Directed Evolution of gp120 Binding Mutants of the Lectin Cyanovirin-N

By

Melissa Ruben

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2013 by the
Graduate Supervisory Committee:

Giovanna Ghirlanda, Chair
James Allen
Rebekka Wachter

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

Cyanovirin-N (CV-N) is a naturally occurring lectin originally isolated from the cyanobacteria *Nostoc ellipsosporum*. This 11 kDa lectin is 101 amino acids long with two binding sites, one at each end of the protein. CV-N specifically binds to terminal Manα1-2Manα motifs on the branched, high mannose Man_9 and Man_8 glycosylations found on enveloped viruses including Ebola, Influenza, and HIV. wt-CVN has micromolar binding to soluble Manα1-2Manα and also inhibits HIV entry at low nanomolar concentrations. CV-N's high affinity and specificity for Manα1-2Manα makes it an excellent lectin to study for its glycan-specific properties. The long-term aim of this project is to make a variety of mutant CV-Ns to specifically bind other glycan targets. Such a set of lectins may be used as screening reagents to identify biomarkers and other glycan motifs of interest. As proof of concept, a T7 phage display library was constructed using P51G-m4-CVN genes mutated at positions 41, 44, 52, 53, 56, 74, and 76 in binding Domain B. Five CV-N mutants were selected from the library and expressed in BL21(DE3) *E. coli*. Two of the mutants, SSDGLQQ-P51Gm4-CVN and AAGRLSK-P51Gm4-CVN, were sufficiently stable for characterization and were examined by CD, T_m, ELISA, and glycan array. Both proteins have CD minima at approximately 213 nm, indicating largely β-sheet structure, and have T_m values greater than 40°C. ELISA against gp120 and RNase B demonstrate both proteins' ability to bind high mannose glycans. To more specifically determine the binding specificity of each protein, AAGRLSK-P51Gm4-CVN, SSDGLQQ-P51Gm4-CVN, wt-CVN, and P51G-m4-CVN were sent to the Consortium for Functional Glycomics (CFG) for glycan array analysis. AAGRLSK-P51Gm4-CVN, wt-CVN, and P51G-m4-CVN, have identical specificities for high mannose glycans containing terminal Manα1-2Manα. SSDGLQQ-P51Gm4-CVN binds to terminal GlcNAcα1-4Gal motifs and a subgroup of high mannose glycans bound by P51G-m4-CVN. SSDGLQQ-wt-CVN was produced to restore anti-HIV activity and has a high nanomolar EC_{50} value compared to wt-CVN's low nanomolar activity. Overall, these experiments show that CV-N Domain B can be mutated and retain specificity identical to wt-CVN or acquire new glycan specificities. This first generation information can be used to produce glycan-specific lectins for a variety of applications.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

ABSTRACT

The 101 amino acid lectin Cyanovirin-N (CV-N) was discovered in an anti-HIV screen. Its low nanomolar anti-HIV activity has made it of great interest. CV-N has broad antiviral activity across a variety of HIV strains and blocks viral infection by binding to high-mannose glycosylations on gp120. This antiviral activity is mediated by multivalent interactions with CV-N's two binding domains. CV-N has a pseudo sequence repeat from 1-50 and 51-101, and two binding domains containing residues 1-38 and 90-101 in Domain A, and 39-89 in Domain B. The protein can also exist as a domain swapped dimer. Several studies were conducted in which either Domain A or Domain B were knocked out, allowing each domain to be studied individually. The binding pocket in Domain B is deep and binds well to Manα1-2Manα. Domain A has a shallower pocket that also binds Manα1-2Manα. P51G-m4-CVN has a stabilizing mutation that also reduces the protein's ability to dimerize, while ΔQ50-m4-CVN and S52P-m4-CVN produce obligate domain swapped dimers. Whether it's two Domain A's, two Domain B's, or one of each domain, two active binding domains are required for anti-HIV activity. Related to CV-N is a family of proteins called CV-N Homologs (CVNH). This group of proteins is predicted to have very similar backbone structures to CV-N and a variety of glycan binding specificities.

BACKGROUND

Glycans and Lectins. Protein-carbohydrate interactions are essential for many *in vivo* biological functions including cell-cell interactions and many host-pathogen interactions. The field of glycobiology is relatively young and much research is being performed to specifically understand the role of glycans in biology and human diseases (Varki, 2009). Proteins can fold into a variety of conformations based on the linear arrangement of their amino acids. Glycans are built from monomeric sugars which can be linked together in a variety of ways, including branched structures, giving glycans far more structural variability than proteins (Varki, 2009).

Unlike cellular proteins, glycans are not directly encoded in an organism's DNA and may be added to proteins as posttranslational modifications. Two common types of glycosylation, N-linked and O-linked glycans, are attached to specific amino acids in protein sequences. During

1

translation, some proteins enter the endoplasmic reticulum (ER) and receive N-linked glycosylations. Further processing occurs in the Golgi complex where the N-linked glycans can be modified and O-linked glycans may be added (Alberts, 2002). However, not every potential glycosylation site will receive a sugar. Glycosylation is partially based on conformational availability of the linkage site and potential glycosylation sites buried in the protein are far less likely to receive glycans (Varki, 2009).

N-linked glycans are attached at Asn when the amino acid sequence is Asn-X-Ser/Thr, where X is any residue except Pro. All N-linked glycans begin as large structures containing two N-acetylglucosamines (GlcNAc), nine mannoses, and three glucoses. They can be processed into high-mannose, complex type, or hybrid-type N-linked glycans. Each type has the basic subunit of an Asn, two GlcNAcs followed by a β1-4 linked mannose, and two branched mannoses attached by α1-3 and α1-6 linkages (Alberts, 2002; Varki, 2009). Figure 1 shows examples of each of the major types of N-linked glycosylations, including high-mannose, complex-type, and hybrid-type.
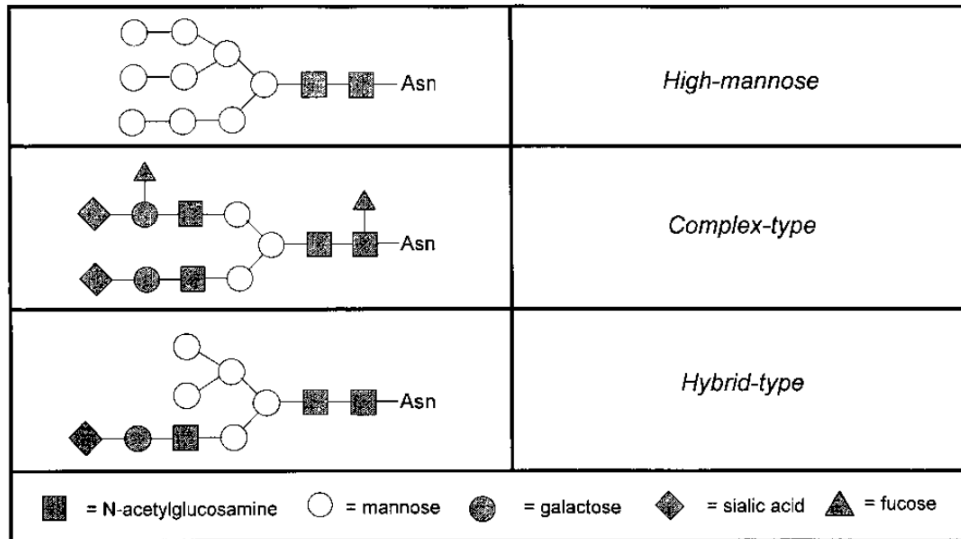
Figure 1: Subgroups of N-linked glycans. This is figure use with permission (Bolmstedt, O'Keefe, Shenoy, McMahon, & Boyd, 2001).

O-linked glycosylations can be attached to Thr and Ser residues. Though specific recognition sequences are not usually known, there is often an abundance of Pro near the

glycosylation sites. One common form of O-linked glycosylations, mucins, occurs in the epithelia and mucus of the respiratory, genitourinary, and gastrointestinal tracts. Mucins contain a large variety of glycan structures and one peptide can have hundreds of glycosylated sites. They can also polymerize, giving viscosity to mucus and helping to protect individuals from pathogenic organisms (Varki, 2009).

Lectins are proteins that bind to glycans, and such proteins are found throughout nature (Alberts, 2002; Arnaud, Audfray, & Imberty, 2013; Varki, 2009). Glycans and lectins often work together to accomplish many important biological tasks, such as pathogen-host binding and immune system function. The binding pockets of many lectins are shallow (Arnaud et al., 2013; C. A. Bewley, 2001; Varki, 2009) and accommodate between one and three terminal monosaccharides (Varki, 2009). Considering the large glycan diversity, many lectins are quite specific for their targets, often associating through hydrogen bonding (Arnaud et al., 2013). Most characterized lectins have μM to low mM binding affinity at a single binding site and commonly act multivalently to produce nM to μM avidity (Arnaud et al., 2013; C. A. Bewley, 2001; Varki, 2009; Shenoy et al., 2002). Multivalent interactions are common among lectins and can be essential for certain tasks such as antiviral activity (Fromme et al., 2007; Y. Liu et al., 2009).

Wild-type CVN (wt-CVN). CV-N is a 101 amino acid, 11 kDa lectin identified during a natural products screen for anti-HIV agents. Wild-type cyanovirin-N (wt-CVN) comes from cyanobacterium *Nostoc ellipsosporum* (blue-green algae) and its ability to prevent HIV infection has made it a well-studied lectin (Boyd et al., 1997). CV-N has internal sequence homology between the first fifty and second fifty amino acids. Sequence alignment of amino acids 1-50 and 51-101, with a one residue space at position 16, reveals 16 strictly conserved residues with 13 more that are modestly mutated (Gustafson et al., 1997). wt-CVN also contains two disulfide bonds, one between residues 8 and 22, and another between residues 58 and 73 (Gustafson et al., 1997). Figure 2 illustrates the sequence homology with dark lines for exactly conserved residues and dotted lines for similar residues.

3

```
1                                                              50
LGKFSQTCYNSAIQGS*VLTSTCERTNGGYNTSSIDLNSVIENVDGSLKWQ
  :  | :|| |:  :  ||     |    |        :  :: |:|:  | |:||:||::
PSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDDHIANIDGTLKYE
51                                                             101
```

Figure 2. Sequence alignment of internal CV-N repeat. There is a sequence gap (*) at residue 16. Solid bars indicate identical amino acids and dashes indicate modestly mutated residues. The large lines indicate disulfide bonds. (Gustafson et al., 1997)

Each of wt-CVN's sequence repeats forms a three-stranded β-sheet, a helical turn, and β-hairpin (L. Liu, Byeon, Bahar, & Gronenborn, 2012; Percudani, Montanini, & Ottonello, 2005). The β-hairpin from one repeat interacts with the three-stranded β-sheet of the other repeat via hydrophobic interactions (C. A. Bewley et al., 1998). Figure 4A shows an NMR structure (PDB: 1IIY) of wt-CVN, with residues 1-50 in yellow and 51-101 in blue.

CV-N also has two binding sites, one at each end of the protein (C. A. Bewley & Otero-Quintero, 2001). Binding Domain A comprises residues 1-38 and 90-101 and Domain B comprises residues 39-89 (Barrientos & Gronenborn, 2002; C. A. Bewley et al., 1998; Yang et al., 1999). As seen in Figure 4B, each binding domain has part of each sequence repeat, with Domain A in green and Domain B in burgundy. CV-N is ~25Å wide and 55Å long with ~40Å between the center of each binding site (C. A. Bewley & Otero-Quintero, 2001), and contains no glycosylations (Boyd et al., 1997; Gustafson et al., 1997).

```
          90        100        10         20         30
Domain A  HIANIDGTLKYELGKFSQTCYNSAIQGS-VLTSTCERTNGGYNTSSIDLNS
Domain B  VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
           40        50         60         70         80
```

Figure 3. Sequence alignment of wt-CVN Domain A and Domain B. Domain A comprises residues 90-101 and 1-38, and Domain B comprises residues 39-89. Figure adapted from (C. A. Bewley, 2001).
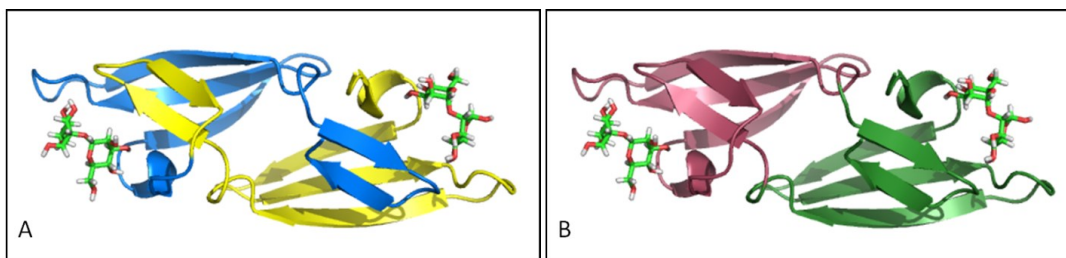
Figure 4. NMR structure of wt-CVN bound to Manα1-2Manα (PDB: 1IIY). N and C termini are in the right side of the protein. A) The pseudo sequence repeat is demonstrated by amino acids 1-50 in yellow and 51-101 in blue; B) Binding Domain A comprises amino acids 1-38 and 90-101 in green, and Domain B comprises amino acids 39-89 in burgundy.

wt-CVN is monomeric in the NMR solution structure, but the solved crystal structure shows

that it can also exist as a domain-swapped dimer (Yang et al., 1999). Domain swapping in CV-N

is facilitated by the helical linker region comprising amino acids 50-56 (Botos et al., 2002;

Fromme et al., 2007; L. Liu et al., 2012; Yang et al., 1999). Figure 5 shows the wt-CVN dimer in

which residues 1-50 and 51'-101' and residues 1'-50' and 51-101 associate, giving the protein

four binding sites. Both regions of the dimer have similar structures to their monomeric form

except for the extended linker region (Botos et al., 2002; Yang et al., 1999).
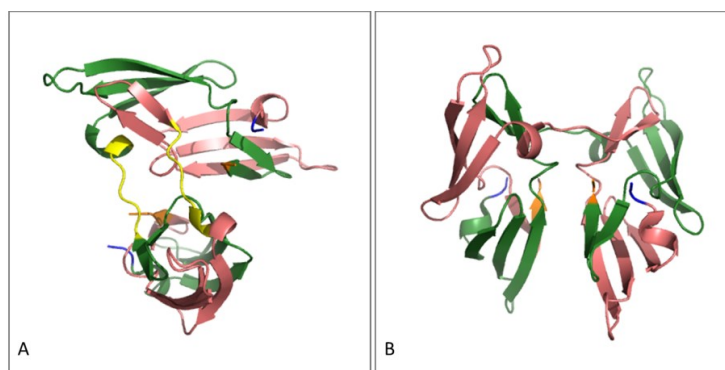
Figure 5. Domain-swapped dimer CV-N crystal structure (PDB: 3GXZ) Glycans were omitted from the figure. CV-N is shown in green and CV-N' is pink. The N terminus of each protein is blue and the C terminus is orange. A) CV-N dimer with the linker region 49-54 in yellow. B) A rotated view of the dimer showing association of CV-N and CV-N'.

wt-CVN Binding. wt-CVN was shown to bind glycoprotein 120 (gp120) on the Human

Immunodeficiency Virus (HIV) envelope and prevent HIV infection and cell-cell transmission in *in*

vitro cellular assays (Boyd et al., 1997). Remarkably, CV-N is effective against a variety of strains

including HIV-1 primary isolates and lab strains, HIV-2, and simian immunodeficiency virus (SIV)

(Boyd et al., 1997).

HIV is a retrovirus belonging to the subgroup *lentiviruses*, which cause diseases that progress slowly (Acheson, 2007). The enveloped virus is covered in glycoprotein spikes made of a transmembrane domain, gp41, which is noncovalently attached to external gp120 (J. Liu, Bartesaghi, Borgnia, Sapiro, & Subramaniam, 2008; O'Keefe et al., 2000). gp120 exists on the viral membrane as a trimer and is heavily glycosylated, with glycosylation accounting for ~50% of the protein's molecular weight (Hu, Mahmood, & Shattock, 2007; Kwong et al., 1998). Of approximately 24 N-linked glycosylations on gp120, 13 are complex type glycans and about 11 are N-linked high mannose sites (Scanlan, Offer, Zitzmann, & Dwek, 2007; Xiong, Fan, & Kitazato, 2010; Yeh, Seals, Murphy, van Halbeek, & Cummings, 1993; Zhu, Borchers, Bienstock, & Tomer, 2000).

HIV infection is mediated by gp120 and gp41 binding to CD4+ cells, usually T-lymphocytes and macrophages expressing chemokine CXCR4 and CCR5, respectively (Acheson, 2007; Esté & Telenti, 2007; Wilen, Tilton, & Doms, 2012). There are also HIV strains that recognize both coreceptors. Host cell entry begins when gp120 binds to CD4, followed by secondary binding to co-receptor CCR5 or CXCR4. CD4 binding causes conformational changes in gp120 and gp41, allowing the helical gp41 trimer to insert itself into the host cell membrane. gp41 then rearranges into a six-helix bundle, allowing fusion of the viral and cell membranes (Alberts, 2002; Eckert & Kim, 2001; J. Liu et al., 2008; Wilen et al., 2012). Figure 6 depicts the HIV entry mechanism.
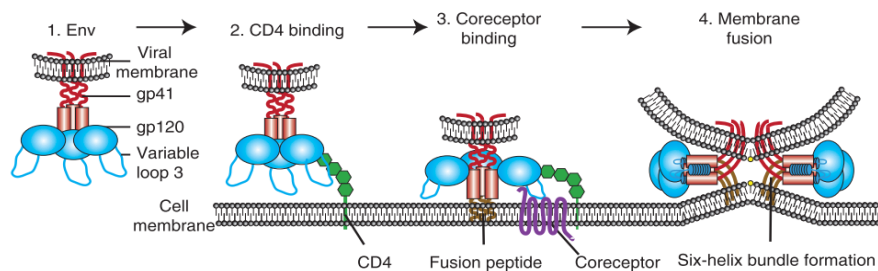
Figure 6. Overview of HIV entry mechanism. To deliver the viral payload into cells, HIV Env, comprised of gp120 and gp41 subunits (1), first attaches to the host cell, bindingCD4 (2). This causes conformational changes in Env, allowing coreceptor binding, which is mediated in part by the V3 loop of Env (3). This initiates the membrane fusion process as the fusion peptide of gp41 inserts into the target membrane, followed by six-helix bundle formation and complete membrane fusion (4) (Wilen et al., 2012).

Several binding studies showed that CV-N does not directly block the CD4 binding site on gp120 or prevent binding of several anti-gp120 antibodies (Boyd et al., 1997; Esser et al., 1999). One exception is human antibody 2G12, which is specific for Manα1-2Man terminal high-mannose glycans (Calarese et al., 2005; Doores, Fulton, Huber, Wilson, & Burton, 2010), and overlaps with CV-N's binding to gp120 (Alexandre et al., 2010; Dey et al., 2000; Esser et al., 1999). To verify the glycan-dependent nature of binding, wt-CVN was incubated with non-glycosylated gp120, but no significant binding was observed (Boyd et al., 1997; O'Keefe et al., 2000; Shenoy, O'Keefe, Bolmstedt, Cartner, & Boyd, 2001). The densely N-glycosylated region where 2G12 and CV-N bind is located nearly opposite of the CD4 binding site on gp120 (Scanlan et al., 2007), consistent with the binding studies. Because CV-N prevents HIV infection by binding to viral high-mannose glycosylations, it can similarly bind high-mannose glycans on other enveloped viruses (C. A. Bewley, 2001; Bolmstedt et al., 2001; O'Keefe et al., 2000) including Ebola (Barrientos et al., 2003), simian immunodeficiency virus (SIV) (Boyd et al., 1997), feline immunodeficiency virus (FIV), measles, human herpes virus 6 (Dey et al., 2000), and influenza (O'Keefe et al., 2003).

Cell fusion assays more specifically demonstrated the binding of wt-CVN to $Man_9GlcNAc_2$ (Man_9) and $Man_8GlcNAc_2$ D1D3 (Man_8) with nanomolar affinity (C. A. Bewley & Otero-Quintero, 2001). Further analysis by NMR determined that wt-CVN binds soluble Manα1-2Manα with low micromolar affinity (C. A. Bewley, 2001; C. A. Bewley & Otero-Quintero, 2001). NMR and isothermal titration calorimetry (ITC) experiments indicate that Domain A has a weaker affinity for Manα1-2Manα than Domain B with ITC yielding K_a values of 6.8 (±4) x 10^5 M^{-1} and 7.2 (±4) x 10^6 M^{-1} respectively (C. A. Bewley, 2001; C. A. Bewley & Otero-Quintero, 2001). CV-N's binding sites also recognize the three trimannose arms of Man_9 (Manα1-2Manα1-2Man, Manα1-2Manα1-3Man, Manα1-2Manα1-6Man) with affinities ranging from millimolar to micromolar (C. A. Bewley, Kiyonaka, & Hamachi, 2002).

The NMR structure of CV-N bound to Manα1-2Manα revealed the amino acid positions involved in sugar binding (C. A. Bewley, 2001; C. A. Bewley & Otero-Quintero, 2001). Domain A uses seven amino acids: K3, Q6, T7, E23, T25, N93, and I94. Domain B uses 10 residues: E41,

N42, D44, S52, N53, E56, T57, K74, R76, and Q78 to bind to dimannose. In order to sequentially identify equivalent residues in each binding site, the sequences for each domain were aligned. Figure 3 shows the alignment of the amino acid sequences for each domain. Domain B has a deeper binding pocket than Domain A (C. A. Bewley, 2001; C. A. Bewley & Otero-Quintero, 2001), which is partially due to smaller, nonpolar residues at positions Ala92, Gly2, and Gly27 in Domain A. The symmetrically equivalent residues in Domain B are Glu41, Ser52, and Gln78, which are polar residues with side chains larger than Ala and Gly (C. A. Bewley, 2001). Domain B has a deeper binding pocket and tighter binding to dimannose than does Domain A (C. A. Bewley, 2001).

Molecular dynamics studies have revealed a cap-and-lock mechanism in Arg76 of Domain B (Margulis, 2005). Arg76 has been observed in three different confirmations in NMR (C. A. Bewley, 2001) and x-ray crystallography studies (Fromme, Katiliene, Fromme, & Ghirlanda, 2008). Once the dimannose is bound in Domain B, Arg76 can form two direct hydrogen bonds with the sugar, or one water-mediated hydrogen bond. The observed variability combined with the supporting computational data suggests a locking mechanism in which Arg76 provides additional lectin interactions with the bound sugar. Margulis et al. and Fromme et al. suggest that this locking mechanism may contribute to the high affinity binding of CV-N to dimannose.

The same molecular dynamics studies have also pointed to Glu41 as being important to the specificity of CV-N for its ligands (Margulis, 2005). The orientation of the hydroxyl group on bound mannose is important for hydrogen bond formation with Glu41 and Margulis et al. suggest that this single residue may play a major role in the protein's specificity.

Hinge Region Mutants. wt-CVN exists as a domain-swapped dimer when it is crystallized (Yang et al., 1999). Several linker region mutants of CV-N have been produced, demonstrating that a single mutation or deletion can strongly influence the protein's oligomeric state. Domain swapping is found in a variety of proteins and involves two or more of the same protein exchanging motifs via a helical linker region (Bennett, Schlunegger, & Eisenberg, 1995). This can be achieved by refolding a high concentration of the protein from denaturing conditions, changing

the pH or temperature (Barrientos, Louis, Botos, et al., 2002; Bennett et al., 1995; Fromme et al., 2007).

wt-CVN has a Pro at position 51, which has fewer allowed confirmations than many other amino acids, and promotes linker extension by causing residues 50-53 to have unfavorable conformations. Dimeric wt-CVN has an extended linker region in each protein, allowing the amino acids to have more favorable conformations (Barrientos, Louis, Botos, et al., 2002). Monomeric wt-CVN can be changed from monomer to dimer when 0.3-1.0 mM protein is denatured in 8 M urea and refolded, or protein of at least 2 mM is incubated at 38°C. Dimeric wt-CVN can be returned to the monomer form by incubation at 38°C for at least 12 hours (Barrientos, Louis, Botos, et al., 2002). A recent paper discusses the thermodynamics and kinetics of domain-swapping in CV-N (L. Liu et al., 2012). Because of a high energy barrier similar to that for unfolding, it is suggested that CV-N transitions from monomer to dimer and visa-versa by unfolding and refolding.

To better understand the hinge region of CV-N, residues 50-56 (Fromme et al., 2007) were mutated in several ways. One such mutant, S52P-CVN was discovered in a T7 phage display library of CV-N genes produced by error-prone-PCR (EP-PCR) (Han, Xiong, Mori, & Boyd, 2002). This mutant is exclusively dimeric because prolines at positions 51 and 52 forced CV-N into the more open hinge confirmation in the domain-swapped dimer (Barrientos, Lasala, Delgado, Sanchez, & Gronenborn, 2004; Barrientos, Louis, Botos, et al., 2002; Han et al., 2002). S52P-CVN has low nanomolar anti-HIV activity (Han et al., 2002) and has a T_m = 53.5 ± 0.5°C (Barrientos et al., 2004).

