# Methods for development and characterization of HIV-1 envelope immunogens

DISSERTATION ZUR ERLANGUNG
DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR
BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG


Vorgelegt von
Benjamin Zimmer
aus
Lauingen (Donau)

im Jahr
2019

# Methods for development and characterization of HIV-1 envelope immunogens

DISSERTATION ZUR ERLANGUNG
DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR
BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG


Vorgelegt von
Benjamin Zimmer
aus
Lauingen (Donau)

im Jahr
2019

Das Promotionsgesuch wurde eingereicht am:
28.02.2019

Die Arbeit wurde angeleitet von:
Prof. Dr. Ralf Wagner

---

Benjamin Zimmer

Meiner Familie

# Content declaration

The work presented in this thesis was undertaken in the laboratory of Prof. Dr. Ralf Wagner at the Institute of Medical Microbiology and Hygiene, University Hospital Regensburg, Germany, between 2014 and 2018. I, Benjamin Zimmer, hereby declare that all the work presented was done by my own with the following exceptions. All MicroScale Thermophoresis experiments and their evaluation were performed by the 2bind molecular interactions GmbH (Am Biopark 11, 93053 Regensburg, Germany). The production of virus-like particles (see **4.1.3**) and their analysis by ELISA (see **4.1.6** and **4.1.10**) was done by Tobias Fischer during his doctoral thesis, performed under my experimental supervision. Poly(lactic acid) particles for MST and ELISA analyses were produced and characterized in the laboratory of Bernard Verrier (Institute of Biology and Chemistry of Proteins (IBCP), University of Lyon, France). The design and development of the BG505 SOSIP gp160 alanine substitution library (see **4.2**) was performed together with Dr. David Peterhoff (laboratory of Prof. Dr. Ralf Wagner). Mapping experiments with the BG505 SOSIP gp160 alanine substitution library (see **4.2.3**) were performed by Iris Ganser during her master`s thesis, which was supervised by Dr. David Peterhoff. Fluorescence-activated cell sorting was performed at the Central FACS Facility (University Hospital Regensburg) of the Regensburg Center for Interventional Immunology (RCI). Next generation sequencing was performed in the laboratory of Prof. Dr. Gunter Meister (Department of Biochemistry I, University of Regensburg, Germany) under the instruction of Norbert Eichner. NGS data were evaluated by an analyzer tool, programmed by Dr. Benedikt Asbach (laboratory of Prof. Dr. Ralf Wagner). This work has not been submitted for any other degree at this university or any other institute of learning. Where required, permission from individual journals for usage of figures was obtained.

# Contents

# Abstract

Currently, most preventive vaccine strategies against the human immunodeficiency virus-1 (HIV-1) focus on the induction of broadly neutralizing antibodies (bNAbs) targeting the viral envelope glycoprotein (Env). However, all attempts to elicit bNAbs by vaccination remained unsuccessful so far. The development of stabilized Env trimers is considered as first success on the road towards broader neutralizing antibody responses. Characterized by a favorable antigenicity profile, meaning the efficient presentation of bNAb epitopes and occlusion of non-neutralizing epitopes, these trimers for the first time elicited autologous tier 2 neutralization in animal models. Novel approaches, including particle-based Env delivery and development of improved Env immunogens are considered important steps to broaden humoral immune responses. This thesis investigated methods for the characterization and development of improved Env immunogens. The first project of this thesis investigated the potential of the biophysical in-solution method MicroScale Thermophoresis (MST) for the analysis of soluble and particle-presented Env proteins under artificial (buffer) and physiological conditions (serum). Using two model systems, 1) a five-member Env/V3 chimeric library with distinct binding capacities towards the monoclonal antibody 447-52D, and 2) stabilized BG505 SOSIP.664 trimers and a panel of quaternary structure-characterizing antibodies, it has been demonstrated that MST is useful tool for the analysis of soluble and particle-presented Env immunogens under artificial and near-physiological conditions (50 % serum). Thereby, results were in general agreement with data obtained from enzyme-linked immunosorbent assay (ELISA), validating MST as a fast, low-sample consuming method for the analysis of Env immunogens. The second project focused on the development of a mammalian cell display approach for mapping of bNAb epitopes, providing valuable information for rational vaccine design. For this purpose, a BG505 SOSIP gp160 based alanine substitution library was developed and applied to proof-of-principle screenings with well-characterized bNAbs. Here, the library demonstrated its capacity for epitope mapping. However, further improvements are necessary to turn the library into a powerful mapping tool. The third project aimed to investigate the influence of single glycan deletions on binding of bNAbs. For this purpose, a cellular BG505 SOSIP gp145 glycan-knockout library was generated and applied to fluorescence-activated cell sorting (FACS). An initial screening with the glycan-dependent bNAb PGT135 resulted in a 35-fold enrichment of the Env variant

N332A in the applied low-affinity gate. This demonstrates that the library could be a useful to tool to identify glycans that contribute to binding of bNAbs.

# Zusammenfassung

Aktuell zielen die meisten präventiven Impfstrategien gegen das humane Immundefizienz-Virus 1 (HIV-1) darauf ab breit-neutralisierende Antikörper (bNAbs) gegen das virale Hüllprotein (Env) zu induzieren. Trotz aller Anstrengungen die diesbezüglich unternommen wurden, blieb die Entwicklung eines Schutz-vermittelnden Vakzins bisher erfolglos. Die Entwicklung von stabilisierten Env Trimeren, welche sich durch ein vorteilhaftes Immunogenitätsprofil auszeichnen – effektive Präsentation von bNAb Epitopen bei gleichzeitiger Vermeidung der Exposition von nicht-neutralisierenden Antikörper Epitopen – wird als erster Erfolg auf dem Weg zu einem schützenden Vakzin angesehen. In Tierstudien, in denen stabilisierte Env Trimere als Immunogene verabreicht wurden, konnte erstmals eine Neutralisation von autologen tier 2 Viren erreicht werden. Basierend auf diesen Erkenntnissen sollen neue innovative Ansätze, wie beispielsweise eine Partikel-vermittelte Env Präsentation und die Entwicklung von neuartigen Env Immunogenen, dazu beitragen eine breitere humorale Immunantwort zu induzieren. Die Ziele dieser Arbeit waren, zu untersuchen welchen Beitrag biophysikalische und Zytometer-basierte Methoden zur Charakterisierung und Entwicklung von verbesserten Env Immunogenen leisten können. Im ersten Teil der Arbeit sollte untersucht werden, ob sich die biophysikalische Messmethode Microscale Thermophoresis (MST) für die Charakterisierung von Antikörper-Env Interaktionen eignet. Hierbei sollte insbesondere überprüft werden, ob die MST zur Analyse von Partikel-präsentierten Env Proteinen herangezogen werden kann. Zudem sollte die Anwendbarkeit der MST unter physiologischen Bedingungen (Serum) getestet werden. Hierzu wurden zwei Modellsysteme verwendet: 1) Eine Modell-Bibliothek bestehend aus fünf chimären Env Proteinen (Env/V3) mit definiertem Bindungsprofil zum monoklonalen Antikörper 447-52D und 2) stabilisierte BG505 SOSIP.664 Trimere und ein Set von Quartärstruktur-charakterisierenden Antikörpern. Die Ergebnisse zeigten, dass die MST in der Lage ist Interaktionen zwischen löslichen und Partikel-präsentierten Env Proteinen und Antikörpern zu analysieren. Dies gilt sowohl für Messungen unter artifiziellen Bedingungen (Puffer) als auch für Messungen unter nahezu physiologischen Bedingungen (50 % Serum). Die Ergebnisse aus den MST Analysen stimmten dabei im Großen und Ganzen mit Ergebnissen aus *enzyme-linked immunosorbent assay* (ELISA) Experimenten überein, wodurch die MST als schnelle und zugleich sparsame Methode zur Analyse von Env Immunogenen validiert werden konnte. Der zweite Teil der Arbeit konzentrierte sich auf die Entwicklung eines auf Säugetierzellen

basierenden *Display* Verfahrens zu Kartierung von bNAb Epitopen. Zu diesem Zweck wurde eine auf dem BG505 SOSIP gp160 Env basierende Alanin-Substitutionsbibliothek entwickelt, welche mit einem Set von bekannten bNAbs charakterisiert wurde. Hierbei zeigte sich sowohl, dass die Bibliothek prinzipiell für Kartierungsexperimente herangezogen werden kann, als auch, dass noch einige Verbesserungen nötig sind, um aus der Bibliothek in ein robustes und präzises Instrument zur Antikörperkartierung zu generieren. Der dritte Teil der Arbeit zielte darauf ab, den Einfluss einzelner Glykan-Deletionen auf die Bindung von bNAbs zu untersuchen. Zu diesem Zweck wurde eine zelluläre BG505 SOSIP gp145-Glykan-*Knockout*-Bibliothek generiert. Diese wurde mittels fluoreszenz-aktivierter Zellsortierung (FACS) auf ihre Bindung zum glykan-abhängigen bNAb PGT135 hin untersucht. Hierbei konnte eine 35-fache Anreicherung der Env Variante N332A im *gate* mit niedriger Affinität zum PGT135 Antikörper beobachtet werden. Dieses Ergebnis zeigt, dass die Bibliothek zukünftig dazu verwendet werden könnte, Glykane zu identifieren die einen Beitrag zur Bindung von bNAbs leisten.

# 1. Introduction

## 1.1 The HIV-1 pandemic

Since its identification in the 1980`s HIV-1, a retrovirus in the genus *Lentiviridae* and causative agent of the acquired immunodeficiency syndrome (AIDS), spread all over the world and became one of the most devastating infectious diseases in recent history[1,2]. The HIV-1 infection is characterized by a progressive depletion of T lymphocytes, carrying the cluster of differentiation 4 (CD4) receptor. The loss of $CD4^+$ T lymphocytes, impairing the cellular immune response, is in close association with an increased susceptibility of infected individuals to opportunistic infections, defining AIDS[3].

The origins of the pandemic are located in west central Africa, were HIV-1 emerged from four cross-species transmission events of simian immunodeficiency viruses (SIVs) to humans, most likely by exposure to infected ape body fluids. The four independent transmission events resulted in four HIV-1 linages, termed groups M (major), N (non-M/non-O), O (outlier) and P (pending). Phylogenetic tree analysis showed, that group M and N are closely related with SIVcpz*Ptt* strains, indicating that they are of chimpanzee origin. Origins of groups P and O are not fully understood, but phylogenetic data support a gorilla origin of both, group P and O[1,4]. Out of the four groups described above, M is responsible for the HIV-1 pandemic. Phylogenetic and statistical analyses indicate that group M emerged in the beginning of the twentieth century, most likely in the area around Kinshasa and it is assumed that early diversification of group M also took place in this area, from which the infection spread globally. Currently, group M lineages are classified by nine subtypes (A–D, F–H, J, K), nowadays predominating in different geographic areas, resulting from a number of founder events[1], see **Figure 1**. Furthermore, there exists a growing number of currently about 100 circulating recombinant forms (CRFs), resulting from multiple infections with different subtypes in the same population[5,1]. Within group M, subtype C accounts for almost 50% of the infections, and thus represents the dominant subtype in the actual HIV-1 pandemic, followed by subtypes A (12 %) and B (11 %), CRF02_AG (8 %), CRF01_AE (5 %), subtypes G (5 %) and D (2 %)[6], as shown in **Figure 1**.

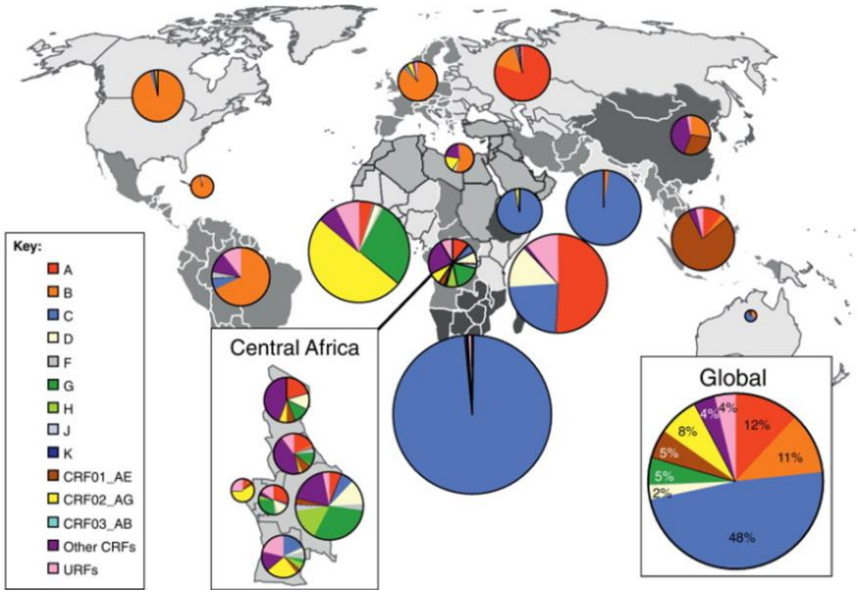



**Figure 1. Geographic distribution of HIV-1 subtypes and recombinants.** Pei charts illustrate the regional percentage distribution of HIV-1 subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs), represented by different colors. The distribution was calculated according to data collected from 2004-2007. Figure adapted from Hemelaar, J. The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med*. 18, 182–92 (2012)[7], with permission from Elsevier, license number 4536020267052.

According to the Joint United Nations Programme on HIV/AIDS (UNAIDS) 77.3 million [59.9–100 million] people have become infected with HIV-1 since the start of the epidemic from who 35.4 million [25.0–49.9 million] died due to AIDS-related illness. In 2017, there were 36.9 million [31.1 million–43.9 million] people living with HIV, with sub-Saharan Africa accounting for about 70 % of the global burden of infection. With major efforts, the number of new infections has been steadily reduced from 3.4 million [2.6–4.4 million] in 1996 (peak of new infections) to 1.8 million [1.4–2.4 million] in 2017, meaning a reduction of 47 %. Furthermore, AIDS-related deaths have been reduced by more than 51 % since the peak in 2004[8], see **Figure 2**.

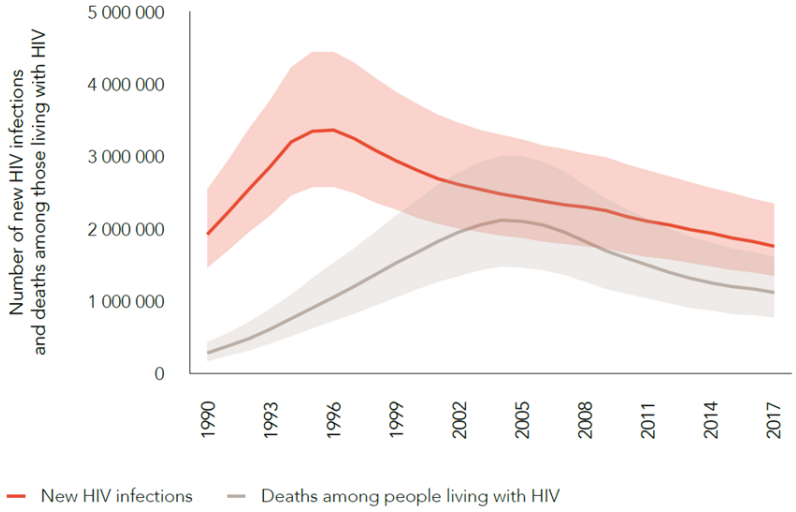



**Figure 2. Number of new HIV infections and deaths among the HIV population.** Global data from 1990-2017, adapted from UNAIDS[9].

Ambitious prevention programs, significant progress in the development of antiretroviral drugs and modern treatment strategies went hand in hand and contributed to this success. However, the story of success in anti-retroviral therapy included three important therapeutic revolutions. The first revolution was the introduction of new classes of antiretroviral drugs in 1995/1996, including protease inhibitors and non-nucleoside reverse-transcriptase inhibitors. These newly introduced anti-retroviral drugs in combination with nucleoside-analog reverse-transcriptase inhibitors, which until then were the only treatment option available, greatly improved the prognosis of an otherwise lethal disease. The combined antiretroviral therapy is nowadays referred to as cART and in 2018 there were about 40 antiretroviral drugs available, including 7 drug classes, which were approved by the US Food and Drug Administration (FDA)[10,11,12]. The second revolution was marked by the realization that HIV-1 treatment contributes to prevention, since studies showed that treatment in serodiscordant couples (one HIV seronegative and one HIV seropositive partner) reduced the risk of sexual transmission by 96 %. The third revolution arised when first studies showed that early cART treatment, also if the number $CD4^+$ cells is more than 500 cells per cubic millimeter, is associated with a significant clinical benefit in terms of a 57 % reduced risk in AIDS and non-AIDS morbidity and mortality[13,14,11]. The findings of these studies had a lasting impact on treatment guidelines. For example, while in the 2013 World Health Organisation (WHO) guidelines a $CD4^+$ cell count of less than or equal to 500 cells per cubic millimeter was one of the criteria for starting therapy, recommendations changed in the 2016 WHO guidelines where therapy for all infected individuals is recommended, regardless of the $CD4^+$ cell count. This is known

as the test and treat strategy[15,16,11]. Further, in 2014 UNAIDS launched the 90-90-90 targets meaning that by 2020 90 % of all people living with HIV know their HIV status, 90 % of people with diagnosed HIV infection will receive sustained cART, and 90% of those on cART will be virally suppressed. These targets were extended to 95-95-95 by 2030. Modeling analyses suggest that achieving these ambitious goals by 2020, enables ending of the epidemic in 2030[17,18]. However, another study predicts that if current diagnosis, treatment and suppression levels are maintained (53-75-77, when the study was published in 2017) a global median of 49 million new infections will occur between 2015 and 2035. The same study predicts that from the 49 million infections 22 million can be averted by reaching 90-90-90 and a further 3.3 million by reaching 95-95-95 targets. Furthermore, the study predicted that a 50 % efficacy vaccine alone, rolled out in 2020 with an annual scale-up of 25 % coverage to a maximum coverage of 70 %, and under current diagnosis and treatment levels, would avert 17 million new infections[19], see **Figure 3**. This highlights that even a moderate efficacy vaccine could significantly contribute to the end the HIV-1 pandemic.

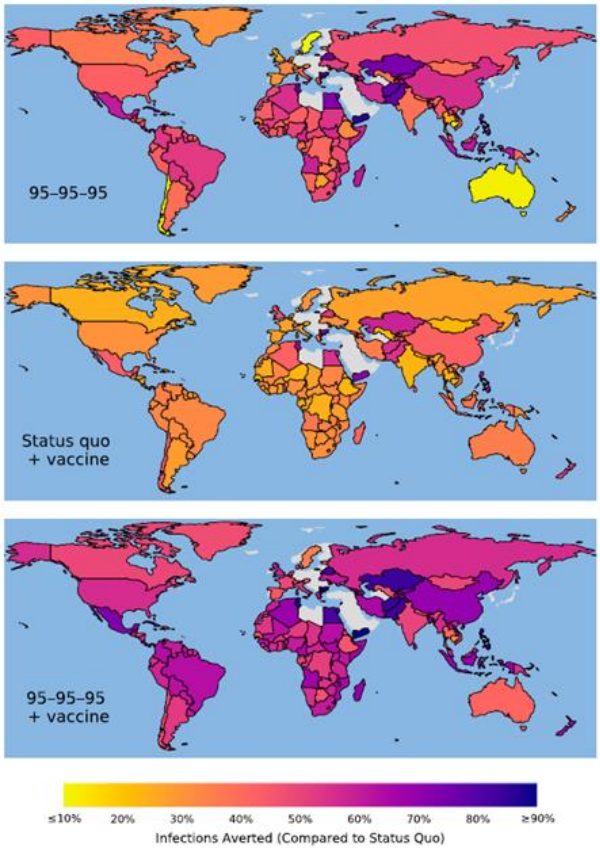



**Figure 3. Calculated averted infections between 2015 and 2030 compared with maintaining status quo diagnosis, treatment and viral suppression levels.** Averted infections by reaching UNAIDS 95-95-95 targets (95 % of infected people know their HIV status, 95 % with diagnosed infection receive cART, 95% of people receiving cART are virally suppressed) **(Top)**, rollout of a vaccine in 2020 **(Middle)** and 95-95-95 targets combined with vaccine enrollment in 2020 **(Bottom)**. Data are based on diagnosis, treatment and viral suppression levels of 53-75-77. Figure from Medlock, J. *et al.* Effectiveness of UNAIDS targets and HIV vaccination across 127 countries. *Proc. Natl. Acad. Sci. U. S. A.* 114, 4017–4022 (2017)[19].

In summary, major efforts in prevention programs and significant progress in anti-retroviral therapy contributed reduce the number of new infections and AIDS-related death. This is also reflected by the current UNAIDS global data, since in 2017 75 % of people living with HIV know their status, 79 % of people living with HIV who know their status are accessing antiretroviral therapy and 81 % of people accessing treatment had suppressed viral loads[8]. However, a strategy with focus on anti-retroviral therapy has some drawbacks since it requires daily medication, a high level of compliance and medical monitoring, still representing major challenges in low-income countries. As described above, a prophylactic vaccine could significantly contribute to prevent further infections with a still not curable disease and therefore the development of such a vaccine is of utmost importance.

## 1.2 HIV-1 vaccine clinical trials

Soon after the discovery of HIV it was expected that an effective vaccine would be rapidly developed but so far only six HIV-1 vaccine candidates have been tested in clinical efficacy trials, with no liscenced vaccine yet, see **Figure 4**. However the story of vaccine paradigms and clinical trials can be separated in three overlapping "waves"[20,21].

The first wave of HIV vaccines was based on the concept that neutralizing antibody responses would be sufficient to protect from infection. Since the viral envelope glycoprotein is the mediator of host cell infection and therefore a main target of a neutralizing antibody response, vaccine development in late 1980s to mid-1990s focused on recombinant Env immunogens. These first recombinant Env immunogens were primarily produced as glycoprotein (gp) 120 proteins, representing soluble monomeric subunits of the Env protein, that is naturally present as trimer of heterodimers on the surface HIV virions, see **1.3.1**[20,22]. From 1998 to 2003, two phase 3 efficacy trials with recombinant bivalent gp120 vaccines were carried out. However, these studies known as VAX004 (gp120 subtype B/B) and VAX003 (gp120 subtype B/E) failed to show efficacy, see **Figure 4**. Although both vaccines induced high titers of neutralizing antibodies (NAbs), these were of limited specificity and therefore could not prevent HIV infection[20,21,23,24].

| Study | Regimen | Participants | Aim | Outcome |
|---|---|---|---|---|
| VAX004 (United States, Netherlands) | *r*gp120 B/B | MSM, high-risk women | bNAbs | No prevention of HIV infection |
| VAX003 (Thailand) | *r*gp120 B/E | Drug users | bNAbs | No prevention of HIV infection |
| Step/HVTN502 (USA) | *r*Ad5 HIV-1 gag/pol/nef B | MSM, high-risk women | $CD8^+$ T-cells | Increased infection risk |
| Phambili/HVTN503 (South Africa) | *r*Ad5 HIV-1 gag/pol/nef B | Heterosexual men, women | $CD8^+$ T-cells | Increased infection risk |
| HVTN505 | * DNA/*r*Ad5 | MSM, transgender women | Ab and T-cells | No infection risk, no efficacy |
| RV144 (Thailand) | * ALVAC-HIV/AIDSVAX B/E gp120 in alum | High risk men and women | Ab and T-cells | 31.2 % vaccine efficacy |

**Figure 4. HIV vaccine clinical trials conducted until today.** Abbreviations: recombinant (r), men who have sex with men (MSM), broadly neutralizing antibodies (bNAbs), prime-boost regimen (*), antibodies (Ab). Freely reproduced from Trovato, M. *et al*. HIV Vaccination: A Roadmap among Advancements and Concerns. *Int. J. Mol. Sci.* 19, 1241 (2018)[21].

