Supporting Information

Keeffe et al. 10.1073/pnas.1108777108

SI Results

CV-N Dimers Are Domain-Swapped in Crystal Structures. To elucidate a mechanism for the enhanced neutralization activity of the dimeric proteins, we solved crystal structures of CVN_2L0 (PDB accession code 3S3Y) and CVN_2L10 (PDB accession code 3S3Z). Both proteins crystallized under many conditions and structures for both were determined from crystals in a trigonal spacegroup. Structures of WT CV-N had previously been solved by others, both in solution (1) (Fig. S1*A*), and from two different crystallization states: from trigonal crystals as described above (2) (Fig. S1*C*), and from tetragonal crystals (3) (Fig. S1*B*). Our CVN_2 linked dimer structures were solved by molecular replacement using the domain-swapped WT CV-N structure from trigonal crystals (PDB ID: 3EZM) (2) as the initial model (Table S4).

The structures contained only half of the CVN_2 dimer in the asymmetric unit, such that the second tandem repeat of CV-N was generated through crystallographic symmetry (Fig. S2*A*). Refinement was complicated by the fact that the two copies of CV-N related by crystallographic symmetry were one polypeptide covalently linked through a flexible polypeptide linker. To account for the termini and linker residues occupying the same locations in the crystal, the free termini (CVN_2L0) residues were modeled in the structure at 50% occupancy. In the CVN_2L10 structure, the free and linked termini appeared to exist in the same conformation due to the long flexible linker and therefore were modeled at 100% occupancy. Linker residues with defined density in all structures were modeled at 50% occupancy (Fig. S2*B*). SDS-PAGE verified that crystals grown under the same conditions contained exclusively CVN_2 and were not contaminated with any cleaved CV-N.

Both structures were determined to be intramolecularly domain-swapped with no major deviations relative to WT CV-N domain-swapped structures (Fig. S2 *A* and *C*). Small differences between the CVN_2 structures and previously published WT CV-N structures were noted near the termini, as expected from the linkage, and also in the region of the domain swap. These differences, however, are minor and are unlikely to be responsible for the observed differences in antiviral activity.

SI Materials and Methods

Crystallization. Protein samples were concentrated to 25 to 30 mg/mL using 5,000 molecular weight cutoff (MWCO) Amicon Ultra concentrators (Millipore). Crystallization conditions were set up using a Mosquito automated nanoliter pipettor (TTP Labtech). Screening was done with 480 conditions in 96-well sitting drop plates using 0.3×0.3 μL drops. Each protein crystallized under many conditions, and suitable crystals were found for data collection from these initial screens. Data were collected on CVN_2L0 crystals grown in 0.1 M sodium Hepes pH 7.5, 0.8 M potassium dihydrogen phosphate, 0.8 M sodium dihydrogen phosphate. The CVN_2L10 dataset was collected on a crystal grown in 0.2 M sodium fluoride and 20% PEG-3350.

Crystallographic Data Collection and Refinement. All crystals were cryoprotected in TMP oil. Data for the CVN_2L0 structure were collected using a MicroMax-007HF X-ray generator with an RAXIS IV++ detector (Rigaku Corp.). The CVN_2L10 dataset was collected on the 12-2 beam line at the Stanford Synchrotron Radiation Lightsource (SSRL). All data were processed using CrystalClear (Rigaku Corp.) and Mosflm (4). The indexed and scaled data were further evaluated using CCP4i (5). Molecular replacement was done with Phaser version 1.3.3 (6) using 3EZM (2) as the starting model. Further refinement was done using Refmac (7) and COOT was used for model building (8). Simulated annealed omit maps were generated using CNS (9, 10). Figures were made using PyMOL (11).

1. Bewley CA (2001) Solution structure of a cyanovirin-N:Manα1-2Manα complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* 9:931–940.
2. Yang F, et al. (1999) Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol* 288:403–412.
3. Botos I, et al. (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* 277:34336–34342.
4. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Prot. Crystallography* 26:27–33.
5. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. in *Acta Crystallogr D Biol Crystallogr* (Collaborative Computational Project, Number 4), pp 760–763.
6. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61:458–464.
7. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
8. Emsley P, Cowtan K (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
9. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2:2728–2733.
10. Brunger AT, et al. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
11. DeLano WL (2008) The PyMOL Molecular Graphics System. (http://www.pymol.org, DeLano Scientific, Palo Alto, CA).

