Chapter 3

Structural characterization of engineered cyanovirin-N dimers

Abstract

Cyanovirin-N (CVN) is an antiviral protein that has broad specificity for enveloped viruses. It specifically binds carbohydrates expressed on envelope proteins and prevents the virus-host cell interactions that are necessary for infection. In the previous chapter, we described dimeric CVN variants (CVN_2s) that contain a tandem repeat of CVN linked through a flexible polypeptide linker. These variants showed up to 10-fold greater efficacy against HIV in a cell-based HIV neutralization assay and also showed broad specificity for different strains and clades of HIV. To investigate the mechanism for this increase in biological function, we have solved four crystal structures of three of these dimeric variants. The structures show that CVN_2 L0, CVN_2 L1, and CVN_2 L10 are all intramolecularly domain-swapped and have a great degree of similarity to wild-type (WT) CVN domain-swapped structures. Although we were able to see some density for the linker in each case, the structures were complicated by the fact that only half of the CVN_2 dimer was in the asymmetric unit in three out of four cases. In the fourth case, a CVN_2 L1 structure, we could clearly distinguish the free termini from the linked termini. Unfortunately, there were no major differences between the linked dimers and WT CVN, and therefore major structural changes do not contribute to the increase in HIV neutralization. We can, however, conclude that linking the termini of two CVN molecules does not adversely affect the structure or function and that the molecules are intramolecularly domain-swapped. Additionally, in the domain-swapped form, the distance between carbohydrate binding sites in CVN become more consistent and these distances are similar to the distance between the binding sites in the broadly neutralizing

anti-HIV antibody, 2G12. By stabilizing this domain-swapped form, we may be positioning the binding sites at the ideal geometric position to best neutralize HIV.

Introduction

Cyanovirin-N (CVN) is a lectin originally isolated from the cyanobacterium *Nostoc ellipsosporum* that has been shown to effectively neutralize various enveloped viruses, including HIV,[1] influenza,[2] and Ebola.[3] CVN specifically binds α1-2 linked oligomannose glycosylation[4] sites on the envelope proteins of these viruses and blocks critical interactions between the virus and the host cell, thus preventing infection.[5]

CVN exists in solution both as a monomer and a domain-swapped dimer (Figure 3-1). The monomer consists of two pseudo-domains that display high sequence homology. The first domain contains residues 1-39 and 90-101, and the second domain contains residues 39-89.[6] CVN includes a three-stranded antiparallel β sheet and a β hairpin in each pseudo-domain. The pseudo-domains are connected through two helical turns. CVN also contains two native disulfide bridges: between residues 8 and 22, and between residues 58 and 73. These two disulfide bridges are located near each end of the molecule and anchor the first strand of the β sheet to the second strand. The dimer contains the same topology, but is domain-swapped at residues 51-53.[7] In this case, the first domain of one chain (A) forms a "monomer-like" structure with the second domain of the other chain (B') in an almost symmetric domain swapping (Figure 3-1B). The two quasi-monomers can sample different orientations relative to each other due to the flexibility of the domain-swap region, and the orientation appears to be pH dependent in crystal structures.[8]

CVN contains two carbohydrate-binding sites of differing affinities for α1-2 mannose: one at each end of the molecule (Figure 3-1A). The high affinity site, located distal from the N- and C-termini, has a dissociation constant in the low nanomolar range,

whereas the lower affinity site has approximately 10-fold weaker affinity.[9] To date, no crystal structures have been solved of CVN with carbohydrate bound to the high affinity site. This is probably due to crystallographic packing, which obstructs this binding site. However, numerous structures have been solved with carbohydrate bound to the low affinity sites, and these structures are very similar to the NMR structures.[10,11]

In solution, CVN exists mainly as a monomer, and NMR structures have been solved of monomeric CVN, both free[6] and bound to carbohydrates.[4] However, all crystal structures of the wild-type (WT) protein to date have yielded domain-swapped structures, a proposed artifact of the crystallographic process.[7,10,12,13] The domain-swapped form is metastable in solution and rapidly coverts to the more stable monomer at physiological temperatures, but is stable for long periods of time at low temperature.[12] However, in the crystallographic conditions of high protein concentration, extreme pH, and high precipitant concentration, the equilibrium is shifted from the purified monomer to the crystallized dimer. Solution structures of isolated dimer are similar to those solved using crystallography.[12] Various constructs have been engineered to modulate the domain-swapping, including variants that preferentially form monomers[12] and those that form dimers in solution.[12,14,15] One of these variants, a five-fold mutant including the P51G mutation, which stabilizes the monomeric state, was solved recently as a monomer using crystallography.[11]

Although a great deal of work has been done to change the preference for monomeric or dimeric protein, there is still a controversy about the effect of dimerization on the antiviral activity of CVN. Because WT domain-swapped dimer converts to monomer during the course of a viral neutralization assay, it is difficult to assay the effect

of dimerization directly. Therefore, various mutants have been generated to try to elucidate the relationship between oligomerization and activity. Kelley *et al*. created an obligate domain-swapped dimer by deleting one of the residues in the domain-swap region and observed a 3.5-fold reduction in the concentration at which half of the viral particles are neutralized (IC_{50}) of HIV fusion. They also showed a similar 3.5-fold reduction in the IC_{50} for purified WT dimer.[14] However, Barrientos *et al*. tested WT monomer, WT dimer, and various engineered mutants and found that regardless of the oligomerization state, all molecules had essentially the same activity against both HIV and Ebola Zaire.[15] Differences in the incubation time and assay conditions could explain the discrepancy, but the question still remains whether dimeric CVN is more effective at neutralizing viruses than monomeric CVN.

In Chapter 2, I described the creation and characterization of obligate dimeric CVN molecules (CVN_2s). These proteins consist of two copies of WT CVN linked with a flexible linker of varying length. We tested linker lengths ranging from 0 amino acids (L0) to 20 amino acids (L20) and found that some of the variants displayed up to 10-fold better HIV neutralization than WT. Specifically, CVN_2 L0, CVN_2 L5, and CVN_2 L10 were significantly more effective than WT. By adding only one amino acid between the N- and C-termini, however, the CVN_2 L1 variant displayed only approximately five-fold better neutralization than WT. While this difference could easily be explained by experimental error in the biological assay, we wanted to investigate the mechanism for the increase in efficacy for CVN_2 L0, CVN_2 L1, and CVN_2 L10 and determine whether there was a structural explanation for the increase in activity. We therefore solved crystal structures of these three variants. We hypothesize that differences in the structures of the

proteins, including potential domain-swapping differences, could account for the changes in activity.

Methods

Protein expression and purification. CVN_2 L0, CVN_2 L1, and CVN_2 L10 were expressed and purified as described in Chapter 2. After gel filtration, the proteins were concentrated using 5,000 MWCO Amicon Ultra concentrators (Millipore) to 25-30 mg/mL.

Crystallization. Crystallization conditions were set up using a Mosquito automated nanoliter pipettor (TTP Labtech) in the Molecular Observatory at Caltech. 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. The best diffracting CVN_2 L0 crystals were grown in 0.1 M sodium HEPES pH 7.5, 0.8 M potassium dihydrogen phosphate, 0.8 M sodium dihydrogen phosphate. CVN_2 L1 data sets were collected on crystals from 0.1 M phosphate-citrate pH 4.2, 2.0 M sodium/potassium phosphate ($P3_221$ structure) and from 0.1 M CHES pH 9.5, 0.2 M lithium sulfate and 1 M potassium/sodium tartrate ($P4_12_12$ structure). The CVN_2 L10 data set was collected on a crystal grown in 0.2 M sodium fluoride and 20% PEG-3350.