The second wave of HIV vaccines started in the early 2000s when the failure of recombinant Env vaccines and new findings, supporting the importance of $CD8^+$ T-cells in controlling viral replication, focused the field towards cytotoxic T-lymphocytes (CTLs)[20]. Two phase 2b vaccine efficacy trials were conducted to explore if cell-mediated immune responses can confer protection from infection or reduce plasma viral load after infection. In both, the Step study (HIV Vaccine Trials Network (HVTN) 502, started in 2004) and the Phambili trial (HVTN503, started in January 2007), see **Figure 4**, vaccinees were immunized with the MRKAd5 HIV-1 vaccine. This vaccine, consisting of a mixture of replication-defective recombinant adenovirus type 5 (Ad5) vectors expressing HIV-1 group-specific antigen (*gag*), polymerase (*pol*) and negative regulatory factor (*nef*) subtype B genes, induced cell-mediated immune responses in a phase I clinical trial[25,26,27]. However, after the first interim analysis of the Step study in September 2007 both studies were stopped, since, despite of eliciting interferon (IFN)-γ enzyme-linked immuno spot assay (ELISPOT) responses in 75 % of the vaccinees, the vaccine did no prevent HIV infection. Instead, analysis of available data revealed an increased risk of infection for male vaccinees who were uncircumsized or Ad5 seropositive pre-vaccination[21,28,29,30].

The third wave began with vaccine efficacy trials that attempted to stimulate both arms of the immune response to ideally induce both, protective antibody and protective cell-mediated immune responses. Until today, two of those studies were conducted and evaluated. From

2009 to 2013 the HVTN enrolled a phase 2b efficacy trial called HVTN505, see **Figure 4**. To adress the viral diversity, the vaccine regimen was based on a heterologous prime/boost approach including immunogens from multiple HIV-1 subtypes. The vaccine regimen consisted of a DNA prime (six DNA plasmids expressing HIV-1 clade B Gag, Pol, and Nef and Env proteins from clades A, B, and C) and a rAd5 boost (four rAd5 vectors expressing a HIV-1 clade B Gag-Pol fusion protein and Env proteins from clades A, B, and C). However, despite of inducing both, cellular and humoral immune responses, the vaccine regimen did not show efficacy in preventing HIV infection[21,31,32]. The RV144 phase 3 clinical trial, conducted in Thailand between 2003 and 2005 and enrolling a total of 16,402 volunteers, was the only clinical trial that demonstrated an, albeit modest, efficacy in preventing HIV infection, see **Figure 4**. The study tested a heterologous prime/boost regimen consisting of two vaccines, matching the in Thailand predominantly circulating HIV-1 subtypes B and CRF01_AE: 1) ALVAC-HIV (vCP1521), a canarypox vector expressing HIV-1 CRF01_AE gp120 Env (linked to the transmembrane-anchoring portion of subtype B gp41) and HIV-1 subtype B Gag and protease was used for priming; and 2) AIDSVAX B/E (same vaccine as in the VAX003 study) formulated with alum adjuvant was used for boosting. Prior to RV144 study, this prime/boost regimen was tested in a phase 2 trial were it induced both, cellular and humoral immune responses and was therefore qualified for advanced testing in the RV144 trial[33,34]. The results of the RV144 trial, published in 2009, demonstrated 31.2 % vaccine efficacy in preventing HIV aquisition after 3.5 years. Further, post-hoc analyses reported that even higher vaccine efficacies (60.5 %) occurred in the first year following vaccination and declined over time[33,35,36]. Analysis of immune correlates, reviewed in[36,37], revealed that the protective effect primarily correlated with non-neutralizing immunoglobulin G (IgG) antibodies targeting the variable 1/variable 2 (V1/V2) regions of the Env protein (see **1.3.1**). The protective effect of the antibody response was most likely mediated by Fc-receptor functions, including antibody-dependent cellular cytotoxicity (ADCC)[38,37]. Further, Env-specific IgG3 immune responses correlated with a decreased risk of HIV transmission. Recently, a study comparing matched Env IgG1 and Env IgG3 antibodies reported that the IgG3 subclass is more potent than the IgG1subclass in mediating virion internalization. This maybe a mechanism by which non-neutralizing Env-specific IgG1 and IgG3 antibody responses contributed to vaccine efficacy in the RV144 trial[39,36]. Polyfunctional $CD4^+$ T-cell responses also correlated with a decreased risk of HIV infection. Two T-cell subsets were identified, significantly correlating with a decreased risk of infection. One subset expressed CD40 ligand (CD40L), interleukin (IL)-2, IL-4, IFN-γ and tumor necrosis factor (TNF)-α,

while the other expressed CD40L, IL-2 and IL-4. Notably, both subsets include CD40L and IL4, which are involved $CD4^+$ T-cell/B-cell interactions, indicating that these T-cell subsets may contribute T-cell help necessary for antibody production[40,36]. Next to correlates of protection in RV144 described above, there was also a strong direct correlate increasing risk of infection. Studies reported that high levels of serum IgGA, binding to specific envelopes increased risk of infection. Further, it has been shown that the Env-specific IgA/IgG ratio directly correlates with infection risk, most likely due to interfering events, since the same study reported that Env specific IgA can do both, inhibit Env binding and block ADCC activity of vaccine-induced IgG targeting the same epitope[41,42,37,36]. Based on the promising results from the RV144 trial a number of follow-on trials were designed, to test if the RV144 regimen can induce protective immune responses in South Africa. For this purpose the RV144 regimen was adapted to subtype C using ALVAC-HIV (vCP2438) in combination with recombinant bivalent subtype C gp120 formulated in MF59 adjuvant. This combination is currently tested in a phase 2b/3 efficacy trial (HVTN702) which was launched in 2016 in South Africa[43,21].

When the results of the RV144 study were published in 2009, the focus of vaccine design was redirected from cell-mediated immune responses towards the humoral immune response, since antibodies, although not neutralizing, were correlated with protection from HIV infection. In parallel, studies of humoral immune responses of HIV-positive individuals using novel high-throughput screening methods resulted in the identification of scores of Env-directed broadly neutralizing antibodies (bNAbs) since beginning of 2009[44]. BNAbs are characterized by potent and broad neutralizing activity (described in detail in **1.4.2**), and the attempts to induce a bNAb response by vaccination further focused vaccine design towards the Env protein and dominate the field until today.

## 1.3 The HIV-1 envelope glycoprotein

The viral envelope glycoprotein (Env) plays a key role in HIV-1 infection, since it mediates the infection of $CD4^+$ host cells, and thus enables viral replication. In addition, Env is the only viral protein on the surface of HIV virions and therefore the only target of the humoral immune response. Thus, the Env protein represents a key component in HIV-1 vaccine design[45,46,47].

### 1.3.1 Biosynthesis and structure of the HIV-1 envelope protein

The HIV-1 Env protein is expressed as gp160 precursor polypeptide that gets extensively glycosylated during synthesis in the endoplasmatic reticulum (ER). Thereby, glycans account for about the half of the molecular mass of the protein. Glycosylation of Env proteins occurs as N-linked glycosylation at the asparagine of NXT/S sequons, with X being any amino acid (AA) except possibly proline or aspartic acid[48,49]. In addition, there are studies reporting some rare O-linked glycosylations of Env proteins, which can occur at serine or threonine residues. But contrary to N-linked glycosylation sites, there is no clear-cut motif in the amino acid sequence which defines their glycosylation[50,51,52]. The gp160 precursor is proteolytically processed by Golgi-associated serin proteases of the Furin family, cleaving at the C-terminus of the amino acid consensus sequence REKR, into non-covalently linked gp120/gp41 subunits, called protomers[53,54]. Three gp120 / gp41 protomers assemble into Env trimers that are transported to the plasma membrane, and thus incorporated into virions. From there, they can mediate host cell entry by a series of complex receptor-triggered conformational changes[55,56], see **1.3.2**.

During the infection of host cells, gp120 and gp41 have different functions, which are defined by their structure. The gp120 subunit, including CD4 and co-receptor binding sites, is heavily glycosylated and has a highly variable surface that can be subdivided in five variable loops (V1-V5) and five constant regions (C1-C5), see **Figure 5A** to **C** [46,57,58]. Structurally, gp120 can be separated in an inner- and outer domain that are connected by the bridging sheet, consisting of four anti-parallel β-sheets. While the inner domain non-covalently connects gp120 to the gp41 subunit, the outer domain is involved in receptor binding[58]. The gp41 subunit anchors the Env protein in the viral membrane and contains the fusion machinery, required for fusion of host and viral membranes. In comparison to gp120, the gp41 subunit is more conserved in its amino acid sequence. The gp41 subunit consists of the fusion peptide (FP), two heptad repeats (HR1, HR2), a conserved disulfide loop (DSL), the membrane-proximal external region (MPER), the transmembrane domain (TM) and the cytoplasmic tail (CT), see **Figure 5A** to **C**[46,59,60].

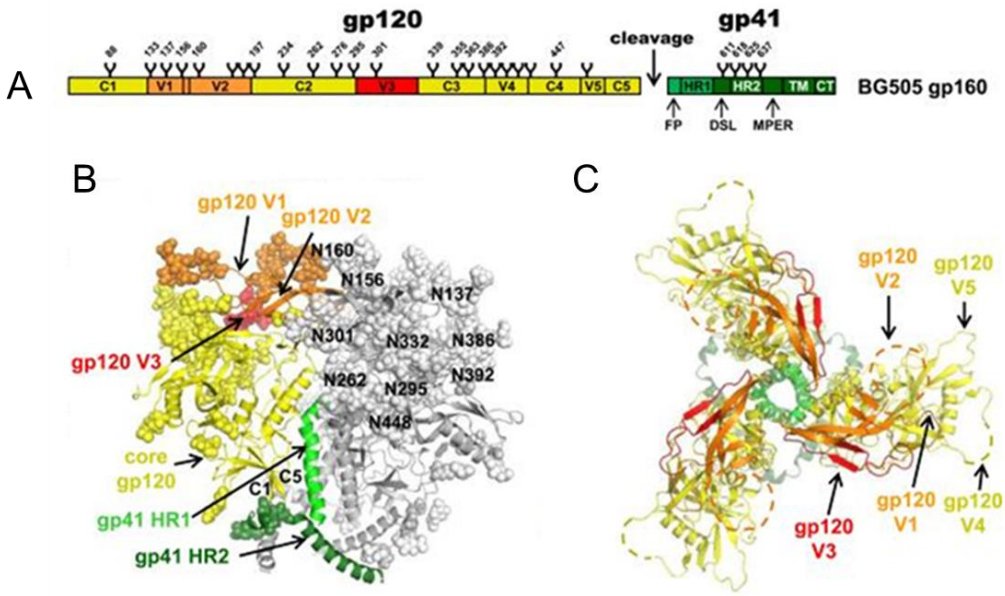



**Figure 5. Schematic representation and overall architecture of the HIV-1 Env trimer. (A)** Schematic view of the BG505 full-length gp160 consisting of gp120 and gp41 subunits. Gp120 includes constant regions 1-5 (C1-C5) and variable loops 1-5 (V1-V5), while gp41 contains the fusion peptide (FP), heptad repeats 1 and 2 (HR1/HR2), the disulfide loop (DSL), the membrane-proximal external region (MPER), the transmembrane domain (TM) and the cytoplasmic tail (CT). Glycan trees and the Furin cleavage site are also indicated. **(B)**, **(C)** Structure of the soluble, cleaved, recombinant HIV-1 Env BG505 SOSIP.664, representing a soluble native-like Env trimer[61]. Note, since the BG505 SOSIP.664 is a soluble, MPER deleted protein, ending at amino acid position 664 (HXB2 numbering), the MPER, TM, and CT are not part of the crystal structure. **(B)** Side view of the Env trimer. For one out of the three protomers, the gp120 core is highlighted in yellow while V3 and V1/V2 loops are colored in red and orange, respectively. Gp41 HR1 and HR2 are labeled in different shades of green. Protein components are depicted according to their secondary structure, while glycans are shown as spheres and are numbered according to their asparagine (N) residues (HXB2 numbering). **(C)** Top view on the Env trimer with variable loops V1-V5 projecting into the periphery. The localizations of V2 and V4 loops are indicated by dashed lines as they are disordered in the structure. Figure adapted from Julien, J.-P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science (80-. ).* 342, (2013)[60], reprinted with permission from AAAS, license number 4536030553645.

### 1.3.2 Function of the HIV-1 envelope

HIV-1 can infect all cells that express its primary receptor CD4 and co-receptors of the chemokine receptor family, including the C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4). Thus, $CD4^+$ T-cells as well as macrophages are susceptible to HIV infection[47,62]. The metastable nature of the pre-fusion state of the protein plays a crucial role in host cell infection, since it allows the protein to undergo irreversible conformational changes required for the fusion process[63].

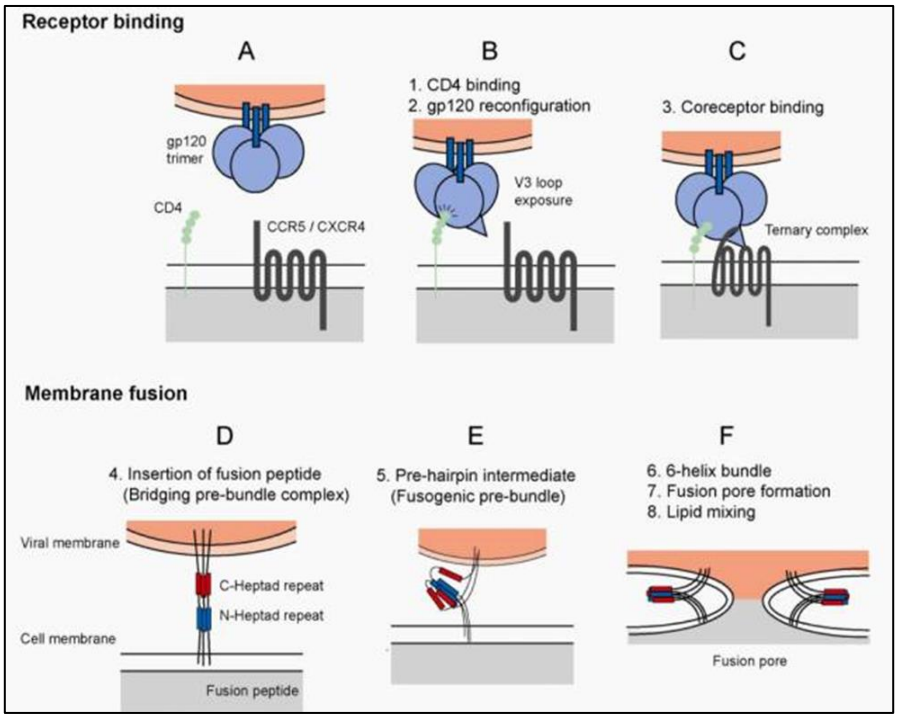



**Figure 6. HIV entry into host cells.** HIV entry into host cells involves two major events, receptor binding and membrane fusion. In the first step, the HIV Env trimer, consisting of three gp120 (blue spheres) and three gp41 (blue bars) subunits comes in proximity to host cells **(A)**. After engagement of the CD4 receptor (green spheres) by the gp120 subunit, containing the CD4 binding site (CD4bs), structural rearrangements result in the exposure of the co-receptor binding site **(B)**. Followed by co-receptor binding **(C)**, resulting in further structural rearrangements, the fusion peptide (FP) is inserted into the host cell membrane **(D)**. The formation of a stable six-helix bundle (6HB) brings the viral and host cell membrane in close proximity. This process is mediated by N and C-terminal α-helical regions of the gp41 subunits, termed heptad repeats 1 and 2 (HR1/HR2), shown in blue and red. During this process, the N-terminal heptad repeats of the three gp41 subunits fold into a central triple-stranded coiled coil and the three C-terminal heptad repeats pack antiparallel into the three groves of the coiled-coil. The process involves a prefusion intermediate **(E)** and results in the formation of the 6HB and the establishment of the fusion pore **(F)**, required to release the viral core into the cytoplasm. Figure from Lobritz, M. A., Ratcliff, A. N. & Arts, E. J. HIV-1 Entry, Inhibitors, and Resistance. *Viruses* 2, 1069–105 (2010)[47].

The initial contact is mediated by the gp120 subunit that binds with its CD4 binding site (CD4bs), consisting of discontinuous residues in the constant regions of gp120, to the N-terminus of CD4, **Figure 6A**, and **B**. This results in the rearrangement of the gp120 subunit and exposure of the highly conserved co-receptor binding site, consisting of regions of C4, β-sheets of V1/V2 and residues of the V3 loop, see **Figure 6B**. The interaction of the co-receptor binding site with CCR5 or CXCR4 co-receptors leads to further structural rearrangements, resulting in the exposure of the FP. The FP, consisting of 15 hydrophobic amino acids located at the N-terminus of gp41, inserts into and destabilizes the host cell membrane (**Figure 6C** and **D**). After insertion of the FP, viral and host cell membrane are brought in close proximity and membrane fusion occurs. This process involves a re-folding event within gp41 that is mediated by N and C-terminal α-helical regions, termed heptad repeat 1 and 2 (HR1/HR2). HR1 and HR2 of each gp41 subunit fold into each other, forming a stable six-helix bundle (6HB). In this configuration, the FP and the transmembrane domain

of gp41 are located at the same end of the molecule. The re-folding event and release of free energy finally results in membrane fusion and formation of a fusion pore (**Figure 6E** and **F**). Lipid mixing results in enlargement of the fusion pore and when a critical size is reached the viral core is released into the cytoplasm[47,64,65]. From there, subsequent steps in HIV replication cycle occur, reviewed in[2,66].

### 1.3.3 Immune evasion mechanisms of the HIV-1 envelope

The viral Env protein is the only viral component on the surface of HIV-1virions, and thus exposed to the selection pressure of the humoral immune response. In order to avoid antibody-mediated neutralization, HIV has evolved a number of escape mechanisms:

**1) HIV-1 variability (Figure 7B):** The high replication rate (~$10^9$ virions are generated per day in an infected individual) and the error-prone reverse transcriptase HIV-1( approximately one nucleotide substitution per genome during a single replication cycle) are responsible for the extensive genetic diversity of HIV-1 and contribute to the adaptation of both, antiretroviral therapy and immune responses[67]. Thereby the envelope is the most variable of all HIV-1 genes[68]. Studies have shown that Env proteins can exhibit 35 % amino acid diversity between subtypes and 20 % within a subtype. Thereby most sequence variation occurs in the gp120 subunit, including variable loops[69,70]. The high variability in these domains is driven by the need of the virus to continuously escape hosts neutralizing antibody responses. The Env variable loops 1-5 (V1-V5) contribute to masking of more conserved Env g120 regions lying underneath, and can tolerate high mutation rates, including insertions and deletions. For example, the length of the most variable loop V1/V2 can range from 50 to 99 amino acids, while V4 and V5 loop lengths range from 19 to 44 and 14 to 36 amino acids, respectively. The length of the V3 loop is more or less constant (34-35 amino acids), since V3 is involved in important trimer stabilizing contacts and in co-receptor engagement. Next to their diversity in amino acid composition and length, variable loops exhibit also high levels of glycosylation and flexibility, and thus contribute to the viral defense to protect conserved Env regions from neutralizing antibody responses [46,71,72,73]. **2) Env glycanshield (Figure 7A):** HIV-1 Env trimers contain a median of 93 N-linked glycans, comprised of a mixture of mainly high-mannose-type and further processed complex-type glycans that account for approximately 50 % of the mass of the protein. The glycan shield consists of a network of interlocked oligosaccharides that occludes most of the polypeptide surface of Env. Thus, it enables HIV-1 to avoid most antibody-mediated neutralization. The glycan shield appears to be evolving by coupling the high HIV mutation rate with high numbers of N-linked glycans

on the surface of Env trimers. This seems to be the primary mechanism by which HIV escapes humoral immune responses since despite of high titers of HIV-reactive antibodies, a sustained viremia is frequently observed during chronic infection[74,75,76,77].

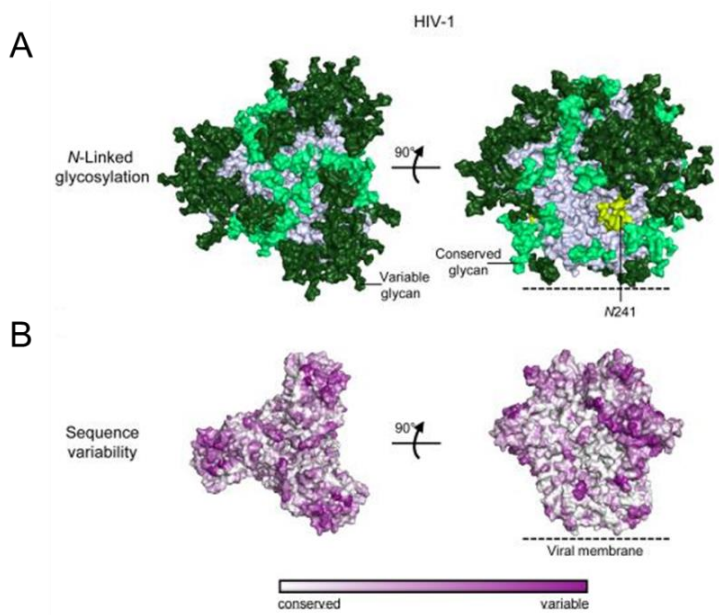



**Figure 7. HIV-1 envelope glycan shield and sequence variability. (A)** N-linked glycans on the pre-fusion closed Env strucure of BG505 SOSIP.664 (PDB ID: 4TVP), shown in top (left) and side view (right). Glycans with a conservation of more than 90 % are shown in light green, while glycans with greater variability (less than 90 % conservation) are depicted in dark green. A conserved glycan residue, not present in BG505, is shown in yellow. **(B)** Sequence variability shown from conserved to variable (white to purple). Figure adapted from Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* 514, 455–61 (2014)[78], with permission of Springer Nature, license number 4536050168052.

**3) Conformational masking:** Immune evasion is also mediated by conformational masking of Env conserved functional sites such as the co-receptor binding site in gp120 that is not fully exposed until the CD4 binding site has bound to the CD4 receptor. Similarly, the fusion machinery in the MPER of gp41 becomes only accessible when the co-receptor binding site is occupied and viral fusion is already initiated. Thus, conformational masking restricts the access of antibodies to these relatively conserved regions of Env and therefore contributes to immune evasion[79,80,81]. **4) Conformational flexibility (Figure 8):** HIV Env pre-fusion trimers are highly dynamic metastable fusion machineries that alternate between closed and more open conformations. This is referred to as “breathing”. It has been shown that trimer opening results in the exposure of some non-neutralizing epitopes, possibly misleading humoral immune responses. However, the relative proportion of pre-fusion conformational states varies within different isolates. For example, in easy-to-neutralize tier 1 isolates a larger proportion of trimers adopt an open conformation, compared to neutralization-resistant tier 2 isolates. In addition, CD4 engagement of the gp120 subunit results in dissociation of the Env trimer, and thus to exposure of epitopes that induce non-neutralizing antibody responses, since these epitopes are occluded in pre-fusion state trimers[82,83,84].

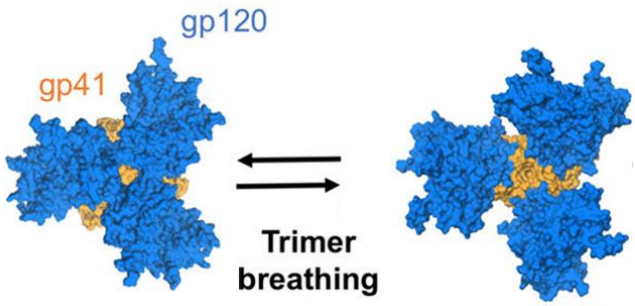



**Figure 8. Conformational states of HIV-1 pre-fusion trimers.** Biophysical data suggest that pre-fusion Env trimers can fluctuate between closed (left) and more open states (right) in a dynamic equilibrium. This is referred to as trimer breathing. Figure adapted with modifications from Ozorowski, G. *et al.* Open and Closed Structures Reveal Allostery and Pliability in the HIV-1 Envelope Spike. *Nature* 547, 360–363 (2017)[85], with permission of Springer Nature, license number 4536060148887.

