UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ENGINEERING CHO CELLS TO IMPROVE THE QUALITY AND IMMUNOGENICITY OF HIV ENVELOPE GLYCOPROTEIN SUBUNIT VACCINES

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Sophia W. Li

June 2020

The Dissertation of Sophia W. Li is approved:

Professor Phillip Berman, chair

Professor Seth Rubin

Professor Rebecca DuBois

Professor Chad Saltikov

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Supplemental Figures

List of Tables

Abstract

Engineering CHO Cells to Improve the Quality and Immunogenicity of HIV Envelope Glycoprotein Subunit Vaccines

Sophia W. Li

A major focus of HIV vaccine development has been improving the envelope glycoprotein 120 (gp120) derived immunogens used in the RV144 HIV vaccine trial. RV144 has been the only human clinical trial to demonstrate modest but significant protection in humans. Over the last 30 years it has been difficult to manufacture Clade B immunogens, such as MN-rgp120 used in the RV144 trial. Clade B immunogens are representative of viruses from North America, Europe and Australia and become proteolyzed when expressed in Chinese Hamster Ovary cells. In this report, we identified complement component 1s (C1s) as the protease responsible for cleavage in the third variable region (V3) at a site recognized by neutralizing antibodies. To solve this problem, we used CRISPR/Cas9 to inactivate the C1s-gene of several CHO cell lines including the standard CHOS, CHOK1 cell lines and the glycan restricted MGAT1^{-} CHO cell line. These novel cell lines provide the means to produce Clade B immunogens as well as other recombinant proteins without proteolysis. Proteolysis of recombinant proteins is a common problem in the biopharmaceutical industry. Another medically significant protein that also becomes proteolyzed during expression in CHO cells is human Factor VIII

(FVIII). Biochemical analysis was carried out to see if unintended cleavage of FVIII could also be attributed to C1s. Differences between the substrate specificity of human thrombin and CHO C1s were elucidated using mutagenesis and homology modeling and showed that FVIII proteolysis was due to another protease. In other studies, we were interested in enhancing the immunogenicity of epitopes in the second variable region (V2) of A244 gp120 that correlated with protection in the RV144 clinical trial. This was accomplished through the characterization of 10C10, a mouse monoclonal antibody elicited by immunization with A244-gp120. 10C10 was shown to have cross-reactivity to clade AE, B and C gp120s and bind to a V2 peptide through protein- and peptide-binding studies. In the RV144 trial, non-neutralizing antibodies to the V2 region were found to be a correlate of protection. Similar to the CH58-like antibodies that have been identified, 10C10 was found to be non-neutralizing and bind to a V2 peptide that assumed an alpha-helical conformation as determined by computational modeling. Preventing proteolysis and targeting antibodies to the 10C10 site will help guide the development of the next generation of vaccine candidates to achieve the goal of making a viable HIV vaccine.

Acknowledgements

I would like to thank my advisor, Dr. Philip Berman, who challenged me to the best scientist I could be. I was given many wonderful opportunities to try new projects and learned so much along the way. I am indebted to all the great people in the Berman lab who contributed to the most engaging scientific discussions and educated me on the best lab techniques: Sara O'Rourke, Kate Mesa, Bin Yu, Meredith Wright, Lucy Yin, Leonardo Ramirez, Gabriel Byrne, Jennie Hutchinson, Rachel Doran, Chelsea Didinger, and Dave Alexander.

Lastly, I want to thank my parents, whom have raised me to become the person I am today; my friends, whom have shown me how to live life to the fullest; and my boyfriend, whom has been my rock through these last four years of my Ph.D. Chris, I could not have asked for a more wonderful partner in my life.

Introduction

Cocktails of anti-retroviral drugs have allowed people infected with HIV-1 to live normal lives in first world countries. However, a lack of access to good healthcare in rural areas of Africa and other third world countries have prevented HIV-infected individuals from receiving proper treatment. 34% of people in east and southern Africa and 60% of people in west and central Africa who are living with HIV are not receiving treatment (Dwyer-Lindgren et al., 2019). AIDS is the most common cause of death in sub-Saharan Africa (Roth et al., 2018). A vaccine would offer protection for an extended period without the need for continued medication and would benefit those who cannot access treatment.

The HIV envelope protein is the major focus of vaccine development efforts. It mediates binding to the CD4 and chemokine cell surface receptors and is the major target of virus neutralizing antibodies. Gp120 (glycoprotein 120) is the monomer form of the envelope protein and is non-covalently attached to gp41 to form the trimeric spike protein on the virus surface. It is approximately 510 amino acids in length, contains nine disulfide bridges and possesses 25 or more predicted N-linked glycosylation sites (PNGS). Gp120 is a complex glycoprotein and contains five constant (C) regions and five variable (V) regions. Constant regions are tucked into the core of the envelope protein (Env) and the sequence of these constant regions remains relatively unchanged. The variable regions of the Env are exposed to the surroundings and undergo a high rate of mutagenesis, allowing the

virus to evade the immune system. The V1/V2 loop occurs at the apex of the virus spike and stabilizes the trimeric structure. It contains multiple glycans, which shield the V3 loop. The V3 loop recognizes the T-cell specific surface co-receptors, CXCR4 or CCR5, while the CD4 binding site spans the in the C4 and V4 regions. The variable loops are important epitopes of Env vaccines as they are highly immunogenic and have been found to be the target of broadly neutralizing antibodies.

The RV144 HIV vaccine trial has been the only human clinical trial to show that vaccination can prevent HIV infection. Although not statistically significant, vaccine efficacy was as high as 60.5% at 6 months but decreased to a statistically significant 31.2% (P=0.04) after three years (Rerks-Ngarm et al., 2009). This vaccine consisted of a canarypox vector vaccine in combination with a bivalent mixture of gp120s from the clade CRF01-AE virus, A244 and the clade B MN virus. In the STEP and Phambili HIV vaccine trials, the adenovirus vector vaccine coding for the HIV internal proteins would ideally elicit protective cytotoxic CD8+ T cells (Buchbinder et al., 2008; Gray et al., 2011). However, these trials were halted after the number of HIV infections increased. Thus, subunit vaccines, like gp120, that primarily stimulate antibody responses appear to have less safety concerns that recombinant virus vector vaccines that elicit cellular immune responses.

Recently, strategies have been developed to produce trimeric envelope vaccines that replicate virus spike structures. These SOSIP trimers include a

disulfide bond between gp120 and gp41 and an isoleucine to proline mutation to keep the gp41 subunits in their prefusion form and stabilize the trimer structure. While engineered trimer Envs have been critical in structural studies to understand the full trimer structure of the HIV envelope protein and trimer-specific epitopes recognized by antibodies, they remain limited in their protective capacity. Env trimers have been unable to neutralize tier 2 viruses or elicit broadly neutralizing antibodies. Thus, it remains necessary to make improvements to the RV144 trial immunogens and increase their efficacy and length of protection.

Although it has been possible to produce recombinant gp120s from many clades of HIV, it has proved difficult to produce gp120 from clade B viruses. Clade B viruses are important because they represent the majority of viruses transmitted in North America, Europe, the Caribbean, and Australia. MN-rgp120, the clade B immunogen used in the RV144 trial, was derived from the MN strain of HIV, the prototypic clade B isolate. Proteolysis of MN-rgp120 and other clade B immunogens occurs when they are expressed in Chinese Hamster Ovary (CHO) cells, the main cellular substrate used in the pharmaceutical industry for expression of recombinant proteins. This proteolysis occurs at the V3 region, a highly immunogenic region that can elicit neutralizing antibodies and broadly neutralizing antibodies such as 447-52D and PGT128. Proteolysis of the epitope will impact how well the immunogen can elicit V3 antibodies. Chapter 2 discusses the use of mass spectrometry to identify complement component 1s (C1s), a serine protease in the complement cascade. CRISPR/Cas9 was used to knockout C1s in the

MGAT1- CHO cell line. MGAT1 is a glycosyltransferase that is necessary for complex glycans formation. Previously, our lab described the MGAT1- CHO cell line which was used to express HIV immunogens with high-mannose glycans. Thus, the $C1s^{-/-}$ MGAT1- CHO cell line described here has both prevented proteolysis and enhanced glycosylation of HIV immunogens. Chapter 3 discusses the transient expression of several clade B Envs in the $C1s^{-/-}$ CHO cell line, the investigation into the proteolysis of human Factor VIII by C1s and determining the substrate specificity of CHO C1s.

Identification of broadly neutralizing antibodies and non-neutralizing antibodies have guided design of vaccine immunogens. We immunized mice with the RV144 immunogen A244 to generate antibodies that could be used to identify other promising A244-like immunogens. 10C10, a non-neutralizing V2-directed mouse monoclonal with cross-clade reactivity to antigens in clades AE, B and C was identified. Non-neutralizing V2 antibodies that mediate antibody-dependent cellular cytotoxicity were found to be a correlate of protection in the RV144 trial. Study of 10C10 and other non-neutralizing V2 antibodies will help us understand how to improve upon immunogens such as A244 to increase the protection provided by these V2 antibodies. Chapter 4 discusses the characterization of the epitope and structure of 10C10.

One of the features of HIV is that it targets and weakens the immune system, leaving people vulnerable to even minor illnesses. We discovered a sequence on the HIV envelope protein that was similar to a sequence found on CTLA4 and

CD28, the negative and positive regulators of T cells. This conserved sequence on CTLA4 and CD28 appears at the site of interaction with its ligands. Upregulation of CTLA4 has been observed in HIV-infected individuals though the reason behind this is unclear. Chapter 5 discusses the CTLA-like sequence found on HIV and our findings that Chonblock, a commercial blocking buffer, was necessary to eliminate false positive binding caused by rabbit serum proteins in ELISA studies.

Work from the two C1s related papers in Chapters 2 and 3 resulted in the $C1s^{-/-}$ CHO and $C1s^{-/-}$ $MGAT1^{-}$ CHO cell lines. The $C1s^{-/-}$ CHO cell line can be used for the expression of any recombinant protein that is susceptible to proteolysis by C1s and is not limited to only expression of HIV immunogens. We also combined knockout of the C1s protease with the knockout of MGAT1 for expression of non-proteolyzed recombinant proteins with high mannose glycans. Knockout of MGAT1, a glycosidase enzyme, restricts the glycosylation of proteins to only high mannose glycans. We previously showed that this enhanced binding of immunogens to glycan-dependent broadly neutralizing monoclonal antibodies such as PG9. The generation of the $C1s^{-/-}$ CHO and $C1s^{-/-}$ $MGAT1^{-}$ CHO cell lines are detailed in the technical reports in chapters 6 and 7.

Chapter 1: Identification and CRISPR/Cas9 knockout of the endogenous C1s protease in CHO cells eliminates aberrant proteolysis of recombinantly expressed proteins

This chapter contains text and figures from the following manuscript:

Li SW, Yu B, Byrne G, Wright M, O'Rourke S, Mesa K, et al. Identification and CRISPR/Cas9 Inactivation of the C1s Protease Responsible for Proteolysis of Recombinant Proteins Produced in CHO Cells. Biotechnology and Bioengineering. 2019 Sep;116(9):2130–45.

Introduction

Despite the availability of anti-retrovirals, there is still an urgent need for a vaccine that protects against HIV (Fauci and Marston, 2014). The HIV envelope (Env) glycoprotein, gp120, was a major component of the multivalent vaccine used in the RV144 clinical trial, the only trial to demonstrate protection in humans (Berman et al., 1999; Rerks-Ngarm et al., 2009). Subsequent studies suggested that protection could be attributed to antibodies to gp120 rather than cellular immune responses (Haynes et al., 2012; O'Connell et al., 2012). There is interest in improving gp120s and Env-based vaccines such as developing a scalable and cost-effective recovery process suitable for large scale manufacturing and increasing the potency and breadth of protection. However, problems with proteolysis have

prevented current research developments from being applied to vaccine immunogens from clade B viruses, which circulate in North America, Europe, and other regions of the world (Junqueira and Almeida, 2016). Here, we address the problem of proteolysis of clade B vaccine immunogens in Chinese Hamster Ovary cells (CHO), the standard cell line used in the biopharmaceutical production of recombinant proteins.

Although proteolysis of gp120 expressed in CHO cells has been observed for over 30 years, the protease responsible for this has never been identified due to the lack of an annotated CHO genome (Berman et al., 1990; Stephens et al., 1990; Clements et al., 1991; Werner and Levy, 1993; Scandella et al., 1993; Schulz et al., 1993; Du et al., 2008; Pugach et al., 2015). When clade B gp120 is produced in CHO cells, it is degraded into 70 kDa and 50 kDa fragments by a secreted, endogenous serine protease. Trimers, such as the B41 SOSIP.664 trimer, also undergo proteolysis at the V3 domain (Pugach et al., 2015). Our experiments are with the gp120 monomer, but our findings should apply to both monomeric and trimeric forms of the vaccine. Proteolysis occurs at the Gly-Pro-Gly-Arg (GPGR↓AF) motif at the crown of the V3 domain of gp120. This sequence is present in 71% of clade B envelope proteins and is also present in other clades of HIV (Foley and Korber, 1996). The V3 domain is important as V3 domain-specific antibodies directed to both the crown and the stem of the V3 loop effectively neutralize multiple strains of HIV (LaRosa et al., 1990; Zolla-Pazner et al., 2004; Burton et al., 2005).

Eliminating this proteolytic activity is necessary for producing unclipped clade B antigens with the correct antigenic structure required to elicit protective antibody responses. Solutions to mitigate proteolysis of clade B envelope proteins are not ideal due to reduced protein recovery, the need for media additives or alteration of the antigen sequence. Purification by a monoclonal antibody that spans the protease cleavage site resulted in low yields of unclipped protein from the proteolyzed, secreted protein (Pitisuttithum et al., 2006; Berman, 2015). Changes to the fermentation process including the lowering of cell culture temperatures; the addition of cell culture additives such as fetal calf serum, EDTA or EGTA (Du et al., 2008; Srivastava et al., 2002; Chakrabarti et al., 2016); and the use of continuous flow bioreactor systems (Srivastava et al., 2003, 2008) were associated with lower protein expression and recovery. Lastly, substitution of the GPGR cleavage motif for GPGQ, a common sequence in other clades, removes the cleavage site recognized by the protease and makes gp120 no long susceptible to proteolysis, but would prevent elicitation of 447-52D-like antibodies (Sullivan et al., 2017; Wen et al., 2018). 447-52D effectively neutralize 92% of 32 primary isolates containing the GPGR motif (Stanfield et al., 2004; Zolla-Pazner et al., 2004). Thus, a strategy to develop a vaccine effective against clade B strains is to retain the GPGR motif and induce 447-52D-like antibodies.

In this paper, we describe the use of conventional column chromatography, tandem mass spectroscopy and the annotated CHO genome to identify complement component 1s (C1s) as the serine protease that proteolyzes clade B gp120s. We use

CRISPR/Cas9 to engineer the $C1s^{-/-}$ $MGAT1^{-}$ CHO cell line to show that proteolysis of gp120 is eliminated with the knockout of the C1s protease. Additionally, we show that knockout of the C1s protease can be combined with knockout of the MGAT1 glycosyltransferase (mannosyl alpha-1,3 -glycoprotein beta-1,2-N-acetylglucosaminyl-transferase) to express gp120 that is unclipped and possesses the oligomannose structures required for the binding of multiple glycan dependent bNAbs. This cell line will enable the development and testing of vaccine concepts based on clade B vaccine immunogens and the production of unclipped, properly glycosylated Env proteins in quantities sufficient for clinical trials.The C1s-deficient CHO cell line is also a proof-of-concept demonstrating that engineered CHO cells lines with knockout of C1s or other CHO proteases can be used to mitigate the proteolysis of biologics.

Materials and Methods

HIV Env Proteins/Antibodies: A synthetic gene encoding BaL-rgp120 was created using GenBank#AY713409. Recombinant MN, A244 and TZ97008 gp120s were produced in HEK293, CHO-S and MGAT1- CHO-S cell lines as described previously (Byrne et al., 2018a; Doran et al., 2018a). PG9, PGT128, VRC01 were produced based on published sequences. PGT121 and 10-1074 were acquired from the NIH AIDS Reagent Program. The 34.1 mAb to the Herpes Simplex Virus glycoprotein D (gD) was produced in-house (O'Rourke et al., 2019). 11G5 was a

gift from Global Solutions for Infectious Diseases (S. San Francisco, CA). The 447-52D antibody was from the NIH AIDS Reagent Program/Dr. Susan Zolla-Pazner.

Protein Expression and Quantification: Cell culture supernatants or purified proteins were run on NuPAGE 4-12% Bis-Tris gels (Thermo Fisher Scientific, Invitrogen, Carlsbad, CA) in MES buffer (Thermo Fisher) and stained with SimplyBlue (Thermo Fisher). For immunoblots, PAGE gels were transferred using iBlot2 (Thermo Fisher). Membranes were blocked with 5% milk for 1 hour. The primary antibody was incubated at 1.5 ug/ml in 5% milk. The secondary antibody used was Peroxidase AffiniPure bovine anti-goat IgG (Jackson ImmunoResearch, West Grove, PA) at a 1:5000 dilution in 5% milk.

Protease inhibition: 5ug of purified MN-rgp120 protein in PBS was mixed with protease inhibitors at a 1:1 w/w ratio. They were further mixed at 1:9 v/v ratio with CHO supernatant and set at 37ºC overnight. Samples were run on SDS-PAGE with and without dithiothreitol (DTT). Inhibitors: N-(1-Naphthalenylsulfonyl)-Ile-Trp-aldehyde (Enzo Life Sciences, Farmingdale, NY), Z-Phe-Ala-fluoromethylketone (Sigma, St. Louis, MO), protease inhibitor cocktail cOmplete Mini (Roche, Basel, Switzerland), 4-aminobenzamidine (Sigma), chymostatin (Sigma).

Enrichment of protease through conventional column chromatography: A 200ml Superdex S200 size exclusion column (GE Healthcare, Sunnyvale, CA) was

used to purify 15 ml of supernatant using 0.05 M Tris buffered saline, pH 8. 5ug of MN-rgp120 was incubated with 1 ml of 20x concentrated pooled fractions and incubated at 37C overnight. Fractions with proteolytic activity were pooled and loaded onto a 5ml QHP anion exchange column (GE). Buffer A was 20mM $NaPO_4$ and buffer B was 20mM $NaPO_4$, 1M NaCl. 1.4ug of MN-rgp120 was incubated with 60x concentrated QHP pooled fractions at 37°C overnight.

Protein Digestion for Mass Spectroscopy: Protease-enriched samples were concentrated 60x in centrifugal filters with a 10 kDa cutoff (EMD Millipore, Hayward, CA). 15ul of 50 mM NH_4HCO_3, 1.5 ul of 100mM DTT, 10ul of protein sample and 0.5 ul milliQ water were combined and incubated at 55°C for 60 minutes. 3ul 100mM iodoacetamide was added and incubated in the dark at room temperature for 45 minutes. 4ul trypsin at 40 ng/ul was added and incubated overnight at room temperature, followed by incubation for 8 hours with 4 ul chymotrypsin at 40 ng/ul. 4.5 ul of 5% formic acid was added and incubated at room temperature for 10 minutes.

Tandem Mass Spectroscopy: Digested peptides were analyzed on an LTQ-Orbitrap Velos Pro-MS (Thermo Fisher) with a 50mm x 2.1 mm Hypersil GOLD VANQUISH C18 UHPLC column (Thermo Fisher). A linear acetonitrile gradient was run from 5 to 10% for 2 minutes, 10% for 3 minutes, 10 to 80% for 50 minutes, 80 to 95% for 2 minutes, at 95% for 3 minutes, from 95 to 5% in 1 minute and at

5% for 4 minutes. Full-scan MS spectra was acquired (300-2000 m/z). Charge states ≥ 1 were fragmented by collision-induced dissociation. Spectra was processed using Proteome Discoverer v1.4.1.14 (Thermo Fisher) and analyzed against the CHO genome (UniProt-CHO). Parameters: no enzyme specified; dynamic carbamidomethyl modification of cysteine, +57.021 Da; dynamic modification of oxidized methionine, +15.995 Da; strict target FDR of 0.01, relaxed target FDR of 0.05.

Development of a MGAT1- CHO cell line to express BaL-rgp120: A MGAT1^{-} CHO-S cell line expressing BaL was developed by a protocol described previously (O'Rourke et al., 2018a), utilizing the MGAT1- CHO-S (Byrne et al., 2018a). The modified pCDNA3.1 expression vector encoded a codon optimized BaL-rgp120 gene and an N-terminal purification tag from Herpes Simplex Virus glycoprotein D (gD) as described previously (Berman 1998).

CRISPR-Cas9 knockout of C1s in the stable BaL-expressing MGAT1^{-} CHO cell line: The C1s gene was sequenced and verified against the NCBI mRNA RefSeq XM_007646821.2 and protein RefSeq XP_007645011.1. Guide RNAs were cloned into the CRISPR Nuclease Vector (GeneArt, Thermo Fisher) and selected using CRISPy Cas9 target finder (Novo Nordisk, Technical University of Denmark): sgRNA1 GATAATCTCAGGAGGCGTCG, sgRNA2 GCCAGTAAACCGCTCTTCGT, sgRNA3 GGGCTCTTATTGACGAGTAC and

sgRNA4 GTTGACAGCCGCTCATGTTG. To sequence indels, 0.5 x 10^6 cells were spun down and boiled in 10 ul of milliQ water. 5 ul of cell lysate was used for PCR. For single cell cloning, cells were diluted to 0.5 cells/well in 96 well tissue-culture treated microplates (Corning, Corning, NY) and grown for 14-21 days at 37C, 8% CO_2 and 85% humidity. For scale-up protein production, media was supplemented at 10% of culture volume every 3 days with a feed consisting of Proyield Cotton CNE50M-UF (FrieslandCampina, Delhi, NY) and TC Yeastolate Ultra-Filtered (BD Biosciences, San Jose, CA). For blue-white screening, PCR products of the genetic knockout from clones was ligated into the T-Vector pMD20 (Takara Bio, Mountain View, CA). One Shot TOP10 Competent cells were transformed and plated in IPTG and XGal.

Electroporation: Using the Maxcyte STX (MaxCyte Inc., Gaithersburg, MD), 80 x 10^6 cells were transfected with 120 ug of CRISPR Vector DNA ul in the OC-400 cuvette. Cells were incubated at 37℃ for 40 minutes. 30 ml of CHO Growth A media (Irvine Scientific, Santa Ana, CA) was added to flasks and transferred to Kuhner shakers at 125 rpm.

Purification of BaL-rgp120: Supernatant from the BaL-rgp120 stable CHO cell line was passed through a 0.2 um filter (Millipore) and purified by the gD tag using an immunoaffinity column consisting of the 34.1 mAb conjugated to AminoLink Plus Coupling resin (Thermo Fisher). Buffer A was 50 mM Tris, 0.5 M NaCl, pH

7.4 and buffer B was 0.1 M glycine, pH 3. 1M Tris was added to eluate at 1:10 ratio of eluate:buffer for neutralization. Subsequent size exclusion chromatography was performed on an S200 gel filtration column (GE) with TBS, pH 8. Protein concentration was measured by BCA assay (Thermo Fisher).

Thrombin proteolysis of BaL-rgp120: Purified BaL-rgp120 expressed in the MGAT1^{-} CHO cell line was digested by thrombin for use in the fluorescence immunoassay comparing monoclonal antibody binding to intact and proteolyzed BaL. 60 ug of BaL-rgp120 was digested with 8.45 units of thrombin in a total volume of 929.5 ul for 92 hours.

Fluorescence immunoassay (FIA): Fluortrac high binding 96 well plates (Grenier Bio One, Kremsmunster, Austria) were coated for 2 hours with 2 ug/ml of 34.1 mAb. Plates were blocked for 1 hour with 1% BSA and washed 4x with PBS-T. Purified protein was incubated for 2 hours. Antibodies were prepared at 10 ug/ml to 0.003 ug/ml in duplicate. After a 4x wash with PBS-T, secondary antibodies were added at a 1:3000 dilution. Alexa Fluor 488 Goat α-human IgG (Jackson ImmunoResearch), Donkey α-goat IgG (Invitrogen), or Goat α-mouse (Invitrogen) IgG were used. After a 4x PBS-T wash, wells were filled with 200 ul PBS. Plates were read on an EnVision plate reader (Perkin Elmer Inc, Waltham, MA) at excitation/emission wavelengths of 353/485 nm. For direct FIAs, protein was

incubated with plates for 2 hours at 2 ug/ml. Subsequent steps were similarly to the capture FIA. Assays were performed in triplicate.

Results

Characterization of the CHO protease responsible for gp120 degradation using protease inhibitors

To characterize the type of protease responsible for gp120 degradation, protease inhibitors were added to growth-conditioned cell culture supernatants from normal CHO cells and to inhibit proteolysis of MN-rgp120. MN-rgp120 has a molecular weight of 120 kDa. Proteolyzed 70k and 50kDa fragments were observed by SDS-PAGE in reduced samples. The cathepsin L inhibitor, N-(1-Naphthalenylsulfonyl)-Ile-Trp-aldehyde; the cysteine protease inhibitor, Z-Phe-Ala-fluoromethylketone; and chymostatin failed to inhibit proteolysis (Fig. 1A). However, proteolysis was successfully inhibited by the serine and cysteine protease inhibitor, complete mini EDTA-free protease inhibitor cocktail, and by serine protease inhibitor, 4-Aminobenzamidine dihydrochloride (4-AMB) (Fig. 1A, 1B). The inhibition of proteolysis by serine protease inhibitors is consistent with studies that have shown gp120 is cleaved by a thrombin-like, serine protease (Clements et al., 1991; Schulz et al., 1993). Proteolysis was also inhibited by the calcium-chelating agent, egtazic acid (EGTA) and has been previously observed (Fig. 1C) (Scandella et al., 1993). These inhibition experiments indicated that the protease in CHO cell supernatant responsible for degrading gp120 was a calcium-dependent serine protease.

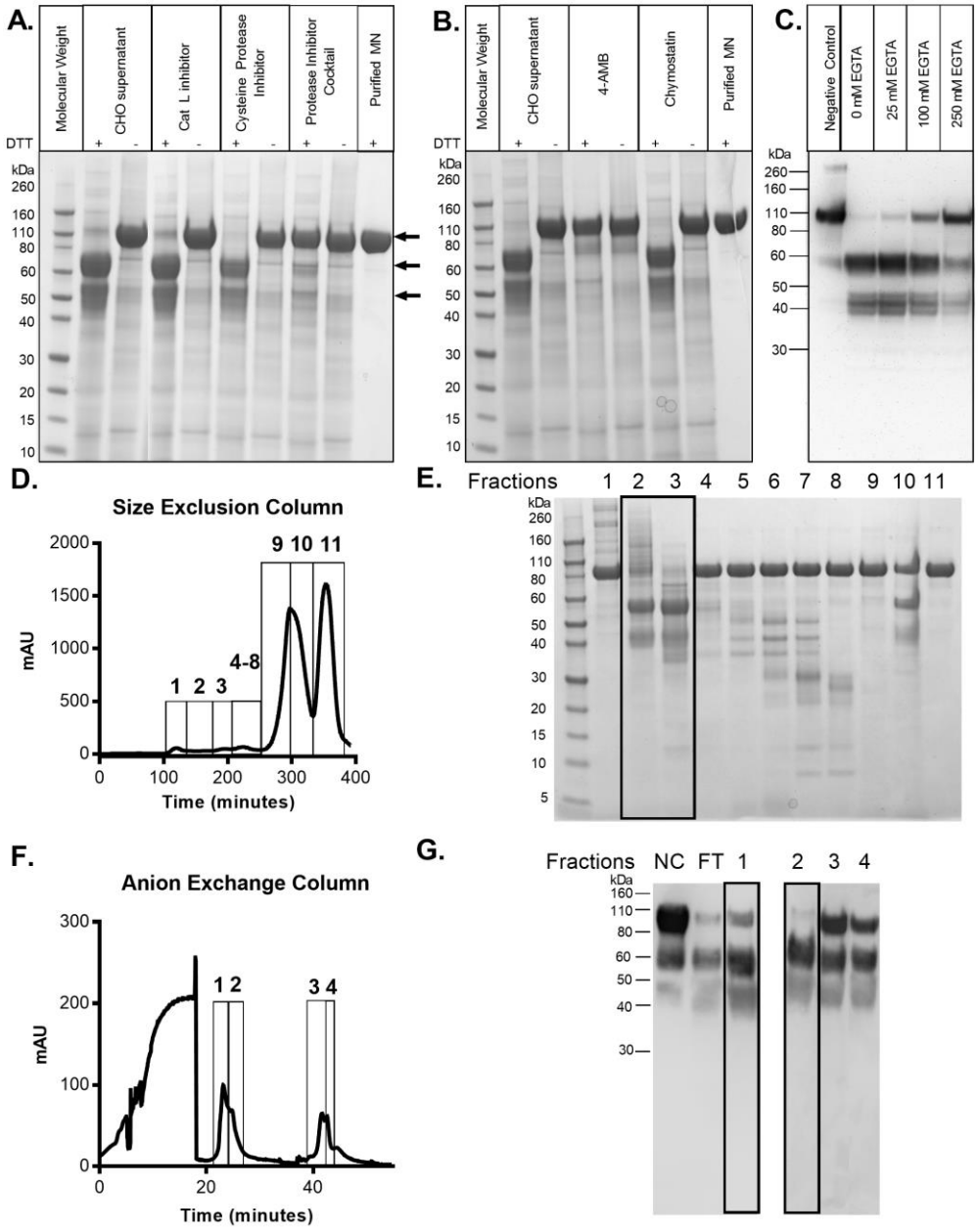

A.
Molecular Weight
CHO supernatant
Cat L inhibitor
Cysteine Protease Inhibitor
Protease Inhibitor Cocktail
Purified MN
DTT
kDa
260
160
110
80
60
50
40
30
20
15
10
B.
Molecular Weight
CHO supernatant
4-AMB
Chymostatin
Purified MN
C.
Negative Control
0 mM EGTA
25 mM EGTA
100 mM EGTA
250 mM EGTA
D.
Size Exclusion Column
mAU
Time (minutes)
1 2 3
4-8
9 10 11
E.
Fractions 1 2 3 4 5 6 7 8 9 10 11
F.
Anion Exchange Column
mAU
Time (minutes)
1 2
3 4
G.
Fractions NC FT 1 2 3 4

Figure 1. Characterization and enrichment of the CHO cell protease responsible for the degradation of clade B, gp120 immunogens.

(A) Purified, unclipped MN-rgp120 produced in HEK293 cells was incubated at 37℃ overnight with growth-conditioned CHO cell culture medium containing the endogenous gp120-cleaving CHO protease. gp120 proteolysis was visualized by the appearance of a 50 kDa and 70 kDa fragment and seen only under reduced conditions by SDS-PAGE. Shown are the results obtained with the cathepsin L inhibitor, N-(1-Naphthalenylsulfonyl)-Ile-Trp-aldehyde; the cysteine protease inhibitor, Z-Phe-Ala-fluoromethylketone; and protease inhibitor cocktail, cOmplete Mini. (B) Panel B shows results obtained with 4-aminobenzamidine, a serine protease inhibitor, and chymostatin, a serine and cysteine protease inhibitor. (C) Panel C shows results obtained with increasing amounts of EGTA, a calcium-chelating agent. In these studies, MN-rgp120 was added to 20X concentrated CHO supernatant and analyzed by immunoblot using a Protein G-purified goat polyclonal antibody raised against gp120. (D) Elution profile of CHO host cell proteins fractionated by size exclusion chromatography using an S200 gel filtration column. (E) Immunoblots showing the mobilities of gp120 protein fragments after overnight treatment at 37℃ of purified MN-rgp120 with fractions from the SEC eluate. (F) Elution profile following anion exchange chromatography with a QHP sepharose column of proteins from fractions 2 and 3 in the SEC. (G) Immunoblots with α-gp120 polyclonal antibody showing the mobilities of gp120 protein fragments after overnight treatment at 37℃ of purified MN-rgp120 with fractions from the QHP column. Fractions 1 and 2 were pooled and 60x concentrated for analysis by tandem MS.

Enrichment of gp120-cleaving protease from CHO supernatant by column chromatography.

The gp120-cleaving protease secreted from CHO cells was enriched using column chromatography to remove CHO host cell proteins before mass spectroscopy analysis. CHO supernatant was first fractionated by size exclusion chromatography (Fig. 1D). Fractions were concentrated, incubated with purified MN-rgp120 overnight to determine fractions containing proteolytic activity.

Fractions 2 and 3 contained the major proteolytic activity, while lesser proteolytic activity was observed in fraction 10 (Fig. 1E). Fractions 2 and 3 were pooled and further separated over a QHP, strong anion exchange column (Fig. 1F). The fractions were concentrated, incubated with purified MN gp120 overnight and run on immunoblot. Fractions 1 and 2 from the QHP eluate displayed greater proteolytic activity than fractions 3 and 4 (Fig. 1G). Subsequently, fractions 1 and 2 were pooled to be run on mass spectroscopy.

Identification of the calcium-dependent serine protease, C1s, by mass spectroscopy.

MS/MS spectroscopy was used to analyze the proteins recovered from the two-column chromatography method described above (Supplemental Fig. 1A). The MS/MS spectrum was analyzed with the Uniprot CHO genome (Xu et al., 2011). The proteins identified were mainly secreted extracellular proteins or those found on the cell surface (Frantz et al., 2010; Park et al., 2017) (Fig. 2A). Only one protease was identified: complement component 1s (C1s). Six unique peptides from C1s were identified (Supplemental Fig. 2), and coverage of the protein was 14.5%, providing high confidence in the identity of this protease. The six unique peptides are labeled by retention time (Supplemental Fig. 1B), and the fragment ion spectrum is shown for peptide SNTLDIVFQTDLTEQR (Supplemental Fig. 1C). Not only was C1s the only protease identified in the sample, it was both a calcium-

dependent protease and a serine protease as predicted by the protease inhibition experiments discussed above.

A.

Uniprot Identifier	Description	Score	Coverage (%)	# Unique Peptides	# Peptides
G3H0E4	Chondroitin sulfate proteoglycan 4	124.16	18.30	14	34
G3HWE4	Nidogen-1	73.91	25.17	10	20
G3GUR0	**Calcium-dependent serine proteinase**	**45.54**	**14.53**	**6**	**9**
G3HN02	Talin-1	13.23	2.60	1	6
G3GXS2	EMILIN-1	5.57	4.62	1	5
G3H354	Heat shock protein HSP 90-alpha	20.54	16.57	3	5
G3H1K9	Alpha-actinin-1	17.42	6.42	4	5
G3I278	Laminin subunit beta-1	14.71	3.90	1	5
G3H8F4	Dystroglycan	9.90	4.37	1	4
G3IFA4	Multiple epidermal growth factor-like domains 6	2.12	4.23	1	4
G3HWE7	Dickkopf-related protein 3	6.44	4.02	1	2
G3I1Y9	Sulfated glycoprotein 1	5.56	8.43	1	2
G3HHR3	Vimentin	13.44	4.94	1	2
G3HC84	Heat shock protein HSP 90-beta	8.39	6.07	1	2
G3H171	G protein-coupled receptor kinase	5.32	2.90	1	2
G3HSX8	Biglycan	3.38	2.98	1	1
G3GYZ1	CD44 antigen	3.72	2.43	1	1
G3ILK7	Calsyntenin-1	4.04	2.16	1	1

B.

Substrate	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'	Cleaved by C1s
BaL gp120	Gly	Pro	Gly	Arg	Ala	Phe	Tyr	Thr	+
Consensus B gp120	Gly	Pro	Gly	Arg	Ala	Phe	Tyr	Thr	+
Consensus A/E gp120	Gly	Pro	Gly	Gln	Val	Phe	Tyr	Arg	-
Consensus C gp120	Gly	Pro	Gly	Gln	Thr	Phe	Tyr	Ala	-
complement component C2	Asn	Leu	Gly	Arg	Arg	Ile	Gln	Ile	+
complement component C4	Gly	Leu	Ala	Arg	Ala	Gln	Glu	Val	+

Figure 2. CHO cell proteins identified in column fractions enriched for CHO protease by tandem MS and protein sequences sensitive to degradation by the C1s protease.

(A) The proteins identified in fractionated CHO cell supernatants by tandem MS were tabulated and listed along with the UniProt identifiers. The score provided by SEQUEST is the sum of the ion scores of all the peptide identified for the protein. Percent coverage indicates the extent of the protein sequence identified from the peptides. The number of unique peptides and total number of peptides are also provided. (B) Alignment of protein sequences from HIV gp120 and CHO cell complement proteins that are sensitive or resistant to degradation by the C1s protease. Standard P1/P1* nomenclature was used to identify conserved residues on either side of the scissle bond**.** Sequences are listed for the V3 crown sequences of BaL-rgp120 and the consensus sequences for clade B, clade CRF01-AE, and Clade C gp120. Also included are alignments of the normal physiologically-relevant substrates of C1s, complement components C2 and C4 (UniProt identifier G3HZE1 and the UniProt identifier UPI0004542963).

C1s is a serine protease in the complement cascade and circulates within the blood plasma. It cleaves the complement proteins, C4 and C2, into their active forms, C4b and C2a, which are required for the progression of the complement cascade and result in opsonization and cell lysis of foreign pathogens (Noris and Remuzzi, 2013). The cleavage site sequences for Chinese hamster C2 and C4 were aligned with the sequence for BaL and consensus sequences for clade B, C and AE viruses. Arginine (R) is required in the P1 position and small amino acids, glycine (G) or alanine (A), are conserved in the P2 position (Fig. 2B). The lack of R at the P1 positions for clade CRF01-AE and clade C gp120s account for their resistance to proteolysis when expressed in CHO cells. While no structural information exists for Chinese hamster C1s, BLAST alignment of the residues in human and Chinese hamster C1s showed 77% homology. The similar cleavage site sequences of clade B gp120 and the complement proteins, C2 and C4, support the observation that these proteins are cleaved by C1s.

The presence of C1s in CHO cells has been previously observed in CHO proteomic experiments, which identified all proteins in CHO supernatant by mass spectroscopy (Baycin-Hizal et al., 2012; Park et al., 2017). Quantitative analysis of these proteins showed that C1s is a relatively highly expressed protein; in the supernatant of batch run cultures, C1s in was in the top 150 most highly expressed proteins over an 8-day culture. The high expression of C1s in CHO supernatant may correlate with the rapid proteolysis observed of clade B gp120s in CHO cells. More recently, a group used next-generation sequencing to observe differential

expression of CHO serine proteases between a CHO-K1 cell line and a CHO-DUXB11 cell line, which produces protein with less proteolytic degradation. C1s was one of the proteases that had significant differences in gene expression (Laux et al., 2018). Thus, our identification of C1s as the gp120-cleaving protease is consistent with other CHO proteomic and protease experiments.

Proteolysis of clade B gp120 occurs in CHO cells, but not in HEK293 cells.

The extent of proteolysis of gp120s antigens with the GPGR motif varies considerably between Envs expressed in CHO cells and another commonly used mammalian cell line, Human Embryonic Kidney (HEK293) cells. To confirm this, the clade B isolates, BaL-rgp120 and EN3-rgp120, both containing the GPGR motif, were transiently expressed in CHO and HEK293 cells. The BaL- and EN3-rgp120s remained intact when expressed in HEK293 cells but were proteolyzed in CHO cells (Fig. 3A), with BaL-rgp120 being completely proteolyzed and EN3-rgp120 partially degraded. The differences in proteolytic activity between CHO and HEK293 cells were consistent with C1s RNA expression data provided by the Human Protein Atlas (Fig. 3B). Although these values are derived from human tissues, the 10-fold higher expression of C1s in ovary tissues over that in kidney tissues may explain why proteolysis of clade B gp120s is observed when expressed in CHO cells, but not in HEK293 cells. Differences in C1s expression have also been observed in mouse tissues where the gene expression of C1s was shown to be 80-fold higher in ovary tissues over kidney tissues (Supplemental. Fig. 3).

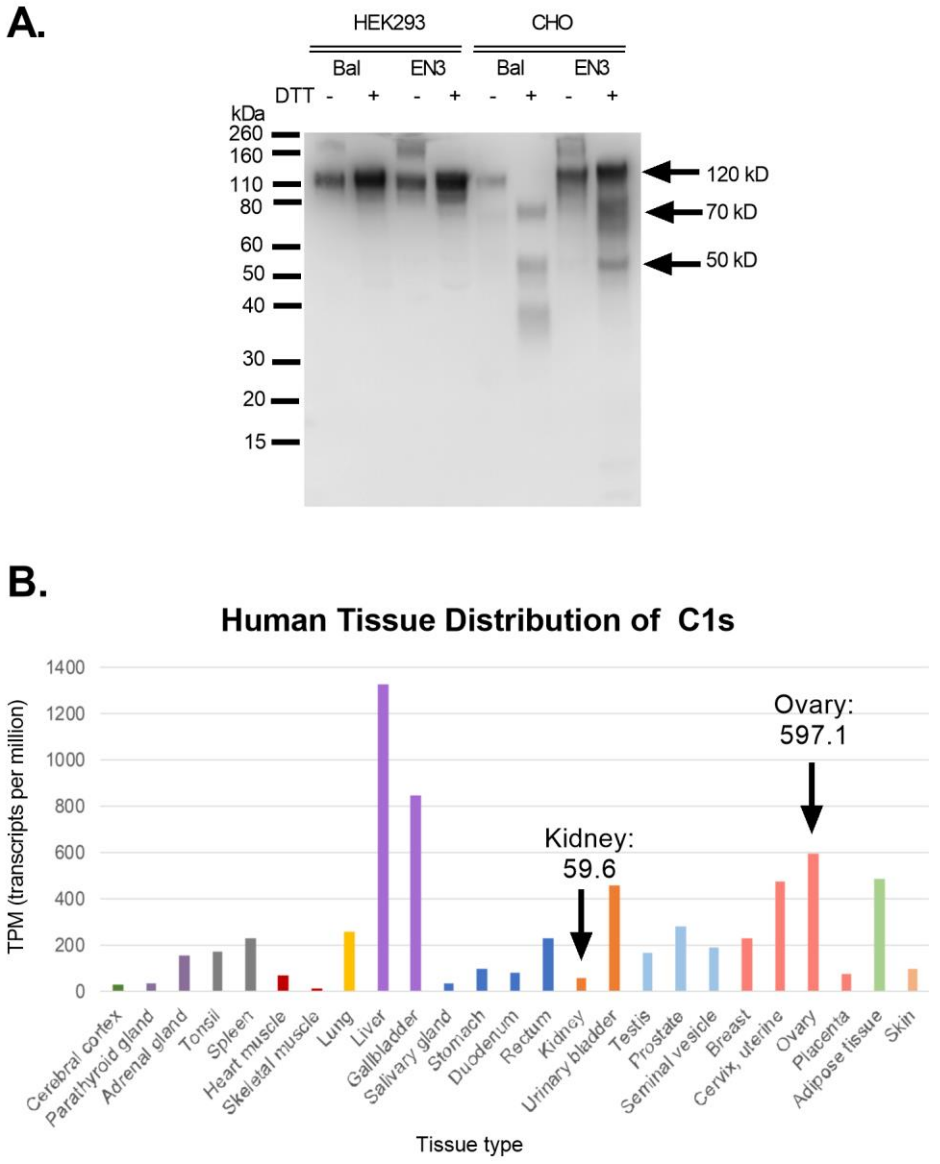


Figure 3. Differential expression of C1s in different cell types.

