

Supplementary Information

Tables

Supplementary Table 1: Disulfide bond features. Table showing the C_α-C_α distance of the cystines in CDR H1 and H3 loops that form a disulfide bond. Details on the position of the cystines are also shown. ^sstretched twist, ^ttwist.

PDB ID	H3 Length	Cys(C _α)-Cys(C _α) (Å)	Cys Position in H1	Cys Position in H3	Relative Position of Cys in H3		H3 Midpoint
					End	Start	
1YC7	10	4.0	32	96	<i>n</i> -8	2	5.0
1ZV5 ^t	12	6.1	33	100B	<i>n</i> -4	8	6.0
1RI8 ^s	17	5.7	33	100B	<i>n</i> -9	8	8.5
1F2X ^t	19	5.7	33	100C	<i>n</i> -10	9	9.5
1RJC ^s	19	5.9	33	100D	<i>n</i> -9	10	9.5
1JTO ^s	24	5.7	33	100E	<i>n</i> -13	11	12.0
1MEL ^s	24	5.7	33	100E	<i>n</i> -13	11	12.0
1XFP ^s	24	5.7	33	100E	<i>n</i> -13	11	12.0
1ZMY ^s	24	5.8	33	100E	<i>n</i> -13	11	12.0
Mean Distance		5.6					
Standard Deviation		0.6					

Supplementary Table 2: Comparison of different loop building techniques for long CDR H3 loops ($n_{H3} \geq 16$ residues) using global rmsds. *LowE*, *LowRMS* and *LowALL* indicates the lowest-scoring CDR H3 loop model, the CDR H3 with the lowest global rmsd in the ten-lowest scoring models and the model with the lowest global-rmsd observed respectively. ^s and ^t indicates structures whose CDR H3 loops adopt the stretched-twist and twist conformations respectively. ^c indicates that the native contains a disulfide bond between CDRs H1 and H3.