A second hinge region mutant, ΔQ50-CVN was produced in which the glutamine at position 50 was removed from the protein leaving a 100 amino acid version of CV-N with a shortened linker (Kelley, Chang, & Bewley, 2002). The shortened linker forced exclusive CV-N domain swapping due to less favorable steric interactions among the remaining amino acids (Barrientos et al., 2004; Kelley et al., 2002). ΔQ50-CVN also has low nanomolar anti-HIV activity and a melting temperature of 50.2 ± 0.5°C (Barrientos et al., 2004).

9

Another hinge region mutant, P51G-CVN, can form a dimer or monomer, but prefers the monomeric state (Barrientos, Louis, Botos, et al., 2002; Fromme et al., 2007) because Gly can achieve more favorable conformations than Pro in the same position. P51G-CVN dimer can be produced under similar conditions as wt-CVN dimer (Barrientos, Louis, Botos, et al., 2002; Fromme et al., 2007). Like the other hinge region mutants, P51G-CVN also has low nanomolar anti-HIV activity (Mori et al., 2002). P51G-CVN is more thermally stable than wt-CVN, S52P-CVN, and ΔQ50-CVN with a T_m = 67.8 ± 0.5°C (Barrientos, Louis, Botos, et al., 2002; Mori et al., 2002). Interestingly, Gly is conserved at position 51 among CV-N homologs (CVNH), which will be discussed later (Percudani et al., 2005).

m4-CVN. With the discovery of two binding sites with different affinities for Manα1-2Manα, clarification was desired to determine if one or two binding sites are required for anti-HIV activity. NMR and ITC experiments indicate that Domain A has a weaker affinity for Manα1-2Manα than Domain B, with ITC yielding K_a values of 6.8 (±4) x 10^5 M^{-1} and 7.2 (±4) x 10^6 M^{-1} respectively (C. A. Bewley, 2001; C. A. Bewley & Otero-Quintero, 2001). To determine the need for Domain A in anti-HIV activity, three mutants were created with two, three, or four modified amino acids in Domain A. m2-CVN with mutations K3N and E23I still bound dimannose, but m3-CVN (K3N, E23I, N93A) and m4-CVN (K3N, T7A, E23I, and N93A) prevented binding of dimannose to Domain A (L. Chang, 2002). These binding studies were performed by NMR and the results also indicated conservation of the wt-CVN fold and Domain B's ability to bind dimannose. However, m4-CVN has a significantly lower thermal stability with T_m = 45°C (Fromme et al., 2007). Potent anti-HIV activity was also retained and low nanomolar IC_{50} values were observed for wt-CVN, m3-CVN, and m4-CVN (L. Chang, 2002). Based on these results, Chang et al concluded that only Domain B, with high affinity binding to dimannose, was required for anti-HIV activity and the lower affinity Domain A was not needed.

P51G-m4-CVN and CVN[mutDB]. To further understand the role of multivalent interactions in CV-N's anti-HIV activity, P51G-m4-CVN was produced (Fromme et al., 2007). This CV-N mutant contains the P51G mutation (Barrientos, Louis, Botos, et al., 2002) to stabilize the protein and discourage dimer formation, and has Domain A inactivated by mutations K3N, T7A, E23I, and

N93A so that it no longer binds Manα1-2Manα (L. Chang, 2002). With these mutations, P51G-m4-CVN has only one active binding site and is expressed almost exclusively as a monomer. Like wt-CVN and P51G-CVN, it can be made into a dimer when it is denatured and refolded at high concentrations (Fromme et al., 2007). Even with the four destabilizing mutations in Domain A, P51G-m4-CVN is quite thermally stable, with an initially reported T_m of 62°C (Fromme et al., 2007).

Anti-HIV activity was measured using a XTT-tetraxolium assay (Fromme et al., 2007; Gulakowski, McMahon, Staley, Moran, & Boyd, 1991). m4-CVN and P51G-m4-CVN were assayed in parallel and m4-CVN demonstrated nanomolar activity while P51G-m4-CVN did not exhibit antiviral activity up to the 1 μM maximum concentration. Upon observing m4-CVN's nanomolar anti-HIV activity, Chang et al. concluded that only a single binding site was required for CVN's antiviral activity (L. Chang, 2002). Monomeric m4-CVN was used for his anti-HIV assay, but the protein likely dimerized during the assay, restoring its antiviral activity (Fromme et al., 2007). Monomeric wt-CVN is capable of converting to the domain-swapped dimer during prolonged incubation at 38°C. Barrientos et al. demonstrated that one third of a 2.2 mM monomeric wt-CVN sample was converted to dimer after 4 days at 38°C (Barrientos, Louis, Botos, et al., 2002). Though wt-CVN was used at much lower concentrations in the antiviral assay, the assay was conducted over six days at 38°C and wt-CVN likely achieved a monomer-dimer equilibrium under these conditions. P51G-m4-CVN is a thermally stable monomer in the assay conditions and clearly demonstrates that at least two active binding domains are required for potent anti-HIV activity (Fromme et al., 2007). Though P51G-m4-CVN does not retain anti-HIV activity, it does bind to gp120 and Manα1-2Man with nanomolar and micromolar affinities respectively [(Fromme et al., 2007) and unpublished data]. This is consistent with an early mutation study establishing that CV-N can bind gp120 without retaining anti-HIV activity (Mori et al., 1997).

Barrientos et al. also produced a CV-N mutant (CVN[mutDB]) containing P51G and E41A, N42A, T57A, R76A, and Q78G to inactivate Domain B (Barrientos, Matei, Lasala, Delgado, & Gronenborn, 2006; Matei, Furey, & Gronenborn, 2008). CVN[mutDB] has a T_m of 72°C and does not

bind gp120 in ELISA or exhibit antiviral activity. However, ITC shows that Domain A does have a $K_d = 4.3 \pm 0.3$ μM for Man_9. Knocking out Domain B reinforces the finding that CV-N requires multivalent interactions for anti-HIV activity.

ΔQ50-m4-CVN and S52P-m4-CVN. To further demonstrate the importance of multivalent interactions in CV-N's anti-HIV activity, my lab also produced mutants ΔQ50-m4-CVN and S52P-m4-CVN (Y. Liu et al., 2009). These mutants combine K3N, T7A, E23I, N93A (m4-CVN), and hinge region mutations designed to force domain swapping. ΔQ50-m4-CVN lacks Pro51 found in wt-CVN, giving it a shorter linker region (Kelley et al., 2002). S52P-m4-CVN has two prolines in the linker region, Pro51 and Ser52Pro (Han et al., 2002). These modifications cause sterically unfavorable amino acid conformations in the hinge region (50-56) allowing these mutants to favor the open linker position and form domain-swapped dimers (Barrientos et al., 2004; Barrientos, Louis, Botos, et al., 2002; Han et al., 2002; Kelley et al., 2002). The dimeric form of each mutant has two active B Domains and no active A Domains. ELISA against gp120 shows that ΔQ50-m4-CVN and S52P-m4-CVN have nanomolar binding worse than monomeric wt-CVN, but better than monomeric P51G-m4-CVN (Y. Liu et al., 2009). ΔQ50-m4-CVN and S52P-m4-CVN also possess anti-HIV activity of $EC_{50} = 320$ nM and ~3000 nM respectively, which are both worse than wt-CVN (Y. Liu et al., 2009).

ΔQ50-m4-CVN and S52P-m4-CVN are not very thermally stable, with T_m of 32.8°C and 39.5°C respectively, which likely contributed to their reduced antiviral activity (Y. Liu et al., 2009). The anti-HIV assay was conducted at 38°C which is near or above their melting points, indicating that much or all of the CV-N was denatured by the end of the six day assay. The combination of four destabilizing Domain A mutations and linker mutations favoring extended linkers destabilize the proteins more than either of the mutations alone (Barrientos et al., 2004; Fromme et al., 2007; Y. Liu et al., 2009). Despite less than ideal protein stability, these two CV-N mutants demonstrate that antiviral activity is the result of multivalent interactions. The results also demonstrate that multivalency can be achieved by two active binding domains in monomeric CV-N or by two active domains in dimeric CV-N.

12

Covalent Dimers. Due to the transient nature of some of the dimer species, and to further examine the avidity effect of multiple binding sites, covalent dimers have been produced in two separate labs.

Matei et al produced a CV-N mutant [CVN$^{\Delta A}$]$_{ssd}$ with inactivated Domain A, P51G linker mutation, and a G2C mutation to allow dimerization by disulfide bond formation (Matei et al., 2010). These mutations created a species with only one active domain in the monomeric form, with two active domains in the disulfide-bonded dimer. [CVN$^{\Delta A}$]$_{ssd}$ has slightly worse anti-HIV activity than P51G-CVN, and the monomeric [CVN$^{\Delta A}$]$_{ssm}$ showed no anti-HIV activity up to 100 nM. While [CVN$^{\Delta A}$]$_{ssd}$ did not have improved anti-HIV activity, it did reinforce the idea that multivalent interactions are required for CV-N's anti-HIV activity.

Keeffe et al produced several wt-CVN dimers in which two proteins were attached in series, C-terminus to N-terminus (CVN$_2$) (Keeffe et al., 2011). A peptide linker between zero and 20 amino acids long was used to attach the two proteins. CVN$_2$ can exist as two monomers attached by the linker, or in a domain-swapped form. Anti-HIV activity was slightly increased for CVN$_2$ over wt-CVN, but only by three to six fold. This group also made trimer (CVN$_3$) and tetramer (CVN$_4$) versions, but these species did not show improved anti-HIV activity. The authors hypothesized that CVN$_3$ and CVN$_4$ did not show improved anti-HIV activity due to possible orientation away from the glycan target or that some of the binding sites may be blocked by the protein itself. Domain knockout versions of CVN$_2$ were also produced, reinforcing the need for multivalent CV-N interactions for anti-HIV activity.

To date, covalent dimer or higher order CV-N structures have not exhibited substantially improved anti-HIV activity, but have further verified the need for at least two active binding domains for antiviral activity.

CVN-Homologs. CV-N has a predominantly β-sheet fold and, as can be seen in Figure 7, sequence searches have revealed a family of similar sequences in a variety of species (Percudani et al., 2005). These CV-N homologs (CVNH) have nine strictly conserved residues, 11 partially conserved residues, and structure recognition methods indicate conserved secondary structure. Most of the conserved residues are located in the protein core, with only four conserved

in the binding pockets. Residues 7 and 93 are conserved in Domain A and 42 and 57 are in

Domain B. The previously discussed m4-CVN mutant takes advantage of this and mutates

residues 3, 7, 23, and 93 to abolish mannose binding activity in Domain A (L. Chang, 2002). A

similar approach was employed with CVN[mutDB], in which E41A, N42A, T57A, R76A, and Q78G

were mutated to abolish binding by Domain B (Barrientos et al., 2006). With the exception of

amino acid 76, all of these positions are highly conserved among CVNH. Interestingly, Gly 41 is

strictly conserved among CVNH, but Glu 41 is native in CV-N.

Similarly, Pro 51 is found in CV-N while Gly 51 is commonly found, though not strictly

conserved, in CVNH (Percudani et al., 2005). This information was considered during

experimental design, and residues conserved in CV-N and CVNH were conserved in my Domain

B mutant library. Because Glu 41 was different from the CVNH consensus and it is important for

CV-N binding (Margulis, 2005), it was mutated in my experiments. Finally, many of the CVNH

sequences lacked the two disulfide bonds found in CV-N.
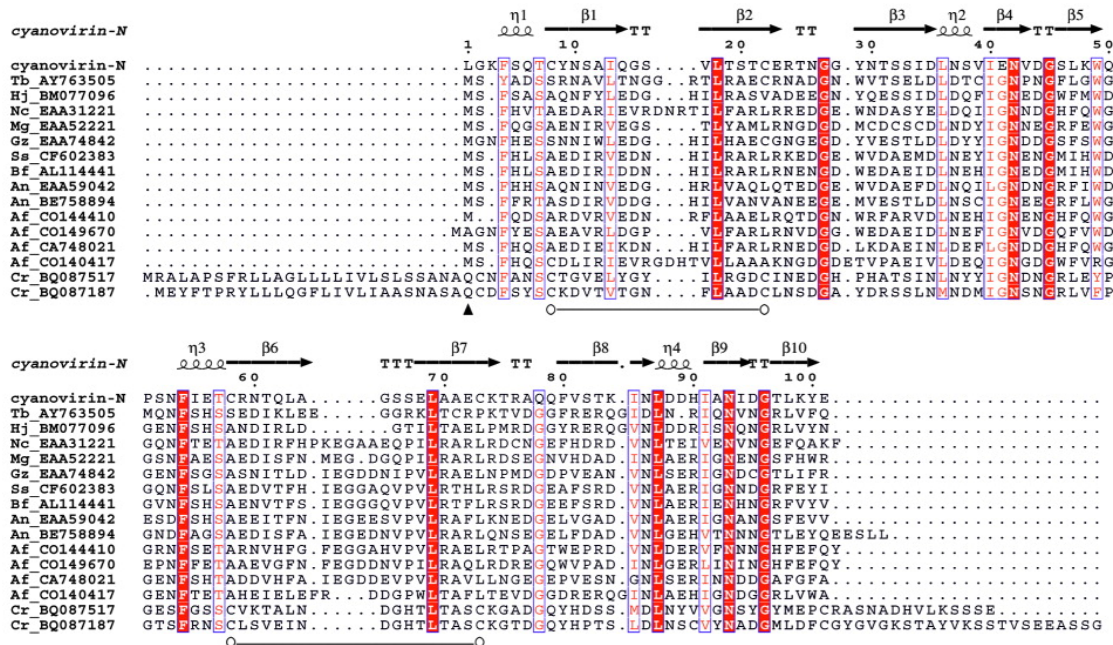
Figure 7. Sequence alignment of CV-N homologs (CVNH). Conserved residues are highlighted.
(Percudani et al., 2005)

To briefly illustrate the structural similarities of CV-N and CVNHs, solution structures of

CrCVNH (from *Ceratopteris richardii*), TbCVNH (from *Tuber borchii*), and NcCVNH (from

Neurospora crassa) are displayed in Figure 8 (Koharudin, Viscomi, Jee, Ottonello, & Gronenborn,

2008). Each of these proteins binds Manα1-2Man and has additional glycan specificities.

TbCVNH and NcCVNH have only one binding site each and CrCVNH has two binding sites.

Consistent with CV-N, two binding sites are required for anti-HIV activity among all three CVNHs

as evidenced by CrCVNH having anti-HIV and TbCVNH and NcCVNH having very weak or non-

existent anti-HIV activity. All three CVNHs have very similar 3-D structures to each other and to

CV-N, with some variation in length, disulfide bonds, and loop regions.
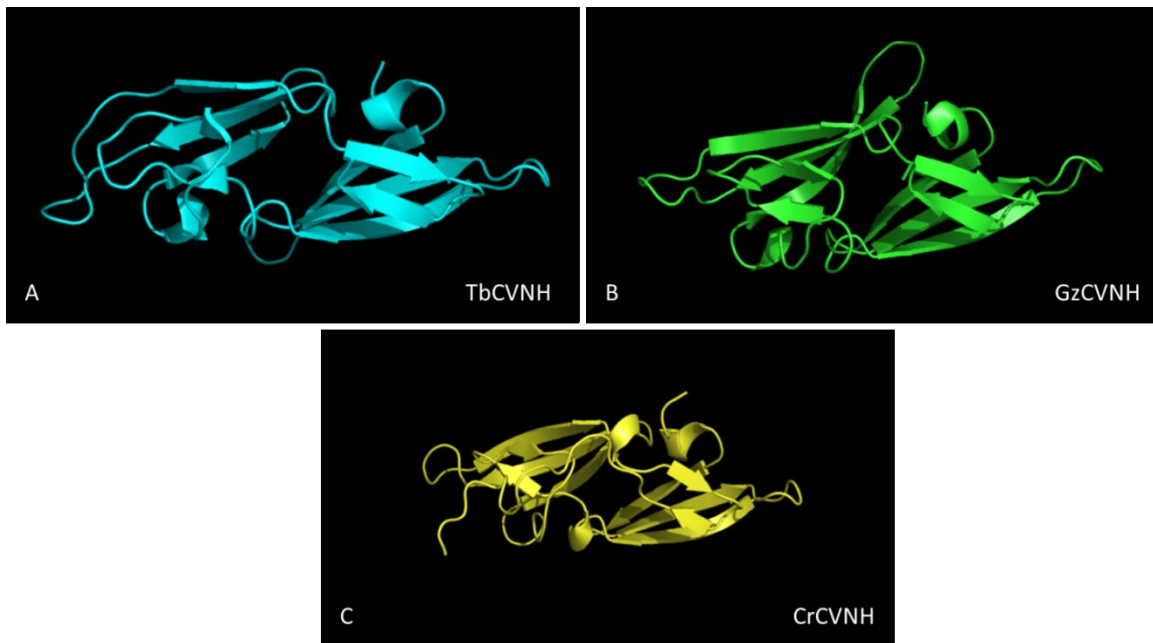
Figure 8. NMR structures of CV-N homologs. A) TbCVNH (PDB: 2JZK) B) GzCVNH (PDB: 2L2F) C) CrCVNH (PDB: 2JZJ)

CHAPTER 2

T7 PHAGE DISPLAY LIBRARY

ABSTRACT

In order to study the CV-N Domain B binding pocket, amino acids 41, 44, 52, 53, 56, 74, and 76 were randomized in P51G-m4-CVN. The mutated genes were used to make a T7 phage display library of CV-N Domain B mutants. The library was selected against gp120 for four rounds of biopanning, followed by one round against RNase B, and a sixth and final round against gp120. Nine DNA sequences were sampled from each of rounds 3-6, for a total of 36 sequences. Seven unique sequences were identified, five of which are full length CV-N sequences and two contain a stop codon. Randomized positions of the unique full length sequences were compared to P51G-m4-CVN. There was very little residue conservation among the mutants when compared to P51G-m4-CVN, with only two instances of residue conservation observed. Conservation of charged versus uncharged amino acids was examined as well as bulky verses small residues. R76 in P51G-m4-CVN was only conserved in LPRANHR-P51Gm4-CVN, but all of the sequences contain bulky, charged amino acids at that position. Conversely, E41 in P51G-m4-CVN was not conserved and the position was consistently filled with uncharged, small amino acids in harmony with CVNH sequence studies. The two truncated sequences were 73 and 75 residues long, including most of the Domain B binding site. However, approximately 25% of each protein is missing and it is unlikely that either sequence could fold properly.

INTRODUCTION

CV-N has been well studied in its anti-HIV and anti-viral capacities, but this project is designed to study CV-N as a lectin and develop a method by which it can be modified to specifically bind other glycan motifs. As proof of concept, P51G-m4-CVN will be selected for binding to gp120. Subsequently, CV-N could be modified to specifically bind other glycans, making it a source of customizable lectins.

T7 Phage Display. Phage display is a powerful tool used to select peptides or proteins with specific characteristics from large libraries of molecules. A gene or gene library of interest can be ligated into a phage capsid gene and expressed on the outside of the phage as part of its

16

protein coat. Phage display is useful because the protein and DNA are kept together throughout the selection process. Once the desired proteins are identified, individual genes reveal the sequences of interest and can be used to produce more protein. T7 phage was chosen for this project, and it has been successfully used by another group studying CV-N (Han et al., 2002). This particular virus has an icosahedral shape (60 nm capsid) with a tail and six tail fibers. It also has a 40 Kb linear, double-stranded DNA genome (Acheson, 2007). T7 is a lytic phage and lyses the host *E. coli* when progeny virions are released (Novagen, 2002; Rosenberg, 1996). T7 also grows relatively quickly compared to other phage (Novagen, 2002).

T7 phage display systems are available in high-copy and low-copy versions from Novagen (Rosenberg, 1996). T7 phage naturally produce two forms of their capsid protein, 10A and 10B. The capsid can be made of one protein or the other or a mixture of both. For phage display, Novagen has modified the T7 genome to only express 10B, including the recombinant gene. A gene of interest can be ligated into the phage genome cloning region beginning at amino acid 348. High copy phage are completely formed from recombinant 10B capsid protein and display peptides up to 39 amino acids long on all 415 capsid proteins. To accommodate longer genes, low copy phage only express 0.1 to 1 copy of a protein or peptide up to about 1200 amino acids in length. CV-N is 101 amino acids long and I selected the Novagen T7Select 1-1 low copy phage display system (Novagen, 2002). In this case, the phage are grown in BLT5615 *E. coli* that also contain a plasmid for producing 10A under control of the *lacUV5* promoter. The phage gene produces capsid 10B, but has been modified to reduce production from wild-type levels. T7 phage grown in this strain incorporate 10A protein produced by the bacterial plasmid and 10B from the phage genome.

Library Selection. As proof of concept, I used recombinant gp120 as the binding target for this protocol. Future work will employ other glycan targets that are not native for wt-CVN. wt-CVN and P51G-m4-CVN are known to bind gp120 at the Manα1-2Manα motif of N-linked high-mannose Man$_8$ and Man$_9$ (C. A. Bewley, 2001; C. A. Bewley & Otero-Quintero, 2001). Recombinant gp120 (rgp120) was obtained from the NIH AIDS Reagent program (NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 gp120 CM, Cat

2968). It was produced in a baculovirus expression system in insect cells, likely similar to the protocol described by Yeh, et al. (Yeh et al., 1993). Their rgp120 contains 22 N-linked, high mannose glycans, and has no O-linked glycans.

A positive selection control was used to retain CV-N mutants that bound to high-mannose glycans. Ribonuclease B (RNase B) has one high-mannose N-linked glycosylation, but is otherwise quite different in structure from gp120 (Goodsell, 2008; J. Liu et al., 2008), making RNase B a suitable positive control. RNase B and ribonuclease A (RNase A) are ~15 kDa and 13.7 kDa enzymes that are found in the liver and degrade RNA (Sigma). These two enzymes are the same with the exception of a single N-linked glycosylation on RNase B and no glycosylation on RNase A. RNase B has a single N-linked glycosylation at Asn34 (Sigma) ranging in structure from Man_5 to Man_9 (Hua et al., 2012; Prien, Ashline, Lapadula, Zhang, & Reinhold, 2009; Sigma). According to Hua et al., Man_5 appears on RNase B almost 50% of the time with Man_8 and Man_9 each accounting for 10% or fewer of the glycan species.

CV-N Gene Library. For phage display and binding studies, I decided to use P51G-m4-CVN (Fromme et al., 2007), which has four mutations that destroy Domain A's ability to recognize Manα1-2Manα (L. Chang, 2002). This arrangement allows me to study Domain B without multivalent interactions from Domain A. Hinge region mutation P51G (Mori et al., 2002) was also incorporated to discourage CV-N from forming domain-swapped dimers. Once Domain B has been characterized, Domain A activity can be restored and modified for the target of interest.

The first amino acid of wt-CVN is Leu. However, genes inserted into the T7 phage gene 10B must be compatible with digestion by EcoRI and HindIII on the 5' and 3' ends respectively. EcoRI's restriction site is GAATTC, leaving AATT as the 5' sticky end, and requiring CV-N to begin with Asn-Ser. Leu was replaced with Ser, and Asn was added to the beginning of the genes. Likewise, the 3' end of the CV-N library genes ended with the HindIII restriction site AAGCTT.

The crystal structure of P51G-m4-CVN (PDB: 2Z21) alone and P51G-m4-CVN (PDB: 2PYS and 2RDK) bound to Manα1-2Manα were previously solved (Fromme et al., 2007). Ten amino acids in Domain B are involved in binding dimannose, including 41, 42, 44, 52, 53, 56, 57,

74, 76, and 78. Residues 42, 57, and 78 are conserved in CVN-homolog sequence alignments (Percudani et al., 2005) and were not mutated in this library. Gly41 was conserved among CVNH, but wt-CVN contains Glu41. Because wt-CVN Glu41 is not consistent with CVNH conserved Gly41, the position was randomized in the T7 phage display library. The remaining six positions (44, 52, 53, 56, 74, 76) were also randomized in the library.

METHODS

T7 Phage Display Library. The P51G-m4-CVN T7 phage display library was created by randomizing amino acids 41, 44, 52, 53, 56, 74, and 76 in the Domain B active site of P51G-m4-CVN. The previously described P51G-m4-CVN (Fromme et al., 2007) is the "wild-type" protein for the library. It contains mutations K3N, T7A, E23I, and N93A (collectively called m4) in Domain A that remove Manα1-2Manα binding ability (L. Chang, 2002). P51G (Mori et al., 2002) stabilizes the protein. To accommodate ligation of library genes into T7 vector arms, Asn was added to the beginning of the gene and the first amino acid was mutated from Leu to Ser. The amino acid sequence is as follows, with x representing points of mutation:

SGNFSQACYNSAIQGSVLTSTCIRTNGGYNTSSIDLNSVIxNVxGSLKWQGxxFIxTCRNTQLAGS
SELAAECxTxAQQFVSTKINLDDHIAAIDGTLKYELEHHHHHH

GenScript (Piscataway, NJ) used site-directed mutagenesis to produce the library of genes. Each of the randomized position was represented by the codon NNS, where N incorporates A, T, G, or C, and S incorporates G or C. The genes were delivered as PCR product with 7-9 extra base pairs on each end to facilitate restriction enzyme digestion.