(A) BaL-rgp120 and clade B, EN3-rgp120 were transiently expressed in HEK293 and CHO cells. Supernatants was analyzed by immunoblot using an α-gp120 polyclonal antibody. Proteolytic degradation was indicated by the appearance of bands at 50 kDa and 70 kDa. (B) RNA seq expression data for C1s in human expression tissues was plotted based on data from the Human Protein Atlas. RNA expression values in transcripts per million for kidney and ovary tissues are labeled.

Development of a stable CHO cell line expressing the BaL-rgp120 for use in screening for C1s gene knockout experiments.

To show that C1s was responsible for proteolysis, we created a stable CHO cell line expressing a proteolysis-sensitive gp120 to be used for protease gene knockout experiments. The gp120 gene from the BaL isolate of HIV (Hwang et al., 1991) was transfected into a recently described CHO cell line, MGAT1^{-} CHO, containing a mutation in the MGAT1 glycosyltransferase gene that limits glycosylation to early high-mannose intermediates in the N-linked glycosylation pathway (Byrne et al., 2018a; O'Rourke et al., 2018a). BaL-rgp120 was selected based on its binding profile to bN-mAbs. In particular, BaL-rgp120, when expressed in MGAT1- CHO cells, binds the prototypic PG9, PGT128, 10-10-74 and VRC01 bN-mAbs that target epitopes in the V1/V2, V3 stem, and CD4 binding site of gp120. Antigenicity of BaL-rgp120 was improved over MN-rgp120, the original clade B vaccine used in the RV144 trial. MN-rgp120 binds strongly to VRC01, minimally to PG9, and does not bind to PGT121, PGT128 or 1010-74. The bN-mAb binding data for MN-rgp120 is shown by Doran et al., and Fig. 7, while the antigenicity of BaL-rgp120 is shown in Figs. 6 and 7 (Doran et al., 2018a). Virion associated Env proteins primarily possess high-mannose glycans (Bonomelli et al., 2011; Doores et al., 2010). Restricting glycosyation improved binding of bN-mAbs dependent on the presence of high-mannose glycans (Sok et al., 2014; Doran et al., 2018a; Byrne et al., 2018a). Using the ClonePix2 colony selection robot, 30,000 rgp120-secreting colonies were imaged and ranked based

on fluorescence intensity and geometric parameters (Fig. 4A). Seven colonies were grown in shake-flask cultures. After eight days, all seven cultures exhibited significant degradation of BaL-rgp120 (Fig. 4B) with productivities of 200-300mg/L for four out of seven cultures (2.4D, 1.5D, 2.3D, 1.3D) (Fig. 4C). In these experiments, clone 1.5d MGAT1^{-} CHO cell line stably expressing BaL-rgp120 was shown to be sensitive to proteolysis and produced sufficient gp120 for use in gene knockout experiments to assess the role of C1s in the proteolysis of clade B Env proteins.

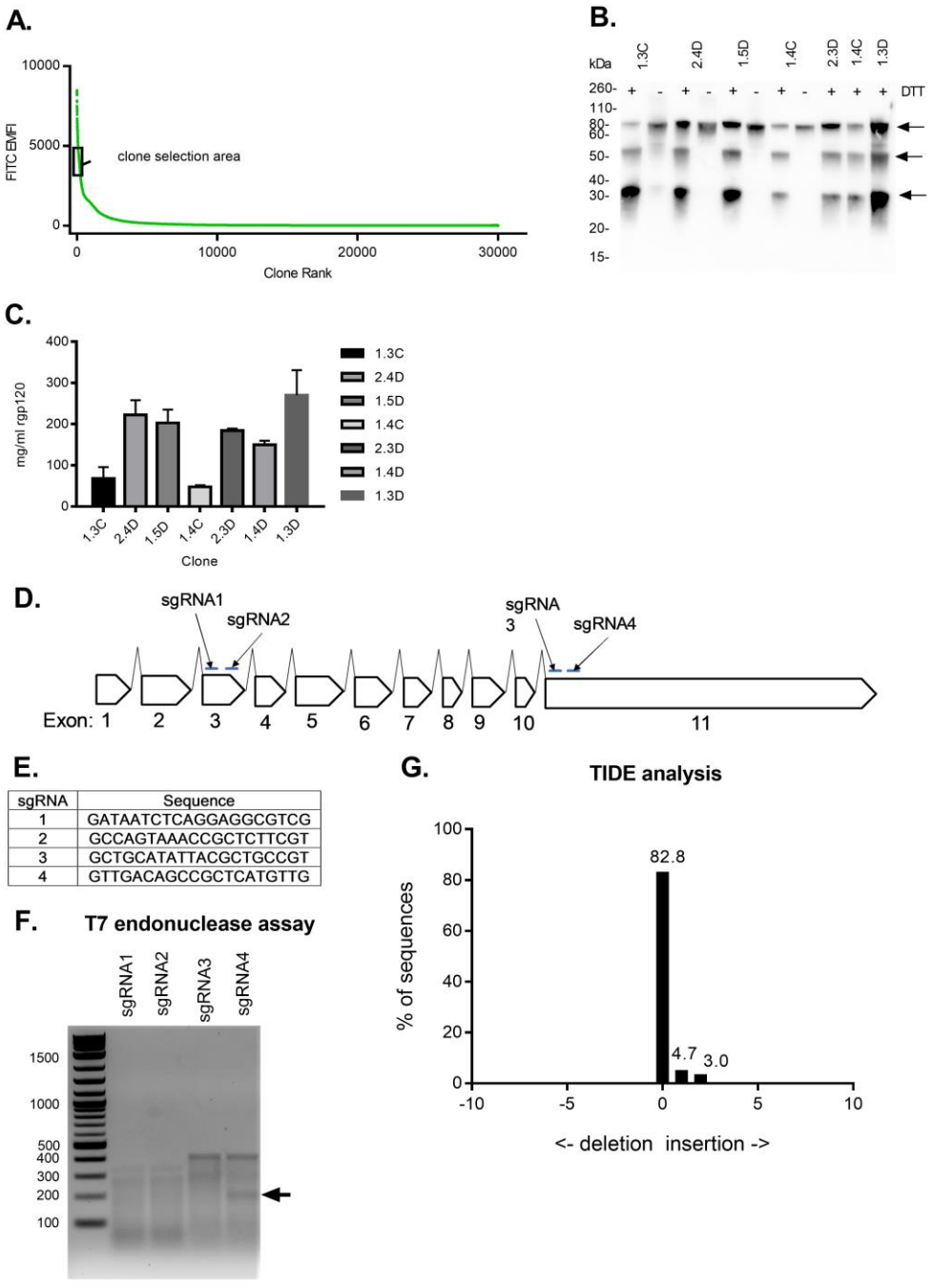

A.
10000
5000
0
FITC EMFI
clone selection area
0
10000
20000
30000
Clone Rank
B.
kDa
1.3C
2.4D
1.5D
1.4C
2.3D
1.4C
1.3D
260-
110-
80-
60-
50-
40-
30-
20-
15-
DTT
C.
400
300
200
100
0
mg/ml rgp120
1.3C
2.4D
1.5D
1.4C
2.3D
1.4D
1.3D
Clone
D.
sgRNA1
sgRNA2
sgRNA 3
sgRNA4
Exon: 1 2 3 4 5 6 7 8 9 10 11
E.
sgRNA | Sequence
1 | GATAATCTCAGGAGGCGTCG
2 | GCCAGTAAACCGCTCTTCGT
3 | GCTGCATATTACGCTGCCGT
4 | GTTGACAGCCGCTCATGTTG
F.
T7 endonuclease assay
sgRNA1
sgRNA2
sgRNA3
sgRNA4
1500
1000
500
400
300
200
100
G.
TIDE analysis
100
80
60
40
20
0
% of sequences
82.8
4.7
3.0
-10
-5
0
5
10
<- deletion insertion ->

Figure 4. Production of a stable cell line expressing BaL-rgp120 and screening colonies for CRISPR/Cas9 knockout of C1s.

(A) 30,000 colonies of MGAT1$^-$ CHO cells were transfected with a plasmid expressing both BaL-rgp120 and G418 (geneticin) resistance. These were analyzed for rgp120 expression using the Clonepix2 robot. Colonies were ranked by exterior mean fluorescence intensity (EMFI), which was plotted as a function of clone rank. Approximately 0.01% of colonies, or 42 colonies, were selected and expanded. (B) Following selection and expansion, supernatants from seven cell lines were assayed by Western blot for the presence of Bal-rgp120 protease cleavage products (indicated by arrows). (C) Quantitative yield of rgp120 from the eight-day batch-fed cultures as determined by ELISA. (D) Diagram of the C1s gene showing intron and exon structure with the exon length to scale and the location of single guide RNAs (sgRNA) as indicated with arrows. (E) Sequences of sgRNAs that were tested for the ability to knockout the C1s gene. (F) Selection of single guide RNA for CRISPR/Cas9 knockout of C1s. Knockout efficiency was qualitatively measured in a population of cells post-transfection by the T7 endonuclease assay for sgRNA1, sgRNA2, sgRNA3 and sgRN4. A 400 bp PCR product was generated around the Cas9 target site using PCR primers from upstream and downstream of the sgRNA sites. The arrow indicates the 200 bp PCR product generated by the T7 endonuclease at the secondary structures formed by indel-containing DNA. (G) The quantitative TIDE analysis for sgRNA4 in a population of cells post-transfection indicated a Cas9-knockout efficiency of 11.4%.

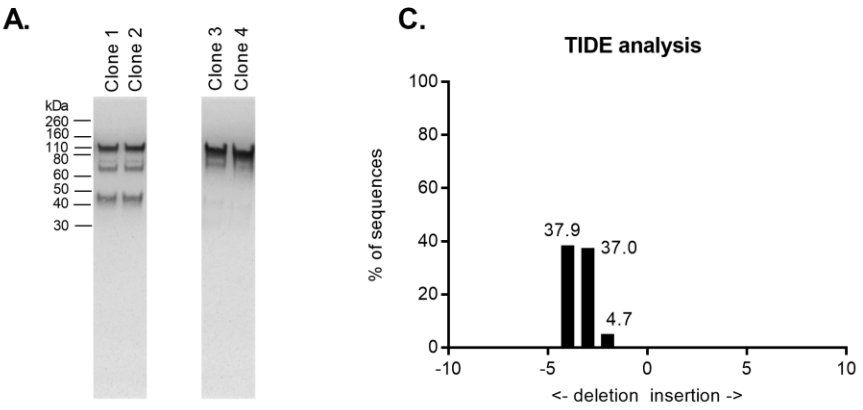
CRISPR/cas9 knockout of C1s in a stable MGAT1$^-$ CHO cell line expressing BaL-rgp120 eliminates proteolysis.

Based on the identification of C1s, CRISPR/Cas9 was used to knock out the C1s gene in the BaL-rgp120 1.5d MGAT1$^-$ CHO cell line. Four guide RNA targets (Fig. 4D, 4E) were selected using CRISPy target selection software (Ronda et al., 2014) and targeted exon 3 and exon 11, the serine protease domain of C1s (Fig. 4D). A screen of the four gRNAs was done to determine gene-editing efficiency as not all gRNAs are equally effective (Peng et al., 2016). Knockout efficiency in the population of cells post-transfection was determined by the T7 endonuclease assay

and TIDE (tracking of indels by decomposition) analysis (Figure 4F, 4G) (Vouillot et al., 2015; Brinkman et al., 2014). Only sgRNA4 exhibited any knockout efficiency as qualitatively shown by the T7 assay (Fig. 4F) and quantitatively by TIDE (Fig. 4G). The indel introduced by sgRNA4 occurs at nucleotide 1627 of the NCBI RefSeq XM_007646821.2. The 11.4% knockout efficiency measured by TIDE analysis suggested that forty clones should be screened to identify clones with the genetic knockout.

Single cell cloning was carried out to isolate clones with Cas9-mediated mutations of the C1s gene. Supernatants from thirty-eight clones transfected with sgRNA4 were recovered between days 18 to 25 and analyzed by immunoblot for the detection of intact and proteolyzed BaL-rgp120 (Fig. 5A). Of thirty-eight clones analyzed, three clones lacked protease activity as indicated by the accumulation of intact gp120 and no visible 50 kDa or 70 kDa fragments. Blue-white lacZ screening was used to individually sequence each of the two C1s alleles present in each clone (Fig. 5C). Sequencing confirmed that clones 1 and 2 that displayed proteolyzed Bal-rgp120 were not gene-edited, while clones 3 and 4 that expressed non-proteolyzed BaL did have Cas9-modified gene edits. Sequencing and TIDE analysis on clone 4 showed two gene-edits: a -3 basepair and -4 basepair deletion (Fig. 5B, 5C). C1s is part of the complement pathway and is not directly invovled in metabolic pathways. We did not observe significant differences in cell doubling times (~25 hours) or protein expression between the normal CHO cell line and the $C1s^{-/-}$ $MGAT1^{-}$ CHO cell line (100-300 mg/L). The clone 4 BaL-rgp120 $C1s^{-/-}$

MGAT1^{-} CHO cell line identified above will be henceforth referred to as BaL-rgp120 C1s$^{-/-}$ MGAT1^{-} CHO cell line and was used for the subsequent antibody-binding experiments.



Clone 1	5'-GTTGACAGCCGCTCATG TTGTGG-3'	WT	
Clone 2	5'-GTTGACAGCCGCTCATG TTGTGG-3'	WT	
Clone 3: edit 1	5'-GTTGACAGCCGCTCATG ------TGG-3'	Δ-3bp	Deletion of TTG
Clone 3: edit 2	5'-GTTGACAGCCGCTCATGTTTGTGG-3'	Δ+1bp	Insertion of T
Clone 4: edit 1	5'-GTTGACAGCCGCTCATG -------GG-3'	Δ-4bp	Deletion of TTGT
Clone 4: edit 2	5'-GTTGACAGCCGCT--------TTTGTGG-3' OR	Δ-3bp	Deletion of CATG, Insertion of T

Figure 5. Screening clones for a stable C1s$^{-/-}$ MGAT^{-} CHO cell line expressing BaL-rgp120.

(A) Out of thirty-eight clones, three were identified with Cas9-mediated genetic knockouts of the C1s gene. For simplicity, data from only two are shown. Immunoblot analysis of BaL gp120 supernatants of clones 1 through 4 is shown. Clones 1 and 2 are representative of CHO cell clones where the C1s gene was unmodified, while clones 3 and 4 are examples of clones where the C1s gene was inactivated. (B) Sanger sequencing was used to characterize the genetic edits created by Cas9-editing in cells from clone 3 and clone 4. (C) TIDE was used to confirm the presence of two mutations in the C1s gene of clone 4.

Proteolyzed Bal-rgp120 displays diminished binding to V3-directed mAbs and bN-mAbs relative to intact BaL-rgp120.

To compare differences in neutralizing antibody binding between intact and proteolyzed gp120, BaL-rgp120 was purified from the parental 1.5D BaL-rgp120 MGAT- CHO cell line and the BaL-rgp120 $C1s^{-/-}$ $MGAT1^{-}$ CHO cell lines (Fig. 6J). However, BaL-rgp120 produced in the stable $MGAT1^{-}$ 1.5d CHO cell line contained a mixture of proteolyzed and non-proteolyzed proteins. To fully distinguish differences in antibody binding between intact and proteolyzed BaL-rgp120, cleaved BaL-gp120 was produced by additional treatment with thrombin. Treatment of gp120 with exogenous thrombin has been shown to cleave rgp120 into fragments of 50 and 70 kDa at the same GPGR sequence as the C1s protease (Clements et al., 1991; Schulz et al., 1993). The intact BaL-rgp120 expressed in the BaL-rgp120 $C1s^{-/-}$ $MGAT1^{-}$ CHO cell line and the thrombin-cleaved BaL-rgp120 expressed in the $MGAT1^{-}$ 1.5d CHO cell line were used in the antibody binding assays.

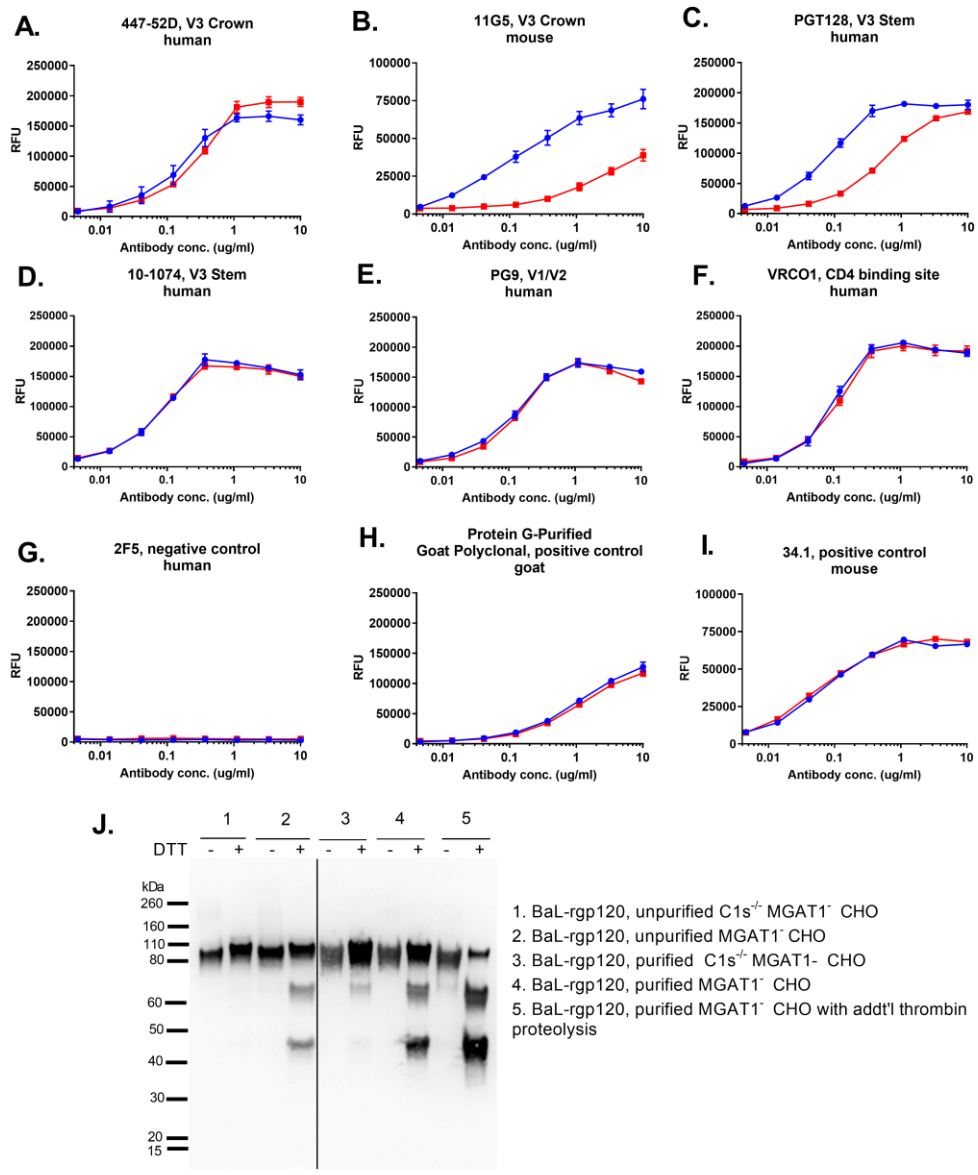


Figure 6. bN-mAb and V3-antibody binding to intact and proteolyzed BaL-rgp120.

A direct coat FIA was performed to measure bNAb binding to BaL-rgp120 produced in C1s$^{-/-}$ MGAT1^{-} CHO cells and normal MGAT1^{-} CHO cells. Gp120s from MGAT1- CHO cells were further cleaved by treatment with thrombin as described in the Results section. The binding of mAbs to proteolyzed BaL-rgp120 (red) and non-proteolyzed BaL-rgp120 is indicated (blue). (A-F) Included were antibodies to the V3 domain (447-52D, PGT128, 10-1074, and 11G5), the V1/V2 domain (PG9) and the CD4 binding site (VRCO1). (G) The human bN-mAb, 2F5,

directed to gp41 was included as a negative control. (H, I) Polyclonal goat antibodies to gp120 and the 34.1 mouse monoclonal antibody to the N-terminal gD purification tag served as positive controls. (J) Immunoblot analysis of the proteins used in these antibody binding assays are shown. The figure shows both protein from supernatant as well as purified protein.

Fluorescence immunoassay (FIA) was used to quantitate differences in antibody binding and determine structural changes due to proteolysis between intact BaL-rgp120 and proteolyzed BaL-rgp120 cleaved in the V3 domain. V3-antibodies had reduced binding due to the loss of the V3 epitope, while binding of bN-mAbs to the V1/V2 domain and CD4 binding site were not affected (Fig. 6A-I). PGT128 displayed the most dramatic loss of binding, followed by the 11G5 mAb. PGT128 recognizes a glycan dependent epitope at the base of the V3 domain, while 11G5 recognizes an epitope that spans the cleavage site in the V3 domain (Pejchal et al., 2011; Nakamura et al., 1992). We did not observe significant differences in binding to the human 447-52D mAb, which recognizes the V3 GPxR crown, or for 10-1074, which recognizes the base of the V3 stem (Zolla-Pazner et al., 2004; Mouquet et al., 2012). Epitopes to the V1/V2 domain and the CD4 binding site were unaffected as shown by the PG9 and CD4 bN-mAbs (McLellan et al., 2011; Zhou et al., 2010). These antibody binding experiments demonstrate that V3-directed antibodies are sensitive to structural changes due to proteolysis at the V3 domain.

Comparative bN-mAb binding of unclipped MN-rgp120 used in the RV144 trial to glycan restricted clade B Bal-rgp120 and clade C TZ97008-rgp120.

Fluorescence immunoassay was used to compare the antigenic structure of unclipped clade B MN-rgp120, similar to that used in the RV144 trial, with unclipped BaL-rgp120 and clade C TZ97008-rgp120 (Doran et al., 2018b). Because both the BaL-rgp120 and the TZ97008-rgp120 were expressed in the MGAT1^{-} CHO cell line, N-linked glycosylation was restricted high-mannose glycans. In contrast, MN-rgp120 was expressed in HEK293 cells to prevent proteolysis and possessed a mixture of hybrid and complex sialic acid containing glycans (Yu et al., 2012). BaL-rgp120 exhibited robust binding to the PG9, PGT128, 10-1074 and VRCO1 (Fig. 7A, 7C, 7D and 7E). In contrast, the clade B MN-rgp120 immunogen exhibited robust binding to VRC01, bound only moderately to PG9, but did not bind to PGT128 or 10-1074. The clade C TZ97008-rgp120 expressed in the MGAT1^{-} CHO cell line also bound to PG9, PGT128, 10-1074, VRCO1, and additionally bound to PGT121 (Fig. 7B). While PGT121 has been found to be a strongly neutralizing antibody (Walker et al., 2011), identifying an antigen that binds to PGT121 has been difficult. Although several bN-mAbs require mannose-5 for binding, PGT121 requires a mixture of high mannose and sialic acid glycans for binding (Mouquet et al., 2012). However, 10-1074 is in the same family as PGT121 and binds to Bal-rgp120 at an overlapping epitope. Further screening of clade B antigens, including additional Tier 2 viruses, may result in identification of an antigen that is additionally recognized by PGT121. BaL-rgp120

in combination with TZ97008-rgp120 expressed in the $C1s^{-/-}$ $MGAT1^{-}$ and $MGAT1^{-}$ CHO cell lines, represent candidates for a bivalent vaccine effective against clade B and clade C viruses.

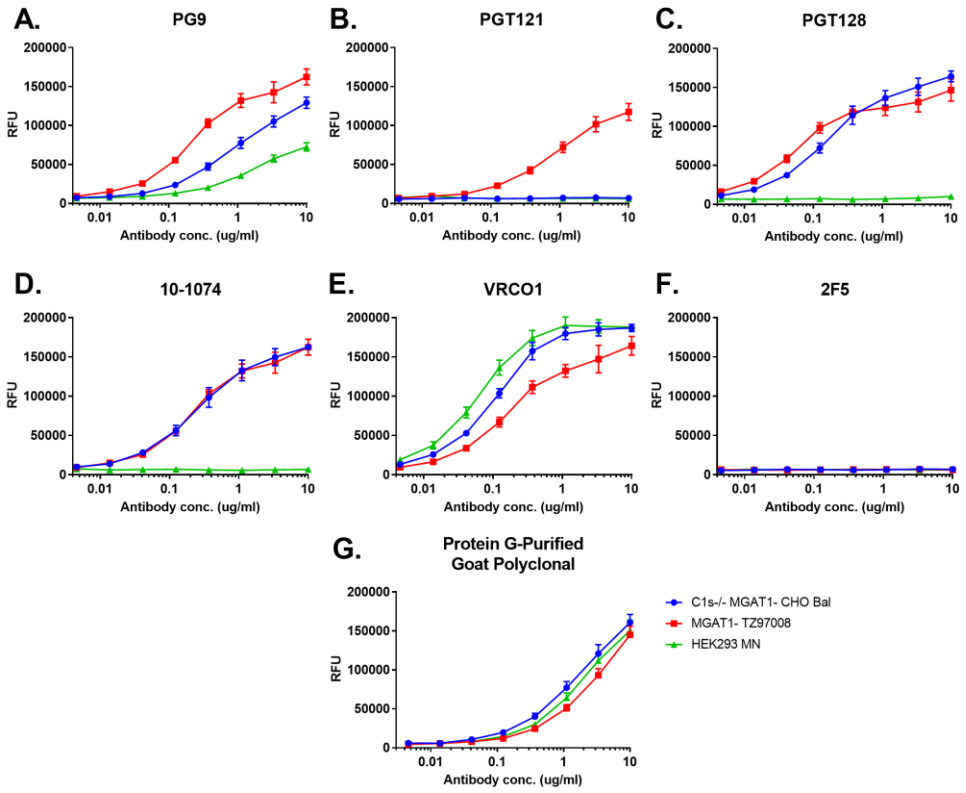


Figure 7. Comparison of bNAb binding to MN-rgp120 used in the RV144 trial with BaL-rgp120 and TZ97008-rgp120 produced in CRISPR-engineered cell lines.

(A-F) Unclipped gp120s possessing normal N-linked glycosylation (MN-rgp120) or N-linked glycosylation restricted to mannose-5 glycans (BaL-rgp120 and TZ97008-rgp120) were compared for the binding of several prototypic bN-mAbs by FIA. Unclipped MN-rgp120 (green) was prepared by expression in HEK293 cells. TZ97008-rgp120 (red) was expressed in MGAT1- CHO cells. BaL-rgp120 (blue) was expressed in $C1s^{-/-}$ $MGAT1^{-}$ CHO cells. (F) The bN-mAb 2F5 was used as a negative control. BaL-rgp120 (blue) was expressed in $C1s^{-/-}$ $MGAT1^{-}$ CHO cells. (G). Polyclonal antibodies to gp120 were used as a positive control.

Discussion

Proteolysis associated with CHO cells, the preferred cell line for biopharmaceutical production, has hindered the development of therapeutic biologics including HIV vaccines. In these studies, we described the identification of C1s as the protease responsible for the cleavage of HIV vaccine immunogens. This conclusion is supported by the finding that knockout of C1s in CHO cells prevents proteolysis of BaL-gp120, a clade B isolate of HIV. Two CRISPR/Cas9 induced mutations (C1s and MGAT1) were combined in a single CHO cell line to improve the quality of a recombinant protein intended for large scale production. This cell line overcomes two major problems in HIV vaccine production: 1) proteolysis of clade B HIV vaccine immunogens expressed in CHO cells and 2) glycan heterogeneity that has impaired the antigenic structure of recombinant Env proteins. The strategy we have developed provides a practical method for the large-scale production of unclipped, glycan restricted, Env proteins to enable the testing of clade B HIV vaccine concepts including monomeric and trimeric envelope proteins, as well as guided immunization strategies. Further studies may also elucidate whether a C1s-deficient CHO cell line or other CHO protease-deficient CHO cell line is beneficial for the expression of other recombinant proteins found to be proteolyzed in CHO cells including but not limited to monoclonal antibodies, human Factor VIII or IFN-γ (Bee et al., 2015; Dorai et al., 2011; Clincke et al., 2011; Kaufman et al., 1988).

The identification of C1s as the protease found in CHO cell culture medium provides an explanation for the proteolysis that has limited the development of recombinant therapeutic proteins. Previously, groups have partially characterized CHO proteases involved in proteolytic degradation and have stated that the CHO protease of interest is a metalloprotease or serine protease based on screening assays with protease inhibitors and limited amino acid sequence from mass spectroscopy experiments. At the time, they were not able to identify the protease due to the lack of an annotated CHO genome (Du et al., 2008; Sandberg et al., 2006; Clincke et al., 2011; Dorai et al., 2011). With the publication of the CHO genome in 2011 (Xu et al., 2011) and common MS/MS techniques, we have determined the identity of the protease and created an engineered knockout cell line to eliminate proteolysis of recombinant proteins. Similar to our work, other groups in recent years have also utilized the CHO genome to identify CHO proteases involved in the degradation of biologics (Cathepsin D, Cathepsin L, matriptase, etc…) or to do large-scale proteomics studies (Bee et al., 2015; Laux et al., 2018; Luo et al., 2019; Park et al., 2017; Valente et al., 2015). Identifying the myriad of proteases or host cell proteins in CHO cells is necessary due to the continuing use of the CHO cell line for expression of therapeutic proteins. The workflow described here (ie enrichment by column chromatography followed by tandem mass spectroscopy) can be used to identify other CHO proteases involved in the degradation of recombinant proteins. Knowledge of these proteases and host cell proteins is also important for designing purification processes that will remove impurities from the

final purified product, as many host cell proteins will co-elute with the product and can be difficult to remove (Valente et al., 2015). We demonstrate separation of C1s by anion exchange chromatography during the enrichment of the protease; adding an anion exchange chromatography step could be used as a method to remove C1s in a protein purification process. Additional studies will determine whether recombinant proteins other than HIV Env such as Factor VIII or IFN-γ are also susceptible to proteolysis by C1s in CHO cells.

Conservation of the GPGR motif recognized by C1s in clade B viruses suggests that this cleavage site provides an advantage in fitness. In previous studies, we observed that cleavage sites recognized by the antigen-processing enzymes, cathepsins S, L and D, were conserved in Envs and occurred at epitopes recognized by bNAbs (Yu et al., 2010). It was suggested that these protease cleavage sites represent an immune escape mechanism whereby key epitopes are degraded during the antigen presentation process, diminishing an effective immune response to these sites. Similarly, deletion of an overlapping cathepsin S cleavage site in MN-rgp120 facilitated cross presentation and increased binding affinity for MHC class 1 molecules (Frey et al., 2018). Preservation of the C1s cleavage site may also facilitate cellular immune responses as viruses have been known to use host cell proteases and complement to promote infectivity and immune escape strategies (Ploegh, 1998; Zipfel et al., 2007). While the GPGR motif is present only in clade B viruses, the protease cleavage sites of cathepsins S, L and D are present in Envs of other clades. Proteolysis of clade B Envs at the GPGR supports the broader

hypothesis that protease cleavage sites located proximal to epitopes recognized by bNAbs on viral proteins may promote immune escape or enhance the viruses' evolutionary fitness.

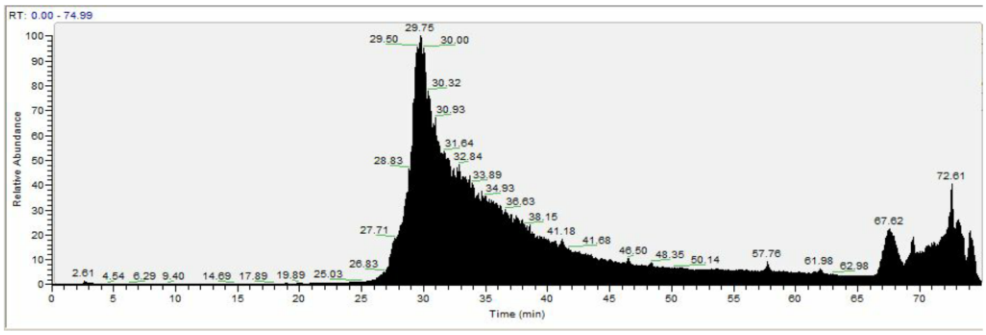
Along with the identification of the C1s protease, we have also shown highly antigenic Env proteins suitable for an HIV vaccine. We recently described glycan-optimized gp120s from the A244 strain of HIV (clade CRF01-AE) and the TZ97008 strain of HIV (clade C) (O'Rourke et al., 2018a; Doran et al., 2018b). Here, we describe a novel $C1s^{-/-}$ $MGAT1^{-}$ CHO cell line expressing unclipped, glycan optimized gp120 from the BaL isolate of HIV. The two gp120s described previously along with the BaL-rgp120 described in this paper represent three components of a trivalent vaccine effective against the three major subtypes (clades B, C, and CRF01_AE) circulating in the developed and developing worlds. Further immunogenicity studies will be required to define the full immunogenic potential of these new candidate vaccines. Preliminary studies on the immunogenicity of mannose-5 glycosylated antigens from the $MGAT1^{-}$ CHO cell line show comparable antibody titers in animal sera (unpublished) and we are currently analyzing the antibody repertoire in sera elicited by these antigens.

In summary, the data presented in this report represents a continuation of efforts to improve the quality and manufacturability of recombinant HIV immunogens. This technology can be used to produce other candidate vaccines that have not been tested because they could not be manufactured cost effectively or expeditiously. The C1s-deficient CHO cell line represents a permanent solution to

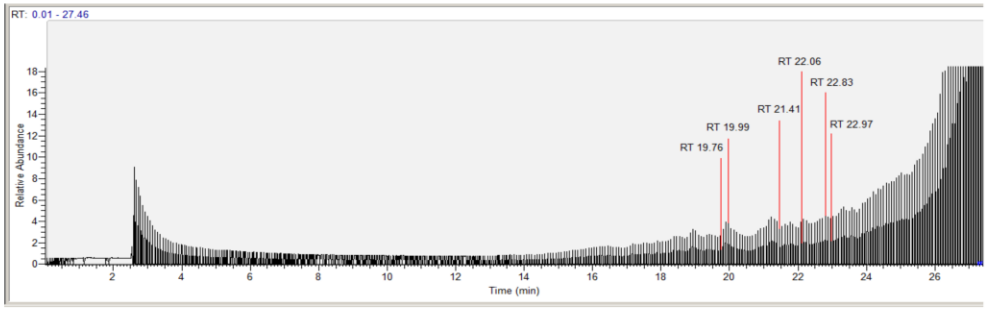
mitigate proteolysis of recombinant HIV immunogens and possibly other recombinant therapeutic proteins susceptible to serine protease activity.

Supplemental Figures

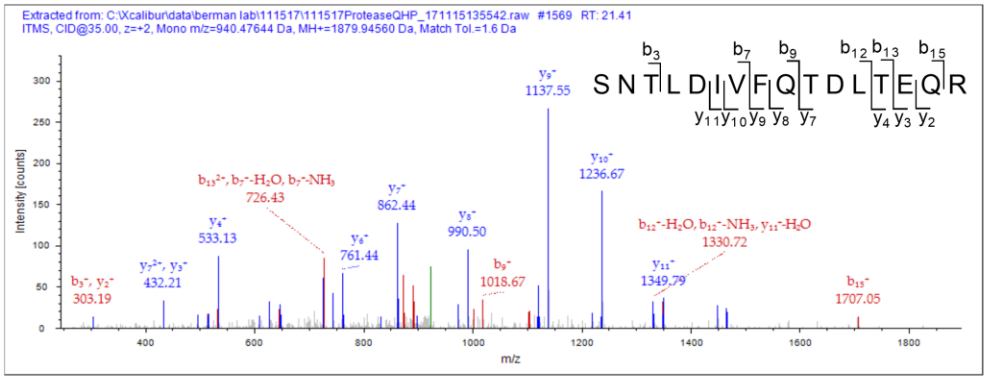
A.



B.



C.



Supplemental Figure 1. Spectra from tandem MS experiments for identification of C1s.

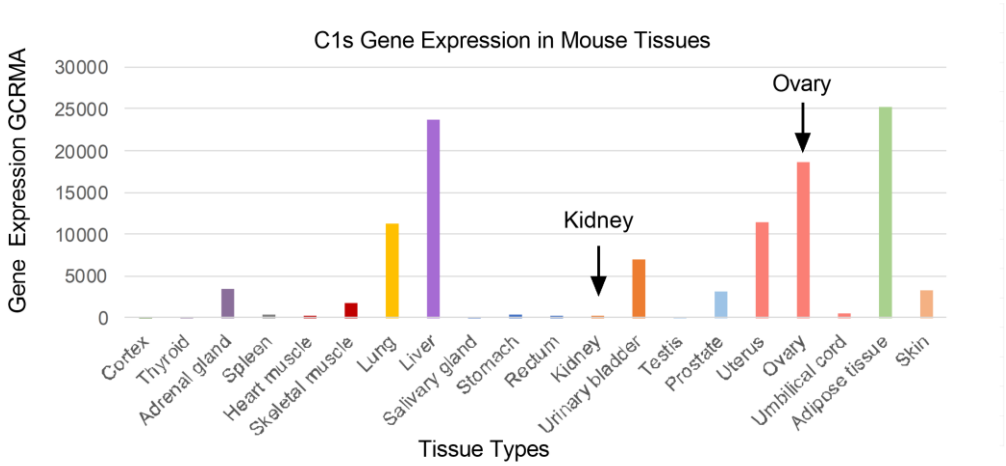
(A) Full-scan spectra of the pooled fractions 1 and 2 from the anion exchange eluate shown in Figure 2D that was enriched for gp120-proteolytic activity (B) Magnified region of the full-scale spectra showing unique peptides of C1s (C)

Fragment ion spectra for peptide SNTLDIVFQTDLTEQR from C1s, one of the unique peptides identified in Figure 3B.

Confidence level	Sequence	XCorr	RT [min]
High	SNTLDIVFQTDLTEQR	5.27	21.41
High	TNFDNDIALVQLK	4.45	19.76
High	NDLGLISGWGR	4.35	19.99
High	FYVAGLVSWGK	4.16	22.06
High	VDDPENTVFGSVIQY	4.02	22.83
High	TWDIEVPEGFGIR	3.92	22.97

Supplemental Figure 2. Unique C1s peptides identified by tandem MS.

The identified peptides all had high XCorr, a measure of how well the spectra matches the predicted peptide sequence. The retention times of the peptide are also listed.



Supplemental Figure 3. C1s Gene Expression in Mouse Tissues.

Kidney tissues had a unitless gene expression value of 221.35 and ovary tissues had a unitless gene expression value of 18748.35. The dataset used is called GeneAtlas GNF1M, gcrma. This gene expression dataset was acquired from BIOGPS. Expression values are derived from fluorescence intensity in the Affymetrix microarray assay used.

Chapter 2: Gene editing in CHO cells to prevent proteolysis and enhance glycosylation: Production of HIV envelope proteins as vaccine immunogens

This chapter contains text and figures from the following manuscript:

Li SW, Wright M, Healey JF, Hutchinson JM, O'Rourke S, Mesa KA, et al. Gene editing in CHO cells to prevent proteolysis and enhance glycosylation: Production of HIV envelope proteins as vaccine immunogens. PLOS ONE. 2020 May 29;15(5):e0233866.

Introduction

The majority of recombinant glycoprotein therapeutics are manufactured in CHO (Chinese Hamster Ovary) cells due to their high productivity (1-10 grams per liter), genetic stability, and ability to be grown in large-scale suspension culture (Huang et al., 2010; Worton et al., 1977; Sinacore et al., 2000). However, many recombinant proteins including monoclonal antibodies, antibody fusion proteins, and IFN-γ are partially degraded or "clipped" by endogenous CHO cell proteases during the cell culture or recovery process (Robert et al., 2009; Clincke et al., 2011; Dorai et al., 2011; Lim et al., 2018; Laux et al., 2018; Luo et al., 2019). This is also the case for glycoprotein 120 (gp120), the monomeric subunit of the HIV-1 envelope protein (Env), used in many of the HIV vaccines tested to date in human vaccine efficacy trials (Flynn et al., 2005; Pitisuttithum et al., 2006; Rerks-Ngarm

et al., 2009; Robinson, 2018). The HIV Env protein mediates virion binding to CD4, the T-cell surface receptor, and to the CXCR4 or CCR5 chemokine receptors (Laurent-Crawford et al., 1993; Bandres et al., 1998; Weissman et al., 1997). Env proteins have been included in most HIV vaccines since they are the major target for virus neutralizing antibodies (Kwong et al., 2013; Mascola and Haynes, 2013; Burton and Hangartner, 2016). HIV isolates are classified into different genetic clades based on impartial sequence analysis (Korber et al., 2000; Robertson, 2000). These include clades C and CFRF01_AE viruses, prevalent in Africa and Asia respectively, and clade B viruses in North America, Europe, the Caribbean and Australia. Because they lack the clade B consensus sequence Gly-Pro-Gly-Arg-Ala-Phe (GPGR/AF) at the crown of the V3 domain, most clade C and CRF01_AE Envs can be produced in CHO cells without proteolysis. In contrast, the V3 domain of most clade B Envs has been shown to be highly sensitive to proteolysis by exogenous thrombin or an unidentified CHO cell protease (Clements et al., 1991; Schulz et al., 1993; Du et al., 2008; Pugach et al., 2015). Env proteins proteolyzed in this manner are difficult to manufacture and purify in quantities required for immunization of populations at high risk for infection (Du et al., 2008; Wen et al., 2018).

Recently, we reported that the major CHO cell protease responsible for cleavage of clade B gp120s was the complement component 1 protease, C1s (Li et al., 2019b). C1s is a serine protease that recognizes the sequence Gly-Pro-Gly-Arg, located in the V3 loop of gp120. This sequence is present in 71% of clade B HIV

strains (Foley and Korber, 1996) and is also present in the Env protein from the clade A/G Z321 isolate, one of the earliest known strains of HIV (Srinivasan et al., 1989). The V3 loop mediates binding to the coreceptors, CXCR4 or CCR5 (Wu et al., 1996). Thus, antibodies to this portion of the V3 region are highly effective at virus neutralization. These antibodies include the monoclonal antibody (mAb), 447-52D, that binds to the crown of the V3 region (Zolla-Pazner et al., 2004), and the glycan-dependent, broadly neutralizing monoclonal antibodies (bN-mAbs), PGT121, PGT128, and 10-1074, that bind to the stem of the V3 region (Walker et al., 2011; Julien et al., 2013; Pejchal et al., 2011; Mouquet et al., 2012). Because the GPGR/AF sequence is part of, or adjacent to, the epitopes recognized by these neutralizing antibodies, it is necessary to keep the V3 loop intact on the HIV Env immunogen in the hopes of eliciting similar, neutralizing antibodies. In our previous study, we showed that CRISPR/Cas9 inactivation of the C1s gene in a stable CHO-S cell line expressing gp120 from the laboratory-adapted isolate, HIV_{BaL}, prevented proteolysis at the GPGPR/AF sequence in the V3 domain. As this cell line is uniquely developed to express BaL-rgp120, it cannot be used for the expression of other recombinant proteins. Therefore, a C1s knockout CHO cell line, that can be used as the cell substrate for the expression of any other recombinant protein, is needed.