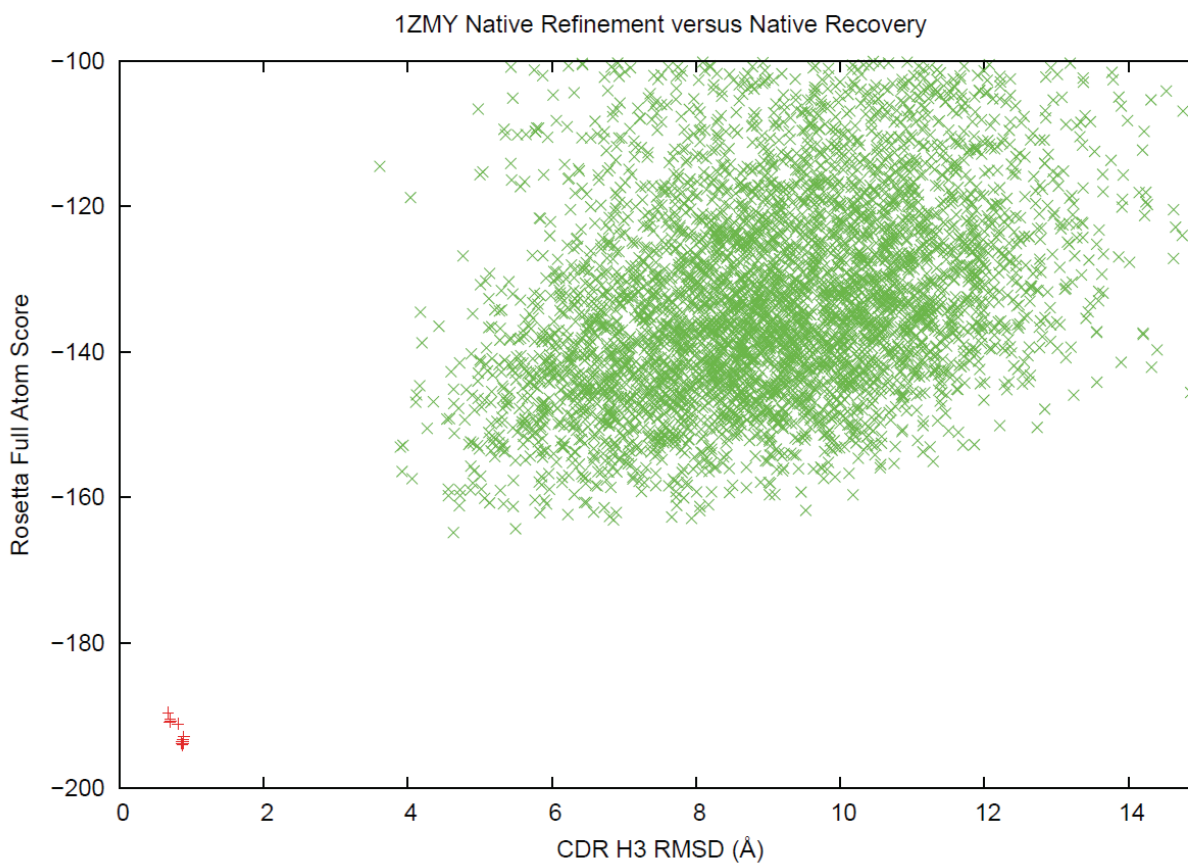
No. of models		No constraints with nine residue fragments only (Å)			Constraints with nine residue fragments only (Å)			Constraints with three residue fragments following nine residue fragments (Å)					
		5,000			5,000			5,000			20,000		
		PDB ID	CDR H3 length	LowE	LowRMS	LowALL	LowE	LowRMS	LowALL	LowE	LowRMS	LowALL	LowE
1ZVH ^s	16	7.3	3.3	2.9	5.3	2.3	2.2	4.5	3.4	2.3	4.9	3.2	2.3
1UOQ ^{s,c}	16	2.6	2.6	2.5	3.0	2.8	2.8	2.8	2.8	2.8	3.1	2.8	2.1
3DWT ^s	16	8.0	3.3	2.6	4.2	3.1	2.5	3.3	3.3	2.6	3.2	3.2	2.6
1QD0 ^s	16	7.4	3.1	2.5	1.5	1.5	1.5	0.9	0.9	0.9	0.9	0.9	0.9
1MVF ^s	17	9.3	4.0	3.5	4.0	3.9	3.2	3.4	3.4	3.4	2.3	2.3	1.8
1R18 ^{s,c}	17	10.0	4.6	4.0	4.8	3.3	2.7	5.1	4.0	2.6	3.6	2.6	2.5
1ZVY ^s	18	6.6	6.6	3.0	3.9	3.1	2.8	4.3	3.3	3.0	4.3	3.4	2.2
1F2X ^{t,c}	19	14.7	9.0	4.5	5.4	4.6	3.5	3.9	3.9	3.3	5.7	3.3	3.0
1RJC ^{s,c}	19	7.1	6.3	4.4	2.5	2.5	2.5	3.6	3.6	2.8	3.6	2.0	2.0
1JTO ^{s,c}	24	10.8	5.2	4.8	4.5	4.1	3.7	6.7	4.3	3.4	5.4	4.3	3.4
1MEL ^{s,c}	24	18.6	6.2	3.6	5.1	4.6	3.5	4.0	4.0	3.8	6.3	4.0	3.3
1XFP ^{s,c}	24	12.9	9.4	5.2	5.0	5.0	3.2	4.2	4.2	3.9	4.5	4.2	3.4
1ZMY ^{s,c}	24	7.8	7.6	4.8	4.6	4.6	3.6	3.9	3.9	3.8	4.0	3.9	3.3
Median		8.0	5.2	3.6	4.5	3.3	2.8	3.9	3.6	3.0	4.0	3.2	2.5

Figure

Supplementary Figure 1: Sequence alignment of cAb V_HH and classical V_HS. The camelid V_HH are listed first with the letter “A” following the four letter PDB ID. Red (*) and black stars (*) indicate alignment positions of newly and previously discovered amino acid mutations. The classical V_HS have the letter “H” following their PDB IDs. The CDR regions are annotated in red. Basic: K,R (red); Hydrophobic: A,F,I,L,M,V,W (blue); Polar: N,Q,S,T (green); H,Y (cyan); C (salmon); Acidic: D,E (magenta); Proline (yellow); Glycine (orange).