Library construction was performed using the T7Select 1-1b Cloning Kit and T7Select Packaging Kit (Novagen) and the protocols below are based on Novagen's kit protocols. CV-N library genes were digested with EcoRI and HindIII (New England Biolabs). The digestion was performed with 20 µL library DNA (50 ng/µl), 24 µL nuclease free water, 1 µL HindIII, and 5 µL 10x NEBuffer2. The reaction was incubated for 40 min at 37°C followed by addition of 1 µL EcoRI and incubation at 37°C for an additional 30 min.

After digestion, the genes were mixed with 5x loading dye (QIAGEN) and run in TAE (40 mM Tris, 19.4 mM acetic acid, 1 mM EDTA, pH 8.5) 1% agarose gel electrophoresis for 1.5-2

hours at 80v. DNA was extracted from the gel using the Promega Wizard SV Gel and PCR Clean-Up System. T4 DNA ligase (New England Biolabs) was used to ligate the P51G-m4-CVN library genes into T7 vector DNA at a molar ratio of 8:1 insert:vector with the reaction being incubated for 16 hours at 16°C.

After ligation, the genes were packaged into T7 phage particles using T7Select Packaging Extract. 5 µL of ligation reaction was added to 25 µL of packaging extract and incubated at room temperature for 2 hours. The reaction was stopped by adding 270 µL LB.

Plaque Assay. Plaque assay was performed on the initial phage library and again after amplification. Agar plates containing 50 µg/mL carbenicillin were prepared and incubated at 37°C for one hour prior to use. LB and phage were mixed to give dilutions of 1:10^3 to 1:10^6 for initial recombinant phage and 1:10^{11} to 1:10^{14} for amplified phage library. 100 µL of each phage dilution were mixed with 3 mL top agarose (1 g tryptone, 0.5 g yeast extract, 0.5 g NaCl, 0.6 g agarose, 100mL) and 250 µL mid-log BLT5615 (add IPTG, 1 mM final concentration, 30 minutes prior to use).

The library was amplified by the liquid lysate method at the 500 mL scale (Novagen, 2002). 500 mL BLT5615 was grown up with 50 µg/mL carbenicillin and induced with IPTG (1 mM final concentration) 30 minutes before use. The phage library was added to the culture, with 100-1000 cells per pfu. The culture was incubated at 37°C and 250 rpm for a few hours until lysis was observed. The lysate was centrifuged for 10 minutes at 8000xg and the supernatant was stored at 4°C.

Biopanning. Selection of the phage display library was performed using a polystyrene 96-well plate (Nunc). The plate was washed a few times with water and tapped dry on a paper towel. The two targets used for biopanning were gp120 (NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 gp120 CM, Cat # 2968) and RNase B (Sigma). gp120 was used for four rounds followed by one round of RNase B and then one more round of gp120. gp120 (11µg/mL) in TBS pH 7.4 (10mM Tris, 0.15mM NaCl), RNase B (10µg/mL in TBS pH 7.4), and TBS-Tween 0.1%, pH 8.0, were prepared. 100 µL of target was added to one well of the plate. The plate was covered and incubated for 3 hours at room temperature. After incubation,

the well was washed three times with TBS (pH 7.4). 200 µL blocking buffer (5% BSA [w/v] in water) was added to each well. The plate was covered and stored overnight at 4°C, washed six times with water, and stored until use at 4°C with 200 µL of water in each well.

The water was removed from the 96-well plate and 60 µL phage library (2.08×10^{12} pfu/mL) and 40 µL TBST pH 8.0 were added to the well and incubated for 1 h at room temperature followed by six washes with TBST. After washing, 200 µL T7 Elution Buffer (10.6 mM Tris, 35.4 mM SDS, pH 7.5) was added to the plate. It was covered and incubated for 20 minutes at room temperature. The T7 Elution Buffer was removed by pipette and added to 50 mL of mid-log BLT5615 (add IPTG, 1 mM final concentration, 30 minutes prior to use). The bacteria and phage were incubated at 37°C and 250 rpm until lysis was observed. The culture was centrifuged at 7740xg (r_{max}) for 10 minutes and the supernatant was saved at 4°C. Plaque assay was performed using phage dilutions $1:10^{10}$ to $1:10^{14}$.

Phage Plaque Sequencing. After six rounds of biopanning, nine plaques per round were selected from the plaque assay plates of rounds 3-6. Each plaque was scraped from the agar plate and prepared for PCR according to the Novagen protocol (Novagen, 2002). Harvested plaques were suspended in 100 µL of 10 mM EDTA, pH 8.0, vortexed, and then incubated at 65°C for 10 minutes. The suspended plaque was centrifuged for 3 minutes. 2 µL of phage lysate was combined with 5 µL Nova*Taq* Buffer (with $MgCl_2$), 1 µL T7SelectUP Primer (5 pmol/µl), 1uL T7SelectDOWN Primer (5 pmol/µl), 1 µl dNTP mix (10 mM of each nucleotide), 1.25 U Nova*Taq* DNA polymerase, and water up to 50 µL total volume. PCR amplification was performed for 35 cycles at 94°C for 50 sec, 50°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 6 minutes. 10 µL of each sample was run on a TAE 1% agarose gel to check its length and make sure there was a CV-N gene in the phage vector. The DNA was also sequenced using the T7Select UP and DOWN primers (Novagen) at the Arizona State University DNA Laboratory (School of Life Sciences).

RESULTS

The phage display library has a theoretical size of 1.28×10^9 different proteins. This is calculated by X^n where X is the number of amino acids possible at that position and n is the number of positions mutated. Based on plaque assay, the initial size of the T7 phage display library was 1.41×10^6 pfu. The initial library was amplified to provide additional copies of each mutant and a larger working volume for the library.

Following six rounds of selection, nine plaques were sequence from the plaque assay rounds 3-6. Of 36 plaques surveyed, two did not have a CV-N gene, and seven unique CV-N sequences were identified. Two of the unique sequences contained stop codons and five were full length sequences. Figures 9-12 show all of the isolated sequences and the parent P51G-m4-CVN for comparison.
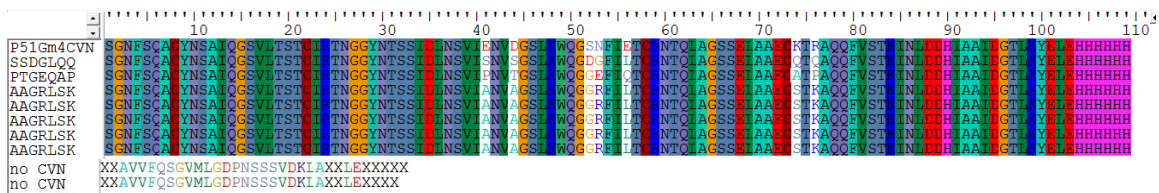

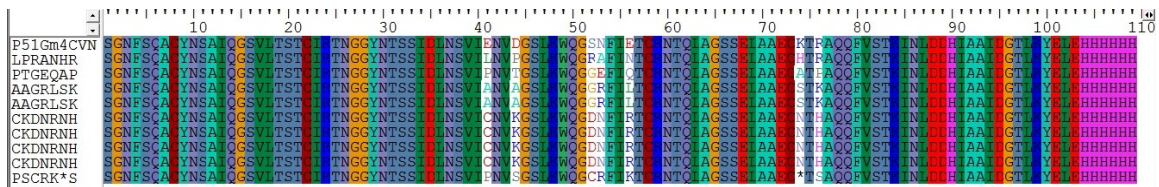*Figure 9*. Mutant CV-N protein sequences isolated from biopanning round 3 (binding gp120)

Figure 10. Mutant CV-N protein sequence isolated from biopanning round 4 (binding gp120)
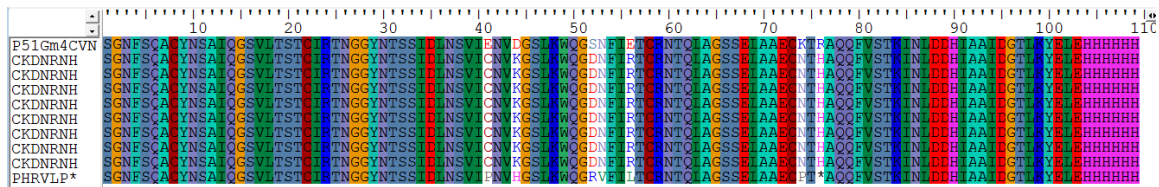
Figure 11. Mutant CV-N protein sequences isolated from biopanning round 5 (binding RNase B)

22
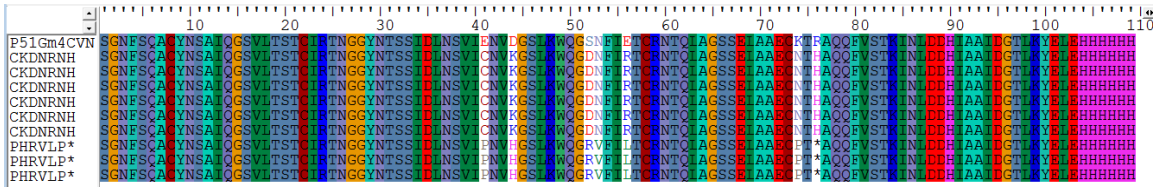
Figure 12. Mutant CV-N protein sequences isolated from biopanning round 6 (binding gp120)

Round 3 gave three different sequences, round 4 gave five sequences, round 5 gave two sequences, and round 6 also gave two sequences. Round four had the greatest sequence diversity, followed by a sharp drop in diversity when the library was selected against RNase B. CKDNRNH-P51Gm4-CVN was the only full length sequence selected from rounds 5 (RNase B) and 6 (gp120), and was the most frequently identified sequence found in 50% of the 36 plaques surveyed. Interestingly, PHRVLP*-P51Gm4-CVN contains a stop codon, but was isolated and enriched above most of the full length sequences. All five of the unique full length sequences are compared to P51G-m4-CVN in Figure 13 and their occurrences are described in Table 1.
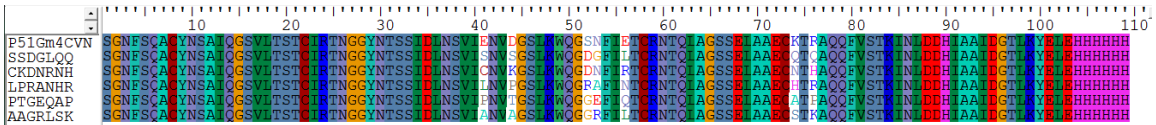
Figure 13. All unique full length sequences compared with P51G-m4-CVN parent sequence.

Table 1

Occurrences of Unique CV-N Sequences in Biopanning Rounds 3-6

CVN gene	Round of Biopanning				Total Occurrences
	3	4	5	6	
CKDNRNH	0	4	8	6	18
AAGRLSK	5	2	0	0	7
PHRVLP*	0	0	1	3	4
PTGEQAP	1	1	0	0	2
SSDGLQQ	1	0	0	0	1
LPRANHR	0	1	0	0	1
PSCRK*S	0	1	0	0	1
No CVN	2	0	0	0	2

As Figure 13 shows, there is very little sequence conservation from the parent protein to the mutants. Likewise, there is not much sequence similarity among the five full length proteins. The limited sequence conservation occurs at position 52 where Asp and Gly occur twice each and in residue 56 where Leu occurs twice. There is no sequence conservation among the

mutants at any other position. When comparing mutant residues to P51G-m4-CVN, the only sequence conservation is in CKDNRNH-P51Gm4-CVN at Asn53 and LPRANHR-P51Gm4-CVN at Arg76.

Table 2 compares the charged residues of P51G-m4-CVN and the full length CV-N mutants. For simplicity, only the mutated positions are shown. Residues 41, 44, 56, 74, and 76 of P51G-m4-CVN are charged, while residues 52 and 53 are polar residues. However, the mutants show an almost opposite pattern. Residues 52, 53, and 76 tend to have charged residues while residues 41, 44, 56, and 74 tend to have non-charged residues. Except for conservation of charged residues at position 76, the trend is reversed from P51G-m4-CVN to the mutants.

Table 2

Charged Residues at Amino Acids 41, 44, 52, 53, 56, 74, and 76 of Mutated CV-N

E	D	S	N	E	K	R
41	44	52	53	56	74	76
S	S	D	G	L	Q	Q
A	A	G	R	L	S	K
C	K	D	N	R	N	H
L	P	R	A	N	H	R
P	T	G	E	Q	A	P

Note: Charged residues are in green, P51G-m4-CVN is the top sequence and the five full length library mutants are below the position numbers.

Mutated residues were also compared based on side chain size. Smaller residues were considered Ala, Ile, Leu, Val, Asn, Cys, Ser, Thr, Asp, and Gly. Larger residues are Phe, Trp, Tyr, Gln, Met, Glu, Arg, His, Lys, and Pro. Table 3 compares parent protein P51G-m4-CVN residues with those of the library mutants. Positions 44, 52, and 53 are smaller amino acids in the parent protein and also tend to be small in the mutants. Position 76 is a larger residue both in the parent and mutant proteins. However, position 41 was large in the parent, but four of five mutants have smaller residues. Similarly, E56 and K74 in P51G-m4-CVN were now mostly smaller residues in the mutants with three of five mutants in each case having smaller residues.

Table 3

Amino Acid Side Chain Size Comparison at Positions 41, 44, 52, 53, 56, 74, and 76 in P51G-m4-CVN and CV-N Mutants

E	D	S	N	E	K	R
41	**44**	**52**	**53**	**56**	**74**	**76**
S	S	D	G	L	Q	Q
A	A	G	R	L	S	K
C	K	D	N	R	N	H
L	P	R	A	N	H	R
P	T	G	E	Q	A	P

Note: Small residues are in blue, P51G-m4-CVN is the top sequence and the five full length library mutants are below.

Figure 14 shows a PyMOL model of the domain B binding pocket for each full length

mutant. The models were created by superimposing mutations on the previously solved P51G-

m4-CVN crystal structure (PDB 2Z21) (Fromme et al., 2007). The actual crystal structure of

P51G-m4-CVN is also included for comparison.
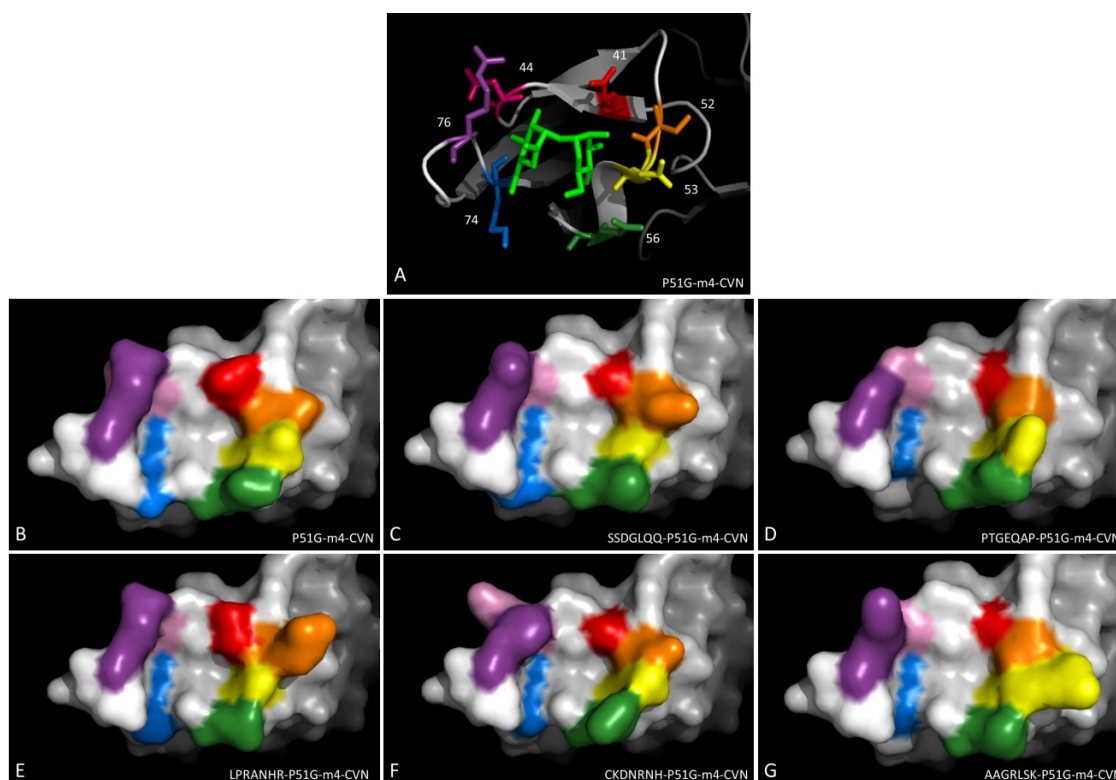
Figure 14. PyMOL models of Domain B library mutations using the P51G-m4-CVN crystal structure (PDB 2PYS). Amino acids 41 are red, 44 are pink, 52 are orange, 53 are yellow, 56 are green, 74 are blue, and 76 are purple. A) P51G-m4-CVNbound to dimannose, B) P51G-m4-CVN, C) SSDGLQQ-P51Gm4-CVN, D) PTGEQAP-P51Gm4-CVN, E) LPRANHR-P51Gm4-CVN, F) CKDNRNH-P51Gm4-CVN, G) AAGRLSK-P51Gm4-CVN.

25

wt-CVN and P51G-m4-CVN have deep Domain B binding pockets that hug the sugar, and cap it in place with Arg76 (Margulis, 2005). The PyMOL images in Figure 14 are possible Domain B conformations for the library mutants. Some mutated amino acids are shorter than those found in 51G-m4-CVN, indicating that some of the binding pockets may be shallower than the wild-type conformation.

The two mutants with stop codons were also modeled in PyMOL. PSCRK*S-P51Gm4-CVN contains a stop codon at position 74 and PHRVLP*-P51Gm4-CVN has a stop codon at 76. These proteins were not completely expressed because of the stop codons. Figure 15 shows truncated versions of the crystal structure of P51G-m4-CVN (PDB 2PYS). The protein is shown at 73 amino acids, 75 amino acids, and the full 101 amino acid versions. Truncation at 73 and 75 eliminates part of the Domain B binding site and half of the second sequence repeat, including the β-hairpin that associates with Domain A.
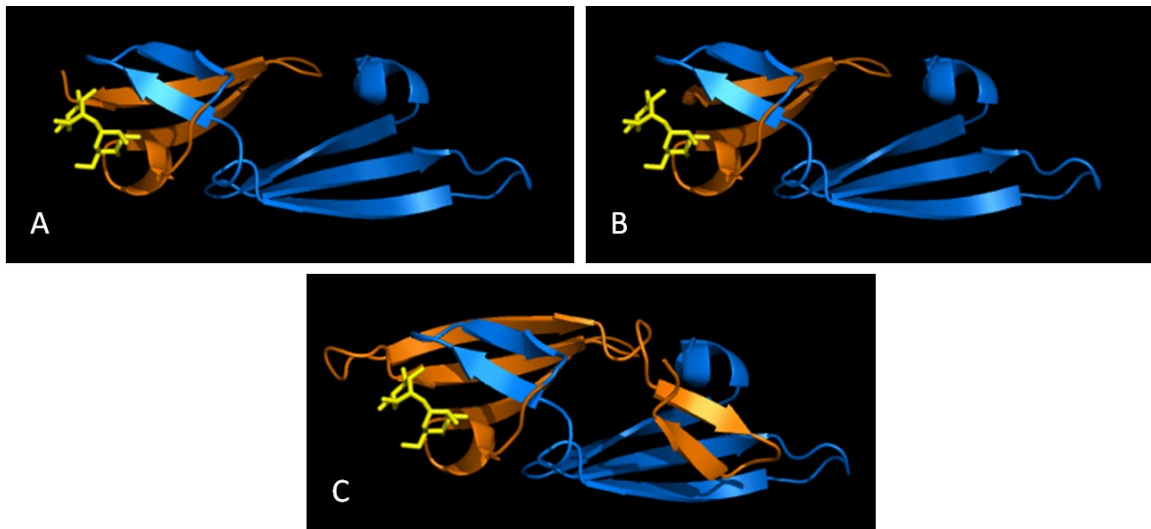
Figure 15. Full length P51G-m4-CVN and truncated mutants. Whole protein view comparison of truncated mutants using the crystal structure of P51G-m4-CVN bound to 2α-Mannobiose (PDB: 2PYS). A) P51G-m4-CVN 75 amino acids long, B) P51G-m4-CVN 73 amino acids long, C) P51G-m4-CVN full length monomer.

DISCUSSION

Biopanning of the T7 phage display library yielded seven unique CV-N sequences, five of them full length and two with stop codons mid-sequence. Of the 36 sequences sampled (nine plaques each of rounds 3-6), round 3 gave three different sequences, round 4 gave five

sequences, round 5 gave two sequences, and round 6 also gave two sequences. There were also two sequences in round 3 that did not contain a CV-N gene. Native T7 phage in the library is likely due to incomplete digestion of vector arms prior to insertion of library genes. Since there were no native phage observed in rounds 4-6, it is assumed that this species was reduced or eliminated from the pool because it didn't specifically bind to any motif on gp120 or RNase B.

More rounds of selection were performed with this library than is typical for phage display (Novagen, 2002; Smith & Petrenko, 1997). Two to four rounds of biopanning are usually performed. As Smith and Petrenko pointed out, there is a balance between stringency and yield. If selection is too strict, then some of the desirable proteins will be lost. In this case, the number of rounds was too high and the phage population was nearly homogeneous. Greater sequence diversity was observed in the first four rounds of selection than in the last two rounds. By the fourth round of biopanning, CKDNRNH-P51Gm4-CVN was significantly enriched and selection against RNase B in the fifth round shifted the population to almost exclusively favor this mutant. Because the sample diversity was so low for rounds five and six, genes from rounds three and four were also sequenced.

Before selecting against RNase B, library mutants were being selected for their ability to bind any part of gp120. There are several different glycans on gp120 and more stringent conditions are required to select for binding to a single glycan type. Making RNase B the target in round two instead of round five would provide a more specific selection for Man_9 while retaining some diversity in the population. Since the only common motif between gp120 and RNase B is Man_9, there should be a low risk of selecting sequences specific for other portions of RNase B or gp120 in the positive control selection.

The five viable protein sequences show limited residue conservation compared to themselves and to P51G-m4-CVN. Table 2 examines the conservation of charged residues and the mutants have an almost opposite configuration of charged versus uncharged residues. P51G-m4-CVN has charged residues at 41, 44, 56, 74, and 76, and polar residues at positions 52 and 53. Three of the mutants also have charged residues at 76, but none at 41, one at 44, three at 52, two at 53, and one each at 56 and 74.

27

Amino acid size was compared for the mutated positions. E41 in P51G-m4-CVN was not conserved in CVNH and is not conserved in the sequences selected by phage display. In both cases, amino acids smaller than Glu are present. Since Gly41 is strictly conserved among CVNH, it is not surprising that the phage display library would also favor smaller residues. Fromme and Margulis describe the importance of steric specificity at E41 for CV-N binding to dimannose. Though small amino acids are conserved at position 41, none of the mutants have retained Glu. Binding studies will shed additional light on the importance of E41 for CV-N's lectin specificity.

Among P51G-m4-CVN and the selected sequences, position 76 is almost exclusively larger, charged amino acids. This suggests that a bulky, charged residue at this position is important to Domain B's ability to bind glycans. Conservation of large, charged amino acids at position 76 agrees well with computational studies (Margulis, 2005) and crystal structures of P51G-m4-CVN (Fromme et al., 2008) that demonstrate a cap and lock mechanism. Among residues 44, 52, and 53, there seems to be a preference for smaller amino acids in agreement with P51G-m4-CVN. Finally, positions 56 and 74 were bulkier amino acids in P51G-m4-CVN, but three of five mutants at each position favor smaller residues. Overall, the selected sequences suggest that Domain B favors smaller, uncharged residues at position 41 and bulky, charged residues at position 76.

Of the seven unique sequences isolated by biopanning T7 phage library, two contain stop codons. PSCRK*S-P51Gm4-CVN and PHRVLP*-P51Gm4-CVN each contain a stop codon at positions 74 and 76 respectively. All four cysteines are still present at positions 8, 22, 58, and 73. However, Figure 15 shows that approximately half of the second sequence repeat is missing as is the β-hairpin that normally associates with Domain A. A third of Domain B's β-sheet is also missing. With 26-28 residues missing in a 101 amino acid protein, the protein is probably not folded correctly and may exhibit some non-specific binding. These truncated sequences were not studied any further.