Data collection and refinement. All crystals except the CHES CVN_2 L1 crystal were cryoprotected in TMP oil. CVN_2 L1 crystals grown in the CHES/tartrate condition

were cryoprotected using the reservoir condition including 20% glycerol. Data for the CVN_2 L0 and L1 structures were collected using a MicroMax-007HF X-ray generator with an RAXIS IV++ detector (Rigaku Corp.). The CVN_2 L10 data set 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.[16] The indexed and scaled data were further evaluated using CCP4i.[17] The molecular replacement for data sets indexed to the $P3_221$ space group were done using 3EZM as the starting model.[7] The molecular replacement of the CVN_2 L1 data in the $P4_12_12$ space group was done using 2Z21 as the starting model.[11] Phaser version 1.3.3 was used for the molecular replacement.[18] Further refinement was done using Coot[19] and Refmac[20] and omit maps were created using CNS.[21,22] Figures were made using PyMOL.[23]

Results

Crystallization. CVN_2 L0, CVN_2 L1, and CVN_2 L10 were chosen for structural characterization. As described in Chapter 2, CVN_2 L0 and CVN_2 L10 were the most active of the engineered dimers, while CVN_2 L1 was less active than CVN_2 L0 due to the single amino acid linker. We therefore solved crystal structures of these proteins to determine whether any major structural differences could explain the changes in the HIV neutralization activity that we observed. All of the proteins were crystallized in 96-well plates with 0.6 μL drops. CVN_2 L0 crystallized in approximately 20 out of the 480 conditions tested. Most of these conditions contained sulfate or phosphate as the precipitant and low pH buffers, although the protein also crystallized well in 20% PEG 3350. CVN_2 L10 crystallized in similar conditions to CVN_2 L0, and crystals were

observed in approximately 35 conditions. CVN_2 L1 crystallized in approximately 35 conditions as well, but in addition to crystal forms seen for CVN_2 L0 or CVN_2 L10, new crystal forms were observed in high pH conditions. Approximately half of the crystal conditions for CVN_2 L0 were above neutral pH, while the CVN_2 L10 conditions were only about one-third above neutral pH. The structure for CVN_2 L1 was therefore determined from both a low pH condition ($P3_221$ space group) and a high pH condition ($P4_12_12$ space group). CVN_2 L0 and CVN_2 L10 structures were solved using the only well-diffracting crystals available, which were from low pH conditions ($P3_221$ space group).

Crystal structure refinement. All of the $P3_221$ space group structures were solved using the domain-swapped WT structure 3EZM.pdb[7] as the model for molecular replacement (Table 3-1). This WT structure was solved from the same space group, and the molecular replacement provided a good initial model. Further refinement on all structures yielded domain-swapped models that fit the density well. Omit maps were calculated for each of the structures, which agreed well with the models. The omit maps did not indicate any major differences for either the backbone or the side chains of the structure. Solvent molecules, including waters and sodium ions, were added to each structure when there were appropriate electron density and hydrogen bonding partners.

The structures solved from $P3_221$ space group crystals contained only half of the CVN_2 dimer in the asymmetric unit, and the second tandem repeat of CVN was generated through crystallographic symmetry. However, because the two copies of CVN are covalently linked through a flexible polypeptide linker, this caused difficulty in the

refinement. In order to properly model the termini as well as the linkage, the linker residues are at 50% occupancy. This is because half of the proteins are oriented in a way that the free termini are in a specific location, while the other half of the proteins have the linkage in that same location. The protein crystallized in both orientations with approximately the same frequency, resulting in symmetry with 50% occupancy of the termini. In the case of CVN_2 L0, as described below, we have not yet been able to fully resolve the linkage. We suspect that the free termini and the linked termini have significantly different conformations that are difficult to distinguish in the density.

In addition to the low pH structure, the CVN_2 L1 structure at high pH ($P4_12_12$ space group) was also determined (Table 3-1). A molecular replacement with 3EZM.pdb was suboptimal, giving a solution and electron density map that did not correlate. However, replacement with a monomeric five-fold variant of WT CVN (2Z21.pdb[11]) gave a model and map that were reasonable. Upon inspection of the map, it was clear that the structure was in fact domain-swapped, similarly to the $P3_221$ structures. The structure model was modified to reflect the domain-swapping of the electron density, and solvent molecules were added.

To confirm that the crystallized protein in each case was in fact CVN_2 and not contamination from WT CVN, we ran an SDS-PAGE gel on crystals grown in the same conditions as those the data sets were collected on. The crystals were rinsed to remove any non-crystallized protein before being denatured. The gel showed that all of the crystals were indeed CVN_2 with no WT contamination (data not shown), allowing us to conclusively determine that indeed only half of the molecule was present in the

asymmetric unit in the $P3_221$ space group cases and that it was reasonable to model in the linker at 50% occupancy.

CVN_2 L0 structure. CVN_2 L0 is a domain-swapped dimer under low pH crystallographic conditions (Figure 3-2). Its structure is remarkably similar to WT CVN, with an RMSD of 0.239 Å. Although the refinement is not complete, we can state with certainty that there are no major disruptions of the structure by directly linking two CVNs together without a linker. There are, however, minor differences in the domain-swapped area compared to WT CVN.

Because only one half of the CVN_2 L0 was in the asymmetric unit, as described above, the electron density at the termini is a composite from both the free and the linked termini. A view of the electron density fit to a free N- and C-terminus shows positive electron density between them (Figure 3-3A), indicating the model does not fit well in this area. We hypothesize that the free termini are in a significantly different conformation from the termini that are linked and that the electron density is a combination of these. Future rounds of refinement will be done to model both conformations separately at 50% occupancy in order to fit the experimental data.

We can clearly see from the electron density that CVN_2 L0 is a domain-swapped dimer under these conditions (Figure 3-3B). Molecular replacement with a monomeric WT CVN resulted in density that also showed clear domain-swapping, indicating that model bias is not responsible for this density.

CVN_2 L1 structures. CVN_2 L1 crystallized with two major morphologies. Because of the difference in the shape of the crystals and because they were indexed to different space groups, we solved two structures of CVN_2 L1: one from the $P3_221$ space group and the other from the $P4_12_12$ space group. Upon molecular replacement, it became clear that both structures were WT-like domain-swapped structures with slightly different orientations of the domains relative to each other (Figure 3-4). It has been shown previously that WT CVN crystallizes in different space groups and different morphologies depending on the pH of the crystallization condition.[8] This appears to be the case here as well.

The $P3_221$ structure, solved from a low pH condition, is very similar to the WT low pH structure, with an RMSD of 0.283 Å. Like the other $P3_221$ structures, there are minor deviations in the domain-swap area, but overall there are no major structural perturbations by linking two termini. The $P4_12_12$ also does not appear to be significantly different from WT CVN when compared to a structure solved at high pH (Figure 3-4B).[12] The RMSD between WT CVN and the $P4_12_12$ is 0.407 Å. In this case as well, there are no major structural changes to CVN_2 L1 that would explain the vast increase in biological activity.

Clear density was visible for a single glycine linker between the termini of the $P3_221$ structure (Figure 3-5A). However, due to the fact that only half of the molecule is in the asymmetric unit, we have modeled the linker in at 50% occupancy. We believe that the free termini have significantly different conformations from the termini that are linked. We have not yet modeled the free termini and trust that this will increase the

reliability of the model. Forty-five water molecules and two sodium ions were placed in the structure with high confidence.