←-----H3----->

101	*	*					*	*	*	*
1bzq.A	EDTAVYCAA	GGYELRDRT					YGQWGGTQ	VTVS		
1f2x.A	EDTAIYACAG	STVASTGWCS	RLRPYD				YHYRGGTQ	VTVSS		
1g9e.A	EDTAVYTCGA	GEGGT					WDSWGGTQ	VTVS		
1hcv.A	EDTAVYTCGA	GEGGT					WDSWGGTQ	VTVSS		
1i3v.A	EDTAVYCAA	KTTTWGGNDP	NN				WNYWGGTQ	VTVSS		
1ieh.A	EDTAVYCAK	YSGGA					LDAWGGTQ	VTVSSQSEOK	LISEEDLNHH	HHH
1jto.A	EDTAIYCAA	DSTIYASYE	CGHGLSTGGY				GYDSWGGTQ	VTVSS		
1kxq.A	EDTGIYCAT	GNSVRLASWE	G				YFYWGGTQ	VTVSS		
1me1.A	EDTAIYCAA	DSTIYASYE	CGHGLSTGGY				GYDSWGGTQ	VTVS		
1mvf.A	EDTAMYCAA	SSRWMDYSAL	TAKA				YNSWGGTQ	VTVSSR		
1op9.A	EDTAMYCAA	TEVAGWPLDI	GI				YDYWGGTE	VTVSS		
1qd0.A	EDTAVYCAA	RPVRVADISL	PVG				FDYWGGTQ	VTVSS		
1ri8.A	EDTAMYCAA	GWSSLGSCGT	NRNR				YNYWGGTQ	VTVSS		
1rjc.A	EDTAMYCAA	DTSTWYRGYC	GTNPNY				FSYWGGTQ	VTVS		
1shm.A	EDTAVYTCGA	GRIGRSVFNL	RRESW				VTWGGTQ	VTVSS		
1sxx.A	EDTAVYCAA	EDRHRIGT					NGYWGGTQ	VTVSS		
1u0q.A	DDTAVYCAV	RMPYSGDYRS	SGT				YDYWGGTQ	VTVSS		
1xfp.A	EDTAIYCAA	DSTIYASYE	CGHGLSTGGY				GYDSWGGTQ	VTVS		
1yc7.A	EDTGMYYCOI	QCGVRSI					REYWGGTQ	VTVS		
1zmy.A	EDTAMYCAA	DSTIYASYE	CGHGLSTGGY				GYDSWGGTQ	VTVSS		
1zv5.A	EDTASYCAA	GYNNGOCA					TRYWGGTQ	VTVS		
1zvh.A	EDTALYCAA	ARQWYIPLN	SYG				YNYWGGTQ	VTVS		
1zvy.A	EDTAIYTCGA	T-RKYVPRF	ALDQSS				YDYWGGTQ	VTV		
2bse.A	EDTAIYCAA	RSGGFSSNRE	L				YDGGGGTQ	VTVSS		
2p49.A	EDTAVYCAA	GGYELRDRT					YGQWGGTQ	VTVSS		
3cfi.A	EDTAVYCAK	-WL--GGR					DWYDRGGTQ	VTVS		
3dwt.A	EDTAIYCAA	VRGYFMRLPS	SHN				FRYWGGTQ	VTVSSR		
3ezj.A	EDTAVYCAV	NVKTWAGMT					RDYWGGTQ	VTVSS		
1a6t.H	EDSAVYCAR	RDDYY					FDVWGGTQ	VT		
1bj1.H	EDTAVYCAK	YPHYGSS	HWY				FDVWGGTQ	VT		
1bq1.H	EDSGVYCLH	GNYP					FDVWGGTQ	VT		
1cgs.H	EDSAVYCTR	GYSS					MDYWGGTQ	VT		
1clz.H	EDTAMYCAR	GLDDG	AW				FAYWGGTQ	VT		
1dba.H	EDTAVYCTR	GDYVNWY					FDVWGGTQ	VT		
1dqq.H	EDTAVYCAS	WG					GDVWGGTQ	VTVSS		
1f58.H	EDTAVYCAR	EEAMPYGNOA	YYYA				MDCWGGTQ	VTVSS		
1fbi.H	EDSAVYCAS	LYYGTSTYGV					LDYWGGTQ	VT		
1fgn.H	EDTAVYCAR	DNSYY					FDYWGGTQ	VT		