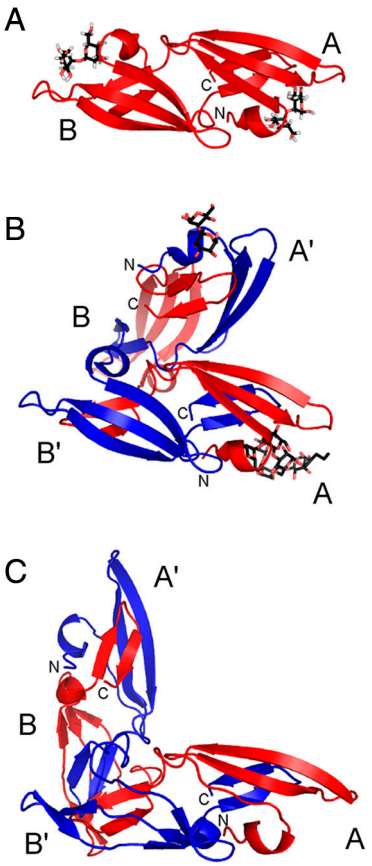

Fig. S1. Structures of WT CV-N. (*A*) NMR structure of the WT CV-N monomer with carbohydrates bound in both the A and B domains (1). Protein is shown in red and carbohydrates are shown as sticks with black carbons. N and C termini and domains are labeled. (*B*) Crystal structure of a WT CV-N-carbohydrate complex obtained from tetragonal crystals (2). Carbohydrates are seen bound to domains A and A′, but not to domains B or B′. The two chains of CV-N are shown in red and blue, and carbohydrates are shown as sticks with black carbons. Domains A and B′ are shown in roughly the same position as domains A and B in (*A*) to facilitate comparison. (*C*) Crystal structure of uncomplexed WT CV-N obtained from trigonal crystals (3). Colors and labels are the same as (*B*). Domains A and B′ are shown in roughly the same position as domains A and B in (*A*) to facilitate comparison.

1. Bewley CA (2001) Solution structure of a cyanovirin-N:Manα1-2Manα complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* 9:931–940.
2. Botos I, et al. (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* 277:34336–34342.
3. Yang F, et al. (1999) Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain-swapping. *J Mol Biol* 288:403–412.

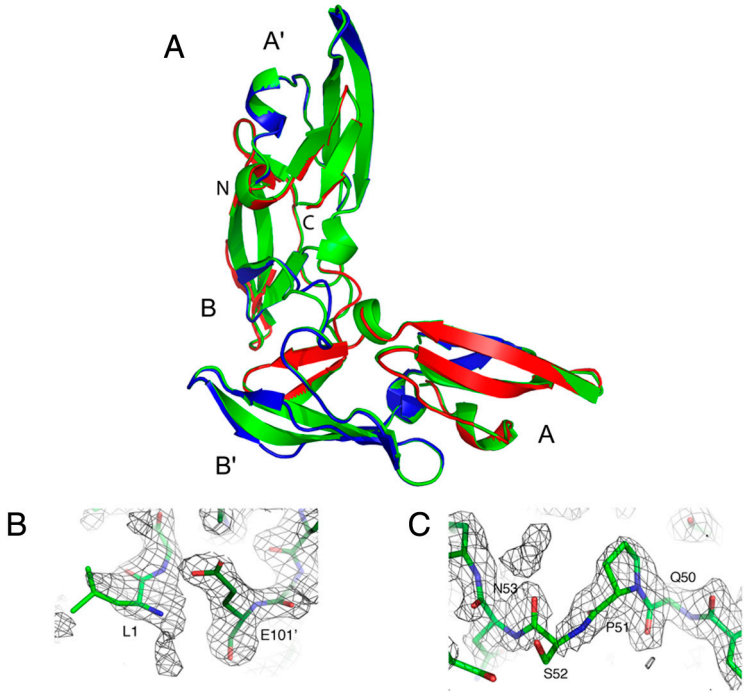

Fig. S2. CV-N dimer crystal structures are domain-swapped and similar to WT CV-N. (*A*) CVN_2L0 structure (green) shown superimposed over a WT CV-N structure (PDB ID 3EZM) (1) (red and blue) (rmsd of 0.31 Å for superposition of all atoms). The only significant difference between the CVN_2L0 structure and the WT structure is in the domain-swapping region. The four domains are labeled as in Fig. 1. (*B*) Annealed omit map of the termini/linker region of CVN_2L0 (residues 1 and 101 of alternative chains) contoured at 1.0σ. Free termini were modeled at 50% occupancy, but there was additional density for the linkage. (*C*) Annealed omit map of the domain-swapped region (residues 50–53) contoured at 1.0σ.