In this paper, we describe the development of two novel, suspension-adapted CHOK1 cell lines useful for the production of biopharmaceutical proteins. The first cell line, $C1s^{-/-}$ CHOK1 2.E7 is a CHOK1 cell line where the C1s gene

has been inactivated by gene editing. The second cell line, C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1 is also a CHO K1 cell line where both the C1s and MGAT1 genes have been inactivated. Inactivation of the MGAT1 gene limits N-linked glycosylation to mannose-5 and earlier intermediates in the N-linked glycosylation pathway. Limiting N-linked glycosylation via the MGAT1 cell line has been shown to improve the antigenic structure of gp120 as measured by the ability to bind multiple bN-mAbs (Morales et al., 2014; Byrne et al., 2018b; Doran et al., 2018b).

To demonstrate the utility of these cell lines, we expressed three novel clinical isolates of clade B gp120s that have potential as improved HIV vaccine immunogens. These gp120s were derived from rare individuals, who either possessed broadly neutralizing antibodies (elite neutralizers) and/or were able to control their virus replication for extended periods of time without anti-retroviral drug treatment (viral controllers and elite controllers) This will allow us to further test the hypothesis that Env proteins from individuals with the elite neutralizer and/or rare controller phenotype possess unique structural features that are particularly effective in stimulating effective, anti-viral immune responses. Additionally, we studied the expression of another recombinant protein, human Factor VIII, known to be susceptible to proteolysis in CHO cells to explore the substrate specificity of C1s and how it differs from thrombin, another enzyme able to cleave the GPGRAF consensus sequence. This work documents the C1s$^{-/-}$ CHOK1 cell line and C1s$^{-/-}$ MGAT1^{-} CHOK1 cell line as novel cell substrates for the production of biopharmaceutical recombinant proteins.

Materials and Methods

Antibodies and purified proteins

The coding sequences of the PG9, PGT128, VRC01 bN-mAbs were obtained from published sequences (available from the NIH AIDS Reagent Program, Germantown, MD) and used to created synthetic genes for stable or transient expression of antibodies in CHO cells. Antibodies were purified using Protein G chromatography. These cell lines have been deposited in the NIH AIDS Reagent Program. The 10-1074 bN-mAb was acquired from the NIH AIDS Reagent Program. The 34.1 mAb to the Herpes Simplex Virus glycoprotein D (gD) was produced in-house (O'Rourke et al., 2019). The human 447-52D mAb was contributed by Dr. Susan Zolla-Pazner (Mount Sinai, New York City, NY). Recombinant rgp120s were cloned with an N-terminal gD tag into a pCF vector (O'Rourke et al., 2019). Plasmids were transformed in 5-alpha competent E. coli cells (New England Biolabs, Ipswich, MA), purified using Plasmid Maxi Kit (Qiagen, Hilden, Germany) and electroporated into CHO cells as described below. Recombinant human factor VIII consisted of Advate (Baxter Healthcare Corporation, Deerfield, Illinois) (Shapiro, 2007), and was acquired by the laboratory of Dr. Peter Lollar (Emory University, Atlanta, GA).

Clade B clinical isolates

Genes encoding gp120s were cloned from cryopreserved plasma and/or peripheral blood mononuclear cells from anti-retroviral therapy naïve individuals.

Clade B Envs (EN2, EN3, EN6, and EN7) were obtained from individuals prescreened for the presence of bN-mAbs and the ability to control virus loads and CD4 counts without antiretroviral therapy. The clade B isolates were obtained from the SCOPE Cohort (Observational Study of the Consequences of the Protease Inhibitor Era cohort) and WIHS cohort (Women's Interagency Health Studies cohort). The Envs were screened for the presence of bN-mAbs using the Simek panel of HIV pseudoviruses (Simek et al., 2009) and individuals with high levels of bN-mAbs were selected for further study. In these studies, Envs were amplified by PCR and screened for infectivity by Monogram Biosciences (South San Francisco, CA) using the Phenosense assay system.

Culture of CHOK1 cells

CHO K1 cells (ATCC® CCL-61) were obtained from the American Type Culture Collection (ATCC, Manassas, VA) and were adapted to suspension culture as described previously (Byrne et al., 2018b). These were maintained in shake flasks (Corning, Corning, NY) using a Kuhner ISF1-x shaker incubator (Kuhner, Birsfelden, Switzerland) at 37℃, 8% CO2, and 125 rpm. Static cultures were maintained in 96 or 24 well cell culture dishes (Corning) and grown in a Sanyo incubator (Sanyo, Osaka, Japan) at 37℃ and 8% CO2. All cell counts were performed using a TC20TM automated cell counter (BioRad, Hercules, CA) with viability determined by trypan blue exclusion (Thermo Fisher Scientific).

For transient protein expression, plasmid DNA was eluted into endotoxin-free water at concentrations greater than 5 mg/ml for optimal transfection efficiency by electroporation using the OC-400 cuvette and the MaxCyte transfection system (MaxCyte, Gaithersburg, MD) as previously described (Byrne et al., 2018b). 1M Sodium Butyrate was added to a final concentration of 1 mM to limit cell doubling. Cells underwent a temperature shift from 37C to 34C. Cells were maintained in CHO Growth A media and 8mM glutamine (GlutaMAX, Gibco) and additionally supplemented at 10% of culture volume every 3 days with a feed consisting of Proyield Cotton CNE50M-UF protein hydrolysate (FrieslandCampina, Amersfoort, Netherlands), CD Efficient Feed C (Gibco) and 16mM glutamine.

For normal cell growth, cells were maintained in serum free CHO Growth A media (Irvine Scientific, Santa Ana, CA) supplemented with 8mM GlutaMAX (Gibco, Gaithersburg, MD). Cells were initially maintained in serum free CD CHO Medium (Gibco) with 8mM GlutaMAX. However, due to clumping of cells, the media was switched to CHO Growth A. The doubling time for cells was calculated using the formula for population doubling time, DT= (T ln 2)/(ln (Xe/Xb)) where T is the incubation time, Xe is the number of cells at the end of the incubation time and Xb is the number of cells at the beginning of the incubation time.

Detection of secreted proteins by immunoblot

The apparent molecular mass of either purified gp120 or unpurified gp120s in cell culture supernatants was determined by sodium dodecylsulfate

polyacryamide gel electrophoresis (SDS-PAGE) on NuPAGE 4-12% Bis-Tris precast gels (Thermo Fisher Scientific, Invitrogen, Carlsbad, CA). Samples reduced by dithiothreitol (DTT) were run on gels in MES (2-(N-morpholino)ethanesulfonic acid) running buffer (Thermo Fisher) and stained with SimplyBlue stain (Thermo Fisher) For immunoblots, PAGE gels were transferred using iBlot 2 (Thermo Fisher). Membranes were blocked with 5% milk for 1 hour. The primary antibody, a polyclonal goat antibody raised against gp120, was incubated with the membrane at 1.5 ug/ml concentration in 5% milk. Blots were washed three times with PBS with 0.05% Tween (PBS-T). The secondary antibody used was the Peroxidase AffiniPure bovine anti-goat IgG (Jackson ImmunoResearch, West Grove, PA) and incubated with the membrane at a 1:5000 dilution in 5% milk. Blots were washed three times with PBS-T and three times with PBS. The chemiluminescent substrate used was WesternBright ECL horseradish peroxidase substrate (Advansta, Menlo Park, CA). Images were taken using a FluorChem Q imager (Alpha Innotech, San Leandro, CA).

Fluorescence immunoassay (FIA)

A fluorescent immunoassay (FIA) was used to measure gp120 concentrations. In this assay, Fluortrac high binding 96 well plates (Grenier Bio One, Kremsmunster, Austria) were coated overnight at 4°C with 2 ug/ml of 34.1 mAb. Plates were blocked for 5 hours with 2% BSA and washed 4x with PBS-T. Growth-conditioned cell culture supernatant was diluted in 1% BSA/PBS-T and

incubated overnight at 4℃. Antibodies were 3x serially diluted from 10 ug/ml to 0.003 ug/ml and incubated for 1.5 hours. After a 4x wash with PBS-T, secondary antibodies were added at a 1:3000 dilution in 1% BSA/PBS-T and incubated for 1.5 hours. Alexa Fluor 488 Goat α-human IgG (Jackson ImmunoResearch), Donkey α-goat IgG (Invitrogen), or Goat α-mouse (Invitrogen) IgG were used. After a 4x PBS-T wash, wells were filled with 50 ul PBS. Plates were read on an EnVision plate reader (Perkin Elmer Inc, Waltham, MA) at excitation/emission wavelengths of 353/485 nm. All assays were performed in triplicate.

CRISPR/Cas9 knockout of the C1s$^{-/-}$ gene in the CHOK1 cell line and the MGAT1^{-} CHOK1 cell line

The C1s gene was isolated from CHO cells by PCR as described previously (Li et al., 2019b) sequenced, and verified against the NCBI mRNA Reference Sequence XM_007646821.2 in Cricetulus griseus and the predicted protein NCBI Reference Sequence XP_007645011.1. For gene inactivation, the guide RNA, GTTGACAGCCGCTCATGTTG, was synthesized and cloned into the CRISPR Nuclease Vector (GeneArt, Thermo Fisher) which expresses Cas9 and an orange fluorescent protein reporter. To ligate sgRNA inserts into the vector, DNA oligos were synthesized (Eurofins Genomics, Louisville, KY). Bacterial amplification of the completed vectors was carried out in One Shot TOP10 Chemically Competent E. coli following the recommended protocol (Thermo Fisher). Culture were grown in 15 ml LB broth overnight at 37℃ at 250 rpm. Minipreps (Qiagen, Redwood City,

CA) were performed using the recommended protocol and submitted for Sanger sequencing (University of California Core Sequencing Facility, Berkeley, CA) with the U6 primer provided in the GeneArt CRISPR kit. 1 liter Maxipreps (Qiagen) were performed using the recommended protocol. Plasmid DNA was eluted into endotoxin-free water at concentrations greater than 5 mg/ml for optimal transfection efficiency by electroporation using the MaxCyte transfection system as previously described (Byrne et al., 2018b).

For single cell cloning, five 96 well flat-bottomed tissue-culture treated microplates were filled with 50 ul of conditioned serum free CHO Growth A media. Cells were serially diluted to 10 cells/ml in fresh CHO Growth A media and 50 ul of the diluted cells were added to each well, for a final concentration of 0.5 cell/well. Growth was monitored daily; any wells with more than a single colony were discarded. Cells were grown for 20 days at 37C, 8% CO2 and 85% humidity. At 50% confluency, cell culture supernatants from each clone were mixed with exogenous gp120 and analyzed by immunoblot to observe proteolytic activity. For the immunoblot, 200 ng of purified, BaL-gp120 was incubated with 16 ul of supernatant at 37℃ for 48 hours. Cell culture supernatants from clones that were unable to cleave purified BaL-rgp120 were then sequenced to verify presence of CRISPR/Cas9-induced indels). To sequence the C1s genes, 0.5 x 10e6 cells were spun down and boiled in 10 ul of milliQ water. 5 ul of cell lysate was used for PCR using OneTaq Hot Start DNA Polymerase (New England Biolabs). Positive clones with confirmed indels in both copies of the C1s gene were moved into 24 well

plates in a volume of 150 ul. At 50% confluence, clones were moved into 6 well plate in a volume of 3 ml in a shake flask.

Proteolysis of recombinant Factor VIII and Dylight 650-labeled Factor VIII by SDS-PAGE

Recombinant human factor VIII from therapeutic drug, Advate was treated with growth conditioned cell culture supernatant from normal (C1s containing) CHOK1 cell supernatants or supernatants from the $C1s^{-/-}$ CHOK1 2.E7 cell line. For these experiments, Advate was diluted in HBS/C/T-80, FVIII dilution buffer (20 mM HEPES/0.15M NaCl/5mM CaCl2/0.01% TWEEN-80, pH 7.4) to 100 ug/ml. Advate was then diluted with either buffer or CHOK1 cell supernatants at a 1:8 ratio. Samples were incubated at 37C for 24 hours. 5 U/ml of bovine thrombin was incubated with 1 ug Advate as the positive control. 0.4 ug Advate was loaded onto both the SDS-PAGE gel and Dylight-labeled Advate gel. SDS-PAGE gels were stained with GelCode Blue stain (Thermo Fisher). Dylight 650-labeled (Thermo Fisher) FVIII was visualized using a 700 nm channel on an Odyssey imaging system (LI-COR Biosciences, Lincoln, NE).

Results

Proteolysis of clade B gp120s in CHO cells.

For these studies, we selected clade B Envs from three unrelated individuals with the rare elite neutralizer/controller phenotype, EN2, EN6 and EN7 (Mesa, et al. in preparation, Hutchinson, et al. in preparation), similar to other elite neutralizer/controllers previously described (Hutchinson et al., 2019; Mesa et al., 2019). One Env from each individual, was selected for further study based on their bN-mAb-binding ability. The physical and chemical characteristics, organized region by region, for the domain lengths, insertions, deletions and glycosylation sites of the three Envs are summarized (Table 1). The gp120 lengths varied from to 474 to 491 amino acids and the number of predicted N-linked glycosylation sites (PNGS) varied from 17 to 24 PNGS. In contrast, the protease resistant control Envs, A244 and EN3, ranged from 499 to 523 amino acids in length and contained from 20 to 25 PNGS (Supp Fig 1). The three Envs contain the protease-sensitive motifs Gly-Pro-Gly-Arg, or GPGR, at the crown of the V3 loop (Clements et al., 1991; Li et al., 2019b) (Table 2). In contrast, the protease resistant A244 isolate and the clade B EN3 possess GPGQ and GPGG at the crown of the V3 domain, respectively.

Table 1. Physical characteristic of elite neutralizer/controller rgp120s used in this study.

		EN2_005		EN6_226		EN7_090	
Domain	**HXB2**	**Length (aa)**	**PNGS**	**Length (aa)**	**PNGS**	**Length (aa)**	**PNGS**
Signal	**1-30**	29	0	29	0	29	0
C1	**31-130**	102	2	100	0	100	2
V1	**131-157**	35	5	25	1	32	5
V2	**158-196**	41	2	38	2	39	2
V1/V2	**131-196**	76	7	63	3	71	7
C2	**197-295**	99	5	99	5	99	6
V3	**296-331**	35	1	35	1	35	1
C3	**332-384**	52	2	52	2	52	3
V4	**385-418**	32	4	31	4	35	4
C4	**419-459**	43	1	41	1	41	0
V5	**460-469**	10	2	11	1	14	2
C5	**470-511**	42	0	42	0	42	0
Extracellular	**512-678**	166	5	166	4	166	4
Transmembrane	**679-699**	22	0	22	0	22	0
Cytoplasmic	**700-856**	157	1	157	1	157	1
gp120	**1-511**	491	24	474	17	489	25
gp41	**512-856**	345	6	345	5	345	5
gp160	**1-856**	836	30	819	22	834	30

The length in amino acids (AA) and number of potential N-linked glycosylation sites (PNGS) are listed for each of the HIV Env domains for the expressed Envs: EN2_005, EN6_226 and EN7_090.

Table 2. V3 domain crown sequences for EN rgp120s

Env	Sequence	Clade	Cleaved by CHO C1s
EN2	GPG**R**VWYA	B	+
EN6	GPG**R**PEGI	B	+
EN7	GPG**R**AFYA	B	+
A244	GPG**Q**VFYR	AE	-
EN3	QPGGAIYA	B	-

V3 domain crown sequences for the transiently expressed Envs are listed, along with the virus clade and summary of susceptibility to proteolysis by CHO C1s.

When EN2, EN6, and EN7 were transiently transfected in normally CHOK1 cells and analyzed by immunoblot, non-reduced gp120 was observed at 120 kDa (Fig 8). After reduction, additional bands of 80 and 50 kDa bands appeared that were indicative of proteolysis at the V3 domain. The same cleavage pattern occurred with Envs from clade B, BaL and MN and other strains, as was previously observed (Du et al., 2008; Pugach et al., 2015; Li et al., 2019b). Full-length EN6 has a lower molecular weight of 105 kDa compared to EN2 and EN7 due to significantly fewer glycans and was cleaved into 60 kDa and 45 kDa fragments. Additionally, we examined the protease sensitivity of two Envs (A244 and EN3) that possessed sequences in the V3 crown that differed from the clade B consensus GPGR sequence. The CRF01_AE A244 isolate from Thailand possesses the GPGQ sequence typical of non-clade B and non-clade F2 Envs as shown in the consensus

sequences for the clades of Group M (Table 3). Although EN3 is a clade B virus, it possesses an unusual QPGG sequence at the crown of the V3 region. As expected, neither Env was proteolyzed when expressed in CHO cells. This data suggests that there is no difference in accessibility to C1s cleavage in the V3 domain between Envs and that all Envs with the GPGR motif will be subject to proteolysis.

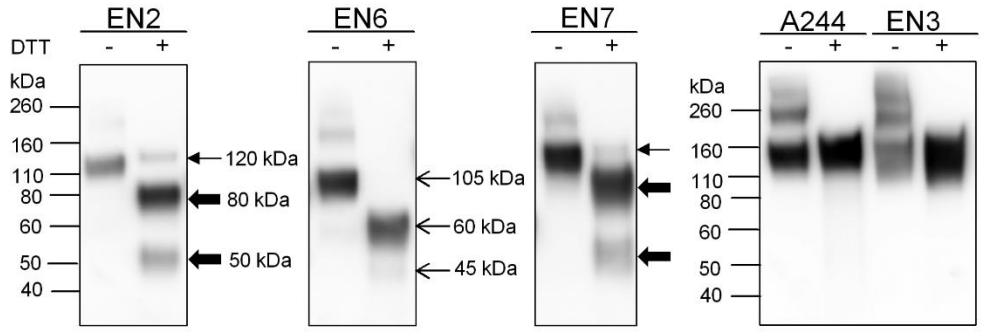


Figure 8. Proteolysis of six HIV Env gp120s expressed in CHO cells as visualized by immunoblot.

Clade B HIV Env gp120s from four individuals with neutralization breadth (EN2, EN6, EN7, EN3) were transiently transfected into normal CHOK1 cells. Cell culture supernatant was analyzed by immunoblot, and samples were run non-reduced and reduced with DTT. Full-length gp120 is indicated with thin arrows and the proteolyzed 80kDa and 50 kDa fragments are indicated with thick arrows for EN2 and EN7. For EN6, full-length gp120 runs as 105 kDa due to the presence of only 17 PNGS and has smaller proteolyzed fragments than EN2 and EN7. Blots were probed with a goat polyclonal serum to gp120. The clade CRF01-AE, A244-rgp120 that lacks the clade B GPGRAF consensus sequence was expressed in normal CHOK1 cells and used as a negative control.

Table 3. V3 consensus sequences

Clade	V3 Consensus Sequence
Con A1	GPG**Q**AFYA
Con A2	GPG**Q**AFYT
Con B	GPG**R**AFYT
Con C	GPG**Q**TFYA
Con D	GPG**Q**ALYT
Con F1	GPG**Q**AFYA
Con F2	GPG**R**AFYA
Con G	GPG**Q**AFYA
Con H	GPG**Q**AFYA
Con AE	GPG**Q**VFYR

HIV Env consensus sequences for the V3 crown are listed for all of the major clades of HIV (from the Los Alamos National Laboratory HIV Sequence Database). Shaded rows have sequences with the GPGR motif recognized by C1s.

Creation of a CHOK1 cell line deficient in C1s

Because previous studies (Byrne et al., 2018b; O'Rourke et al., 2018b; Li et al., 2019b) suggested that stable cell CHO-S lines (ThermoFisher) expressing gp120 tended to form clumps in suspension cultures, we initially surveyed gp120 expression in several other cell lines including DG44 and CHOK1. Of these cell lines, CHOK1 had better growth in our suspension culture system. We used CRISPR/Cas9 gene editing to knockout the C1s protease in a suspension-adapted CHOK1 cell line. The C1s gene contains 11 exons, with exon 11 coding for the serine protease domain (Fig 9A) (Hammond et al., 2012). In a previous study, we screened potential guide RNAs targeting various exons within the C1s gene for knockout efficiency and found one that efficiently knocked out C1s proteolytic activity in the CHO-S cell line (Li et al., 2019b). In this study, we used the same guide RNA targeting exon 11 to inactivate the C1s gene in CHOK1 cells. The procedure for creating the $C1s^{-/-}$ CHOK1 cell line is described in detail in the Methods and Materials section. Using this strategy, 32 clones transfected with the CRISPR/Cas 9 guide RNA were analyzed by immunoblot. The supernatants of these clones were screened for the ability to cleave exogenous purified rgp120 and of these clones, three appeared to lack C1s protease activity (Fig 9B). The C1s gene from these three clones were PCR-amplified from genomic DNA extracted from CHO cells and all three were verified as successful knockouts by Sanger sequencing. To select the highest producing clone, clones 2.E7, 2.E9 and 3.E8 were transiently transfected with BaL-rgp120 (Fig 9C). Of these, clone 2.E7 was selected

for use as the C1s$^{-/-}$ CHOK1 cell line. The C1s gene has two copies in CHO cells (Li et al., 2019b). A different indel was observed in each copy: an insertion of one thymine in one gene and an insertion of two thymines in the other (Fig 9D and Table 4). To verify that genetic inactivation of the C1s gene prevented proteolysis of gp120, we expressed the genes encoding three different clade B gp120s, EN2, EN6 and EN7, in the C1s$^{-/-}$ CHOK1 2.E7 cell line and compared their sensitivity to proteolysis with the same proteins expressed in the normal CHOK1 cell line possessing an intact C1s gene (Fig 10A). As expected, inactivation of the C1s gene prevented proteolysis of all three proteins.

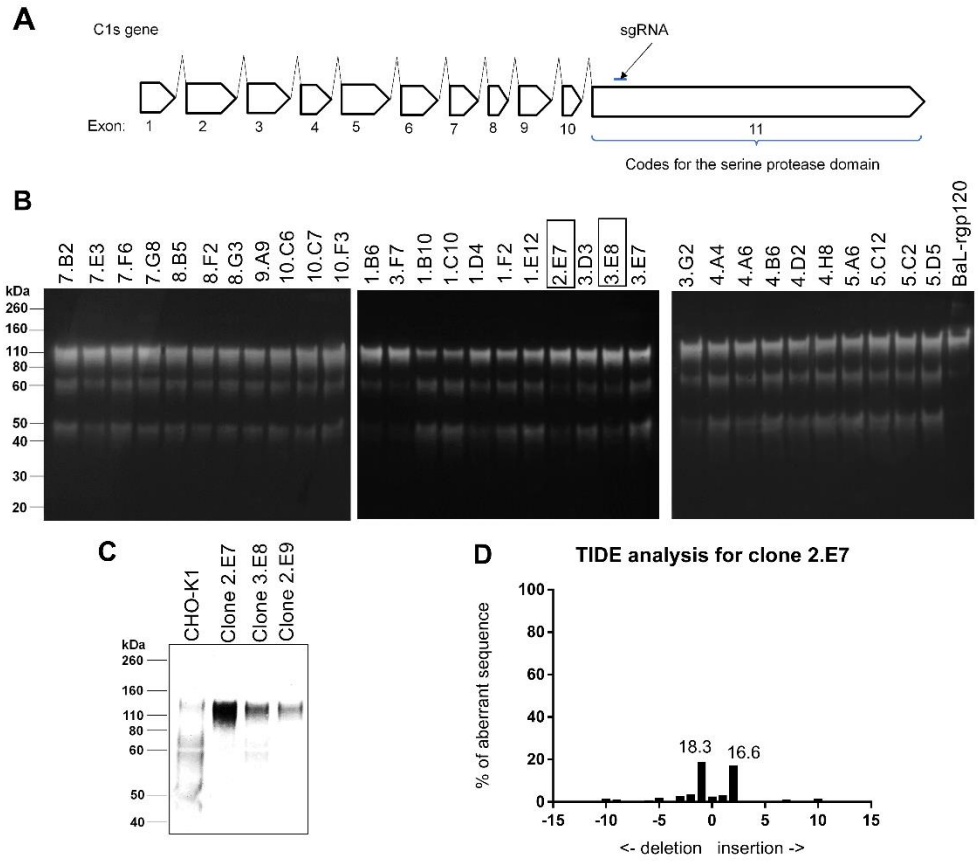


Figure 9. Selection of C1s$^{-/-}$ CHOK1 2.E7.

A. A schematic for the C1s gene, containing 11 exons. The single guide RNA targeting exon 11 used to direct Cas9 is indicated by the arrow. B. Clones with both alleles of C1s knocked out were screened by Western Blot. 100 ng of purified BaL-rgp120 was incubated with 15 ul of supernatant from 24 well plates overnight at 37℃. Samples were run reduced with DTT. Clones marked with a black box have knockouts of both alleles as verified by sequencing. C. Bal-rgp120 was transiently transfected into C1s$^{-/-}$ CHOK1 clones 2.E7, 2.E9, and 3.E8 and cultures were grown for 6 days. 5 ul of supernatant was run reduced with DTT on Western Blot. D. The sequence around the expected indel site was analyzed by TIDE (Tracking of Indels by Decomposition, tide.deskgen.com) for clone 2.E7 to check for knockout of both alleles. The indel site was amplified by PCR from genomic DNA extracted from the CHO cell clone and Sanger sequenced. The peaks from Sanger sequencing chromatograms are decomposed to determine the proportion and identity of the indels present in the clone.

Table 4. C1s allele sequences for C1s$^{-/-}$ CHOK1 Clone 2.E7

Wildtype C1s sequence	GTTGACAGCCGCTCATG↓TTGTGG	
Mutation in Allele 1	GTTGACAGCCGCTCATG↓**TT**TTGTGG	Inserted TT
Mutation in Allele 2	GTTGACAGCCGCTCATG↓~~T~~TTGTGG	Deleted T

The sequence around the indel site was PCR amplified and cloned into a plasmid in order to sequence the two different mutations for each allele. The sequences of the two mutations observed at the expected indel site resulting in knockout of C1s in the CHOK1 cell line. The arrow indicates where the expected indel should occur, or three nucleotides before the protospacer adjacent motif (PAM). In this case, the PAM site is the last three nucleotides, TGG. Letters in bold indicate an insertion while letters with strikethrough indicate deletion.

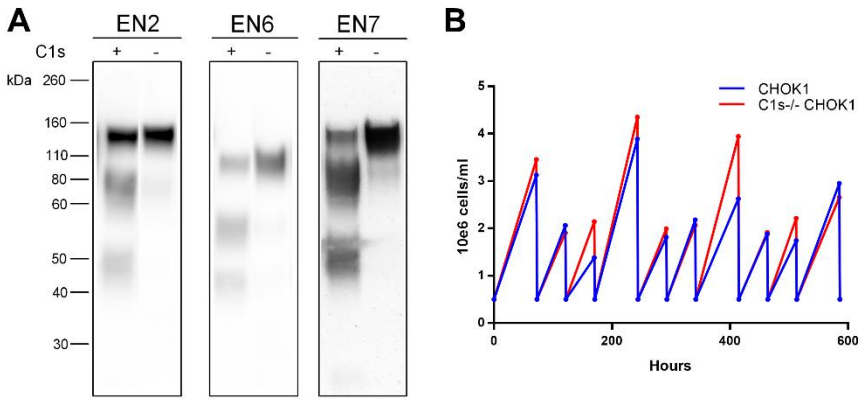


Figure 10. Characteristics of the C1s$^{-/-}$ CHOK1 2.E7 cell line.

A. Comparison of proteolysis of clade B gp120s expressed in CHO cells with and without the presence of C1s. EN2, EN6 and EN7 rgp120s were transiently transfected in the CHOK1 and the C1s$^{-/-}$ CHOK1 2.E7 cell line by electroporation. rgp120s expressed in the CHOK1 cell line are indicated with (+) and those expressed in the C1s$^{-/-}$ CHOK1 2.E7 cell line are indicated with (-). Supernatants were analyzed by immunoblot using an anti-rgp120 polyclonal antibody. Cell culture supernatant was assayed by immunoblot with a goat polyclonal serum to gp120. B. The viable cell densities, or VCD, plotted against time for the CHOK1 cell line and the C1s$^{-/-}$ CHOK1 2.E7 cell line. VCD was counted prior to splitting cells every two or three days for ten passages.

Next, we examined the genetic stability of the C1s$^{-/-}$ CHOK1 2.E7 cell line. Viable cell densities were recorded over ten passages every two to three days (Fig 10B). The C1s$^{-/-}$ CHOK1 2.E7 cell line had a doubling time of 24.5 hours which was comparable to, and slightly better than the parental, suspension-adapted CHOK1 cell line, which had a doubling time of 26.6 hours. The shorter doubling time of the C1s$^{-/-}$ CHOK1 2.E7 cell line is likely due to the clonal selection process for the C1s$^{-/-}$ CHOK1 2.E7 line compared to the heterogenous CHOK1 cell line.

Thus, the C1s$^{-/-}$ CHOK1 2.E7 cell line is a robust cell line and inactivation of the C1s gene did not have deleterious effects on cell growth.

Creation of the C1s$^{-/-}$ MGAT1^{-} CHOK1 cell line

Previous studies have shown that many broadly neutralizing monoclonal antibodies to HIV recognize glycan dependent epitopes in gp120 (Walker et al., 2009, 2011; Julien et al., 2013; Sok et al., 2014). Moreover, these studies showed that efficient binding of these antibodies required the mannose-5 form of N-linked glycosylation rather than the normal mature sialic acid containing form of N-linked glycosylation. The shorter, high-mannose glycans more closely resemble the type of glycosylation on authentic HIV virions which imparts enhanced binding of bN-mAbs (Doores et al., 2010; Bonomelli et al., 2011; Byrne et al., 2018b; Doran et al., 2018a). Previously, we showed that gene editing could be used to inactivate the MGAT1^{-} gene, a critical intermediate in the N-linked glycosylation pathway (Byrne et al., 2018b). Inactivation of this gene resulted in enrichment for mannose 5 and earlier intermediates in the N-linked glycosylation pathway. Therefore, it was of considerable interest to create a CHOK1 cell line deficient in both the MGAT1 and C1s genes for the production of clade B gp120s. For this purpose, we used CRISPR/Cas9 to inactivate the C1s gene in an MGAT1^{-} CHOK1 cell line made in a similar manner to the MGAT1^{-} CHO-S cell line described in a previous paper from our lab (Byrne et al., 2018b). The resulting C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1 cell line was used to express several clade B rgp120s.

We tested the ability of the EN2, EN6 and EN7 rgp120s to bind a panel of bN-mAbs representative of major glycan dependent epitopes. We have previously shown that gp120s expressed in the MGAT1- CHO-S cell line displayed properly truncated oligomannose glycans (Byrne et al., 2018b). EN2 rgp120 expressed in the $C1s^{-/-}$ $MGAT1^{-}$ CHOK1 1.A1 cell line had a lower molecular mass of about 120 kDa compared to EN2 gp120 expressed in the $C1s^{-/-}$ CHOK1 2.E7 cell lines with a mass of 95 kDa (Fig 11C). This was expected since the loss of multiple sialic acid residues per glycan has been shown to cause a visible change in molecular weight (Eggink et al., 2010; Byrne et al., 2018b). We found that Envs produced in the $C1s^{-/-}$ $MGAT1^{-}$ CHOK1 1.A1 cell line (Fig 11A) displayed substantially greater binding to the glycan-dependent bN-mAbs, PG9 and PGT128 (Walker et al., 2009; Pejchal et al., 2011), compared to the same Envs expressed in the $C1s^{-/-}$ CHOK1 2.E7 cell line (Fig 11B). Binding of the glycan dependent 10-1074 bN-mAb and the glycan independent VRC01 bN-mAb remained the same in both cell lines (Mouquet et al., 2012; Zhou et al., 2010).

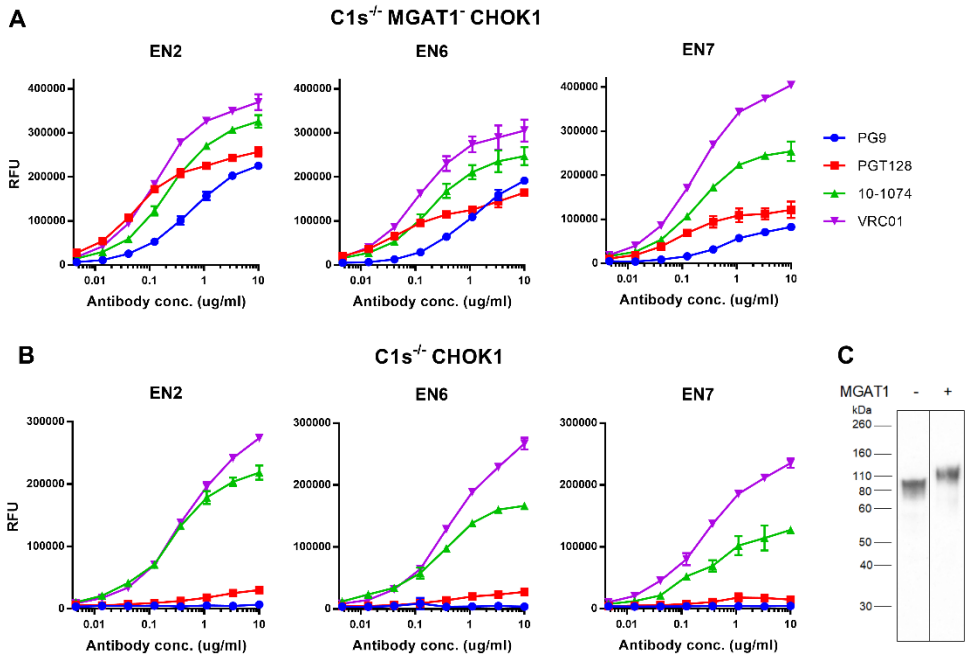


Figure 11. Enhanced binding of glycan-dependent bN-mAbs to clade B rgp120s expressed in cell lines deficient in C1s and MGAT1.

A. Plasmids encoding three different gp120s (EN2, EN6, and EN7) were transfected into the C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1. Diluted cell culture supernatants were collected after 4 days, captured with the 34.1 mAb for a capture fluorescence immunoassay (FIA) and assayed against a panel of bN-mAbs. B. The same EN2, EN6 and EN7 rgp120 monomers were transiently expressed in the C1s$^{-/-}$ CHOK1 2.E7 cell line to show differences in antigenicity due to knockout of MGAT1 when compared to the C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1 cell line. All assays were performed in triplicate. C. EN2 was expressed in the C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1 cell line, indicated with a (-), and the C1s$^{-/-}$ CHOK1 2.E7 cell line, indicated with a (+). Growth-conditioned cell supernatant was immunoblotted with an anti-rgp120 goat polyclonal antibody.

To explore the effect of proteolysis by the C1s protease on antigenicity of rgp120, we compared EN6 rgp120 expressed in the C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1 cell line (not proteolyzed EN6) and MGAT1^{-} CHOK1 cell line (proteolyzed EN6) (Fig 12). As expected, we found that the binding of EN6 rgp120 to PG9 and VRC01

was not affected by the C1s protease inactivation as these bN-mAbs recognize the V1/V2 domain and CD4 binding site, respectively. However, binding of proteolyzed EN6 to PGT128, which recognizes the V3 stem, was significantly decreased as is consistent with the presence of the C1s site. Similarly, binding of proteolyzed EN6 to bN-mAb 10-1074, which recognizes the V3 stem, and mAb 447-52D, which recognizes the V3 crown, was also diminished (Zolla-Pazner et al., 2004). Thus, expression of Envs in the $C1s^{-/-}$ $MGAT1^{-}$ CHOK1 1.A1 cell line prevented proteolysis of the rgp120s at the V3 loop and promoted binding to glycan-dependent and V3 bN-mAbs.

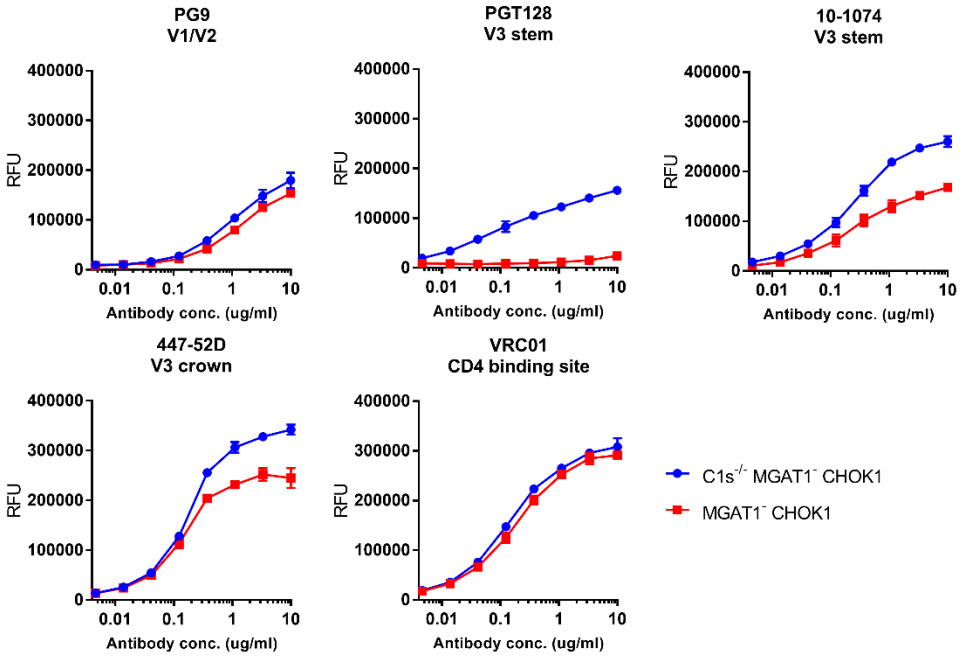


Figure 12. Effect of proteolysis by C1s on the binding of neutralizing monoclonal antibodies and broadly neutralizing monoclonal antibodies to EN6 rgp120.

Plasmids encoding EN6 rgp120 were transiently expressed in the C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1 and MGAT1^{-} CHOK1 cell lines. Growth-conditioned cell culture supernatants were diluted for a 34.1 mAb capture FIA. All assays were performed in triplicate. Indicated under the antibody names are the epitope recognized by the respective antibodies.

Partial proteolysis of Factor VIII thrombin cleavage sites in gp120-Factor VIII chimeric constructs

Next, we wanted to explore the substrate specificity of C1s by measuring its effect on other recombinant proteins susceptible to proteolysis. For these studies, we selected human factor VIII (FVIII). Human FVIII is a large protein of 2332 amino acids, is part of the coagulation pathway, and is necessary for the formation of fibrin found in blood clots (Lenting et al., 1998). It undergoes both maturational cleavages mediated at a furin-cleavage site and by thrombin as well as artifactual proteolysis mediated by an unknown protease when expressed in CHO cells (Lenting et al., 1998; Hultin and Jesty, 1981; Kaufman et al., 1988). FVIII is activated through cleavage by thrombin at R372 (Arg372), R740 and R1689 (Fig 13A) (Warren et al., 1999). Similar to rgp120, artifactual proteolysis of FVIII is mediated by an endogenous CHO protease. This proteolysis occurs at FVIII's three thrombin cleavage sites (Kaufman et al., 1988). Both thrombin and C1s are serine proteases, and since thrombin is able to cleave the GPGRAF sequence in gp120, we wanted to examine the possibility that C1s was the endogenous protease responsible for artifactual proteolysis of recombinant Factor VIII when expressed in CHO cells.

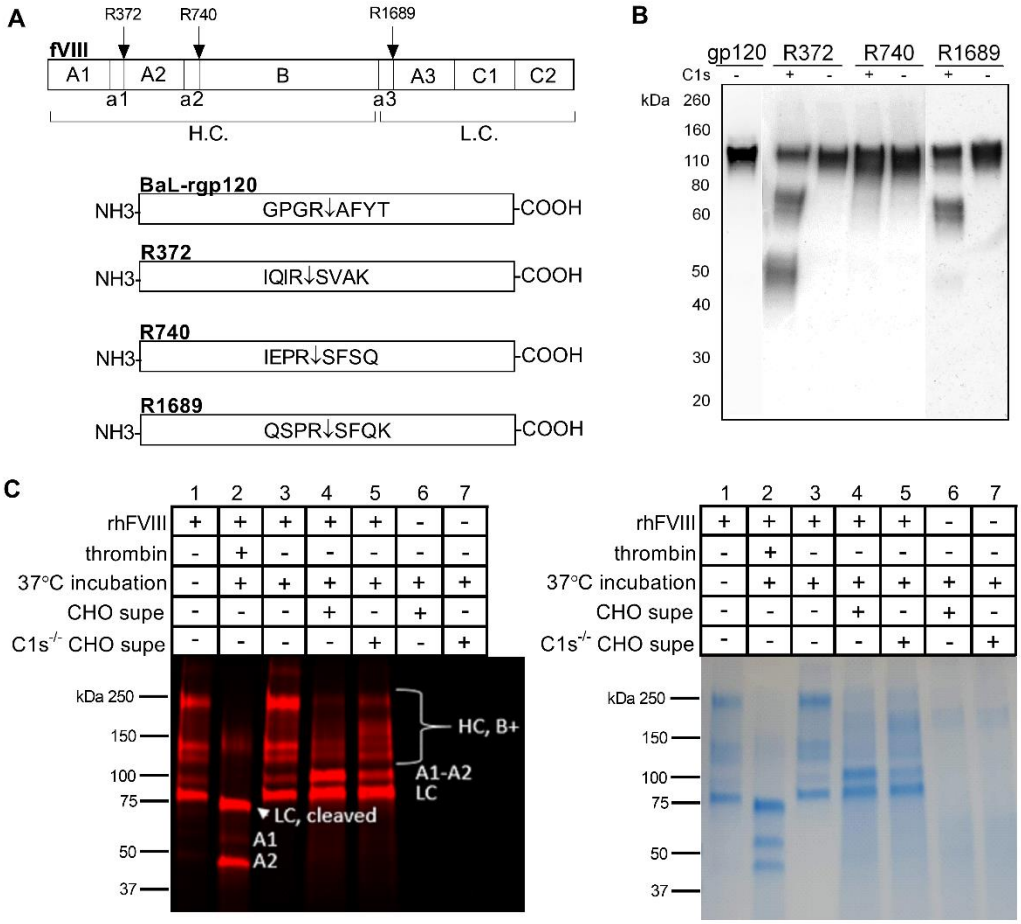


Figure 13. Limited proteolysis of recombinant human Factor VIII by CHO cell-derived C1s.