**5) Low spike density:** HIV presents only a low number of Env proteins on its surface (~14/per virion), thereby impeding bivalent antibody binding by inter-spike crosslinking. The primarily monovalent binding suffers from reduced potency, compared to bivalent binding. Thus, the low spike density contributes to immune evasion of Env[86]. In addition, virions bear non-functional gp120-gp41 monomers, most likely resulting from trimer dissociation along the axis of trimerization, and gp41 stumps that result from shedding of non-covalently linked gp120 subunits. Both, gp120-gp41 monomers and gp41 stumps expose non-functional Env epitopes, and thus contribute to immune evasion by misleading antibody responses[87,88].

The features and mechanisms described above enable HIV-1 to continuously escape from host humoral immune response. Further, they represent major hurdles that have to be considered within the development of a protective vaccine targeting the HIV-1 Env protein, see **1.5**.

## 1.4 Env directed antibody responses

### 1.4.1 Antibody responses during acute and chronic infection

Env-directed antibody responses can be classified into three groups that occur at different timepoints during the course of HIV infection[89]. The first group includes Env-binding but non-neutralizing antibodies that arise within weeks after transmission. Virion-antibody immune complexes are the first form of detectable B-cell responses that occur about one week after plasma virus detection, followed by first free plasma anti-gp41 antibodies a few days later. In contrast, anti-gp120 antibodies primarily targeting V3 can be detected about four weeks after onset of viremia. However, these early antibodies have no detectable effect on plasma viral load and there is also no evidence of antibody-induced Env selection[90]. The antibodies described above, typically bind to immunodominant Env epitopes on HIV virions that are not present on native functional trimers, such as gp41 stumps, resulting from gp120 shedding. Therefore, they are not able to neutralize functional trimers. Nevertheless, the may

have some antiviral activity, since antiviral activity is not restricted to neutralization (meaning viral inactivation by binding and blocking of functional Env trimers, and thus preventing infection of host cells) but also includes other antibody functions such as ADCC or antibody-dependent cell-mediated virus inhibition (ADCVI)[89,91].

The second group includes antibodies that can neutralize HIV-1 in a strain specific manner (NAbs). These antibodies typically arise after several weeks to months after infection. The reasons for the slow development of neutralizing antibodies remain elusive. Presentation of immunodominant epitopes (for example gp41 stumps) that divert antibody responses towards non-neutralizing epitopes, and impaired $CD4^+$ T cell help may contribute to the delayed induction of neutralizing responses. As their name implies NAbs are capable to neutralize the infecting (autologous) strain but cannot neutralize most other (heterologous) strains. NAbs typically target epitopes in variable loops or other regions with relatively high sequence variation, which is also the reason why HIV can easily escape by amino acid substitution, insertions/deletions or an evolving glycan shield, see **1.3.3** [89,91].

However, in some individuals the sustained co-evolution between NAbs and viral escape after years results in the development antibodies with increased neutralization breadth and potency. These antibodies are termed broadly neutralizing antibodies and represent the third group of Env-directed antibody responses, which is described separately in the following chapter.

### 1.4.2 Broadly neutralizing antibodies

During HIV-1 infection between 10 % and 50 % of all infected individuals develop broadly neutralizing antibodies that can neutralize diverse viral isolates. A much smaller fraction (around 1 %), referred to as elite neutralizers, develops bNAbs with exceptional breadth and potency (bNAbs in the context of this thesis). Some of these bNAbs were capable to neutralize over 90 % of isolates tested in pseudovirus assays[92,93]. However, even infected individuals with broadly neutralizing serum activity do not benefit from the existing bNAbs as the autologous virus continues to escape from neutralization[94].The reasons why only a small proportion of infected individuals develop potent broadly neutralizing responses are rather complex and involve a number of factors. For example, high viral loads, low $CD4^+$ counts, a higher frequency of T-follicular helper (Tfh) cells in the periphery and the duration of infection have been associated with the development of bNAb responses[92,93,95].

Typically, bNAbs arise between two and four years after infection in the chronic phase of disease and display a number of unusual features[96,95]. **1) High degree of somatic hypermutation:** A high level of somatic hypermutation (SHM) is a common feature shared by bNAbs. Whereas most affinity matured human antibodies carry about 15-20 somatic mutations, bNAbs are more extensively mutated with up to 40-110 somatic mutations. Typically, mutations accumulate in the complementarity determining region (CDR) loops of the antibodies since they usually contact the antigen. In addition, mutations in the canonical framework regions (FWRs), required for scaffolding of CDRs and commonly less tolerant to mutations, seem to play a critical role in both, breadth and potency of most bNAbs[79,97]. **2) Long HCDR3 loops:** Some, but not all bNAbs exhibit an unusual long heavy chain complementarity-determining region 3 (HCDR3) with loop lengths between 20 to 34 amino acids (average in humans: 16 amino acids). Long HCDR3 loops are typical features that enable V1/V2 and V3 directed bNAbs to penetrate the Env glycan shield. Similar is observed for MPER directed bNAbs, where long HCDR3 loops help to reach highly conserved hydrophobic residues of gp41. However, the lower frequency of human B-cells encodes for long HCDR3s. In addition, antibodies with long HCDR3s have the potential of auto-reactivity, thus B-cells encoding for this antibodies are frequently deleted by negative selection. Taken together, it is seen as challenging to induce antibodies with long HCDR3s by vaccination[98,79]. **3) Poly-/Autoreactivity:** Despite poly-/autoreactivity is normally associated with negative B-cell selection**,** some bNAbs also display poly-/autoreactivity. Polyreactivity likely contributes to increase the overall affinity via heteroligation. In addition, lipid binding seems to essentially contribute to the general neutralization mechanism of MPER-directed bNAbs. Thereby, the initial low affinity contact to the viral lipid membrane seems to result in the exposure of lipid-immersed MPER epitopes that are subsequently bound with high affinity[79,99].

Until 2009 only a limited amount of bNAbs was described in literature. The first generation bNAbs were isolated in the 1990s by phage display and from human hybridomas and resulted in the identification of conserved Env epitopes (see **Figure 9**), such as the CD4bs (bNAb: b12[100,85]), the MPER region (bNAbs: 4E10[101,102], 2F5[101,103,104]) and a glycan epitope at the outer domain (OD) of gp120 (bNAb: 2G12[101,105]). However, all epitopes of the bNAbs described above are present within a single gp120/gp41 protomer, resulting from monomeric antigens used for selection and screening[79,92]. Since 2009, the development of new screening procedures and technologies, including single B-cell culture and functional screenings with tier 2 neutralization as primary selection criterion, resulted in the identification of scores of

bNAbs, which reflect the second generation of bNAbs. Compared to first generation bNAbs, second generation bNAbs were in general of greater breadth and potency[44,92]. Since a screening with neutralization as primary selection criterion is not restricted to antibody-binding to already existing antigens, like phage display or B-cell sorting, some of the isolated antibodies defined new bNAb epitopes (see **Figure 9**). These include the gp120 V2-glycan site at the trimer apex, a region at the gp120/gp41 interface and the gp120 V3-glycan epitope, overlapping the 2G12 epitope. Since the V3-glycan epitope has a central glycan at position 332 it is also referred to as N332-glycan supersite[44,92,89,106]. Notably, multiple epitopes of the newly identified bNAbs were characterized/identified using BG505 SOSIP.664 trimers[61,56] (described in detail in **1.5.1**). BG505 SOSIP.664 trimers have been proven valuable for structural studies since using these trimers in cryo electron microscopy and x-ray crystallography approaches resulted in the first authentic high-resolution structures of a HIV-1 Env trimer. For example, a new cluster of quaternary structure-specific antibodies (including PGT151[107,108]) was mapped to the gp120-gp41 interface of BG505 SOSIP.664 trimers[56].

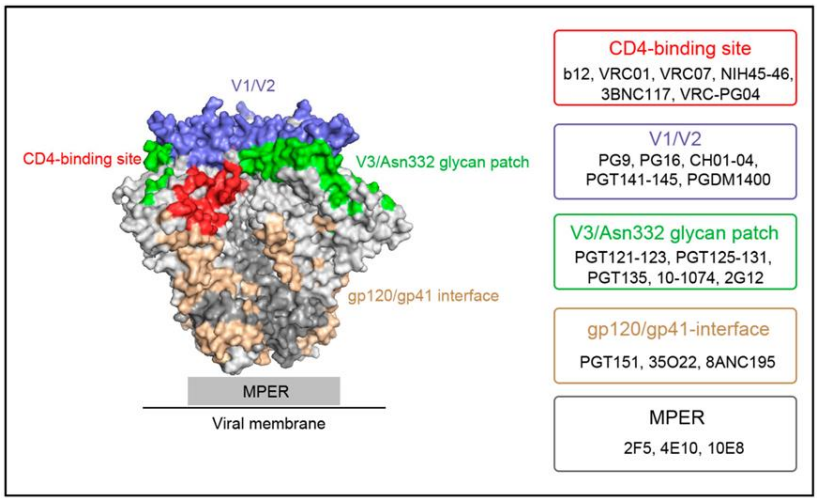



**Figure 9. bNAb epitopes on the Env trimer.** Major epitopes of broadly neutralizing antibodies (bNAbs) on the surface of BG505 SOSIP.664 (PDB ID: 4TVP. Epitopes are shown in different colors: CD4 binding site (red); V1/V2 (V2-glycan) (blue); V3-glycan (green); gp120/gp41 interface (wheat); MPER (illustrated as grey rectangle). Unspecified regions of gp120 and gp41 moieties are shown in light and dark grey, respectively. Representative bNAbs for each epitope are listed in boxes (right). Figure from Zhang, Z. *et al.* Antiviral Therapy by HIV-1 Broadly Neutralizing and Inhibitory Antibodies. *Int. J. Mol. Sci.* 17, 1901 (2016)[109].

In summary, several findings described above greatly impacted HIV vaccine design and focused vaccine design towards the Env protein. These findings include that 10 % to 50 % of all infected individuals can develop a broadly neutralizing antibody response, with a smaller number developing bNAbs with the capacity to neutralize over 90 % of tested isolates. This shows that the human immune system is basically able to generate broadly neutralizing antibody responses[91]. Further, a large number of bNAbs have been isolated since 2009 and mapping of these antibodies to newly available immunogens, such as the BG505 SOSIP.664,

resulted in the discovery of new bNAb epitopes. The knowledge, gained from the discovery and mapping bNAbs, was and is of utmost importance in development of next generation Env immunogens, described in **1.5**. That the induction of bNAbs is a desirable goal in HIV vaccine development is highlighted by passive immunization and challenge experiments. It has been shown that passive immunization of bNAbs can protect macaques from infection with chimeric simian-human immunodeficiency viruses (SHIVs) encoding HIV-1 envelope genes in an SIV backbone[110,111,112]. This indicates that bNAbs, if present in sufficient titers prior to infection, can protect from infection.

## 1.5 Advanced Env vaccine design and efforts towards the induction of cross-neutralizing antibody responses

The first HIV vaccines were monomeric gp120 subunit vaccines that failed to confer protection by neutralization in phase 3 efficacy trials[23,24], see **1.2**. Although gp120 monomers may display epitopes of potent bNAbs such as the CD4 binding site, they failed to induce neutralizing antibody responses. The presentation of immunodominat non-neutralizing epitopes, such as gp120 epitopes normally occluded by inter-subunit interaction, and the inability to present potent quaternary structure and gp41 epitopes may have contributed to this outcome[56]. However, the only efficacy trial demonstrating modest efficacy (RV144[33]) was based on a gp120 subunit vaccine. Notably, a primary correlate of protection was a non-neutralizing antibody response directed against V1/V2, see **1.2**. However, to overcome limitations of gp120 described above, HIV Env vaccine design focused on the production of soluble trimeric proteins. A general strategy for the production of soluble trimers involved a stop codon before the gp41 transmembrane domain to generate soluble proteins consisting of the ectodomain of gp41 ($gp41_{ECTO}$) and gp120 subunits. Further, they included the inactivation of the Furin cleavage site (e.g. REKR to SEKS) to produce covalently linked gp120 and gp41 subunits that otherwise would rapidly dissociate into gp120 subunits and a trimerized form of gp41. When tested in humans, these Env trimers elicited only weak neutralizing antibody responses, limited to easy-to-neutralize tier 1 strains[56,113,114,115]. The limitations of the first soluble Env trimers were revealed more than a decade after their development, when studies have shown that these trimers adopt non-native conformations with gp120 subunits dangling from trimerized gp41 thereby exposing non-neutralizing epitopes[56,116,117].

### 1.5.1 Stabilized trimers to focus immune responses towards bNAb epitopes

A hallmark in HIV Env vaccine design was the development of the soluble cleaved BG505 SOSIP.664[61] trimer, based on the clade A transmitted/founder isolate BG505[118]. Many modifications in the wild-type BG505 sequence (see **Figure 10**), resulted in BG505 SOSIP.664 gp140, representing trimers with improved stability, homogeneity and solubility. The sequence modifications included an optimized Furin cleavage site, REKR to RRRRRR (R6), to enhance cleavage and therefor proper protein folding, disulfide-bond forming cysteines at amino acid positions 501 and 605 (referred to as "SOS") for covalent linkage of gp120-gp41$_{ECTO}$ subunits and a helix-breaking point substitution isoleucine to proline at amino acid position 559 (I559P), allowing gp41$_{ECTO}$ subunits to remain in the pre-fusion ground state. Furthermore, the MPER region was deleted to improve homogeneity and solubility (C-terminal residue at amino acid position 664) and a glycan was introduced at amino acid position 332 (T332N), since the binding of several bNAbs is depending on the presence of this glycan. Negative stain electron microscopy confirmed that affinity and size exclusion chromatography (SEC) purified BG505 SOSIP.664 trimers were highly homogeneous with more than 95 % adopting a tri-lobed propeller shape, nowadays defining the characteristics of native-like soluble trimers[56,61]. With melting temperatures of about 67 °C, the trimers were highly stable and different biochemical and biophysical assays (ELISA, surface plasmon resonance spectroscopy (SPR), biolayer interferometry (BLI), isothermal titration calorimetry (ITC)) confirmed that the BG505 SOSIP.664 trimers have desirable antigenic properties since they show binding to all known bNAbs (with the exception of MPER directed bNAbs) but no or very poor binding to non-neutralizing antibodies[61,119,120,56]. In an immunization study with rabbits, BG505 SOSIP.664 trimers elicited strong and consistent NAb responses against the autologous tier 2 virus BG505.T332N, whereas BG505 gp120 monomers and uncleaved gp140 trimers induced weaker or undetectable NAb responses. Notably, autologous neutralization was observed the first time for Env-based immunogens. In an immunization study with rhesus macaques BG505 SOSIP.664 trimers also induced a NAb response against the autologous virus but antibody titers were lower than in rabbits. However, heterologous neutralization of a tier 2 clade C isolate was not achieved[121,122]. Although both, rabbits and macaques, developed autologous neutralization, the epitopes mainly contributing to autologous neutralization differed between the species as it was reported in a very recent published study[123]. According to the study, a previously

described glycan-hole site in BG505 SOSIP.664 (amino acid positions 241/289) accounts for about 50 % of autologous neutralization in rabbits. However, the study identified another important epitope in the C3 region of Env (termed C3/465) that accounts for about 25 % of the neutralizing responses in rabbits. Remarkably, the same study was able to show that the neutralizing response in macaques is dominated by the newly identified C3/465 epitope[123,124].

Meanwhile, the SOSIP.664 design was successfully transferred to other isolates with properties similar to BG505. In addition, different strategies were applied to further stabilize BG505 SOSIP trimers and to reduce the conformational flexibility in order to avoid/reduce undesired exposure of non-neutralizing epitopes related to trimer breathing (for example non-neutralizing V3 epitopes) and CD4-induced open conformations[56]. These strategies include chemical crosslinking of trimers as well as further stabilization by applying further stabilizing modifications[125,56]. Further, findings from the SOSIP design resulted in alternative designs for trimer stabilization, like the native-flexible linker design (NFL). The NFL design essentially reflects the SOSIP design, with the major difference that gp120 and gp41 subunits are covalently linked by a flexible glycine-serine (GS) linker instead of a disulfide bond. Thus, NFL trimers do not require Furin cleavage to adopt a native-like conformation[126,127]. By direct comparison in an immunization study with macaques, BG505 NFL trimers and BG505 SOSIP.664 trimers induced approximately equivalent neutralizing antibody responses[128]. In addition, the same study investigated the immunogenicity of advanced BG505 SOSIP.664 trimers designed for further stabilization, reduced V3 loop exposure and minimized CD4 reactivity/CD4-induced non-nAb epitope exposure[129,130,131]. Interestingly, despite of reduced V3 responses, there was no difference in absolute tier 2 NAb responses between further modified versions and the basis version of BG505.SOSIP.664 trimers[128].

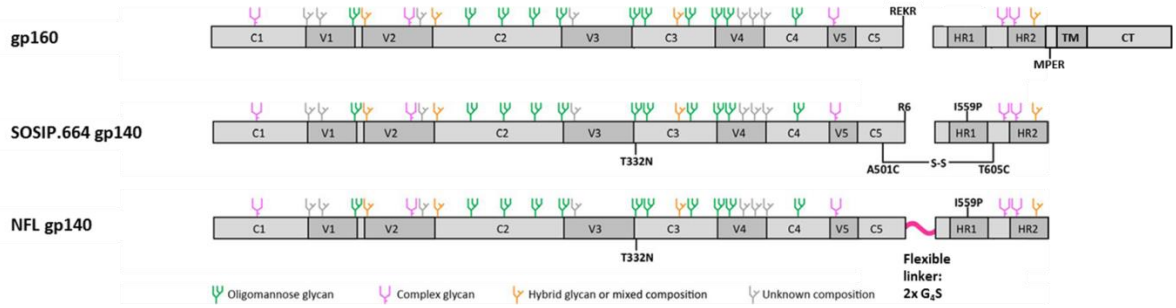



**Figure 10. Schematic presentation of BG505 gp160 wildtype (WT), BG505 SOSIP.664 and native flexible linker (NFL) Env variants. Top**: BG505 gp160 wildtype. Constant regions (C1-C5), variable loops (V1-V5) as well as heptad repeats 1 and 2 (HR1, HR2), membrane-proximal external region (MPER), transmembrane domain (TM) and cytoplasmic tail (CT) are shown in grey. Glycan trees are depicted in green (oligomannose-type), purple (complex-type), orange (hybrid- or mixed-types) and grey (unknown composition). The REKR cleavage site is also indicated. **Middle**: BG505 SOSIP.664. MPER, TM and CT are truncated to enable expression of a soluble trimer. SOSIP modifications include the helix-breaking isoleucine to proline substitution at amino acid position 559 (I559P) to stabilize the pre-fusion ground state of $gp41_{ECTO}$ subunits, and further disulfide-bond forming cysteine substitutions at positions 501 (A501C) and 605 (T605C) to covalently link gp120 and gp41 subunits. Further modifications include the replacement of the original REKR Furin cleavage site by six arginine (R6) to enhance Furin cleavage and the introduction of a glycan site at amino acid position 332 (T332N) to restore the central glycan of the V3-glycan bNAb epitope. **Bottom**: NFL design. The Furin cleavage site is substituted by two flexible glycine-serine linkers (2x $G_4S$, in purple) that covalently link gp120 and gp41 subunits and allow a cleavage-independent folding into native-like trimers. Importantly, the NFL design also includes the I559P substitution. Figure adapted with modifications from Sanders, R. W. & Moore, J. P. Native-like Env trimers as a platform for HIV-1 vaccine design. *Immunol. Rev.* 275, 161–182 (2017)[56].

Although stabilized trimers were designed to focus the humoral immune response towards bNAb epitopes, the induction of bNAbs remained unsuccessful so far. Nevertheless, stabilized trimers are worth to be tested in human trials since, as learned from the RV144 trial, protection from infection is not limited to the (although highly desirable) induction of bNAbs. Native-like trimers, such as BG505 SOSIP.664, expose V1/V2 loops that were the target of one correlate of protection, non-neutralizing antibodies mediating ADCC, in the RV144 trial. However, they do not expose immunogenic epitopes such as the "backside" of gp120, like the gp120 immunogens in RV144 probably did. Thus, stabilized trimers maybe useful to focus immune responses towards protective epitopes (in context of ADCC) by avoiding distraction of antibody responses to immunodominant sites not correlated with ADCC. However, if this theory holds true has to be shown in future clinical studies.

### 1.5.2 Sequential immunization to guide B-cells towards bNAb responses

In infected patients the development of bNAbs usually takes years of co-evolution between the immune system and viral escape, see **1.4.2**. Sequential immunization strategies basically aim to mimic this process by using suitable Env immunogens. For this purpose, rational vaccine design focuses on the development of germline-targeting Env immunogens that are

able to activate naive B-cells, expressing receptors of bNAb germline precursors. Once activated and expanded, sequentially immunized intermediate Env variants should guide B-cells towards bNAb responses[132,133,121]. Adressing B-cell receptors of bNAb germline precursors by specially designed antigens became from particular interest since immunization studies with (stabilized) trimers as only immunogens (e.g. BG505 SOSIP.664 in rabbits and macaques[122]) failed to induce bNAb responses. A reason that maybe contributed to the outcome of such studies is that BG505 SOSIP.664 (and maybe also other Env immunogens) does not react with most germline reverted precursors of bNAbs[121,134], and thus failed to efficiently stimulate B-cell receptors of bNAb germline precursors. However, sequential immunization studies once may help to solve this problem. The potential of these approaches were highlighted in a recent study that aimed for the induction VRC01 class antibodies, targeting the CD4 binding site and representing highly potent bNAbs. In the study, an engineered outer domain (eOD) construct GT8[135,136], based on gp120 and containing modifications that enable binding of multiple predicted germline precursors of VRC01 class antibodies, was used to prime transgenic mice expressing germline VRC01 heavy chains. Boosting with an intermediate Env (BG505-GT3) and then with BG505 SOSIP.664 containing a glycan deletion at amino acid position 276 (N276D) resulted in the elicitation of highly mutated antibodies. Analyses revealed that a significant part of the mutations were shared with mature VRC01 class antibodies. Interestingly, the elicited antibodies were able to broadly neutralize a panel of near-native tier 2 isolates (lacking the glycan at amino acid position 276). However, heterologous neutralization of fully-native viruses containing the N276 glycan was limited to one isolate[137]. Similar results were obtained in a study in which transgenic mice, expressing the germline reverted version of the PGT121 family precursor (epitope: V3-glycan), were primed and sequentially boosted with a set of successive mutated Env antigens[133]. In addition, a recent study published the design of a germline antibody binding variant of BG505 SOSIP.664. 17 individual point substitutions, including the deletion of several glycans (also the glycan at position 276), and the deletion of seven amino acids in the V2 loop resulted in a stabilized trimer termed BG505 SOSIP.v4.1-GT1. This trimer can bind multiple germline reverted bNAb precursors *in vitro* and activate B-cells in transgenic mice, expressing germline VRC01 and PGT121 precursors[138]. Future sequential immunization studies have to prove the suitability of this protein as germline priming antigen for the induction of bNAbs.