CHAPTER 3

CYANOVIRIN-N MUTANT EXPRESSION AND PURIFICATION

ABSTRACT

Five full length CV-N sequences were isolated from the T7 phage display library. These genes were modified and inserted into pET26b+ expression vector and BL21(DE3) *E. coli* for protein expression. Each purified protein was examined using MALDI-TOF MS to compare the actual and calculated molecular weights. All five sequences were expressed with an N-terminal *pelB* leader sequence, which targets the protein to the periplasm, and a C-terminal His-tag. The CV-N mutant genes were also PCR modified to contain an N-terminal His-tag and (TEV) cleavage site. Under typical culture conditions, a large amount of CV-N in each BL21-(DE3) preparation was located in inclusion bodies, with a smaller amount going into the soluble fraction. The predominant localization of protein to inclusion bodies is consistent with other CV-N mutants expressed in this lab. Cold growth with benzyl alcohol promotes increased production of soluble, folded protein in the periplasmic fraction and eliminates the need for purification under denaturing conditions. SSDGLQQ-P51Gm4-CVN was the most studied of the mutants, and a SSDGLQQ-wt-CVN mutant was also created to restore anti-HIV activity. AAGRLSK-P51Gm4-CVN and PTGEQAP-P51Gm4-CVN were isolated from the soluble cell fraction and both are soluble under native conditions. CKDNRNH-P51Gm4-CVN was the only full length sequence identified in rounds five and six of phage library selection, but CKDNRNH-P51Gm4-CVN and LPRANHR-P51Gm4-CVN are generally insoluble under native conditions and it was very difficult to obtain any soluble sample of either protein.

INTRODUCTION

Protein Expression. Proteins of interest can be recombinantly produced in a variety of biological systems including bacteria, yeast, insect and mammalian cells. Each system has its advantages and disadvantages for expressing recombinant proteins, but *E. coli* is the most commonly used and best studied system.

Cells use signal sequences to target proteins to specific intracellular locations. A system common in recombinant protein expression is the 22 amino acid, N-terminal *pelB* leader

sequence expressed with the protein and enzymatically cleaved upon the protein's arrival in the periplasm (Baneyx, 2004). *pelB* interacts with SecB, which transports the partially folded protein to the cytoplasmic membrane for export to the periplasmic space. Within the periplasm, a variety of chaperones are available to promote proper disulfide bond formation and protein folding. Because correct disulfide bond formation is critical to wt-CVN's anti-HIV function (Mori et al., 1997), many protocols target the protein to and extract it from the periplasmic cell fraction (Barrientos et al., 2004; Barrientos et al., 2006; Boyd et al., 1997; Mori et al., 1998; Mori et al., 1997). Though many groups isolate CV-N from the periplasm, CV-N tends to localize in insoluble inclusion bodies as evidenced by high protein yields of 40 mg/L or more when protein is recovered under denaturing conditions (Colleluori et al., 2005; Fromme et al., 2007; Xiong et al., 2010). wt-CVN, P51G-m4-CVN, S52P-m4-CVN, and ΔQ50-m4-CVN have been successfully isolated from inclusion bodies (Barrientos, Louis, Botos, et al., 2002; Barrientos, Louis, Hung, et al., 2002; Fromme et al., 2007; Y. Liu et al., 2009).

Several strategies are available to promote or increase formation of soluble recombinant protein in *E. coli*, including cooler culture temperature and addition of benzyl alcohol (de Marco, Vigh, Diamant, & Goloubinoff, 2005; Horvath et al., 1998; Shigapova et al., 2005). Benzyl alcohol makes the cellular membrane more fluid and induces production of chaperone proteins (de Marco et al., 2005; Horvath et al., 1998; Shigapova et al., 2005). These molecular chaperones help refold misfolded and aggregated recombinant protein, yielding a higher amount of soluble, folded protein.

His-tag Cleavage. Since there are many different proteins expressed in each bacterial cell, protein tags are often used to help isolate the recombinant protein of interest. One of the most commonly used tags is the six histidine, or 6His-tag. Placed on the N or C terminus, this tag has a high affinity for Ni^{2+} and other metal ions. At times it is also advantageous to remove the tag after the protein has been purified since tags can detrimentally affect protein folding (Kapust, Tozser, Copeland, & Waugh, 2002). Several proteases are available to cleave 6His-tags. These enzymes recognize specific protein sequences, which may be included between the His-tag and protein sequences. A protease with high sequence specificity is tobacco etch virus protease

(TEV), which recognizes the seven residues Glu-Asn-Leu-Tyr-Phe-Gln-Gly/Ser, with cleavage occurring between Gln and Gly/Ser (Kapust et al., 2002). Some TEV constructs also have a His-tag and will bind to a Ni column to facilitate purification of cleaved protein from the TEV and cleaved tags.

METHODS

Non-cleavable C-terminal His-tag Mutants. The CV-N genes isolated from the selection needed some modifications before they could be properly expressed in *E. coli*. The restriction sites EcoRI and HindIII were used for phage display, but were not ideal for protein expression. All five of the full length CV-N genes isolated in T7 phage selection were modified to include the NdeI restriction site and *pelB* leader sequence on the 5' end and restriction site XhoI at the 3' end. This gene design has been described previously (Barrientos, Louis, Botos, et al., 2002; Fromme et al., 2007; Mori et al., 1998). SSDGLQQ-P51Gm4-CVN, AAGRLSK-P51Gm4-CVN, PTGEQAP-P51Gm4-CVN, CKDNRNH-P51Gm4-CVN, and LPRANHR-P51Gm4-CVN genes were synthesized by GenScript and arrived in the pUC57 cloning vector. For each gene, 1 µg of DNA was digested with NdeI and XhoI (New England Biolabs) and purified by 1% agarose 1xTAE (1 mM EDTA, 1 M glacial acetic acid, 40.0 mM Tris, pH 8.5) gel electrophoresis. The CV-N gene fragment was extracted from the gel using Promega's Wizard SV Gel and PCR Clean-up System. pET26b+ (Novagen) was digested and purified in the same manner. Ligation was performed using T4 DNA ligase (New England Biolabs), 2 µL 10x T4 DNA Ligase Reaction Buffer, 100 ng pET26b+ and 15 ng CV-N genes in a 20 µl total volume. The reaction was incubated for 10 minutes at room temperature before transformation into BL21-(DE3) cells (Stratagene). Transformed cells were plated on agar containing 15 µg/mL kanamycin and grown overnight at 37°C. Several colonies were grown in liquid culture to make glycerol stocks and verify the CV-N gene sequence. 10 mL overnight culture of transformed cells was treated with Promega Wizard Plus SV Miniprep DNA Purification System to extract the DNA. The DNA was sent to Arizona State University's DNA lab for sequencing.

PCR Modification for N-terminal His-tag, TEV cleavable Mutants. In order to study the mutants without a His-tag, the CV-N genes were modified to include an N-terminal His-tag and

TEV cleavage site, and the *pelB* leader sequence was removed. Three primers were designed to PCR modify the mutant genes. The starting gene was the form isolated from the phage library with EcoRI and HindIII restriction sites and the C-terminal His-tag. Three overhanging primers were designed, requiring two sequential PCR reactions. The PCR primers are as follows (from Integrated DNA Technologies):

Outside Forward 5'- AGTCAACATATGCACCATCACCATCACCATGAGAACCTGTAC - 3'; Inside Forward 5' - CATCACCATGAGAACCTGTACTTCCAGTCAGGCAACTTTAGCCAG - 3'; Reverse 5' - TGACCTGACCTCGAGTCATTCATATTTCAGGGTGCCATCG - 3'.

The first PCR reaction used primers Reverse and Inside Forward. A 50 µl reaction was prepared including 1 µl dNTP (10 mM each base), 2.5 µl of each primer (10 µM), 1 ng template DNA, 0.5 µl Phusion Hot Start II DNA polymerase (New England Biolabs), 10 µl HF buffer, and water. The reaction proceeded at 98°C for 30 seconds followed by 35 cycles of 98°C for 30 seconds, 52.3°C for 30 seconds, and 72°C for 30 seconds. The reaction was held at 72°C for 6 minutes followed by storage at 4°C. After PCR, the reaction was run on a 1% agarose TAE gel to purify the gene and confirm its size. The modified CV-N gene was extracted from the gel, as described above, and used for the second PCR reaction.

The second PCR reaction called for primers Reverse and Outside Forward. The template DNA was the purified product of the first PCR reaction, and the remaining reagents were the same as described above. The reaction proceeded at 98°C for 30 seconds followed by 35 cycles of 98°C for 30 seconds, 52.9°C for 30 seconds, and 72°C for 30 seconds. The reaction was held at 72°C for 6 minutes followed by storage at 4°C. The reaction was gel purified and gene size was confirmed.

After the second PCR modification, the CVN genes and pET26b(+) were digested with NdeI and XhoI. The digested gene was purified by Promega's Wizard SV Gel and PCR Clean-up System while the vector was purified by agarose gel electrophoresis. The modified CV-N gene and pET26b(+) were ligated as described above and the plasmid was transformed into BL21-(DE3) cells.

SSDGLQQ-wt-CVN. After mutant SSDGLQQ-P51Gm4-CVN was expressed in *E. coli* and characterized, the Domain B mutations were made to wt-CVN without P51G or the four Domain A inactivating mutations (L. Chang, 2002). The SSDGLQQ-wt-CVN gene was designed to include an N-terminal His-tag and TEV cleavage site. GenScript produced the gene and sent it in pUC57. I ligated the gene into pET26b+ and transformed the plasmid into BL21(DE3) as described previously.

Protein Expression and Purification from Inclusion Bodies. This protocol was used for wt-CVN, P51G-m4-CVN, all N-term His-tag library mutants, SSDGLQQ-wt-CVN, and all C-term His-tag library mutants. An agarose plate containing 15 μg/mL kanamycin was streaked from a bacteria stock and grown overnight at 37°C. A single colony was picked and grown overnight in 10 mL LB pH 7.0 at 37°C and 250 rpm. The overnight culture was added to 1L of LB and grown at 37°C and 250 rpm until OD ~1.0-1.5, followed by addition of IPTG at 1 mM final concentration. The culture was further incubated for 3-4 hours at 37°C and 250 rpm, or overnight at 30°C and 200 rpm. The cells were harvested by centrifugation at 4500xg and the supernatant was discarded. The pellet was used immediately or stored at -20°C until use.

The cell pellet was resuspended in 100 mM Tris, pH 7.0, and sonicated on ice for 5 minutes followed by 5 minutes rest. The sonication was done four times total and the lysate was centrifuged at 17000xg for 30min. The supernatant was discarded and the pellet was resuspended in clarifying buffer (100 mM Tris, 2 M urea, pH 7.0). The pellet was again centrifuged at 17000xg for 30 minutes and the clarifying buffer step was repeated. The pellet was then resuspended in 100 mM Tris, pH 8.0, followed again by 30 minutes of centrifugation. The final resuspension was in denaturing binding buffer (8 M urea, 0.5 M NaCl, 20 mM imidazole, 20 mM Tris, pH 7.5) and the supernatant was filtered in preparation for Ni affinity chromatography.

A 5 mL GE Healthcare HisTrap HP Ni column was prepared by equilibrating the resin with denaturing binding buffer (8 M urea, 0.5 M NaCl, 20 mM imidazole, 20 mM Tris, pH 7.5). The filtered protein was applied to the column followed by wash with denaturing binding buffer. The protein was then eluted by denaturing elution buffer (8 M urea, 0.5 M NaCl, 500 mM imidazole, 20 mM Tris, pH 7.5).

Following the Ni column, fractions were run on SDS-PAGE to determine protein purity and location. Fractions containing CV-N were pooled, DTT (5 mM final concentration) was added and the sample was incubated for at least 30 minutes at room temperature. The protein was dialyzed in 2L buffer (2M urea, 50mM Tris, pH 8.0) for 24 hours at room temperature, then in 2L Tris buffer (100 mM NaCl, 10mM Tris, pH 8.0) for 48 hours at room temp or 4°C. After 48 hours, the protein sample was moved to a fresh 2L of Tris buffer for another 48 hours.

To remove any remaining impurities and remove any dimeric protein, the protein was run through a GE HiLoad 16/600 Superdex 75pg gel filtration column attached to an Agilent 1260 Infinity bio-inert system under isocratic conditions in PBS, pH 7.4.

Protein Purification from the Soluble Cytoplasm Fraction. This protocol was used for AAGRLSK-P51Gm4-CVN (C-term 6His-tag), CKDNRNH-P51Gm4-CVN (C-term 6His-tag), PTGEQAP-P51Gm4-CVN (C-term 6His-tag) and LPRANHR-P51Gm4-CVN (C-term 6His-tag). Harvested cells were resuspended in 100 mM Tris, pH 7.0, and put on ice. They were lysed by sonication for four cycles of 1 minute of sonication followed by 1 minute rest. The cell lysate was centrifuged at 24000xg for 45 minutes. The supernatant was filtered before Ni column purification. A 5 mL GE Healthcare HisTrap HP Ni column was prepared by equilibrating the resin with native binding buffer (0.5 M NaCl, 20 mM imidazole, 20 mM Tris, pH 7.5). The supernatant was applied to the column followed by wash with native binding buffer. The protein was then eluted by native elution buffer (0.5 M NaCl, 500 mM imidazole, 20 mM Tris, pH 7.5) followed by analysis on SDS-PAGE.

Periplasmic Protein Expression and Extraction. This protocol was used for wt-CVN, P51G-m4-CVN, SSDGLQQ-P51Gm4-CVN (C-term 6His-tag), CKDNRNH-P51Gm4-CVN (C-term 6His-tag), and LPRANHR-P51Gm4-CVN (C-term 6His-tag). BL21(DE3) was grown to OD_{600} ~ 0.6 at 37°C and 250 rpm. Benzyl alcohol (20 mM final concentration) and 0.4 mM IPTG were added followed by 17 hours of incubation at 20°C. The cells were harvested and resuspended in 30 mM Tris, 20% sucrose, and 0.5 mM EDTA (pH 7.5) for 10 minutes at room temperature with stirring. The cells were collected by centrifugation at 10,000 rpm at 4°C, followed by resuspension in cold

5 mM MgSO$_4$ and stirring for 20 minutes on ice. The cells were centrifuged again and the supernatant contained the periplasmic fraction.

Protein localization for *pelB*. The protein localization protocol was taken from Novagen's pET System Manual (Novagen, 2002). Briefly, samples were taken of the total cell, medium, periplasm, soluble cytosol, and insoluble cytosol. Each sample was prepared by addition of one part sample, one part 2x SDS sample buffer, heated for three minutes at >85°C, and then stored at -20°C until they were run in an SDS-PAGE.

Protein localization for no *pelB*. The mutants with N-terminal His-tag and TEV cleavage site do not contain the *pelB* leader sequence and are therefore not likely to be found in the periplasm. A simplified protein localization protocol was used for these mutants. Harvested cells were either stored at -20°C until use or prepared directly after harvest. The media was discarded and the cell pellet was resuspended in 100mM Tris, pH 7.0, and sonicated three times for five minutes each, with five minutes between each time. The lysate was clarified by centrifugation at 24000xg for 30 minutes. The supernatant containing the soluble protein fraction was saved. The pellet was washed twice in 100mM Tris, 2 M urea, pH 7.0, and each time centrifuged at 24000xg for 30 minutes. The wash supernatant was saved. A third wash was performed with 100 mM Tris, pH 7.0 followed by centrifugation at 24000xg for 30 minutes. The supernatant was saved. Finally, the pellet was resuspended in denaturing binding buffer (8 M urea, 0.5 M NaCl, 20 mM imidazole, 20 mM Tris, pH 7.5) and incubated overnight at 4°C before centrifugation at 31000xg. The pellet was discarded and the supernatant was saved. 5 µl of supernatant from each step was mixed with 5 µl of 2x SDS Sample Buffer (Invitrogen) and heated for a few minutes and run on SDS-PAGE.

SDS-PAGE. Samples were prepared by mixing 5 µL of protein with 5 µL 2x Sample Buffer (Invitrogen) and heating the mixture to at least 80°C for 10-20 minutes and placed into a 16 % acrylamide SDS-PAGE. After running the gel, it was stained with Coomassie Brilliant Blue (0.25 g Coomassie Brilliant Blue R-250, 100 mL destaining solution) and destained (10% glacial acetic acid, 45% water, 45% methanol). The Mark 12 Unstained Standard (Life Technologies) was used as a molecular weight marker.

TEV cleavage. For CV-N mutants with an N-terminal, cleavable His-tag, the folded protein was incubated 1:100 with 2 mg/mL TEV (Sigma) or 1:1000 with TEV produced in-house. Following the reaction, the cleaved protein was separated from uncleaved-CVN, cleaved His-tags, and TEV by Ni column under native conditions. A 5 mL GE Healthcare HisTrap HP Ni column was prepared by equilibrating the resin with native binding buffer (0.5 M NaCl, 20 mM imidazole, 20 mM Tris, pH 7.5). The filtered protein was applied to the column followed by wash with native binding buffer. The protein was then eluted by native elution buffer (0.5 M NaCl, 500 mM imidazole, 20 mM Tris, pH 7.5). Cleaved protein was analyzed by SDS-PAGE.

MALDI-TOF MS. After a protein was purified, its molecular weight was assessed by MALDI-TOF MS. Sinapic acid was suspended in 50% acetonitrile and mixed 1:1 or 1:10 with 1-2 µl of protein sample and the mixture was spotted onto a 100 well stainless steel plate and analyzed by an Applied Biosystems Voyager System 4320.

RESULTS

SSDLQQ-m4-CVN was produced with two different His-tag configurations. Figure 16 shows the various protein constructs, with the His-tags in red, the *pelB* leader sequence in blue, and TEV recognition sequence in green. The first sequence shows the *pelB* leader sequence for illustrative purposes, but the purified protein is the second sequence, with the *pelB* sequence cleaved by the cellular enzymes upon arrival in the periplasmic space. The lower sequence shows the N-terminal His-tag and TEV cleavage site.

KYLLPTAAAGLLLLAAQPAMA – CV-N Gene – LEHHHHHH
CV-N Gene – LEHHHHHH
HHHHHHENLYFQ – CV-N Gene

Figure 16. Histidine tag and *pelB* leader sequence placement on CV-N mutants. The first line is the *pelB* sequence attached to the CV-N gene that has a C-term His-tag. The second line is the expressed version in which the *pelB* sequence has been cleaved from CV-N. The third sequence is the N-terminal His-tag with TEV cleavage site. The last residue required for TEV recognition is the S at the beginning of CV-N.

Following expression, various cell fractions were analyzed for the presence of CV-N. Figure 17 shows protein localization of N-terminal His-tag SSDGLQQ-P51Gm4-CVN. CV-N can be seen in the soluble cytoplasmic fraction and first wash step, but ~75% of the protein appears to be located in inclusion bodies as seen in lane 12 of the gel.
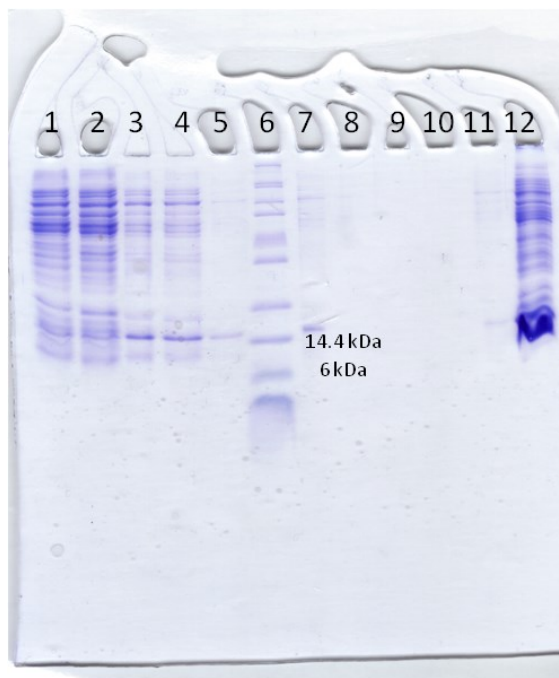
Figure 17. SDS-PAGE of protein localization for 12.4 kDa SSDGLQQ-P51Gm4-CVN with N-terminal His-tag. 1,2) soluble fraction; 3,4) first 2 M urea wash; 5,7) second 2 M urea wash; 6) Mark 12 standard from Invitrogen; 11) Tris wash; 12) Inclusion bodies 8 M urea fraction

For the C-terminal His-tag version of SSDGLQQ-P51Gm4-CVN, CV-N can be seen in Figure 18 in the soluble cytoplasmic portion and in inclusion bodies. The periplasmic and media fractions are very faint and it is hard to tell if there is any significant amount of CV-N in these fractions.
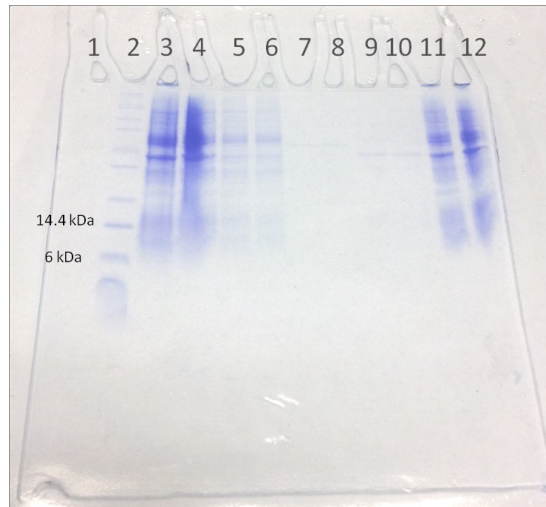
Figure 18. SDS-PAGE showing protein localization for 11.8 kDa C-terminal His-tag SSDGLQQ-P51Gm4-CVN. 2) mark 12 standard, 3,4) inclusion bodies, 5,6) soluble cytoplasmic fraction, 7,8) periplasmic fraction, 9,10) media, 11) Total cell before IPTG induction, 12) Total cell at time of cell harvest.

For proteins extracted from inclusion bodies, 20-40mg of CV-N per prep was usually isolated from Ni column purification. If the protein had an N-terminal TEV cleavage site, the protein was refolded into native conditions and cut by TEV. As assessed by SDS-PAGE, the enzyme cut 50%-70% of cut protein. The cut protein was purified by Ni column to remove uncut CV-N, cleaved tags, and TEV.

Because it was successfully purified from inclusion bodies in large quantities, SSDGLQQ-P51Gm4-CVN became the most studied of the five CV-N mutants. It can be refolded from inclusion bodies as an N-terminal and C-terminal His-tag construct, and the C-terminal 6His-tag construct can also be produced by the benzyl alcohol method.

AAGRLSK-P51Gm4-CVN (C-term 6His-tag) was purified from the soluble fraction. The N-term 6His-tag version was more difficult to refold from inclusion bodies and a limited amount of this protein was produced due to precipitation.

Unfortunately, even with C-term 6His-tag and benzyl alcohol expression conditions, CKDNRNH-P51Gm4-CVN and LPRANHR-P51Gm4-CVN were extremely difficult to purify as soluble proteins. CKDNRNH-P51Gm4-CVN was the only full length gene isolated from T7 phage display selection rounds five and six, and was the most frequently observed mutant in round four. CKDNRNH-P51Gm4-CVN with a C-term His-tag was not successfully isolated from the soluble

cytoplasmic fraction. This protein does not behave well and is quite difficult to prepare as a soluble protein.

LPRANHR-P51Gm4-CVN (N-term 6His-tag) and PTGEQAP-P51Gm4-CVN (N-term 6His-tag) were not successfully refolded from inclusion bodies and precipitated significantly under native conditions. PTGEQAP-P51Gm4-CVN (C-term 6His-tag) was successfully recovered from the soluble fraction.

P51G-m4-CVN, wt-CVN, and SSDGLQQ-P51Gm4-CVN (C-term 6His-tag) were produced using the benzyl alcohol method for periplasmic protein production. As seen in Figure 19, about one third of the protein was found in the periplasmic, soluble cytosolic, and inclusion body fractions. This is a significant increase in periplasmic protein production over the standard 37°C expression method.
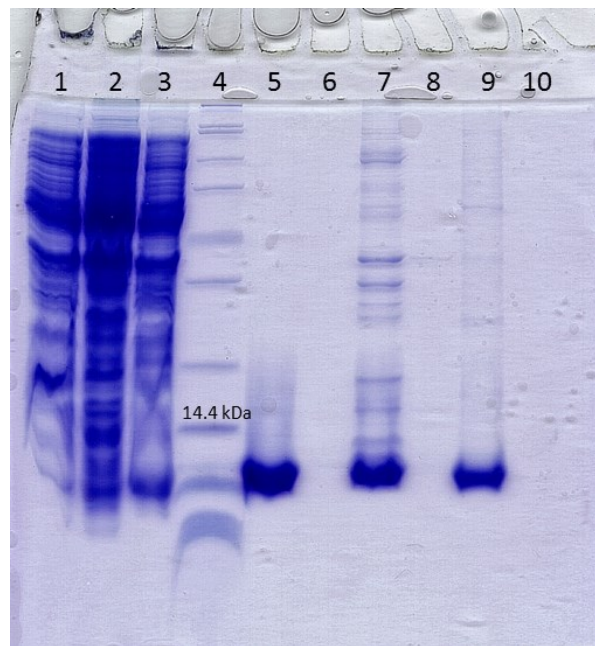
Figure 19. Gel of benzyl alcohol growth of P51G-m4-CVN with *pelB* and C-term 6His-tag. Lane 1 is the total cell before benzyl alcohol or IPTG. Lane 2 is the total cell with benzyl alcohol, but without IPTG. Lane 3 is total cell that has benzyl alcohol and IPTG. Lane 4 is the Mark 12 standard (Life Technologies). Ni column fractions 2 and 3 are in lanes 5 and 6 for the periplasmic fraction, lanes 7 and 8 for the soluble cytosolic fraction, and lanes 9 and 10 for inclusion bodies. Gel from Haiyan Sun.