In the $P4_12_12$ structure, 104 water molecules were added. There was also a CHES molecule near the high affinity carbohydrate binding site of one CVN domain, which broke the symmetry of the molecule, causing the entire CVN_2 L1 to be in the asymmetric unit, unlike the $P3_221$ structures. The free termini and the linked termini were clearly distinguishable in the initial electron density map, and therefore, we were able to model them separately (Figure 3-6). In addition to different conformations for the linked residues, there was also clear density for the single glycine in the linker.

In both crystal forms, CVN_2 L1 forms a domain-swapped dimer similar to WT CVN (Figure 3-5B, Figure 3-6C). The conformations of residues involved in the swap in the $P4_12_12$ structure are almost identical to the WT structure, but there are small differences in the conformations from the $P3_221$ structure compared to WT. While these differences may be real, we do not expect that they explain the significant increase in activity of CVN_2 L1 over WT CVN in the neutralization assay, as described in Chapter 2.

Both CVN_2 L1 structures have unexpectedly high R_{free} values (Table 3-1). Omit maps on each structure indicate that there were no major problems with either the backbone or the side chains. Because the termini and the linkage are somewhat unstructured and the model is not perfectly matched to the electron density in this area, we may be getting some model bias. However, we do not expect these small deviations to have a significant impact on the overall structures. The addition of more water and solvent molecules may also decrease the R_{free} and give more reasonable statistics.

CVN_2 L10 structure. The crystal structure of CVN_2 L10 indicates that this protein is also very similar to WT CVN (Figure 3-7). The two structures have an RMSD of 0.353 Å and the only significant differences between the two structures are in the domain-swap area and some of the backbone phi and psi angles of the β strands. The CVN_2 L10 structure forms more optimal β strands in many cases than previously solved WT structures.

Like all of the other $P3_221$ structures, only half of the CVN_2 L10 molecule is in the asymmetric unit. However, because the 10 amino acid linker is long and flexible, the N- and C-termini appear to be identical when free and when linked. We were therefore able to fit four of the linker residues (two glycines on each terminus) at 50% occupancy while leaving Leu1 and Glu101 as 100% occupied (Figure 3-8A). There was no clear density for the six residues in the center of the linker, so they are not included in the structure.

CVN_2 L10 is also definitively domain-swapped and contains several structural water molecules to stabilize this conformation (Figure 3-8B). The CVN_2 L10 data set was molecularly replaced with a monomeric CVN model, and domain-swapping density was clear, indicating in this case, as in the previous cases, the domain swapping is not an artifact of the replacement and refinement process.

Conclusions

We solved four crystal structures of three different CVN_2 variants with linkers containing zero, one, and ten amino acids. We had hoped that significant structural changes or differences in the domain-swapping of the variants would partially explain the

increase in biological activity described in Chapter 2. However, the structures are all remarkably similar to domain-swapped WT CVN crystal structures. All four structures are intramolecularly domain-swapped and show varying amounts of density for the flexible polypeptide linker. The RMSDs for the structures as compared to WT were all less than 0.5 Å and the minor differences were typically observed in the domain-swap region.

Complicating the structure refinement was the fact that three of the four structures contained only half of the CVN_2 molecule in the asymmetric unit. In these cases, the molecule could align in two possible orientations: one with the free termini in a given location and another with the linked termini in the same location. This led to an additional plane of symmetry where the two halves of the CVN_2 were generated by crystallographic symmetry and the free and linked termini were each represented by the same density at 50% occupancy each.

The crystal structures of three CVN_2 molecules showed no major differences from WT CVN. Because the structures are remarkably similar, we can be confident that linking two repeats of CVN does not negatively affect the structure and does not cause any major perturbations. In addition, the linkage stabilizes the domain-swapped form over the monomeric form due to the steric restraints provided by a short linker and the increase in local concentration and therefore the proteins form obligate domain-swapped dimers. All of the structures are intramolecularly domain-swapped, and we feel confident that this is the biologically relevant conformation in solution. If the crystals had contained intermolecularly domain-swapped CVN_2 protein, we could have inferred that the molecule in solution was monomeric-like. We know by gel filtration that the protein is

monomeric or intramolecularly domain-swapped because we can effectively separate intermolecularly domain-swapped CVN_2. Although we are confident that the proteins are indeed domain-swapped in solution, additional studies are necessary to confirm this finding. We hope that NMR experiments will help determine the solution state of the molecules. In addition, we are looking into cleavage assays in which the protein is cleaved at the linker and then run on a gel filtration column. If the protein is indeed forming a domain-swapped dimer in solution, as expected, we would see dimer elute from a gel filtration column. If, on the other hand, the protein contains two independent monomer-like domains, we would expect that after cleavage, each would elute as a monomer on a gel filtration column.

By stabilizing the domain-swapped dimer, we believe we are rigidifying the distances between the four carbohydrate binding sites in CVN_2. Although the two domains may sample various conformations with respect to each other, as demonstrated by the differences between the $P3_221$ and $P4_12_12$ structures, the carbohydrate binding sites in both cases are brought together and held in close proximity in the domain-swapped form. For example, in the domain-swapped dimer crystal structure of WT CVN in the $P4_12_12$ space group, each of the four carbohydrate binding sites is approximately 30 to 40 Å from the other sites (Figure 3-9). This geometry may be ideal for interacting with gp120 glycosylation, and by stabilizing this form we may be increasing the affinity of interaction. Additional support for this mechanism comes from the crystal structure of 2G12, a broadly neutralizing anti-HIV antibody that is also specific to the glycosylation on gp120.[24] Unlike the standard "Y" shaped antibodies, 2G12 contains a domain swap in the Fab region, which brings the two carbohydrate binding sites approximately 35 Å

apart, a similar distance to the distances in domain-swapped CVN (Figure 3-9B).[25] Instead of being highly flexible, the antigen binding region of 2G12 is fixed to enhance the interaction with gp120. The similarity in the spacing between binding sites in domain-swapped CVN and 2G12 indicates that 30 to 40 Å spacing may be biologically relevant for gp120 glycosylation binding and that efficacy may be increased by stabilizing the domain-swapped form of CVN.

Acknowledgements

I would like to thank Len Thomas for help with data collection and refinement and Pavle Nikolovski and the Molecular Observatory for setting up crystal trays. I would also like to thank Professors Pamela Bjorkman, Doug Rees, and Bil Clemons for helpful discussions regarding structure refinement. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

References

1. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., 2nd, Buckheit, R. W., Jr., Nara, P. L., Pannell, L. K., Sowder, R. C., 2nd & Henderson, L. E. (1997). Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* **41**, 1521-30.

2. O'Keefe, B. R., Smee, D. F., Turpin, J. A., Saucedo, C. J., Gustafson, K. R., Mori, T., Blakeslee, D., Buckheit, R. & Boyd, M. R. (2003). Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* **47**, 2518-25.

3. Barrientos, L. G., O'Keefe, B. R., Bray, M., Sanchez, A., Gronenborn, A. M. & Boyd, M. R. (2003). Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* **58**, 47-56.

4. Bewley, C. A. (2001). Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* **9**, 931-40.

5. Mori, T. & Boyd, M. R. (2001). Cyanovirin-N, a potent human immunodeficiency virus-inactivating protein, blocks both CD4-dependent and CD4-independent binding of soluble gp120 (sgp120) to target cells, inhibits sCD4-induced binding of sgp120 to cell-associated CXCR4, and dissociates bound sgp120 from target cells. *Antimicrob Agents Chemother* **45**, 664-72.

6. Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M. & Gronenborn, A. M. (1998). Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat Struct Biol* **5**, 571-8.

7. Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., Clore, G. M. & Wlodawer, A. (1999). Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol* **288**, 403-12.

8. Botos, I. & Wlodawer, A. (2003). Cyanovirin-N: a sugar-binding antiviral protein with a new twist. *Cell Mol Life Sci* **60**, 277-87.

9. Bewley, C. A. & Otero-Quintero, S. (2001). The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *J Am Chem Soc* **123**, 3892-902.

10. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R. & Wlodawer, A. (2002). Structures of the complexes of a

potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* **277**, 34336-42.

11. Fromme, R., Katiliene, Z., Giomarelli, B., Bogani, F., Mc Mahon, J., Mori, T., Fromme, P. & Ghirlanda, G. (2007). A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry* **46**, 9199-207.

12. Barrientos, L. G., Louis, J. M., Botos, I., Mori, T., Han, Z., O'Keefe, B. R., Boyd, M. R., Wlodawer, A. & Gronenborn, A. M. (2002). The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **10**, 673-86.

13. Botos, I., Mori, T., Cartner, L. K., Boyd, M. R. & Wlodawer, A. (2002). Domain-swapped structure of a mutant of cyanovirin-N. *Biochem Biophys Res Commun* **294**, 184-90.

14. Kelley, B. S., Chang, L. C. & Bewley, C. A. (2002). Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. *J Am Chem Soc* **124**, 3210-1.

15. Barrientos, L. G., Lasala, F., Delgado, R., Sanchez, A. & Gronenborn, A. M. (2004). Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity. *Structure* **12**, 1799-807.

16. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Prot. Crystallography* **26**.

17. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3.

18. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**, 458-64.

19. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.

20. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55.

21. Brunger, A. T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat Protoc* **2**, 2728-33.

22. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R.

J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21.

23. DeLano, W. L. (2008). The PyMOL Molecular Graphics System. http://www.pymol.org, DeLano Scientific, Palo Alto, CA.

24. Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J. P. & Katinger, H. (1996). Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* **70**, 1100-8.

25. Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M. R., Stanfield, R. L., Roux, K. H., Kelly, J. W., Rudd, P. M., Dwek, R. A., Katinger, H., Burton, D. R. & Wilson, I. A. (2003). Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**, 2065-71.

Table 3-1. Crystallographic statistics.

	CVN_2 L0	CVN_2 L1	CVN_2 L1	CVN_2 L10
Data collection				
Space group	$P3_221$	$P3_221$	$P4_12_12$	$P3_221$
Cell dimensions				
a,b,c (Å)	47.9, 47.9, 78.7	47.5, 47.5, 78.6	60.6, 60.6, 147.6	48.0, 48.0, 79.3
α,β,γ (deg)	90, 90, 120	90, 90, 120	90, 90, 90	90, 90, 120
Resolution (Å) *	2.0 (2.11 - 2.0)	2.0 (2.11 - 2.0)	2.1 (2.21 - 2.1)	1.75 (1.84 - 1.75)
No. reflections	42309	50188	206257	59304
Unique reflections	7456	7347	16878	11158
R_{merge} (%) *	5.1 (22.8)	3.7 (10.8)	4.3 (27.6)	10.3 (38.4)
$I/\sigma I$ *	23.6 (5.6)	33.3 (13.5)	42.9 (9.2)	12.7 (4.2)
Completeness (%) *	100.0 (100.1)	99.9 (99.9)	100.0 (100.0)	100.0 (100.1)
Redundancy *	5.7 (5.7)	6.8 (6.7)	12.2 (12.3)	5.3 (5.2)
Refinement				
Resolution (Å)	23.9 - 2.0	28.4 - 2.0	27.1 - 2.1	36.8 - 1.75
No. reflections				
Working set	6687	6642	15109	10060
Test set	341	340	851	528
R_{work}/R_{free}	22.0 / 24.9%	21.8 / 27.4%	22.0 / 28.8%	18.8 / 21.2%
No. atoms				
Protein	796	785	1567	823
Solvent	60	47	118	89
B-factors				
Protein	31.2	31.0	32.4	19.3
Water	34.0	34.5	34.9	28.6
R.m.s. deviations				
Bond lengths (Å)	0.012	0.011	0.011	0.011
Bond angles	1.375	1.307	1.215	1.351
Ramachandran plot				
Favored (%)	89	90.2	88.3	98.2
Allowed (%)	11	9.8	12	10.8
Generously allowed (%)	0	0	0	0
Disallowed (%)	0	0	0	0

* Last shell is shown in parentheses.

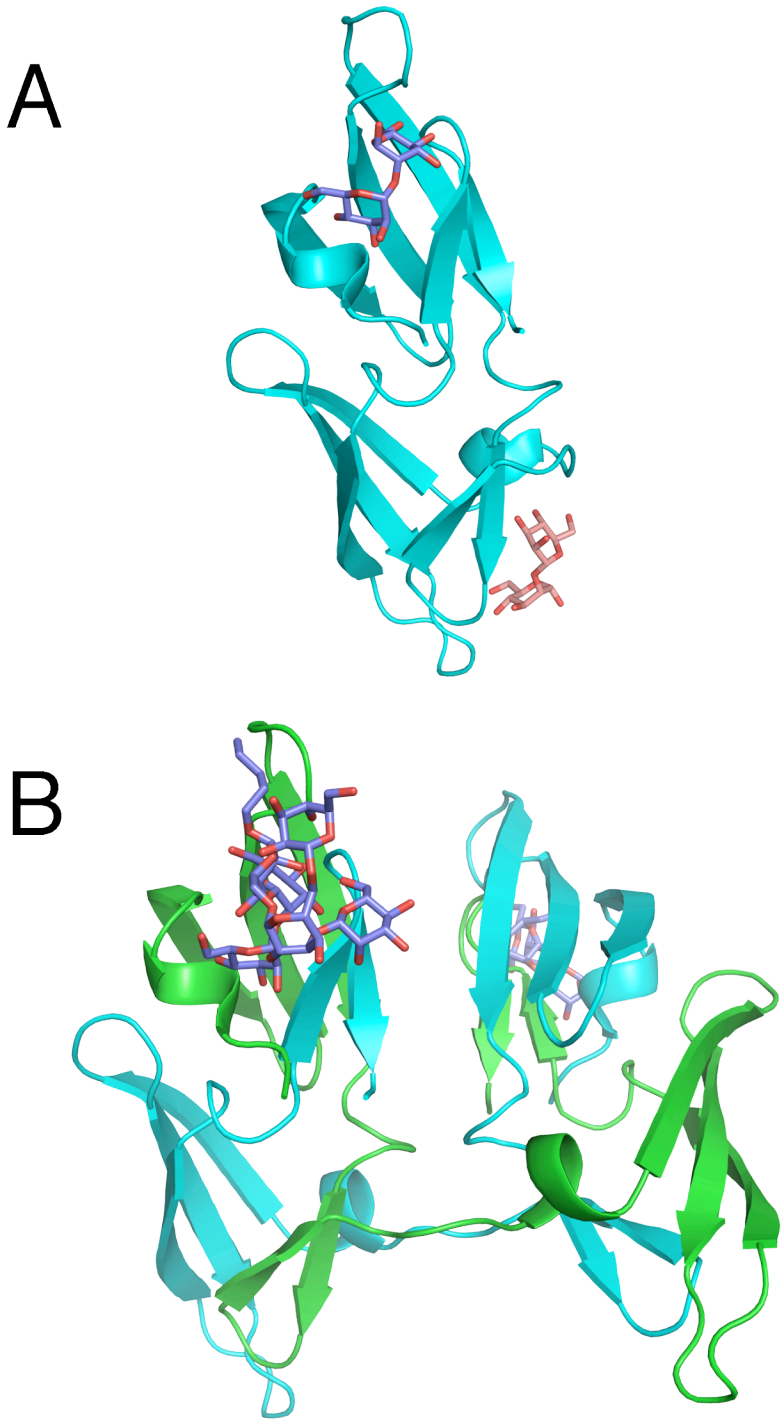

Figure 3-1. Wild-type CVN structures. In solution, CVN can exist as either a monomer (A)[4] or a domain-swapped dimer (B).[10] CVN is shown in green and cyan ribbons to indicate protein chains. Carbohydrates bound to the high affinity site are shown with orange carbons (present only in A) and carbohydrates bound to the low affinity site are shown with blue carbons (present in both A and B). The monomer and the left half of the dimer are in approximately the same orientation.