### 1.5.3 Approaches to improve durability and quality of Env-directed antibody responses

One important observation in the RV144 trial was that the initial vaccine efficacy of 60.5 % during the first 12 months after vaccination, declined to 31.2 % in the follow-up time of 42 months. Further, the initial vaccine efficacy corresponded to a time when most immune responses, including Env IgG/IgG3 and ADCC responses (correlates of protection, see **1.2**) were higher[36]. The observations highlight that the magnitude of Env-directed immune responses were critical for vaccine efficacy, and further that Env-directed immune responses are limited in their durability. Thus, a general aim in HIV vaccine development is to induce durable Env-directed immune responses of high magnitude that ideally also contribute to the development of bNAbs. Formulation of Env immunogens in potent adjuvants is a likely way to address the limitations described above. In the RV144 trial proteins for boosting were administered in alum (see **1.2**), representing one of the oldest (discovered in the 1920s) and most widely used adjuvants[139]. Although alum failed to induce long-lasting antibody titers in the RV144 trial, it is a highly safe adjuvant, normally known to generate high and long-lasting antibody titers via type 2 T-helper (Th) cell immune responses. However, there are also other types of adjuvants available, like for example the squalene based oil-in-water emulsion MF59 (improves humoral and cell-mediated immunity, generates high antibody titers), monophosphoryl lipid A (toll-like receptor 4 agonist, generates Th1 immune responses), ISCOMATRIX (lipid-based adjuvant formation containing immunostimulatory saponin QS21, generates high and long-lasting antibody titers, induction of balanced Th1 and Th2 responses as well as cell-mediated immunity) and many others [139]. For example, MF59 is currently tested in the phase 2b/3 efficacy trial HVTN702, see **1.2** and ISCOMATRIX is widely used is in studies analyzing humoral immune responses induced by stabilized trimers[140]. Presentation of Env proteins on the surface of particles is another possibility that is expected to improve the durability and quality of antibody responses, since they offer several advantages over soluble proteins, including: 1) Delivery of many antigen copies to an antigen-presenting cell at once, possibly resulting in enhanced T-cell priming; 2) crosslinking of B-cell receptors by higher avidity interactions, and thus enhanced B-cell activation. Enhanced B-cell activation supported by Tfh-cells is expected to result in improved germinal center (GC) reactions, including affinity maturation by somatic hypermutation, which in turn lead to antibody responses with increased durability and breadth[141,142]. The potential of particle-based approaches is highlighted in a recent immunization study where macaques were immunized

with clade C 16055 NFL trimers, arrayed in a well-ordered manner on the surface liposomes. Interestingly, trimer exposing liposomes were generated by capturing histidine-tagged 16055 NFL trimers to liposomes with incorporated $Ni^{+2}$ lipids. The results of the study showed that liposome-presented Env induced superior GC responses (larger GCs, greater content of Tfh cells in GCs of liposome immunized animals) and higher autologous tier 2 neutralizing antibody titers compared to the soluble version of the proteins[143]. However, there exist also other approaches to present Env on particles, including for example virus-like particles (VLPs) and poly(lactic acid) particles (PLA particles)[141,56,144]. Generation of Env-carrying VLPs, produced by coexpression of HIVs group-specific antigen (Gag) and Env proteins in mammalian cells, represents one of the initial efforts to present Env trimers on the surface of particles. The advantages of VLPs are that Env proteins are presented in their membrane-bound state, and further that the Gag protein itself is immunogenic and elicits cytotoxic T lymphocyte responses. However, there are also some challenges with VLPs. These include low Env density and presentation of a mixture of native and non-native forms of Env, as known from HIV virions (see **1.3.3**). Both problems have been addressed. Proteases for example were successfully used to digest non-functional Env molecules (greater protease vulnerability) from the surface of JRFL Env-carrying VLPs, while Env with native conformation remained intact. The amount of VLP incorporated Env can be increased by truncation of the CT[141,56]. However, experiments with cell-surface expressed Env molecules showed that the deletion of the CT diminishes binding of trimer-specific bNAbs and results in the exposure of non-neutralizing epitopes[145]. PLA particles consist of biodegradable polymers that have been licensed for human use and offer features which makes them interesting for vaccine design in general. For example, PLA particles allow the adsorption of antigens to the surface of the particles via electrostatic interactions or encapsulation of compounds into the particles or a combination of both[146]. Several studies reviewed in[146], highlight the potential of PLA particles as vaccine carriers. For example, PLA-adsorbed HIV p24 protein induced high antibody titers in mice, rabbits and macaques[147]. Further, in an immunization study with rabbits, PLA-adsorbed p24 induced similar antibody titers as MF59-formulated p24[148]. In conclusion, the immunogenic properties of Env can be greatly enhanced by potent adjuvant formulations and particle presentation or a combination of both.

## 1.6 Biophysical and biochemical characterization of Env proteins

Characterizing the antigenicity profile of Env immunogens is of great importance in HIV vaccine design, especially in case of native-like Env immunogens. Usually two important components are involved in the characterization of (native-like) Env immunogens: 1) methods and technologies that allow the precise quantification of Env-antibody interactions and 2) a set of quaternary structure-characterizing antibodies, including bNAbs and non-neutralizing antibodies. Commonly used assays for example are ELISA, SPR, BLI and less often also ITC[56,61,119,120]. Each of the methods named above has its individual benefits and drawbacks. ELISA for example is an easy to establish, widely used and robust method that enables measurement of equilibrium binding affinities but it requires surface immobilization. Depending on the assay setup, surface immobilization can affect the structure and thereby the antigenicity profile of the immobilized ligand[125]. Establishment of SPR assays is quite labor-intensive and, like ELISA, requires surface immobilization of one interaction partner. However, when successfully established the method provides valuable data about on- and off-rates of an interaction[149]. Similar is true for BLI assays. The benefits of ITC are that the method is label-free and measurement takes place in solution. By measuring the reaction enthalpy, this method provides information about dissociation constants and stoichiometry of an interaction. Drawbacks of the method include preparation time, high concentration and amount of sample material and buffer limitations[149].

MicroScale Thermophoresis represents a relatively new and powerful technique for the analysis of molecular interactions (explained in detail in **4.1.1**)[149,150]. MST can measure equilibrium binding affinities and provides several advantages, including low sample consumption, essentially no limitation in respect to molecule size, and the ability to measure affinities in solution and under physiological conditions, such as serum[149]. Thus, the method offers great opportunities for the analysis of Env-antibody interactions that could be only limitedly applicable in some of the methods described above. These include particularly the potential to characterize the antigenicity profile of particle-presented Env proteins and to analyze Env-antibody interactions in the presence of serum. Both aspects play essential roles in vaccine design. For example, native-like trimers immobilized to particles should retain their native-like structure and favorable antigenicity profile, which therefore has to be analyzed by suitable methods. Another, often neglected question is how the presence of serum

affects the antigenicity profile of native-like Env trimers since in most studies the antigenicity profile is assessed under artificial conditions in buffer. However, it is well documented that a stabilized soluble Env protein from subtype B (B41) can suffer from proteolytical damage of the V3 region, mediated by serum proteases in the medium of producer cells[151]. Since proteolytical damage may influence the favorable antigenicity profile of native-like trimers and with the possibility that this may also happen in the presence of human serum, it is reasonable to address this question. Thus, one objective of this thesis was to test if MST can meet the requirements to investigate Env-antibody interaction under these rather complex conditions described above.

# 2. Objectives

The induction of broadly neutralizing antibodies by Env immunogens is a highly desirable goal in HIV-1 vaccine design that remained unsuccessful so far[152]. The development of stabilized native-like trimers with desired antigenic properties (binding to bNAbs, no/reduced binding to non-neutralizing antibodies) represented a hallmark in HIV-1 Env vaccine design[61]. Tested in animal models, these trimers for the first time induced neutralizing antibody responses against the autologous tier 2 virus[121,122]. With the intention to elicit more durable and broader NAb responses Env immunogens have been presented on different kinds of particles[141,143,56,144,153]. However, it is essential that the antigenicity profile of particle-presented Env is preserved and comparable to the corresponding soluble protein. Further, it is reasonable to test if Env immunogens, soluble or particle-presented, retain their antigenicity profile under physiological conditions, such as serum. One objective of this thesis was to test if MicroScale Thermophoresis, allowing affinity measurements in solution, is a useful technique for the characterization of soluble and particle-presented Env proteins in buffer and under physiological conditions. Therefore, a five-member Env/V3 chimeric model library and the monoclonal antibody (mAb) 447-52D as well as native-like BG505.SOSIP.664 trimers and a panel of quaternary structure-characterizing antibodies were used to establish protocols for MST analysis of soluble and particle-presented Env proteins. Thereby, MST results were compared to ELISA data that were collected in parallel.

The design of Env immunogens requires knowledge about amino acid residues that are critical for antibody binding. A mammalian cell display of Env libraries containing Env variants with consecutive alanine substitutions is one approach that can be used for identification of such amino acid residues and the generation of such a library reflects the second objective of this thesis. The reliability of the results of an Env cell-surface display mapping technology is strongly dependent on a homogeneous well-ordered Env trimer population on the cell membrane. Thus, the library was generated based on the BG505 isolate that is known to benefit from intrinsic trimer enhancing propensities[127]. In addition, a part of the library containing all variants with alanine substitutions that result in the knockout of N-linked glycosylation sites was used to for the generation of a stable cell line based BG505 SOSIP gp145 glycan-knockout library. This library should be used to investigate antibody-glycan interactions via a previously described cell sorting based approach[154] allowing the isolation of Env antigens according to their antibody binding properties.

# 3. Materials and methods

Unless stated otherwise, all chemicals used in this work were purchased from Sigma-Aldrich, Merck or Carl Roth. General plastic material was purchased from BD Falcon, Eppendorf and Sarstedt.

## 3.1 Human codon optimization

All HIV-1 envelope variants used in this work were optimized for human codon usage using the online tool "GeneArt GeneOptimizer"[155,156] (GeneArt/ThermoFisher Scientific).

## 3.2 Molecular biology

### 3.2.1 General molecular biology

Unless stated otherwise, cloning of all DNA constructs was performed using standard cloning protocols[157]. Restriction digestions of vectors and inserts were carried out using restriction endonucleases from New England Biolabs (NEB) or Fermentas according to manufacturer`s instructions. Inserts for cloning were generated either by digestion of plasmids with restriction endonucleases or polymerase chain reaction (PCR) amplification using the Phusion High Fidelity Polymerase. Briefly, primer extension PCR with appropriate oligonucleotides was used for the attachment of restriction sites and sequences encoding for protein tags, whereas overlap extension PCR was used to introduce additional mutations. Amplified inserts as well as digested vectors and inserts were purified via agarose gel electrophoresis, followed by gel extraction of desired DNA fragments using the QIAquick Gel Extraction Kit. Ligation of digested vectors and inserts (molar ratio 1:3 vector to insert) was performed using the T4 DNA ligase according to manufacturer`s recommendations. The ligation mixture was transformed into chemically competent *Escherichia coli* (*E. coli)* DH10B or DH5α according to standard protocols[158]. Insert containing colonies were identified by an analytical PCR using GoTaq® Green Master Mix and subsequently used to inoculate LB medium containing the appropriate antibiotic. Plasmid DNA was purified from LB cultures using the GeneJET™ Plasmid Miniprep Kit and DNA concentration and purity were determined by photometric quantification (absorbance at 260 and 280 nm) using a NanoDrop ND-1000 photometer. Finally, DNA was verified by Sanger sequencing (Seqlab, Göttingen, Germany or GATC, Konstanz, Germany). Cloning inserts into QuickLigation (QL) vectors containing a control of

cell death toxin B (CcdB) cassette[159] was carried out by golden gate cloning as described previously[160]. Briefly, a reaction mixture containing the pQL13 plasmid, the insert of interest and the type IIs endonuclease Esp3I was incubated for 45 min at 37 °C in a PCR cycler. After incubation, a ligation mixture containing T4 DNA ligase was added and the reaction was incubated at alternating temperatures (25 $C^0$, 37 $C^0$, 25 $C^0$) for 45 min, each. The reaction mixture was transformed into chemically competent *E. coli* and further analysis was carried out as described above. All oligonucleotides used for cloning and sequencing are listed in **Appendix Table 2**. A detailed catalogue of all vectors used in this thesis is shown in **Appendix Table 3**. Oligonucleotides and elements used for cloning of the BG505 alanine substitution library are separately listed, see **3.2.3**.

| | |
|---|---|
| Phusion® High-Fidelity DNA Polymerase | New England Biolabs, Ipswich, USA, # M0530 |
| QIAquick Gel Extraction Kit | Qiagen, Hilden, Germany, # 28706 |
| T4 DNA Ligase | New England Biolabs, # M0202 |
| *E. coli* DH10B | F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara, leu)7697 araD139 galU galK nupG rpsL λ-[161] |
| *E. coli* DH5α | F- supE44 ΔlacU169 (φ80 lacZΔM15) hsdR1 recA1 endA1 gyrA96 thi-1 relA1[158] |
| Lysogeny broth (LB) medium | 1 % (w/v) NaCl, 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, pH 7.5 |
| GoTaq® Green Master Mix | Promega, Fitchburg, USA, # M712 |
| GeneJET™ Plasmid Miniprep Kit | ThermoFisher Scientific, Waltham, USA, # K0503 |
| NanoDrop ND-1000 photometer | PEQLAB Biotechnologie GmbH, Erlangen, Germany |
| Endonuclease Esp3I | ThermoFisher Scientific, # ER0452 |

### 3.2.2 Preparation of plasmid DNA for transfection of human cells

Depending on the amount needed, plasmid DNA was isolated from LB cultures using QIAGEN Midi, Maxi or Mega Kits according to manufacturer`s instructions. Concentration and purity of isolated plasmid DNA was determined by photometric quantification (absorbance at 260 and 280 nm) using a NanoDrop ND-1000 photometer.

| | |
|---|---|
| QIAGEN Plasmid Midi, Maxi, Mega Kit | Qiagen, # 12143, # 12163, # 12183 |

### 3.2.3 Cloning of the BG505 SOSIP alanine substitution library

The BG505 SOSIP alanine substitution library was cloned based on the full-length BG505 sequence (GenBank # DQ208458.1), which was adapted to human codon usage and further modified by adding the previously described SOSIP modifications, the R6 cleavage site and the T332N substitution[61]. Starting with amino acid position 34 (HXB2 numbering), every amino acid (except SOSIP modifications, R6 cleavage site and natural occurring alanines and cysteines) of the extracellular part of the protein, including the transmembrane domain (gp145, AA 34-712) was consecutively substituted for an alanine, while the intracellular part of the protein, represented by the cytoplasmic tail remained unmodified. Besides, for being able to quickly transform the BG505 SOSIP gp160 alanine substitution library into either a gp145 or a gp140 library, elements were implemented to allow this switching by simple and robust molecular biology tools. The cloning procedure of the library included five major steps, which will be described in detail in the sections below: 1) cloning of the *library template*, encoding for the extracellular part of the protein; 2) cloning of *sequence segment exchange cassettes* into the *library template* which were later on substituted by double stranded *linker oligos* including the alanine substitutions; 3) cloning of the CT encoding sequence into the target *expression vector*; 4) cloning of the *sequence segment exchange cassettes* into the target *expression vector*; 5) linker ligation of *linker oligos* into cloning cassettes to introduce alanine substitutions, thereby generating 615 BG505 SOSIP gp160 variants encoding for consecutive alanine substitutions in the extracellular part of the protein. All elements, including oligonucleotides, gBlock sequences and DNA linkers required for cloning and sequencing of the BG505 SOSIP gp160 alanine substitution library are listed in **Appendix Table 4**, **Appendix Table 5** and **Appendix Table 6**, respectively.

**Cloning of the BG505 SOSIP R6 library template**

The library template was cloned in the pSEVA silent vector (see **Appendix Table 3**). In a first step, a KpnI site in the multiple cloning site (MCS) of the vector was removed to later on enable cloning of library cassettes. To generate the library template, the 3´ part of the extracellular BG505 SOSIP sequence was synthesized as gBlock sequence, (gBlockXhoI, obtained from GeneArt/ThermoFisher) including NcoI, XhoI, SfiI and AarI (type IIs endonuclease) restriction sites, and cloned into pSEVA silent via HindIII/NheI. To complete the BG505 SOSIP library template, the 5´ part of the BG505 SOSIP sequence was cloned into the pSEVA silent vector via HindIII/NcoI (see **Figure 11**).

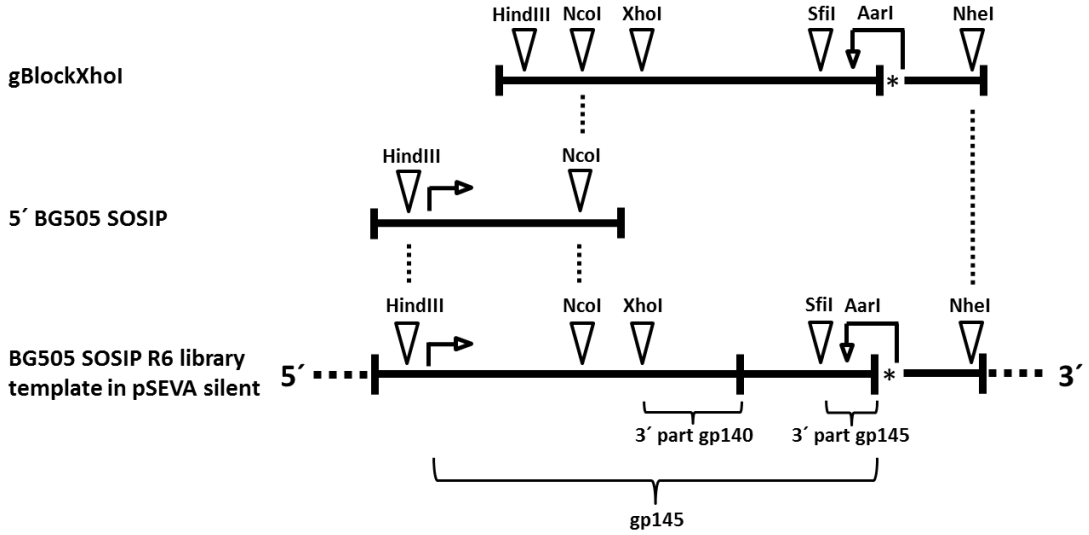



**Figure 11. Cloning of the library template encoding for the extracellular part of BG505 SOSIP R6.** The gBlockXhoI sequence, encoding for the 3´ part of the BG505 SOSIP sequence and including NcoI, XhoI, SfiI and AarI restriction sites, was cloned into pSEVA silent via HindIII and NheI. In the next step, HindIII and NcoI were used to clone the 5´ part of the BG505 SOSIP sequence into the pSEVA silent vector, thereby completing the library template.

### Cloning of exchange cassettes

The BG505 SOSIP R6 library template in pSEVA silent was used to generate exchange cassettes to later on enable linker ligation in the pLib1.1 vector. Starting 3´ to the signal peptide (SP) encoding sequence, oligonucleotide pairs (see **Appendix Table 5**) were used in vector PCRs to generate 38 vectors containing 38 consecutive exchange cassettes. The exchange cassettes were designed to allow the exchange of 54 bp, each, thereby together covering the entire BG505 SOSIP sequence encoding for the extracellular part of the protein, see **Figure 12A**. It is important to note that sequences flanking a particular exchange cassette correspond to the initial BG505 SOSIP library template. During the vector PCR, the oligonucleotides introduced BsaI (type IIs endonuclease) and KpnI restriction sites (see **Figure 12B**). The latter, were used to digest and re-ligate the amplified vectors, whereas the BsaI sites were required for later on linker ligation in the pLib1.1 expression vector.

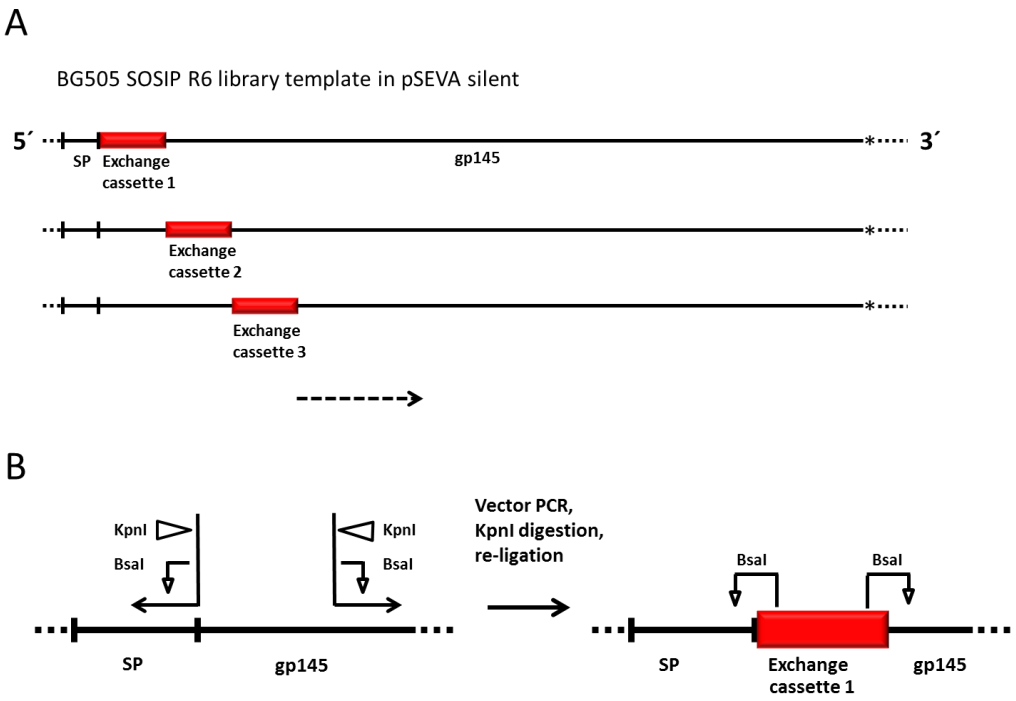



**Figure 12. Cloning of exchange cassettes based on the BG505 SOSIP R6 library template in pSEVA silent. (A)** Starting 3´ to the signal peptide (SP) encoding sequence, oligonucleotide pairs were used in vector PCRs to generate 38 vectors, containing 38 consecutive exchange cassettes. Each cassette was designed to allow the exchange of 54 bp, thereby together covering the entire of the BG505 SOSIP sequence, encoding for the extracellular part of the protein. **(B)** Cloning and design of the exchange cassettes. Exchange cassettes contained KpnI and BsaI restriction sites for digestion and re-ligation of vectors (KpnI) and later on linker ligation in the pLib1.1 expression vector (BsaI).

### Cloning of expression vector pLib1.1

First, a BsaI site in the ampicillin resistance of the pcDNA5/FRT/TO vector was removed by introducing a silent mutation and a KpnI site in the MCS of the vector was also removed. The sequence encoding for the CT and the 3´ part of gp140 was synthesized as a gBlock sequence (gBlockCT) including XhoI, KpnI and AarI restriction sites. PCR was used to extend the 3´ part of gp140 with sequences encoding for a glycine-serine linker, a hexa histidine-tag ($His_6$-tag) and stop codons to enable purification of soluble proteins. The modified gBlockCT was cloned into pcDNA5/FRT/TO ΔBsaIΔKpnI via NotI/ApaI and the KpnI and XhoI sites in the gBlockCT sequence were used to introduce a SfiI restriction site, the 3´ part of gp145 and stop codons. The resulting vector was further referred to as pLib1.1 see **Figure 13**. Next, the BG505 SOSIP R6 library template in pSEVA silent, containing the exchange cassettes were cloned into the pLib1.1 vector via HindIII and AarI, resulting in a total number of 38 pLib1.1 vectors containing exchange cassettes for linker ligation, the CT sequence and the 3´ prime parts of gp145 and gp140, respectively. The SfiI and XhoI sites together with the 3´ parts of

gp145 and gp140 offer the option to easily switch the gp160 into gp145 or gp140 constructs, since corresponding restriction sites are also present in the BG505 SOSIP R6 library template.