1. Yang F, et al. (1999) Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain-swapping. *J Mol Biol* 288:403–412.

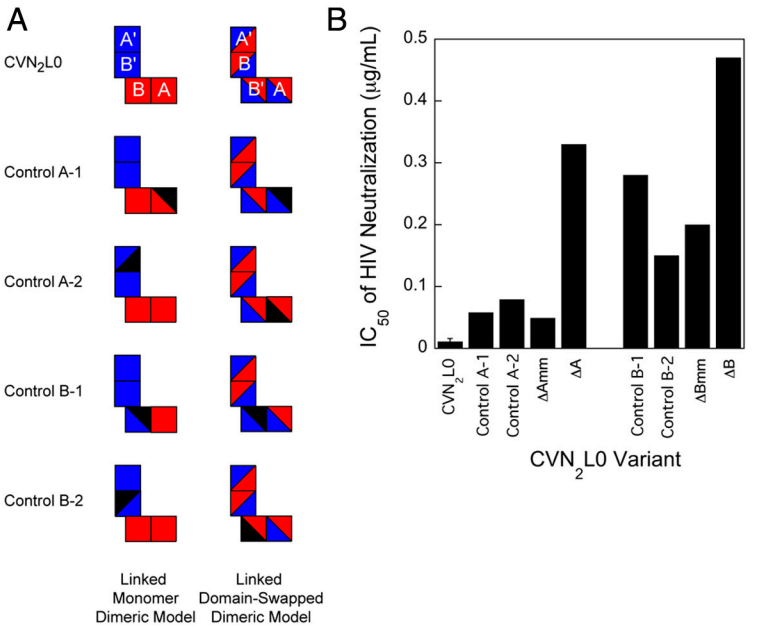

Fig. S3. Partial binding site knockouts do not significantly affect the neutralization potency of CVN_2L0. (*A*) Each binding site in the domain-swapped dimer is composed of residues from both chain A (red) and chain B (blue). In the monomeric model, each binding site is made of residues entirely in one CV-N repeat (either red or blue). To test the hypothesis that an entire binding site needs to be knocked out in order to see a significant decrease in the potency of HIV neutralization, we created control constructs that contained only half binding site knockouts (see Table S5 for mutations). Schematics for ΔAmm, ΔA, ΔBmm, and ΔB can be found in Fig. 4 in the main text. (*B*) Mutations to half of the A binding site (black triangles in Control A-1 and Control A-2) do not significantly affect the potency compared to wild-type CVN_2L0. The results are similar for the B binding site (Control B-1 and Control B-2), although the half binding site knockout in Control B-1 has a fairly substantial effect on the protein. Overall, the potency of the monomeric model deletion variants (ΔAmm and ΔBmm) are similar to the individual half binding site knockouts, indicating they contain two half-site knockouts, whereas the variants based on the domain-swapped dimer (ΔA and ΔB) show significant decreases in potency, indicating they are full binding site knockouts.

Table S1. CV-N oligomer linker sequences

Variant	Linker sequence
CVN_2L0	None
CVN_2L1	G
CVN_2L3	GSG
CVN_2L5	GGSGG
CVN_2L6	GSGGSG
CVN_2L7	$(GGS)_2G$
CVN_2L8	$(GGS)_2GG$
CVN_2L9	GGSGGGSGG
CVN_2L10	$(GGSGG)_2$
CVN_2L20	$(GGSGG)_4$
CVN_3L0	None
CVN_3L5	GGSGG
CVN_3L10	$(GGSGG)_2$
CVN_4*	None/$(GGSGG)_4$/None

*(CVN_2L0)-L20-(CVN_2L0).

Table S2. HIV neutralization IC_{50} values for CV-N, CV-N oligomers, and CV-N dimer binding site knockouts*