A. The FVIII gene is shown with the three thrombin cleavage sites, R372, R740 and R1689. The heavy chain and light chain are labeled H.C. and L.C. Chimeric gp120-FVIII constructs were created by replacing the C1s cleavage site in the V3 crown of gp120 with thrombin cleavage sites from Factor VIII. The three thrombin cleavage sites on human Factor VIII were cloned into BaL-rgp120 to replace the GPGR↓AFYT sequence. Three cloned constructs were created with each of the three thrombin cleavage sites. B. Transient expression and cleavage of the chimeric gp120-Factor VIII constructs R372, R740 and R1689 in either the CHOK1 cell line or the C1s$^{-/-}$ CHOK1 2.E7 cell line. Chimeric constructs expressed in the CHOK1 cell line are indicated with (+) and those expressed in the C1s$^{-/-}$ CHOK1 2.E7 cell line are indicated with (-). Cell culture supernatant was assayed by immunoblot, and samples were all run reduced with DTT. C. Recombinant human Factor VIII (rhFVIII), also known as Advate, was labeled with DyLight650 and tested for proteolysis following treatment with CHO cell and C1s$^{-/-}$ CHOK1 2.E7 cell

supernatants. Shown are analyses of FVIII fragments by SDS-PAGE. Thrombin was incubated with rhFVIII as a positive control to observe the expected protein fragments after proteolysis at R372, R740 and R1689. Factor VIII fragments are labeled on the immunoblot as HC (heavy chain), LC (light chain), and the A1 and A2 domains.

In a pilot study, we created chimeric gp120-FVIII constructs where the GPGR motif in the V3 loop was replaced by each of the FVIII cleavage sites to observe whether the FVIII thrombin cleavage sites were sensitive to proteolysis by C1s. The V3 sequence, GPGR/AFYT, was replaced with each of the three thrombin cleavage sites, R372 (IQIR/SVAK), R740 (IEPR/SFSQ), and R1689 (QSPR/SFQK), such that these sequences would be accessible to C1s and could be assayed by the same immunoblot methodology used to study the proteolysis of rgp120 (Fig 13A). The chimeric rgp120-Factor VIII constructs were transiently transfected in the CHOK1 and $C1s^{-/-}$ CHOK1 2.E7 cell line. We found that C1s, secreted by the normal CHOK1 cells, was able to cleave the chimeric rgp120s. The R372 and R1689 thrombin cleavage sites in the gp120-Factor VIII chimeric constructs were cleaved by CHO C1s, but the R740 thrombin cleavage site was resistant to cleavage (Fig 13B).

Resistance of recombinant human Factor VIII to proteolysis by CHO C1s

While the chimeric constructs provided some information on the substrate specificity of C1s, we were interested in treating Factor VIII in its native structure with C1s to observe its susceptibility to C1s-mediated proteolysis. In these studies,

we utilized Advate, a recombinant FVIII therapeutic protein. FVIII was incubated with growth-conditioned supernatants from CHOK1 cells and C1s$^{-/-}$ CHOK1 2.E7 cells to determine whether proteolysis of FVIII occurred after treatment with CHOK1 cell supernatants and whether supernatants from cells lacking C1s prevented the proteolysis. FVIII is composed of a glycosylated 200 kDa heavy chain and an 80 kDa light chain, held together by a divalent metal ion (Fig 13A) (Pipe, 2009). In lane 2, Advate was cleaved by thrombin to show the expected cleavage products (Fig 13C). The heavy chain was cleaved in two places resulting in three fragments: the A1 domain, the A2 domain and the B domain. The 80 kDa light chain was cleaved to a 73 kDa fragment. In lanes 3 and 4, proteolysis of the heavy chain was observed at the furin cleavage sites at R1313 and R1648 for Advate treated with supernatants from both the C1s-containing and C1s-deficient cell lines. This indicated the presence of a furin-like protease secreted from CHO cells which was not C1s (Lind et al., 1995). Slightly less proteolysis of the heavy chain for the Advate + C1s$^{-/-}$ CHOK1 2.E7 supernatant than with Advate + CHOK1 supernatant was observed, indicating that some proteolysis was caused by C1s. The expected FVIII fragments of the cleaved light chain at 73 kDa, the A1 domain at 50 kDa and the A2 domain at 43 kDa after cleavage by thrombin were not observed when FVIII was incubated with CHO supernatant, suggesting that CHO C1s was not involved in the proteolysis of FVIII despite the fact that the R372 and R1689 sites was susceptible to cleavage in the gp120-FVIII chimeric constructs. The accessibility of the thrombin cleavage sites is different in Advate, the recombinant

human Factor VIII protein, compared to those in the chimeric gp120-FVIII constructs we created. In the gp120-FVIII constructs, the thrombin cleavage sites were placed in the V3 loop, a long, extended loop of approximately 35 residues. Normally, the V3 loop extends past the core of gp120 to bind to the co-receptor on T-cells. However, this also makes the loop accessible to proteolysis. Conversely, the thrombin cleavage sites on human Factor VIII only serve to activate the protein. These cleavage sites might not be readily proteolyzed by C1s due to tertiary and quarternary structure imparted by the multiple, globular domains of Factor VIII. In sum, we did not observe the proteolysis of recombinant Factor VIII when expressed in CHO cells as described by Kaufmann et al. (Kaufman et al., 1988).

The substrate specificity of CHO C1s

While the substrate specificity of thrombin has been well-studied, the substrate specificity of C1s produced in CHO cells has not been determined. Here, we compare the substrate specificity of CHO C1s to human thrombin to observe differences in the amino acids they favor and the substrates they will cleave. Because human C1s shares a 77% amino acid sequence identity with CHO C1s and has a known substrate specificity and solved crystal structure, we used human C1s to determine the specificity of CHO C1s (Gaboriaud et al., 2000; Kerr et al., 2005; Pang et al., 2017). Human C1s prefers substrates with small amino acids such as glycine and alanine in the P2 position, as shown in the cleavage site sequence logo from the MEROPS peptidase database (Fig 14A) (Rawlings et al., 2014). The

common nomenclature for protease substrates, P4-P3-P2-P1-P1'-P2'-P3'-P4', is used where P1 is the residue immediately before the scissile bond. CHO C1s also shows a similar substrate specificity for small amino acids in the P2 position based on experiments with HIV Env and the physiological relevant substrates of C1s, C2 and C4 (Table 5). CHO C1s cleaves HIV Envs with the GPGR motif as well as the complement component C2 protein, both of which have glycine in the P2 position. CHO C1s also cleaves complement component C4 which has an alanine in the P2 position. Thus, CHO C1s efficiently cleaves substrates with a small amino acid in the P2 position.

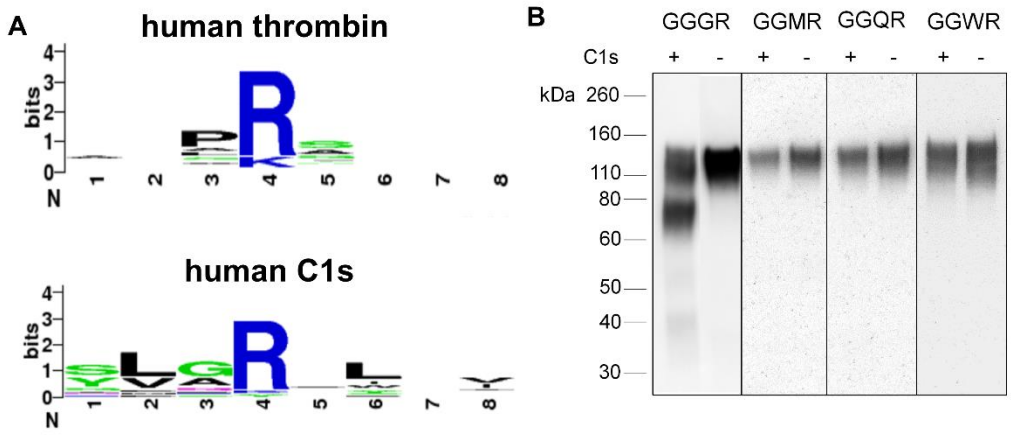


Figure 14. Determining the substrate specificity of C1s.

A. The cleavage site sequence logos for human thrombin and human C1s are graphed based on data from the MEROPS database (Rawlings et al., 2014). Positions 1 through 8 represent positions P4-P3-P2-P1-P1'-P2'-P3'-P4' as per conventional nomenclature for substrates of proteases. B. Expression of gp120 variants with amino acid substitutions in the V3 crown sequence. GGSR, GGMR, GGQR and GGWR were substituted for GPGR in the V3 crown of BaL-rgp120. BaL-rgp120 was transiently expressed in the CHOK1 cell line and $C1s^{-/-}$ CHO 2.E7 cell line. Mutagenesis constructs expressed in the CHOK1 cell line are indicated with (+) and those expressed in the $C1s^{-/-}$ CHOK1 2.E7 cell line are indicated with (-). Cell culture supernatant was assayed by immunoblot, and all samples were reduced with DTT.

Table 5. Known substrates of CHO C1s and the substrates tested for proteolysis by CHO C1s

Substrate	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'	Cleaved by CHO C1s	Side Chain Type
Clade B BaL gp120	Gly	Pro	**Gly**	Arg	Ala	Phe	Tyr	Thr	+	small
complement component C2	Asn	Leu	**Gly**	Arg	Arg	Ile	Gln	Ile	+	small
complement component C4	Gly	Leu	**Ala**	Arg	Ala	Gln	Glu	Val	+	small
Factor VIII R372	Ile	Gln	**Ile**	Arg	Ser	Val	Ala	Lys	+	medium
Factor VIII R740	Ile	Glu	**Pro**	Arg	Ser	Phe	Ser	Gln	-	bulky
Factor VIII R1689	Gln	Ser	**Pro**	Arg	Ser	Phe	Gln	Lys	+	bulky
GGGR-GGGG	Gly	Gly	**Gly**	Arg	Gly	Gly	Gly	Gly	+	small
GGMR-GGGG	Gly	Gly	**Met**	Arg	Gly	Gly	Gly	Gly	-	long
GGQR-GGGG	Gly	Gly	**Gln**	Arg	Gly	Gly	Gly	Gly	-	long
GGWR-GGGG	Gly	Gly	**Trp**	Arg	Gly	Gly	Gly	Gly	-	long, bulky

Shown are the amino acids in the P4-P4’ positions of the substrates, the susceptibility to proteolysis by CHO C1s, and a descriptor for the amino acid in the P2 position.

Testing for the proteolysis of the thrombin cleavage sites on Factor VIII further revealed the nuanced substrate specificity of CHO C1s. As discussed previously, CHO C1s recognized two out of three thrombin cleavage sites, R372 and R1689, in the gp120-Factor VIII chimeric constructs (Table 5). R372 has the sequence IQIR/SVAK and has an isoleucine in the P2 position. This sequence was recognized by CHO C1s due to the similarity between isoleucine and alanine, which have short, aliphatic side chains. R1689 has the sequence QSPR/SFQK with a serine in the P3 and a proline in the P2 position. Based on the premise that C1s prefers small amino acids in the P2 position, it was surprising to see that CHO C1s recognized and cleaved this particular sequence due to the larger and more bulky proline in the P2 position. It is possible that the presence of the serine, a small amino acid, in the P3 position allowed C1 to accommodate larger amino acids in the P2 position. The third thrombin cleavage site, R740, which has the sequence IEPR/SFSQ with a glutamine in the P3 position and a proline in the P2 position, was not cleaved by CHO C1s. The long-chained, negatively charged glutamine in combination with the proline made this substrate inaccessible to CHO C1s.

Mutagenesis of the P2 position of gp120-Factor VIII constructs was done to determine the extent by which proteolysis is controlled by the residue in the P2 position (Fig 14B). The P4 position was kept as arginine, while all other residues in the P4-P4' positions other than P2 were mutated to glycine, to create a GGXRGGGG sequence (Fig 7B). Glycine (G), methionine (M), glutamine (Q) and tryptophan (W) were mutated into the P2 position based on peptide microarray

substrate specificity data for human C1s from Gosalia et al (Gosalia et al., 2005). The peptide microarray data showed that C1s would not cleave substrates with methionine, glutamine or tryptophan in the P2 position, but could cleave substrates with glycine in the P2 position. gp120-Factor VIII constructs with serine, methionine, glutamine and tryptophan in the P2 position were transiently expressed in the CHOK1 and C1s$^{-/-}$ CHOK1 2.E7 cell lines to determine whether CHO C1s could cleave these substrates. The construct with glycine in the P2 position was cleaved, but the constructs with methionine, glutamine and tryptophan were not cleaved. This data further supported the conclusion that CHO C1s could cleave substrates with small amino acids including glycine and alanine in the P2 position but would be obstructed by substrates with a larger residue.

Thrombin has been shown to cleave substrates including clade B gp120s, coagulation factor proteins, and fibronectin and can recognize substrates with not only proline in the P2 positions, but also hydrophobic and non-aromatic amino acids including glycine, alanine, valine, leucine, isoleucine and glutamine based on data from the MEROPS peptidase database (Fig 14A) (Clements et al., 1991; Gallwitz et al., 2012). However, we have shown that CHO C1s has a more limited specificity as it cannot cleave the R740 thrombin cleavage site in our gp120-Factor VIII chimeric constructs. Additionally, we showed that CHO C1s is specifically hindered by the residue in the P2 position, as it cannot cleave substrates with methionine, glutamine or tryptophan in the P2 position. These experiments have shown that CHO C1s has a narrower specificity than thrombin but in certain cases,

can accommodate amino acids such as isoleucine or proline in the P2 position when flexibility of the substrate is imparted by a smaller residue in the P3 position.

Structural analysis of the CHO C1s substrate binding site

To understand how the structure of CHO C1s determines its substrate specificity, we studied the previously solved crystal structure of human C1s in complex with the substrate, gigastasin (PDB 5UBM) (Fig 15A) (Pang et al., 2017). The residues Y610 and F526 interact with the P4 residue F62 to form a hydrophobic box, a feature also seen in coagulation factor proteases that favors a bulky, hydrophobic residue in the P4 position (Stubbs and Bode, 1994). The P3 residue K63 protrudes out and away from the protease. The P2 residue C64 is small and restricted by the bulky F526 and H475. The P1 residue R65 is in the S1 pocket in which D626 forms the bottom and G663 and S627 form the sides. The catalytic serine S632 is situated adjacent to the S2 pocket. The charged K629 potentially obstructs unsuitably sized substrates from entering the substrate binding site. The C-terminal side of the scissile bond is moderately restricted by W476 and the more distal residues P458, W457, W459 and R578.

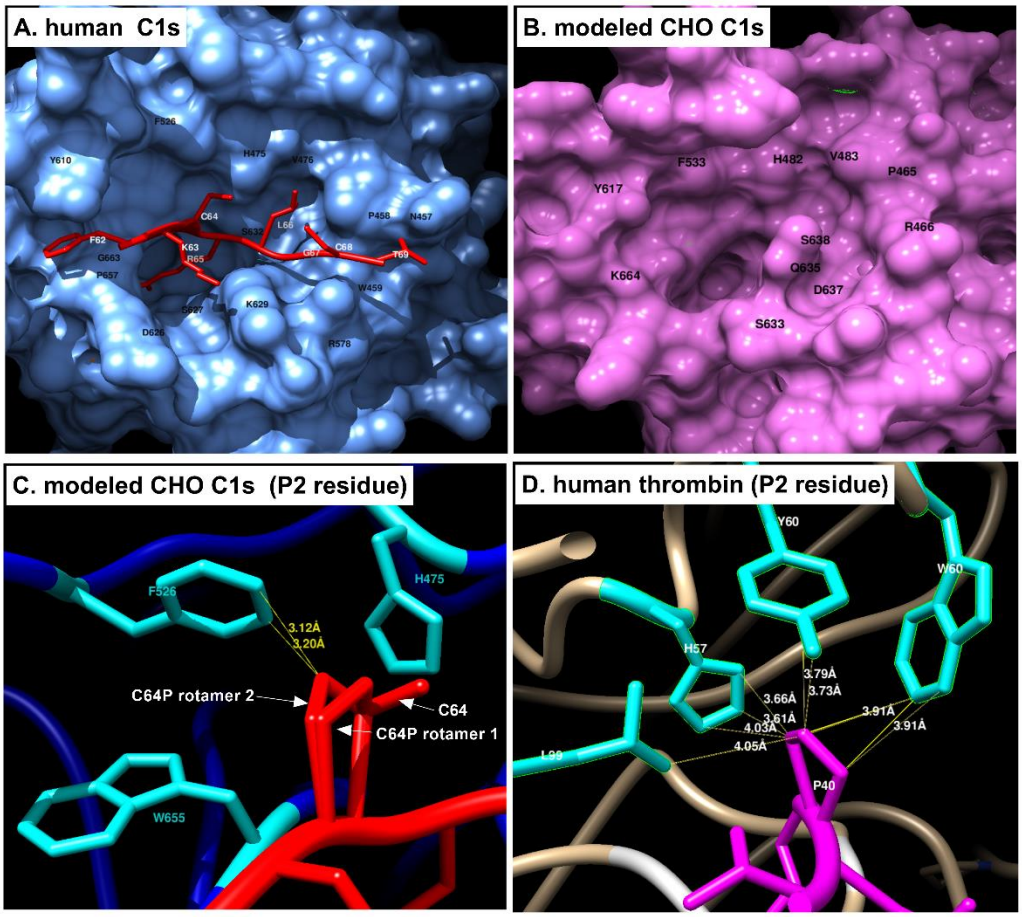


Figure 15 Homology modeling of the structure of the CHO C1s active site.

A. The crystal structure of human C1s with the substrate gigastasin (PDB 5UBM) was visualized using Chimera. The P4-P4' residues of gigastasin are in red, while the protease is in blue. Residues on the protease that interact with the substrate are labeled in black. The P4-P4' residues of the substrate are labeled in white. B. A homology model of CHO C1s was created with SWISS-model using the sequence of CHO C1s and the crystal structure of human C1s as a template. In black are the residues of CHO C1s that likely interact with the substrate based on the homology to human C1s and predicted van der Waals interactions. C. A zoomed in view of Cysteine64 (C64) in the P2 position of gigastasin (red) with human C1s (blue). Residues within 5 Angstroms of C64 are in cyan. C64 was computationally mutated into a proline. Shown are two possible rotamers using the Dunbrack 2010 rotamer library in Chimera. Parameters for clashing were as follows: atoms with VDW overlap >= 0.4 angstroms, subtract 0.2 from overlap for potentially H-bonding pairs. Distances in yellow are those involved in clashes. D. A zoomed in view of human thrombin (tan) with its substrate, protein C (magenta) (PDB

4D7Y). Distances between Proline40 of the substrate were identified with default contact parameters.

A model of CHO C1s was made using SWISS-model using the crystal structure of human C1s as a template (PDB 1ELV) (Fig 15B). This structure was used in the molecular replacement for the above crystal structure of human C1s with gigastasin. The homology model of CHO C1s is similarly structured to human C1s with minor differences. The hydrophobic box, consisting of Y617 and F533, are still conserved and the positions of the F533 and H482 are similar to the corresponding residues in human C1s, limiting the size of the residue in the P2 position. The S1 pocket is clearly visible and is bounded by D637 at the bottom and S638 at the side. The catalytic serine is present in approximately the same place, adjacent to the S1 pockets. A glutamine, Q535, now blocks the entrance to the P1 pocket instead of the lysine in human C1s. The C-terminal side of the scissile bond remain relatively unchanged.

A closer look into the interactions between the P2 residue of the substrate and surrounding residues of the protease explains the ability of thrombin, but not C1s to cleave residues with a proline in the P2 position. Mutating C64, the P2 residue of gigastasin, to proline shows potential clashing (Fig 15C). One of the two rotamer forms of proline that tilts towards F526 clashes with it and has a van der Waals overlap between 0.4 and 0.6 angstroms. A more severe clash would have a van der Waals overlap greater than 0.6 angstroms. This moderate overlap can sometimes be resolved by a smaller residue such as a serine in the P3 position to

accommodate larger residues in the P2 position as was observed with the Factor VIII R1689 cleavage site (Fig 6B and Table 5). When substrate protein C is complexed with thrombin, the proline, P40, in the P2 position does not clash with the surrounding residues and distances between the proline and the surrounding residues are greater than those in C1s (Fig 15D).

Discussion

Although CHO cells are widely used for the production of most protein therapeutics, limited proteolysis, or clipping, is a common problem in large scale production of recombinant proteins. Clipping can be cell line-specific or protein-specific, and there are multiple CHO cell lines with varying chromosomal structures and genome sequences that are still in the process of being characterized for protease activity (Wurm and Wurm, 2017). The assembly of the CHO cell genome in 2011 and the discovery of gene-editing technology has provided insights into the best ways to engineer CHO cells to optimize the expression of recombinant proteins (Xu et al., 2011). The zinc-finger matriptase knockout cell line demonstrates the ongoing need for protease-inactivated cell lines (Laux et al., 2018). A glutamine synthetase knockout cell line has been shown to enhance the clone selection process and increase productivity, while a transcriptomics approach to CHO cell engineering led to knockout of the repressors of STAT1, a transcription factor that elicits a cytokine response, that enhanced resistance of CHO cells to viral contamination (Laux et al., 2018; Fan et al., 2012; Stolfa et al., 2018; Chiang et al., 2019). Recent papers have identified Cathepsin D as a protease detrimental to the expression of recombinant proteins from CHO cells and used zinc-finger nucleases to knockout matriptase to create a protease-knockout cell line (Laux et al., 2018; Lim et al., 2018). A total of four CRISPR/Cas9 gene-edited cell lines have been developed by our lab for the production of recombinant proteins (Table 6). Here, we describe the identity of another protease (C1s) that results in the degradation of

another recombinant protein, specifically clade B, HIV Env proteins used as vaccine immunogens. The creation of the C1s$^{-/-}$ CHOK1 2.E7 cell line and the C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1 cell line, and the analysis of C1s substrate specificity has created a practical solution to solve the proteolysis problem in the large-scale production of clade B, HIV vaccine immunogens and other recombinant proteins.

Table 6. CRISPR-Cas9 gene edited CHO cell lines for the expression of recombinant proteins.

Cell line	**Journal Ref.**
MGAT1^{-} CHO-S	Byrne et al. (2018)
C1s$^{-/-}$ MGAT1^{-} CHO-S (stably expressing BaL-rgp120)	Li et al. (2019)
C1s$^{-/-}$ CHOK1 2.E7	shown in this paper
C1s$^{-/-}$ MGAT1^{-} CHOK1 1.A1	shown in this paper

These cell lines, except the C1s-/- MGAT1- CHO-S cell line, are available for transient or stable transfection of any recombinant protein. All cell lines were produced in our lab.

One of the existing solutions to prevent proteolysis of gp120s is mutating Arg315 of the GPGR motif to a Glutamine (Q). This mutation has been done for both gp120 monomers and SOSIP trimers found to undergo proteolysis, and highlights the applicability of the C1s-knockout cell lines (Sullivan et al., 2017; Wen et al., 2018). While the R315Q mutation prevents the elicitation of antibodies

dependent on arginine such as the neutralizing mAb 447-52D, it does not affect elicitation of antibodies to the V3 stem such as PGT128 or PGT121. At the same time, the R315Q mutation removes the GPGR motif, a highly-conserved motif in clade B strains and as a result, will prevent elicitation of potentially neutralizing antibodies. Lastly, changing the Env sequence may not be ideal in certain studies. For example, antibodies in human sera from elite neutralizers or controllers should be studied against the exact viral Env sequences. Otherwise, Envs to the Arg315 may be overlooked.

A number of groups are creating stable CHO cell lines expressing HIV envelope proteins for clinical trials. One such study hopes to evaluate different adjuvants with a multivalent vaccine using the RV144 subtype AE immunogen A244 and a subtype B immunogen, designated 6240 (Wen et al., 2018). Other groups are expressing Env trimers in CHO cells such as gp140 trimer nanoparticles and SOSIP trimers (Dey et al., 2018; Bale et al., 2018; Kong et al., 2016). In most cases, these vaccines were produced in normal CHO cells that incorporated complex, sialic acid-containing glycans that inhibited the binding of multiple, glycan-dependent bN-mAbs such as PG9, PGT121, and PGT128 (Raska et al., 2010; Byrne et al., 2018b). These cell lines contained C1s which would degrade the final protein product, particularly when expressed at large scale for long fermentation runs. With these cell lines deficient in MGAT1 and C1s, we have shown that gene editing can solve both the problems of inappropriate glycosylation

and limited proteolytic cleavage and that these cell lines can be utilized for production of HIV immunogens.

In these studies, we explored the differences between the substrate specificity of C1s and thrombin. Although thrombin has long been known to cleave rgp120 at the GPGR/AF sequence in the V3 region, it was interesting to discover that C1s also cleaved this sequence (Clements et al., 1991; Schulz et al., 1993). Although C1s did not cleave at the thrombin cleavage sites in the context of the FVIII protein, it was able to cleave two of three Factor VIII thrombin cleavage sites in the gp120 V3 domain. While C1s and thrombin share some similar substrate specificity, this demonstrates a conformational or accessibility requirement for C1s cleavage. These overlapping substrate specificities reflect the crosstalk observed between proteases and substrates in the coagulation and complement pathways (Markiewski et al., 2007). Proteases in the coagulation system, thrombin, coagulation factors XIa, Xa, and IXa, and plasmin were found to cleave C3 and C5, substrates of the complement system (Amara et al., 2010). Conversely, MASP1 and MASP2 of the lectin pathway in the complement system can cleave fibrinogen into fibrin and activate both Factor VIII and prothrombin (Conway, 2015). It is thought that coagulation factor proteases evolved from complement proteases in vertebrates and that there is an ancestral group of protein domains from which complement proteases of arthropods and vertebrates originate (Krem and Cera, 2002). Despite the fact that only thrombin can cleave substrates with bulkier proline in the P2 position of the substrate, this lineage shows that these serine proteases from

seemingly two very different systems are evolutionarily, structurally and functionally similar.

With the knowledge of the substrate specificity of C1s, future issues of proteolysis with recombinant proteins can be avoided. The prevalence of proteases in the regulation of physiological processes suggests there will be continued recombinant expression of proteases and protease substrates. Of the approximately 20,000 proteins in the human proteome, 570 are proteases, or about 2.6% of the proteome (Quesada et al., 2009; Ponomarenko et al., 2016). Proteases regulate physiological processes such as endocrine hormone maturation, the blood coagulation cascade, the complement pathway, and virus maturation. Dysfunctional proteases or genetic variants of substrates that become resistant to proteolysis can result in fatal disease including hemophilia, lupus, etc (Walport et al., 1998). The $C1s^{-/-}$ CHOK1 2.E7 and $C1s^{-/-}$ $MGAT1^{-}$ CHOK1 1.A1 cell lines described in this report can be used for the expression of all recombinant proteins susceptible to proteolysis by C1s. The finding that CHO cell-derived C1s only cleaves substrates with a small amino acid prior to the arginine in the P1 position will help predict C1s cleavage sites in other proteins. Data mining of the MEROPS database for protein substrates with the sequences Gly-Arg or Ala-Arg in exposed loop regions may reveal a number of proteins that cannot be expressed in the normal CHO cell line. This information will guide how these proteins are expressed whether it be through the use of the $C1s^{-/-}$ CHOK1 cell line, mutating away protease

cleavage sites as was done for the CHO production of MN-rgp120 or use of an alternate cell line (Wen et al., 2018).

In summary, we have shown that gene editing can solve the problems of proteolysis and glycosylation heterogeneity encountered in the expression of HIV vaccine immunogens. In principle, this same approach can be employed to eliminate proteases and other glycan modifying enzymes that interfere with the manufacturing and production of other recombinant proteins of pharmaceutical interest.

Supporting information

		A244		EN3_071	
Domain	HXB2	Length (aa)	PNGS	Length (aa)	PNGS
Signal	30-Jan	25	0	29	0
C1	31-130	91	0	101	2
V1	131-157	34	4	38	4
V2	158-196	41	2	37	2
V1/V2	131-196	75	6	75	6
C2	197-295	101	5	101	6
V3	296-331	33	1	33	1
C3	332-384	52	1	52	3
V4	385-418	28	4	38	5
C4	419-459	41	1	41	1
V5	460-469	11	2	11	1
C5	470-511	42	0	42	0
gp120	1-511	499	20	523	25

Supplemental Figure 4. Physical characteristic of A244, EN3 rgp120s used in this study.

The length in amino acids (AA) and number of potential N-linked glycosylation sites (PNGS) are listed for each of the HIV Env domains for the expressed Envs: A244, EN3_071.

Chapter 3. A monoclonal antibody, 10C10, elicited by A244-gp120 and recognizes a conserved epitope in the V2 domain that correlated with protection in the RV144 HIV vaccine trial

Introduction

A major focus of HIV vaccine research is the identification of epitopes recognized by antibodies that confer protection from HIV. To date, the RV144 HIV vaccine trial has been the only study to show that immunization can prevent HIV infection (Rerks-Ngarm et al., 2009). Protection did not correlate with the presence of neutralizing antibodies or anti-viral cytotoxic lymphocyte mediated immune responses. Rather, protection correlated with cross-clade non-neutralizing antibodies to the V1/V2 domain of the HIV envelope protein (Haynes et al., 2012). More specifically, in the RV144 trial with AIDSVAX B/E immunogens MN and A244, antibodies to K169 were associated with protection. Thus far, V2 non-neutralizing antibodies continue to be the only correlate of protection as shown in recent SIV studies (Barouch et al., 2012; Roederer et al., 2014; Vaccari et al., 2016).

Two main groups of antibodies to the V2 region have been observed. The first group of antibodies are broadly neutralizing monoclonal antibodies (bN-mAbs) to the V1/V2 as represented by PG9-like antibodies. These antibodies are neutralizing, broadly cross-reactive and glycosylation-dependent (Walker et al., 2009; McLellan et al., 2011; Pancera et al., 2013). They exhibit long complementary determining regions (CDR) H3 loops that bind with high-affinity

to glycans and recognize a four-beta strand scaffold structure. The second group of antibodies are non-neutralizing antibodies to the V2 loop as represented by CH58-like antibodies. Conversely, the main mechanism of action for these antibodies is antibody dependent cellular cytotoxicity (ADCC) (Liao et al., 2013; van Eeden et al., 2018). They are less cross-reactive than bN-mAbs, glycosylation-independent and recognize linear peptide sequences.

In this study, we describe a highly cross-reactive, non-neutralizing monoclonal antibody called 10C10 that was elicited in a mouse immunized with A244 gp120, a major component of AIDSVAX B/E vaccine used in the RV144 clinical trial (Rerks-Ngarm et al., 2009). 10C10 cross-reacted with envelope proteins from three different clades, depended on K169 for binding, and bound to the A244 V2 158-178 peptide in an α-helical structure. Furthermore, we use computational tools to model a structure for the complex of 10C10 Fv with the A244 V2 158-178 peptide.

Materials and Methods

Recombinant proteins and synthetic peptides

rgp120s from the A244, MN, Bal, CN97001, ZM651, TZ97008, CN98005, ZA97012, IN98025, IN98026, TZ97005, ZA97010, and ZM233 isolates were expressed by transient expression in HEK293 (ATCC) cells and purified by immunoaffinty chromatography as described previously (Smith et al., 2010). Synthetic peptides were synthesized by GenScript Inc (Piscataway NJ, USA).

Antibodies

The 10C10 monoclonal antibody was isolated from a mouse immunized with A244 rgp120 by standard techniques based on the method of Milstein and Kohler (Köhler and Milstein, 1976). Briefly, Balb/c mice (Jackson Laboratory, ME USA) were primed with 20ug A244 rgp120 (HEK293) in Sigma Adjuvant System (SAS) injected intraperitoneal (IP) and subcutaneous (SC) routes, and similarly boosted at days 14, 34, 63 and 128 days. Spleens were harvested for fusion at day 130. The CH58 mAb was kindly provided by Barton Haynes and Larry Liao of the Duke University Center for HIV/ AIDS Vaccine Innovation (CHAVI), (Durham, N. Carolina). The VRC01 bN-mAb was provided by the NIH AIDs Reagent program (Germantown, MA) and the PG9 broad neutralizing monoclonal antibody (bN-mAb) was purchased from Polymun Scientific GmbH (Vienna, Austria). Purified anti-human Fc and anti-mouse Fc polyclonal antibodies coupled to Alexa-488 dye were from Life Technologies (Calabasas, CA). Serum from individuals who

participated in the RV144 clinical trial were kindly provided by Nelson Michael and Jerome Kim of the Military HIV Vaccine Research Program, (Walter Reed Army Institute of Research, Rockville, MD) and Peter Gilbert (Statistical Data Management Center of the HIV Vaccine Trials Network; University of Washington, Seattle, WA) in connection with the RV144 correlates of protection group. Serum were pre-screened for binding to synthetic peptides from the V2 domain by ELISA.

Fluorescent ImmunoAssay

For antibody binding to synthetic peptides, Grenier Bio-One Fluortrac 600 microtiter plates (Kremsmunster, Austria) were coated with 2ug/mL of peptide overnight in PBS at 4°C. Plates were blocked in PBS + 2.5% BSA, then washed 4 times with PBS containing 0.05% Tween-20. Serial dilutions of CH58 or 10C10 were added in a range from 20ug/mL to 0.0001ug/mL and incubated. For competition studies, the competing antibody was added at the indicated concentration to the diluent used for the test antibody. After incubation and washing, fluorescently conjugated anti-Hu or anti-Mu (Invitrogen, Carlsbad, CA) was added at a 1:3000 dilution. Plates were incubated for 90 minutes with shaking, then washed as before. Plates were imaged in the Perkin Elmer Envision plate spectrophotometer (Waltham, MA) at excitation 395nm and emission 490nm. All steps except coating were carried out at room temperature on a shaking platform for 90 minutes. All dilutions were done in blocking buffer.

Biacore

Surface plasmon resonance measurements were carried out on a Biacore X100+ instrument (GE, Chicago, IL). Experiments to determine kinetic constants for antibody affinity were conducted at 25°C at a flow rate of 10ul/min using a reference subtraction. Kinetic measurements were carried out by using anti-human or anti-mouse Fc capture. Capture surfaces were prepared on CM5 chip. Anti-Fc antibodies were immobilized at 9000 RU and were used to capture CH58 or 10C10 at 70-110 RU. The biosensor surface was regenerated with a 30s pulse of 10mM Glycine pH 2.5 between injections. The running buffer and diluent was HBS-EP+. Control IgG bound to the reference surface as a background reference. Buffer blanks were injected over the sensor surface in triplicate. Double referencing protocols were employed to consistency across the injection series. All Biacore sensor grams were processed using Biacore Evaluation software to determine ka, kd, and kD from processed data sets. Data from all flow cells were globally fit to a 1:1 Langmuir binding model that included a mass transport term. Chi2 values less than 1 and Rmax <100 were established as quality cutoffs for acceptable data.

Method for Computation modeling of 10C10

ABodyBuilder was used to model the 10C10 Fv, using the 10C10 Fv sequence as input. It uses the Structural Antibody database primarily composed of human and mouse antibody structures to model antibody structures (Leem et al., 2016). PEP-

FOLD3, which computationally generates *de novo* structures of peptides using a probabilistic method, was used to model the peptide (Lamiable et al., 2016). Parapred, a paratope prediction program, was used to predict the CDR residues that would be part of the paratope and interact with the peptide (Liberis et al., 2018). Haddock 2.4 was used to dock the 10C10 Fv and V2 peptide (Dominguez et al., 2003). Residues identified as part of the paratope by Parapred and scored at least a 0.67 out of 1 were added as "active" residues, or sites of interaction to constrain docking. The CDR regions plus two residues before and after the CDR and the whole peptide were made fully-flexible. Three distances were measured between the heavy chain and light chain and added as restraints to deter separation of the heavy and light chain during docking. Additional parameters for docking: Random exclusion of ambiguous restraints was turned off, number of structures of rigid body docking (5000), number of structures for semi-flexible refinement (400), number of structures of water refinement (400), automatically define backbone dihedral restraints from structure (only for alpha helix), number of MD steps during first rigid body cooling stage (1000), number of MD steps during first rigid body cooling stage (1000), number of MD steps during second cooling stage with flexible sidechains at interface (2000), number of MD steps during third cooling stage with fully flexible interface (1000), number of structures to analyze (200) (Ambrosetti et al., 2020). Please also refer to the Bonvin Lab tutorial, "HADDOCKing of the p53 N-terminal peptide to MDM2". All other options were set to default.

Virus Neutralization Assay

Virus neutralization studies were carried out using the TZM-bl assay described by Montefiori (Montefiori et al., 2012). A plasmid for the expression of pseudovirions from the TH023 isolate of HIV-1 was kindly provided by Drs. Victoria Polonis and Jerome Kim of the Military HIV Research Program (Walter Reed Army Institute of Research, Silver Spring, MD).

Statistical Methods

Statistically significant differences in the binding of antibodies in sera from the RV144 trial to the peptides recognized and not by the 10C10 mAb were determined using the Mann-Whitney non-parametric test and applying a two tailed P value (GraphPad Software, Prism 7).

Results

Isolation of 10C10, a non-neutralizing mouse monoclonal antibody

Mice were immunized at day 14, 34, and 175 with A244 rgp120 expressed in HEK293 cells. 10C10 came from a stable hybridoma cell line and showed reactivity to full-length A244 rgp120 and A244 V1/V2 fragments. Because 10C10 was an antibody with specificity to the V2 region, we compared it to PG9 and CH58, two human monoclonal antibodies to the V2 region, in a pseudovirus neutralization assay with the clade AE TH023 and clade C MW965 viruses. PG9 is a broadly neutralizing monoclonal antibody (bN-mAb) that can neutralize tier 2 viruses and recognizes two glycans and a strand within the V1/V2 loops (Walker et al., 2009; McLellan et al., 2011). CH58 neutralizes some tier 1 strains including AE. TH023 and C.SHIV1157ipEL-p, but not B.MN. CH58 was also shown to have a linear epitope consisting of a V2 peptide from residues 169-182 (Liao et al., 2013). Similar to CH58, 10C10 showed low neutralization against viruses TH023 and MW965 (Fig. 16), while PG9 demonstrated strong neutralization against TH023 and MW965 as expected for a bN-mAb.

	TH023	MW965
10C10	8.3	23.03
CH58	10.9	>88
PG9	0.05	0.25

Figure 16. Pseudovirus neutralization by 10C10 as compared with CH58 and PG9.

Neutralization of tier 1 pseudoviruses, TH023 and MS965, by 10C10, CH58 and PG9 was tested. IC50 values reported in ug/mL antibody.

Using the Pierce Rapid Isotyping strip test for mouse IgG1, IgG2a and IgG2b antibodies, 10C10 was found to be isotype IgG2b. The mouse IgG2b isotype, along with IgG2a and IgG3, is one of the predominant isotypes involved in ADCC and mediates ADCC through interaction with the mouse FcγRIV receptor (Nimmerjahn et al., 2010). The mouse FcγRIV receptor shares a 63% sequence identity to the human FcγRIII receptor, which has been shown to be indispensable for ADCC by natural killer cells (Mechetina et al., 2002; Yeap et al., 2016). Thus, based on the low neutralizing ability and IgG2b isotype of 10C10, we verified that 10C10 is a non-neutralizing or low-neutralizing antibody.

10C10 has cross-clade activity for clades CRF01_AE, B and C rgp120s

We determined the cross-clade activity of 10C10 by observing its recognition of rgp120s from three different clades: clade AE, B and C. A244 rgp120 from the RV144 trial is a clade AE Env (Fig. 17). The lab-adapted BaL and MN

gp120 are clade B Envs, and CN97001 and ZM651 are clade C Envs. 10C10 bound well to clade AE A244 rgp120 and both clade C Envs, CN97001 and ZM651. It bound moderately well to the clade B Envs, BaL and MN. Because 10C10 recognized the clade C Envs as well as it did to A244 rgp120, we also tested binding of 10C10 to additional clade C isolates to further study this cross-reactivity. 10C10 binding was measured against a larger panel of eight clade C isolates: one Chinese isolate, CN98005; two Indian isolates, IN98025 and IN98026; two Tanzanian isolates, TZ97005 and TZ97008; two South African isolates, ZA97020 and ZA97012; and one Zambian isolate, ZM233. Most of these isolates exhibited intermediate to high binding affinity to 10C10, while only one, the ZA97013 isolate, exhibited no specific binding activity.

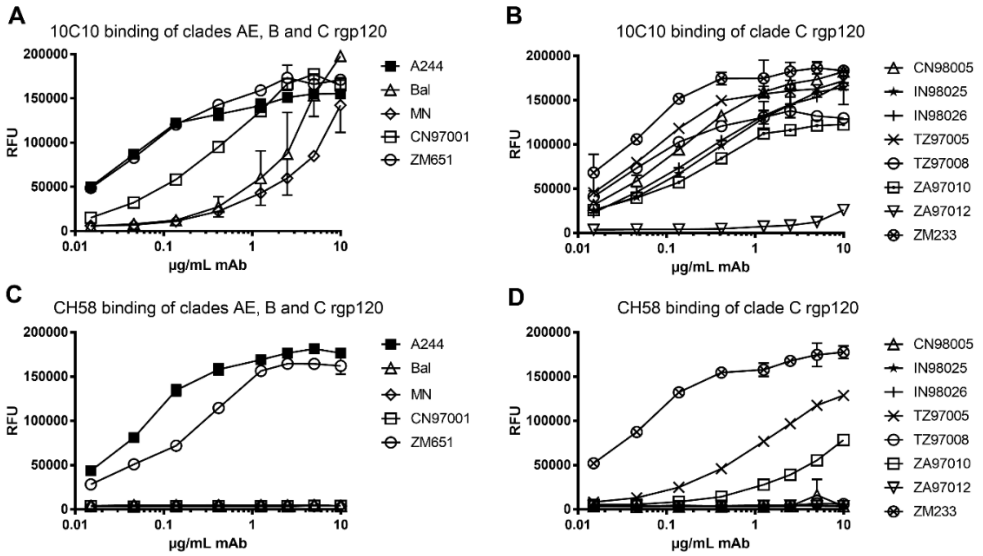


Figure 17. Binding of 10C10 and CH58 to Clade AE, B and C rgp120s.

A. Capture fluorescence immunoassay (FIA) was used to measure binding of 10C10 to clade AE, B and C rgp120s. rgp120s were captured with 34.1, a mouse monoclonal antibody to the gD tag. Alexa Fluor A488 conjugated to an α-mouse antibody was used for detection. B. Binding of 10C10 to eight additional clade C rgp120s. C. As a comparison, binding of CH58 to the same clades AE, B and C rgp120s was measured. D. Binding of CH58 to the eight additional clade C rgp120s.

We then compared binding of the same rgp120 immunogens to CH58, another V2 peptide binding antibody. CH58 is human non-neutralizing monoclonal antibody that binds to the V2 loop in a helix-loop conformation (Liao et al., 2013). CH58 did not show the same cross-clade activity as 10C10. CH58 bound well to the clade AE A244 and two of the clade C Envs, ZM651 and ZM233. It only bound moderately to two other clade C Envs, YZ97005 and ZA97010, and did not bind to either of the two clade B Envs or the other six clade C Envs. Overall, 10C10 exhibited greater cross-reactivity to clade AE, B and C rgp120s than CH58.