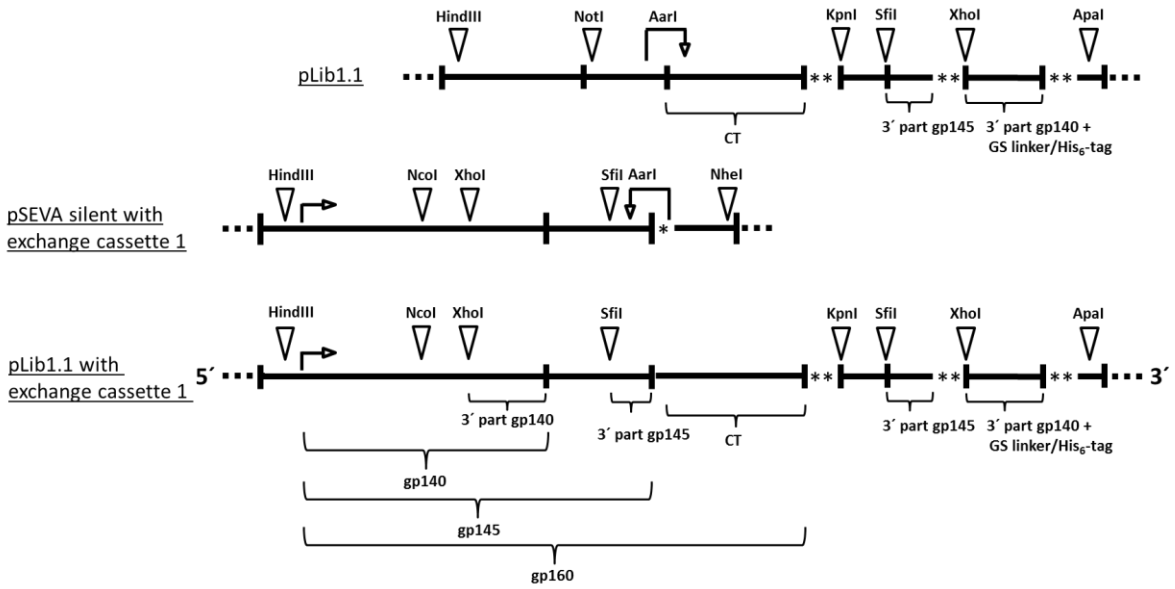



**Figure 13. Cloning of expression vector pLib1.1 with exchange cassettes.** BsaI and KpnI sites in the ampicillin resistance and the MCS of the pcDNA5/FRT/TO vector were removed by mutation. The gBlockCT sequence encoding for the Env cytoplasmic tail (CT) and the 3´ part of gp140, including XhoI, KpnI and AarI restriction sites, was further modified by extending the 3´ part of gp140 with sequences encoding for a glycine-serine (GS) linker a hexa histidine-tag ($His_6$-tag) and stop codons. NotI and ApaI were used to clone the modified gBlockCT into the pcDNA5/FRT/TO ΔBsaIΔKpnI vector. The KpnI and XhoI sites in the gBlockCT were then used to introduce the 3´ part of gp145 and a SfiI site and the resulting vector was further referred to as pLib1.1. HindIII and AarI were used to clone the BG505 SOSIP R6 library template with exchange cassettes (in pSEVA silent) into pLib1.1, like it is exemplarily shown for exchange cassette 1. In total, 38 pLib1.1 vectors were generated, containing exchange cassettes for linker ligation, the CT sequence and the 3´ parts of gp140 and gp145, respectively.

### Insertion of alanine substitutions via linker ligation

The final step in generating the BG505 SOSIP alanine substitution library was the integration of codons encoding for alanine substitutions into pLib.1.1 vectors with exchange cassettes. This was achieved via BsaI digestion of the corresponding pLib.1.1 vectors and subsequent ligation of DNA linkers (see **Appendix Table 6**), encoding for the respective BG505 SOSIP sequence with an alanine substitution at a particular position, see **Figure 14**. The linker sequences, representing annealed oligonucleotides with appropriate overlaps for particular BsaI digested exchange cassettes, were obtained from GeneArt/ThermoFisher. Taken together, BsaI digestion and linker ligation were used for seamless cloning of 615 BG505 SOSIP R6 variants in which amino acid positions 34 - 712 (according to HXB2 numbering) were consecutively substituted for an alanine with exception of SOSIP modifications, the R6 cleavage site and naturally occurring alanines and cysteines. Finally, plasmid DNA was

purified and verified by Sanger sequencing (Seqlab, Göttingen, Germany or GATC, Konstanz, Germany).

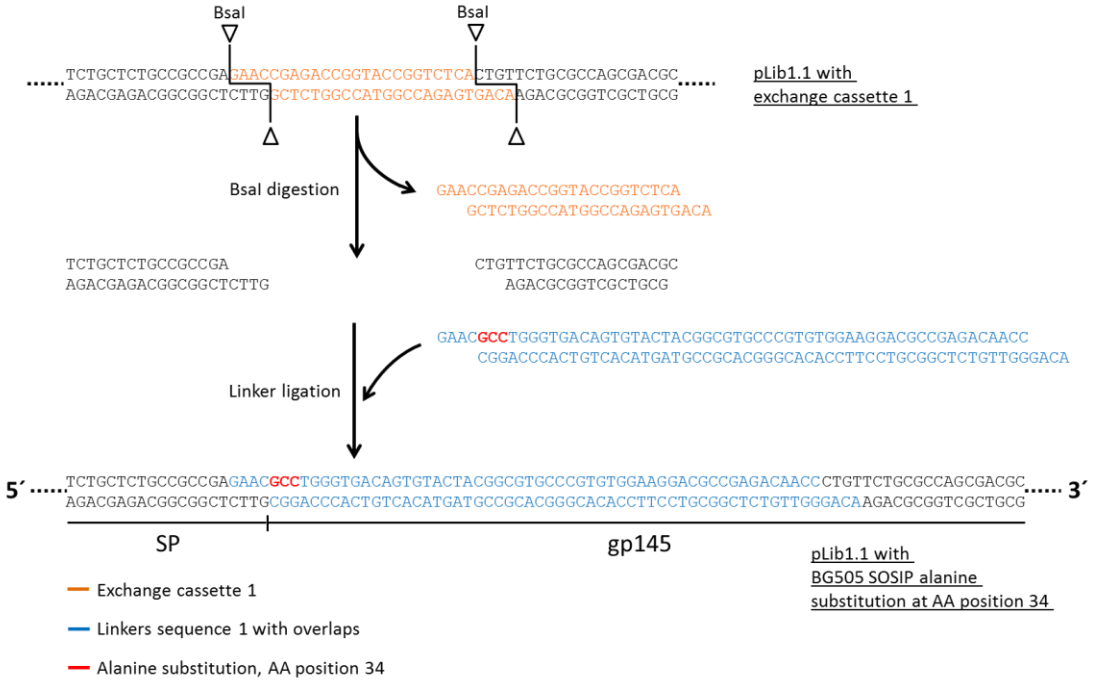



**Figure 14. Insertion of alanine substitutions via linker ligation.** To complete cloning of the BG505 SOSIP alanine substitution library, DNA linkers encoding for the respective BG505 SOSIP sequence with an alanine substitution at a particular position were cloned into pLib1.1 vectors. For this purpose, pLib1.1 vectors with exchange cassettes were digested with BsaI and DNA linkers (annealed oligonucleotides with appropriate overlaps for BsaI digested exchange cassettes) were ligated into the respective pLib1.1 vectors. The cloning procedure is exemplarily shown for pLib.1.1 with exchange cassette 1 and DNA linker 1, which encodes for the BG505 SOSIP sequence with an alanine substitution at amino acid (AA) position 34 (first amino acid after the Env signal peptide (SP), according to HXB2 numbering).

### 3.2.4 Magnetic beads purification

PCR fragments, generated during next generation sequencing (NGS) sample preparation were purified by using AMPure XP magnetic beads. To maximize DNA yield, all steps were performed in 1.5 ml DNA LoBind Tubes. Briefly, 1.8 times the sample volume of magnetic beads was added to DNA samples in $H_2O$ and thoroughly mixed by pipetting. After incubation for 5 min at room temperature (RT), samples were placed in a magnetic rack until the solution turned clear. Next, the supernatant was removed and the pellets were washed two times with 500 µl freshly prepared 70 % ethanol. To this end, the ethanol was added to the pellet and incubated for 30 sec while turning the cups in the magnetic rack. Subsequently, the supernatant was discarded without disturbing the pellet and the washing procedure was repeated. After the second washing step, the ethanol was carefully aspirated by pipetting and

the pellets were air-dried for 3 min. To elute DNA, 30 µl $H_2O$ were added, pellets were vortexed for 15 sec, centrifuged and placed in the magnetic rack for 1 min. Finally, the supernatant was transferred into a new tube.

| | |
|---|---|
| Agencourt AMPure XP | Beckman Coulter, # A63881 |
| 1.5 ml DNA LoBind Tubes | Eppendorf, Hamburg, Germany, # 0030108051 |

### 3.2.5 Next generation sequencing

NGS was used to quality control the BG505 SOSIP gp145 glycan-knockout library (see **3.3.5**), and further to identify Env variants, which were enriched during FACS analysis (see **3.3.8**). In total, *env* genes isolated from four different genomic DNA (gDNA) samples were simultaneously analyzed in one experiment using the Illumina MiSeq platform[162,163]. The preparation of the NGS experiment required several steps, which are described in the following sections.

**Generation of the amplicon library**

Two consecutive PCR reactions (PCR1 and PCR2) were used to generate an amplicon library, consisting of PCR fragments about 300 bp in size (amplicons), including NGS-specific primer binding sites, indices and adapter sequences required for NGS. PCR reactions were performed based on the protocols described in detail in the PhD thesis of Julia Koop[164]. PCR1 was used to amplify *env* genes from the gDNA templates (for gDNA preparation see **3.3.8**). Briefly, 12 primer pairs (see **Appendix Table 7**) were used to generate overlapping PCR amplicons about 300 bp in size, together covering the entire BG505 SOSIP gp145 sequence and containing NGS-specific primer binding sites. Since PCR1 was performed on four different gDNA templates this resulted in a total number of 48 amplicons (12 amplicons per gDNA template). The resulting amplicons were purified from agarose gels using the QIAquick Gel Extraction Kit and further purified by magnetic beads purification, see **3.2.4**. During PCR2 TruSeq Small RNA PCR Index Primers (see **Appendix Table 8**) were used to add index and adapter sequences to the amplicons generated in PCR1. Adapter sequences are required to hybridize samples to the surface of the Illumina flow cell, whereas indices are needed to differentiate between the analyzed samples. To this end, all 12 amplicons generated during PCR1, using a particular gDNA template, were barcoded with the very same index sequence. In total, four different indices were used (1 index per 12 related amplicons), resulting in a

total number of 48 barcoded amplicons. After gel and magnetic bead purification, the concentration of each of the amplicons was quantified using the Agilent 2100 Bioanalyzer and the Agilent High Sensitivity DNA Kit and an amplicon library was generated by pooling all 48 amplicons in an equimolar ratio to a final concentration of 30 nM.

| | |
|---|---|
| QIAquick Gel Extraction Kit | Qiagen, # 28706 |
| Agilent High Sensitivity DNA Kit | Agilent, Santa Clara, USA, # 5067-4626 |
| Agilent 2100 Bioanalyzer Instrument | Agilent |

**Next generation sequencing of the amplicon library**

Prior to loading samples on the NGS flow cell, the amplicon library was diluted and subsequently denatured. Since the utilized amplicon library is characterized by low sequence diversity, 30 % PhiX control were spiked-in during library dilution to mitigate sequencing challenges. Briefly, a mixture composed of the amplicon library (70 %) and PhiX control (30 %) with a final concentration of 2 nM in 10 µl $H_2O$ was created and subsequently denatured by adding 10 µl freshly prepared 0.2 N NaOH solution. After incubation for 5 min at RT, the mixture was diluted with chilled HT1 hybridization buffer to a final concentration of 10 pM in a total volume of 2 ml. Next, 1 ml of the 10 pM dilution was loaded on the flow cell and the NGS run was started using a previously designed MiSeq control software sample sheet with following parameter settings: 1) FastQ only; 2) sample prep kit: TruSeq Small RNA. The number of cycles in the sample sheet were manually adjusted for 2x 151 reads to enable sequencing from both the 5´and 3´end. The NGS run was performed at the laboratory of Prof. Dr. Gunter Meister under instruction of Norbert Eichner.

| | |
|---|---|
| MiSeq Reagent Kit v2 (300 cycles) | Illumina, San Diego, USA, # MS-102-2002 |
| PhiX Control v3 | Illumina, # FC-110-3001 |
| MiSeq NGS platform | Illumina |

**Analysis of NGS data**

Sequence analysis of NGS results was performed using CLC Genomics Workbench 7. After importing FastQ files into CLC, files containing corresponding forward and reverse reads were paired. Overlapping pairs were merged into one file without altering the predefined settings of the program and subsequently aligned with a reference sequence. Next, results were exported as SAM files and further investigated using an analyzer tool (programmed by

Dr. Benedikt Asbach) that was designed to count every codon at each position, delivering a table with the absolute numbers of amino acids.

| | |
|---|---|
| CLC Genomics Workbench Version 7.0.3 | Qiagen |

# 3.3 Cell biology

## 3.3.1 Cultivation of human cells

**Adherent cells**

All adherent cell lines were cultivated at 37 °C and 5 % $CO_2$ in cell culture flasks with culture medium DMEM, supplemented with additives listed in **Table 1**. Cells were grown to a confluency of 80 % and subcultured in a 1:10 ratio twice a week. For subculturing cells were washed with PBS and detached with Trypsin/EDTA solution, followed by resuspension in the respective culture medium.

| | |
|---|---|
| Cell culture flasks T25, T75, T175 | Sarstedt, Nümbrecht, Germany, # 83.3910.002, # 83.3911.002, # 83.3912.002 |
| Dulbecco`s Modified Eagle`s Medium (DMEM) | Gibco/ThermoFisher Scientific, # 11995-065 |
| Phosphate buffered saline (PBS) | 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.47 mM $KH_2PO_4$, pH 7.4 |
| Trypsin/EDTA | PAN-Biotech, Aidenbach, Germany, # P10-023500 |

**Table 1. Cell lines and culture media.**

| Cell line | Culture medium |
|---|---|
| Human Embryonic Kidney (HEK) 293T, (ATCC, # CRL-11268) | DMEM + 10 % FBS + 1 % Pen/Strep |
| Flp-In™ T-REx™ 293, (ThermoFisher Scientific, # R78007) | DMEM + 10 % FBSΔtet + 1 % Pen/Strep + 100 µg/ml zeocin + 15 µg/ml blasticidin |
| Flp-In™ T-REx™ 293 CMV_Furin+, clone [4_2][165] | DMEM + 10 % FBSΔtet + 1 % Pen/Strep + 100 µg/ml zeocin + 15 µg/ml blasticidin + 400 µg/ml geniticin sulfate |
| Stable cell lines generated from Flp-In™ T-REx™ 293 cells | DMEM + 10 % FBSΔtet + 1 % Pen/Strep + 100 µg/ml hygromycin B + 15 µg/ml blasticidin |
| Stable cell lines generated from Flp-In™ T-REx™ 293 CMV_Furin+ clone [4_2] cells | DMEM + 10 % FBSΔtet + 1 % Pen/Strep + 100 µg/ml hygromycin B + 15 µg/ml blasticidin + 400 µg/ml geniticin sulfate |
| **Additives** | **Supplier** |
| Fetal bovine serum (FBS) for HEK293T cell cultivation | Gibco/ThermoFisher Scientific, # 10270-106 |
| Tetracycline free FBS for Flp-In™ T-REx™ 293 cell cultivation (FBSΔtet) | Biochrom, Berlin, Germany, # S0115 |
| Penicillin/Streptomycin (Pen/Strep) | 10,000 U/ml penicillin, 10 mg/ml streptomycin, PAN Biotech, # P06-07100 |
| Zeocin | 100 mg/ml, InvivoGen, San Diego, USA, # ant-zn |
| Blasticidin | 10 mg/ml, InvivoGen, # ant-bl |
| Hygromycin B | 100 mg/ml, InvivoGen, # ant-hg |
| Geniticin sulfate (G418) | 100 mg/ml, InvivoGen, # ant-gn |

### 293-F suspension cells

FreeStyle™ 293-F suspension cells (293F-cells) were cultured at 37 °C, 8 % $CO_2$ and 90 rpm in 1 L Corning® Erlenmeyer cell culture flasks containing 300 ml FreeStyle™ 293 expression medium supplemented with 0.5 % Pen/Strep. For maintenance, cells were subcultured between cell densities of 0.1-1.6 Mio/ml. For subculturing, the respective amount of cells was centrifuged for 5 min at 100 g and RT, the supernatant was discarded and the cell pellet was resuspended in fresh medium.

| | |
|---|---|
| FreeStyle™ 293-F suspension cells | ThermoFisher Scientific, # R79007 |
| FreeStyle™ 293 expression medium | ThermoFisher Scientific, # 12338018 |
| Corning® Erlenmeyer cell culture flasks | Sigma-Aldrich, St. Louis, USA, # CLS431147 |
| Penicillin/Streptomycin (Pen/Strep) | PAN Biotech, # P06-07100 |

### 3.3.2 Transient transfection of human cells

**HEK293T cells**

HEK293T cells were transiently transfected using polyethyleneimine (PEI). Depending on the experiment, different transfection scales were used to obtain the desired amount of transfected cells, see **Table 2**. Cells were seeded one day before transfection to reach a confluency of about 80 % at the day of transfection. Prior to transfection, the medium was exchanged to DMEM, free of serum and antibiotics (see **Table 2**, volume DMEM for transfection). The transfection mix was prepared by dissolving DNA containing plasmids and PEI (dissolved in $H_2O$ to a final concentration of 1 mg/ml, adjusted to pH 7 with NaOH) in a 1:4 (w/w) ratio in DMEM without supplements. After incubation for 20 min at RT, the transfection mix was added to the cells. Five to six hours post transfection, the medium was exchanged to HEK293T culture medium, see **Table 1**. In case of BG505 SOSIP transfections, plasmids encoding for Env (pcDNA5/FRT/TO, pLib1.1) and Furin (pcDNA3.1(+)QL) were cotransfected at a ratio of 3:1 (w/w).

| | |
|---|---|
| Polyethyleneimine (PEI) Max linear, MW 40,000 | Polysciences, Warrington, USA, # 24765 |
| 6-well, 96-well tissue culture plates | Sarstedt, Nümbrecht, Germany, # 83.3920, # 83.3924 |

**Table 2. Transfection of different cell culture scales.**

| Culture vessel | Vol. culture medium (ml) | Number of seeded cells per well/flask | Vol. DMEM for transfection reaction (ml) | DNA (µg) | PEI (µg) | Vol. transfection mix (ml) |
|---|---|---|---|---|---|---|
| 96-well | 0.1 | $3 \times 10^4$ | 0.03 | 0.24 | 0.96 | 0.03 |
| 6-well | 2 | $5 \times 10^5$ | 1 | 2 | 8 | 0.1 |
| T175 | 30 | $1.2 \times 10^7$ | 30 | 30 | 120 | 3 |

**293-F suspension cells**

FreeStyle™ 293-F suspension cells were cultivated at 37 °C, 8 % $CO_2$ and 90 rpm in 1 L Corning® Erlenmeyer cell culture flasks containing 300 ml FreeStyle™ 293 expression medium supplemented with 0.5 % Pen/Strep. For transfection, cells were grown to a cell concentration of 1 Mio/ml at the day of transfection. Prior to transfection, cells were harvested by centrifugation (5 min at 100 g and RT). The supernatant was discarded and the pellet was resuspended in 300 ml FreeStyle™ 293 expression medium without supplements. The transfection mix was prepared by dissolving 300 µg plasmid DNA in 6.5 ml DMEM, free

of serum and antibiotics. Simultaneously 1.2 ml PEI solution (PEI diluted to a final concentration of 1 mg/ml in PEI buffer, pH 7.5) was added to 6.5 ml DMEM without supplements. After incubation for 5 min at RT, PEI and DNA solutions were mixed, incubated for 20 min at RT and then added to the cells. Five to six hours post transfection, the medium was exchanged to FreeStyle™ 293 expression medium supplemented with 0.5 % Pen/Strep.

| | |
|---|---|
| Polyethyleneimine (PEI), linear, MW 25,000 | Polysciences, Warrington, USA, # 23966 |
| PEI buffer | 25 mM HEPES pH 7.5, 150 mM NACl |

### 3.3.3 Stable transfection of Flp-In™ T-REx™ 293 CMV_Furin+ cells

The constitutive Furin overexpressing cell line Flp-In™ T-REx™ 293 CMV_Furin+ [clone 4_2], which was previously established in our lab (master`s thesis Dominik Fiegle[165]), was used to generate stable cell lines with inducible Env expression regulated by the Tet operator/repressor system. Flp-In™ T-REx™ 293 CMV_Furin+ cells were transfected as described for HEK293T cells in **3.3.2**, with the difference that the plasmid mixture contained the respective *env* variant in pLib1.1 and the pOG44 helper plasmid, encoding for the recombinase, in a pLib1.1/pOG44 ratio of 4:1 (w/w). Five to six hours post transfection the medium was exchanged to DMEM + 10 % FBSΔtet + 1 % Pen/Strep. 24 h post transfection the medium was refreshed. 48 h post transfection the cells were detached by pipetting and directly transferred into cell culture flasks containing selection medium to select for cells with stable integration events. In case of 6-well transfections, cells were transferred to T75 flasks. After 24 hours the selection medium was refreshed and selection was carried out for 24 days with medium being refreshed every 4 days.

| | |
|---|---|
| Selection/culture medium | DMEM + 10 % FBSΔtet + 1 % Pen/Strep + 100 µg/ml hygromycin B + 15 µg/ml blasticidin + 400 µg/ml geniticin sulfate |

### 3.3.4 Induction of Env expression in stable cell lines

Cells were seeded one day before induction to reach a confluency of about 80 % at the day of induction. Depending on the experiment, different amounts of cells were seeded to obtain the

desired amount of cells. For the amount of seeded cells in different cell culture scales, see **Table 2**. To induce Env expression, the medium was refreshed to the respective cell culture medium (see **Table 1**) containing 1 µg/ml doxycycline.

| | |
|---|---|
| Doxycycline | Sigma-Aldrich, # D9891 |

### 3.3.5 Generation of the BG505 SOSIP gp145 glycan-knockout library expressed by stable cell lines

The BG505 SOSIP contains 28 N-linked potential glycosylation sites per gp145 protomer (84 glycosylation sites per trimer) which are defined by (NXT/S) sequons in the amino acid sequence, with X being any amino acid except possibly proline or aspartic acid[48,49]. The BG505 SOSIP gp160 alanine substitution library (see **3.2.3**) was used as a template to generate a BG505 SOSIP gp145 glycan-knockout library, including all variants which encode for alanine substitutions in the glycosylation sequons. Since the alanine substitution of N and T/S should prevent the glycosylation of the asparagine in the sequon, both knockout variants were included into the library, resulting in two BG505 SOSIP glycan-knockout variants for a particular glycosylation site and a total number of 56 BG505 SOSIP glycan-knockout variants. Further, alanine variants were included which knockout two potential O-linked glycosylation sites at positions T499 and T606 (HXB2 numbering), respectively[50,51]. O-linked glycosylation can take place on serine or threonine residues and, contrary to N-linked glycosylation sites, there is no clear-cut motif in the amino acid sequence which defines their glycosylation[52]. Using the respective BG505 SOSIP gp160 alanine substitution variants as starting material, each of the 58 Env variants described above was cloned from gp160 to gp145 constructs by simple restriction digest with SfiI and subsequent re-ligation as described in **3.2.3**. The resulting BG505 SOSIP gp145 constructs were stably transfected into the constitutive Furin overexpressing cell line Flp-In™ T-REx™ 293 CMV_Furin+ (see **3.3.3**), thereby generating stable cell lines with doxycycline inducible Env expression, each expressing a particular Env variant with a single glycan deletion in the gp145 protomer (3 deletions per trimer). After completing the generation of cell lines, cells of each cell line were cryopreserved and stored in liquid nitrogen according to manufacturer`s recommendations for Flp-In™ T-REx™ 293 cells (see **3.3.1**, **Table 1**). Furthermore, cells of all cell lines were

pooled to equal numbers and the resulting pool was further referred to as BG505 SOSIP gp145 glycan-knockout library.