1for.H	EDSAVYCAR	SGNYPYA					MDYWGGTQ	VT		
1fpt.H	DDSAVYCAR	DFYDYDVG					FDYWGGTQ	VT		
1g9m.H	DDTAVYFCAV	VYEGEAGEGE	YRNN			G	FLKHWGGTQ	VTVTS		
1hzh.H	ADTAVYCAR	VGPYSWDDSP	QDNY				YMDVWKGTT	VTVSS		
1igm.H	EDTAIYCAK	HRVSVLTC					FDWGGTQ	VT		
1igt.H	EDTAMYCAR	HGGYYA					MDYWGGTQ	VT		
1iqd.H	DDTAVYCAV	PDPDA					FDIWGGTQ	VTVSS		
1jhl.H	EDSAVYCTR	DDNYGA					MDYWGGTQ	VT		
1jpt.H	EDTAVYCAR	DTAAY					FDYWGGTQ	VT		
1k4c.H	EDSAVYCAR	ERGDGY					FAYWGGTQ	VT		
1kb5.H	EDSAVYCAR	SRTDLYY					FDYWGGTQ	VT		
1kem.H	EDSATYCAR	WGSYA					MDYWGGTQ	VT		
1mcp.H	EDTAIYCAR	NYGSS	TWY				FDVWGGTQ	VT		
1mlb.H	EDSAVYCAR	GDBG					YGYWGGTQ	VT		
1nca.H	EDTATFFCAR	GEDNFGSL					SDYWGGTQ	VT		
1qbl.H	EDTAVYCAV	YDYG					FDYWGGTQ	VT		
1tet.H	EDTAVYCAR	RSWY					FDVWGGTQ	VT		
1vfa.H	DDTAVYCAR	ERDYR					LDYWGGTQ	VT		
1wc7.H	EDSATYCVR	DKGSYGNYP	AW				FAYWGGTQ	VT		
1ynt.H	EDSAIYCAR	SSTWY					FDYWGGTQ	VT		
1z3g.H	EDSAVYCAR	GWD					VAYWGGTQ	VTVSA		
1zan.H	EDTAVYCAR	DGGYSSST	LYA				MDAWGGTQ	VT		
1ztx.H	EDSAVYCAR	SASYGDY					ADYWGHTQ	VT		
2adf.H	EDTAVYCAR	DNPYYA					LDYWGGTQ	VT		
2adg.H	EDTAMYCAR	HEDGNWNY					FDYWGGTQ	VT		
2aep.H	EDSATYCAR	VDYGTN					YDYWGGTQ	VT		
2ai0.H	EDTAVYCAR	HDDYGGKS	PYF				FDVWGGTQ	VTASS		
2aj3.H	ADTAVYCAR	YHRHFIRGPI	S				FDYWRGGTQ	VT		
2aju.H	EDTAVYCAR	YDYGNT					GDYWGGTQ	VT		
2b2x.H	EDTAVYCTR	GFGDGGY					FDVWGGTQ	VT		
2b4c.H	DDTAVYCAR	DFGPDWEDGP	SYDGSGRGF				FDVWGGTQ	VTVSS		
2b4n.H	EDTAVYCAR	GVFGE					FDYWGGTQ	VT		
2c1p.H	EDSAVYCAR	DDYD					GAFWGGTQ	VTVSA		
2cju.H	EDSATYCAR	GYG	AW				FAYWGGTQ	VT		
2ddg.H	EDSAVYCAP	Y					GGYWGGTQ	VTVSS		
2dqu.H	EDTAMYCAR	VSHYDGSRDW	Y				FDVWGGTQ	VT		
2fbj.H	EDTALYCAR	LHYYGY					NAYWGGTQ	VT		
2fd6.H	EDSAVYCAR	WGPHWY					FDVWGGTQ	VT		
2fjg.H	EDTAVYCAR	FVFFLPYA					MDYWGGTQ	VT		
2fjh.H	EDTAVYCAR	WGHSTSPWA					MDYWGGTQ	VT		
2g5b.H	EDSATYCVR	DNGSDY	RWY				FDVWGGTQ	VT		
2h1p.H	EDTALYCAR	RDSSASLY					FDYWGGTQ	VT		
2h2p.H	EDTALYCAR	LYYGYWY					FDVWGGTQ	VT		
2jel.H	EDSAIYCAR	VMGEQY					FDVWGGTQ	VT		