Figure 3-2. CVN_2 L0 crystal structure compared to WT CVN. The CVN_2 L0 crystal structure (green) and domain-swapped WT CVN (3EMZ.pdb[7]) (blue and cyan) have an RMSD of 0.239 Å. The structures are very similar with small deviations in the β strands and the domain swap area.

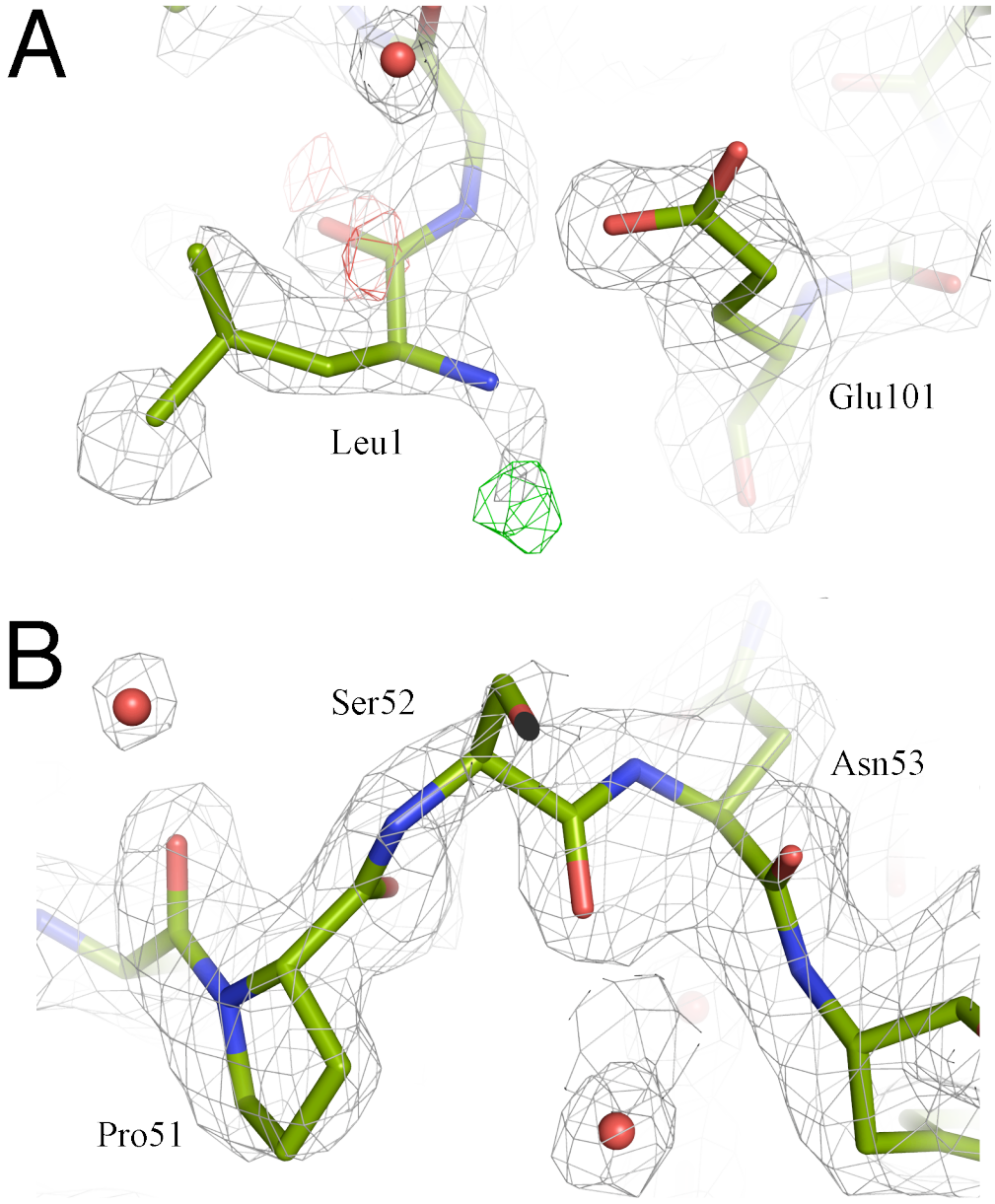

Figure 3-3. CVN_2 L0 structure. Only half of the CVN_2 dimer was found in the asymmetric unit. The other half is generated through crystallographic symmetry, and the linker and free termini are each 50% occupied. The 2Fo-Fc electron density map is shown as gray mesh. (A) The termini of CVN_2 L0 are not well defined. There is positive density where the linkage may occur. Further refinement is necessary to fit the model to the data. Additionally, the free N- and C-terminal residues probably have very different conformations from those that are linked. Positive density from the Fo-Fc map is shown in green, and negative density is shown in red. (B) CVN_2 L0 is domain-swapped in the crystal structure. The density in the swapped area is clear and definitive.