10C10 epitope consists of a V2 peptide

To determine the epitope of 10C10, we first aligned V2 sequences of the rgp120s that were tested for binding to 10C10 (Fig. 18). Lower binding of 10C10 to rgp120 seemed to be correlated to differences at residues M161, K169, and V172. Given the location of these mutations, we tested binding of 10C10 to V2 peptides to identify the minimal V2 peptide required for binding. 14-22 mer peptides spanning residues 158-189 were screened for binding to 10C10 (Fig. 19A). The minimal peptide required for binding was A244 V2 158-178. This peptide consists of the second half of the B strand, the B/C junction, and the C strand (Fig. 20A). The truncated A244 V2 160-174 did not bind, which suggested that residues S158, F159 or L175-F176-Y177-K178 were essential for binding. Peptides further downstream in the V2 region and spanning the CH58 epitope including peptides 171-188, C/D junction 173-194 and V2 178-192 did not bind. Furthermore, binding of 10C10 to a linear peptide without glycans indicated that 10C10 would be similar to the glycan-independent CH58. PG9 has been shown to bind only to glycan-containing scaffolds or peptides (McLellan et al., 2011; Alam et al., 2013; Amin et al., 2013).

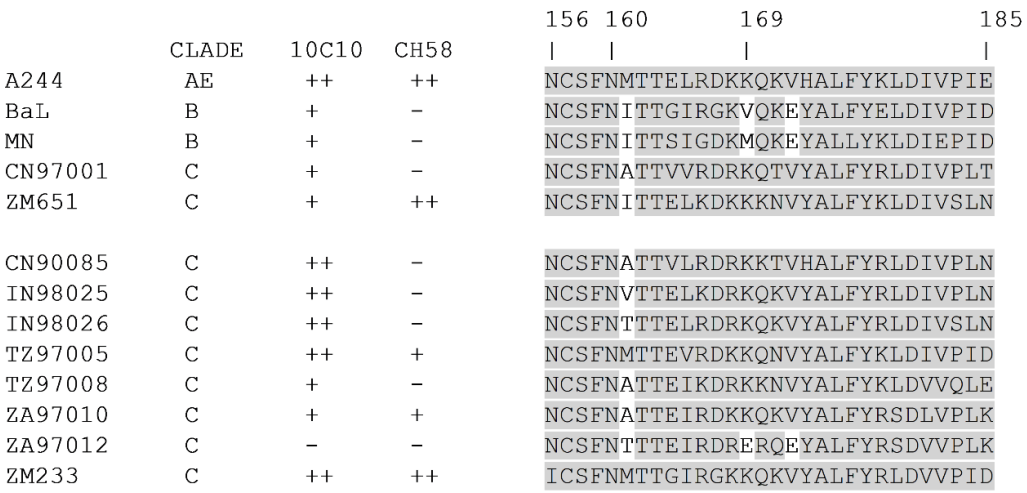


Figure 18. Sequence alignment of V2 sequences from clade AE, B and C rgp120s.

V2 sequences from HXB2 156 to 185 are aligned for the clade AE, B and C rgp120s that were tested for binding against 10C10. Strong (++), intermediate (+), and low (-) binding of 10C10 and CH58 to the different rgp120s is listed. Residues in white differ from the corresponding residue in that position for A244.

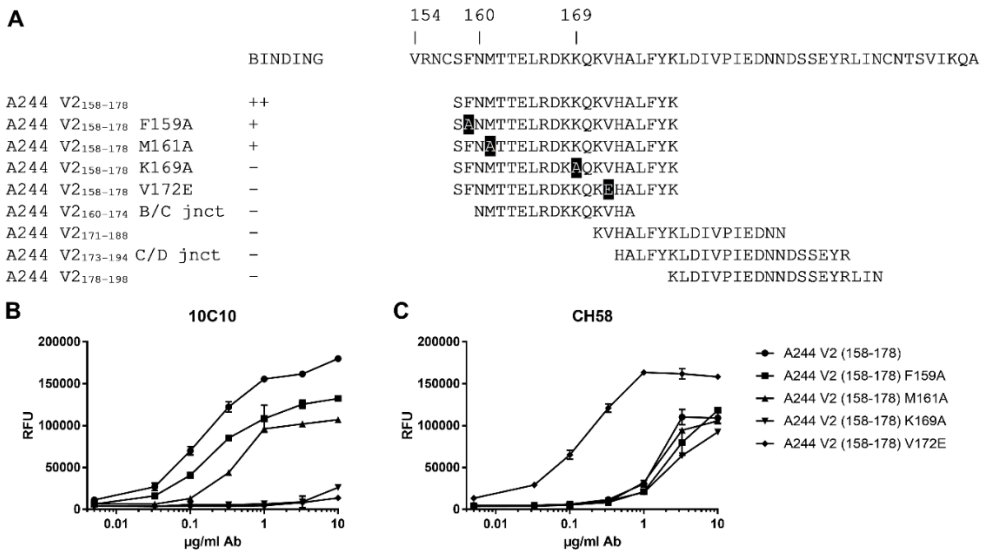


Figure 19. Identification of residues involved in the epitope of 10C10.

A. Shown are peptides used in peptide binding studies by direct FIA with 10C10. Peptides are aligned with the A244 V2 sequence shown at the top. Strong (++), intermediate (+), and low (-) binding of 10C10 to the peptides is listed. F159A, M161A, K169A and V172E peptides were used to identify specific residues that interact with 10C10. The peptides spanning residues 160 to 198 were used to identify the minimal peptide required for 10C10 binding. B, C. Binding curves of 10C10 and CH58 to A244 V2 158-178 peptide and mutated peptides F159A, M161A, K169A and V172E.

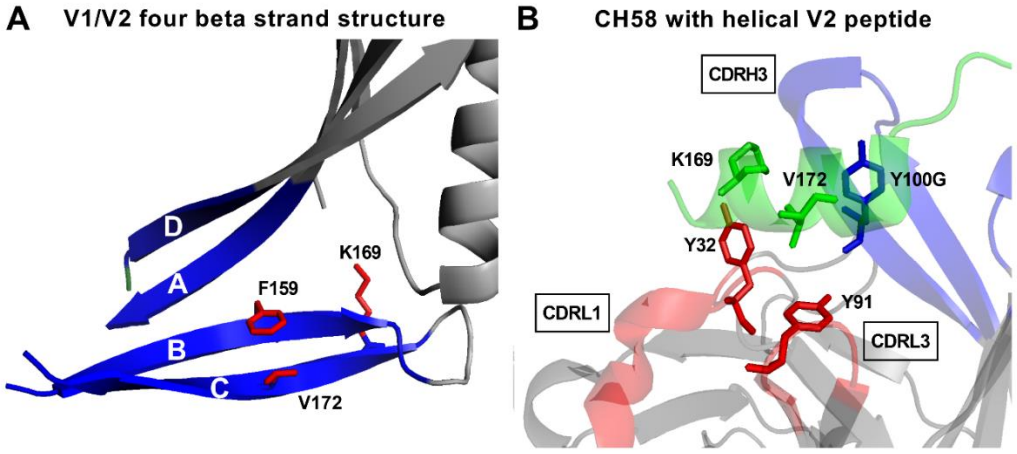


Figure 20. V1/V2 structures

A. The four strand V1/V2 loop region was visualized in Pymol. Strands A, B, C and D are colored blue and labeled in white. The figure is adapted from the structure of the V1/V2 scaffold of ZM109 with the PG9 Fab (McLellan et al., 2011). B. Shown is the α-helical A244 V2 peptide 165-186 from the structure of the peptide with the CH58 Fab (Liao et al., 2013). The V2 peptide is in green. The heavy chain and light chain CDRs are in blue and red, respectively.

We then performed mutagenesis of F159, M161, K169, and V172 on the A244 V2 peptide 158-178 (Fig. 19A, 19B). Given the possibility that the V2 peptide could form a four stranded β-sheet domain, we thought F159 on the B strand could be involved in strong hydrophobic interaction with the C strand (Fig. 20A). However, the mutation F159A only resulted in a small reduction in binding. We also observed that 10C10 was able to bind to the peptide 158-178 but could not bind the shorter peptide 160-174. Since F159 did not seem to play a major role in the interaction between 10C10 the A244 V2 peptide 158-178, this indicated one or multiple residues within residues 175-178 (LFYK) were an essential part of the epitope.

Based on the sequence alignment done in Fig. 3 and the differences in sequence observed at residues M161, K169 and V172, we expected that these three residues were important for 10C10 binding. M161 and K169 were mutated to alanine. V172 was mutated to glutamic acid, the commonly observed residue associated with loss of binding as seen in strains BaL, MN and ZA97012 (Fig. 18). While M161A resulted in a moderate loss of binding, mutations K169A and V172E completely abolished binding.

K169 and V172 are known to play important roles in antibody-antigen recognition. K169 is a highly conserved residue of the C-strand that is recognized by both neutralizing such as PG9 and CAP256 and non-neutralizing antibodies such as CH58 and is shown to be a site of immune pressure (McLellan et al., 2011; Moore et al., 2011; Rolland et al., 2012; Liao et al., 2013). In the structure of CH58 in complex with α-helical V2 peptide, V172 makes crucial hydrophobic interactions with three tyrosines in CDRH3, CDRL1, and CDRL3 (Fig. 20B). In sum, we identified A244 V2 158-178 as the minimal peptide important for binding and determined residues K169, V172 and residues 167-170 to be absolutely necessary for binding.

Binding to CH58 decreased after mutagenesis of F159, M161, K169 and V172 (Fig. 19C). All mutations resulted in a significant drop in binding with

K169A having the most detrimental effect. Thus, 10C10 and CH58 were dependent on some of the same residues, indicating overlapping epitopes. However, the K169A mutation did not completely abolish binding of CH58 in our experiments as was observed for CH58 in other studies (Liao et al., 2013). A decrease in binding of CH58 for the F159 and M161 peptides was unexpected, as CH58 was shown to recognize the peptide 165-186 which does not contain residues F159 and M161. CH58 may have non-specific binding for the N-terminus of the A244 V2 158-178 peptide which could explain the unexpected decrease in binding with F159 and M161A peptides as well as the partial binding observed with peptide K169A.

Competitive binding of 10C10 to CH58, PG9, and VRC01

We next carried out studies to determine whether the 10C10 antibody had similar epitopes to other functionally significant antibodies to the V1/V2 domain by competitive binding (Fig. 21A). The non-neutralizing V2 monoclonal antibody CH58 and bN-mAbs PG9 and VRC01 were assayed for binding to A244 rgp120 expressed in the GNTI$^-$ HEK293 cell line. High-mannose glycans on rgp120 were required in order to observe differences in PG9 binding (Morales et al., 2014; Doran et al., 2018a). First, we measured the ability of 10C10 to inhibit the binding of CH58. 10C10 effectively inhibited the binding of CH58 to A244 rgp120, confirming the epitope of 10C10 overlaps with the epitope of CH58. We showed with the previous experiment that 10C10 binds to residues 158-178 and Liao et al.,

showed that CH58 binds to residues 165-186. Both these antibodies targeted the residues at and around K169.

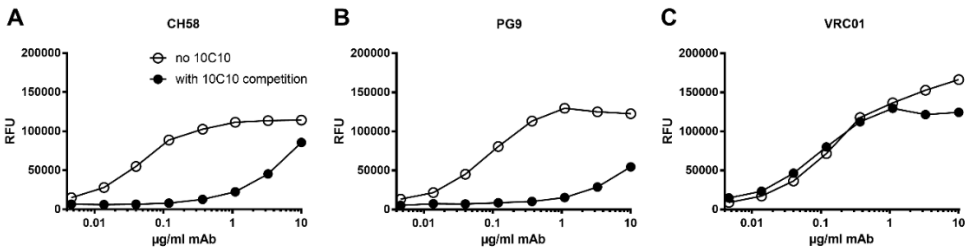


Figure 21. Competition of 10C10 with V1/V2 mAbs and bN-mAbs

A. Binding of CH58 to A244 rgp120 was measured by FIA. A244 rgp120 was expressed in the GNTI$^-$ HEK 293 cell line. 20 ug/ml 10C10 was added to compete for CH58 binding. B, C. 10C10 was used to compete with PG9 and VRC01 for binding to A244 rgp120.

Next, we competed PG9 with 10C10. PG9 is a broadly neutralizing monoclonal antibody (bN-mAb) that depends on K169 and two glycans at residues 160 and either 156 or 173 (Fig. 21B). We found that 10C10 was effective at competing with PG9 for binding to A244 rgp120 expressed in the GNTI$^-$ cell line. GNTI$^-$ A244 rgp120 was used since binding of PG9 depends on high mannose glycans at 156 and 160 (Doores et al., 2010; Amin et al., 2013; Doran et al., 2018a). As a negative control, 10C10 was used to compete with VRC01 for binding for A244 rgp120 (Fig. 21C). VRC01 is a bN-mAb that targets the CD4 binding site (Wu et al., 2010; Li et al., 2011). 10C10 did not affect binding as was expected since these antibodies do not target the same epitope.

We also carried out surface plasmon resonance studies to characterize the kinetics of 10C10 binding in comparison with the CH58 mAb. In these studies, we measured the binding of both mAbs to A244 gp120 (Supp. Fig. 5). We found that 10C10 bound with an kD of approximately 1.4×10^{-8} M whereas the binding to CH58 was somewhat weaker but with a similar magnitude at a kD of 6.2×10^{-8} M.

Modeling of 10C10 Fv and A244 V2 158-178 peptide

A structure of the 10C10 Fv and A244 V2 158-178 peptide was created by modeling the Fv and peptide separately, and then docking them together. The Fv of 10C10 was sequenced using an reverse transcriptase – polymerase chain reaction (RT-PCR) method with universal primers for the antibody constant regions followed by a template switch primer to generate cDNA (Meyer et al., 2019). ABodyBuilder selects templates for modeling based off antibody sequence alignment and uses the corresponding antibody structures to construct an antibody model (Leem et al., 2016). The framework regions were modeled to within 0.8 A with at least 75% confidence (Supp. Fig. 6A, 6B). CDRH1, CDRH2, CDRL1, CDRL2 and CDRL3 and the three residues before and after each CDR were modeled to within 1.7 A with at least 75% confidence. CDRH3 was modeled with limited accuracy; one residue within the CDRH3 and the two residues before it could not be modeled accurately. The residues following the CDRH3 were modeled accurately. Because the CDRH3 was short, it would be unlikely to play a major role

in the recognition of the V2 peptide. We allowed the ambiguous CDRH3 region to be resolved by refinement during the docking process with the V2 peptide.

To model the V2 peptide, we used *de novo* peptide structure prediction (Lamiable et al., 2016). The V2 peptide was predicted to have an α-helical structure from residues M161 to Y178 (Supp. Fig. 6B, 6C). The V2 peptide exhibited an alpha-helical structure of 4 turns that was longer than the 3 turns observed for the V2 peptide 165-186 peptide with CH58.

Known sites of interaction, or "active" residues were used to guide docking of the V2 peptide to the modeled Fv structure (Fig. 22) (Dominguez et al., 2003). Residues of the paratope were predicted using a machine learning algorithm generated based off the Structural Antibody Database and made the active residues of 10C10 (Supp. Fig. 7) (Liberis et al., 2018). Residues indicated as active on the peptide were K169 and V172 based off the mutagenesis experiment. H173 was also made an active residue because it is often involved in salt bridge interactions (Liao et al., 2013; van Eeden et al., 2018).

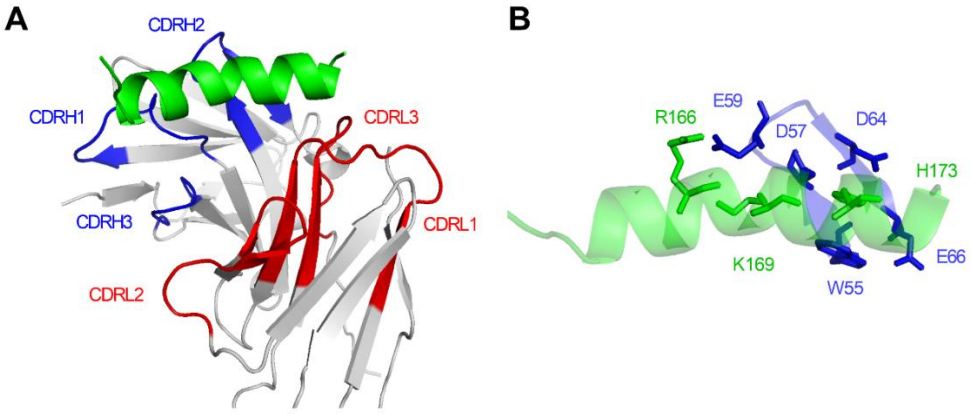


Figure 22. Modeled 10C10 Fv with V2 peptide 158-178.

A. The complex of the 10C10 Fv and V2 peptide 158-178 was docked using Haddock and visualized using Pymol. The V2 peptide is in green. The heavy chain and light chain CDR regions are in blue and red, respectively. B. A zoomed-in view interactions taking place between the cationic V2 peptide and anionic CDRH2.

From the model of 10C10 Fv with the V2 peptide, we observed that CDRH2 is the primary CDR loop of 10C10 that makes multiple contacts with the V2 peptide. This differs from CH58 and CAP228 in that for these antibodies, the CDRH3 makes most of the contacts with the V2 peptide while the CDRL2 containing the ED or DDxD motifs made additional contacts (Liao et al., 2013; van Eeden et al., 2018). The CDRH2 loop residues W55, D57, D64 and E66 are involved in hydrophobic and salt bridge interactions with R166, K169 and H173. These positively charged residues cluster along one side of the peptide towards the CDRH2. CDRH1 Y37 and Y38 along with CDRL2 Y166 hold the N-terminal of the peptide in place while the CDRL3 F242 grasps the C-terminus.

RV144 sera binding to A244 V2 158-178 peptide

To assess the relevance of the epitope recognized by the 10C10 mAb to the immune response directed towards the V1/V2 domain in humans, we measured binding of serum from fifty-seven RV144 vaccinated individuals to the A244 V2 158-178 peptide and to the same peptide with the K169A mutation (Fig. 23A). Antibodies elicited by the RV144 immunogens, A244 and MN rgp120, were able to bind to both peptides. This suggested that serum from RV144 vaccinees also contained 10C10 antibodies that could recognize the 158-178 epitope. Additionally, the data showed that serum antibodies could not bind as well when the lysine was not present at position 169. To better observe the differences between these two groups, binding of serum to the V2 wildtype peptide and the V2 K169A peptide was plotted for each participant (Fig. 23B). The difference observed between binding to the V2 wildtype peptide and V2 K169A peptide was also plotted (Fig. 8C). Two-thirds of individuals had serum that bound better to the V2 wildtype peptide than to the V2 K169A peptide. This could be interpreted as there being higher titers of antibodies that recognize K169 than those that recognize A169. Alternatively, antibodies that recognize K169 may be able to cross-react with A169, leading to higher levels of binding to the peptide with K169.

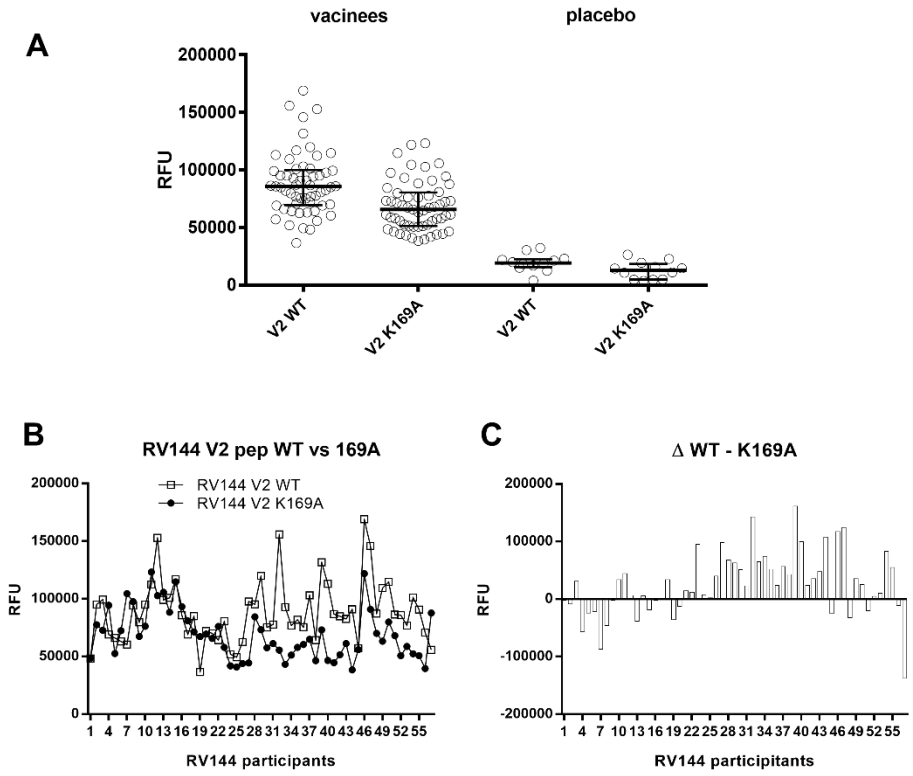


Figure 23. Recognition of RV144 serum antibodies to K169.

A. The serum from individuals of the RV144 trial were tested for binding to either the A244 V2 peptide 158-178 (V2 WT) or the A244 V2 peptide 158-178 with the K169A mutation (V2 K169A). Serum from the fifty-seven vaccinated RV144 participants (RV144) and serum from twelve placebo-given RV144 participants (placebo) were tested by FIA in duplicate. Median and interquartile range are plotted. B. For each of the fifty-seven vaccinated individuals, binding of serum to the V2 wildtype peptide and to the V2 K169A peptide is plotted. C. For each of the fifty-seven individuals, the subtracted difference between binding of serum to the V2 wildtype peptide and the V2 K169A peptide is plotted.

Discussion

Previously identified non-neutralizing V2 antibodies, CH58 and CAP228, were isolated from an RV144 vaccinee and an HIV-infected individual from the CAPRISA cohort, respectively. Here, we show that the RV144 immunogen A244-rgp120 alone contains the epitope that is able to elicit a cross-clade, non-neutralizing V2 antibody that recognizes a helical V2 peptide spanning residues 158-178. The footprint of 10C10 overlaps epitopes of bN-mAbs such as PG9 and non-neutralizing antibodies such as CH58 and is centered around K169. This residue was also highlighted as one of several in the V2 domain that are subject to immune escape in the RV144 trial (Rolland et al., 2012). All of the above indicates that 10C10 is representative of the antibodies directed toward a region that correlated with protection in the RV144 trial.

The V2 peptide 158-178 has a strong helical propensity that promotes the grouping of three positively charged amino acids (R166, K169, H173) along one side of the peptide. These cations can form several strong salt bridge interactions with the 10C10 CDRH2. The clustering of several salt bridge interactions may contribute to the cross-reactivity of 10C10, given that the cationic region from 166 to 171 is well-conserved across clades (van Eeden et al., 2018; Duerr and Gorny, 2019). Additionally, the helical conformation of the V2 loop is thought to be present when the V1/V2 loops are displaced in the open conformation of HIV Env (Ozorowski et al., 2017; Wang et al., 2018; Wibmer et al., 2018; Duerr and Gorny, 2019). The strong helical propensity of the V2 peptide 158-178 may stabilize this

open conformation of the V1/V2 loop. The open conformation of V1/V2 loops allows for exposure of residues with less restriction from glycans. V1/V2 residues that are shielded by glycans while in the closed conformation would become "unshielded", which may promote development of antibodies against V2 linear peptide sequences.

Several antibody structures in complex with the V1/V2 domain have been determined and strongly suggest the presence of two discrete V2 structures. The crystal structures of the V2 domain with 10C10, CH58, CH59, or with CAP226-16H show the V2 loop can retain an alpha-helical structure, while structure of the V1V2 domain with antibodies such as PG9 or 830A show that the V2 loop can instead be part of a beta-barrel structure (Liao et al., 2013; McLellan et al., 2011; Pan et al., 2015; Wibmer et al., 2018). The range in antibodies elicited by the two, or possibly more, conformations of the V2 domain illustrates the complexity of the adaptive immune system in response to HIV.

With the 10C10 structure, we have a better understanding of the structural basis for the inability of 10C10 to bind certain strains. 10C10 could not bind ZA9798 due to the K169E and V172E variants. The replacements of a positively charged lysine and hydrophobic valine with two negatively charged glutamic acids dramatically affects the charge distribution around these residues. The glutamic acids would repel the negatively charged 10C10 CDRH2, thereby disrupting the ionic interactions necessary to mediate 10C10 binding to the V2 loop. We also observed that for the BaL and MN strains when only one glutamic acid is

introduced, binding is also decreased. This highlights the need for a strongly cationic V2 region to elicit antibodies with anionic CDR loops.

Mouse antibodies and human antibodies differ in combinatorial germline gene diversity and junctional diversity (Zemlin et al., 2003; Shi et al., 2014). Differences are most dramatic for the CDRH3 region. Human antibodies have CDRH3 in the range of 1-35 residues and a mean of 15.2, while mouse antibodies have CDRH3 in the range of 1-21 residues with a mean of 11.5 when using IMGT-defined lengths (Zemlin et al., 2003). The 10C10 CDRH3 had a length of three amino acids, shorter than most human CDRH3 lengths and in the lower end for mouse CDRH3 lengths as well. Including the two residues flanking it, the 10C10 CDRH2 contains four negatively charged amino acids: two glutamic acids and two aspartic acids. This negatively charged loop recapitulated the negatively charged amino acids present in the long CDRH3 loops and the ED or DDxD motifs seen on the human CDRL2 loops of CH58 and CAP228 (Liao et al., 2013; Wiehe et al., 2014; van Eeden et al., 2018). Anionic, aspartic acid-rich CDRs favor the cationic, lysine-rich region from residues 166 to 171 of the HIV Env, regardless of the animal species that antibodies originate from.

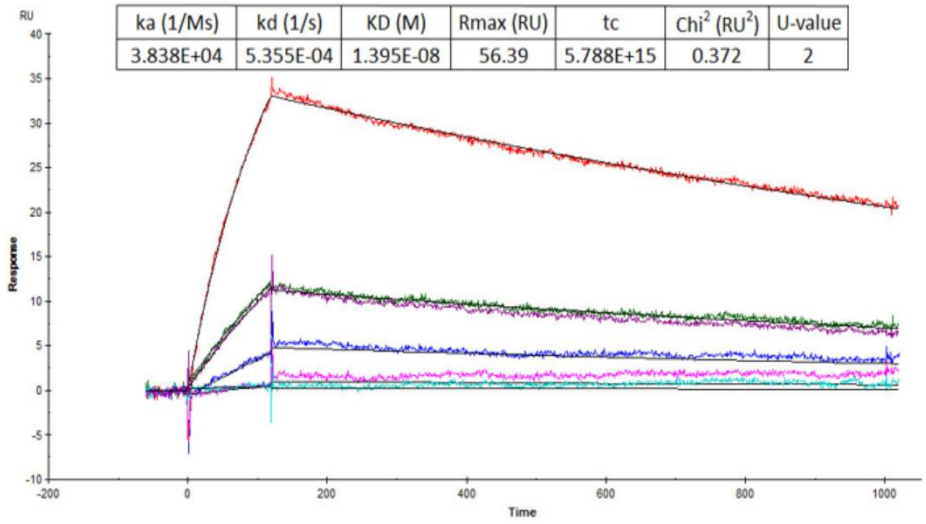
The specificity, magnitude and breadth of the vaccine-elicited V2 antibodies continue to be part of the criteria required to demonstrate improvement over the RV144 vaccine. V1/V2 antibody responses to the HVTN 097 and HVTN 100 pox-virus prime-protein boost vaccines were compared to each other as well as to the response to the RV144 vaccine (Shen et al., 2020). Ad26 vectors + gp120

and Ad26 mosaic vectors + gp140, similar to those being currently being tested in HVTN 705, were also for their ability to generate V2 antibodies and whether there was a correlate of protection (Vaccari et al., 2016; Barouch et al., 2018). Our increased understanding of the epitopes and structural basis of binding for non-neutralizing antibodies to the V2 loop will support the ongoing vaccine development.

Supporting Information

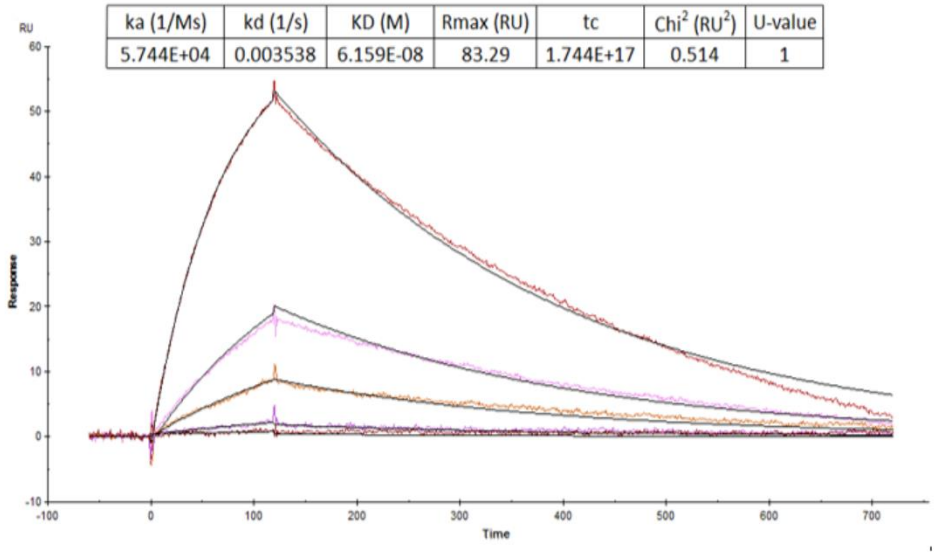
A **10C10 binding to A244 rgp120**

ka (1/Ms)	kd (1/s)	KD (M)	Rmax (RU)	tc	Chi^2 (RU^2)	U-value
3.838E+04	5.355E-04	1.395E-08	56.39	5.788E+15	0.372	2



B **CH58 binding to A244 rgp120**

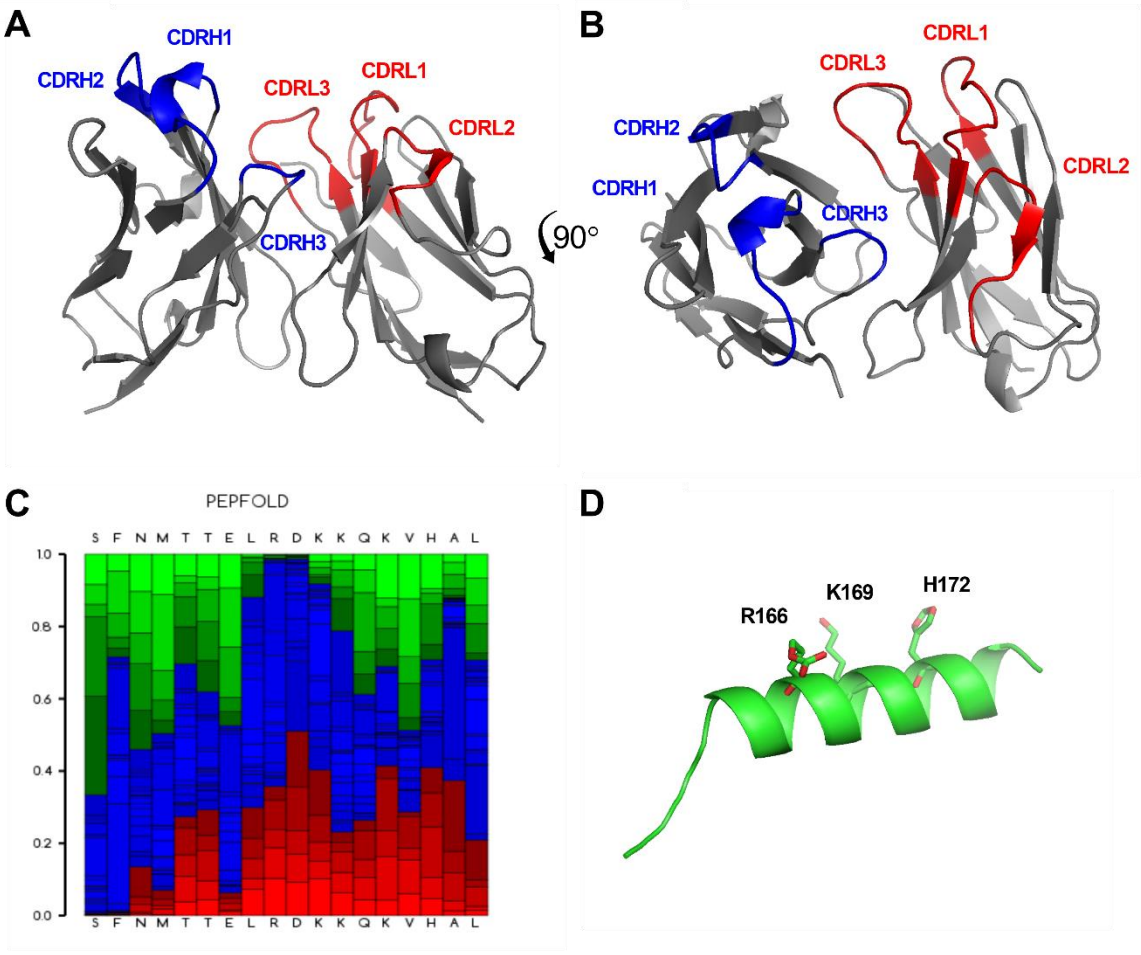
ka (1/Ms)	kd (1/s)	KD (M)	Rmax (RU)	tc	Chi^2 (RU^2)	U-value
5.744E+04	0.003538	6.159E-08	83.29	1.744E+17	0.514	1



Supplemental Figure 5. Kinetics of 10C10 and CH58 to A244 rgp120

A. Shown is the sensorgram of 10C10 binding to A244 rgp120. HIV envelope analyte concentrations was prepared using 3- to 10-fold dilutions, starting at 0.5uM.

B. Shown is the sensorgram of CH58 binding to A244 rgp120.



Supplemental Figure 6. Modeling and prediction of 10C10 Fv and A244 V2 158-178 peptide.

A. The 10C10 Fv was modeled using ABodyBuilder and visualized in Pymol. The heavy chain and light chain CDR loops are in blue and red, respectively. B. The secondary structure of the V2 peptide 158-178 was predicted using PEPFOLD3. Helix, coil and extended conformations for each residue in the peptide are in red, blue and green. C. 3D structure of the predicted helical peptide for V2 158-178 peptide by PEPFOLD3.

CDRH1		CDRH2		CDRH3	
Residue	**Score**	**Residue**	**Score**	**Residue**	**Score**
A	0.006908	W	0.932063	N	0.578826
S	0.030671	I	0.136628	N	0.158097
G	0.116662	D	0.892248	L	0.278848
F	0.37931	P	0.282315	D	0.511837
N	0.403707	E	0.86979	N	0.380133
I	0.212767	N	0.860497	W	0.0491504
K	0.817234	G	0.177158	G	0.0312921
G	0.901531	D	0.791682		
Y	0.84421	T	0.243621		
Y	0.960252	E	0.740837		
M	0.0237059				

CDRL1		CDRL2		CDRL3	
Residue	**Score**	**Residue**	**Score**	**Residue**	**Score**
T	0.00887094	I	0.0311456	Y	0.0277171
C	0.00288062	Y	0.614023	C	0.00628759
S	0.0092512	R	0.848979	L	0.0836534
A	0.0049029	T	0.12448	Q	0.0738457
S	0.0538501	S	0.218892	G	0.935189
S	0.21013	N	0.581454	S	0.890537
D	0.253054	L	0.175895	T	0.780582
I	0.184999	A	0.147413	F	0.958503
R	0.884908	S	0.36004	P	0.174484
S	0.665292	G	0.0746585	L	0.661789
N	0.774441	V	0.0443755	M	0.0193867
Y	0.850079			F	0.00419738
L	0.0203445			G	0.0019411
H	0.0422917				
W	0.00373734				
Y	0.0117615				

Supplemental Figure 7. Parapred prediction for residues of the paratope of 10C10

Parapred was used to predict the probability for each residue in the CDR that it would be part of the paratope of 10C10. Residues include CDR loops as determined by Chothia numbering scheme and two residues before and after the

CDR. The score listed is out of 1.0. Scores above 0.67 were used as “active” residues when docking with HADDOCK.

Chapter 4: Investigation of a CTLA4-like conserved sequence on HIV Env

Introduction

HIV infects T cells through recognition of the CD4 receptor, resulting in a loss of CD4+ helper T cells and the inability to control HIV infection. Positive and negative regulation of T cells is mediated by CD28 (Cluster of Differentiation 28) and CTLA4 (cytotoxic T-lymphocyte associated protein 4), two receptors found on the cell surface of T cells. CTLA4 has been found to be upregulated on HIV-specific CD4+ T cells in people infected with HIV, but lower or absent in elite controllers, whom are able to control their viral loads without treatment (Leng et al., 2002; Kaufmann et al., 2007). This suggested that there might be a relationship between HIV infection and regulation of T cells through CTLA4. While CTLA4 has been identified as an indicator of immunosuppression, the means by which CTLA4 is upregulated is unknown.

Viral genomes have been shown to consist of genetic sequences that enhance their fitness, also known as molecular mimicry. The V2 loop of HIV Env contains a conserved tripeptide motif, LDV/I, which mimics the natural ligands of α4β7 (Arthos et al., 2008). The integrin, α4β7, directs HIV to the gut where HIV has been found to severely deplete CD4+ T cells. In nature, the genomes of viruses will encode proteins that mimic the hosts' chemokines, cytokines or cytokine receptors in order to evade the hosts' immune system (Alcami, 2003). For example, African swine fever virus encodes a homologue of an NF-κB inhibitor, blocking the expression of cytokines normally induced by NF-κB (Dixon et al., 2019).

Similarly, a rabbit myelin basic protein (MBP) peptide sequence TTH**YGSLPQ**K was found to be similar to a Hepatitis B Virus Polymerase (HBVP) sequence IGC**YGSLPQ**E. Antibodies elicited by rabbits immunized with the HBVP peptide were cross-reactive with the MBP peptide (Fujinami and Oldstone, 1985). MBP is the second most abundant protein in the central nervous system, and consequently caused an inflammatory response localized to the central nervous system (Boggs, 2006). Thus, unique viral sequences present in the HIV genome can play a role in perturbing the immune response to HIV infection.

In this study, we have identified a conserved MYAPPI sequence on the HIV envelope protein (Env) similar to the hexapeptide MYPPPY motif found on CTLA4 and CD28. The MYPPPY sequence is conserved on both CTLA4 and CD28 and is the site of interaction between CTLA4 and CD28 and their ligands, CD80 and CD86 (Fig. 24) (Peach et al., 1995; Pentcheva-Hoang et al., 2004). The CD80 and CD86 ligands are located on antigen-presenting cells and mediate the activation and deactivation of T cells. Antibodies to CTLA4 and CD28 can interfere with normal signaling pathways and modify disease progression. For example, the anti-CTLA4 therapeutic antibody, Ipilimumab, blocks signaling through the CTLA4 receptor. It has been shown to be an effective tumor suppressor treatment, resulting in improved survival (Lipson and Drake, 2011; Wolchok et al., 2013). We hypothesized that the similarity of the conserved MYAPPI sequence found on HIV Env to the functionally important MYPPPY sequence on CTLA4 and CD28 may play a role in the activation or deactivation of T cells. We hypothesized that

antibodies elicited by HIV during infection could cross-react with CTLA4 or CD28 and possibly alter T cell function.

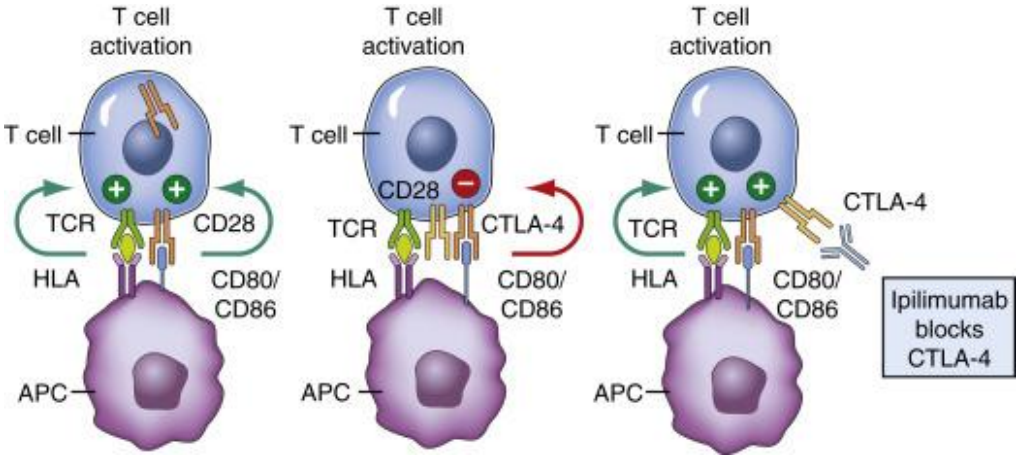


Figure 24. Activation and deactivation of T-cells through the CTLA4 and CD28 receptors.

(adapted from Bell et al., Oral, Head and Neck Oncology and Reconstructive Surgery (2018)) TCR – T-cell receptor. HLA – human leukocyte antigen. APC – antigen-presenting cells. Activation of T-cell requires two sets of signals: the TCR needs to recognize the antigen presented by the HLA and the CD28 receptor needs to bind to either the CD80 or the CD86 ligand. Following T-cell activation, CTLA moves from intracellular compartments to the cell surface, binds to CD80 and CD86 and prevents further activation of T-cells. Ipilimumab, a therapeutic antibody, blocks the CTLA4 receptor to enhance the immune response against tumors.

Methods

Fluorescence Immunoassay (FIA): Fluortrac high binding 96 well plates (Grenier Bio One, Kremsmunster, Austria) were coated with peptide or recombinant protein overnight at 4°C. Plates were blocked for 5 hours with 2% BSA and washed 4x with PBS-T. Rabbit serum diluted in PBS-T was added to wells and incubated for 90 minutes, followed by a 4X PBS-T wash. Alexa Fluor 488 goat anti-rabbit antibody (Invitrogen, Waltham, mA) was incubated at a 1:3000 dilution for 90 minutes. After a 4x PBS-T wash, wells were filled with 50 ul PBS. Plates were read on an EnVision plate reader (Perkin Elmer Inc, Waltham, MA) at excitation/emission wavelengths of 353/485 nm. Assays were performed in triplicate.

Results

Identification and conservation of the MYAPPI sequence

The MYAPPI sequence (HBX2 numbering amino acids 432-439) is located in the C4 region of gp120, adjacent to the CD4 binding site. Alignment of residues 432-441 from consensus sequences for each clade within group M of HIV-1 shows a high level of conservation (Fig. 25). Also, shown are the conserved CTLA4 and CD28 motif, MYPPPY, that mediate binding to the ligands CD80 and CD86, found on antigen-presenting cells (Pentcheva-Hoang et al., 2004).