### 3.3.6 Flow cytometry

Transiently transfected cells (see **3.3.2**, HEK293T cells) or doxycycline induced stable cell lines (see **3.3.4**) were analyzed by flow cytometry 48 h post transfection/induction. Briefly, cells were washed and detached with an appropriate amount of chilled FACS buffer and 7 x $10^4$ to 1 x $10^5$ cells were transferred to a 96-well round bottom plate and centrifuged for 5 min at 500 g and 4 °C. After removal of the supernatant, cells were stained with the respective Alexa647- or Alexa488-labeled antibodies (for antibody labeling see **3.4.7**). For this purpose, antibodies were diluted in FACS buffer to desired concentrations and cells were stained with 30 µl antibody dilution per well for 30 min at 4 °C. Experiments with the antibody 17b, in some cases, required pre-incubation of cells with the soluble CD4 receptor (sCD4) to trigger the CD4-induced (CD4i) state of cell-surface presented Env molecules. For this purpose, cells were incubated with 25 µl sCD4, diluted in FACS buffer to a final concentration of 8.6 µg/ml, for 30 min at 4 °C. Next, 5 µl of 17b were added at desired concentrations and the cells were incubated for another 30 min at 4 °C. After antibody incubation, three washing cycles were performed to remove unbound antibody. Washing was performed by adding 200 µl FACS buffer per well to the cells, followed by centrifugation for 5 min at 500 g and 4 °C. Subsequently, cells were resuspended in 200 µl FACS buffer and analyzed using the Attune NxT flow cytometer. Cells were gated for single (FSC-A/FSC-H plot), living cells (FSC-A/SSC-A plot) as shown in **Figure 15**.

| | |
|---|---|
| FACS buffer | PBS, 1 % (v/v) heat inactivated FBS, 2 mM EDTA, 0.1 % (w/v) $NaN_3$ |
| Fetal bovine serum (FBS) | Gibco/ThermoFisher Scientific, # 10270-106 |
| Soluble CD4 receptor (sCD4) | National Institutes of Health (NIH) AIDS Reagent Program, Germantown, Maryland, USA, # 4615 |
| 96-well tissue culture plates (round bottom) | Sarstedt, Nümbrecht, Germany, # 83.3925 |
| Attune NxT flow cytometer | ThermoFisher Scientific |
| Anti-Env antibodies | See **3.4.6**, **Table 4** |

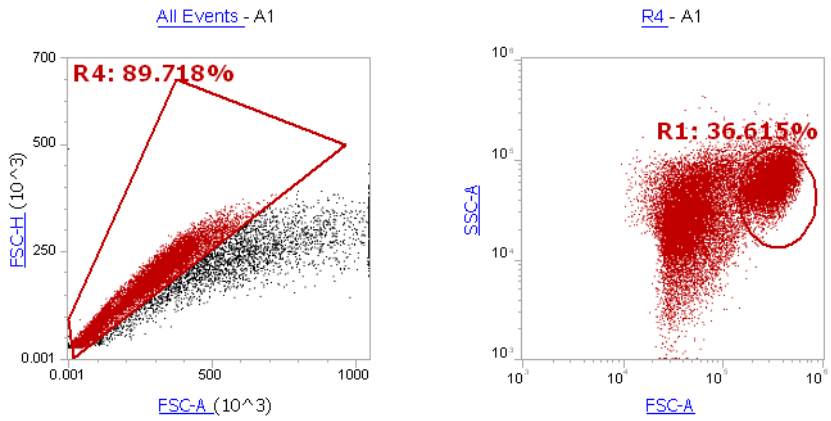



**Figure 15. Gating strategy for single (FSC-A/FSC-H plot, left panel), living cells (FSC-A/SSC-A plot, right panel).**

### 3.3.7 Flow cytometry-based antibody titration

For flow cytometry-based titration experiments, HEK293T cells transiently transfected with plasmids encoding for BG505 SOSIP gp145/160 (see **3.3.2**) or doxycycline induced stable cell lines expressing BG505 SOSIP gp145/160 (see **3.3.4**), were stained with fluorescent-labeled antibodies at desired concentrations and analyzed as described in **3.3.6**. Mock transfected HEK293T cells (pcDNA5/FRT/TO without *env* + pcDNA3.1(+) QL with *furin*) and the constitutive Furin overexpressing cell line Flp-In™ T-REx™ 293 CMV_Furin+ (doxycycline induced, without integrated *env* gene) were used as negative controls for transiently transfected cells and stable cell lines, respectively. The mean fluorescence intensity (MFI) of Env expressing cells stained with the titration antibody was corrected for the antibody background signal by subtracting the respective negative control signals (negative controls were titrated with the very same antibody concentrations) and is further referred to as "referenced" MFI. In addition, to compare antibody binding with regard to varying Env surface expression levels (between transiently transfected cells, stable cell lines and Env variants) the referenced MFI was further normalized for Env surface expression. To this end, cells of each variant were stained in separate wells with 250 nM mAb 2G12 which was used as reference antibody and the respective referenced MFI signal of the titration antibody was divided by the respective referenced MFI signal of 2G12 and is further referred to as "relative MFI". For curve fit, MFI values described above were fitted using non-linear least squares regression (one site binding, hyperbola; GraphPad Prism 5).

GraphPad Prism, version 5.04 GraphPad Software, La Jolla, USA

### 3.3.8 FACS analysis of the BG505 SOSIP gp145 glycan-knockout library

1.2 x $10^7$ cells of the BG505 SOSIP gp145 glycan-knockout library (pool of single cell lines, see **3.3.5**) were seeded in a T175 flask and induced with doxycycline the following day (see **3.3.4**). 48 h post induction, cells were washed with chilled FACS buffer, subsequently detached with 10 ml chilled FACS buffer and centrifuged for 5 min at 500 g and 4 °C. After removal of the supernatant, cells were co-stained with Alexa488-labeled PGT135 (screening antibody) and Alexa647-labeled VRC01 (reference antibody for normalization of Env surface expression). Briefly, 1 ml staining solution containing 100 nM PGT135 and 250 nM VRC01 diluted in FACS buffer was prepared and cells were gently resuspended in the staining solution and incubated for 30 min at 4 °C. After incubation, three washing cycles were performed. For this purpose, 10 ml FACS buffer was added to the cells and after gentle resuspension cells were centrifuged for 5 min at 500 g and 4 °C. After the last washing step, cells were resuspended in 600 µl PBS (optimal concentration for sorting 2 x $10^7$ cells/ml) and singularized by passing cells through a 30 µm pre-separation filter. 50 µl of the sample material was collected to serve as input control and the remaining sample was sorted at the Regensburg Center for Interventional Immunology (RCI) using the FACS Aria IIu device. The instrument was set to "single cell mode" to discard two-target-events and obtain the most accurate counts of the sorting procedure[160]. The following gating strategy was applied: 1) FSC-A/SSC-A plot to gate for living cells (P1); 2) FSC-A/FSC-W plot to gate for single cells (P2); 3) FITC-A/APC-A plot with sorting gate P3, allowing the sorting of cells with the lowest PGT135 signal in relation to their VRC01 signal, and thus the selection of cells with lowest affinity to PGT135 in relation to their Env surface expression; 4) FSC-A/SSC-A plot with gate P4, representing the sorted, living cells that were collected in 100 µl PBS. In total, about 1.5 x $10^7$ cells were sorted from which about 3 x $10^5$cells were sorted into gate P4. The genomic DNA from about 2-3 x $10^5$ cells of the input control and sorted cells was recovered using the QIAamp DNA Mini Kit (protocol for DNA purification from blood and body fluids) according to manufacturer`s instructions. Finally, the genomic DNA was further analyzed by NGS, see **3.2.5**.

| | |
|---|---|
| FACS buffer | PBS, 1 % (v/v) heat inactivated FBS, 2 mM EDTA, 0.1 % (w/v) $NaN_3$ |
| Anti-Env antibodies | See **3.4.6**, **Table 4** |
| Antibody labeling | See **3.4.7** |
| Pre-Separation Filters (30 µm) | Miltenyi Biotec, Bergisch Gladbach, Germany, # 130-041-407 |
| FACS Aria IIu | Becton Dickinson (BD), Franklin Lakes, USA |
| QIAamp DNA Mini Kit | Qiagen, # 51306 |

# 3.4 Protein biochemistry

## 3.4.1 Env constructs

Gp140 Env/V3 constructs were generated by deleting the transmembrane domain from the respective gp145 containing plasmids via PCR. Additionally, gp140 Env/V3 variants were engineered via PCR to contain a flexible glycine-serine linker and Flag/hexa histidine-tag (Flag/$His_6$-tag) at the C-terminus of gp41 by adding amino acid sequence SGGGGSDYKDDDDKHHHHHH and the stop codon. Furthermore, the Furin cleavage site was altered from REKR to REKS via overlap extension PCR. For generation of gp120 constructs, gp41 was deleted by PCR and in contrast to gp140 constructs no tag was added. Env constructs were re-cloned into the pQL13 vector by golden gate cloning, see **3.2.1**. The previously described BG505 SOSIP.664 $gp140^{61}$ was engineered to contain a glycine-serine linker and a $His_6$-tag at the C-terminus of the gp41 ectodomain ($gp41_{ECTO}$) by adding amino acid sequence GSHHHHHH and the stop codon. Cloning was performed by XhoI digestion and subsequent re-ligation of membrane-bound BG505 SOSIP gp145 in pLib1.1 as described in **3.2.3**.

## 3.4.2 Expression and purification of soluble Env proteins

Soluble gp120 and gp140 proteins were expressed into the supernatant of 293-F suspension cells. For this purpose, 293-F suspension cells were transiently transfected with plasmids encoding for gp120 or gp140 Env/V3 chimeras (pQL13) or BG505 SOSIP.664 (pLib1.1) (see **3.4.1**), as described in **3.3.2**. In case of BG505 SOSIP.664 transfection, plasmids encoding for Env (pLib1.1) and Furin (pcDNA3.1(+)QL) were cotransfected at a ratio of 3:1 (w/w) to confer proper gp120-$gp41_{ECTO}$ cleavage. 72 h post transfection, culture supernatants were harvested by centrifugation for 10 min at 1,000 g and 4 °C, followed by filtration with Filtropur 0.22 µm filters. For affinity purification of gp120, culture supernatants were loaded onto a 5 ml lectin affinity column at a constant flow rate of 1 ml/min at 4 °C. Unbound

protein was washed out with 5 column volumes (CV) PBS, 1 mM EDTA, 1 mM EGTA. Env proteins were eluted with PBS, 1 mM EDTA, 1 mM EGTA, 1 M methyl-α-D-mannopyranoside and protein containing fractions were subsequently concentrated to a volume of 0.5 to 1.0 ml using 30 kDa Amicon® Ultra-4 centrifugal filters. Affinity purification of His-tagged gp140 was performed in a similar way with the exceptions of using a 5 ml Ni-sepharose column with corresponding buffers for washing (PBS, 0.5 M NaCl, 0.03 M imidazole, pH 7.4) and elution (PBS, 0.5 M NaCl, 0.5 M imidazole, pH 7.4) and 100 kDa Amicon® Ultra-15 centrifugal filters for concentration. Env proteins were purified to monomer (gp120) or trimer size homogeneity (gp140) using a HiPrep 16/60 Sephacryl S-300 HR size exclusion chromatography column at a constant flowrate of 0.5 ml/min with running buffer HBS-EP or PBS for gp120 and gp140, respectively. Gp120 monomer or gp140 trimer containing fractions were pooled, concentrated (only in case of gp140 trimers) using 30 kDa Amicon® Ultra-4 centrifugal filters and stored at -80 °C. Protein concentrations were determined by photometric quantification (absorbance at 280 nm) using a NanoDrop ND-1000 photometer. Photometric quantification was performed in consideration of extinction coefficients and molecular weights of gp120 monomers and gp140 trimers (see **Table 3**) which were calculated based on the corresponding amino acid sequences using the ExPASy protParam tool[166].

| | |
|---|---|
| Filtropur 0.22 µm filters | Filtropur L 0.2 LS, Sarstedt, Nümbrecht, Germany, # 83.1827.001 |
| Agarose-bound *Galanthus nivalis* lectin for 5 ml affinity column | Vector Laboratories, Burlingame, USA, # AL-1243 |
| Methyl-α-D-mannopyranoside | Merck Millipore, # 462711 |
| 30 kDa Amicon® Ultra-4 centrifugal filters | Merck Millipore, # UFC803024 |
| 5 ml Ni-sepharose column, HisTrap™ High Performance 5 ml | GE Healthcare Life Sciences, # 17-5248-01 |
| 100 kDa Amicon® Ultra-15 centrifugal filters | Merck Millipore, # UFC910024 |
| Size exclusion chromatography column, HiPrep 16/60 Sephacryl S-300 HR | GE Healthcare Life Sciences, # 17116701 |
| HBS-EP | 0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005 % (v/v) Surfactant P20 |

**Table 3. Extinction coefficients and molecular weights of Env proteins calculated using ExPASy protParam tool.**

| Env Protein | Extinction coefficient gp120 monomer | Molecular weight (Da) gp120 monomer | Extinction coefficient gp140 trimer | Molecular weight (Da) gp140 trimer |
|---|---|---|---|---|
| MN | 68965 | 54582 | 427620 | 229559 |
| HXB2 | 65985 | 54618 | 418680 | 229668 |
| CDC42 | 72975 | 54503 | 439650 | 229322 |
| RF | 67475 | 54337 | 423150 | 228823 |
| SF33 | 68965 | 54477 | 427620 | 229244 |
| BG505 SOSIP.664 | | | 335175 | 215306 |

### 3.4.3 Production and purification of HIV-1 Gag VLPs

VLPs were produced by transient transfection of 293-F cells, see **3.3.2**, with plasmids encoding for optimized HIV-1 Gag (TeeGag1[167] in pcDNA3.1(+)) and cleavage competent gp145 versions of the Env/V3 chimeras (in pQL13) at a ratio of 2:1 (w/w). 48 h post-transfection VLP containing supernatants were cleared from cells by centrifugation for 15 min at 3,000 g and 4 $^0$C. To concentrate VLPs, 30 ml of supernatants were loaded onto 5 ml 30 % (w/w) sucrose in PBS cushion and ultra-centrifuged for 2 h at 100,000 g and 4 $^0$C. The pellets were resuspended in an appropriate amount of PBS overnight at 4 $^0$C. To separate VLPs according to their density, VLPs were loaded onto a 10 to 50 % (w/w) sucrose in PBS gradient (2 ml of each, 10 %, 20 %, 30 %, 40% and 50 %) and centrifuged for 2.5 h at 100,000 g and 4 $^0$C. 22 fractions of 550 µl each were collected and analyzed for their density (weighing each fraction, using an electronic analytical balance) and co-banding of Gag and Env (analyzing 10 µl of each fraction by Western blot, see **3.4.5**). VLP containing fractions were pooled, diluted in 35 ml PBS and centrifuged for 2 h at 100,000 g and 4 $^0$C. Finally, VLP pellets were resuspended in 400 µl PBS containing 15 % (w/w) trehalose over night at 4 $^0$C and stored at - 80 $^0$C. To approximately quantify the amount of Gag in the VLP preparations a densitometric analysis was performed. For this purpose, 3 µl of the VLP preparations and BSA reference samples were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), see **3.4.5**. After coomassie staining, see **3.4.4**, the band intensities were densitometrically analyzed using ImageJ. Further, to semi-quantitavely analyze the Env concentrations in the VLP preparations a slot blot experiment was performed. Briefly, 0.25 – 2 µl of the VLP preparations and a serial dilution of a soluble gp140 standard protein (96ZM651) were loaded onto a slot blot, see **3.4.5**, and the band intensities were analyzed by densitometric analysis, using ImageJ[168].

| | |
|---|---|
| Centrifuge tubes for sucrose cushion: 38.5 ml, 25 x 89 mm | Beckman Coulter, Brea, USA, # 326823 |
| Centrifuge tubes for sucrose in PBS gradient: 14 ml, 14 x 95 mm | Beckman Coulter, # 344060 |
| Ultracentrifuge: Optima™ L-90K with rotors SW 32 Ti and SW 40 Ti for 38.5 ml and 14 ml tubes, respectively | Beckman Coulter |
| Electronic analytical balance: A200S | Sartorius AG, Göttingen, Germany |
| Bovine serum albumin (BSA), fraction V | Biomol, Hamburg, Germany, # 01400 |
| ImageJ 1.48v | National Institutes of Health (NIH), Bethesda, USA |

### 3.4.4 Blue Native PAGE

The purity of gp120 monomers and gp140 trimers was analyzed by Blue native polyacrylamide gel electrophoresis (BN-PAGE). In brief, 3 µg Env proteins were mixed with 2x loading dye and directly loaded on precast SERVAGel$^{TM}$ N 4-16 native Gels. BN-PAGEs were performed in running buffers for 2 h at 200 V, followed by staining with Coomassie blue solution.

| | |
|---|---|
| 2x loading dye | 100 mM MOPS, 100 mM Tris-HCl, pH 7.7, 40 % glycerol, 0.2 % Coomassie Brilliant Blue R-250 |
| Coomassie Brilliant Blue R-250 | AppliChem GmbH, Darmstadt, Germany, # A1092 |
| SERVAGel$^{TM}$ N 4-16, Vertical Native Gels 4-16 % | SERVA Electrophoresis, Heidelberg, Germany, # 43253 |
| Anode buffer | 50 mM Bis Tris-HCl pH 7.0 |
| Cathode buffer | 50 mM Tricine, 15 mM Bis Tris |
| Coomassie blue solution | 0.125 % (w/v) Coomassie Brilliant Blue R-250, 50 % (v/v) ethanol, 7 % (v/v) acetic acid |

### 3.4.5 SDS-PAGE, Western Blot and Slot blot analysis

Proteins were separated according to their size via SDS-PAGE, as described previously[169]. Prior to loading on SDS gels (12.5 %), samples were mixed with 2x Laemmli buffer and incubated for 5 min at 95 °C. PageRuler$^{TM}$ Plus Prestained Protein Ladder was used as molecular weight marker. Electrophoresis was started by applying 90 V for 30 min (pre-run). Afterwards, the voltage was adjusted to 140 V, until the tracking dye reached the bottom of the gel. Western Blot transfer of SDS-PAGE separated proteins to nitrocellulose membranes was performed using the BlueFlash-L semi-dry blotting unit (SERVA Electrophoresis,

Heidelberg, Germany), according to manufacturer`s instructions (1.5 mA/cm$^2$ gel for 1 h). For slot blots, protein solutions were transferred to nitrocellulose membranes using the Bio-Dot® SF microfiltration apparatus (Bio-Rad, Hercules, USA). Briefly, the membrane was incubated for 10 min in transfer buffer and placed in the Bio-Dot apparatus, including five layers of transfer-buffer-moistened Whatman blotting papers. A vacuum of 800 mbar was applied and after a washing step with 250 µl transfer buffer protein samples, diluted to a final volume of 200 µl in PBS, were absorbed onto the membrane, followed by a further washing step with transfer buffer. Both, nitrocellulose membranes from Western Blot and slot blot experiments, were blocked in TBS-M5 overnight at 4 °C. For immunodetection, membranes were washed four times with TBS-T, followed by incubation with primary antibodies in TBS-M1 for 1h at RT. After washing with TBS-T, membranes were incubated with secondary antibodies (in TBS-M1) for 1 h at RT. Membranes were washed with TBS-T, followed by incubation with ECL substrate solution for 5 min at RT. After removal of the substrate solution the chemiluminescent signal of the peroxidase reaction was detected with a ChemiluxPro device (Intas, Göttingen, Germany).

| | |
|---|---|
| 2x Laemmli buffer | 125 mM Tris, 2 % (w/v) SDS, 10 % (v/v) β-mercaptoethanol, 1 mM EDTA, 10 % (w/v) glycerol, 0.01 % (w/v) bromophenol blue, pH 6.8 |
| PageRuler™ Plus Prestained Protein Ladder | ThermoFisher Scientific, # 26620 |
| Nitrocellulose membranes, Amersham™ Protran® 0.2 µm | GE Healthcare Life Sciences, # 10600001 |
| Transfer buffer | 25 mM Tris, 150 mM glycine, 10 % (v/v) methanol |
| Whatman™ 3MM Chr Blotting Paper | GE Healthcare Life Sciences, # 3030-917 |
| Tris buffered saline (TBS) | 150 mM NaCl, 50 mM Tris/HCl, pH 7.4 |
| TBS-M5 | TBS, 5 % (w/v) skim milk powder |
| TBS-M1 | TBS, 1 % (w/v) skim milk powder |
| TBS-T | TBS, 0.3 % (v/v) Tween-20 |
| ECL substrate solution | 100 mM Tris/HCl pH 8.5, 12.5 mM luminol, 1.98 mM coumaric acid, 305 µl/l 30 % (v/v) $H_2O_2$ |
| Primary α-gp41antibody cocktail for Env detection | Human monoclonal antibodies10E8, 2F5, 4E10 and 5F3, see **Table 4**, each at 1 µg/ml |
| Primary antibody for p24/p55 detection | Clone CB-13/5[170,171], monoclonal mouse IgG1 from hybridoma cell line, dilution 1:1,000. |
| Secondary antibodies | Polyclonal horseradish peroxidase (HRP)-conjugated rabbit anti-human and rabbit anti-mouse immunoglobulin (Ig) fractions (1:2,000 each), Dako, Agilent, Santa Clara, USA, # P0214, # P0260) |

### 3.4.6 Anti-Env antibodies

All human monoclonal antibodies targeting the HIV-1 Env protein that were used in this work are listed in **Table 4**. Monoclonal antibodies PG9, PGT145, PGT135, PGT151, VRC01, 17b, F105, and 10E8 were produced in house. Briefly, 293-F cells were transfected with the corresponding heavy and light chain genes (both in CMVR expression vectors) in a 1:1 (w/w) ratio, using PEI transfection, see **3.3.2**. 72 h post transfection, culture supernatants were harvested by centrifugation for 10 min at 1,000 g and 4 °C, followed by filtration with Filtropur 0.22 µm filters. The supernatants were loaded onto a 1 ml HiTrap® rProtein A FF column at a constant flowrate of 1ml/min. After washing with 5 CV PBS, antibodies were eluted with 5 CV 0.1 M glycine pH 3.2, followed by neutralization with 1 M Tris-HCl pH 9. In case of PGT145 elution was performed with 0.1 M glycine pH 3.2, 0.1 M fructose. Purified mAbs were buffer exchanged to PBS or 0.1 M $NaHCO_3$, 0.5 M NaCl, pH 8.3 in case of PGT145, using PD-10 columns. Antibody concentrations were determined by photometric quantification (absorbance at 280 nm) using a NanoDrop ND-1000 photometer. For MST and ELISAs antibody concentrations were recorded in nM with the assumption that the average size of an IgG molecule is approximately 150 kDa.

| | |
|---|---|
| CMVR VRC01 mAb heavy chain and light chain expression vectors | NIH AIDS Reagent Program, Germantown, Maryland, USA, # 12035, # 12036 |
| Filtropur 0.22 µm filters | Filtropur L 0.2 LS, Sarstedt, Nümbrecht, Germany, # 83.1827.001 |
| HiTrap® rProtein A FF, 1 ml | GE Healthcare Life Sciences, # 17-5079-01 |
| PD-10 Columns | GE Healthcare Life Sciences, # 17-0851-01 |

**Table 4. Human monoclonal antibodies targeting the HIV-1 Env protein.**

| Antibody | Specificity | Supplier |
|---|---|---|
| 10E8[172,173] | gp41 MPER | In house production |
| 4E10[101,102] | gp41MPER | Polymun Scientific, Klosterneuburg, Austria, # AB004 |
| 2F5[101,104,103] | gp41 MPER | Polymun Scientific, Klosterneuburg, Austria, # AB001 |
| 5F3[101,174] | gp41 | Polymun Scientific, Klosterneuburg, Austria, # AB010 |
| PGT151[107,108] | gp41-gp120 interface | In house production |
| PG9[175,176,177] | V1/V2 | In house production |
| PGT145[178,179] | V1/V2 | In house production |
| PGT135[178,180] | N332 glycan supersite | In house production |
| 2G12[101,105] | N332 glycan supersite | Polymun Scientific, Klosterneuburg, Austria, # AB002 |
| VRC01[181,182] | CD4 binding site | In house production |
| F105[183,184] | CD4 binding site | In house production |
| 17b[185,186] | CD4 induced | In house production |
| 447-52D[187,188] | V3 loop | Polymun Scientific, Klosterneuburg, Austria, # AB014 |