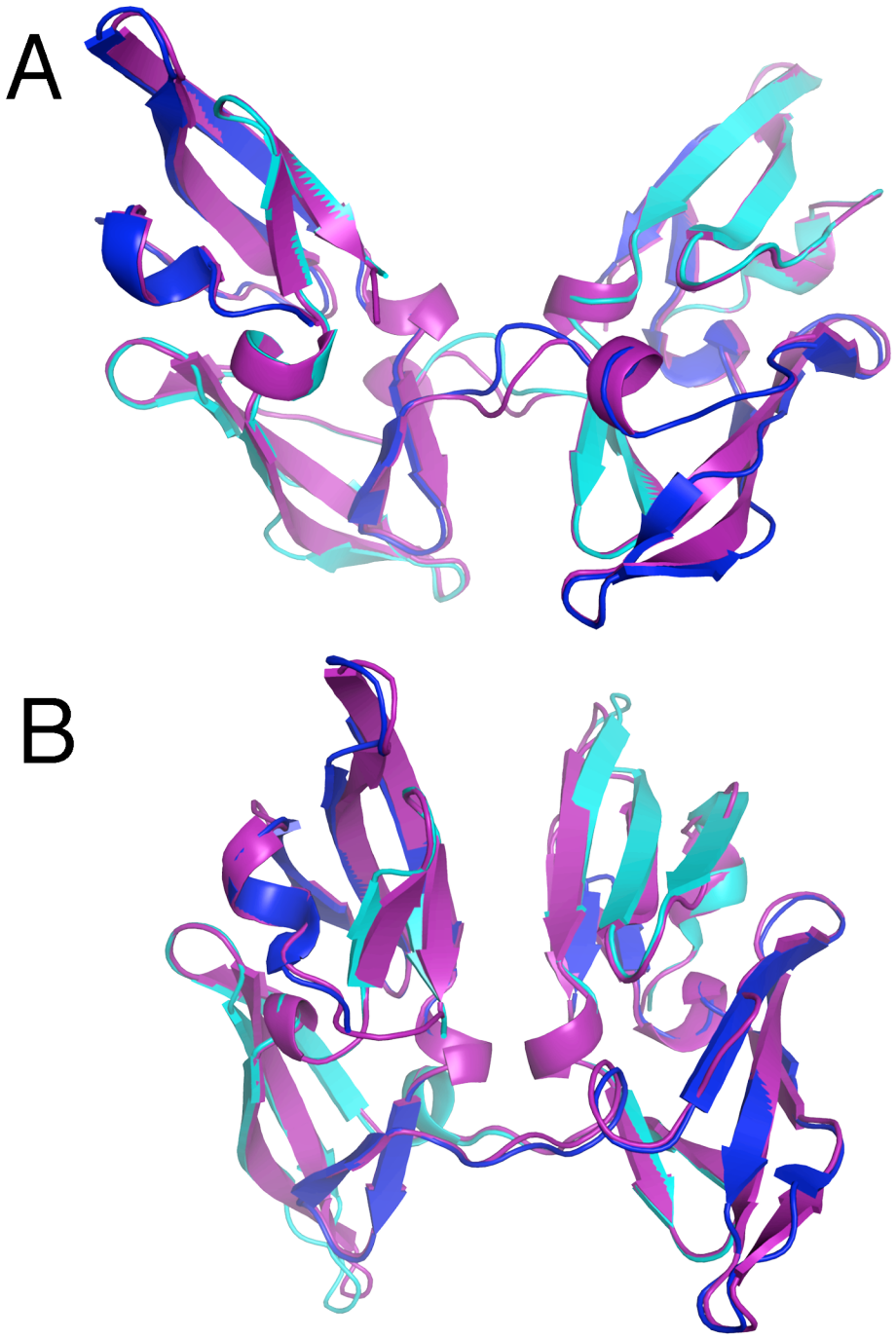

Figure 3-4. CVN_2 L1 crystal structures compared to WT CVN. CVN_2 L1 structures are shown in purple and the WT CVN structures are shown in blue and cyan. (A) CVN_2 L1 in the $P3_221$ space group, solved in low pH conditions, overlaid with domain-swapped WT CVN (3EZM.pdb).[7] The RMSD of these structures is 0.283 Å, and the major differences are seen in the domain swap area. (B) CVN_2 L1 in the $P4_12_12$ space group, solved in high pH conditions, overlaid with domain-swapped WT CVN (1L5B.pdb).[12] These structures have a 0.407 Å RMSD. The one residue linker can be seen in the P4 structure in the domain on the left.

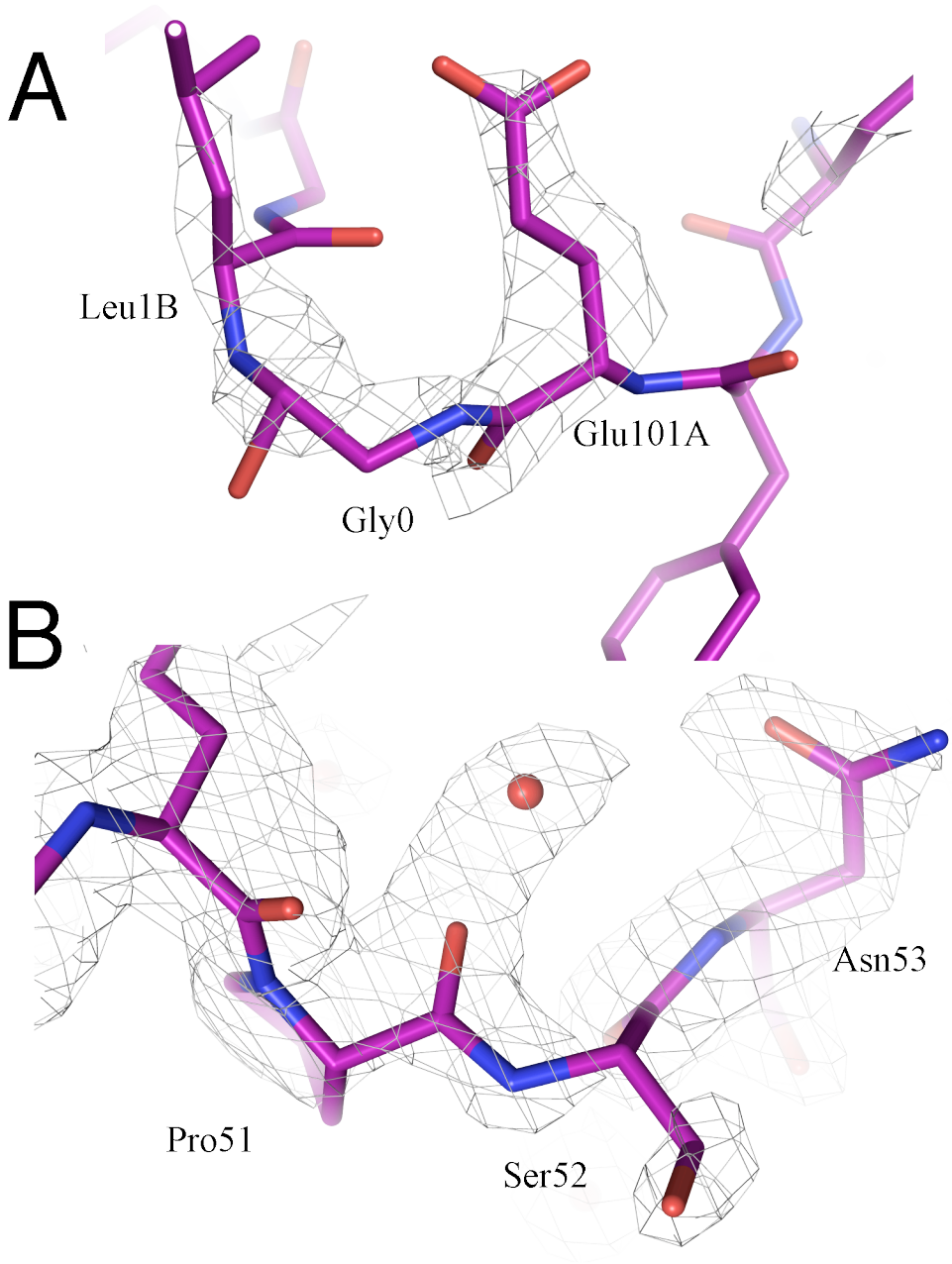

Figure 3-5. CVN_2 L1 $P3_221$ structure. Only half of the CVN_2 molecule was in the asymmetric unit. The free termini and the linked termini are both represented with the same density, at 50% occupancy each. (A) The N- and C-termini of CVN_2 L1 with a 2Fo-Fc electron density map contoured at 1 σ showing clear density for the single glycine residue linker. Gly0 shown in the figure is only 50% occupied. (B) The CVN_2 L1 in this crystal structure is clearly domain-swapped as evidenced by clear electron density in the domain swap region.