Group	Sequence
Consensus A1	GQA**MYAPPI**QG
Consensus A2	GRA**MYAPPI**AG
Consensus B	GKA**MYAPPI**RG
Consensus C	GRA**MYAPPI**AG
Consensus D	GKA**MYAPPI**EG
Consensus F1	GRA**MYAAPI**AG
Consensus F2	GRA**MYAPPI**AG
Consensus G	GQA**MYAPPI**AG
Consensus H	GQA**MYAPPI**KG
CTLA4	VEL**MYPPPY**YL
CD28	IEV**MYPPPY**LD

Figure 25. Conservation of MYAPPI sequence in all clades of HIV-1 group M.

Sequences are from the Los Alamos National Laboratory HIV sequence database. Listed are amino acids from HXB2 positions 431-441 for consensus sequences for clades of HIV-1 group M. Sequences for human CTLA4 (uniprot# P16410) are from positions 131-141 and for human CD28 (uniprot #P10747) are from positions 113-124. In bold is the MYAPPI sequence in the consensus HIV Env sequences and the hexapeptide MYPPPY motif for CTLA4 and CD28.

We hypothesized that antibodies elicited by HIV could cross-react and bind to the MYPPY motifs of CTLA4 and CD28. For an antibody to be cross-reactive to both the MYAPPI sequence on HIV and the MYPPPY sequence on CTLA4 and CD28, these epitopes should be able to bind to and elicit antibodies. Previously, Nakamura et al. identified a mouse monoclonal antibody, 13H8, that recognizes the GKAMYAPPI sequence of the MN strain of the monomeric HIV envelope protein, MN-rgp120 (Nakamura et al., 1993). This suggests that the MYAPPI motif is both capable of eliciting antibodies and is available for binding by antibodies. Similarly, CTLA4 and CD28 can elicit antibodies at the MYPPPY motif. The anti-CTLA4 therapeutic antibody, Ipilimumab, was raised against human CTLA4 and disrupts the interaction of CTLA4 with its ligands at the MYPPPY motif (Keler et al., 2003). Ipilimumab is an effective treatment for cancer as it blocks CTLA4 from "turning off" the T cell, resulting in prolonged T cell activation. Thus, the MYAPPI sequence on gp120 and the MYPPPY sequence on CTLA4 and CD28 are epitopes which can elicit and be recognized by antibodies.

Preliminary binding of sera from gp120-immunized rabbits to CTLA4 and CD28

Because these similar hexapeptide sequences appeared on both the CTLA4 and CD28 receptors and on gp120, we tested whether antibodies elicited by gp120 could also bind to CTLA4 and CD28 (Fig. 26). Cross-reactive antibodies to both epitopes could have the ability to alter receptor function and affect T-cell activity. We first tested whether antibodies from gp120-immunized rabbits could bind to CTLA4 and CD28 peptides containing the MYPPPY motif and three residues before and after the motif (Fig. 27A). For positive controls, peptides from A244 and MN, the two gp120 immunogens in the rabbit immunization, were used. Rabbit sera from nine immunogens with the highest signal were plotted on a heat map. Pools of pre-immunization rabbit sera was used as a negative control (Fig. 27B). Certain immunogens had multiple rabbits display antibody-binding properties including gp120 108060_022 Q655R, 108048-gp140 and 108048-rgp120 Q655R.

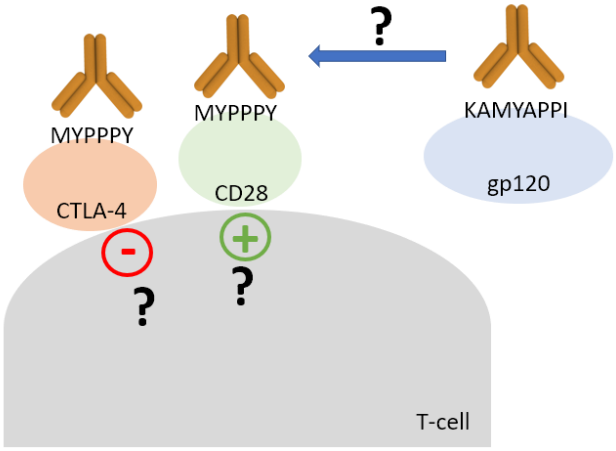


Figure 26. A possible mechanism for immunosuppression in HIV-infected individuals.

HIV infection results in antibodies elicited by gp120. Antibodies specific for the MYAPPI sequence may cross-react with the MYPPPY sequence on either CTLA4 or CD28. Blocking CTLA4 or CD28 can alter normal T-cell function, possibly resulting in immunosuppression.

A

Peptide	Peptide sequence
MN C4	GKAMYAPPIEGQ
A244 C4	GQAMYAPPISGT
hCTLA4	GELMYPPPYYLG
hCD28	GEVMYPPPYLDN

B

Rabbit #	
313	gp120 108060_022 Q655R
314	gp120 108060_022 Q655R
315	gp120 108060_022 Q655R
316	gp120 108060_022 Q655R
pre-im	
365	MN-D533
366	MN-D533
367	MN 533-430_436 deletion
368	MN 533-430_436 deletion
pre-im	
389	gD-MN rgp120
375	full length A244
376	full length A244
377	full length A244 in GNTI
378	full length A244 in GNTI
379	full length A244 in GNTI
380	full length A244 in GNTI
pre-im	
343	CN97001 D179N gp120
344	CN97001 D179N gp120
345	CN97001 D179N gp120
346	CN97001 D179N gp120
pre-im	
324	108048-gp140
325	108048-gp140
pre-im	
340	108048-rgp120 Q655R
341	108048-rgp120 Q655R
342	108048-rgp120 Q655R
pre-im	

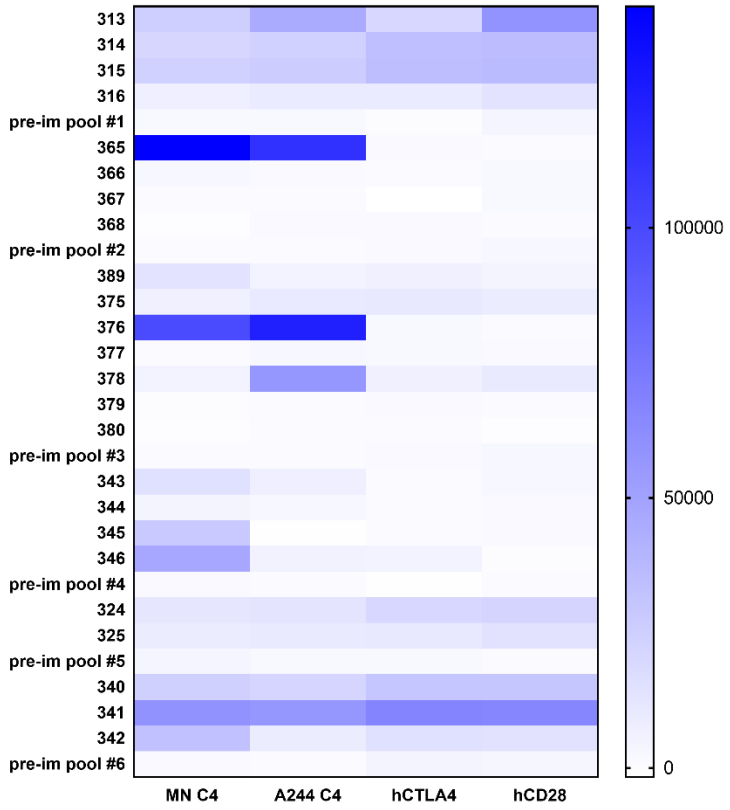


Figure 27. Preliminary binding data of sera from gp120-immunized rabbits to CTLA4 and CD28 peptides.

A. Peptide sequences for MN, A244, CTLA4 and CD28. B. Rabbits were immunized with gp120 and gp140 immunogens. Sera from the final day of immunization were used for binding studies by FIA. Peptides were coated at 5 ug/ml overnight at 4°C. Rabbit serum was diluted 1:100 in PBS-T and detected by Alexa Fluor A488 goat anti-rabbit antibody.

False positive binding data due to insufficient blocking power of 2% BSA

Gp120-immunized rabbit sera, gp120-immunized human sera and HIV-infected human sera were next tested against recombinant CTLA4 and CD28. However, during these assays, it was discovered that background signal was significant for the negative control, consisting of a blank PBS well that was blocked with 2% BSA (Bovine Serum Albumin) and probed with sera and secondary antibody. This differed from the negative control used in the earlier peptide binding assay where the negative control was a peptide-coated well probed with pre-immune rabbit serum which did not exhibit background signal. Chonblock, a commercial blocking buffer, has been shown to be a better blocking buffer for assays using serum antibodies (Waritani et al., 2017). Chonblock was tested alongside 2% BSA and showed no background signal in the negative control when tested with purified gp120 probed by purified human antibody, VRC01, by FIA (Fig. 28). Binding of VRC01 to gp120 was comparable when either Chonblock or 2% BSA were used.

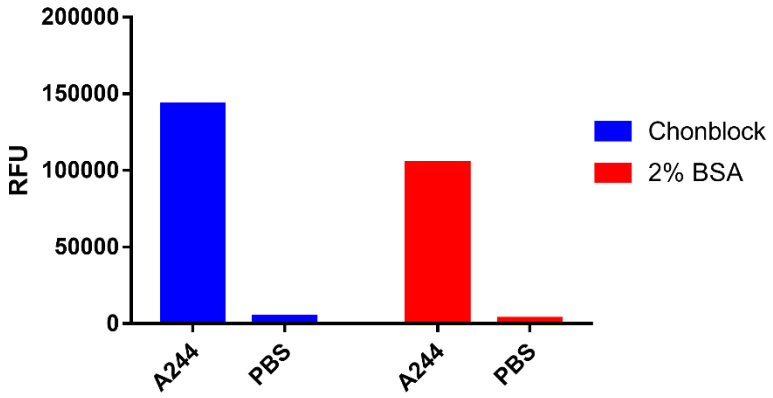


Figure 28. Comparison of Chonblock to 2% BSA in FIA with purified gp120 and purified human antibody.

Purified A244 rgp120 was blocked with either Chonblock or 2% BSA, probed with human broadly neutralizing antibody VRC01 at 10 ug/ml and detected with goat anti-human antibody. The negative control was a well filled with PBS and no antigen, probed with VRC01, and detected with goat anti-human antibody.

We repeated binding of rabbit sera to CTLA4 and CD28 peptides and CTLA4 and CD28 recombinant proteins and compared use of 2% BSA and Chonblock (Fig. 29). Serum from rabbit #341 (immunized with 108048-rgp120 Q665R) was used because it had the highest signal in the earlier experiment. When Chonblock was used as the blocking buffer, no signal was observed for the well with rabbit serum binding to peptide (Fig. 29A). #341 is a peptide-coated well probed with rabbit serum. #341 pre-immune is a peptide-coated well probed with pre-immune rabbit serum. PBS is an uncoated well and is probed with rabbit serum. For the wells that had 2% BSA as the blocking buffer, both the wells with antigen (#341) and without antigen (PBS) had signal, indicating that there was a false

positive signal caused by serum proteins binding to the well plate (Fig. 29B). The pre-immune sample did not have detectable signal and was shown to be an insufficient negative control compared to using no antigen (PBS) as a negative control. Similar to binding with the CTLA4 and CD28 peptide, no binding of rabbit serum to CTLA4 and CD28 recombinant protein was observed when using Chonblock (Fig. 29C). High background signal due to serum proteins was again observed when using 2% BSA when testing binding of rabbit serum to recombinant CTLA4 and CD28 (Fig. 29D). This demonstrated that Chonblock, but not 2% BSA could sufficiently block nonspecific binding of serum proteins.

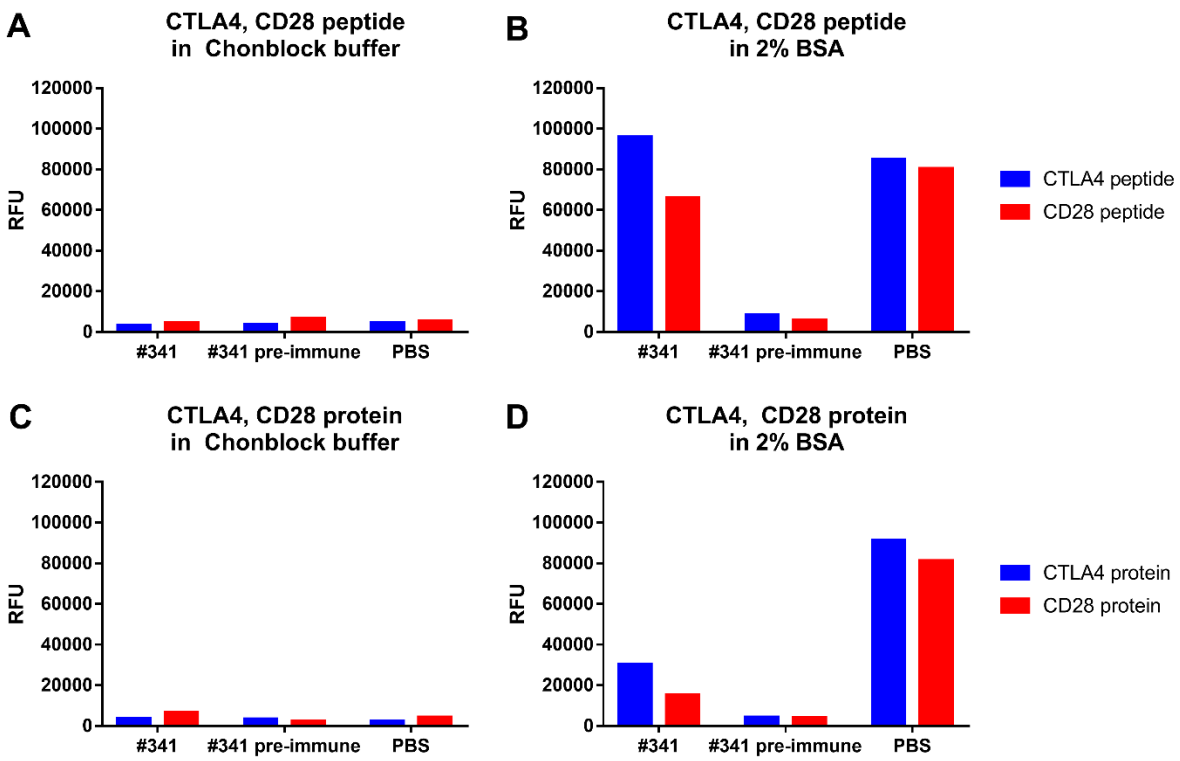


Figure 29. Comparison of Chonblock and 2% BSA with CTLA4/CD28 peptide and protein by FIA.

A. Binding of sera from a gp120-immunized rabbit to CTLA4 and CD28 peptide was tested by FIA. Peptides were blocked with Chonblock, probed with rabbit sera at a 1:30 dilution, and detected using an Alexa Fluor A488 goat anti-rabbit secondary antibody. Rabbit #341 was immunized with 108048-rgp120 Q665R. On the graph, #341 is a peptide-coated well that is probed with rabbit serum. The #341 pre-immune sample is a peptide-coated well that is probed with the #341 pre-immune serum. PBS is the negative control containing an antigen-uncoated well that is probed with rabbit serum. B. CTLA4 and CD28 peptide was blocked with 2% BSA, probed with rabbit sera and detected with goat anti-rabbit antibody. C. CTLA4 and CD28 proteins were blocked with Chonblock, probed with rabbit sera and detected with goat anti-rabbit antibody D) CTLA4 and CD28 protein were blocked with 2% BSA, probed with rabbit sera and detected with goat anti-rabbit antibody.

Discussion

In these studies, we tested whether antibodies by vaccination with gp120 could also bind to the T cell surface receptors, CTLA4 and CD28. While we did not see antibodies to HIV cross-react with CTLA4 or CD28, there are other mechanisms of action that can be tested which might suggest that HIV can interfere with normal CTLA4 or CD28 function and cause the immunosuppressive effects observed in HIV-infected individuals. One possible mechanism of action may occur through direct binding of the MYAPPI site on HIV Env to the ligands CD80 and CD86 on antigen-presenting cells. A pull-down assay or measurement of biomolecular interaction by Octet could be used to observe this possible interaction. If CTLA4 or CD28 are not the direct binding partners of the MYAPPI sequence of gp120, the cell lysates of peripheral blood mononuclear cells could be tested for possible interaction with MYAPPI sequence. α4β7, the gut-homing integrin that binds to the LVD/I motif of the HIV envelope protein, was identified from cell lysate of natural killer cells that was passed over a gp120 affinity column (Arthos et al., 2008). Similarly, the cell lysates of peripheral blood mononuclear cells could be passed over a MYAPPI peptide affinity column.

BSA was an ineffective blocking agent when animal sera were used in FIA assays. A high level of background signal in the negative control was due to serum proteins binding to the well plate when antigen was not coated on the plate. Nonspecific, false positive reactions in ELISAs can be caused by hydrophobic binding of immunoglobulin components from serum samples onto plastic well

plates as well as ionic interactions between immunoglobulin and antigen (Terato et al., 2014). BSA may be a suitable blocking agent when purified antibodies are used to probe antigens. However, it does not seem sufficiently block the nonspecific binding of proteins found in animal serum. Chonblock, a commercially available blocking buffer, and 100% normal goat serum were the only buffers found to block nonspecific binding of immunoglobulin to well plates (Waritani et al., 2017). From our experiments, we found that Chonblock was able to effectively block nonspecific binding of serum proteins in rabbit serum and eliminate false positive reactions.

Technical Report: Use of CRISPR/Cas9 for the development of a C1s$^{-/-}$ CHO-K1 cell line for HIV-1 vaccine production

Principle Investigator:	Phillip W. Berman, Ph.D. Berman Lab, University of California Santa Cruz 1156 High Street, MS-SOE2 Santa Cruz, CA 95064
Study Site:	Berman Lab, University of California Santa Cruz 1156 High Street, MS-SOE2 Santa Cruz, CA 95064
Report Date:	August 5th, 2019
Report Authors:	Sophia W. Li Phillip W. Berman, Ph.D.

Abstract

This report describes the knockout of C1s (complement component 1s) using CRISPR/Cas9 to create a $C1s^{-/-}$ CHO-K1 cell line for use in the production of HIV-1 vaccine immunogens. The purpose of these knockouts is to eliminate proteolysis of immunogens containing a Gly-Pro-Gly-Arg motif recognized by endogenous CHO C1s.

Introduction

Engineering better HIV immunogens continues to be the goal of the HIV vaccine field which includes the use of genetically modified, CRISPR-engineered cell lines. Most recently, the RV144 trial in Thailand demonstrated 31% efficacy in reducing the probability of infection with the use of HIV envelope proteins from two different strains. However, one of these strains belonging to a subtype of HIV called clade B, is cleaved by an endogenous protease when expressed in Chinese Hamster Ovary cells (CHO), a mammalian cell line. CHO cells are standard cell line used in the biopharmaceutical industry for the production of protein therapeutics, and in order to make scalable quantities of protein for clinical trials and commercial use of the vaccine, it is necessary to express HIV envelope proteins in CHO cells.

In this report, we discuss the CRISPR/Cas9 knockout of the C1s protease in the CHO-K1 cell line for the production of HIV vaccine immunogens. We previously described the stable Bal-rgp120 expressing $C1s^{-/-}$ $MGAT1^{-}$ CHO-S cell line (Li et

al., 2019a). Clade B rgp120s were shown to be cleaved at the V3 loop following the Arginine in the Glycine-Proline-Glycine-Arginine, or GPGR, motif when expressed in CHO cells. The protease was identified as C1s, a serine protease in the complement pathway and we showed that the knockout of C1s eliminated proteolysis of rgp120. The $C1s^{-/-}$ CHO-K1 cell line described here can be used for transient transfections or to create stable cell lines expressing any envelope protein.

Materials and Methods

Cells and antibodies

CHO-K1 cells were obtained from ATCC (ATCC, Manasses, VA) as adherent cells prior to their adaptation to suspension in the Berman lab.

Cell culture conditions

Cells were maintained in shake flasks (Corning, Corning, NY) using a Kuhner ISF1-x shaker incubator (Kuhner, Birsfelden, Switzerland) at 37°C, 8% CO_2, and 125 rpm. Static cultures were maintained in 96 or 24 well cell culture dishes (Corning) and grown in a Sanyo incubator (Sanyo, Osaka, Japan) at 37°C and 8% CO_2. All cell counts were performed using a TC20TM automated cell counter (BioRad, Hercules, CA) with viability determined by trypan blue exclusion (Thermo Fisher Scientific).

Cell culture media

For normal cell growth, cells were maintained in CHO Growth A media (Irvine Scientific, Santa Ana, CA) supplemented with 8mM GlutaMAX (Gibco, Gaithersburg, MD). Cells were initially maintained in CD CHO Medium (Gibco) with 8mM GlutaMax. However, due to clumping of cells, the media was switched to CHO Growth A. For protein production, 1M Sodium Butyrate was added to a final concentration of 1 mM to stall cell doubling. Cells were maintained in CHO Growth A media and 8mM GlutaMax and additionally supplemented at 10% of

culture volume every 3 days with a feed consisting of Proyield Cotton CNE50M-UF (FrieslandCampina, Delhi, NY), CD Efficient Feed C (Gibco) and 16mM glutamine.

Gene sequencing

The C1s gene was sequenced and verified against the NCBI mRNA RefSeq XM_007646821.2 and protein RefSeq XP_007645011.1. Guide RNA GTTGACAGCCGCTCATGTTG was cloned into the GeneArt CRISPR Nuclease Vector with OFP Reporter (GeneArt, Thermo Fisher). To sequence indels, 0.5 x 10e6 cells were spun down and boiled in 10 ul of milliQ water. 5 ul of cell lysate was used for PCR.

CRISPR/Cas9 target design and plasmid preparation

The C1s gene was sequenced and verified against the NCBI mRNA Reference Sequence XM_007646821.2 in Cricetulus griseus and the predicted protein NCBI Reference Sequence XP_007645011.1. The guide RNA GTTGACAGCCGCTCATGTTG was cloned into the CRISPR Nuclease Vector (GeneArt, Thermo Fisher) which expresses Cas9 and an orange fluorescent protein reporter. To ligate sgRNA inserts into the vector, DNA oligos were synthesized (Eurofins Genomics, Louisville, KY). Bacterial amplification of the completed vectors was carried out in One Shot TOP10 Chemically Competent E. coli following the recommended protocol (Thermo Fisher). Culture were grown in 15

ml LB broth overnight at 37℃ at 250 rpm. Minipreps (Qiagen, Redwood City, CA) were performed using the recommended protocol and submitted for Sanger sequencing (University of California Core Sequencing Facility, Berkeley, CA) with the U6 primer provided in the GeneArt CRISPR kit. 1 liter Maxipreps (Qiagen) were performed using the recommended protocol. Plasmid DNA was eluted into endotoxin-free water at concentrations greater than 5 mg/ml for optimal transfection efficiency.

Electroporation

Plasmids were transfected by electroporation using the Maxcyte STX scalable transfection system (MaxCyte Inc., Gaithersburg, MD). Cells were spun down at 250g for 10 minutes and resuspended in MaxCyte EP buffer (MaxCyte) at a density of 200 x 10e6 cells/mL. 80 x 10e6 cells and 120 ug of CRISPR Nuclease Vector DNA were combined for a total volume of 400 ul in the OC-400 processing assembly (MaxCyte). The CHO protocol in the MaxCyte STX software was used for transfections. After electroporation, cells were transferred into a 125 ml Erlenmeyer flask and incubated without shaking at 37℃ for 40 minutes. 15 ml of pre-warmed CD CHO media (Gibco) was added to the flasks, and the flasks were transferred to Kuhner shakers at 125 rpm.

Plating, expansion, and culture of CRISPR transfected CHO-K1 cells

For single cell cloning, five 96 well flat-bottomed tissue-culture treated microplates were filled with 50 ul of conditioned CHO Growth A media. Cells were serially diluted to 10 cells/ml in fresh CHO Growth A media and 50 ul of the diluted cells were added to each well, for a final concentration of 0.5 cell/well. Growth was monitored daily; any wells with more than a single colony were discarded. Cells were grown for 20 days at 37C, 8% CO2 and 85% humidity. At 50% confluency, cell culture supernatant was run on western blot to observe proteolytic activity and for selection of clones with successful knockout. 200 ng of purified, BaL-gp120 was incubated with 16 ul of supernatant at 37°C for 48 hours. Positive clones were moved into 24 well plates in a volume of 150 ul. At 50% confluence, clones were moved into 6 well plate in a volume of 3 ml.

Western blot

Samples of cell culture supernatants containing gp120 or purified proteins were run on NuPAGE 4-12% Bis-Tris precast gels (Thermo Fisher Scientific, Invitrogen, Carlsbad, CA) in MES running buffer (Thermo Fisher), with reduction by dithiothreitol (DTT), and stained with SimplyBlue stain (Thermo Fisher) For immunoblots, PAGE gels were transferred using iBlot 2 (Thermo Fisher Scientific). Membranes were blocked with 5% milk for 1 hour. The primary antibody, a polyclonal goat antibody raised against gp120, was incubated with the membrane at 1.5 ug/ml concentration in 5% milk. Blots were washed three times with PBS

with 0.05% Tween (PBS-T). The secondary antibody used was the Peroxidase AffiniPure bovine anti-goat IgG (Jackson ImmunoResearch, West Grove, PA) and incubated with the membrane at a 1:5000 dilution in 5% milk. Blots were washed three times with PBS-T and three times with PBS. The chemiluminescent substrate used was WesternBright ECL horseradish peroxidase substrate (Advansta, Menlo Park, CA). Images were taken using a FluorChem Q imager (Alpha Innotech, San Leandro, CA).

Materials Documentation

CHO Growth A media: Lot #94120180302

Gibco Glutamax: Lot# 1930108

CD CHO: Lot#1552716

Electroporation OC-400 cuvette: Lot #LM 235456

Electroporation buffer: Lot AC12974263

L-Glutamine: Lot#23317006

WFI Water: Lot #AC10220323, Lot# AD15533268

Results

Target design and cleavage of C1s$^{-/-}$ MGAT1^{-} CHO-K1

CRISPR/Cas9 was used to knockout the C1s protease in the suspension-adapted CHO-K1s cell line. Previously, the guide RNA GTTGACAGCCGCTCATGTTG was shown to have the highest knockout efficiency among 4 guide RNAs in knocking out C1s in the C1s$^{-/-}$ MGAT1^{-} CHO-S cell line (Li et al., 2019a). This same guide RNA was used again in knocking out C1s in the CHO-K1s cell line. The guide RNA was incorporated into the GeneArt CRISPR Nuclease vector, which we have utilized in previous studies (Fig. 30A, 30B) (Byrne et al., 2018b; Li et al., 2019a). Specifically, this sequence targets exon 11 of the C1s gene, which codes for the serine protease domain (Fig. 31). The CRISPR Nuclease vector containing the guide RNA GTTGACAGCCGCTCATGTTG was electroporated into the CHO-K1 cell line to knockout C1s.

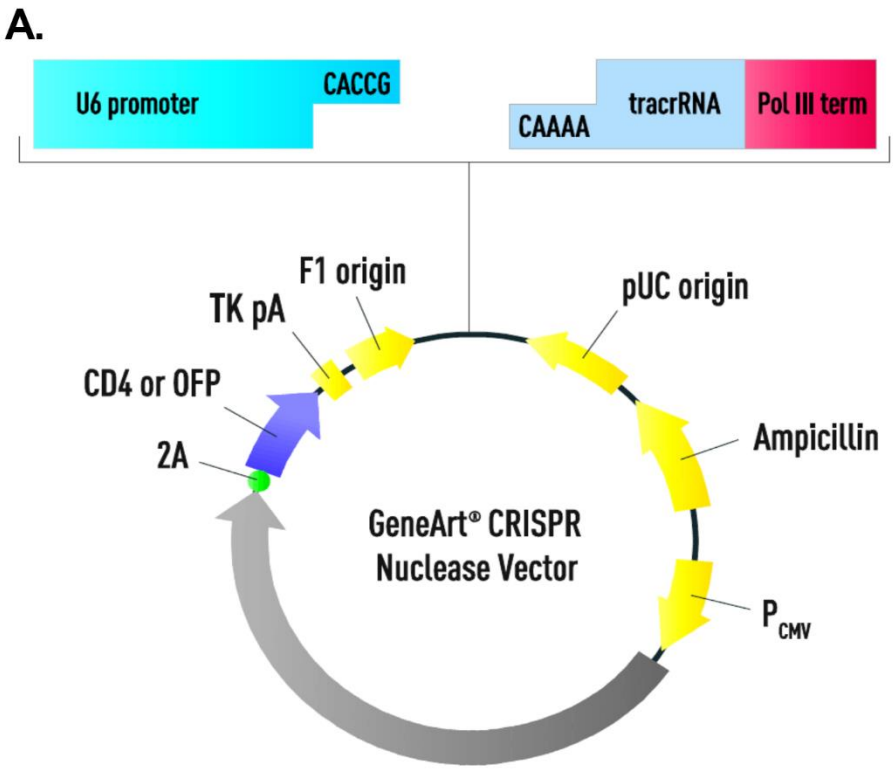


Figure 30. GeneArt CRISPR Nuclease Vector.

A. The vector contains the Cas9 nuclease and guide RNA. The CMV promoter drives expression of Cas9 and OFP, while the guide RNA expression is driven by the U6 promoter. The 2A peptide linker connects Orange Fluorescence Protein (OFP) to Cas9 allowing for selection by fluorescence. B. The target sequence is shown along with additional 3' overhangs of GTTTT and CGGTC to mediate ligation into the vector at the corresponding 5' overhangs CACCG and AAAAC.

TCTGTGGGGTACCCACTGAGCCCATTGCATTACAGCAGAGGATA
TTTGGAGGATTCCCTGCAAAGATCCAGAGTTTTCCTTGGCAAGTC
TTCTTTGAGTCCCCACGGGCCGGTGGGGCTCTTATTGACGAGTA
CTGGGT**GTTGACAGCCGCTCATGTTG**TGGAGGGAAACTCTGACC
CATCTATGTATGTGGGGACCACATTTGTGAGAATGGAACATCTGG
CGAATGCCCAGAGGCTCACCGCTGAACGTGTGATTATTCATCCA
GGCTGGAAGCCAGCGGATGACCTAGAAACACGGACAAATTTTGA
CAATGACATTGCACTGGTGCAGCTGAAAGACCCCGTGAAAATGG
GGCCCACTGTCTCCCCCATCTGCCTGCCAGGTACCTCCTCAGAG
TACAACCCCTCAAAGAATGACCTGGGACTGATCTCAGGGTGGGG
CCGAACAGAGAAGAGAAATATTGTTCCCCAACTCAAAGGGGCAA
AGTTACCTGTGACCTCTTTAGAGAAGTGCCAACAGGTGAAAGGG
GAGAACTCCAAAGTGAGGGCGGATGACTACGTTTTCACCAGCAA
CATGATCTGTGCTGGAGAGAAAGGTGTTGATAGCTGTCAGGGGG
ACAGTGGTGGGGCTTTTGCTTTGCAGGTCCCCAATGTAAAGGAC
CCCAAATTCTATGTGGCAGGCCTAGTGTCCTGGGGGAAAAAGTG
TGGGACCTATGGAATCTACACAAAGGTAAAGAACTACATGGATTG
GATCGTGAAGACGATGCAGGAGAATAGTGTCCCCAGTAAGGACT
AACCCATACATACGGCCAGACCCTCCCACCAGTGCATGCTCTCA
GACCCTTATGATAGTCCCATTATCTCACAATGATGGAGAGAAGAT
GCAGGAGTACGATTAACACGGAACTTGATTGCCAAGACACCTGG
ATGGAAGGAGGGTCTGATCCAGGTGTGTTGATTATCCAGGTGTG
AATTCCTGCCTGTTACTGGTAACAACATGGCAGAGGCTCTTCTGT
CTTACACCATTCTACAGGGGTTTCTTAACTCCCCCCCTTTTACCT
GCTTGGGATTCCATTTGTAACCCTTGGAGTCCCTTATGCTTACAA
TAAAGCGTATTTCTAATCTGGCTGACCTCAGGGGAGGGAAATGA
CAGGGCACCTTGTCATGGATGGAACTCAGGCATCAGCAGCTCCT
CTTTAACCATCATGTAAACAGGAAATTCAGTCTTTCTTCTTCCTGA
GACTTACTGAATACCATGAGGTTAGGATGGACTGTTTTAGTTTCC
TTTGTGAGTTGTCTCAATCTGTCTCCTAAAAATTAATTTTCTTATGA
TTTGCATTCCTTTGTACCATGCAATATGATGACAGTTAAAAAAAAA
TCGAACAT

Figure 31. Targeting exon 11 of the C1s gene for CRISPR/Cas9 Knockout.

Shown above is the sequence of exon 11 of the C1s gene in CHO cells. In bold is the target sequence used in the CRISPR Nuclease Vector. The protospacer adjacent motif or PAM is underlined.

Selection of clones with successful knockout

A flowchart of the steps for the two rounds of CRISPR/Cas9 needed to knockout both alleles of C1s is shown (Fig. 32). Three days after transfection with the CRISPR Nuclease vector, cells underwent single cell dilution to 0.5 cells/well in 96 well plates in order to grow homogeneously genetic clones. At the same time, cells were sampled and analyzed by TIDE (Tracking of Indels by DEcomposition) to estimate knockout efficiency, which was shown to be between 12-20% (Fig. 33) (Brinkman et al., 2014). From 10 plates, 33 clones were identified and screened by western blot to identify clones without proteolytic activity (Fig. 34A). Purified BaL-rgp120 was incubated with supernatant from the clones taken from the 96 well plates and incubated at 37°C for 48 hours to allow proteolysis to proceed. From these clones, two clones, Clones 5 and Clones 6, had knockouts in one of the two alleles of C1s. TIDE analysis was performed on clone 5 to determine indel formation; the wildtype sequence along with a -1 basepair deletion were seen in equal proportion, indicating knockout of one allele (Fig. 34B). The sequence around the indel site was further sequenced and showed the mutation in more detail. One allele was indeed the wildtype sequence, while the other allele had a deleted thymidine (Fig. 34C).

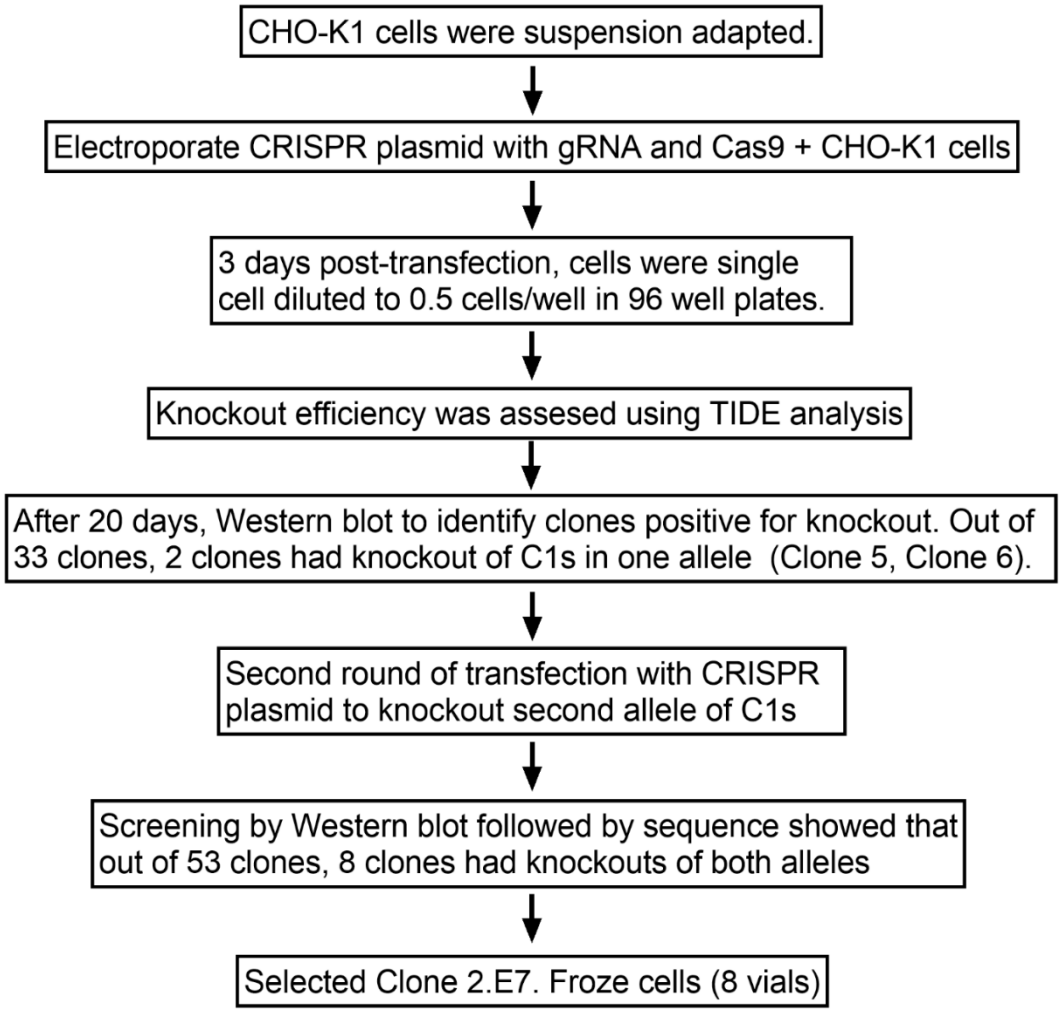


Figure 32. Flow chart of C1s gene editing in the CHO-K1 cell line and clone selection strategy.

The first round of knockout resulted in only clones with knockouts of one allele of C1s. Two such clones were selected which were called Clone 5 and Clone 6. Subsequently, a second round of knockout was done to knockout the second allele of C1s, resulting in the selection of Clone 2.E7, a $C1s^{-/-}$ CHO-K1 clone.

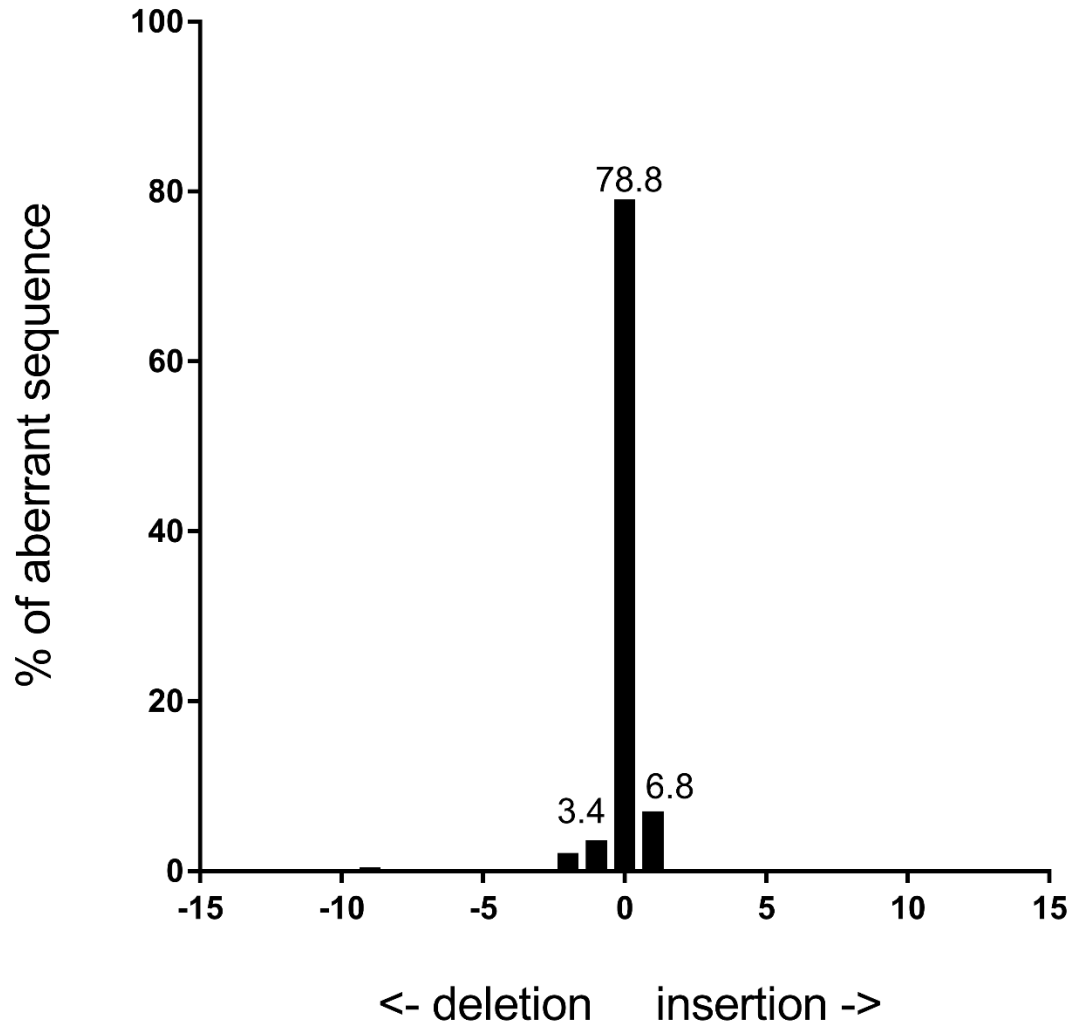


Figure 33. Post-transfection knockout efficiency in the CHO-K1 cell line.

TIDE analysis was used to determine knockout efficiency of the CHO-K1 cell line after transfection with the CRISPR Nuclease Vector. 0.5e6 cells were sampled after transfection. Cells were lysed and the sequence around the expected mutation site was PCR amplified. Sanger sequencing results were analyzed by TIDE. The x-axis indicates the size of the insertion or deletion. The y-axis indicates the percentage of cells with the respective indel.

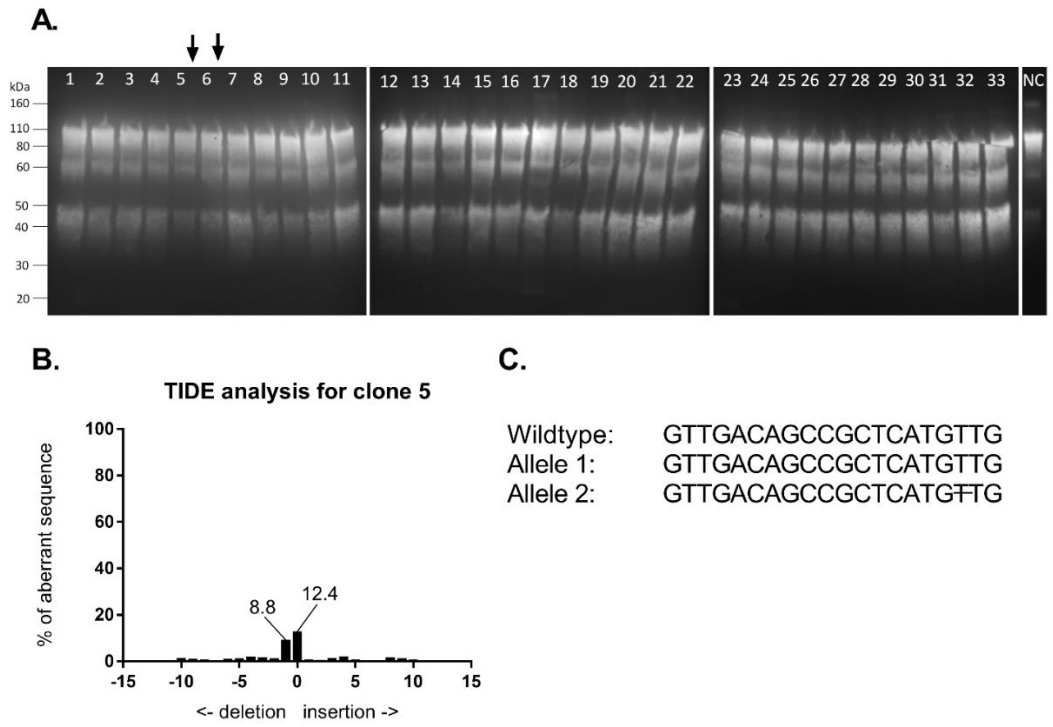


Figure 34. Knockout of one of two alleles of C1s in first round of CRISPR/Cas9.