MALDI-TOF MS was performed on purified CV-N mutants to ensure the protein size was in agreement with calculated values. Table 4 shows the calculated molecular weight and the observed molecular weights by MALDI-TOF MS.

Table 4

Calculated and Observed Protein Molecular Weight Values of CV-N Proteins

His-tag location	Protein	Calculated MW (Daltons)	MW by MALDI-TOF MS (Daltons)
C-terminal	SSDGLQQ-P51Gm4-CVN	11,765	11,727
	AAGRLSK-P51Gm4-CVN	11,733	11,695
	CKDNRNH-P51Gm4-CVN	11,917	11,920
	PTGEQAP-P51Gm4-CVN	11,730	11,733
	LPRANHR-P51Gm4-CVN	11,895	11,898
	P51G-m4-CVN	11,935	11,886
	wt-CVN	12,078	12,069
N-terminal	SSDGLQQ-P51Gm4-CVN	12,449	12,474
	AAGRLSK-P51Gm4-CVN	12,417	12,484
	PTGEQAP-P51Gm4-CVN	12,414	12,487
	CKDNRNH-P51Gm4-CVN	12,601	12,625
No His-tag (TEV cut)	SSDGLQQ-wt-CVN	10,843	10,869
	AAGRLSK-P51Gm4-CVN	10,668	10,693
	SSDGLQQ-P51Gm4-CVN	10,700	10,725
	CKDNRNH-P51Gm4-CVN	10,852	10,869

Note: MALDI-TOF MS was used to measure the protein samples.

Each of the proteins was analyzed by analytical gel filtration to assess its oligomeric state. Figure 20 shows a gel filtration profile for wt-CVN, P51G-m4-CVN, and each of the CV-N mutants. P51G-m4-CVN, wt-CVN, AAGRLSK-P51Gm4-CVN, SSDGLQQ-P51Gm4-CVN, and SSDGLQQ-wt-CVN were largely monomeric with large peaks between 36 and 40 minutes. CKDNRNH-P51Gm4-CVN was a very low concentration, but there appears to be a monomer and dimer peak, but it is hard to tell the predominant species from this gel filtration run. PTGEQAP-P51Gm4-CVN appears mostly dimeric with a large peak at 33 minutes. LPRANHR-P51Gm4-CVN has what appears to be a dimer peak at about 33 minutes, but the signal is relatively low.
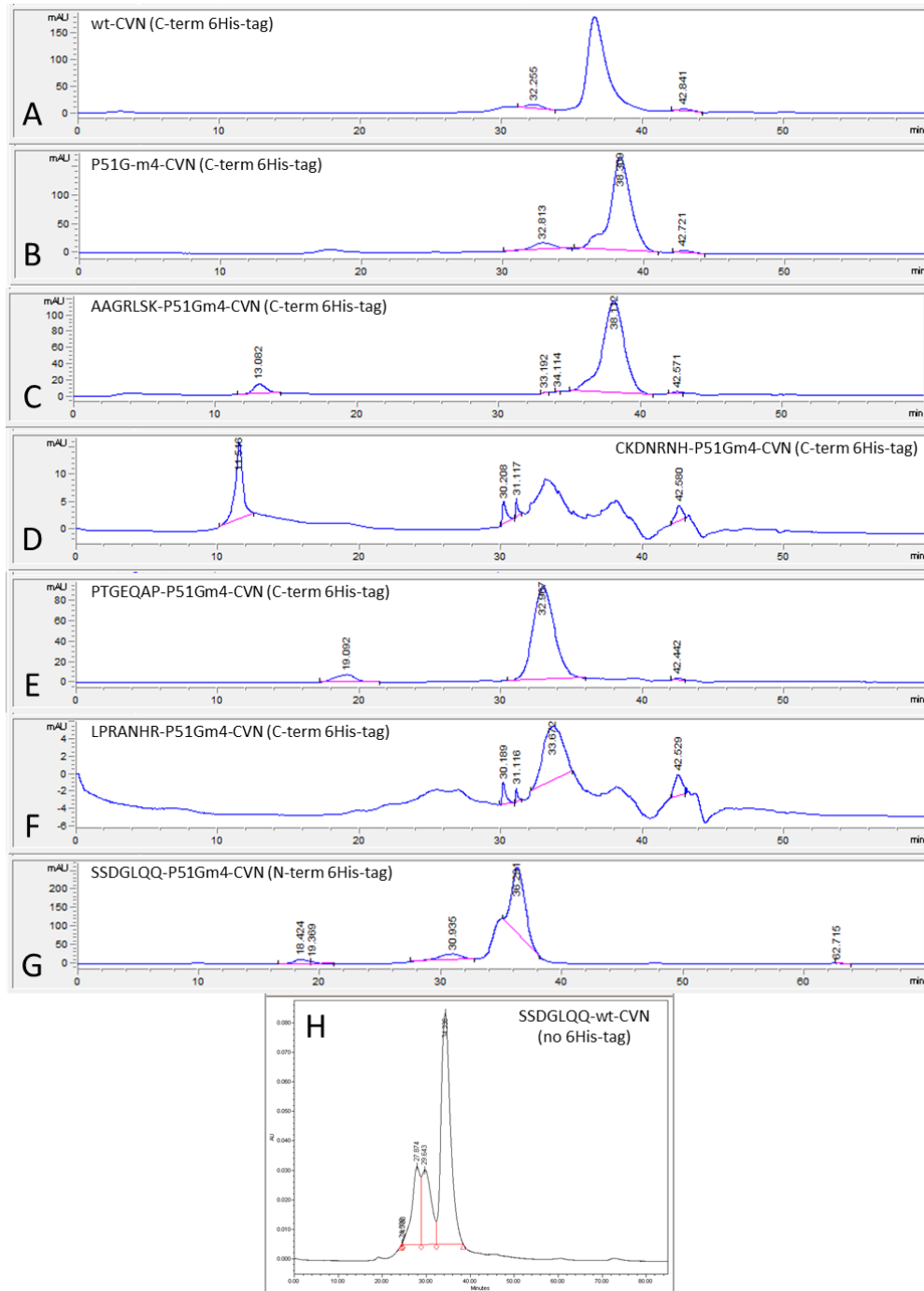
Figure 20. Gel filtration profiles of CV-N mutants on a Superdex 75 10/300 GL gel filtration column. In A-G, peaks between 36 and 40 minutes are monomeric protein and peaks near 33 minutes are dimeric protein. A) wt-CVN (C-term 6His-tag), B) P51G-m4-CVN (C-term 6His-tag), C) AAGRLSK-P51Gm4-CVN (C-term 6His-tag), D) CKDNRNH-P51Gm4-CVN (C-term 6His-tag), E) PTGEQAP-P51Gm4-CVN (C-term 6His-tag), F) LPRANHR-P51Gm4-CVN (C-term 6His-tag), G) SSDGLQQ-P51Gm4-CVN (N-term 6His-tag), H) SSDGLQQ-wt-CVN (no 6His-tag).

41

DISCUSSION

Five CV-N mutant genes were identified by selecting the T7 phage display library against gp120 and RNase B. A non-cleavable C-terminal 6His-tag was incorporated into the phage library design and was present during selection. Each gene was modified to include the *pelB* leader sequence and was ligated into pET26b(+) at NdeI and XhoI. The *pelB* leader sequence targets CV-N to the periplasm, where the *pelB* tag is removed and the protein has a better environment in which to fold properly. When wt-CVN, P51G-m4-CVN, and the library mutants are expressed in BL21(DE3) at 30°C or 37°C, most of the protein is observed in inclusion bodies, while some is in the soluble cytoplasmic fraction, and very little is observed in the periplasmic fraction. Predominant inclusion body localization for CV-N is consistent with previous mutants produced in this lab (Fromme et al., 2007; Y. Liu et al., 2009) and large amounts of CV-N purified from inclusion bodies in other labs (Colleluori et al., 2005; Xiong et al., 2010). However, a significant amount of protein can be harvested from the periplasmic fraction when the culture is grown at 20°C and benzyl alcohol (20 mM final concentration) is added at the time of induction. The benzyl alcohol method saves much time and effort in protein purification because the protein is already natively folded, eliminating the refolding steps, and it is more easily extracted from the cells.

Each of the mutants was also PCR modified to have cleavable, N-terminal His-tags. The N-terminal His-tag version includes a TEV cleavage site between the His-tag and CV-N, and does not have a *pelB* leader sequence. These CV-N mutants were found most abundantly in inclusion bodies and the lack of *pelB* eliminates the option of harvesting these proteins from the periplasm using benzyl alcohol and 20°C culture conditions.

Despite containing the stabilizing P51G linker mutation, PTGEQAP-P51Gm4-CVN and LPRANHR-P51Gm4-CVN seem to be predominantly dimeric by analytical gel filtration. This could be due to specific mutations in each protein. Overall, the C-term 6His-tag construct with *pelB* leader sequence seems to yield more soluble protein than the N-term 6His-tag construct without *pelB*. This is especially true when the culture is grown with benzyl alcohol at 20°C. Two of the five CV-N mutants did not express well in BL21(DE3) despite being grown under two different conditions and trying two different 6His-tag placements. This problem may lie in the T7 phage

display library design, which included the C-term 6His-tag as part of the CV-N proteins displayed on-phage. Having an unstructured portion of the protein may have encouraged non-specific binding by some poorly folded mutants. To reduce non-specific binding, future generations of the phage display library will not contain a 6His-tag as part of the protein gene. Such a modification may be added after selection for purification convenience.

CHAPTER 4

CHARACTERIZATION OF CYANOVIRIN-N MUTANTS

ABSTRACT

Five full length CV-N mutants were obtained from the phage library. Only a very small amount of soluble CKDNRNH-P51Gm4-CVN was obtained and observed by CD and T_m, revealing a T_m = 37.9 ± 0.27°C. PTGEQAP-P51Gm4-CVN is soluble under native conditions, but CD indicated random coil structure and the protein was not characterized further. SSDGLQQ-P51Gm4-CVN and AAGRLSK-P51Gm4-CVN were purified in sufficient amounts to allow characterization by CD, T_m, ELISA, glycan array, and ITC. Both proteins have CD minima near 213 nm, which is similar to P51G-m4-CVN and indicative of β-sheet structure. AAGRLSK-P51Gm4-CVN and SSGDLQQ-P51Gm4-CVN are also thermally stable with T_m = 41.1 ± 0.37°C and T_m = 45.8 ± 0.28°C respectively. ELISA binding assays indicate binding specificity for gp120 and RNase B. To more specifically determine their target glycans, wt-CVN, P51G-m4-CVN, AAGRLSK-P51Gm4-CVN and SSDGLQQ-P51Gm4-CVN were submitted to the Consortium for Functional Glycomics (CFG) for glycan array screening. This study indicated identical specificity for AAGRLSK-P51Gm4-CVN, P51G-m4-CVN, and wt-CVN. However, SSDGLQQ-P51Gm4-CVN was specific for some different high-mannose glycans and the GlcNAcα1-4Gal motif. ITC experiments were performed with P51G-m4-CVN and SSDGLQQ-P51Gm4-CVN against 2α-Mannobiose and Man9 to further confirm the altered specificity of SSDGLQQ-P51Gm4-CVN compared to P51G-m4-CVN. P51G-m4-CVN binds 2α-Mannobiose with a K_d of 141 μM. Two of five CV-N mutants from the T7 phage display library were successfully characterized. The results demonstrate the ability of wt-CVN's Domain B to be mutated and retain specificity for the parent glycans and for additional glycans. This technique can be used to make CV-N specific for other glycans and build a repertoire of useful lectins.

INTRODUCTION

Circular Dichroism (CD). CD shines polarized light through a protein sample in a cuvette. The resulting spectra are indicative of secondary structures. Figure 21 shows typical spectra shape for alpha helix, beta sheet, and random coil protein structures. The same instrument can

be used to monitor a protein sample as it is heated. The shift in ellipticity at specific wavelengths can be monitored to determine protein unfolding.
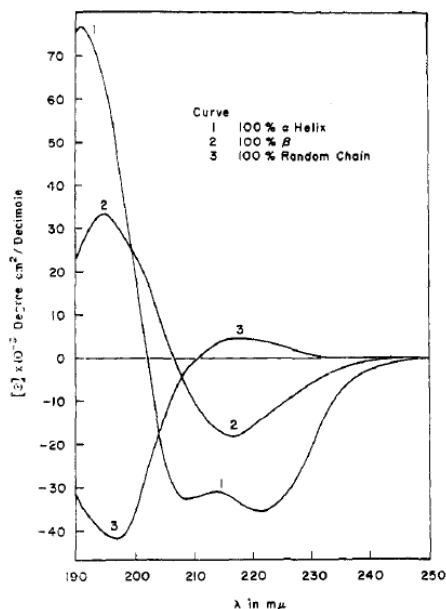

Figure 21. Typical Spectra of 1) 100% α-helix, 2) 100% β-sheet, and 3) 100% random coil. (Greenfield, 1969)

Glycan Array. Microarray technology has made it possible to study a large number of molecules quickly and efficiently. It is very valuable in areas such as glycomics, proteomics and genomics in which high-throughput methods are essential. Microarrays are typically produced by covalently binding molecules of interest to a solid surface, in this case glycans to glass. Hundreds of different glycans can be spotted or printed onto a single slide (Falciani, 2007). The Consortium for Functional Glycomics (CFG), a core laboratory at Emory University (Atlanta, GA), offers microarray screening against more than 600 glycans ("Consortium For Functional Glycomics"). CFG uses amine-reactive N-hydroxysuccinimide (NHS)-activated slides (Blixt et al., 2004). Microarray screening at CFG enables rapid assessment of lectin specificity and provides valuable information for further binding studies. In the case of the five CV-N mutants, CFG screening reveals the specificity of each mutant to the Manα1-2Man motif and identifies any cross-reactivity.

XTT anti-HIV Assay. The XTT anti-HIV assay (Fromme et al., 2007; Gulakowski et al., 1991) has been used to demonstrate CV-N's ability to inhibit HIV infection *in vivo*. The assay uses CEM-SS cells incubated with HIV-1 and various concentrations of CV-N for 6-7 days at

45

37°C. Viable cells are detected by their ability to metabolize 2,3-bis[2-methoxy-4-nitro-5-sulfophenyl]-5-[(phenylamino)carbonyl]-2H-tetrazolium hydroxide (XTT) into formazan (XTT formazan) (Scudiero, 1988) at the end of the incubation period. The colored XTT formazan is quantified by spectrophotometric analysis at 450 nM (Gulakowski et al., 1991).

Computational Docking. In addition to laboratory experiments, the five CV-N library mutants were also examined computationally. A flexible docking method was used to examine each mutant's binding to dimannose (Bolia et al., 2012; Bolia, Gerek, & Ozkan, 2012; Bolia, Woodrum, Ozkan, & Ghirlanda, 2013). Some modeling protocols use a rigid protein with little or no consideration of how the protein and ligand move in response to one another. On the other hand, molecular dynamics (MD) studies extensively examine backbone and side-chain movement (Adcock et al., 2006). MD calculations are time consuming and expensive. Flexible docking is in between rigid docking and MD, accounting for backbone and side-chain movement, but in a less comprehensive manner than MD studies. For this study, BP-Dock starts with known protein crystal structures and perturbs each amino acid, one at a time, to see how the protein backbone responds. The protein and glycan are then docked.

METHODS

CD and Thermal Denaturation (T_m). CD was performed on a Jasco J-815 Spectrometer. CV-N samples of 10-60 µM were prepared in a 1 mm cuvette in 10 mM KH_2PO_4, pH 7.0 (pH adjusted with KOH), and scanned from 260-190 nm with data pitch 0.5 nm and scanning speed 50 nm/min. Thermal denaturation temperatures (T_m) were also determined by observing the samples from 4°C-80°C or 4°C-96°C. Spectra were taken every 2°C with a scanning speed of 100 nm/min, ramp of 1°C/min, data pitch of 0.5 nm, and scanning 260-190 nm.

ELISA. This protocol is based on one previously described by Boyd, et al (Boyd et al., 1997). A polystyrene 96-well plate (Nunc) was coated with 100 µL per well gp120 (1.1 µg/mL in PBS) or 100 µL RNase B (10 µg/mL in PBS) and incubated for two hours at 37°C. It was washed four times with PBS-Tween 0.05% (PBST) and 200 µL 3% BSA was incubated in the plate overnight at 4°C. The plate was washed with PBST and 100 µL of protein was added and incubated for 30 min at room temperature. After being washed, the plate was incubated with 100

μL anti-CV-N rabbit polyclonal antibodies (1.5 mg/mL diluted 1:1000 in PBS) for 30 min. The anti-CV-N antibodies were kindly donated by Toshiyuki Mori and Barry O'Keefe at NCI. The plate was washed and then incubated for 30 min with 100 μl of the secondary antibody, stabilized goat anti-rabbit HRP conjugate (Pierce, diluted in PBS to 0.01 μg/mL). After the final wash, 100 μL TMB (0.2 g/L 3,3',5,5'- tetramethylbenzidine and H_2O_2 0.01%, KPL Cat: 507600) was added to each well and the reaction was quenched with 100 μL 2 M sulfuric acid. The plate was read at 450 nm. The following reagent was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 gp120 CM, Cat # 2968, from DAIDS, NIAID.

NMR. SSDGLQQ-P51Gm4-CVN (N-term 6His-tag) used for NMR was grown in 1 L of M9 minimal media, with ^{15}N ammonium chloride (Cambridge Isotope Laboratories) being the only available nitrogen source in the media. M9 minimal media was prepared in two parts. 16 g $Na_2HPO_4*7H_2O$, 3.5 g KH_2PO_4, 0.6 g NaCl, 1.25g NH_4Cl (^{15}N labeled), and water were mixed for a total volume of 1 L. The solution was checked to ensure it was between pH 7.0 and 7.5, and then it was autoclaved. After the media cooled, the following filter-sterilized reagents were added: 2 mL 1M $MgSO_4$, 100 μL 1M $CaCl_2$, 20 mL 20% glucose, 0.5 mL 1 M thiamin, 1 mL 1000x trace metals (10 mM each $FeCl_3*6H_2O$, $CuSO_4*5H_2O$, $MnSO_4*H_2O$, $ZnSO_4*7H_2O$), 1 mL 30 mg/mL kanamycin. Bacteria were grown overnight in 10 mL LB and then added to 1 L of M9 media. SSDGLQQ-P51Gm4-CVN was purified from inclusion bodies according to the protocol described in Chapter 3. The protein was dialyzed into 10 mM KH_2PO_4, pH 6.0, and then concentrated to 118.2 μM. Finally, 400 μl concentrated protein was mixed with 20 μl D_2O. Titration was performed using 25 mM 2α-Mannobiose (Sigma M1050) on a Varian VNMRS 500 MHz NMR. Two molar equivalents of 2α-Mannobiose were added.

Glycan Array. SSDGLQQ-P51Gm4-CVN (N-term 6His-tag), AAGRLSK-P51Gm4-CVN (C-term 6His-tag), P51G-m4-CVN (C-term 6His-tag), and wt-CVN (C-term 6His-tag) were prepared for screening at the Consortium for Functional Glycomics. Protein was concentrated to 2 mg/mL in 1 mL 0.1mM sodium bicarbonate, pH 8.3. 2.5 mg 5-(and-6)-Carboxytetramethylrhodamine succinimidyl ester dye (C-1171 Invitrogen) was dissolved in 280 μL DMF. 25 μL dye was added

to the protein and stirred at room temperature for 1 hour. Excess dye was removed by size exclusion chromatography on a Bio-Rad DG10 desalting column equilibrated with PBS pH 7.4.

Samples were sent to the Consortium for Functional Glycomics for microarray glycan screening. SSDGLQQ-P51Gm4-CVN was analyzed on the Mammalian Printed Array version 5.0 and all other proteins were analyzed on Mammalian Printed Array version 5.1. Each array is printed on a glass slide with over 600 different glycans as previously described (Blixt et al., 2004; "Consortium For Functional Glycomics,").

Anti-HIV Assay. wt-CVN and SSDGLQQ-wt-CVN, both without His-tags, were sent to the National Cancer Institute (NCI) lab of Barry O'keefe (Frederick, MD) for an anti-HIV activity assay. The assay uses CEM-SS cells incubated with HIV and eight concentrations of CV-N for six days at 5% CO_2 and 37°C. At the end of the incubation period, viable cells were detected by addition of 50 µl XTT, incubation at 37°C for 4 hours, followed by spectrophotometric analysis at 450 nm (Fromme et al., 2007; Gulakowski et al., 1991).

Computational Protocol. The Backbone Perturbation-Dock (BP-Dock) flexible docking method (Bolia et al., 2012; Bolia, Gerek, & Ozkan, 2012; Bolia, Woodrum, Ozkan, & Ghirlanda, 2013) was used to examine dimannose binding by P51G-m4-CVN (PDB: 2RDK), CVN^{mutDB} (PDB: 3CZZ), and the five full-length CV-N mutants produced in my T7 phage display library. Briefly, each αC is a node and each is treated as if there is a spring in between. The structures are compared and similar structures are grouped together/eliminated. Amino acid side-chain conformations are then considered. All of these different structures will be considered when docking to the ligand.

RESULTS

Of the five full length CV-N sequences isolated from the T7 phage display library, only SSDGLQQ-P51Gm4-CVN and AAGRLSK-P51Gm4-CVN were significantly characterized.

Secondary structure and T_m of the CV-N mutants were examined using CD. The CD spectra have minima at approximately 213 nm, indicating mainly β-sheet structure for each protein. Figure 22 gives spectra for each of the mutants produced.
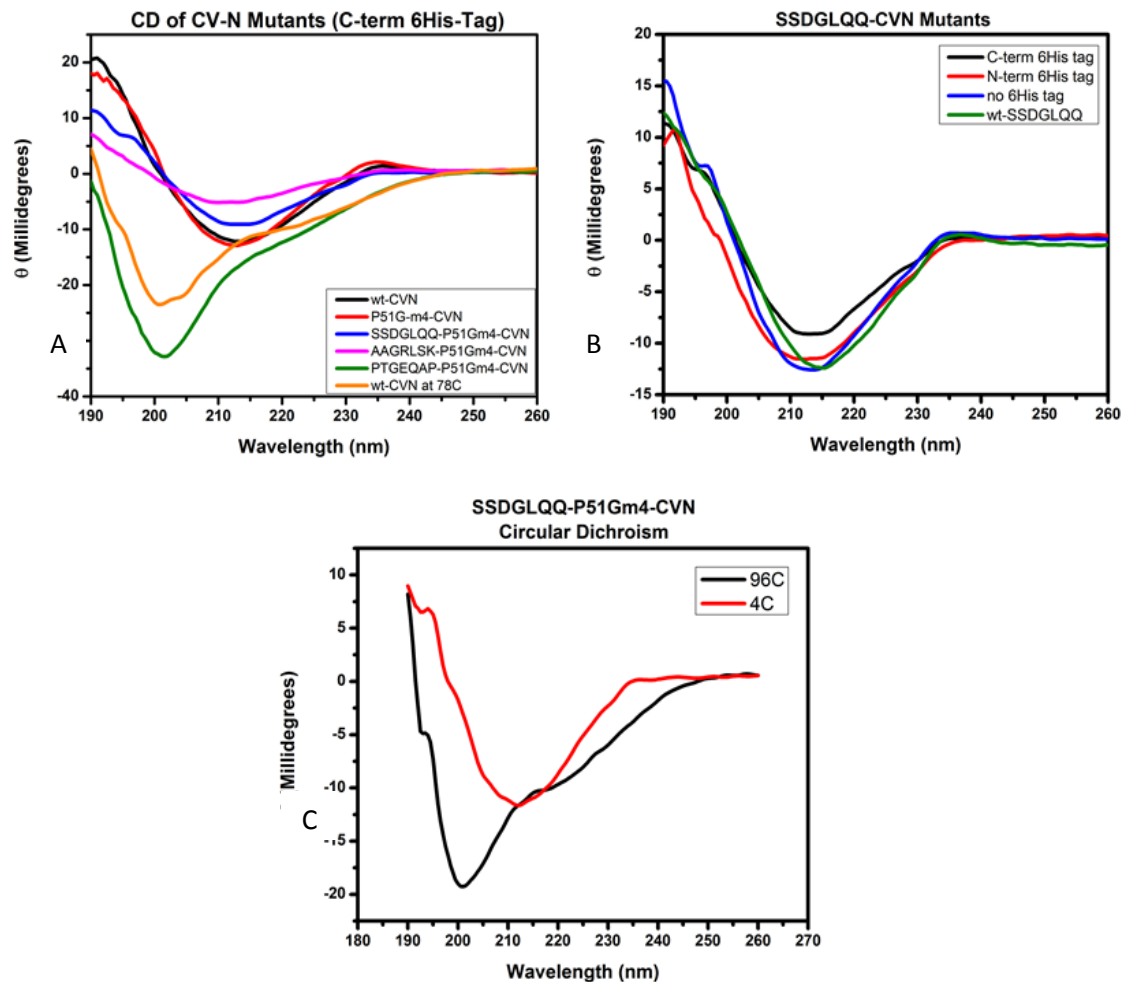
Figure 22. CD spectra of CV-N mutants, wt-CVN, and P51G-m4-CVN. A) CD spectra of C-term 6His-tag versions of mutants and wt-CVN and P51G-m4-CVN. B) SSDGLQQ-wt-CVN, and C-term, N-term, and no tag versions of SSDGLQQ-P51Gm4-CVN. C) SSDGLQQ-P51Gm4-CVN at 4°C and 96°C.