Variant	IC_{50} (nM)	IC_{50} (μg/mL)
WT CV-N	4.4 ± 2.6	0.054 ± 0.032
CVN_2L0	0.49 ± 0.23	0.011 ± 0.005
CVN_2L1	0.67 ± 0.11	0.016 ± 0.002
CVN_2L3	0.68 ± 0.04	0.016 ± 0.001
CVN_2L5	0.91 ± 0.34	0.022 ± 0.008
CVN_2L6	0.50 ± 0.14	0.012 ± 0.003
CVN_2L7	0.56 ± 0.13	0.013 ± 0.003
CVN_2L8	0.40 ± 0.09	0.009 ± 0.002
CVN_2L9	0.41 ± 0.02	0.010 ± 0.001
CVN_2L10	0.76 ± 0.32	0.018 ± 0.008
CVN_2L20	2.7	0.067
CVN_3L0	0.38 ± 0.17	0.013 ± 0.006
CVN_3L5	0.86	0.030
CVN_3L10	1.3	0.046
$CVN_2L0_{\Delta Amm}$	2.1	0.049
$CVN_2L0_{\Delta A}$	14	0.33
$CVN_2L0_{\Delta Bmm}$	8.7	0.20
$CVN_2L0_{\Delta B}$	20	0.47
$CVN_2L0_{\Delta A\|B}$	1,000	23
$CVN_2L0_{\Delta A\times B}$	74	1.7
$CVN_2L0_{\Delta AA}$	26	0.61
$CVN_2L0_{\Delta BB}$	600	14
$CVN_2L0_{\Delta ABB}$	>1,000	>23
$CVN_2L0_{\Delta AAB}$	>1,000	>23

*The reported error of IC_{50} values (±) represents the standard deviation of at least three independent trials ($N \geq 3$).

Table S3. IC_{50}s (nM) of CV-N and HIV neutralizing antibodies against various HIV strains from clades A, B, and C

		IC_{50} (nM)									
Clade	Envelope	4E10	2G12	2F5	b12	PG 9	PG 16	VRC01	WT CV-N	CVN_2L0	CVN_2L10
A*	DJ263.8	NT	NT	NT	NT	NT	NT	NT	7.5	0.46	0.48
	Q23.17	120	>330	49	>330	NT	0.014	0.41	15	1.4	2.6
	Q842.d12	93	>330	57	>330	0.20	0.18	0.34	19	2.8	3.3
	Q259.d2.17	95	>330	71	>330	NT	0.068	0.69	160	14	41
	3718.v3.c11	77	>330	23	>330	0.47	0.068	2.5	64	4.0	8.9
	0330.v4.c3	39	4.7	63	>330	0.13	0.14	0.41	2.7	0.11	0.25
	3415.v1.c1	160	14	240	170	0.53	0.14	0.34	5.7	0.28	0.44
B†	SF163.LS	2.0	4.0	0.67	0.067	NT	NT	NT	16	1.1	1.7
	PV0.4	43	8.0	>330	>330	80	49	4.4	4.1	0.2	0.49
	CAAN5342.A2	18	>330	24	>330	80	122	5.8	34	9.5	14.6
	WITO4160.33	2.0	7.3	4.0	21	0.33	0.095	0.41	1.9	0.10	0.12
	AC10.2.29	2.0	>330	8.7	13	0.53	0.22	5.8	5.0	0.27	0.79
	SC422661.8	6.0	14	4.7	1.3	20	120	0.62	4.2	0.24	0.42
	6535.3	1.3	13	13	9.3	4.1	195	7.1	18	1.4	2.6
	THRO4156.18	2.0	>330	>330	3.3	273	51	36	7.8	0.59	0.74
	REJO4541.67	4.7	>330	4.0	4.7	0.13	NT	0.21	11	0.50	0.68
	TRJO4551.58	30	>330	>330	>330	3.9	11	0.62	4.5	0.13	0.35
	QH0692.42	9.3	19	6.7	2.0	>665	>2,900	7.5	14	1.0	2.4
	TRO.11	2.0	2.7	>330	>330	239	6.8	2.4	9.6	0.53	0.85
	RHPA4259.7	46	>330	80	0.67	180	3.0	0.27	12	1.1	2.0
C‡	MW965.26	NT	NT	NT	NT	NT	NT	NT	7.1	0.40	0.89
	ZM197M.PB7	3.3	>330	82	130	6.7	8.1	4.3	4.3	0.41	0.49
	ZM249.PL1	14	>330	>330	21	NT	0.81	0.62	23	1.9	2.3
	ZM53M.PB12	47	>330	>330	170	0.40	0.025	4.6	19	1.4	3.4
	ZM214M.PL15	27	>330	>330	20	1530	>2,900	1.7	29	1.5	2.6
	Du156.12	1.3	>330	>330	5.3	0.33	0.16	0.62	24	2.0	3.8
	Du442.1	4.7	>330	>330	1.3	2.9	0.95	>68	5.0	0.28	0.48
	Du172.17	2.0	>330	>330	6.7	3.3	0.41	>68	3.3	0.33	0.38
	CAP45.2.00.G3	17	>330	>330	4.7	0.020	0.027	44	1.2	0.17	0.41
	CAP210.2.00.E8	8.0	>330	>330	140	NT	0.47	>48	17	1.4	1.0
	ZM233M.PB6	8.0	>330	>330	>330	0.23	0.0034	5.3	4.6	0.24	0.29
	ZM109F.PB4	4.0	>330	>330	>330	1.0	88	1.5	18	2.8	5.5
	ZM135M.PL10a	4.0	>330	>330	>330	NT	>2,900	4.9	17	1.9	3.1
	Geometric Mean	11	110	76	34	3.7	6.2	3.4	10	0.74	1.3
	Arithmetic Mean	29	240	180	150	124	347	12	18	1.6	3.3

NT, not tested.