A. CHO-K1 clones with knockout of C1s were screened by Western Blot. 200 ng of purified BaL-rgp120 was incubated with 15 ul of supernatant from 24 well plates for 48 hours at 37℃. Samples were run reduced with DTT. Arrows indicate Clones 5 and 6, which contained knockout of one allele of C1s. B. For Clone 5, the sequence around the expected indel site was analyzed by TIDE to mutations at the expected indel site. The indel site was amplified by PCR and went through Sanger sequencing. D. The sequence around the indel site was PCR amplified and cloned into a plasmid in order to sequence the mutation that occurred.

A second round of CRISPR/Cas9 knockout was needed to knockout the second allele of C1s. Another round of transfection and single cell cloning similar to the first was completed. Supernatant was incubated with purified Bal-rgp120 and run on Western blot to observe clones that had the second allele of C1s knocked

out (Fig. 35A). Out of the 32 clones screen, clones 2.E7, 3.E8, 3.G2, and 4.A6 had both alleles of C1s knocked out. An additional 11 clones were screened, and 4 clones identified with knockout of both alleles of C1s including Clone 2.E9 (data not shown). Clones 2.E7, 3.E8, and 2.E9 were scaled up to shake flasks and transiently transfected with a plasmid expressing BaL-rgp120, a clade B Env with the GPGR motif, to observe expression of Envs in these cell lines with complete knockout of C1s (Fig. 35B). Clone 2.E7 was selected based on its high expression of rgp120. Additional TIDE analysis and sequencing showed the Cas9-mediated mutations in more detail (Fig. 35C, 35D). One allele had a +2 basepair insertion with two extra thymidines, while the other allele showed a -1 basepair deletion of a thymidine.

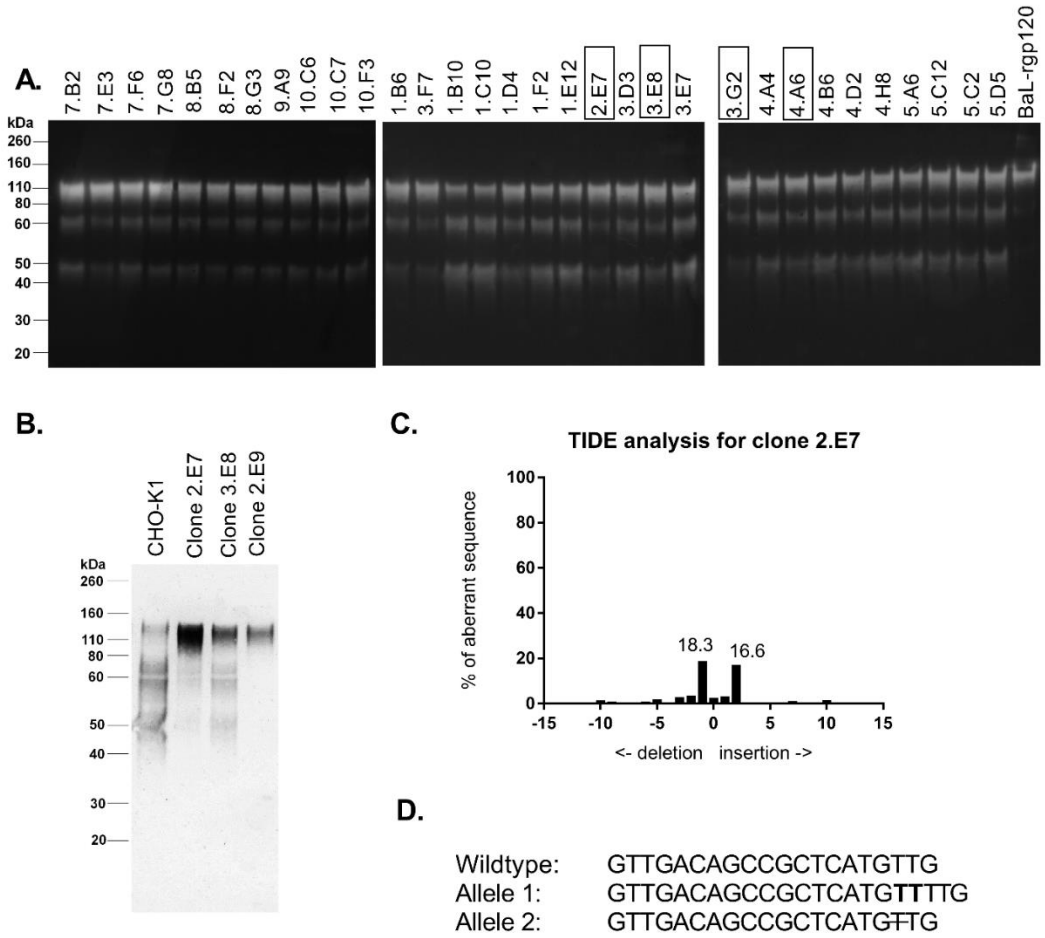


Figure 35. Selection of Clone 2.E7, a C1s-/- CHO-K1 clone.

A. Clones with both alleles of C1s knocked out were screened by Western Blot. 100 ng of purified BaL-rgp120 was incubated with 15 ul of supernatant from 24 well plates overnight at 37℃. Samples were run reduced with DTT. Clones marked with a black box have knockouts of both alleles as verified by sequencing. B. Bal-rgp120 was transiently transfected into C1s-/- CHO-K1 clones 2.E7, 2.E9, and 3.E8 and cultures were grown for 6 days. 5 ul of supernatant was run on Western Blot and all samples were reduced with DTT. C. The sequence around the expected indel site was analyzed by TIDE for clone 2.E7 to check knockout of both alleles. The indel site was amplified by PCR and went through Sanger sequencing. D. The sequence around the indel site was PCR amplified and cloned into a plasmid in order to sequence the two different mutations for each allele.

Discussion

Using CHO cells as the cellular substrate for expression of HIV immunogens is desirable due to its high productivity, ease of scalability and common use in the manufacturing of therapeutic proteins. However, endogenous proteases can be problematic when expressed proteins are recognized as substrates by these proteases. Proteolysis of rgp120s in CHO cells by an unknown serine protease has been documented, but the protease had never been identified (Stephens et al., 1990; Clements et al., 1991; Du et al., 2008; Pugach et al., 2015). In our recent work, we identified C1s as the serine protease cleaving rgp120 at the Gly-Pro-Gly-Arg motif and showed that knockout of C1s in CHO cells using CRISPR/Cas9 could eliminate proteolysis of recombinantly expressed immunogens (Li et al., 2019a).

To support the production of all variants of immunogens, we have utilized CRISPR/Cas9 to genetically modify the CHO-K1 cell line to express non-proteolyzed, HIV Envs. This $C1s^{-/-}$ CHO-K1 cell line can be used to express HIV envelope proteins including but not limited to rgp120, rgp140, trimers or envelopes for sequential immunizations. The same guide RNA used in our previous paper to knockout C1s in the Bal-rgp120 expressing $C1s^{-/-}$ $MGAT1^{-}$ CHO-S cell line was used to knockout C1s in the $MGAT1^{-}$ CHO-K1 cell line. Two rounds of CRISPR/Cas9 knockout was needed to completely knockout both alleles of C1s. This resulted in the final selection of 1 clone called Clone 2.E7, a $C1s^{-/-}$ CHO-K1 clone. Knockout of the C1s protease in the CHO-K1 cell line allows for the

expression of non-proteolyzed clade B Envs with the GPGR motif and for the use of clade B Envs as candidates for HIV vaccines.

Technical Report: Use of CRISPR/Cas9 for the development of a C1s$^{-/-}$ MGAT1^{-} CHO-K1 cell line for HIV-1 vaccine production

Principle Investigator:	Phillip W. Berman, Ph.D. Berman Lab, University of California Santa Cruz 1156 High Street, MS-SOE2 Santa Cruz, CA 95064
Study Site:	Berman Lab, University of California Santa Cruz 1156 High Street, MS-SOE2 Santa Cruz, CA 95064
Report Date:	July 30th, 2019
Report Authors:	Sophia W. Li Phillip W. Berman, Ph.D.

Abstract

This report describes the knockout of two enzymes, C1s (complement component 1s) and MGAT1 (mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase), using CRISPR/Cas9 to create a $C1s^{-/-}$ $MGAT1^{-}$ CHO-K1 cell line for use in the production of HIV-1 vaccine immunogens. The purpose of these knockouts is to eliminate proteolysis of immunogens containing a Gly-Pro-Gly-Arg motif recognized by endogenous CHO C1s and to restrict glycosylation to high mannose glycans thus improving binding of immunogens to glycan-dependent broadly neutralizing antibodies.

Introduction

Engineering better HIV immunogens continues to be the goal of the HIV vaccine field which includes the use of genetically modified, CRISPR-engineered cell lines. Most recently, the RV144 trial in Thailand demonstrated 31% efficacy in reducing the probability of infection with the use of HIV envelope proteins from two different strains. However, one of these strains, belonging to a subtype of HIV called clade B, is cleaved by an endogenous protease when expressed in Chinese Hamster Ovary cells (CHO), a mammalian cell line. CHO cells are standard cell line used in the biopharmaceutical industry for the production of protein therapeutics, and in order to make scalable quantities of protein for clinical trials and commercial use of the vaccine, it is necessary to express HIV envelope proteins in CHO cells.

In this report, we discuss the CRISPR/Cas9 knockout of the C1s protease in the MGAT1^{-} CHO-K1 cell line for the production of HIV vaccine immunogens. We previously described the stable Bal-rgp120 C1s$^{-/-}$ MGAT1^{-} CHO-S cell line (Li et al., 2019a). Clade B rgp120s were shown to be cleaved at the V3 loop following the Arginine in the Glycine-Proline-Glycine-Arginine, or GPGR, motif when expressed in CHO cells. The protease was identified as C1s, a serine protease in the complement pathway and we showed that the knockout of C1s eliminated proteolysis of rgp120. The C1s$^{-/-}$ MGAT1^{-} CHO-K1 cell line described here can be used for transient transfections or stable cell line expressing other envelope proteins. Furthermore, the MGAT1- knockout will restrict glycosylation of the expressed immunogens to high mannose glycans and improve binding to glycan-dependent broadly neutralizing antibodies.

Materials and Methods

Cells and antibodies

CHO-K1 cells were obtained from ATCC (ATCC, Manasses, VA) as adherent cells prior to their adaptation to suspension in the Berman lab. MGAT1- CHOK1s cells. The MGAT1- CHO-K1 cell line ids described in "Technical Report: Use of CRISPR/Cas9 for the development of a CHO-K1 MGAT1- deficient cell line suitable for HIV-1 vaccine production".

Cell culture conditions

Cells were maintained in shake flasks (Corning, Corning, NY) using a Kuhner ISF1-x shaker incubator (Kuhner, Birsfelden, Switzerland) at 37°C, 8% CO_2, and 125 rpm. Static cultures were maintained in 96 or 24 well cell culture dishes (Corning) and grown in a Sanyo incubator (Sanyo, Osaka, Japan) at 37°C and 8% CO_2. All cell counts were performed using a TC20TM automated cell counter (BioRad, Hercules, CA) with viability determined by trypan blue exclusion (Thermo Fisher Scientific).

Cell culture media

For normal cell growth, cells were maintained in CHO Growth A media (Irvine Scientific, Santa Ana, CA) supplemented with 8mM GlutaMAX (Gibco, Gaithersburg, MD). Cells were initially maintained in CD CHO Medium (Gibco) with 8mM GlutaMax. However, due to clumping of cells, the media was switched

to CHO Growth A. For protein production, 1M Sodium Butyrate was added to a final concentration of 1 mM to stall cell doubling. Cells were maintained in CHO Growth A media and 8mM GlutaMax and additionally supplemented at 10% of culture volume every 3 days with a feed consisting of Proyield Cotton CNE50M-UF (FrieslandCampina, Delhi, NY), CD Efficient Feed C (Gibco) and 16mM glutamine.

Gene sequencing

The C1s gene was sequenced and verified against the NCBI mRNA RefSeq XM_007646821.2 and protein RefSeq XP_007645011.1. Guide RNA GTTGACAGCCGCTCATGTTG was cloned into the GeneArt CRISPR Nuclease Vector with OFP Reporter (GeneArt, Thermo Fisher). To sequence indels, 0.5 x 10e6 cells were spun down and boiled in 10 ul of milliQ water. 5 ul of cell lysate was used for PCR.

CRISPR/Cas9 target design and plasmid preparation

The C1s gene was sequenced and verified against the NCBI mRNA Reference Sequence XM_007646821.2 in Cricetulus griseus and the predicted protein NCBI Reference Sequence XP_007645011.1. The guide RNA GTTGACAGCCGCTCATGTTG was cloned into the CRISPR Nuclease Vector (GeneArt, Thermo Fisher) which expresses Cas9 and an orange fluorescent protein reporter. To ligate sgRNA inserts into the vector, DNA oligos were synthesized

(Eurofins Genomics, Louisville, KY). Bacterial amplification of the completed vectors was carried out in One Shot TOP10 Chemically Competent E. coli following the recommended protocol (Thermo Fisher). Culture were grown in 15 ml LB broth overnight at 37℃ at 250 rpm. Minipreps (Qiagen, Redwood City, CA) were performed using the recommended protocol and submitted for Sanger sequencing (University of California Core Sequencing Facility, Berkeley, CA) with the U6 primer provided in the GeneArt CRISPR kit. 1 liter Maxipreps (Qiagen) were performed using the recommended protocol. Plasmid DNA was eluted into endotoxin-free water at concentrations greater than 5 mg/ml for optimal transfection efficiency.

Electroporation

Plasmids were transfected by electroporation using the Maxcyte STX scalable transfection system (MaxCyte Inc., Gaithersburg, MD). Cells were spun down at 250g for 10 minutes and resuspended in MaxCyte EP buffer (MaxCyte) at a density of 200 x 10e6 cells/mL. 80 x 10e6 cells and 120 ug of CRISPR Nuclease Vector DNA were combined for a total volume of 400 ul in the OC-400 processing assembly (MaxCyte). The CHO protocol in the MaxCyte STX software was used for transfections. After electroporation, cells were transferred into a 125 ml Erlenmeyer flask and incubated without shaking at 37℃ for 40 minutes. 15 ml of pre-warmed CD CHO media (Gibco) was added to the flasks, and the flasks were transferred to Kuhner shakers at 125 rpm.

Plating, expansion, and culture of CRISPR transfected CHO-K1 cells

For single cell cloning, five 96 well flat-bottomed tissue-culture treated microplates were filled with 50 ul of conditioned CHO Growth A media. Cells were serially diluted to 10 cells/ml in fresh CHO Growth A media and 50 ul of the diluted cells were added to each well, for a final concentration of 0.5 cell/well. Growth was monitored daily; any wells with more than a single colony were discarded. Cells were grown for 20 days at 37C, 8% CO2 and 85% humidity. At 50% confluency, cell culture supernatant was run on western blot to observe proteolytic activity and for selection of clones with successful knockout. 200 ng of purified, BaL-gp120 was incubated with 16 ul of supernatant at 37°C for 48 hours. Positive clones were moved into 24 well plates in a volume of 150 ul. At 50% confluence, clones were moved into 6 well plate in a volume of 3 ml.

Western blot

Samples of cell culture supernatants containing gp120 or purified proteins were run on NuPAGE 4-12% Bis-Tris precast gels (Thermo Fisher Scientific, Invitrogen, Carlsbad, CA) in MES running buffer (Thermo Fisher), with reduction by dithiothreitol (DTT), and stained with SimplyBlue stain (Thermo Fisher) For immunoblots, PAGE gels were transferred using iBlot 2 (Thermo Fisher Scientific). Membranes were blocked with 5% milk for 1 hour. The primary antibody, a polyclonal goat antibody raised against gp120, was incubated with the membrane

at 1.5 ug/ml concentration in 5% milk. Blots were washed three times with PBS with 0.05% Tween (PBS-T). The secondary antibody used was the Peroxidase AffiniPure bovine anti-goat IgG (Jackson ImmunoResearch, West Grove, PA) and incubated with the membrane at a 1:5000 dilution in 5% milk. Blots were washed three times with PBS-T and three times with PBS. The chemiluminescent substrate used was WesternBright ECL horseradish peroxidase substrate (Advansta, Menlo Park, CA). Images were taken using a FluorChem Q imager (Alpha Innotech, San Leandro, CA).

Materials Documentation

CHO Growth A media: Lot #94120180302

Gibco Glutamax: Lot# 1930108

CD CHO: Lot#1552716

Electroporation OC-400 cuvette: Lot #LM 235456

Electroporation buffer: Lot AC12974263

L-Glutamine: Lot#23317006

WFI Water: Lot #AC10220323, Lot# AD15533268

Results

Target design and cleavage of C1s$^{-/-}$ MGAT1^{-} CHO-K1

CRISPR/Cas9 was used to knockout the C1s protease in the suspension-adapted MGAT1^{-} CHO-K1s cell line. Previously, the guide RNA GTTGACAGCCGCTCATGTTG was shown to have the highest knockout efficiency among 4 guide RNAs in knocking out C1s in the C1s$^{-/-}$ MGAT1^{-} CHO-S cell line (Li et al., 2019a). This same guide RNA was used again in knocking out C1s in the MGAT1^{-} CHO-K1s cell line. The guide RNA was incorporated into the GeneArt CRISPR Nuclease vector, which we have utilized in previous studies (Fig. 36A, 36B) (Byrne et al., 2018b; Li et al., 2019a). Specifically, this sequence targets exon 11 of the C1s gene, which codes for the serine protease domain (Fig. 37). The CRISPR Nuclease vector containing the guide RNA GTTGACAGCCGCTCATGTTG was electroporated into the MGAT1- CHO-K1 cell line to combine knockouts of the MGAT1 and C1s enzymes.

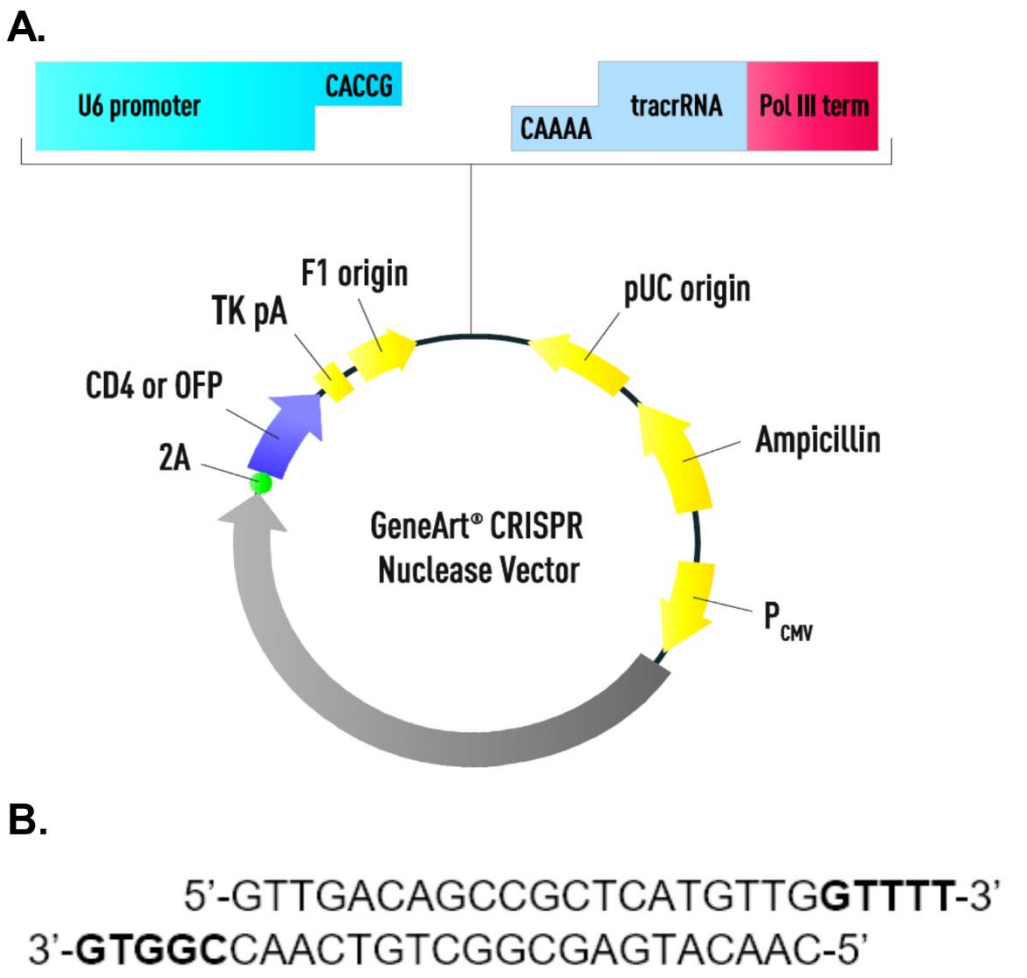


Figure 36. GeneArt CRISPR Nuclease Vector.

A. The vector contains the Cas9 nuclease and guide RNA. The CMV promoter drives expression of Cas9 and OFP, while the guide RNA expression is driven by the U6 promoter. The 2A peptide linker connects Orange Fluorescence Protein (OFP) to Cas9 allowing for selection by fluorescence. B. The target sequence is shown along with additional 3' overhangs of GTTTT and CGGTC to mediate ligation into the vector at the corresponding 5' overhangs CACCG and AAAAC.

TCTGTGGGGTACCCACTGAGCCCATTGCATTACAGCAGAGGATA
TTTGGAGGATTCCCTGCAAAGATCCAGAGTTTTCCTTGGCAAGTC
TTCTTTGAGTCCCCACGGGCCGGTGGGGCTCTTATTGACGAGTA
CTGGGT**GTTGACAGCCGCTCATGTTG**TGGAGGGAAACTCTGACC
CATCTATGTATGTGGGGACCACATTTGTGAGAATGGAACATCTGG
CGAATGCCCAGAGGCTCACCGCTGAACGTGTGATTATTCATCCA
GGCTGGAAGCCAGCGGATGACCTAGAAACACGGACAAATTTTGA
CAATGACATTGCACTGGTGCAGCTGAAAGACCCCGTGAAAATGG
GGCCCACTGTCTCCCCCATCTGCCTGCCAGGTACCTCCTCAGAG
TACAACCCCTCAAAGAATGACCTGGGACTGATCTCAGGGTGGGG
CCGAACAGAGAAGAGAAATATTGTTCCCCAACTCAAAGGGGCAA
AGTTACCTGTGACCTCTTTAGAGAAGTGCCAACAGGTGAAAGGG
GAGAACTCCAAAGTGAGGGCGGATGACTACGTTTTCACCAGCAA
CATGATCTGTGCTGGAGAGAAAGGTGTTGATAGCTGTCAGGGGG
ACAGTGGTGGGGCTTTTGCTTTGCAGGTCCCCAATGTAAAGGAC
CCCAAATTCTATGTGGCAGGCCTAGTGTCCTGGGGGAAAAAGTG
TGGGACCTATGGAATCTACACAAAGGTAAAGAACTACATGGATTG
GATCGTGAAGACGATGCAGGAGAATAGTGTCCCCAGTAAGGACT
AACCCATACATACGGCCAGACCCTCCCACCAGTGCATGCTCTCA
GACCCTTATGATAGTCCCATTATCTCACAATGATGGAGAGAAGAT
GCAGGAGTACGATTAACACGGAACTTGATTGCCAAGACACCTGG
ATGGAAGGAGGGTCTGATCCAGGTGTGTTGATTATCCAGGTGTG
AATTCCTGCCTGTTACTGGTAACAACATGGCAGAGGCTCTTCTGT
CTTACACCATTCTACAGGGGTTTCTTAACTCCCCCCCTTTTACCT
GCTTGGGATTCCATTTGTAACCCTTGGAGTCCCTTATGCTTACAA
TAAAGCGTATTTCTAATCTGGCTGACCTCAGGGGAGGGAAATGA
CAGGGCACCTTGTCATGGATGGAACTCAGGCATCAGCAGCTCCT
CTTTAACCATCATGTAAACAGGAAATTCAGTCTTTCTTCTTCCTGA
GACTTACTGAATACCATGAGGTTAGGATGGACTGTTTTAGTTTCC
TTTGTGAGTTGTCTCAATCTGTCTCCTAAAAATTAATTTTCTTATGA
TTTGCATTCCTTTGTACCATGCAATATGATGACAGTTAAAAAAAAA
TCGAACAT

Figure 37. Targeting exon 11 of the C1s gene for CRISPR/Cas9 Knockout.

Shown above is the sequence of exon 11 of the C1s gene in CHO cells. In bold is the target sequence used in the CRISPR Nuclease Vector. The protospacer adjacent motif or PAM is underlined.

Selection of clones with successful knockout

A flowchart of the steps taken from electroporation to cell culture in 125 ml shake flasks is shown (Fig. 38). Three days after transfection with the CRISPR Nuclease vector, cells underwent single cell dilution to 0.5 cells/well in 96 well plates in order to grow homogeneously genetic clones. At the same time, cells were sampled and analyzed by TIDE (Tracking of Indels by DEcomposition) to estimate knockout efficiency, which was shown to be between 6-17% (Fig. 39). From 10 plates, 30 clones were identified. 24 clones were run on western blot to identify clones without proteolytic activity (Fig. 40). Purified BaL-rgp120 was incubated with supernatant from the clones taken from the 96 well plates and incubated at 37°C for 48 hours to allow proteolysis to proceed. Of the clones, only Clone A did not display a progression of proteolysis, suggesting knockout of C1s was successful.

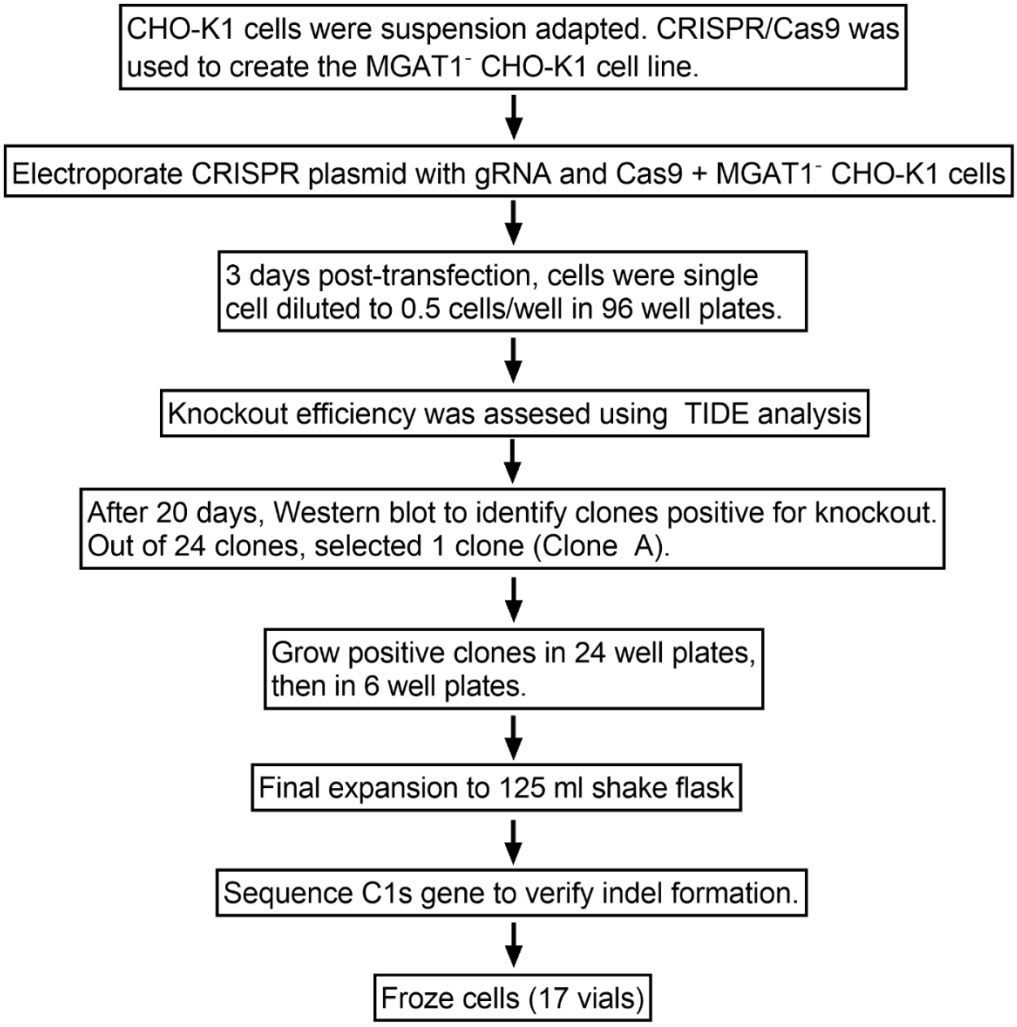


Figure 38. Flow chart of C1s gene editing in the MGAT1$^-$ CHO-K1 cell line and clone selection strategy.

The flowchart shows the steps taken to knockout the C1s gene, select for a CHO cell clone with a successful knockout of C1s, verify indel formation by Cas9 in the C1s gene, and freeze down cells.

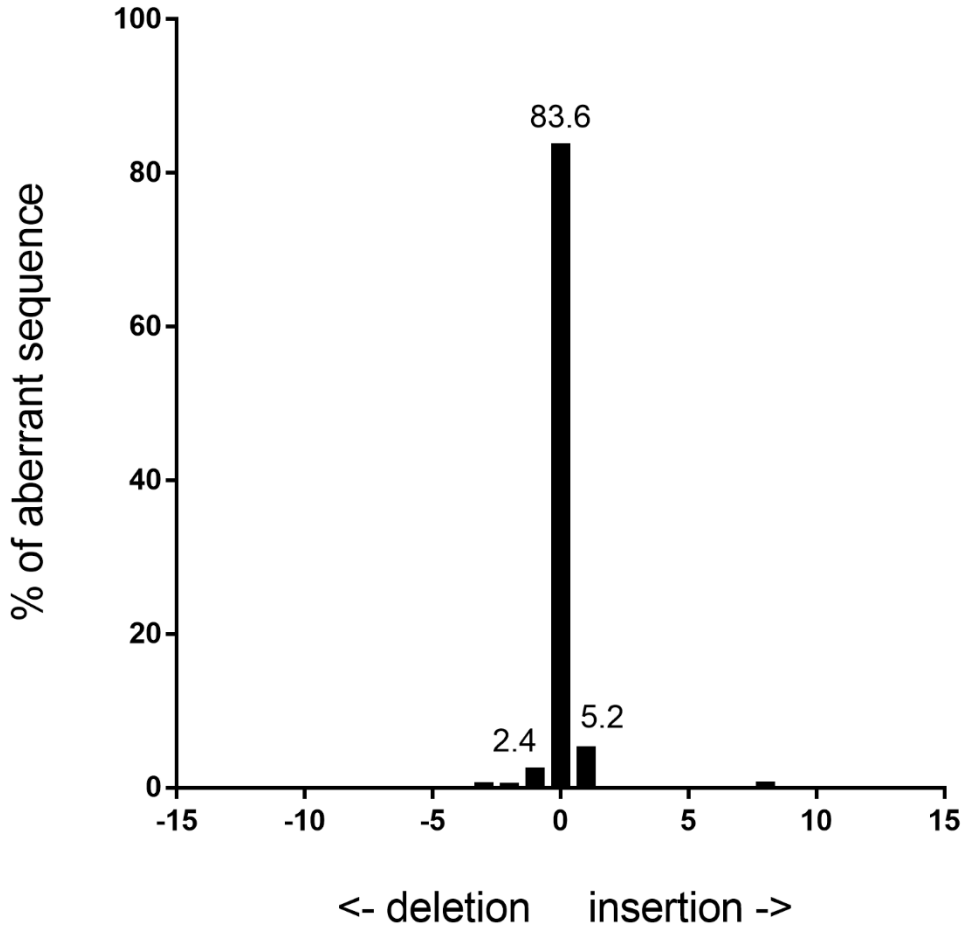


Figure 39. Post-transfection knockout efficiency of the C1s gene.

TIDE analysis was used to determine knockout efficiency of the MGAT1- CHO-K1 cell line after transfection with the CRISPR Nuclease Vector. 0.5e6 cells were sampled after transfection. Cells were lysed and the sequence around the expected mutation site was PCR amplified. Sanger sequencing results were analyzed by TIDE. The x-axis indicates the size of the insertion or deletion. The y-axis indicates the percentage of cells with the respective indel.

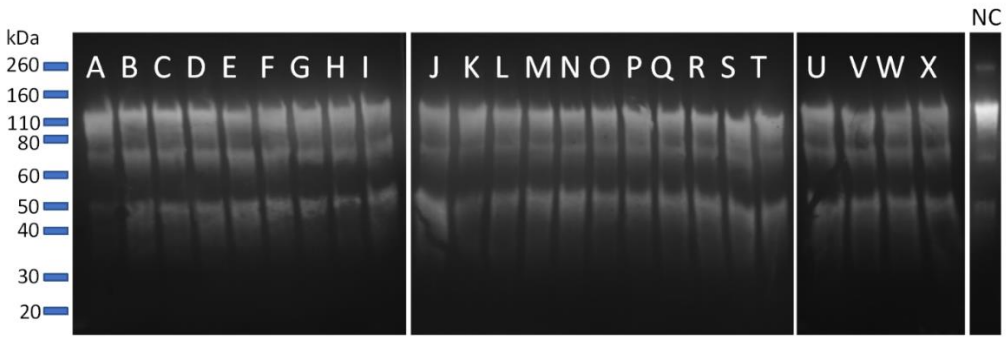


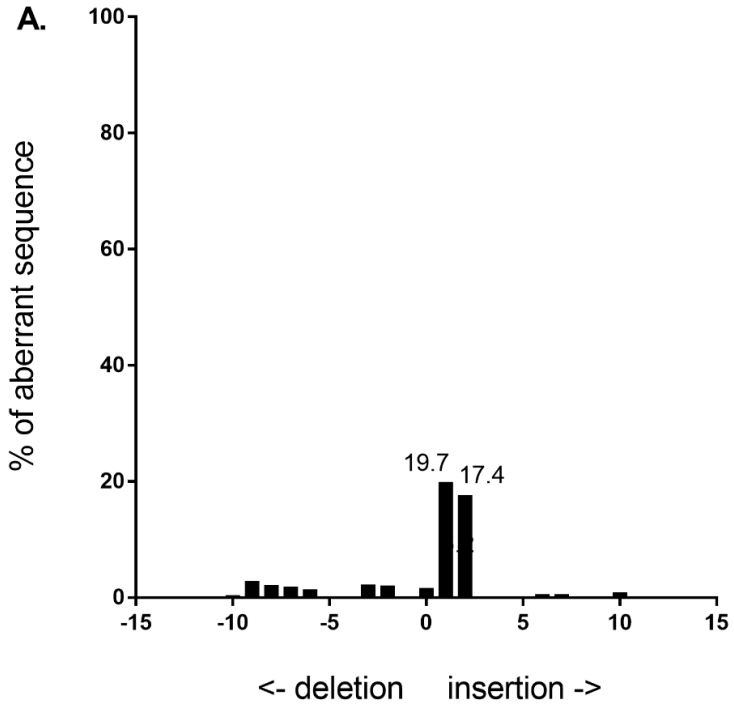
Figure 40. Selection of Clone A from single cell clones after transfection of MGAT1^{-} CHO-K1.

24 single cell clones were grown in 96 well plates for 20 days until >50% confluent. 16 ul of supernatant was removed and assayed for proteolytic activity by western blot. Supernatant was incubated with 200 ng of purified BaL-rgp120, which is shown as the negative control above. All samples were run reduced with DTT. The purified Bal-rgp120 has a molecular weight of approximately 120 kDa and is proteolyzed into a 70 kDa and 50 kDa fragment.

The Cas9-mediated indel in Clone A was sequenced to verify the success of the knockout. Sanger sequencing was used to sequence the nucleotides around the expected indel site. TIDE analysis showed the presence of two unique indels of a +1 and +2 two insertion, indicating two unique alleles as was seen previously (Fig. 41A) (Li et al., 2019a). These two mutations were observed at about a 50:50 ratio, again indicating two alleles from a single cell rather than a heterogeneous population of cells. Further sequencing showed that one allele has an inserted thymidine, while the other allele has two inserted thymidines (Fig. 41B). Exon 11 codes for the serine protease domain of the C1s protease and is 264 amino acids

long (Fig. 42A) The indels observed in the two alleles results in truncated serine protease domains that are 57 and 68 amino acids long (Fig. 42B, 42C).

A.



B.

Wildtype sequence	GTTGACAGCCGCTCATGTTG	
Allele 1	GTTGACAGCCGCTCATG TTTG	Inserted T
Allele 2	GTTGACAGCCGCTCATG TTTTG	Inserted TT

Figure 41. Analysis of indel formation for Clone A, a C1s$^{-/-}$ MGAT1^{-} CHO-K1 clone.

A. Cells from Clone A were lysed and the nucleotide sequence around the expected indel site was PCR amplified. Sanger sequencing of the PCR product was analyzed by TIDE. B. The amplified PCR product around the expected indel site was cloned, transformed into E. coli and sequenced in order to determine the specific mutations on each of the two alleles.

Figure 42. NHEJR induced changes to exon 11 of the C1s gene.

A. The 264 amino acid sequence of exon 11 of the C1s gene. B. Allele 1 with an inserted thymidine results in a truncated 57 amino acid sequence due to an early stop codon. C. Allele 2 with two inserted thymidines results in a truncated 68 amino acid sequence

Discussion

Using CHO cells as the cellular substrate for expression of HIV immunogens is desirable due to its high productivity, ease of scalability and common use in the manufacturing of therapeutic proteins. However, endogenous proteases can be problematic when expressed proteins are recognized as substrates by these proteases. Proteolysis of rgp120s in CHO cells by an unknown serine protease has been documented, but the protease had never been identified (Stephens et al., 1990; Clements et al., 1991; Du et al., 2008; Pugach et al., 2015). In our recent work, we identified C1s as the serine protease cleaving rgp120 at the Gly-Pro-Gly-Arg motif and showed that knockout of C1s in CHO cells using CRISPR/Cas9 could eliminate proteolysis of recombinantly expressed immunogens (Li et al., 2019a).

To support the production of all variants of immunogens, we have utilized CRISPR/Cas9 to genetically modify the CHO-K1 cell line to express non-proteolyzed, HIV Envs with high-mannose glycans. This $C1s^{-/-}$ $MGAT1^{-}$ CHO-K1 cell line can be used to express HIV envelope proteins including but not limited to rgp120, rgp140, trimers or envelopes for sequential immunizations. The same guide RNA used in our previous paper to knockout C1s in the Bal-rgp120 expressing $C1s^{-/-}$ $MGAT1^{-}$ CHO-S cell line was used to knockout C1s in the $MGAT1^{-}$ CHO-K1 cell line. Screening by western blot of 24 clones resulted in the selection of 1 clone called Clone A, a successful $C1s^{-/-}$ $MGAT1^{-}$ CHO-K1 clone. As seen in our previous study, indels were formed in both alleles of C1s, indicating complete

inactivation of the protease. Knockout of the C1s protease in the CHO-K1 cell line allows for the expression of non-proteolyzed clade B Envs with the GPGR motif and for the use of clade B Envs as candidates for HIV vaccines.

Bibliography

Alam, S.M., Dennison, S.M., Aussedat, B., Vohra, Y., Park, P.K., Fernández-Tejada, A., Stewart, S., Jaeger, F.H., Anasti, K., Blinn, J.H., et al. (2013). Recognition of synthetic glycopeptides by HIV-1 broadly neutralizing antibodies and their unmutated ancestors. PNAS *110*, 18214–18219.

Alcami, A. (2003). Viral mimicry of cytokines, chemokines and their receptors. Nature Reviews Immunology *3*, 36–50.

Amara, U., Flierl, M.A., Rittirsch, D., Klos, A., Chen, H., Acker, B., Brückner, U.B., Nilsson, B., Gebhard, F., Lambris, J.D., et al. (2010). Molecular Intercommunication between the Complement and Coagulation Systems. The Journal of Immunology *185*, 5628–5636.

Ambrosetti, F., Jiménez-García, B., Roel-Touris, J., and Bonvin, A.M.J.J. (2020). Modeling Antibody-Antigen Complexes by Information-Driven Docking. Structure *28*, 119-129.e2.

Amin, M.N., McLellan, J.S., Huang, W., Orwenyo, J., Burton, D.R., Koff, W.C., Kwong, P.D., and Wang, L.-X. (2013). Synthetic glycopeptides reveal the glycan specificity of HIV-neutralizing antibodies. Nat. Chem. Biol. *9*, 521–526.

Arthos, J., Cicala, C., Martinelli, E., Macleod, K., Ryk, D.V., Wei, D., Xiao, Z., Veenstra, T.D., Conrad, T.P., Lempicki, R.A., et al. (2008). HIV-1 envelope

protein binds to and signals through integrin α 4 β 7 , the gut mucosal homing receptor for peripheral T cells. Nat Immunol *9*, 301–309.

Bale, S., Martiné, A., Wilson, R., Behrens, A.-J., Le Fourn, V., de Val, N., Sharma, S.K., Tran, K., Torres, J.L., Girod, P.-A., et al. (2018). Cleavage-Independent HIV-1 Trimers From CHO Cell Lines Elicit Robust Autologous Tier 2 Neutralizing Antibodies. Front Immunol *9*, 1116.

Bandres, J.C., Wang, Q.F., O'Leary, J., Baleaux, F., Amara, A., Hoxie, J.A., Zolla-Pazner, S., and Gorny, M.K. (1998). Human Immunodeficiency Virus (HIV) Envelope Binds to CXCR4 Independently of CD4, and Binding Can Be Enhanced by Interaction with Soluble CD4 or by HIV Envelope Deglycosylation. J Virol *72*, 2500–2504.

Barouch, D.H., Liu, J., Li, H., Maxfield, L.F., Abbink, P., Lynch, D.M., Iampietro, M.J., SanMiguel, A., Seaman, M.S., Ferrari, G., et al. (2012). Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. Nature *482*, 89–93.

Barouch, D.H., Tomaka, F.L., Wegmann, F., Stieh, D.J., Alter, G., Robb, M.L., Michael, N.L., Peter, L., Nkolola, J.P., Borducchi, E.N., et al. (2018). Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19). The Lancet *392*, 232–243.

