-	1.9
2ADF	1.89	2.33	-	3
2JEL	1.90	2.80	-	1.4
1YNT	4.00	0.51	0.51	3.6
1DBA	0.77	1.17	1.02	4
2B2X	3.87	1.41	-	7.8
1CLZ	1.01	1.26	4.32	2
2CJU	5.13	1.37	1.37	1.2
1FOR	3.81	2.22	2.05	2.1
1KB5	2.99	2.86	4.50	3
2AJU	3.52	1.01	1.10	2.6
1ZTX	2.60	2.97	-	1.9
1MCP	3.19	1.11	-	2.9
1NCA	11.56	0.82	-	3.4
2FJG	2.82	2.49	-	2
2H1P	1.96	2.28	-	1.7
1FPT	3.62	3.14	3.72	2.2

2ADG	4.82	1.52	4.61	4.2
1IGM	1.92	5.53	-	4.5
2FJH	3.85	1.07	4.37	3
2G5B	0.84	0.84	3.65	3
2H2P	4.17	4.17	-	1.6
1FBI	4.33	4.63	-	5.8
1BJ1	5.28	0.58	0.50	2.8
1WC7	1.52	1.48	1.87	2.7
1ZAN	4.90	6.37	-	4.3
2AJ3	4.60	4.57	-	3.8
2DQU	5.02	0.35	0.35	3.1
1F58	2.58	2.00	-	13.6
1HZH	0.74	1.19	0.74	6.4
1G9M	0.69	0.78	0.68	4.6
2B4C	-	-	-	5.6
Mean	3.12	2.07	1.98	2.96
STD	1.88	1.40	1.44	2.08
Median	3.12	1.68	1.31	2.65

Table S9 Detailed results of the FREAD variants on the Bound-Free set (CDR-L1 to CDR-H2).

			CDR-L1			
Free	Bound	Free-Bound	FREAD	FREAD-S	ConFREAD	Best
1NGZ	1N7M	2.35	2.51	2.3	2.3	2.13
1D5I	1D6V	0.27	0.31	2.79	0.98	0.27
2A6J	2A6I	1.2	1.8	1.34	1	0.71
1Q9K	1Q9Q	1.7	1.4	1.76	2.43	1.28
1KCV	1KCS	0.34	0.52	1.1	0.95	0.34
1CR9	1CU4	0.56	1.65	2.1	1.86	0.39
1GGC	1GGI	1.43	1.83	2.15	2.03	1.41
1CGS	2CGR	5.25	5.94	5.94	-	5.64
1NBV	1CBV	1.25	3.01	3.48	2.16	1.67
1HIL	1IFH	1.59	1.27	1.69	1.69	1.23
1OAQ	1OAU	0.35	0.95	0.72	-	0.31
1MNU	1MPA	2.43	2.55	2.95	2.62	2.04
Mean		1.56	1.98	2.36	1.8	1.45
STD		1.38	1.49	1.38	0.63	1.48
			CDR-L2			
Free	Bound	Free-Bound	FREAD	FREAD-S	ConFREAD	Best
1NGZ	1N7M	1.28	1.39	1.27	0.96	0.77
1D5I	1D6V	0.39	0.44	0.4	0.4	0.32
2A6J	2A6I	1.05	1.37	1.14	1.14	0.76
1Q9K	1Q9Q	0.77	0.79	1.14	1.14	0.49
1KCV	1KCS	0.26	0.29	1.03	0.29	0.25
1CR9	1CU4	0.43	0.45	0.46	0.44	0.32
1GGC	1GGI	1.48	1.55	1.44	1.55	1.39
1CGS	2CGR	2.53	3.11	3.2	3.2	2.52
1NBV	1CBV	0.73	1.23	0.73	0.73	0.25
1HIL	1IFH	1.34	1.32	1.34	1.34	1.09
1OAQ	1OAU	0.14	0.25	0.36	0.36	0.1
1MNU	1MPA	1.28	1.4	1.56	1.56	0.74
Mean		0.97	1.13	1.17	1.09	0.75
STD		0.67	0.79	0.76	0.81	0.68
			CDR-L3			
Free	Bound	Free-Bound	FREAD	FREAD-S	ConFREAD	Best
1NGZ	1N7M	2.3	2.38	2.63	2.01	1.5
1D5I	1D6V	0.37	1.07	0.77	2.21	0.26
2A6J	2A6I	0.95	1.08	1.28	1.36	0.54
1Q9K	1Q9Q	0.46	0.41	0.41	0.43	0.29
1KCV	1KCS	0.93	0.9	0.55	0.71	0.44
1CR9	1CU4	0.64	1.24	1.23	1.14	0.38
1GGC	1GGI	1.28	1.89	1.12	1.1	0.55
1CGS	2CGR	3.92	3.82	4.17	4.29	3.03
1NBV	1CBV	1.02	1.94	2.05	1.08	0.8

Free	Bound	Free-Bound	FREAD	FREAD-S	ConFREAD	Best
1HIL	1IFH	0.57	1.95	0.62	0.85	0.52
1OAQ	1OAU	0.86	1.3	1.01	0.57	0.24
1MNU	1MPA	1.29	1.57	1.84	1.94	1.11
Mean		1.22	1.63	1.47	1.47	0.81
STD		0.99	0.88	1.08	1.05	0.79

CDR-H1

Free	Bound	Free-Bound	FREAD	FREAD-S	ConFREAD	Best
1NGZ	1N7M	1.61	2.47	1.87	1.42	0.71
1D5I	1D6V	0.39	0.41	0.92	0.37	0.36
2A6J	2A6I	0.68	1.44	1.77	0.87	0.57
1Q9K	1Q9Q	0.81	0.71	1.56	0.49	0.42
1KCV	1KCS	0.89	1.34	0.83	-	0.68
1CR9	1CU4	1.55	2.32	2.57	-	2.32
1GGC	1GGI	2.03	2.97	1.97	1.97	1.41
1CGS	2CGR	5.63	4.55	4.55	-	4.55
1NBV	1CBV	2.34	2.27	2.38	1.85	1.85
1HIL	1IFH	3.27	3.27	3.27	3.27	2.85
1OAQ	1OAU	0.18	0.84	0.89	1.07	0.15
1MNU	1MPA	0.84	0.82	1.92	1.33	0.68
Mean		1.69	1.95	2.04	1.4	1.38
STD		1.53	1.25	1.07	0.89	1.31

CDR-H2

Free	Bound	Free-Bound	FREAD	FREAD-S	ConFREAD	Best
1NGZ	1N7M	1.77	1.82	2.46	2.46	1.3
1D5I	1D6V	1.99	2.24	2.76	2.76	1.37
2A6J	2A6I	2.17	7.97	1.22	0.39	0.39
1Q9K	1Q9Q	0.42	0.41	0.74	0.74	0.33
1KCV	1KCS	0.83	1.02	1.18	1.18	0.68
1CR9	1CU4	1.09	0.84	1.81	0.9	0.69
1GGC	1GGI	1.18	1.4	1.4	-	0.77
1CGS	2CGR	4.3	5.08	4.25	5.18	3.48
1NBV	1CBV	1.65	1.12	1.51	2.21	1.12
1HIL	1IFH	2.27	2.62	1.87	-	1.67
1OAQ	1OAU	0.23	1.64	0.28	0.3	0.15
1MNU	1MPA	0.93	3.88	1.24	1.24	0.65
Mean		1.57	2.5	1.73	1.73	1.05
STD		1.09	2.18	1.04	1.48	0.89