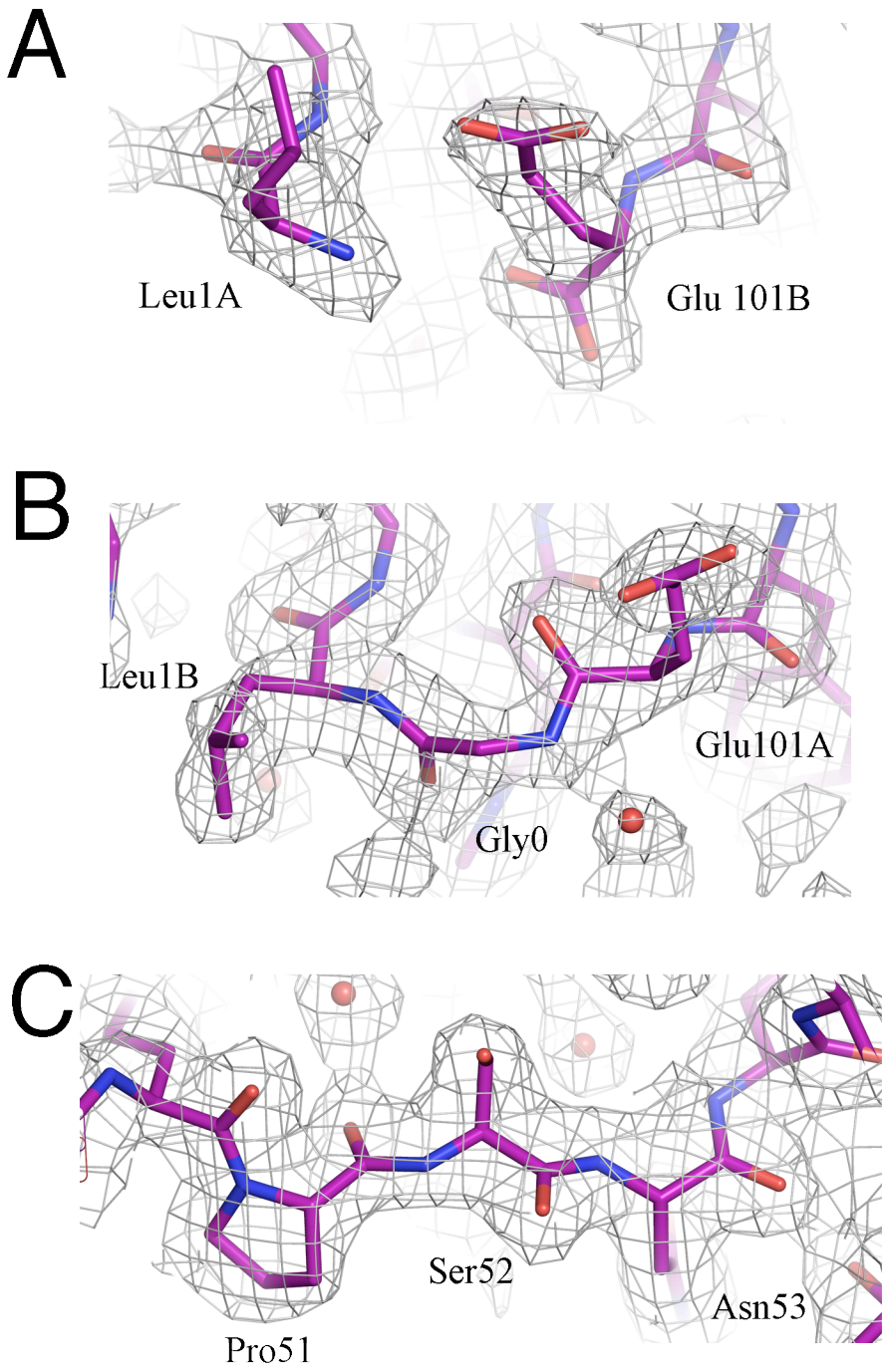

Figure 3-6. CVN_2 L1 $P4_12_12$ structure. The entire CVN_2 L1 dimer was within the asymmetric unit in the $P4_12_12$ crystal structure. The free termini were clearly distinguishable from the linkage as determined by the 2Fo-Fc electron density contoured to 1.0 σ. (A) The free N- and C-termini of CVN_2 L1. (B) The linked N- and C-termini of CVN_2 L1 and clear density for the single glycine linker. (C) Well defined density is seen in both chains for the domain swap residues 51 through 53.

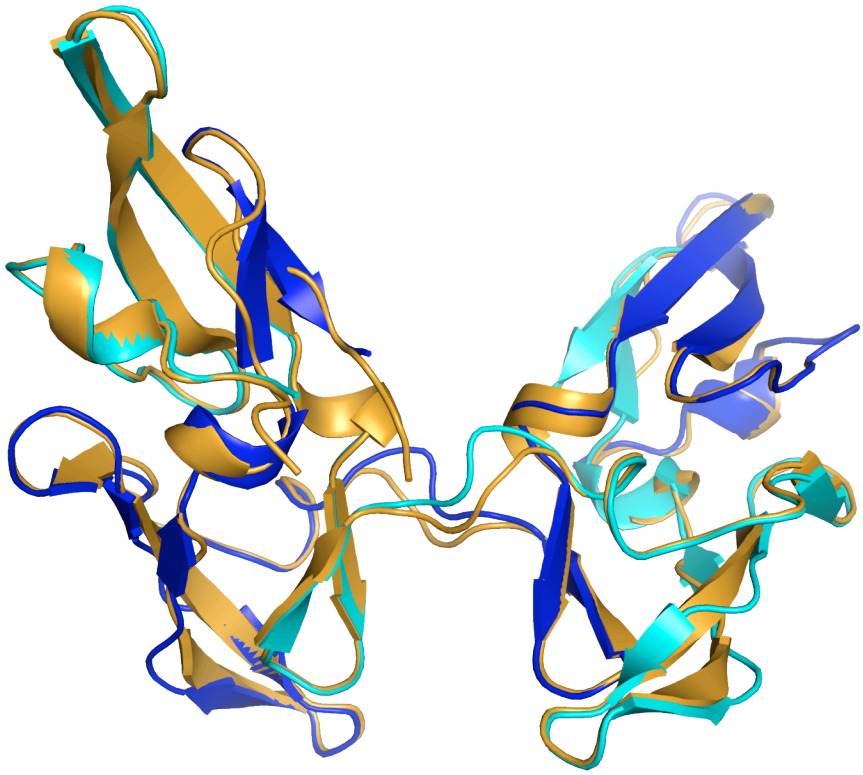

Figure 3-7. CVN_2 L10 structure compared to WT CVN. CVN_2 L10 is shown in orange and the domain-swapped WT CVN (3EZM.pdb)[7] is shown in blue and cyan. These structures have an RMSD of 0.353 Å. Four of the ten linker residues have electron density and are shown in the left domain of the structure. The other six residues are presumed to be disordered and are therefore not modeled in this structure.