Baycin-Hizal, D., Tabb, D.L., Chaerkady, R., Chen, L., Lewis, N.E., Nagarajan, H., Sarkaria, V., Kumar, A., Wolozny, D., Colao, J., et al. (2012). Proteomic Analysis of Chinese Hamster Ovary Cells. J Proteome Res *11*, 5265–5276.

Bee, J.S., Tie, L., Johnson, D., Dimitrova, M.N., Jusino, K.C., and Afdahl, C.D. (2015). Trace levels of the CHO host cell protease cathepsin D caused particle formation in a monoclonal antibody product. Biotechnology Progress *31*, 1360–1369.

Berman, P.W. (2015). Strategies on Improving the AIDSVAX Vaccine (NIAID, Bethesda, MD).

Berman, P.W., Gregory, T.J., Riddle, L., Nakamura, G.R., Champe, M.A., Porter, J.P., Wurm, F.M., Hershberg, R.D., Cobb, E.K., and Eichberg, J.W. (1990). Protection of chimpanzees from infection by HIV-1 after vaccination with recombinant glycoprotein gp120 but not gp160. Nature *345*, 622–625.

Berman, P.W., Huang, W., Riddle, L., Gray, A.M., Wrin, T., Vennari, J., Johnson, A., Klaussen, M., Prashad, H., Köhne, C., et al. (1999). Development of Bivalent (B/E) Vaccines Able to Neutralize CCR5-Dependent Viruses from the United States and Thailand. Virology *265*, 1–9.

Boggs, J.M. (2006). Myelin basic protein: a multifunctional protein. Cell. Mol. Life Sci. *63*, 1945–1961.

Bonomelli, C., Doores, K.J., Dunlop, D.C., Thaney, V., Dwek, R.A., Burton, D.R., Crispin, M., and Scanlan, C.N. (2011). The Glycan Shield of HIV Is Predominantly Oligomannose Independently of Production System or Viral Clade. PLoS One *6*.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Research *42*, e168–e168.

Buchbinder, S.P., Mehrotra, D.V., Duerr, A., Fitzgerald, D.W., Mogg, R., Li, D., Gilbert, P.B., Lama, J.R., Marmor, M., Rio, C. del, et al. (2008). Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. The Lancet *372*, 1881–1893.

Burton, D.R., and Hangartner, L. (2016). Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. Annu Rev Immunol *34*, 635–659.

Burton, D.R., Stanfield, R.L., and Wilson, I.A. (2005). Antibody vs. HIV in a clash of evolutionary titans. PNAS *102*, 14943–14948.

Byrne, G., O'Rourke, S.M., Alexander, D.L., Yu, B., Doran, R.C., Wright, M., Chen, Q., Azadi, P., and Berman, P.W. (2018a). CRISPR/Cas9 gene editing for the creation of an MGAT1-deficient CHO cell line to control HIV-1 vaccine glycosylation. PLOS Biology *16*, e2005817.

Byrne, G., O'Rourke, S.M., Alexander, D.L., Yu, B., Doran, R.C., Wright, M., Chen, Q., Azadi, P., and Berman, P.W. (2018b). CRISPR/Cas9 gene editing for the creation of an MGAT1-deficient CHO cell line to control HIV-1 vaccine glycosylation. PLoS Biol *16*.

Chakrabarti, S., Barrow, C.J., Kanwar, R.K., Ramana, V., and Kanwar, J.R. (2016). Studies to Prevent Degradation of Recombinant Fc-Fusion Protein Expressed in Mammalian Cell Line and Protein Characterization. Int J Mol Sci *17*.

Chiang, A.W.T., Li, S., Kellman, B.P., Chattopadhyay, G., Zhang, Y., Kuo, C.-C., Gutierrez, J.M., Ghazi, F., Schmeisser, H., Ménard, P., et al. (2019). Combating viral contaminants in CHO cells by engineering innate immunity. Scientific Reports *9*, 1–15.

Clements, G.J., Price-Jones, M.J., Stephens, P.E., Sutton, C., Schulz, T.F., Clapham, P.R., McKeating, J.A., McClure, M.O., Thomson, S., and Marsh, M. (1991). The V3 loops of the HIV-1 and HIV-2 surface glycoproteins contain proteolytic cleavage sites: a possible function in viral fusion? AIDS Res. Hum. Retroviruses *7*, 3–16.

Clincke, M.-F., Guedon, E., Yen, F.T., Ogier, V., and Goergen, J.-L. (2011). Characterization of metalloprotease and serine protease activities in batch CHO

cell cultures: control of human recombinant IFN-γ proteolysis by addition of iron citrate. BMC Proc *5*, P115.

Conway, E.M. (2015). Reincarnation of ancient links between coagulation and complement. Journal of Thrombosis and Haemostasis *13*, S121–S132.

Dey, A.K., Cupo, A., Ozorowski, G., Sharma, V.K., Behrens, A.-J., Go, E.P., Ketas, T.J., Yasmeen, A., Klasse, P.J., Sayeed, E., et al. (2018). cGMP production and analysis of BG505 SOSIP.664, an extensively glycosylated, trimeric HIV-1 envelope glycoprotein vaccine candidate. Biotechnology and Bioengineering *115*, 885–899.

Dixon, L.K., Islam, M., Nash, R., and Reis, A.L. (2019). African swine fever virus evasion of host defences. Virus Res *266*, 25–33.

Dominguez, C., Boelens, R., and Bonvin, A.M.J.J. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. *125*, 1731–1737.

Doores, K.J., Bonomelli, C., Harvey, D.J., Vasiljevic, S., Dwek, R.A., Burton, D.R., Crispin, M., and Scanlan, C.N. (2010). Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. PNAS *107*, 13800–13805.

Dorai, H., Santiago, A., Campbell, M., Tang, Q.M., Lewis, M.J., Wang, Y., Lu, Q.-Z., Wu, S.-L., and Hancock, W. (2011). Characterization of the proteases

involved in the N-terminal clipping of glucagon-like-peptide-1-antibody fusion proteins. Biotechnology Progress *27*, 220–231.

Doran, R.C., Tatsuno, G.P., O'Rourke, S.M., Yu, B., Alexander, D.L., Mesa, K.A., and Berman, P.W. (2018a). Glycan modifications to the gp120 immunogens used in the RV144 vaccine trial improve binding to broadly neutralizing antibodies. PLOS ONE *13*, e0196370.

Doran, R.C., Yu, B., Wright, M., O'Rourke, S.M., Yin, L., Richardson, J.M., Byrne, G., Mesa, K.A., and Berman, P.W. (2018b). Development of a Stable MGAT1- CHO Cell Line to Produce Clade C gp120 with Improved Binding to Broadly Neutralizing Antibodies. Front. Immunol. *9*.

Du, S.X., Xu, L., Viswanathan, S., and Whalen, R.G. (2008). Inhibition of V3-specific cleavage of recombinant HIV-1 gp120 produced in Chinese hamster ovary cells. Protein Expr. Purif. *59*, 223–231.

Duerr, R., and Gorny, M.K. (2019). V2-Specific Antibodies in HIV-1 Vaccine Research and Natural Infection: Controllers or Surrogate Markers. Vaccines *7*, 82.

Dwyer-Lindgren, L., Cork, M.A., Sligar, A., Steuben, K.M., Wilson, K.F., Provost, N.R., Mayala, B.K., VanderHeide, J.D., Collison, M.L., Hall, J.B., et al. (2019). Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. Nature *570*, 189–193.

van Eeden, C., Kurt Wibmer, C., Scheepers, C., Richardson, S.I., Nonyane, M., Lambson, B., Mkhize, N.N., Vijayakumar, B., Sheng, Z., Stanfield-Oakley, S., et al. (2018). V2-Directed Vaccine-like Antibodies from HIV-1 Infection Identify an Additional K169-Binding Light Chain Motif with Broad ADCC Activity. Cell Rep *25*, 3123-3135.e6.

Eggink, D., Melchers, M., Wuhrer, M., van Montfort, T., Dey, A.K., Naaijkens, B.A., David, K.B., Le Douce, V., Deelder, A.M., Kang, K., et al. (2010). Lack of complex N-glycans on HIV-1 envelope glycoproteins preserves protein conformation and entry function. Virology *401*, 236–247.

Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., and Frye, C.C. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. Biotechnol. Bioeng. *109*, 1007–1015.

Fauci, A.S., and Marston, H.D. (2014). Ending AIDS — Is an HIV Vaccine Necessary? New England Journal of Medicine *370*, 495–498.

Flynn, N.M., Forthal, D.N., Harro, C.D., Judson, F.N., Mayer, K.H., Para, M.F., and rgp120 HIV Vaccine Study Group (2005). Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. J. Infect. Dis. *191*, 654–665.

Foley, B., and Korber, B. (1996). Global variation in the HIV-1 V3 region. MS K710, Los Alamos National Laboratory III: 77-139.

Frantz, C., Stewart, K.M., and Weaver, V.M. (2010). The extracellular matrix at a glance. J Cell Sci *123*, 4195–4200.

Frey, B.F., Jiang, J., Sui, Y., Boyd, L.F., Yu, B., Tatsuno, G., Billeskov, R., Solaymani-Mohammadi, S., Berman, P.W., Margulies, D.H., et al. (2018). Effects of Cross-Presentation, Antigen Processing, and Peptide Binding in HIV Evasion of T Cell Immunity. The Journal of Immunology *200*, 1853–1864.

Fujinami, R.S., and Oldstone, M.B. (1985). Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity. Science *230*, 1043–1045.

Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G.J., and Fontecilla-Camps, J.C. (2000). Crystal structure of the catalytic domain of human complement C1s: a serine protease with a handle. EMBO J *19*, 1755–1765.

Gallwitz, M., Enoksson, M., Thorpe, M., and Hellman, L. (2012). The Extended Cleavage Specificity of Human Thrombin. PLoS One *7*.

Gosalia, D.N., Salisbury, C.M., Ellman, J.A., and Diamond, S.L. (2005). High Throughput Substrate Specificity Profiling of Serine and Cysteine Proteases Using Solution-phase Fluorogenic Peptide Microarrays. Molecular & Cellular Proteomics *4*, 626–636.

Gray, G.E., Allen, M., Moodie, Z., Churchyard, G., Bekker, L.-G., Nchabeleng, M., Mlisana, K., Metch, B., Bruyn, G. de, Latka, M.H., et al. (2011). Safety and

efficacy of the HVTN 503/Phambili Study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study. The Lancet Infectious Diseases *11*, 507–515.

Hammond, S., Kaplarevic, M., Borth, N., Betenbaugh, M.J., and Lee, K.H. (2012). Chinese hamster genome database: an online resource for the CHO community at www.CHOgenome.org. Biotechnol. Bioeng. *109*, 1353–1356.

Haynes, B.F., Gilbert, P.B., McElrath, M.J., Zolla-Pazner, S., Tomaras, G.D., Alam, S.M., Evans, D.T., Montefiori, D.C., Karnasuta, C., Sutthent, R., et al. (2012). Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial. New England Journal of Medicine *366*, 1275–1286.

Huang, Y.-M., Hu, W., Rustandi, E., Chang, K., Yusuf-Makagiansar, H., and Ryll, T. (2010). Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. Biotechnol. Prog. *26*, 1400–1410.

Hultin, M.B., and Jesty, J. (1981). The activation and inactivation of human factor VIII by thrombin: effect of inhibitors of thrombin. Blood *57*, 476–482.

Hutchinson, J.M., Mesa, K.A., Alexander, D.L., Yu, B., O'Rourke, S.M., Limoli, K.L., Wrin, T., Deeks, S.G., and Berman, P.W. (2019). Unusual Cysteine Content in V1 Region of gp120 From an Elite Suppressor That Produces Broadly Neutralizing Antibodies. Front Immunol *10*.

Hwang, S.S., Boyle, T.J., Lyerly, H.K., and Cullen, B.R. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science *253*, 71–74.

Julien, J.-P., Sok, D., Khayat, R., Lee, J.H., Doores, K.J., Walker, L.M., Ramos, A., Diwanji, D.C., Pejchal, R., Cupo, A., et al. (2013). Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. PLoS Pathog. *9*, e1003342.

Junqueira, D.M., and Almeida, S.E. de M. (2016). HIV-1 subtype B: Traces of a pandemic. Virology *495*, 173–184.

Kaufman, R.J., Wasley, L.C., and Dorner, A.J. (1988). Synthesis, Processing, and Secretion of Recombinant Human Factor VI11 Expressed in Mammalian Cells. Journal of Biological Chemistry *263*, 6352–6392.

Kaufmann, D.E., Kavanagh, D.G., Pereyra, F., Zaunders, J.J., Mackey, E.W., Miura, T., Palmer, S., Brockman, M., Rathod, A., Piechocka-Trocha, A., et al. (2007). Upregulation of CTLA-4 by HIV-specific $CD4^+$ T cells correlates with disease progression and defines a reversible immune dysfunction. Nature Immunology *8*, 1246–1254.

Keler, T., Halk, E., Vitale, L., O'Neill, T., Blanset, D., Lee, S., Srinivasan, M., Graziano, R.F., Davis, T., Lonberg, N., et al. (2003). Activity and Safety of

CTLA-4 Blockade Combined with Vaccines in Cynomolgus Macaques. The Journal of Immunology *171*, 6251–6259.

Kerr, F.K., O'Brien, G., Quinsey, N.S., Whisstock, J.C., Boyd, S., Banda, M.G. de la, Kaiserman, D., Matthews, A.Y., Bird, P.I., and Pike, R.N. (2005). Elucidation of the Substrate Specificity of the C1s Protease of the Classical Complement Pathway. J. Biol. Chem. *280*, 39510–39514.

Köhler, G., and Milstein, C. (1976). Derivation of specific antibody-producing tissue culture and tumor lines by cell fusion. Eur. J. Immunol. *6*, 511–519.

Kong, L., He, L., de Val, N., Vora, N., Morris, C.D., Azadnia, P., Sok, D., Zhou, B., Burton, D.R., Ward, A.B., et al. (2016). Uncleaved prefusion-optimized gp140 trimers derived from analysis of HIV-1 envelope metastability. Nat Commun *7*.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S., and Bhattacharya, T. (2000). Timing the Ancestor of the HIV-1 Pandemic Strains. Science *288*, 1789–1796.

Krem, M.M., and Cera, E.D. (2002). Evolution of enzyme cascades from embryonic development to blood coagulation. Trends in Biochemical Sciences *27*, 67–74.

Kwong, P.D., Mascola, J.R., and Nabel, G.J. (2013). Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. Nature Reviews Immunology *13*, 693–701.

Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., and Tufféry, P. (2016). PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. *44*, W449-454.

LaRosa, G.J., Davide, J.P., Weinhold, K., Waterbury, J.A., Profy, A.T., Lewis, J.A., Langlois, A.J., Dreesman, G.R., Boswell, R.N., Shadduck, P., et al. (1990). Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science *249*, 932–935.

Laurent-Crawford, A.G., Krust, B., Rivière, Y., Desgranges, C., Muller, S., Kieny, M.P., Dauguet, C., and Hovanessian, A.G. (1993). Membrane expression of HIV envelope glycoproteins triggers apoptosis in CD4 cells. AIDS Res. Hum. Retroviruses *9*, 761–773.

Laux, H., Romand, S., Nuciforo, S., Farady, C.J., Tapparel, J., Buechmann-Moeller, S., Sommer, B., Oakeley, E.J., and Bodendorf, U. (2018). Degradation of recombinant proteins by Chinese hamster ovary host cell proteases is prevented by matriptase-1 knockout. Biotechnology and Bioengineering *0*.

Leem, J., Dunbar, J., Georges, G., Shi, J., and Deane, C.M. (2016). ABodyBuilder: Automated antibody structure prediction with data-driven accuracy estimation. MAbs *8*, 1259–1268.

Leng, Q., Bentwich, Z., Magen, E., Kalinkovich, A., and Borkow, G. (2002). CTLA-4 upregulation during HIV infection: association with anergy and possible target for therapeutic intervention. AIDS *16*, 519–529.

Lenting, P.J., van Mourik, J.A., and Mertens, K. (1998). The Life Cycle of Coagulation Factor VIII in View of Its Structure and Function. Blood *92*, 3983–3996.

Li, S., Yu, B., Byrne, G., Wright, M., O'Rourke, S., Mesa, K., and Berman, P.W. (2019a). Identification and CRISPR/Cas9 Knockout of the Endogenous C1s Protease in CHO Cells Eliminates Aberrant Proteolysis of Recombinantly Expressed Proteins. Biotechnology and Bioengineering.

Li, S.W., Yu, B., Byrne, G., Wright, M., O'Rourke, S., Mesa, K., and Berman, P.W. (2019b). Identification and CRISPR/Cas9 Inactivation of the C1s Protease Responsible for Proteolysis of Recombinant Proteins Produced in CHO Cells. Biotechnology and Bioengineering *116*, 2130–2145.

Li, Y., O'Dell, S., Walker, L.M., Wu, X., Guenaga, J., Feng, Y., Schmidt, S.D., McKee, K., Louder, M.K., Ledgerwood, J.E., et al. (2011). Mechanism of Neutralization by the Broadly Neutralizing HIV-1 Monoclonal Antibody VRC01. J. Virol. *85*, 8954–8967.

Liao, H.-X., Bonsignori, M., Alam, S.M., McLellan, J.S., Tomaras, G.D., Moody, M.A., Kozink, D.M., Hwang, K.-K., Chen, X., Tsao, C.-Y., et al. (2013). Vaccine

Induction of Antibodies against a Structurally Heterogeneous Site of Immune Pressure within HIV-1 Envelope Protein Variable Regions 1 and 2. Immunity *38*, 176–186.

Liberis, E., Velickovic, P., Sormanni, P., Vendruscolo, M., and Liò, P. (2018). Parapred: antibody paratope prediction using convolutional and recurrent neural networks. Bioinformatics *34*, 2944–2950.

Lim, A., Doyle, B.L., Kelly, G.M., Reed-Bogan, A.M., Breen, L.H., Shamlou, P.A., and Lambooy, P.K. (2018). Characterization of a cathepsin D protease from CHO cell-free medium and mitigation of its impact on the stability of a recombinant therapeutic protein. Biotechnology Progress *34*, 120–129.

Lind, P., Larsson, K., Spira, J., Sydow-Bäckman, M., Almstedt, A., Gray, E., and Sandberg, H. (1995). Novel Forms of B-Domain-Deleted Recombinant Factor VIII Molecules. European Journal of Biochemistry *232*, 19–27.

Lipson, E.J., and Drake, C.G. (2011). Ipilimumab: An Anti-CTLA-4 Antibody for Metastatic Melanoma. Clin Cancer Res *17*, 6958–6962.

Luo, H., Tie, L., Cao, M., Hunter, A.K., Pabst, T.M., Du, J., Field, R., Li, Y., and Wang, W.K. (2019). Cathepsin L Causes Proteolytic Cleavage of Chinese-Hamster-Ovary Cell Expressed Proteins During Processing and Storage: Identification, Characterization, and Mitigation. Biotechnology Progress *35*, e2732.

Markiewski, M.M., Nilsson, B., Ekdahl, K.N., Mollnes, T.E., and Lambris, J.D. (2007). Complement and coagulation: strangers or partners in crime? Trends in Immunology *28*, 184–192.

Mascola, J.R., and Haynes, B.F. (2013). HIV-1 neutralizing antibodies: understanding nature's pathways. Immunol Rev *254*, 225–244.

McLellan, J.S., Pancera, M., Carrico, C., Gorman, J., Julien, J.-P., Khayat, R., Louder, R., Pejchal, R., Sastry, M., Dai, K., et al. (2011). Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature *480*, 336–343.

Mechetina, L., Najakshin, A., Alabyev, B., Chikaev, N., and Taranin, A. (2002). Identification of CD16-2, a novel mouse receptor homologous to CD16/FcγRIII. Immunogenetics *54*, 463–468.

Mesa, K.A., Yu, B., Wrin, T., Petropoulos, C.J., Pogson, G.H., Alexander, D.L., Perez, G., O'Rourke, S.M., Sinangil, F., Robinson, J., et al. (2019). Ancestral sequences from an elite neutralizer proximal to the development of neutralization resistance as a potential source of HIV vaccine immunogens. PLOS ONE *14*, e0213409.

Meyer, L., López, T., Espinosa, R., Arias, C.F., Vollmers, C., and DuBois, R.M. (2019). A simplified workflow for monoclonal antibody sequencing. PLoS One *14*.

Montefiori, D.C., Karnasuta, C., Huang, Y., Ahmed, H., Gilbert, P., de Souza, M.S., McLinden, R., Tovanabutra, S., Laurence-Chenine, A., Sanders-Buell, E., et al. (2012). Magnitude and Breadth of the Neutralizing Antibody Response in the RV144 and Vax003 HIV-1 Vaccine Efficacy Trials. J Infect Dis *206*, 431–441.

Moore, P.L., Gray, E.S., Sheward, D., Madiga, M., Ranchobe, N., Lai, Z., Honnen, W.J., Nonyane, M., Tumba, N., Hermanus, T., et al. (2011). Potent and Broad Neutralization of HIV-1 Subtype C by Plasma Antibodies Targeting a Quaternary Epitope Including Residues in the V2 Loop. J Virol *85*, 3128–3141.

Morales, J.F., Morin, T.J., Yu, B., Tatsuno, G.P., O'Rourke, S.M., Theolis, R., Mesa, K.A., and Berman, P.W. (2014). HIV-1 Envelope Proteins and V1/V2 Domain Scaffolds with Mannose-5 to Improve the Magnitude and Quality of Protective Antibody Responses to HIV-1. J. Biol. Chem. jbc.M114.554089.

Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J.F., Halper-Stromberg, A., Gnanapragasam, P.N.P., Spencer, D.I.R., Seaman, M.S., et al. (2012). Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. Proc Natl Acad Sci U S A *109*, E3268–E3277.

Nakamura, G.R., Byrn, R., Rosenthal, K., Porter, J.P., Hobbs, M.R., Riddle, L., Eastman, D.J., Dowbenko, D., Gregory, T., Fendly, B.M., et al. (1992). Monoclonal Antibodies to the Extracellular Domain of HIV-1IIIB gp160 that

Neutralize Infectivity, Block Binding to CD4, and React with Diverse Isolates. AIDS Research and Human Retroviruses *8*, 1875–1885.

Nakamura, G.R., Byrn, R., Wilkes, D.M., Fox, J.A., Hobbs, M.R., Hastings, R., Wessling, H.C., Norcross, M.A., Fendly, B.M., and Berman, P.W. (1993). Strain specificity and binding affinity requirements of neutralizing monoclonal antibodies to the C4 domain of gp120 from human immunodeficiency virus type 1. J Virol *67*, 6179–6191.

Nimmerjahn, F., Lux, A., Albert, H., Woigk, M., Lehmann, C., Dudziak, D., Smith, P., and Ravetch, J.V. (2010). FcγRIV deletion reveals its central role for IgG2a and IgG2b activity in vivo. Proc Natl Acad Sci U S A *107*, 19396–19401.

Noris, M., and Remuzzi, G. (2013). Overview of Complement Activation and Regulation. Semin Nephrol *33*, 479–492.

O'Connell, R.J., Kim, J.H., Corey, L., and Michael, N.L. (2012). Human Immunodeficiency Virus Vaccine Trials. Cold Spring Harb Perspect Med *2*, a007351.

O'Rourke, S.M., Byrne, G., Tatsuno, G., Wright, M., Yu, B., Mesa, K.A., Doran, R.C., Alexander, D., and Berman, P.W. (2018a). Robotic selection for the rapid development of stable CHO cell lines for HIV vaccine production. PLOS ONE *13*, e0197656.

O'Rourke, S.M., Byrne, G., Tatsuno, G., Wright, M., Yu, B., Mesa, K.A., Doran, R.C., Alexander, D., and Berman, P.W. (2018b). Robotic selection for the rapid development of stable CHO cell lines for HIV vaccine production. PLoS ONE *13*, e0197656.

O'Rourke, S.M., Yu, B., Morales, J.F., Didinger, C.M., Alexander, D.L., Vollmers, C., and Berman, P.W. (2019). Production of a recombinant monoclonal antibody to Herpes Simplex Virus glycoprotein D for immunoaffinity purification of tagged proteins. J. Immunol. Methods *465*, 31–38.

Ozorowski, G., Pallesen, J., de Val, N., Lyumkis, D., Cottrell, C.A., Torres, J.L., Copps, J., Stanfield, R.L., Cupo, A., Pugach, P., et al. (2017). Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. Nature *547*, 360–363.

Pan, R., Gorny, M.K., Zolla-Pazner, S., and Kong, X.-P. (2015). The V1V2 Region of HIV-1 gp120 Forms a Five-Stranded Beta Barrel. J Virol *89*, 8003–8010.

Pancera, M., Shahzad-ul-Hussan, S., Doria-Rose, N.A., McLellan, J.S., Bailer, R.T., Dai, K., Loesgen, S., Louder, M.K., Staupe, R.P., Yang, Y., et al. (2013). Structural basis for diverse N-glycan recognition by HIV-1–neutralizing V1–V2–directed antibody PG16. Nature Structural & Molecular Biology *20*, 804–813.

Pang, S.S., Wijeyewickrema, L.C., Hor, L., Tan, S., Lameignere, E., Conway, E.M., Blom, A.M., Mohlin, F.C., Liu, X., Payne, R.J., et al. (2017). The Structural Basis for Complement Inhibition by Gigastasin, a Protease Inhibitor from the Giant Amazon Leech. The Journal of Immunology *199*, 3883–3891.

Park, J.H., Jin, J.H., Lim, M.S., An, H.J., Kim, J.W., and Lee, G.M. (2017). Proteomic Analysis of Host Cell Protein Dynamics in the Culture Supernatants of Antibody-Producing CHO Cells. Scientific Reports *7*, 44246.

Peach, R.J., Bajorath, J., Naemura, J., Leytze, G., Greene, J., Aruffo, A., and Linsley, P.S. (1995). Both Extracellular Immunoglobin-like Domains of CD80 Contain Residues Critical for Binding T Cell Surface Receptors CTLA-4 and CD28. J. Biol. Chem. *270*, 21181–21187.

Pejchal, R., Doores, K.J., Walker, L.M., Khayat, R., Huang, P.-S., Wang, S.-K., Stanfield, R.L., Julien, J.-P., Ramos, A., Crispin, M., et al. (2011). A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. Science *334*, 1097–1103.

Peng, R., Lin, G., and Li, J. (2016). Potential pitfalls of CRISPR/Cas9-mediated genome editing. The FEBS Journal *283*, 1218–1231.

Pentcheva-Hoang, T., Egen, J.G., Wojnoonski, K., and Allison, J.P. (2004). B7-1 and B7-2 Selectively Recruit CTLA-4 and CD28 to the Immunological Synapse. Immunity *21*, 401–413.

Pipe, S.W. (2009). Functional roles of the factor VIII B domain. Haemophilia *15*, 1187–1196.

Pitisuttithum, P., Gilbert, P., Gurwith, M., Heyward, W., Martin, M., van Griensven, F., Hu, D., and Tappero, J.W. (2006). Randomized, Double-Blind, Placebo-Controlled Efficacy Trial of a Bivalent Recombinant Glycoprotein 120 HIV-1 Vaccine among Injection Drug Users in Bangkok, Thailand. J Infect Dis *194*, 1661–1671.

Ploegh, H.L. (1998). Viral Strategies of Immune Evasion. Science *280*, 248–253.

Ponomarenko, E.A., Poverennaya, E.V., Ilgisonis, E.V., Pyatnitskiy, M.A., Kopylov, A.T., Zgoda, V.G., Lisitsa, A.V., and Archakov, A.I. (2016). The Size of the Human Proteome: The Width and Depth. Int J Anal Chem *2016*.

Pugach, P., Ozorowski, G., Cupo, A., Ringe, R., Yasmeen, A., de Val, N., Derking, R., Kim, H.J., Korzun, J., Golabek, M., et al. (2015). A native-like SOSIP.664 trimer based on an HIV-1 subtype B env gene. J. Virol. *89*, 3380–3395.

Quesada, V., Ordóñez, G.R., Sánchez, L.M., Puente, X.S., and López-Otín, C. (2009). The Degradome database: mammalian proteases and diseases of proteolysis. Nucleic Acids Res *37*, D239–D243.

Raska, M., Takahashi, K., Czernekova, L., Zachova, K., Hall, S., Moldoveanu, Z., Elliott, M.C., Wilson, L., Brown, R., Jancova, D., et al. (2010). Glycosylation

Patterns of HIV-1 gp120 Depend on the Type of Expressing Cells and Affect Antibody Recognition. J. Biol. Chem. *285*, 20860–20869.

Rawlings, N.D., Waller, M., Barrett, A.J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res *42*, D503–D509.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., et al. (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N. Engl. J. Med. *361*, 2209–2220.

Robert, F., Bierau, H., Rossi, M., Agugiaro, D., Soranzo, T., Broly, H., and Mitchell-Logean, C. (2009). Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. Biotechnology and Bioengineering *104*, 1132–1141.

Robertson, D.L. (2000). HIV-1 Nomenclature Proposal. Science *288*, 55d–555.

Robinson, H.L. (2018). HIV/AIDS Vaccines: 2018. Clin Pharmacol Ther *104*, 1062–1073.

Roche (2004). The Complete Guide for Protease Inhibition.

Roederer, M., Keele, B.F., Schmidt, S.D., Mason, R.D., Welles, H.C., Fischer, W., Labranche, C., Foulds, K.E., Louder, M.K., Yang, Z.-Y., et al. (2014).

Immunological and virological mechanisms of vaccine-mediated protection against SIV and HIV. Nature *505*, 502–508.

Rolland, M., Edlefsen, P.T., Larsen, B.B., Tovanabutra, S., Sanders-Buell, E., Hertz, T., deCamp, A.C., Carrico, C., Menis, S., Magaret, C.A., et al. (2012). Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env-V2. Nature *490*, 417–420.

Ronda, C., Pedersen, L.E., Hansen, H.G., Kallehauge, T.B., Betenbaugh, M.J., Nielsen, A.T., and Kildegaard, H.F. (2014). Accelerating Genome Editing in CHO Cells Using CRISPR Cas9 and CRISPy, a Web-Based Target Finding Tool. Biotechnol Bioeng *111*, 1604–1616.

Roth, G.A., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., et al. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet *392*, 1736–1788.

Sandberg, H., Lütkemeyer, D., Kuprin, S., Wrangel, M., Almstedt, A., Persson, P., Ek, V., and Mikaelsson, M. (2006). Mapping and partial characterization of proteases expressed by a CHO production cell line. Biotechnol. Bioeng. *95*, 961–971.

Scandella, C.J., Kilpatrick, J., Lidster, W., Parker, C., Moore, J.P., Moore, G.K., Mann, K.A., Brown, P., Coates, S., Chapman, B., et al. (1993). Nonaffinity Purification of Recombinant gp120 for Use in AIDS Vaccine Development. AIDS Research and Human Retroviruses *9*, 1233–1244.

Schulz, T.F., Reeves, J.D., Hoad, J.G., Tailor, C., Stephens, P., Clements, G., Ortlepp, S., Page, K.A., Moore, J.P., and Weiss, R.A. (1993). Effect of Mutations in the V3 Loop of HIV-1 gp120 on Infectivity and Susceptibility to Proteolytic Cleavage. AIDS Research and Human Retroviruses *9*, 159–166.

Shapiro, A.D. (2007). Anti-hemophilic factor (recombinant), plasma/albumin-free method (octocog-alpha; ADVATE®) in the management of hemophilia A. Vasc Health Risk Manag *3*, 555–565.

Shen, X., Laher, F., Moodie, Z., McMillan, A.S., Spreng, R.L., Gilbert, P.B., Huang, Y., Yates, N.L., Grunenberg, N., Juliana McElrath, M., et al. (2020). HIV-1 Vaccine Sequences Impact V1V2 Antibody Responses: A Comparison of Two Poxvirus Prime gp120 Boost Vaccine Regimens. Scientific Reports *10*, 1–14.

Shi, B., Ma, L., He, X., Wang, X., Wang, P., Zhou, L., and Yao, X. (2014). Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST. Theor Biol Med Model *11*, 30.

Simek, M.D., Rida, W., Priddy, F.H., Pung, P., Carrow, E., Laufer, D.S., Lehrman, J.K., Boaz, M., Tarragona-Fiol, T., Miiro, G., et al. (2009). Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. J. Virol. *83*, 7337–7348.

Sinacore, M.S., Drapeau, D., and Adamson, S.R. (2000). Adaptation of mammalian cells to growth in serum-free media. Mol Biotechnol *15*, 249–257.

Smith, D.H., Winters-Digiacinto, P., Mitiku, M., O'Rourke, S., Sinangil, F., Wrin, T., Montefiori, D.C., and Berman, P.W. (2010). Comparative Immunogenicity of HIV-1 Clade C Envelope Proteins for Prime/Boost Studies. PLoS One *5*.

Sok, D., Doores, K.J., Briney, B., Le, K.M., Saye-Francisco, K.L., Ramos, A., Kulp, D.W., Julien, J.-P., Menis, S., Wickramasinghe, L., et al. (2014). Promiscuous Glycan Site Recognition by Antibodies to the High-Mannose Patch of gp120 Broadens Neutralization of HIV. Science Translational Medicine *6*, 236ra63-236ra63.

Srinivasan, A., York, D., Butler, D., Jannoun-Nasr, R., Getchell, J., McCORMICK, J., Ou, C.-Y., Myers, G., Smith, T., Chen, E., et al. (1989). Molecular Characterization of HIV-1 Isolated from a Serum Collected in 1976: Nucleotide Sequence Comparison to Recent Isolates and Generation of Hybrid HIV. AIDS Research and Human Retroviruses *5*, 121–129.

Srivastava, I.K., Stamatatos, L., Legg, H., Kan, E., Fong, A., Coates, S.R., Leung, L., Wininger, M., Donnelly, J.J., Ulmer, J.B., et al. (2002). Purification and Characterization of Oligomeric Envelope Glycoprotein from a Primary R5 Subtype B Human Immunodeficiency Virus. Journal of Virology *76*, 2835–2847.

Srivastava, I.K., Stamatatos, L., Kan, E., Vajdy, M., Lian, Y., Hilt, S., Martin, L., Vita, C., Zhu, P., Roux, K.H., et al. (2003). Purification, Characterization, and Immunogenicity of a Soluble Trimeric Envelope Protein Containing a Partial Deletion of the V2 Loop Derived from SF162, an R5-Tropic Human Immunodeficiency Virus Type 1 Isolate. Journal of Virology *77*, 11244–11259.

Srivastava, I.K., Kan, E., Sun, Y., Sharma, V.A., Cisto, J., Burke, B., Lian, Y., Hilt, S., Biron, Z., Hartog, K., et al. (2008). Comparative evaluation of trimeric envelope glycoproteins derived from subtype C and B HIV-1 R5 isolates. Virology *372*, 273–290.

Stanfield, R.L., Gorny, M.K., Williams, C., Zolla-Pazner, S., and Wilson, I.A. (2004). Structural Rationale for the Broad Neutralization of HIV-1 by Human Monoclonal Antibody 447-52D. Structure *12*, 193–204.

Stephens, P.E., Clements, G., Yarranton, G.T., and Moore, J. (1990). A chink in HIV's armour? Nature *343*, 219.

Stolfa, G., Smonskey, M.T., Boniface, R., Hachmann, A.-B., Gulde, P., Joshi, A.D., Pierce, A.P., Jacobia, S.J., and Campbell, A. (2018). CHO-Omics Review:

The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based Bioproduction. Biotechnol J *13*, e1700227.

Stubbs, M.T., and Bode, W. (1994). Coagulation factors and their inhibitors. Curr. Opin. Struct. Biol. *4*, 823–832.

Sullivan, J.T., Sulli, C., Nilo, A., Yasmeen, A., Ozorowski, G., Sanders, R.W., Ward, A.B., Klasse, P.J., Moore, J.P., and Doranz, B.J. (2017). High-Throughput Protein Engineering Improves the Antigenicity and Stability of Soluble HIV-1 Envelope Glycoprotein SOSIP Trimers. J Virol *91*.

Terato, K., Do, C.T., Cutler, D., Waritani, T., and Shionoya, H. (2014). Preventing intense false positive and negative reactions attributed to the principle of ELISA to re-investigate antibody studies in autoimmune diseases. Journal of Immunological Methods *407*, 15–25.

Vaccari, M., Gordon, S.N., Fourati, S., Schifanella, L., Liyanage, N.P.M., Cameron, M., Keele, B.F., Shen, X., Tomaras, G.D., Billings, E., et al. (2016). Adjuvant-dependent innate and adaptive immune signatures of risk of SIVmac251 acquisition. Nat. Med. *22*, 762–770.

Valente, K.N., Lenhoff, A.M., and Lee, K.H. (2015). Expression of difficult-to-remove host cell protein impurities during extended Chinese hamster ovary cell culture and their impact on continuous bioprocessing. Biotechnology and Bioengineering *112*, 1232–1242.

Vouillot, L., Thélie, A., and Pollet, N. (2015). Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. G3 (Bethesda) *5*, 407–415.

Walker, L.M., Phogat, S.K., Chan-Hui, P.-Y., Wagner, D., Phung, P., Goss, J.L., Wrin, T., Simek, M.D., Fling, S., Mitcham, J.L., et al. (2009). Broad and Potent Neutralizing Antibodies from an African Donor Reveal a New HIV-1 Vaccine Target. Science *326*, 285–289.

Walker, L.M., Huber, M., Doores, K.J., Falkowska, E., Pejchal, R., Julien, J.-P., Wang, S.-K., Ramos, A., Chan-Hui, P.-Y., Moyle, M., et al. (2011). Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature *477*, 466–470.

Walport, M.J., Davies, K.A., and Botto, M. (1998). C1q and systemic lupus erythematosus. Immunobiology *199*, 265–285.

Wang, H., Barnes, C.O., Yang, Z., Nussenzweig, M.C., and Bjorkman, P.J. (2018). Partially Open HIV-1 Envelope Structures Exhibit Conformational Changes Relevant for Coreceptor Binding and Fusion. Cell Host & Microbe *24*, 579-592.e4.

Waritani, T., Chang, J., McKinney, B., and Terato, K. (2017). An ELISA protocol to improve the accuracy and reliability of serological antibody assays. MethodsX *4*, 153–165.

Warren, D.L., Morrissey, J.H., and Neuenschwander, P.F. (1999). Proteolysis of Blood Coagulation Factor VIII by the Factor VIIa−Tissue Factor Complex: Generation of an Inactive Factor VIII Cofactor †. Biochemistry *38*, 6529–6536.

Weissman, D., Rabin, R.L., Arthos, J., Rubbert, A., Dybul, M., Swofford, R., Venkatesan, S., Farber, J.M., and Fauci, A.S. (1997). Macrophage-tropic HIV and SIV envelope proteins induce a signal through the CCR5 chemokine receptor. Nature *389*, 981–985.

Wen, Y., Trinh, H.V., Linton, C.E., Tani, C., Norais, N., Martinez-Guzman, D., Ramesh, P., Sun, Y., Situ, F., Karaca-Griffin, S., et al. (2018). Generation and characterization of a bivalent protein boost for future clinical trials: HIV-1 subtypes CR01_AE and B gp120 antigens with a potent adjuvant. PLoS One *13*.

Werner, A., and Levy, J.A. (1993). Human immunodeficiency virus type 1 envelope gp120 is cleaved after incubation with recombinant soluble CD4. J Virol *67*, 2566–2574.

Wibmer, C.K., Richardson, S.I., Yolitz, J., Cicala, C., Arthos, J., Moore, P.L., and Morris, L. (2018). Common helical V1V2 conformations of HIV-1 Envelope expose the α4β7 binding site on intact virions. Nature Communications *9*, 1–14.

Wiehe, K., Easterhoff, D., Luo, K., Nicely, N.I., Bradley, T., Jaeger, F.H., Dennison, S.M., Zhang, R., Lloyd, K.E., Stolarchuk, C., et al. (2014). Antibody

Light-Chain-Restricted Recognition of the Site of Immune Pressure in the RV144 HIV-1 Vaccine Trial Is Phylogenetically Conserved. Immunity *41*, 909–918.

Wolchok, J.D., Hodi, F.S., Weber, J.S., Allison, J.P., Urba, W.J., Robert, C., O'Day, S.J., Hoos, A., Humphrey, R., Berman, D.M., et al. (2013). Development of ipilimumab: a novel immunotherapeutic approach for the treatment of advanced melanoma. Ann N Y Acad Sci *1291*, 1–13.

Worton, R.G., Ho, C.C., and Duff, C. (1977). Chromosome stability in CHO cells. Somatic Cell Genet. *3*, 27–45.

Wu, L., Gerard, N.P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A.A., Desjardin, E., Newman, W., et al. (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. Nature *384*, 179–183.

Wu, X., Yang, Z.-Y., Li, Y., Hogerkorp, C.-M., Schief, W.R., Seaman, M.S., Zhou, T., Schmidt, S.D., Wu, L., Xu, L., et al. (2010). Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. Science *329*, 856–861.

Wurm, F.M., and Wurm, M.J. (2017). Cloning of CHO Cells, Productivity and Genetic Stability—A Discussion. p.

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., et al. (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology *29*, 735–741.

Yeap, W.H., Wong, K.L., Shimasaki, N., Teo, E.C.Y., Quek, J.K.S., Yong, H.X., Diong, C.P., Bertoletti, A., Linn, Y.C., and Wong, S.C. (2016). CD16 is indispensable for antibody-dependent cellular cytotoxicity by human monocytes. Scientific Reports *6*, 1–22.

Yu, B., Fonseca, D.P.A.J., O'Rourke, S.M., and Berman, P.W. (2010). Protease Cleavage Sites in HIV-1 gp120 Recognized by Antigen Processing Enzymes Are Conserved and Located at Receptor Binding Sites. J Virol *84*, 1513–1526.

Yu, B., Morales, J.F., O'Rourke, S.M., Tatsuno, G.P., and Berman, P.W. (2012). Glycoform and Net Charge Heterogeneity in gp120 Immunogens Used in HIV Vaccine Trials. PLOS ONE *7*, e43903.

Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J.A., Schroeder, H.W., and Kirkham, P.M. (2003). Expressed Murine and Human CDR-H3 Intervals of Equal Length Exhibit Distinct Repertoires that Differ in their Amino Acid Composition and Predicted Range of Structures. Journal of Molecular Biology *334*, 733–749.

Zhou, T., Georgiev, I., Wu, X., Yang, Z.-Y., Dai, K., Finzi, A., Kwon, Y.D., Scheid, J.F., Shi, W., Xu, L., et al. (2010). Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. Science *329*, 811–817.

Zipfel, P.F., Würzner, R., and Skerka, C. (2007). Complement evasion of pathogens: Common strategies are shared by diverse organisms. Molecular Immunology *44*, 3850–3857.

Zolla-Pazner, S., Zhong, P., Revesz, K., Volsky, B., Williams, C., Nyambi, P., and Gorny, M.K. (2004). The Cross-Clade Neutralizing Activity of a Human Monoclonal Antibody Is Determined by the GPGR V3 Motif of HIV Type 1. AIDS Research and Human Retroviruses *20*, 1254–1258.