### 3.4.7 Protein labeling

Antibodies used in MST, flow cytometry or FACS analysis were labeled with fluorescent dyes Alexa Fluor™ 647 or Alexa Fluor™ 488 according to manufacturer`s instructions. For MST analysis VLPs were labeled using the Monolith NT™ Protein Labeling Kit RED-NHS according to manufacturer`s instructions, except for step B in which 10 µl of VLPs were mixed with 190 µl of 10.9 µM DY-647P1-NHS-Ester in PBS pH 7.4 containing 2.6 % (v/v) DMSO. After an incubation time of 30 min at RT in the dark, VLPs were eluted into HBS-EP buffer. Antibodies used in serum ELISAs were biotinylated using the EZ-Link™ NHS-LC-Biotin reagent. Briefly, EZ-Link™ NHS-LC-Biotin was dissolved in dimethylformamide to a final concentration of 10 mM, followed by adding the biotin reagent in a 75 molar excess to the antibodies in PBS (40.5 µl 10 mM EZ-Link™ NHS-LC-Biotin per 1 mg antibody). After an incubation time for 1 h at RT, the buffer was exchanged to PBS using NAP™-5 gelfiltration columns and the antibody concentration was determined using a NanoDrop ND-1000 photometer (absorbance at 280 nm).

| | |
|---|---|
| Alexa Fluor™ 647 Protein Labeling Kit | ThermoFisher Scientific, # A20173 |
| Alexa Fluor™ 488 Protein Labeling Kit | ThermoFisher Scientific, # A10235 |
| Monolith NT™ Protein Labeling Kit RED-NHS | NanoTemper Technologies, Munich, Germany, # MO-L001 |
| DY-647P1-NHS-Ester | Dyomics, Jena, Germany, # 647P1-01 |
| HBS-EP | 0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005 % (v/v) Surfactant P20 |
| EZ-Link™ NHS-LC-Biotin | ThermoFisher Scientific, # 21336 |
| NAP™-5 gelfiltration columns | GE Healthcare Life Sciences, # 17-0853-01 |

### 3.4.8 Poly(lactic acid) particles

Poly(lactic acid) particles (PLA particles) with adsorbed BG505 SOSIP.664 trimers were produced and characterized in the lab of our cooperation partner Bernard Verrier as described below. Empty PLA particles, prepared by nanoprecipitation method without any surfactant or stabilizer, were purchased from Adjuvatis. For adsorption of BG505 SOSIP.664 trimers, the protein was diluted in Dulbecco`s phosphate-buffered saline (DPBS) to final a concentration of 100 µg/ml and incubated with PLA particles (10 mg/ml in DPBS) at a 1:1 v/v ratio for 2 h at RT under moderate end-overhead stirring. To calculate the amount of adsorbed BG505 SOSIP.664 trimers, unbound protein was collected in the supernatant by centrifugation at 10,000 g for 10 min and quantified using the Micro BCA™ Protein Assay Kit. The absorbance was measured at 562 nm using a microplate reader. To further characterize the particle preparations, the hydrodynamic diameter and size distribution (polydispersity index) were determined by dynamic light scattering (DLS). For this purpose, particle preparations were diluted in 1 mM NaCl, 0.33 µm filtered, and analyzed using the Zetasizer Nano ZS (measurement at 25 °C, scattering angle of 173°). Each particle preparation was analyzed in four independent measurements and the mean value was calculated. Furthermore, the electrophoretic mobilities (zeta potential) were measured using the Zetasizer Nano ZS but at an angle of 12.8°. In summary, two batches of PLA particles with adsorbed BG505 SOSIP.664 trimers were produced. The average hydrodynamic diameter in the two independent preparations was 175 ± 12 nm in size (average size of two batches of empty control particles: 136 ± 2 nm) and the average polydispersity index was 0.197 ± 0.028, indicating a relatively homogeneous size distribution (control particles: 0.091 ± 0.011). The average zeta potential of the two particle preparations was -50 ± 5 mV (control particles: -45 ± 10 mV) and the amount of BG505 SOSIP.664 trimers was calculated about around 47 ± 3 trimers per particle.

| | |
|---|---|
| PLA particles (i-Particles®) | Adjuvatis, Lyon, France |
| Dulbecco`s phosphate-buffered saline (DPBS) | Gibco/ThermoFisher Scientific, # 14190-086 |
| Micro BCA™ Protein Assay Kit | ThermoFisher Scientific, # 23235 |
| Multiskan™ FC microplate photometer | ThermoFisher Scientific |
| Zetasizer Nano ZS | Malvern Panalytical, Malvern, UK |

### 3.4.9 ELISA

**ELISA with soluble gp120 and gp140 Env/V3 chimeras**

Nunc MaxiSorp® flat-bottom 96-well plates were coated with 200 ng *Galanthus nivalis* lectin (GNL) in PBS per well over night at 4 °C. After coating, plates were washed with three well-volumes PBS-T using the Tecan HydroFlex microplate washer and subsequently blocked with 200 µl blocking buffer per well for 2 h at RT. Plates were washed with three well-volumes PBS-T and incubated with 50 ng Env in PBS per well for 1h at RT. After washing with six well-volumes PBS-T, 50 µl mAb 447-52D in binding buffer per well was titrated and incubated for 1h at RT. Plates were washed with ten well-volumes PBS-T to remove unbound antibodies and antibody binding was detected by incubating plates with 50 µl polyclonal horseradish peroxidase (HRP)-conjugated rabbit anti-human Ig (1:5,000 in binding buffer) per well for 30 min at RT. After washing with ten well-volumes PBS-T, plates were developed with 50 µl 3,3′,5,5′-tetramethylbenzidine (TMB) substrate per well, which was freshly prepared by mixing TMB A and TMB B in a 20:1 ratio. The colorimetric reaction was stopped by adding 25 µl 1 M $H_2SO_4$ per well and the absorbance was measured at 450 nm using a microplate reader. The resulting data were corrected by subtracting negative control signals (wells without Env proteins, titrated with the very same antibody concentrations) and fitted using non-linear least squares regression (one site binding, hyperbola), GraphPad Prism 5.

| | |
|---|---|
| Nunc MaxiSorp® flat-bottom 96-well plates | ThermoFisher Scientific, # 442404 |
| *Galanthus nivalis* lectin (GNL) | Sigma-Aldrich, # L8275 |
| PBS-T | PBS, 0.05 % (v/v) Tween-20 |
| Blocking buffer | PBS, 5 % (w/v) skim milk powder, 0.1 % (v/v) Tween-20 |
| Binding buffer | PBS, 1 % (w/v) BSA |
| Anti-Env antibody 447-52D | See **3.4.6**, **Table 4** |
| Detection antibody | Polyclonal HRP-conjugated rabbit anti-human Ig fraction, Dako, Agilent, Santa Clara, USA, # P0214 |
| TMB A | 30 mM tri-potassium citrate-monohydrate adjusted to pH 4.1 with 10 % (w/v) citric acid |
| TMB B | 0.24 % (w/v) TMB, 10 % (v/v) acetone, 90 % (v/v) ethanol, 80 mM $H_2O_2$ |
| HydroFlex microplate washer | Tecan, Männedorf, Switzerland |
| Microplate reader Model 680 | Bio-Rad Laboratories, Hercules, USA |
| GraphPad Prism, version 5.04 | GraphPad Software, La Jolla, USA |

**His-capture and direct-ELISA with BG505 SOSIP.664 trimers**

His-capture ELISAs with BG505 SOSIP.664 trimers were performed as described above for ELISAs with Env/V3 chimeras, with the difference that ELISA plates were coated with 100 ng 6x-His tag antibody in PBS per well for capturing of His-tagged BG505 SOSIP.664 trimers. Furthermore, a different antibody panel was titrated, including VRC01, PGT145, PGT151, F105 and 2G12. For direct-ELISAs, 50 ng Env in PBS per well were coated to ELISA plates over night at 4 °C and after washing with three well-volumes PBS-T, plates were blocked with 200 µl blocking buffer per well for 2 h at RT. Next, plates were washed with three well-volumes PBS-T, followed by titration of 50 µl mAbs in binding buffer per well. MAbs were incubated for 1 h at RT and all subsequent steps were carried out as described for ELISAs with Env/V3 chimeras.

| | |
|---|---|
| 6x-His Tag Monoclonal Antibody (HIS.H8) | ThermoFisher Scientific, # MA1-21315 |
| Anti-Env antibodies | See **3.4.6**, **Table 4** |

**PLA and VLP-ELISA**

PLA and VLP-ELISAs were essentially performed as described for direct-ELISAs. For PLA ELISAs, plates were coated with 50 ng PLA-adsorbed BG505 SOSIP.664 trimers in PBS per well over night at 4 °C and subsequent steps were carried out as described for direct-ELISAs. In case of ELISAs with VLP-presented gp145 Env/V3 chimeras, plates were coated with VLPs in PBS, containing a total amount of 25 ng Env per well (Env concentrations determined by slot blot titration, see **3.4.3**). All subsequent steps were performed as described

for direct-ELISAs with the differences that all washing steps were performed with three well-volumes PBS instead of PBS-T and the blocking buffer did not contain Tween-20.

| | |
|---|---|
| PLA-adsorbed BG505 SOSIP.664 trimers | See **3.4.8** |
| VLP-presented gp145 Env/V3 chimeras | See **3.4.3** |

**Serum ELISA**

Serum ELISAs of Env proteins and VLPs were performed as described for the respective ELISAs in buffer, with the difference that biotinylated mAbs were incubated in presence of 50 % human serum (100 % serum mixed v/v with PBS). Further, HRP-conjugated streptavidin (1:10,000 in binding buffer) was used for detection.

| | |
|---|---|
| Antibody biotinylation | See **3.4.7** |
| Human serum | Sigma-Aldrich, # H4522 |
| Streptavidin-POD conjugate | Roche, Basel, Switzerland, # 11 089 153 001 |

## 3.5 MicroScale Thermophoresis

MST experiments were performed at 25 °C in premium coated glass capillaries on a Monolith NT.115Pico device. The measurement conditions for the analysis of the Env/V3 chimeras (gp120 monomers, gp140 trimers and VLP-presented gp145) in buffer or 50 % human serum are shown in **Table 5** and **Table 6**, respectively. BG505 SOSIP.664 trimers (soluble or adsorbed to PLA particles) were analyzed in buffer or 50 % human serum under conditions described in **Table 7**. For measurement, the fluorescent-labeled interaction partner was kept at constant concentration in reaction buffer and the unlabeled interaction partner was used as titrant in a 1:2 dilution series. Measurement results were analyzed using the NT Analysis software (analysis from best signal to noise ratio). To calculate the fraction bound ratio, the baseline corrected normalized fluorescence (ΔFNorm) value of each point was divided by the amplitude of the fitted curve. The resulting values, ranging from 0 to 1 (0=unbound, 1=bound), were processed with KaleidaGraph and fitted using the $K_D$ fit formula, derived from the law of mass action.

| | |
|---|---|
| Monolith NT.115Pico | NanoTemper Technologies, Munich, Germany |
| Monolith™ NT.115 MST Premium Coated Capillaries | NanoTemper, # MO-K005 |
| NT Analysis software, version 1.5.41 | NanoTemper Technologies |
| KaleidaGraph, version 4.5 | Synergy Software, Reading, Pennsylvania, USA |
| Human serum | Sigma-Aldrich, # H4522 |
| HBS-EP | 0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005 % (v/v) Surfactant P20 |
| Anti-Env antibodies and protein labeling | See **3.4.6**, **Table 4** and **3.4.7** |

**Table 5. MST measurement conditions for Env/V3 chimeras in buffer.**

| | **gp120 monomers** | **gp140 trimers** | **VLP-presented gp145** |
|---|---|---|---|
| **Fluorescent-labeled interaction partner** | Alexa647-labeled 447-52D at 0.25 - 0.5 nM | Alexa647-labeled 447-52D at 0.05 nM | DY-647P1-labeled VLPs with Env concentrations < 1 ng/µl |
| **Titrant** | MN: 50 nM - 2 pM<br>HXB2: 1556 nM - 50 pM<br>CDC42: 826 nM - 25 pM<br>RF: 920 nM - 28 pM<br>SF33: 1101 nM - 34 pM | MN: 25 nM - 1 pM<br>HXB2: 100 nM - 3 pM<br>CDC42: 100 nM - 3 pM<br>RF: 100 nM - 3 pM<br>SF33: 2569 nM - 78 pM | 447-52D: 1803 nM - 55 pM |
| **Reaction buffer** | HBS-EP | | |
| **MST power** | 80 % | | |
| **LED power** | 30 - 85 % | 100 % | 2 – 20 % |

**Table 6. MST measurement conditions for Env/V3 chimeras in 50 % human serum.**

| | **gp120 monomers** | **gp140 trimers** | **VLP-presented gp145** |
|---|---|---|---|
| **Fluorescent-labeled interaction partner** | Alexa647-labeled 447-52D at 0.5 nM | | DY-647P1-labeled VLPs with Env concentrations < 1 ng/µl |
| **Titrant** | MN: 800 nM - 24 pM<br>HXB2: 1556 nM - 50 pM<br>CDC42: 800 nM - 24 pM<br>RF: 800 nM - 24 pM<br>SF33: 800 nM - 24 pM | MN: 25 nM - 1 pM<br>HXB2: 100 nM - 3 pM<br>CDC42: 100 nM - 3 pM<br>RF: 100 nM - 3 pM<br>SF33: 100 nM - 3 pM | 447-52D: 900 nM - 27 pM |
| **Reaction buffer** | 50 % HBS-EP, 50 % human serum (v/v) | | |
| **MST power** | 80 % | | |
| **LED power** | 5 - 7 % | 10 - 15 % | 10 – 20 % |

**Table 7. MST measurement conditions for soluble and PLA-adsorbed BG505 SOSIP.664 trimers in buffer and 50 % human serum.**

| | **BG505 SOSIP.664 trimers** | | **PLA-adsorbed BG505 SOSIP.664 trimers** | |
|---|---|---|---|---|
| | **Buffer** | **50 % human serum** | **Buffer** | **50 % human serum** |
| **Fluorescent-labeled interaction partner** | Alexa647-labeled mAbs VRC01, PGT145, PGT151, F105 and 2G12 at 0.25 nM | | | |
| **Titrant** | BG505 SOSIP.664: 150 nM - 5 pM | | PLA-adsorbed BG505 SOSIP.664 trimers: 96 nM - 3 pM | |
| **Reaction buffer** | HBS-EP | 50 % HBS-EP, 50 % human serum (v/v) | 50 % HBS-EP, 50 % PBS (v/v) | 50 % PBS, 50 % human serum (v/v) |
| **MST power** | 80 % | | | |
| **LED power** | 15 - 30 % | 12 - 20 % | 30 % | 15 % |

# 4. Results

## 4.1 MicroScale Thermophoresis and ELISA analysis of Env proteins

One objective of this thesis was to explore the opportunities of MST to analyze Env-antibody interactions of soluble and particle-presented Env proteins in buffer and presence of serum. Using two different model systems, a comprehensive and detailed analysis of Env-antibody interactions was performed. The first model system included a five-member Env/V3 chimeric model library with distinct binding capacities to monoclonal antibody 447-52D (see **4.1.2**). This model system was used to step-by-step establish MST measurements of different oligomeric and particle-presented forms of Env, including gp120 monomers, gp140 trimers and VLP-presented gp145 variants. Thereby, Env proteins were analyzed in both, buffer and human serum. In order to test if MST measurements affect the structure of stabilized trimers and, as well to include a relevant Env immunogen in MST analyses, BG505 SOSIP.664 trimers and set of quaternary structure-characterizing antibodies were used as a second model system (see **4.1.12**). Further, based on promising results, MST was utilized to characterize PLA particle-presented BG505 SOSIP trimers. Like for experiments with Env/V3 chimeras, all experiments were performed in buffer and human serum. In addition, ELISA experiments were performed in parallel to benchmark results obtained from MST measurements.

### 4.1.1 MicroScale Thermophoresis

The optical based in solution method MST utilizes the physical phenomenon of thermophoresis, describing the directed movement of molecules in a temperature gradient, which is in this case established by an infrared laser (see **Figure 16A**). Thermophoresis depends on molecule size, charge, and the hydration shell. Since at least one of these parameters is affected by the binding of an interaction partner, this technique can be used to distinguish unbound and bound states and hence determine the equilibrium binding affinities of molecular interaction partners at microliter scale[189,190]. As a detection prerequisite, molecules either have to be fluorescently labeled or intrinsically fluorescent, such as proteins with one or several tryptophan residues. The fluorescence within the observation window is monitored over time (see **Figure 16B**). The initial fluorescence is typically monitored over five seconds before laser activation. Switching on the laser initializes a steep drop of the

fluorescence signal, reflecting temperature dependent changes in the quantum yield of the fluorophore. After this phase, called Temperature Jump or T-Jump, thermophoresis forces molecules to leave the observation window, leading to a decrease of overall fluorescence until thermophoresis is counterbalanced by mass diffusion. Thermophoresis is terminated upon deactivation of the laser. Ultimately, back-diffusion of molecules and reverse T-Jump result in increase of fluorescence in the observation window. In a typical MST experiment, a serial dilution of the ligand is prepared and mixed with a constant concentration of labeled target molecule to establish different ratios of the binding partners. The samples are loaded into capillaries and analyzed in the instrument by subsequent scanning of each capillary. The resulting movement profiles (MST time traces) of the different ratios of interaction partners are plotted in one graph (see **Figure 16C**). The normalized fluorescence (FNorm) is calculated from the ratio between the fluorescence after heating ($F_{hot}$) and before heating ($F_{cold}$). Plotting the baseline corrected FNorm (ΔFNorm) against ligand concentrations and subsequent curve fitting results in binding graphs and corresponding affinity values (see **Figure 16D**).

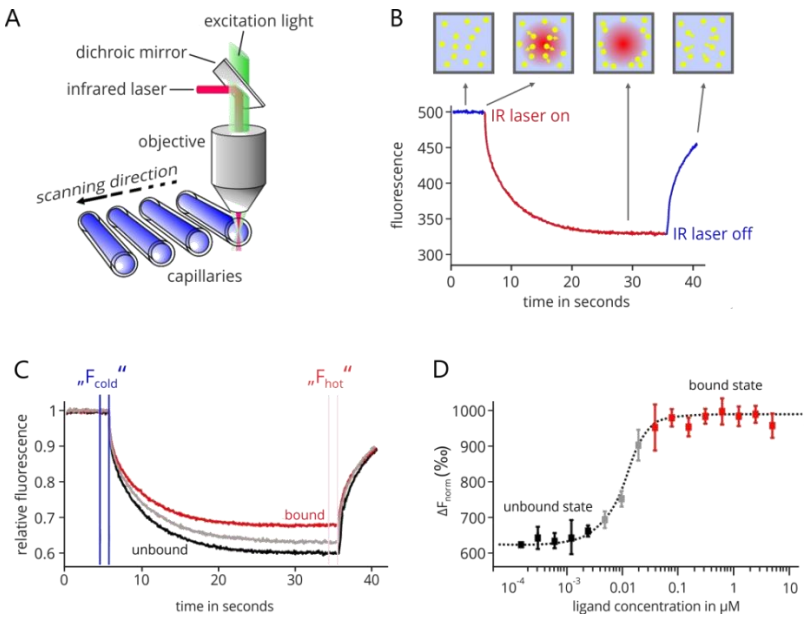



**Figure 16. (A) Technical setup of MicroScale Thermophoresis (MST).** Thin glass capillaries are filled with sample molecules. Fluorescence signals originating from the optically visible interaction partner in the sample are monitored via optics. A thermal gradient is applied by a laser source. Changes in fluorescence can be used to monitor thermophoretic movement of the molecules in solution. **(B)** MST time trace – movement profile of molecules in a temperature gradient. Fluorescence is monitored over time. After an initial 5 sec cold phase (laser off), the laser is switched on and establishes the temperature gradient. After the T-Jump phase, in which the fluorescent dye decreases its signal yield upon heat induction, the thermophoretic movement takes place. After 30 sec the laser is turned off and the molecules diffuse back. **(C)** Results of a typical MST experiment: The MST time traces of 16 capillaries containing the same concentration of optical visible interaction partner and an increasing concentration of the unlabeled ligand are recorded and plotted on one graph. **(D)** The normalized fluorescence (FNorm) of the MST traces is calculated from the ratio between the fluorescence after heating ($F_{hot}$) and before heating ($F_{cold}$). Finally, the baseline corrected FNorm (ΔFNorm) is plotted against the concentration of the ligand and data points are fitted to obtain binding parameters such as the binding affinity. The figure was kindly provided by 2bind.

### 4.1.2 The Env/V3 model system

To investigate Env-antibody interactions by MST and ELISA a previously described model system consisting of five Env/V3 chimeras and the mAb 447-52D[187,188] (Core epitope: amino acid motif GPGR in the V3 crown) was used. The five Env/V3 chimeras are based on the backbone of the subtype C isolate 96ZM651 (GenBank # AF286224.1), with the autologous V3 loop substituted by the V3 loops of the isolates MN, HXB2, CDC42, RF and SF33 (see **Figure 17**). It has been shown recently that 447-52D binds with distinct binding capacities to membrane-bound gp145 versions of the Env/V3 chimeras in cytometer-based titration assays. Furthermore, ELISA titrations have shown that the affinity ranking, with variant MN depicting the best binding, closely followed by HXB2 and CDC42, while variant RF exhibits only low binding and SF33 no binding, is transferable from membrane-bound to soluble trimeric gp140 versions of the Env/V3 chimeras[154]. Thus, this model system represents a valuable tool to investigate Env-antibody interactions by MST and ELISA.

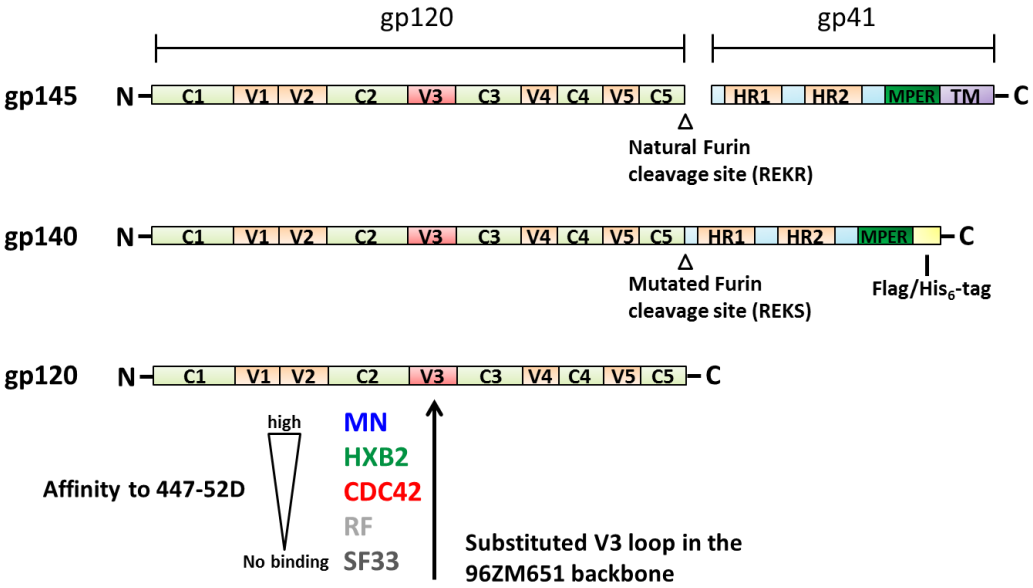



**Figure 17. Schematic presentation of soluble monomeric gp120, trimeric gp140 and membrane-bound gp145 Env/V3 chimeras.** Constant regions (C1-C5), variable loops (V1-V5) as well as heptad repeats 1 and 2 (HR1, HR2), membrane-proximal external region (MPER) and transmembrane domain (TM) are indicated. The intrinsic V3 loop of the HIV-1 isolate 96ZM651 was substituted by the V3 loops of the isolates MN, HXB2, CDC42, RF and SF33, resulting in five chimeric Env variants with distinct binding affinities to the antibody 447-52D. To create soluble gp140 versions of the proteins the gp41 TM was deleted and a C-terminal Flag/$His_6$-tag was added for affinity purification. Furthermore, the natural cleavage site was changed from REKR to REKS to prevent cleavage. To create monomeric gp120 versions of the protein the gp41 domain was completely deleted.