Thermal denaturation studies were also performed. Table 5 gives the thermal denaturation temperatures based on a two-state model.

Table 5

Thermal Denaturation Temperatures Followed at 200 nm

Protein	T_m (°C)
wt-CVN C-term 6H	51.9 ± 0.20°C
SSDGLQQ-wt-CVN no 6H	52.9 ± 0.25°C
P51G-m4-CVN C-term 6H	55.1 ± 0.20°C
AAGRLSK-P51Gm4-CVN C-term 6H	41.1 ± 0.37°C
SSDGLQQ-P51Gm4-CVN C-term 6H	45.8 ± 0.28°C
SSDGLQQ-P51Gm4-CVN N-term 6H	49.2 ± 0.32°C
CKDNRNH-P51Gm4-CVN	37.9 ± 0.27°C
AAGRLSK-P51Gm4-CVN	42.5 ± 0.32°C
P51G-m4-CVN no 6H	46.4 C
SSDGLQQ-P51Gm4-CVN no 6H	55.5 ± 0.26°C
AAGRLSK-P51Gm4-CVN no 6H	47.2 ± 0.18°C

As expected, SSDGLQQ-P51Gm4-CVN was more thermally stable than SSDGLQQ-wt-CVN, which does not have the stabilizing P51G mutation. Interestingly, SSDGLQQ-P51Gm4-CVN and SSDGLQQ-wt-CVN have melting temperatures similar to P51G-m4-CVN and wt-CVN respectively. It is also noted that SSDGLQQ-CVN is more thermally stable than AAGRLSK-P51Gm4-CVN and CKDNRNH-P51Gm4-CVN. Figure 23 follows each protein's denaturation by CD at 200 nm.
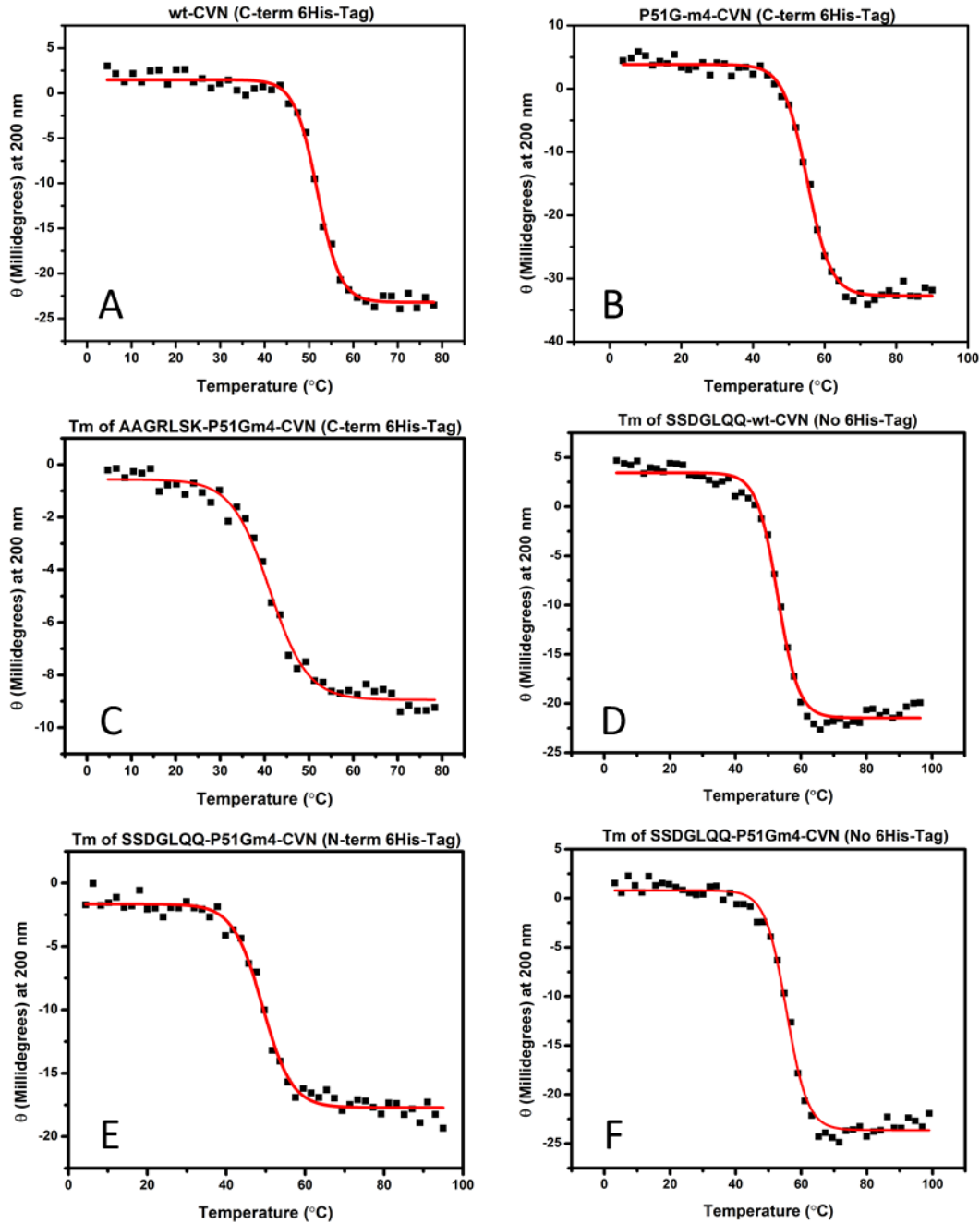
Figure 23. T_m curves of CV-N, P51G-m4-CVN, and CV-N mutants at 200 nm. A) wt-CVN (C-term 6His-tag) has a T_m = 51.9 ± 0.20°C. B) P51G-m4-CVN (C-term 6His-tag) has a T_m = 55.1 ± 0.20°C. C) AAGRLSK-P51Gm4-CVN (C-term 6His-tag) has a T_m = 41.1 ± 0.37°C. D) SSDGLQQ-wt-CVN (no 6His-tag) has a T_m = 52.9 ± 0.25°C. E) SSDGLQQ-P51Gm4-CVN (N-term 6His-tag) has a T_m = 49.2 ± 0.32°C. F) SSDGLQQ-P51Gm4-CVN (no 6His-tag) has a T_m = 55.5 ± 0.26°C.

ELISA was performed against gp120 and RNase B. Figure 24 and Table 6 indicate

nanomolar or low micromolar EC_{50} values for SSDGLQQ-P51Gm4-CVN (N-term 6His-tag),

SSDGLQQ-wt-CVN (tag), P51G-m4-CVN (C-term 6His-tag), and wt-CVN (C-term 6His-tag). SSDGLQQ-P51Gm4-CVN has 3.7 ± 0.5 µM binding compared to 84 ± 32 nM for P51G-m4-CVN, and 0.98 ± 1.3 nM for wt-CVN. SSDGLQQ-wt-CVN (no tag), AAGRLSK-P51Gm4-CVN (C-term 6His-tag and N-term 6His-tag) and SSDGLQQ-P51Gm4-CVN (C-term 6His-tag, and no tag) also bound gp120, but the data could not be fit.
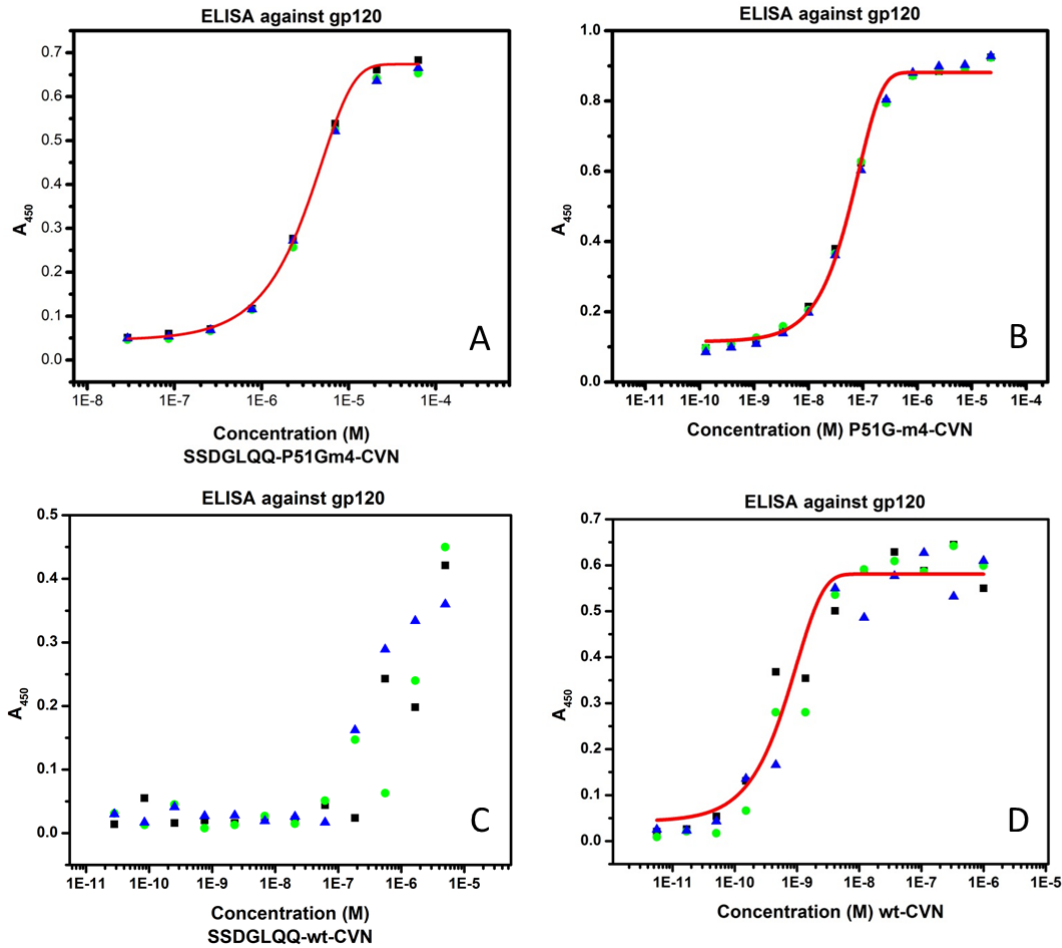
Figure 24. ELISA binding curves. A) SSDGLQQ-P51Gm4-CVN (N-term 6His-tag), B) P51G-m4-CVN (C-term 6His-tag), C) SSDGLQQ-wt-CVN (no tag), and D) wt-CVN (no tag) against gp120, with EC_{50} of 3.7 ± 0.5 µM and 84 ± 32 nM respectively. The EC_{50} is 0.98 ± 1.3 nM for wt-CVN (no tag). The SSDGLQQ-wt-CVN curve was not fit, but likely has a micromolar EC_{50}.

Table 6

ELISA Binding Data for wt-CVN and CV-N Mutants Against RNase B and Recombinant gp120

Protein	Gp120 EC_{50}
SSDGLQQ-P51Gm4-CVN (N-term)	3.7 ± 0.5 µM
P51G-m4-CVN (C-term)	84 ± 32 nM
wt-CVN (no tag)	0.98 ± 1.3 nM

ELISA against RNase B was also performed to detect binding specificity for high-mannose glycans. Though the data was not fit, each protein also bound to RNase B. RNase B has a single high-mannose glycan ranging from Man_5 to Man_9 in size (Hua et al., 2012; Prien et al., 2009). CV-N mutants bind to both proteins suggesting specificity for high-mannose glycans, the only common feature between gp120 and RNase B.
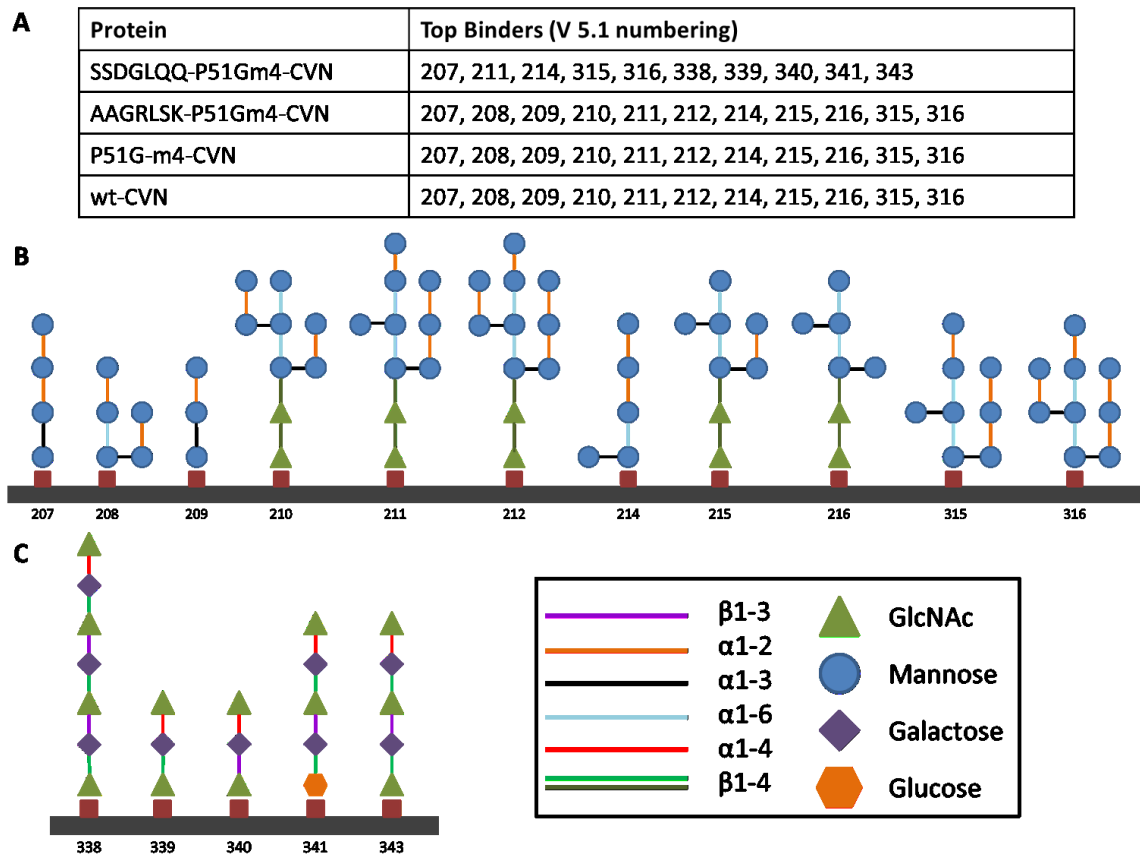
SSDGLQQ-P51Gm4-CVN binding to 2α-Mannobiose was also qualitatively examined by NMR. After adding two molar equivalents of 2α-Mannobiose to SSDGLQQ-P51Gm4-CVN, no changes in peak intensity or peak shifting were observed. Though no binding was observed, the HSQC in Figure 26 shows that the protein is folded.

HSQC of SSDGLQQ-P51Gm4-CVN
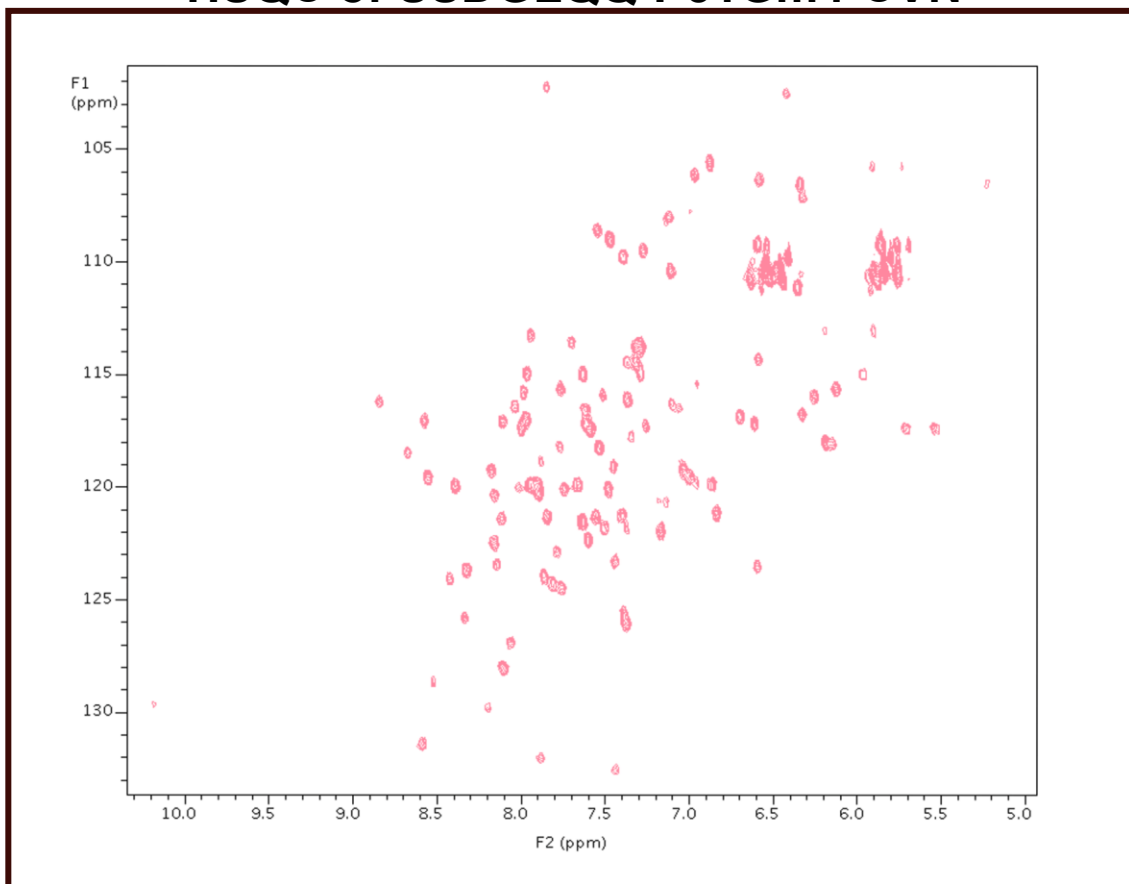
Figure 25. HSQC showing folded ^{15}N labeled SSDGLQQ-P51Gm4-CVN (N-term 6His-tag). Two molar equivalents of 2α-Mannobiose were titrated into the sample and no binding was observed.

To more broadly determine their binding specificities, wt-CVN (C-term 6His-tag), P51G-m4-CVN (C-term 6His-tag), SSDGLQQ-P51Gm4-CVN (N-term 6His-tag), and AAGRLSK-P51Gm4-CVN (C-term 6His-tag) were sent to CFG for glycan array analysis. SSDGLQQ-P51Gm4-CVN was screened on Mammalian Printed Array version 5.0. The remaining proteins were screened on Mammalian Printed Array version 5.1. These versions of the array are nearly identical. Version 5.0 has 611 glycans and version 5.1 has 610 glycans, with the only difference being the absence of Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13 (glycan 56 on version 5.0) on version 5.1 of the array.

Protein	Top Binders (V 5.1 numbering)
SSDGLQQ-P51Gm4-CVN	207, 211, 214, 315, 316, 338, 339, 340, 341, 343
AAGRLSK-P51Gm4-CVN	207, 208, 209, 210, 211, 212, 214, 215, 216, 315, 316
P51G-m4-CVN	207, 208, 209, 210, 211, 212, 214, 215, 216, 315, 316
wt-CVN	207, 208, 209, 210, 211, 212, 214, 215, 216, 315, 316

Figure 26. Glycans significantly bound by CV-N proteins on the CFG mammalian glycan array. A) Glycan array hits for wt-CVN, P51G-m4-CVN, AAGRLSK-P51Gm4-CVN, and SSDGLQQ-P51Gm4-CVN. The glycans are listed in numerical order and are all numbered according to version 5.1 of the CFG mammalian glycan array. B) Pictorial representation of relevant mannose glycan structures on CFG mammalian glycan array version 5.1. C) Pictorial representation of relevant GlcNAcα1-4Gal glycan structures on CFG mammalian glycan array version 5.1.

54

As seen in Figure 27 and Appendix A, SSDGLQQ-P51Gm4-CVN binds well to two glycan motifs and has low levels of background binding. The 10 glycans bound most strongly to SSDGLQQ-P51Gm4-CVN contain either Manα1-2Manα1-2Man or GlcNAcα1-4Gal terminal motifs, indicating dual specificity. The comprehensive list of glycans strongly bound by SSDGLQQ-P51Gm4-CVN, AAGRLSK-P51Gm4-CVN, P51G-m4-CVN, and wt-CVN can be found in Appendix A.

AAGRLSK-P51Gm4-CVN, P51G-m4-CVN, and wt-CVN have identical glycan specificity and bound best to the same 11 high mannose glycans. These include glycans 207, 211, 214, 315, and 316 bound by SSDGLQQ-P51Gm4-CVN and an additional six mannose glycans. In harmony with previous studies (C. Bewley, 2002; C. A. Bewley, 2001), P51G-m4-CVN recognizes glycans containing Manα1-2Man.

To demonstrate anti-HIV activity, the seven Domain B mutations in SSDGLQQ-P51Gm4-CVN were made on wt-CVN, producing SSDGLQQ-wt-CVN. This protein has the new Domain B while retaining a native Domain A and linker region. SSDGLQQ-wt-CVN was sent to NCI for XTT anti-HIV assay, and the results are in Table 8. Wt-CVN has an EC_{50} of 3 nM while SSDGLQQ-wt-CVN has 500 nM activity. Consistent with the ELISA binding data, SSDGLQQ-wt-CVN has less potent anti-HIV activity than wt-CVN.

Table 7

Anti-HIV Activity of wt-CVN, SSDGLQQ-wt-CVN, and P51G-m4-CVN as Measured by XTT Assay

Protein	EC_{50}
wt-CVN	3 nM
SSDGLQQ-wt-CVN	500 nM
P51G-m4-CVN*	>>5000 nM

Note: P51G-m4-CVN data comes from a previous assay (Fromme et al., 2007).

In addition to the laboratory characterizations, all five CV-N library mutants, SSDGLQQ-wt-CVN, P51G-m4-CVNC, and CV-N[mutDB] (Barrientos et al., 2006) were examined computationally for binding to Manα1-2Man and Manα1-2Manα1-2Man. Both dimannose and trimannose were examined based on CFG results for AAGRLSK-P51Gm4-CVN and SSDGLQQ-P51Gm4-CVN. CV-N[mutDB] is the negative control, its Domain B being unable to bind to Manα1-2Man, and P51G-

m4-cvn is the positive control because it is the parent protein of my CV-N library mutants. Table 8

gives ΔG values for each protein.

Table 8

*Computational Data for Nine CV-N Proteins in Which Domain B Was Docked With Manα1-2Man and Manα1-2Manα1-2Man*

Protein (no tag)	Manα1-2Man	Manα1-2Manα1-2Man
wt-CVN (1IIY) Domain B	-12.94 kcal/mol	-13.77 kcal/mol
P51G-m4-CVN	-17.25 kcal/mol	-17.72 kcal/mol
CV-N^{mutDB}*	-9.28 kcal/mol	
SSDGLQQ-wt-CVN	-14.76 kcal/mol	-15.82 kcal/mol
SSDGLQQ-P51Gm4-CVN	-12.61 kcal/mol	-13.75 kcal/mol
CKDNRNH-P51Gm4-CVN	-11.67 kcal/mol	
PTGEQAP-P51Gm4-CVN	-10.01 kcal/mol	
AAGRLSK-P51Gm4-CVN	-11.77 kcal/mol	-12.33 kcal/mol
LPRANHR-P51Gm4-CVN	-9.06 kcal/mol	

Note: This data was provided by Ashini Bolia. *protein originally described by Gronenborn (Barrientos et al., 2006).