Supplementary Figure 2: Plot of Rosetta energies of models versus RMSD from the native crystal structure in Ångstroms, comparing refinement of crystal conformation (red) and *ab initio* loop building (green) of the CDR H3 loop in the native crystal environment.

Rosetta Command Lines

The protocols used in this paper are available in Rosetta 3.2 (Fall 2010) using the following command lines, where `$ROSETTA3` and `$ROSETTA3DB` are environment variables pointing to the Rosetta 3.0 source and database paths. `$RosettaAntibody` is the environment variable pointing to the RosettaAntibody directory, typically `$ROSETTA3/antibody`.

CDR Grafting

```
$ROSETTA3/antibody_mode.default.linuxgccrelease
  -database $ROSETTA3DB -nstruct 1
  -s FR02.pdb -native 1mvf.pdb -in::file::fullatom -out::file::fullatom -
  out::prefix aa
  -graft_h1 -graft_h2 -graft_h3 -camelid
  -ex1aro -ex1 -ex2
  -detect_disulf true -run::rebuild_disulf true -run::find_disulf true
  -mute core -unmute protocols.AntibodyModeler -out::level 550
```

CDR H3 Loop Modeling

```
$ROSETTA3/antibody_mode.default.linuxgccrelease
  -database $ROSETTA3DB -nstruct 5000
  -s FR02.pdb -in::file::native 1mvf.pdb -in::file::fullatom -out::file::fullatom -
  out::prefix aa
  -model_h3 -camelid -constraints::cst_file camelid.cst
  -loops::strict_loops -loops::frag_sizes 9 3 -loops::frag_files
  aaFR02_09_05.200_v1_3 aaFR02_03_05.200_v1_3
  -ex1aro -ex1 -ex2
  -mute core -unmute protocols.AntibodyModeler -out::level 550
  -in::fix_disulf FR02.disulf -run::rebuild_disulf [optional flags: only if H1 and
  H3 both have CYS]
```

Master Script

The cAb V_HH sequence should be provided in a file, e.g. `query_h.fasta`. A master script executed from the directory containing the FASTA sequence will carry out all steps necessary for camelid antibody modeling:

```
perl -P $RosettaAntibody/scripts/ram_buildloop_wrapper.pl pdb1xyz_chothia.pdb 1 1 1 1
  1 1 5000 query_h.fasta camelid
```

Fragments

Protein fragments must be built for loop building. H3 sequences plus one N-terminal and two C-terminal step residues can be submitted to the Robetta server (<http://www.robetta.org>) or generated with the `nn_make.pl` script included in the Rosetta2.x distribution.