*Clade A 4E10, 2G12, 2F5, and b12 data from West et al., 2010 (1).

†Clade B 4E10, 2G12, 2F5, and b12 data from Li et al., 2005 (2).

‡Clade C 4E10, 2G12, 2F5, and b12 data from Li et al., 2006 (3).

1. West AP, Jr., et al. (2010) Evaluation of CD4-CD4i antibody architectures yields potent, broadly cross-reactive antihuman immunodeficiency virus reagents. *J Virol* 84:261–269.
2. Li M, et al. (2005) Human immunodeficiency virus type 1 *env* clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 79:10108–10125.
3. Li M, et al. (2006) Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular *env* clones from acute and early heterosexually acquired infections in Southern Africa. *J Virol* 80:11776–11790.

Table S4. Data collection and refinement statistics for CVN_2 crystal structures

	CVN_2L0	CVN_2L10
Data collection		
Space group	$P3_221$	$P3_221$
Cell dimensions		
a, b, c, Å	47.9, 47.9, 78.7	48.0, 48.0, 79.3
α, β, γ, °	90, 90, 120	90, 90, 120
Resolution, Å	2.0 (2.11–2.0)	1.75 (1.84–1.75)
No. reflections	42,309	59,304
Unique reflections	7,456	11,158
R_{merge}, %	5.1 (22.8)	10.3 (38.4)
$I/\sigma I$	23.6 (5.6)	12.7 (4.2)
Completeness (%)	100.0 (100.1)	100.0 (100.1)
Redundancy	5.7 (5.7)	5.3 (5.2)
Refinement		
Resolution, Å	23.9–2.0	36.8–1.75
No. reflections		
Working set	6,687	10,060
Test set	341	528
R_{work}/R_{free}, %	22.1/24.9	18.8/21.2
No. atoms		
Protein	796	823
Solvent	60	89
B factors,Å^2		
Protein	31.2	19.3
Water	34.0	28.6
r.m.s. deviations		
Bond lengths, Å	0.012	0.011
Bond angles, °	1.375	1.351
Ramachandran plot		
Favored, %	89	89.2
Allowed, %	11	10.8
Generously allowed, %	0	0
Disallowed, %	0	0

Table S5. CVN_2L0 binding site knockout variant mutations

Variant	Mutations in 1st CV-N repeat	Mutations in 2nd CV-N repeat
$CVN_2L0_{\Delta Amm}$	K3N, T7A, E23I, N93A	
$CVN_2L0_{\Delta A}$	K3N, T7A, E23I	N93A
$CVN_2L0_{\Delta Bmm}$	E41A, N42A, T57A, R76A, Q78G	
$CVN_2L0_{\Delta B}$	E41A, N42A	T57A, R76A, Q78G
Control A-1	K3N, T7A, E23I	
Control A-2		N93A
Control B-1	E41A, N42A	
Control B-2		T57A, R76A, Q78G
$CVN_2L0_{\Delta A\|B}$	K3N, T7A, E23I, E41A, N42A	T57A, R76A, Q78G, N93A
$CVN_2L0_{\Delta A\times B}$	E41A, N42A, N93A	K3N, T7A, E23I, T57A, R76A, Q78G
$CVN_2L0_{\Delta AA}$	K3N, T7A, E23I, N93A	K3N, T7A, E23I, N93A
$CVN_2L0_{\Delta BB}$	E41A, N42A, T57A, R76A, Q78G	E41A, N42A, T57A, R76A, Q78G
$CVN_2L0_{\Delta ABB}$	E41A, N42A, T57A, R76A, Q78G, N93A	K3N, T7A, E23I, E41A, N42A, T57A, R76A, Q78G
$CVN_2L0_{\Delta AAB}$	K3N, T7A, E23I, E41A, N42A, N93A	K3N, T7A, E23I, T57A, R76A, Q79G, N93A