My coworker Ashini Bolia has kindly provided PyMOL models displaying potential

hydrogen bonding between each of the five library mutants and dimannose. Figure 27 shows

hydrogen bonds with the CV-N background and side-chains. Dimannose forms hydrogen bonds

with the SSDGLQQ-P51Gm4-CVN backbone at residues Asn42, Ser44, Asp52, Gly53, and

Gln74. The side-chains of Ser41, Gln74, and Gln76 also form hydrogen bonds with the glycan, for

a total of seven amino acids interacting with the glycan. AAGRLSK-P51Gm4-CVN's backbone

interacts with dimannose at Asn42, Gly52, Arg53, and Ser74. The side-chains of Thr57 and

Ser74 are also involved, totaling five amino acids interacting with the glycan. PTGEQAP-

P51Gm4-CVN's model shows dimannose interacting with the protein backbone at Asn42 and

Gly52, and side-chains of Asn42 and Glu53. This gives PTGEQAP-P51Gm4-CVN only three

amino acids interacting with the glycan.

Though CKDNRNH-P51Gm4-CVN and LPRANHR-P51Gm4-CVN were not very soluble in

the lab, PyMOL models show dimannose interacting with six and seven amino acids,

respectively, in Domain B. CKDNRNH-P51Gm4-CVN's model shows potential hydrogen bonding

with dimannose at residues Cys41, Asn42, Asn53, Thr57, Thr75, and Asn74. LPRANHR-

P51Gm4-CVN potentially interacts with dimannose at Asn42, Arg52, Ala53, Thr57, His74, Thr75,

and Arg76.
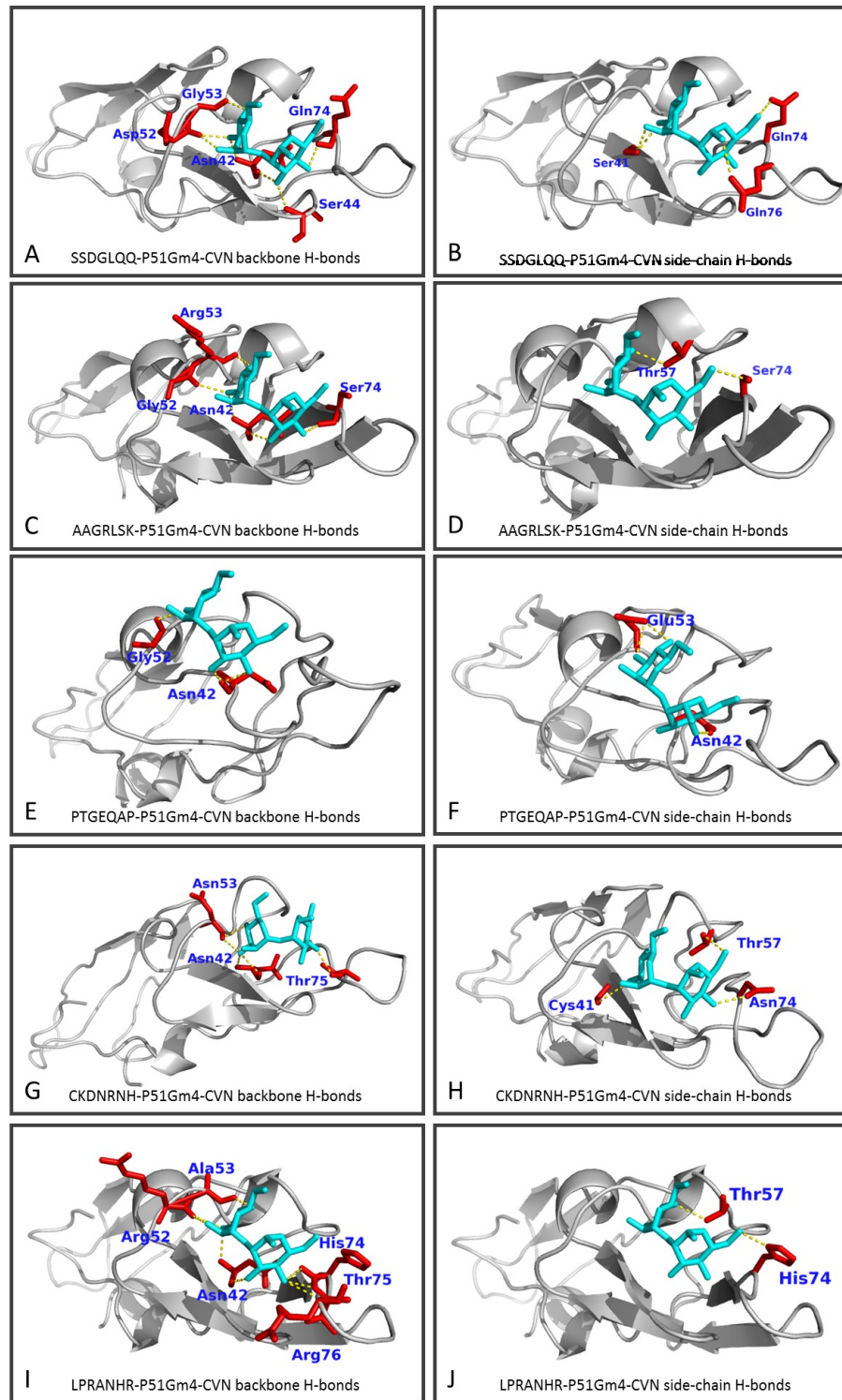
Figure 27. PyMOL models of CV-N library mutant hydrogen bonds with dimannose. A, C, E, G, and I show hydrogen bonding between the CV-N mutant backbone and dimannose. B, D, F, H, and J show hydrogen bonding between CV-N side-chains and dimannose. These models were created in PyMOL and were kindly provided by Ashini Bolia.

DISCUSSION

Five full length CV-N mutants were isolated from the T7 phage display library. By CD, AAGRLSK-P51Gm4-CVN, SSDGLQQ-P51Gm4-CVN, and SSDGLQQ-wt-CVN have minima near 213 nm, indicating mostly β-sheet structure. This is in agreement with wt-CVN and P51G-m4-CVN, indicating similar secondary structure.

Thermal denaturation temperatures were also obtained by CD. SSDGLQQ-P51Gm4-CVN (no 6His-tag) and P51G-m4-CVN (C-term 6His-tag) have T_m of 55°C and SSDGLQQ-wt-CVN (no 6His-tag) and wt-CVN (C-term 6His-tag) both have T_m near 52°C. SSDGLQQ-P51Gm4-CVN and its wild-type version behave similarly to their parent proteins. They are easily refolded from inclusion bodies and have similar T_m and CD minima. Interestingly, the thermal denaturation temperature of SSDGLQQ-P51Gm4-CVN varies by as much as 10°C based on the location or absence of the 6His-tag. The no 6His-tag version has the highest T_m at 55.5 ± 0.26°C, the N-term 6His-tag protein has T_m of 49.2 ± 0.32°C, and the C-term 6His-tag version has a T_m of 45.8 ± 0.28°C.

AAGRLSK-P51Gm4-CVN was the next most stable mutant, with T_m values of 47.2 ± 0.18 °C without a tag, and 41.1 ± 0.37 °C with a C-term 6His-tag. As with SSDGLQQ-CVN, 6His-tag placement seems to influence the protein's thermal stability.

PTGEQAP-P51Gm4-CVN was soluble but not folded at room temperature under native conditions. Its CD minima near 202 nm indicates a random coil structure not typical of CV-N and it appears to be mostly or all dimer by gel filtration. LPRANHR-P51Gm4-CVN was not soluble under any of the expression or refolding conditions attempted. CKDNRNH-P51Gm4-CVN was quite difficult to obtain as a folded protein and only a small amount of the protein was sufficiently soluble to study by CD. It has a T_m value of 37.9 ± 0.27°C, which is too low for anti-HIV assay. CKDNRNH-P51Gm4-CVN has five cysteines, four of them found in wt-CVN, and one at position 41 in Domain B.

Overall, it seems that the 6His-tag affects the thermal stability of CVN mutants. Considering that C-term 6His-tag was the least stable version of SSDGLQQ-P51Gm4-CVN and AAGRLSK-P51Gm4-CVN, it is likely that including the C-term 6His-tag in the T7 phage display

library may have yielded a higher percentage of poorly folded proteins than a library without a 6His-tag.

In addition to thermal destabilization, the 6His-tag likely facilitates an increased level of non-specific binding. Despite being poorly folded, PTGEQAP-P51Gm4-CVN and CKDNRNH-P51Gm4-CVN bind to gp120 and RNase B. This binding is likely achieved nonspecifically.

PyMOL models furnished by Ashini demonstrate potential hydrogen bonds between each of the five library mutants and dimannose. SSDGLQQ-P51Gm4-CVN interacts with the glycan at five backbone amino acids throughout the Domain B cleft, and three side-chain amino acid interactions, for a total of seven amino acids interacting well with dimannose. AAGRLSK-P51Gm4-CVN has fewer amino acids interacting with dimannose, but its five amino acids hydrogen bond with the glycan throughout the cleft, pointing toward good binding. PTGEQAP-P51Gm4-CVN has only three amino acids predicted to hydrogen bond with dimannose. With Gly52 and Glu53 on one end and Asn42 on the other, the glycan seems to be held at its ends by the protein, probably resulting in weaker binding than SSDGLQQ-P51Gm4-CVN or AAGRLSK-P51Gm4-CVN.

Based on Ashini's computational data in Table 8, Domain B of all five library mutants ranges from -9.06 kcal/mol for LPRANHR-P51Gm4-CVN to -12.61 kcal/mol for SSDGLQQ-P51Gm4-CVN when docked with Manα1-2Man. The negative control CV-N^{mutDB} has -9.28 kcal/mol interaction while positive control P51G-m4-CVN has -17.25 kcal/mol. This dimannose docking study indicates that LPRANHR-P51Gm4-CVN and PTGEQAP-P51Gm4-CVN weakly bind or don't bind dimannose as compared to CV-N^{mutDB} (Barrientos et al., 2006). SSDGLQQ-P51Gm4-CVN, AAGRLSK-P51Gm4-CVN, and CKDNRNH-P51Gm4-CVN should bind, even if weakly, to dimannose, though none of the mutants approached the -17.25 kcal/mol of P51G-m4-CVN.

SSDGLQQ-P51Gm4-CVN (N-term 6His-tag) has an ELISA (gp120) EC_{50} of 3.7 ± 0.5 µM compared to 84 ± 32 nM for P51G-m4-CVN and 0.98 ± 1.3 nM for wt-CVN. SSDGLQQ-P51Gm4-CVN bound gp120 and RNase B, indicating specificity for mannose glycans. ELISA data could

not be fit for SSDGLQQ-wt-CVN and AAGRLSK-P51Gm4-CVN, but both proteins may have micromolar EC_{50} against gp120.

CFG glycan array screening of P51G-m4-CVN, wt-CVN, and AAGRLSK-P51Gm4-CVN indicates identical specificity by all three proteins for high-mannose glycans. Previously, Green et al sent P51G-CVN to CFG for glycan array screening on the Mammalian Printed Array version 4.1 (Patsalo, Raleigh, & Green, 2011). This version of the array contains only 465 glycans compared to 610 on version 5.1. P51G-CVN bound best to mannose glycans, consistent with P51G-m4-CVN and wt-CVN. AAGRLSK-P51Gm4-CVN retains a large, charged amino acid at position 76, similar to wt-CVN and P51G-m4-CVN. The modest R76K mutation likely helps AAGRLSK-P51Gm4-CVN retain identical glycan specificity to its parent protein because it can cap and specifically bind the glycan (Margulis, 2005).

Interestingly, the glycan array screen revealed a preference of SSDGLQQ-P51Gm4-CVN for Manα1-2Manα1-2Man terminal glycan motifs. This could be due, in part, to its slightly shallower Domain B binding site and a R76Q mutation that may slightly reduce the ability of 76 to cap the bound glycan (Margulis, 2005). Unlike P51G-m4-CVN, wt-CVN, and AAGRLSK-P51Gm4-CVN, SSDGLQQ-P51Gm4-CVN binds to two distinct glycan motifs.

SSDGLQQ-P51Gm4-CVN also has binding specificity for GlcNAcα1-4Gal. The second terminal glycan motif has been identified most frequently in human gastrointestinal epithelial and other mucosal tissues (Kawakubo et al., 2004; Lee et al., 2008). This glycan motif acts as an antibacterial agent against *Helicobacter pylori*, a common gastric pathogen (Kawakubo et al., 2004). GlcNAcα1-4Gal blocks an enzyme important to the synthesis of a cholesterol cell wall component. The glycan is expressed below the surface of gastric mucosa, allowing *H. pylori* to colonize the top portion of mucosa, but preventing further infection.

The protein selection method described in this experiment was too broad to select for exact glycan specificities, but it did demonstrate the ability to mutate CV-N Domain B and, in AAGRLSK-P51Gm4-CVN, retain glycan specificity identical to P51G-m4-CVN, while changing the glycan specificity of SSDGLQQ-P51Gm4-CVN. Future CV-N libraries should be selected against

the specific glycan of interest instead of a heterogeneous target. This will ensure that the protein

binds to a specific glycan.

REFERENCES

Acheson, N. H. (2007). *Fundamentals of Molecular Virology*. Hoboken, NJ: John Wiley & Sons.

Adcock, S. A., McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev, 106*(5), 1589-1615.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002). *Molecular Biology of The Cell* (Fourth ed.). New York, NY: Garland Science.

Alexandre, K. B., Gray, E. S., Lambson, B. E., Moore, P. L., Choge, I. A., Mlisana, K., . . . Morris, L. (2010). Mannose-rich glycosylation patterns on HIV-1 subtype C gp120 and sensitivity to the lectins, Griffithsin, Cyanovirin-N and Scytovirin. *Virology, 402*(1), 187-196.

Arnaud, J., Audfray, A., & Imberty, A. (2013). Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chem Soc Rev*. Advance online publication. doi: 10.1039/c2cs35435g

Baneyx, F. (Ed.). (2004). *Protein Expression Technologies: Current Status and Future Trends*. Wymondham, UK: Horizon Biosciences.

Barrientos, L. G., & Gronenborn, A. M. (2002). The domain-swapped dimer of cyanovirin-N contains two sets of oligosaccharide binding sites in solution. *Biochem Biophys Res Commun, 298*(4), 598-602.

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*(10), 1799-1807.

Barrientos, L. G., Louis, J. M., Botos, I., Mori, T., Han, Z., O'Keefe, B. R., . . . 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*(5), 673-686.

Barrientos, L. G., Louis, J. M., Hung, J., Smith, T. H., O'Keefe, B. R., Gardella, R. S., . . . Gronenborn, A. M. (2002). Design and initial characterization of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N. *Proteins, 46*(2), 153-160.

Barrientos, L. G., Matei, E., Lasala, F., Delgado, R., & Gronenborn, A. M. (2006). Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Eng Des Sel, 19*(12), 525-535.

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*(1), 47-56.

Bennett, M. J., Schlunegger, M. P., & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci, 4*(12), 2455-2468.

Bewley, C. (2002). Site-specific Discrimination by Cyanovirin-N for α-Linked Trisaccharides Comprising the Three Arms of Man8 and Man9. *Journal of Molecular Biology, 322*(4), 881-889.

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*(10), 931-940.

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*(7), 571-578.

Bewley, C. A., Kiyonaka, S., & Hamachi, I. (2002). Site-specific discrimination by cyanovirin-N for alpha-linked trisaccharides comprising the three arms of Man(8) and Man(9). *J Mol Biol, 322*(4), 881-889.

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*(17), 3892-3902.

Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., . . . Paulson, J. C. (2004). Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A, 101*(49), 17033-17038.

Bolia, A., Gerek, Z. N., Keskin, O., Ozkan, S. B., & Dev, K. K. (2012). The binding affinities of proteins interacting with the PDZ domain of PICK1. *Proteins, 80*(5), 1393-1408.

Bolia, A., Gerek, Z. N., & Ozkan, S. B. (2012). BP-Dock: a flexible docking scheme for exploring protein-ligand interactions. Manuscript submitted for publication.

Bolia, A., Woodrum, B., Ozkan, B., & Ghirlanda, G. (2013). Dissecting the determinants for dimannose binding in Cyanovirin. Manuscirpt in preparation.

Bolmstedt, A. J., O'Keefe, B. R., Shenoy, S. R., McMahon, J. B., & Boyd, M. R. (2001). Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. *Mol Pharmacol, 59*(5), 949-954.

Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., . . . Wlodawer, A. (2002). Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem, 277*(37), 34336-34342.

Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., . . . 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*(7), 1521-1530.

Calarese, D. A., Lee, H. K., Huang, C. Y., Best, M. D., Astronomo, R. D., Stanfield, R. L., . . . Wilson, I. A. (2005). Dissection of the carbohydrate specificity of the broadly neutralizing anti-HIV-1 antibody 2G12. *Proc Natl Acad Sci U S A, 102*(38), 13372-13377.

Chang, L. (2002). Potent Inhibition of HIV-1 Fusion by Cyanovirin-N Requires Only a Single High Affinity Carbohydrate Binding Site: Characterization of Low Affinity Carbohydrate Binding Site Knockout Mutants. *Journal of Molecular Biology, 318*(1), 1-8.

Colleluori, D. M., Tien, D., Kang, F., Pagliei, T., Kuss, R., McCormick, T., . . . Romano, J. W. (2005). Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. *Protein Expr Purif, 39*(2), 229-236.

Consortium For Functional Glycomics.
  http://functionalglycomics.org/static/consortium/resources.shtml

de Marco, A., Vigh, L., Diamant, S., & Goloubinoff, P. (2005). Native folding of aggregation-prone
  recombinant proteins in Escherichia coli by osmolytes, plasmid- or benzyl alcohol-
  overexpressed molecular chaperones. *Cell Stress & Chaperones, 10*(4), 329-339.

Dey, B., Lerner, D. L., Lusso, P., Boyd, M. R., Elder, J. H., & Berger, E. A. (2000). Multiple
  antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120
  interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *J Virol,
  74*(10), 4562-4569.

Doores, K. J., Fulton, Z., Huber, M., Wilson, I. A., & Burton, D. R. (2010). Antibody 2G12
  recognizes di-mannose equivalently in domain- and nondomain-exchanged forms but only
  binds the HIV-1 glycan shield if domain exchanged. *J Virol, 84*(20), 10690-10699.

Eckert, D. M., & Kim, P. S. (2001). Mechanisms of viral membrane fusion and its inhibition. *Annu
  Rev Biochem, 70*, 777-810.

Esser, M. T., Mori, T., Mondor, I., Sattentau, Q. J., Dey, B., Berger, E. A., . . . Lifson, J. D. (1999).
  Cyanovirin-N binds to gp120 to interfere with CD4-dependent human immunodeficiency
  virus type 1 virion binding, fusion, and infectivity but does not affect the CD4 binding site
  on gp120 or soluble CD4-induced conformational changes in gp120. *J Virol, 73*(5), 4360-
  4371.

Esté, J. A., & Telenti, A. (2007). HIV entry inhibitors. *The Lancet, 370*(9581), 81-88.

Falciani, F. (Ed.). (2007). *Microarray Technology Through Applications*. New York, NY: Taylor &
  Francis Group.

Fromme, R., Katiliene, Z., Fromme, P., & Ghirlanda, G. (2008). Conformational gating of
  dimannose binding to the antiviral protein cyanovirin revealed from the crystal structure at
  1.35 A resolution. *Protein Sci, 17*(5), 939-944.

Fromme, R., Katiliene, Z., Giomarelli, B., Bogani, F., Mc Mahon, J., Mori, T., . . . Ghirlanda, G.
  (2007). A monovalent mutant of cyanovirin-N provides insight into the role of multiple
  interactions with gp120 for antiviral activity. *Biochemistry, 46*(32), 9199-9207.

Goodsell, D. S. (2008). Ribonuclease A. *RCSB Protein Data Bank*. doi:
  10.2210/rcsb_pdb/mom_2008_9

Greenfield, N. a. F., Gerald. (1969). Computed Circular Dichroism Spectra for the Evaluation of
  Protein Conformation. *Biochemistry*, 4108-4116.

Gulakowski, R. J., McMahon, J. B., Staley, P. G., Moran, R. A., & Boyd, M. R. (1991). A
  semiautomated multiparameter approach for anti-HIV drug screening. *J Virol Methods,
  33*(1-2), 87-100.

Gustafson, K. R., Sowder, R. C., 2nd, Henderson, L. E., Cardellina, J. H., 2nd, McMahon, J. B.,
  Rajamani, U., . . . Boyd, M. R. (1997). Isolation, primary sequence determination, and
  disulfide bond structure of cyanovirin-N, an anti-HIV (human immunodeficiency virus)
  protein from the cyanobacterium Nostoc ellipsosporum. *Biochem Biophys Res Commun,
  238*(1), 223-228.

Han, Z., Xiong, C., Mori, T., & Boyd, M. R. (2002). Discovery of a stable dimeric mutant of cyanovirin-N (CV-N) from a T7 phage-displayed CV-N mutant library. *Biochem Biophys Res Commun, 292*(4), 1036-1043.

Horvath, I., Glatz, A., Varvasovszki, V., Torok, Z., Pali, T., Balogh, G., . . . Vigh, L. (1998). Membrane physical state controls the signaling mechanism of the heat shock response in Synechocystis PCC 6803: identification of hsp17 as a "fluidity gene". *Proc Natl Acad Sci U S A, 95*(7), 3513-3518.

Hu, Q., Mahmood, N., & Shattock, R. J. (2007). High-mannose-specific deglycosylation of HIV-1 gp120 induced by resistance to cyanovirin-N and the impact on antibody neutralization. *Virology, 368*(1), 145-154.

Hua, S., Nwosu, C. C., Strum, J. S., Seipert, R. R., An, H. J., Zivkovic, A. M., . . . Lebrilla, C. B. (2012). Site-specific protein glycosylation analysis with glycan isomer differentiation. *Anal Bioanal Chem, 403*(5), 1291-1302.

Kapust, R. B., Tozser, J., Copeland, T. D., & Waugh, D. S. (2002). The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun, 294*(5), 949-955.

Kawakubo, M., Ito, Y., Okimura, Y., Kobayashi, M., Sakura, K., Kasama, S., . . . Nakayama, J. (2004). Natural antibiotic function of a human gastric mucin against Helicobacter pylori infection. *Science, 305*(5686), 1003-1006.

Keeffe, J. R., Gnanapragasam, P. N., Gillespie, S. K., Yong, J., Bjorkman, P. J., & Mayo, S. L. (2011). Designed oligomers of cyanovirin-N show enhanced HIV neutralization. *Proc Natl Acad Sci U S A, 108*(34), 14079-14084.

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*(13), 3210-3211.

Koharudin, L. M., Viscomi, A. R., Jee, J. G., Ottonello, S., & Gronenborn, A. M. (2008). The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. *Structure, 16*(4), 570-584.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature, 393*(6686), 648-659.

Lee, H., Wang, P., Hoshino, H., Ito, Y., Kobayashi, M., Nakayama, J., . . . Fukuda, M. (2008). Alpha1,4GlcNAc-capped mucin-type O-glycan inhibits cholesterol alpha-glucosyltransferase from Helicobacter pylori and suppresses H. pylori growth. *Glycobiology, 18*(7), 549-558.

Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G., & Subramaniam, S. (2008). Molecular architecture of native HIV-1 gp120 trimers. *Nature, 455*(7209), 109-113.

Liu, L., Byeon, I. J., Bahar, I., & Gronenborn, A. M. (2012). Domain swapping proceeds via complete unfolding: a 19F- and 1H-NMR study of the Cyanovirin-N protein. *J Am Chem Soc, 134*(9), 4229-4235.

Liu, Y., Carroll, J. R., Holt, L. A., McMahon, J., Giomarelli, B., & Ghirlanda, G. (2009). Multivalent interactions with gp120 are required for the anti-HIV activity of Cyanovirin. *Biopolymers, 92*(3), 194-200.

Margulis, C. J. (2005). Computational study of the dynamics of mannose disaccharides free in solution and bound to the potent anti-HIV virucidal protein cyanovirin. *J Phys Chem B, 109*(8), 3639-3647.

Matei, E., Furey, W., & Gronenborn, A. M. (2008). Solution and crystal structures of a sugar binding site mutant of cyanovirin-N: no evidence of domain swapping. *Structure, 16*(8), 1183-1194.

Matei, E., Zheng, A., Furey, W., Rose, J., Aiken, C., & Gronenborn, A. M. (2010). Anti-HIV activity of defective cyanovirin-N mutants is restored by dimerization. *J Biol Chem, 285*(17), 13057-13065.