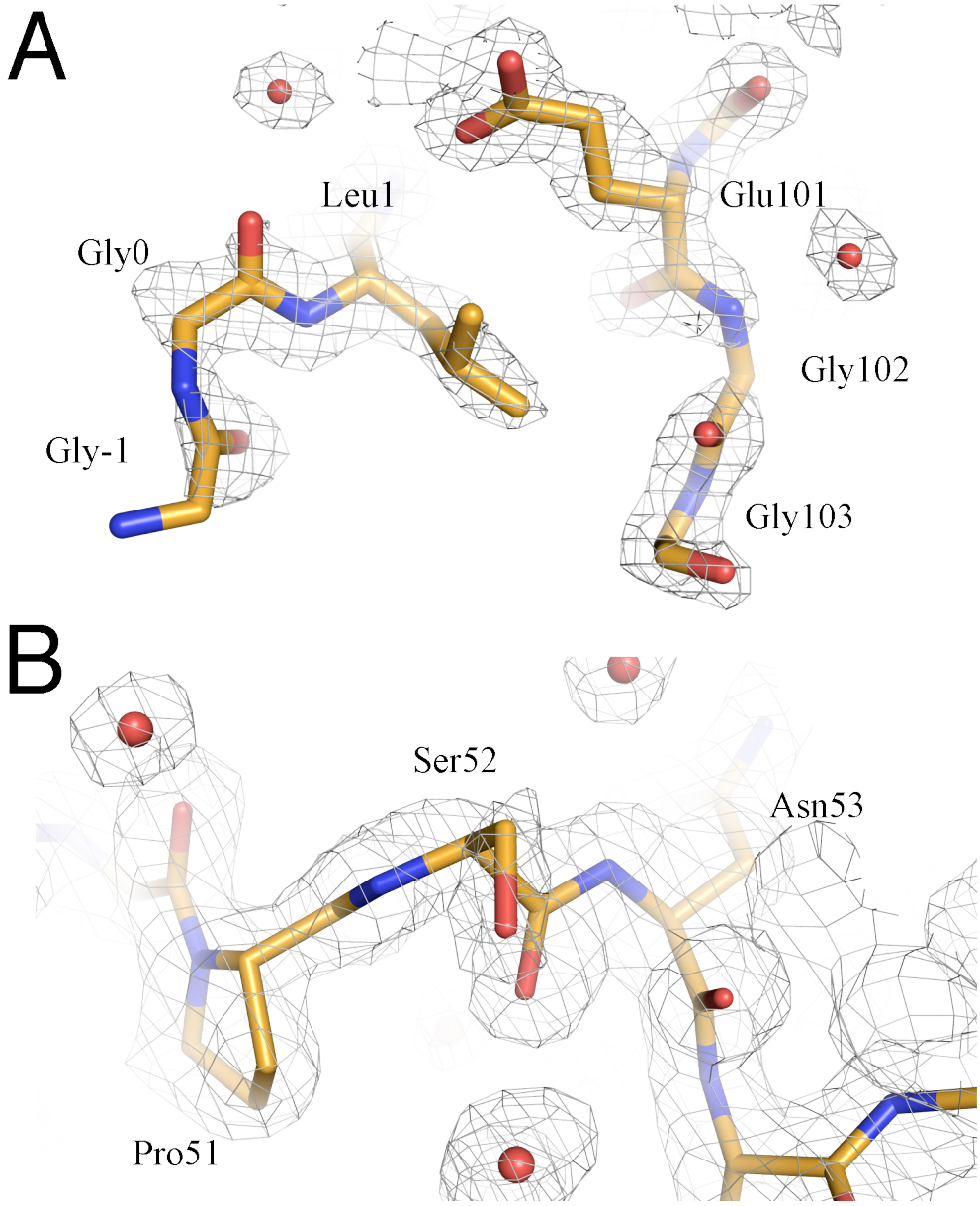

Figure 3-8. CVN_2 L10 structure. Only one copy of CVN was in the asymmetric unit of this structure, therefore the linker residues are only 50% occupied. (A) The termini of the crystal structure with the 2Fo-Fc map contoured to 1.0 σ. Leu1 and Glu101 are both occupied at 100%, whereas the four linker residues with visible density are at 50% occupancy. (B) This structure is domain-swapped as evidenced by the clear electron density in the domain-swapping region (residues 51-53).

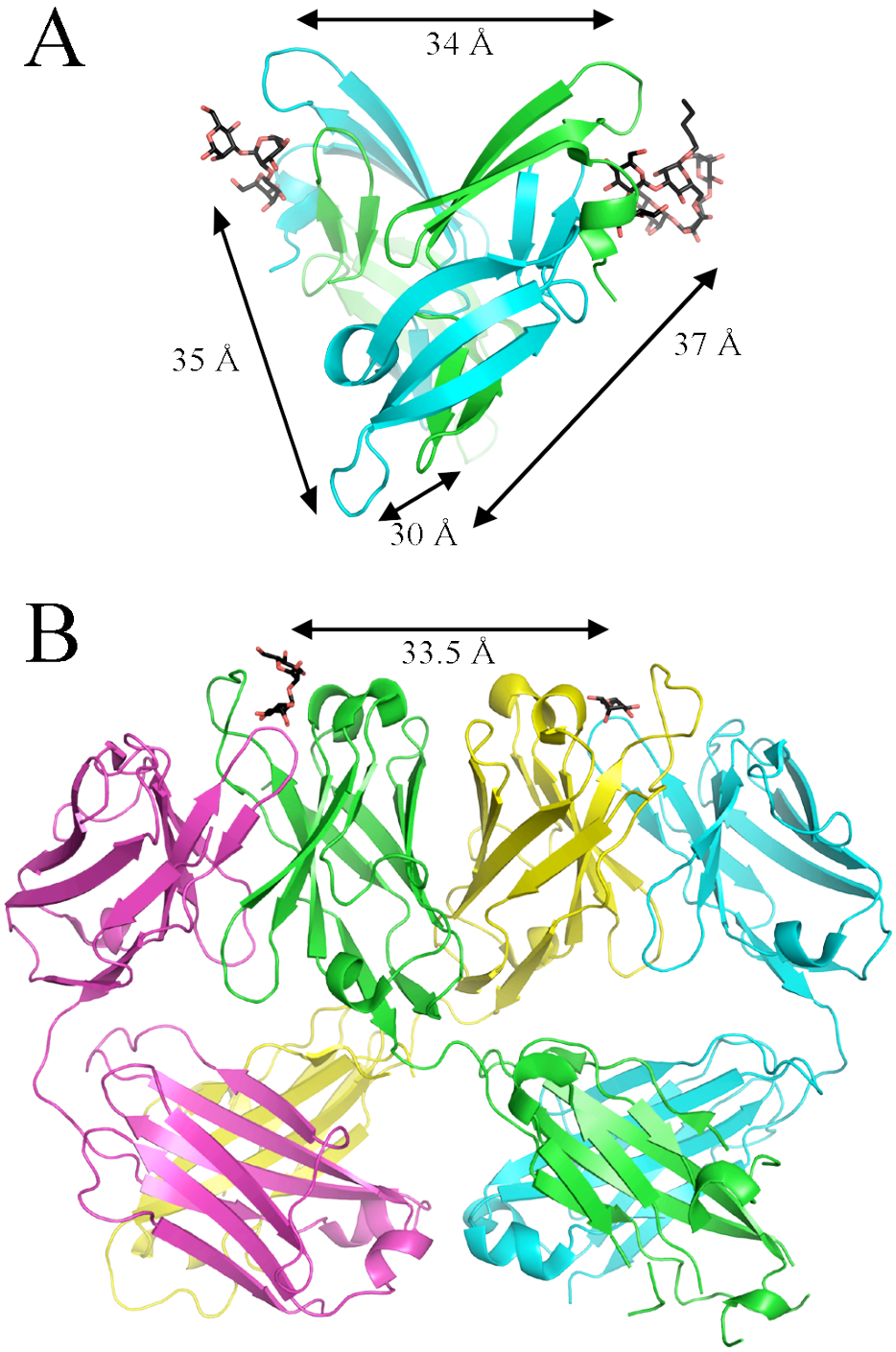

Figure 3-9. Carbohydrate binding site spacing in CVN and the 2G12 anti-HIV Fab. (A) Each of the four carbohydrate binding sites in the WT CVN crystal structure ($P4_12_12$ space group)[10] is approximately 30 to 40 Å from the other sites. (B) The 2G12 Fab, which is specific to carbohydrates on gp120 and is broadly neutralizing, has an unusual domain-swapped form in the crystal structure.[25] This domain-swapping rigidifies the carbohydrate binding sites with respect to each other and holds them approximately 35 Å apart.