Mori, T., Barrientos, L. G., Han, Z., Gronenborn, A. M., Turpin, J. A., & Boyd, M. R. (2002). Functional homologs of cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. *Protein Expr Purif, 26*(1), 42-49.

Mori, T., Gustafson, K. R., Pannell, L. K., Shoemaker, R. H., Wu, L., McMahon, J. B., & Boyd, M. R. (1998). Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein Expr Purif, 12*(2), 151-158.

Mori, T., Shoemaker, R. H., Gulakowski, R. J., Krepps, B. L., McMahon, J. B., Gustafson, K. R., . . . Boyd, M. R. (1997). Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. *Biochem Biophys Res Commun, 238*(1), 218-222.

Novagen. (2002). T7Select System Manual.

O'Keefe, B. R., Shenoy, S. R., Xie, D., Zhang, W., Muschik, J. M., Currens, M. J., . . . Boyd, M. R. (2000). Analysis of the interaction between the HIV-inactivating protein cyanovirin-N and soluble forms of the envelope glycoproteins gp120 and gp41. *Mol Pharmacol, 58*(5), 982-992.

O'Keefe, B. R., Smee, D. F., Turpin, J. A., Saucedo, C. J., Gustafson, K. R., Mori, T., . . . Boyd, M. R. (2003). Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother, 47*(8), 2518-2525.

Patsalo, V., Raleigh, D. P., & Green, D. F. (2011). Rational and computational design of stabilized variants of cyanovirin-N that retain affinity and specificity for glycan ligands. *Biochemistry, 50*(49), 10698-10712.

Percudani, R., Montanini, B., & Ottonello, S. (2005). The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins, 60*(4), 670-678.

Prien, J. M., Ashline, D. J., Lapadula, A. J., Zhang, H., & Reinhold, V. N. (2009). The high mannose glycans from bovine ribonuclease B isomer characterization by ion trap MS. *J Am Soc Mass Spectrom, 20*(4), 539-556.

Rosenberg, A. (1996). T7Select Phage Display System: A powerful new protein display system based on bacteriophage T7. *inNovations, 6*.

Ruben, M., Sun, H., Maxwell, J., Woodrum, B., O'Keefe, B., & Ghirlanda, G. (2013). Directed evolution of gp120 binding proteins. Manuscript in preparation.

Scanlan, C. N., Offer, J., Zitzmann, N., & Dwek, R. A. (2007). Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature, 446*(7139), 1038-1045.

Scudiero, D. (1988). Evaluation of a Soluble Tetrazolium/Formazan Assay for Cell Growth and Drug Sensitivity in Culture Using Human and Other Tumor Cell Lines. *Cancer Research, 48*, 4827-4833.

Shenoy, S. R., Barrientos, L. G., Ratner, D. M., O'Keefe, B. R., Seeberger, P. H., Gronenborn, A. M., & Boyd, M. R. (2002). Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. *Chem Biol, 9*(10), 1109-1118.

Shenoy, S. R., O'Keefe, B. R., Bolmstedt, A. J., Cartner, L. K., & Boyd, M. R. (2001). Selective interactions of the human immunodeficiency virus-inactivating protein cyanovirin-N with high-mannose oligosaccharides on gp120 and other glycoproteins. *J Pharmacol Exp Ther, 297*(2), 704-710.

Shigapova, N., Torok, Z., Balogh, G., Goloubinoff, P., Vigh, L., & Horvath, I. (2005). Membrane fluidization triggers membrane remodeling which affects the thermotolerance in Escherichia coli. *Biochem Biophys Res Commun, 328*(4), 1216-1223.

Sigma. *Technical Bulletin: Ribonuclease B Glycoprotein Standard, Proteomics Grade*. In Sigma (Ed.), (Vol. Product Number R1153).

Smith, G. P., & Petrenko, V. A. (1997). Phage Display. *Chem Rev, 97*(2), 391-410.

Varki, A., Cummings, R., Esko, J., Freeze, H., Stanley, P., Bertozzi, C., Hart, G., Etzler, M. (Eds.). (2009). *Essentials of Glycobiology* (Second ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Wilen, C. B., Tilton, J. C., & Doms, R. W. (2012). HIV: Cell Binding and Entry. *Cold Spring Harb Perspect Med, 2*(8). doi: 10.1101/cshperspect.a006866

Xiong, S., Fan, J., & Kitazato, K. (2010). The antiviral protein cyanovirin-N: the current state of its production and applications. *Appl Microbiol Biotechnol, 86*(3), 805-812.

Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., . . . Wlodawer, A. (1999). Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol, 288*(3), 403-412.

Yeh, J. C., Seals, J. R., Murphy, C. I., van Halbeek, H., & Cummings, R. D. (1993). Site-specific N-glycosylation and oligosaccharide structures of recombinant HIV-1 gp120 derived from a baculovirus expression system. *Biochemistry, 32*(41), 11087-11099.

Zhu, X., Borchers, C., Bienstock, R. J., & Tomer, K. B. (2000). Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry, 39*(37), 11194-11204.

APPENDIX A

CONSORTIUM FOR FUNCTIONAL GLYCOMICS DATA

APPENDIX A

  Appendix A contains raw data from the Consortium for Functional Glycomics glycan array. The 10 or 11 glycans with the highest average RFU for each CV-N mutant are described. Please note that SSDGLQQ-P51Gm4-CVN was screened on Mammalian Printed Array version 5.0 and all other proteins were screened on Mammalian Printed Array version 5.1. These versions of the array are nearly identical. Version 5.0 has 611 glycans and version 5.1 has 610 glycans, with the only difference being the absence of Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13 (glycan 56 on version 5.0) on version 5.1 of the array.
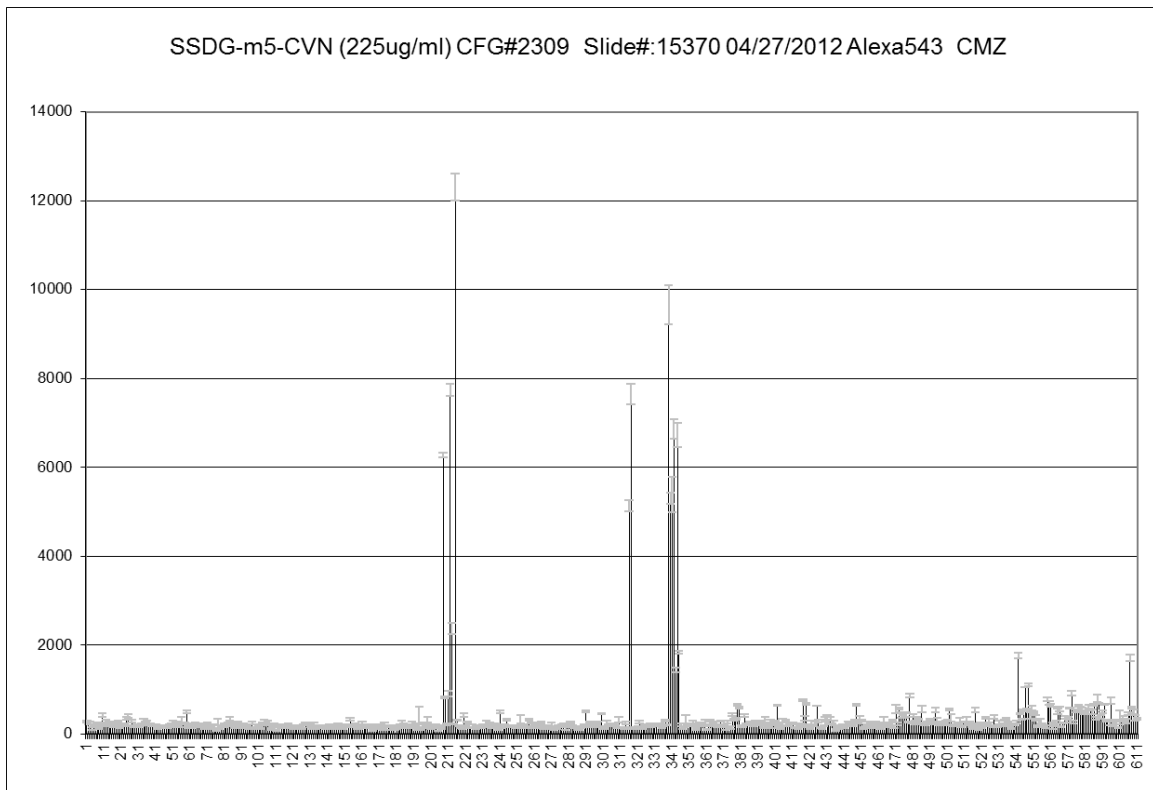
Figure 28. Glycomics array results for SSDGLQQ-P51Gm4-CVN (N-term 6His-tag) on CFG Mammalian Printed Array version 5.0. Two specificities were identified. 1) Manα1-2Manα1-2Man; 2) GlcNAcα1-4Galβ1-4GlcNAcβ and GlcNAcα1-4Galβ1-3GlcNAcβ.

Table 9

Top Hits for SSDGLQQ-P51Gm4-CVN (N-term 6His-tag) on CFG Mammalian Printed Array V 5.0

Chart Number	SSDG-m5-CVN (225ug/ml) CFG#2309 Slide#:15370 04/27/2012 Alexa543 CMZ	Average RFU	StDev
215	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	12306	592
339	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	9654	882
212	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	7736	278
317	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	7651	461
342	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	6869	438
344	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	6728	553
208	Mana1-2Mana1-2Mana1-3Mana-Sp9	6275	104
341	GlcNAca1-4Galb1-3GlcNAcb-Sp0	5390	803
340	GlcNAca1-4Galb1-4GlcNAcb-Sp0	5298	241
316	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	5135	238

Note: Version 5.0. SSDGLQQ-P51Gm4-CVN is specific for glycans containing Manα1-2Manα and GlcNAcα1-4Galβ
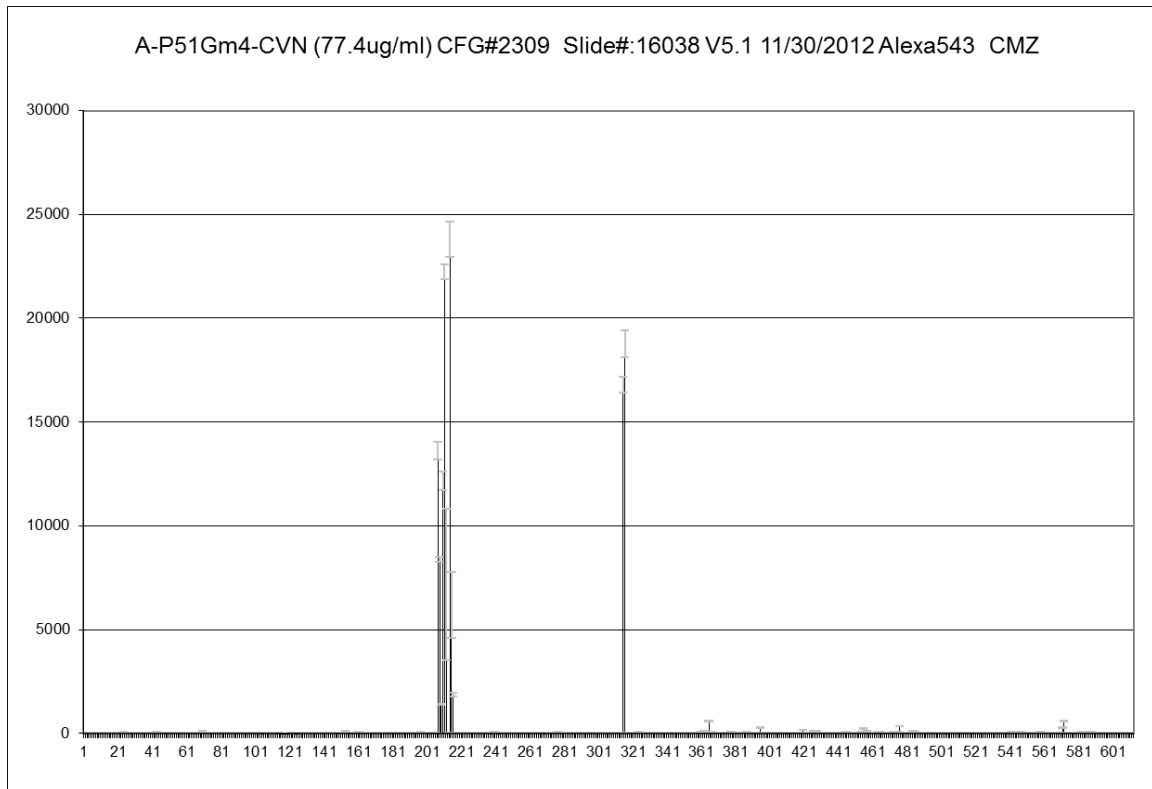
Figure 29. Glycomics array results for AAGRLSK-P51Gm4-CVN (C-term 6His-tag) on CFG Mammalian Printed Array version 5.1.

Table 10

Top Hits for AAGRLSK-P51Gm4-CVN (C-term 6His-tag) on GFC Mammalian Printed Array V 5.1

Chart Number	A-P51Gm4-CVN (77.4ug/ml) CFG#2309 Slide#:16038 V5.1 11/30/2012 Alexa543 CMZ	Average RFU
214	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	23805
211	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	22227
316	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	18774
315	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	16784
207	Mana1-2Mana1-2Mana1-3Mana-Sp9	13632
210	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	12168
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	8399
212	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	7184
215	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	6190
216	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	1867
209	Mana1-2Mana1-3Mana-Sp9	1395
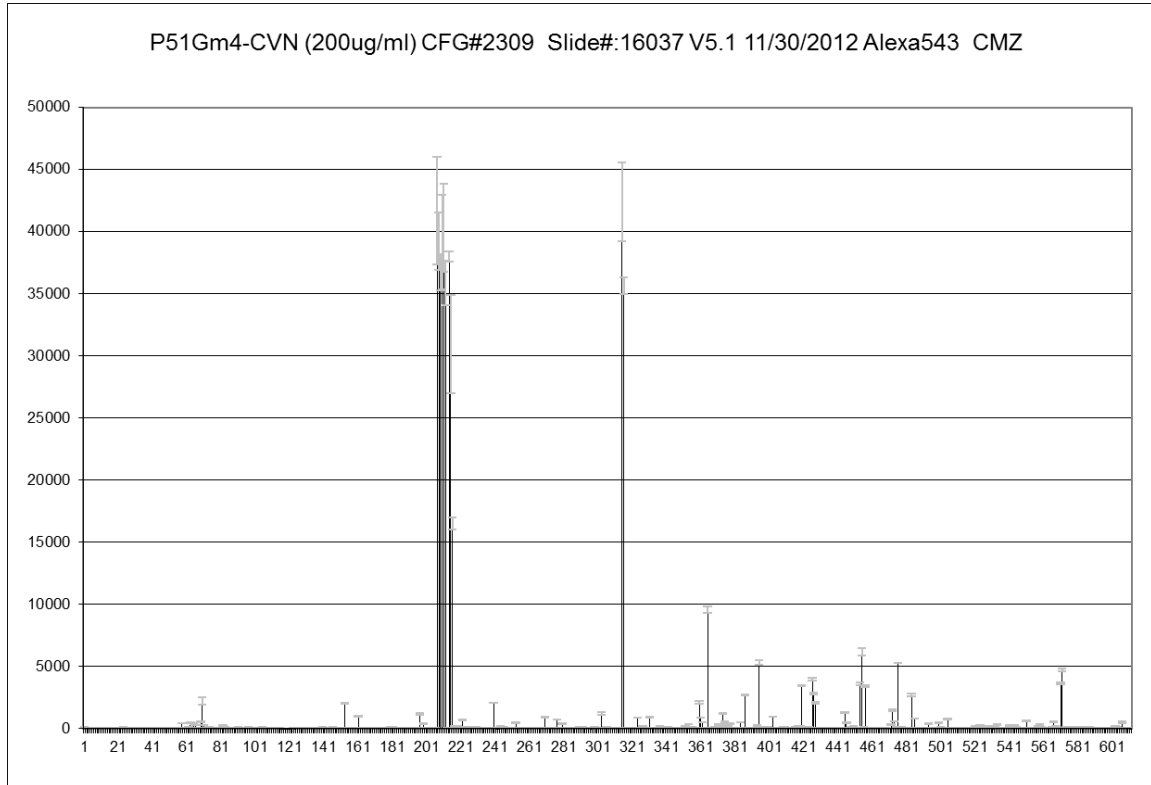
Figure 30. Glycomics array results for P51G-m4-CVN (C-term 6His-tag) on CFG Mammalian Printed Array version 5.1.

Table 11

Top Hits for P51G-m4-CVN (C-term 6His-tag) on GFC Mammalian Printed Array V 5.1

Chart Number	P51Gm4-CVN (200ug/ml) CFG#2309 Slide#:16037 V5.1 11/30/2012 Alexa543 CMZ	Average RFU
315	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	42394
207	Mana1-2Mana1-2Mana1-3Mana-Sp9	41677
211	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	40286
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	39218
210	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	39138
214	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	38018
209	Mana1-2Mana1-3Mana-Sp9	36735
212	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	35885
316	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	35646
215	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	30941
216	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	16499
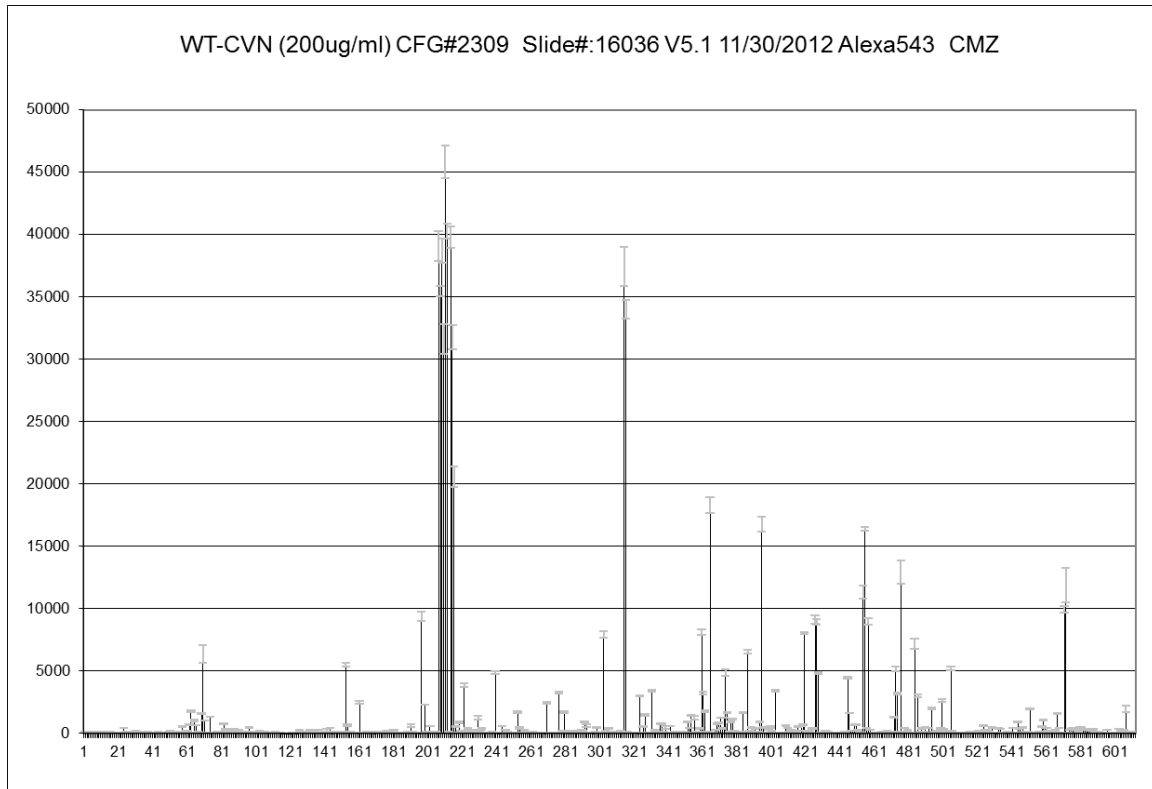
Figure 31. Glycomics array results for wt-CVN (C-term 6His-tag) on CFG Mammalian Printed Array version 5.1.

Table 12

Top Hits for wt-CVN (C-term 6His-tag) on GFC Mammalian Printed Array V 5.1

Chart Number	WT-CVN (200ug/ml) CFG#2309 Slide#:16036 V5.1 11/30/2012 Alexa543 CMZ	Average RFU
211	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	45829
212	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	40272
214	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	39790
207	Mana1-2Mana1-2Mana1-3Mana-Sp9	39071
209	Mana1-2Mana1-3Mana-Sp9	38695
315	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	37406
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	35476
316	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	34006
215	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	31762
210	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	31625
216	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	20562

APPENDIX B

PERMISSION FOR FIGURES

FIGURE 1 PERMISSION

March 11, 2013

Melissa Ruben
Arizona State University
Chemistry and Biochemistry
Tempe, AZ 85287-1604

Email: mruben@asu.edu

Dear Ms. Ruben:

This is to grant you permission to reproduce the following figure in your thesis titled "Directed Evolution of gp120 Binding Mutants of the Lectin Cyanovirin-N" for Arizona State University/ProQuest/UMI:

> Figure 1 from Anders J. Bolmstedt, Barry R. O'Keefe, Shilpa R. Shenoy, James B. McMahon, and Michael R. Boyd, Cyanovirin-N Defines a New Class of Antiviral Agent Targeting N-Linked, High-Mannose Glycans in an Oligosaccharide-Specific Manner, *Mol Pharmacol* May 1, 2001 59:949-954

Permission is also granted to include the figure in any future revisions and editions of your dissertation, including non-exclusive world rights in all languages, and to the publication of your dissertation by UMI. UMI may supply copies of your dissertation on demand.

Sincerely yours,

Richard Dodenhoff
Journals Director

American Society for Pharmacology and Experimental Therapeutics

FIGURE 2 PERMISSION

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Melissa Ruben
Customer address	Arizona State University
	Tempe, AZ 85287-1604
License number	3104860799632
License date	Mar 09, 2013
Licensed content publisher	Elsevier
Licensed content publication	Biochemical and Biophysical Research Communications
Licensed content title	Isolation, Primary Sequence Determination, and Disulfide Bond Structure of Cyanovirin-N, an Anti-HIV(Human Immunodeficiency Virus) Protein from the Cyanobacterium*Nostoc ellipsosporum*
Licensed content author	Kirk R. Gustafson,Raymond C. Sowder,Louis E. Henderson,John H. Cardellina,James B. McMahon,Umamaheswari Rajamani,Lewis K. Pannell,Michael R. Boyd
Licensed content date	8 September 1997
Licensed content volume number	238
Licensed content issue number	1
Number of pages	6
Start Page	223
End Page	228
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	4

FIGURE 3 PERMISSION

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Melissa Ruben
Customer address	Arizona State University
	Tempe, AZ 85287-1604
License number	3104870002143
License date	Mar 09, 2013
Licensed content publisher	Elsevier
Licensed content publication	Structure
Licensed content title	Solution Structure of a Cyanovirin-N:Man! 1-2Man! Complex: Structural Basis for High-Affinity Carbohydrate-Mediated Binding to gp120
Licensed content author	Carole A Bewley
Licensed content date	October 2001
Licensed content volume number	9
Licensed content issue number	10
Number of pages	10
Start Page	931
End Page	940
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic

FIGURE 6 PERMISSION

March 20, 2013

Melissa Ruben
mruben@asu.edu

Dear Melissa,

In response to your request regarding permission to adapt the following material:

Item:	Fig 1
Title:	*Cold Spring Harbor Perspectives in Medicine*
Article:	HIV: Cell Binding and Entry
Author:	Wilen et al
Copyright:	2012

To be used in:

Title:	PhD Thesis
Author:	Melissa Ruben
Publisher:	University of Arizona
Publication date:	2013

I confirm that we are willing to grant non-exclusive rights in the English language, subject to the following conditions:

1. Full acknowledgment shall be made to the source and correct references given. <u>Copyright holder is Cold Spring Harbor Laboratory Press.</u>   Please cite this article as *Cold Spring Harb Perspect Med 2012;2:a006866*
2. The permission of the author is obtained, where possible.
3. Permission is limited to the proposed publication, on a non-exclusive, ***one-time-only basis***, with distribution rights in the English language only. This permission relates to publication in the English language for print and electronic rights only. ***Rights do not apply to revised editions.***
4. Fee is waived for reprinting in PhD thesis. If material is used in any other publication, you must reapply for permission and fee will be determined at that time.

*Please note that we do not grant blanket permissions for future editions or translations. If you need any of the aforementioned, please reapply and include details of usage.

By Carol C. Brown   **Date: 3/20/13**
For Cold Spring Harbor Laboratory Press, 500 Sunnyside Blvd., Woodbury NY 11797, USA

Tel. (516) 422-4038/ e-mail: brown@cshl.edu / Fax (516) 422-4095

FIGURE 7 PERMISSION

FIGURE 21 PERMISSION