### 4.1.3 Production of soluble and VLP-presented Env/V3 chimeras for MST and ELISA analysis

For the step-by-step analysis of Env-antibody interactions by MST and ELISA, Env/V3 chimeras were produced in different oligomerization states (gp120 monomers, gp140 trimers) and as HIV-1 Gag VLP-presented gp145 Env/V3 chimeras. Soluble gp120 and gp140 Env/V3 (inactivated cleavage site, REKS) chimeras were produced by transient transfection of 293F suspension cells with the respective Env expressing plasmids and purified to monomer/trimer homogeneity using affinity and size exclusion chromatography, as described in **3.4.2**. SEC profiles with indicated monomer and trimer fractions are shown in **Appendix Figure 1**. Pooled monomer and trimer fractions were quality controlled by BN-PAGE (see **3.4.4**). As shown in **Figure 18**, Env/V3 chimeras of gp120 and gp140 trimers have been successfully purified. However, both, gp120 and gp140 preparations contained some dimer/aggregate contaminations that could not be separated. As depicted in **Appendix Figure 1**, the protein peaks flowed smoothly into each other, which did not allow a clear-cut fractionation of monomers and trimers without substantial loss of protein. Of note, contrary to gp140 Env/V3 chimeras, it was not possible to purify tagged gp120 Env/V3 chimeras since addition of the Flag/$His_6$-tag in this case resulted in aggregation of the proteins.

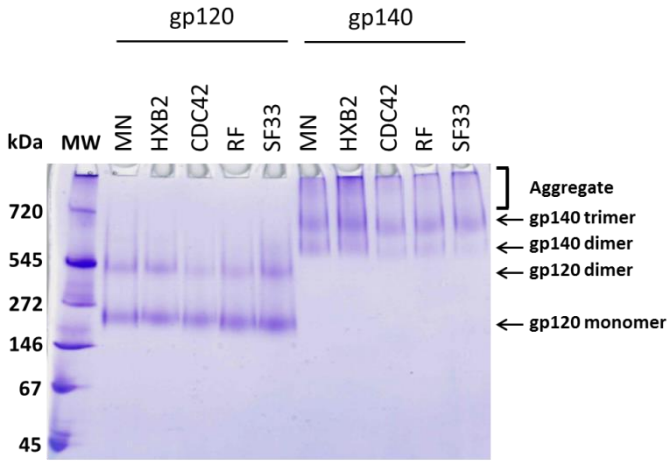



**Figure 18. BN-PAGE analysis of affinity/SEC purified gp120 and gp140 Env/V3 chimeras.** Molecular weight (MW) markers are indicated.

VLPs were produced by cotransfection of 293F cells with Gag and cleavage competent (REKR) gp145 Env/V3 chimeras expressing plasmids and purified from cell culture supernatants using sucrose cushion and sucrose density gradient ultracentrifugation (**3.4.3**). Gag and Env concentrations in the VLP preparations were semi-quantitatively assessed by densitometric quantification of SDS-gel and slot blot analyses (**3.4.3**). For quality control of final VLP preparations a Western blot was performed. In the Western blot an α-gp41 antibody

cocktail and an α-p24/p55 antibody were used for Env and Gag detection, respectively (**3.4.5**). As shown in **Figure 19**, all Env/V3 chimeras-carrying VLPs showed a signal for Gag and Env, indicating the successful incorporation of Env into membrane of VLPs. Blank VLPs (Gag w/o, no cotransfection of Env) that were used as control exhibited a Gag, but no Env signal. The supernatant of cells that were transfected with an Env but without a Gag expressing plasmid (Env w/o) was purified according the VLP purification protocol and used as additional control. As expected, this control resulted in neither, an Env nor a Gag signal. Further, a strong signal at 35 kDa compared to the signal at 130 kDa indicates that the bulk of Env protein in VLP-presented Env/V3 chimeras was cleaved into gp120 (not visible since anti-gp41 antibodies were used for detection) and gp41 subunits.

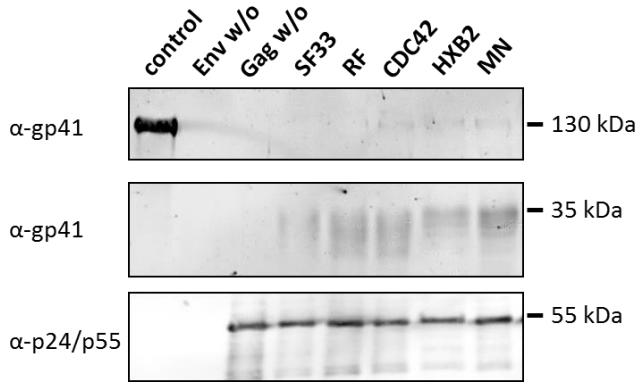



**Figure 19. Analysis of Env/V3 pseudotyped VLPs.** VLP samples were loaded on a reducing SDS gel (loading amount: ~ 200 ng Gag, according to densitometrical analysis) and analyzed by immunoblotting for presence of Gag (α-p24/p55 antibody CB-13/5) and Env (α-gp41 mix). Co-banding of Gag and Env in Env/V3 VLPs **(lane 4-8)** indicates incorporation of Env into the membrane of VLPs. A strong gp41 signal at 35 kDa, compared to the signal at 130 kDa, indicates that the bulk of protein is cleaved. A soluble gp140 standard protein (96ZM651) **(lane 1)**, Env without cotranfected Gag (Env w/o) **(lane 2)** and blank Gag VLPs without cotransfected Env (Gag w/o) **(lane 3)** were used as positive and negative controls, respectively. Data were generated during the doctoral thesis of Tobias Fischer under my experimental supervision.

In summary, Env/V3 chimeras of gp120 monomers, gp140 trimers and VLP-presented gp145 Env have been successfully purified and were subsequently used for MST and ELISA analyses.

### 4.1.4 Interactions between mAb 447-52D and soluble, monomeric gp120 Env/V3 chimeras in buffer

In the first step, interactions between mAb 447-52D and monomeric gp120 Env/V3 chimeras were analyzed in buffer, since these interactions represent the most simple in the Env/V3 model system. For MST analyses, the Alexa647-labeled 447-52D was kept at constant nanomolar concentrations and monomeric Env/V3 chimeras were used as titrants, see **3.5** with measurement conditions in **Table 5**. In order to avoid labeling of Env proteins, this experimental setup was chosen for MST analyses of soluble Env proteins in general. Using common protocols, labeling takes place statistically, mainly at lysine residues of a protein. This in turn can affect antibody epitopes either by blocking lysines required for antibody binding or maybe by sterical hindrance due to lysines carrying a fluorescent dye close to antibody epitopes. In addition, general experiences in our group with fluorescent-labeled antibodies (flow cytometry, FACS) showed that antibodies commonly well-tolerate labeling with fluorescent dyes. MST data from titrations were processed as following: In order to calculate a fraction that relates the amount of bound antibodies to the amount of unbound antibodies, the ΔFNorm value of each point was divided by the amplitude of the fitted curve, resulting in values from 0 to 1 (0=unbound, 1=bound), processed using the KaleidaGraph 4.5 software and fitted using the $K_D$ fit formula derived from the law of mass action (**3.5**). The resulting binding curves (see **Figure 20A**) with corresponding equilibrium dissociation constants ($K_D$)s, see **Table 8**, showed that 447-52D bound with distinct binding affinities to the gp120 Env/V3 chimeras, with MN being the best binding variant, followed by CDC42 and HXB2 which bound with similar affinity. Both, RF and SF33 exhibited marginal/no binding. Next, gp120 Env/V3-447-52D interactions were analyzed by ELISA. In general, following typical ELISA setups for the analyses Env-antibody interactions described in literature[61,127,125], soluble Env proteins were immobilized by capturing of Env proteins, whereas antibodies were used as titrants. For the analysis of gp120 Env/V3-447-52D interactions, gp120 Env/V3 chimeras were captured via GNL while 447-52D was titrated (**3.4.9**). GNL was used for capturing since gp120 Env/V3 chimeras were expressed without a His-tag, representing one of the commonly used tags for protein capturing (e.g. via coated anti-His antibodies). However, as shown in **Figure 20B** with corresponding $K_D$-values in **Table 8**, ELISA results validated affinity data measured by MST since the results were highly comparable. In both, MST and ELISA, variant MN exhibited the best binding, closely

followed by CDC42 and HXB2 which bound with similar affinity to 447-52D, whereas RF and SF33 bound only marginally or not at all.

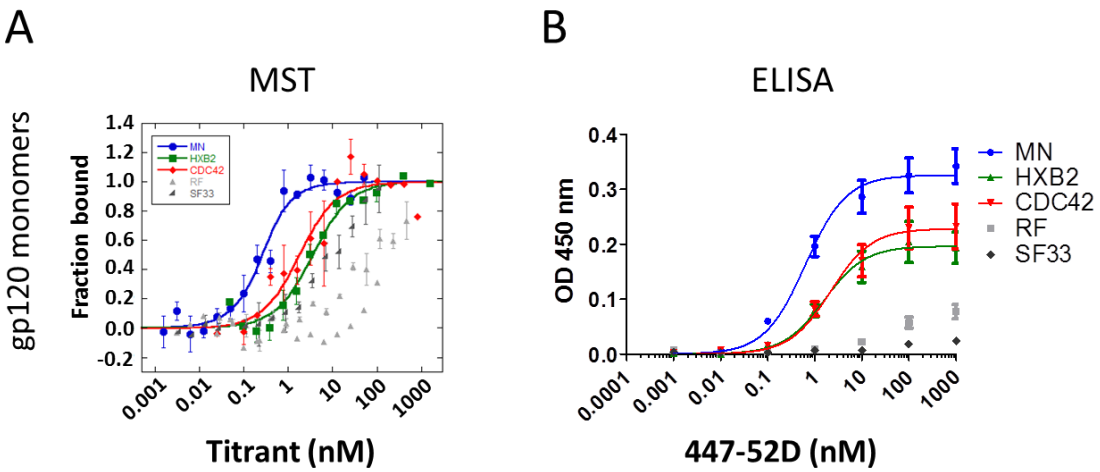



**Figure 20. Interaction of mAb 447-52D with monomeric gp120 Env/V3 chimeras in buffer, quantified by MST and ELISA.** For MST analysis **(A)**, gp120 Env/V3 chimeras diluted in buffer were titrated to the Alexa647-labeled 447-52D which was kept at constant concentration. Results of the MST measurements are presented as normalized binding curves of three independent experiments, error bars reflect the standard error of the mean (SEM) of the independent experiments. **(B)** Validation of MST by ELISA: Binding of 447-52D to monomeric gp120 Env/V3 chimeras was assessed by a lectin (*Galanthus nivalis*) capture-ELISA (negative control: Wells coated with lectin, incubated with PBS). 447-52D, diluted in binding buffer, was titrated to immobilized Env proteins at indicated concentrations and a HRP-conjugated polyclonal rabbit anti-human Ab was used for readout. ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

### 4.1.5 Interactions between mAb 447-52D and soluble, trimeric gp140 Env/V3 chimeras in buffer

The next step included the analysis of interactions between 447-52D and trimeric gp140 Env/V3 chimeras. MST measurements were performed in buffer as described above and under measurement conditions summarized in **3.5**, **Table 5**. The resulting binding curves (**Figure 21A**) showed that the affinity ranking, as determined for gp120 Env/V3 chimeras before, was mostly preserved also in measurements with gp140 Env/V3 chimeras. However, the binding affinity of variants MN, HXB2 and CDC42 to 447-52D can be considered as similar since binding curves were in close range to each other. Further, whereas the monomeric gp120 variant RF showed no saturated binding to 447-52D, this was the case for the trimeric version of the protein. $K_D$-values, calculated from the MST binding curves of MN, HXB2, CDC42, and also RF indicate that 447-52D binds trimeric gp140 Env/V3 chimeras with higher affinities, compared to monomeric gp120 versions of the proteins (see **Table 8**). However, the $K_D$-values determined for gp140 Env/V3 chimeras MN, HXB2 and CDC42 can be only seen as an orientation as these variants bound with affinities that seemed

to be far beneath the concentration of the fluorescent-labeled 447-52D (0.05 nM), and thus at the detection limit (using lower amount of labeled 447-52D was not possible without falling below the required fluorescent signal). GNL capture-ELISAs (**3.4.9**) with gp140 Env/V3 chimeras produced comparable results. Within ELISA of gp140 Env/V3 chimeras, variants MN, HXB2 and CDC42 bound with similar affinities to 447-52D, while variant RF and SF33 showed reduced and no binding, respectively, see **Figure 21B**. In general, as also observed in MST analysis, 447-52D seems to bind with higher affinities to trimeric gp140 Env/V3 chimeras compared to gp120 versions of the proteins. However, in ELISA this was limited to Env/V3 chimeras HXB2, CDC42 and RF, while gp140 MN exhibited a binding affinity comparable to gp120 MN (see **Table 8**). Importantly, the absolute $K_D$-values calculated from the interactions between gp140 Env/V3 chimeras and 447-52D (throughout all following analyses) differed between MST and ELISA. Thus, all calculated affinity values are only comparable within the respective method used.

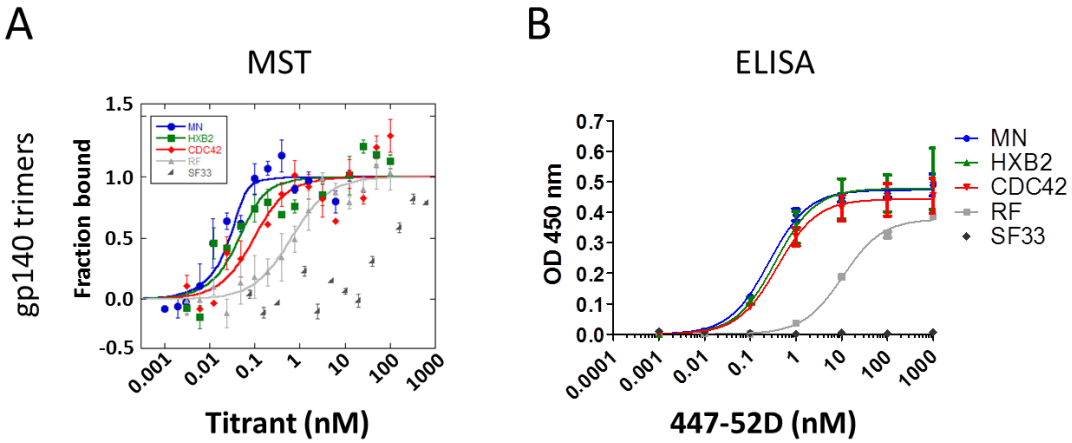



**Figure 21. Interaction of mAb 447-52D with trimeric gp140 Env/V3 chimeras in buffer, quantified by MST and ELISA.** For MST analysis **(A)**, gp140 Env/V3 chimeras diluted in buffer were titrated to the Alexa647-labeled 447-52D which was kept at constant concentration. Results of the MST measurements are presented as normalized binding curves of three independent experiments, error bars reflect the SEM of the independent experiments. **(B)** Validation of MST by ELISA: Binding of 447-52D to trimeric gp140 Env/V3 chimeras was assessed by a lectin (*Galanthus nivalis*) capture-ELISA (negative control: Wells coated with lectin, incubated with PBS). 447-52D, diluted in binding buffer, was titrated to immobilized Env proteins at indicated concentrations and a HRP-conjugated polyclonal rabbit anti-human Ab was used for readout. ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

### 4.1.6 Interactions between mAb 447-52D and VLP-presented gp145 Env/V3 chimeras in buffer

As described in **1.5.3**, there exist several reasons to use particle-presented Env immunogens for vaccination. However, especially in case of particle-presented native-like trimers it has to be ensured that they retain their favorable antigenicity profile, which has therefore to be analyzed. Nevertheless, analysis of the interaction between particle-presented Env and antibodies is rather complex. For example, depending on the particle type, surface immobilization of particles may result in the disruption of particles and/or disturbance of the structure of presented Env proteins. Since measurements in solution avoid this kind of problem from the outset, it was reasonable to test if MST is capable to measure affinity interactions between antibodies and particle-presented Env proteins. Since VLPs are relatively easy to produce, HIV-1 Gag VLPs decorated with gp145 Env/V3 chimeras were chosen as a model system for particle-presented Env proteins. MST analysis of the interaction between 447-52D and the VLP-presented Env/V3 chimeras in buffer was performed using VLP-presented Env/V3 chimeras as fluorescently-labeled interaction partners while 447-52D was used as titrant (for measurement conditions see **3.5**, **Table 5**). This experimental setup was chosen because VLPs contain -next to gp145 Env proteins- all cellular components, including membrane proteins present on the cell surface, which allow an efficient labeling with fluorescent dyes via NHS-chemistry. In turn, using the antibody as titrant allows precise characterization of $K_D$-values, since antibody concentrations can be easily determined. The resulting binding curves (**Figure 22A**) with corresponding $K_D$-values (**Table 8**) showed that the VLP variants MN, CDC42 and HXB2 bound with similar affinities to 447-52D, whereas marginal/no binding was detected for variants RF, SF33 and blank VLPs (**Appendix Figure 2A**). This in general reflected the results of measurements with gp140 Env/V3 chimeras, with the difference that variant RF displayed only indicated binding to 447-52D in the VLP-context, whereas saturated binding to 447-52D was observed for the gp140 RF trimer. To validate MST results, VLP-ELISA experiments were performed (**3.4.9**, **VLP-ELISA**), in which VLP-presented Env/V3 chimeras were adsorbed to the surface of ELISA plates and 447-52D was used as titrant. ELISA binding curves (**Figure 22B**) and corresponding the $K_D$-values (**Table 8**) showed that VLP variants MN, HXB2 and CDC42 bound with comparable affinities to 447-52D, whereas a lower affinity was measured for VLP variant RF and no binding to SF33 and blank VLPs (Gag w/o). ELISA data basically reflected results obtained from MST measurements, with the difference that saturated binding of 447-52D to VLP

variant RF was only detected in ELISA. Of note, as described above, $K_D$-values were only comparable within one method. Interestingly, regarding the absolute binding signal, VLP-ELISA binding curves showed a different picture compared to ELISAs with gp120 monomers and gp140 trimers, see **Figure 23**. In ELISAs with gp120 Env/V3 chimeras, the best binding variant MN also showed the highest overall signal. Similar was true for MN in ELISAs with gp140 Env/V3 chimeras, also not as distinctly as in gp120 ELISAs. However, despite of high affinity binding (according to the calculated $K_D$) in VLP-ELISA, MN exhibited the lowest absolute binding signal. This can most likely be explained by the fact that Env concentrations in VLP preparations were determined by semi-quantitative slot blot titration. Thus, the amount of VLP-presented Env used for coating ELISA plates slightly varied within the different VLP preparations, resulting in varying absolute binding signals. In summary, results demonstrated that MST is capable to measure the affinities of the interactions between VLP-presented Env/V3 chimeras and 447-52D as confirmed by ELISA. In conclusion, MST represents a valuable tool to investigate interactions of antibodies and particle-presented Env molecules.

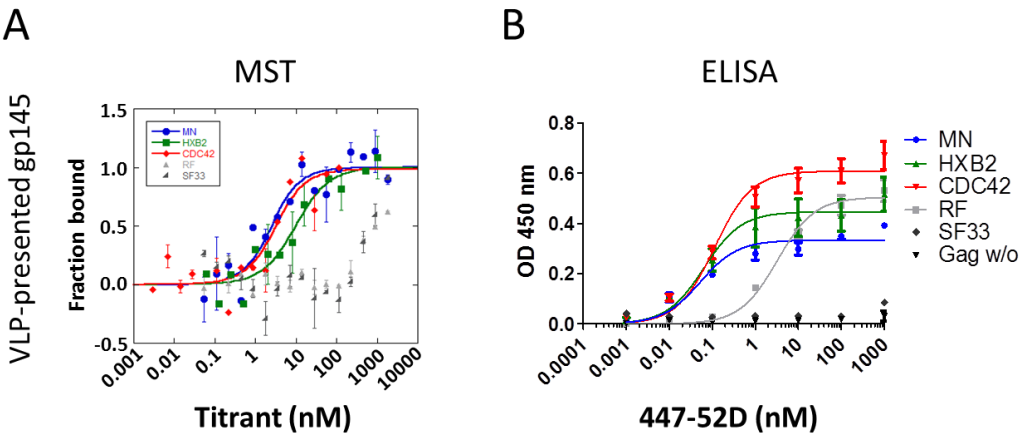



**Figure 22. Interaction of mAb 447-52D with VLP-presented Env/V3 chimeras in buffer, quantified by MST and ELISA. (A)** MST analysis of VLP-presented Env/V3 chimeras in buffer. 447-52D, diluted in buffer, was titrated to DY-647P1-labeled VLPs which were kept at constant concentration. Results of the MST measurements are presented as normalized binding curves of two independent experiments, error bars reflect the SEM of the independent experiments. **(B)** Validation of MST by ELISA: Binding of 447-52D to VLP-presented Env/V3 chimeras was assessed by direct-ELISA. VLPs were directly coated to ELISA plates (negative controls: Wells coated with blank VLPs (Gag w/o) and PBS, respectively). 447-52D, diluted in binding buffer, was titrated to coated VLPs at indicated concentrations and a HRP-conjugated polyclonal rabbit anti-human Ab was used for readout. ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments. VLP-ELISA data were generated during the doctoral thesis of Tobias Fischer under my experimental supervision.

### 4.1.7 Summary: Interactions between Env/V3 chimeras and mAb 447-52D in buffer

In summary, experiments with the Env/V3 model system have shown that MST provided consistent data within the overall expected affinity ranking (MN≥CDC42/HXB2>RF>SF33) of the interactions between gp120, gp140 and VLP-presented gp145 Env/V3 chimeras and 447-52D. Further, MST-derived data haven been successfully validated by ELISA experiments, see **Figure 23**. However, the calculated $K_D$-values were only comparable within the respective measurement method used, see **Table 8**. In general, 447-52D bound with higher overall affinities to soluble trimeric gp140 Env/V3 chimeras compared to monomeric gp120 versions of the protein, in both, MST and ELISA analyses. VLP analyses were excluded from this comparison since different experimental setups were used for their analysis in both, MST and ELISA. These included labeling of VLPs instead of the antibody in MST analyses (potentially affecting the 447-52D epitope) and direct coating of VLPs (instead of lectin capture) in the VLP ELISA setup.

**Table 8. Apparent equilibrium dissociation constants of the 447-52D-Env/V3 chimera interactions in buffer.** $K_D$-values in nM represent the mean value, ± the SEM of n experiments. Detectable, but not saturated signals were excluded from $K_D$-fitting and are therefore designated as "binding indicated". Background-level binding is designated as "no binding" (n.b.). ELISA $K_D$-values were calculated by using non-linear least squares regression (one site binding, hyperbola; GraphPad Prism 5). MST $K_D$-values are computed by the $K_D$-fit model of KaleidaGraph. $R^2$-values of the respective fits are included in brackets.

| | **gp120 monomers** | | **gp140 trimers** | | **VLP-presented gp145** | | |
|---|---|---|---|---|---|---|---|
| **Env/V3 chimera** | MST (n=3) | ELISA (n=2) | MST (n=3) | ELISA (n=2) | MST (n=2) | ELISA (n=2) | **Affinity ranking** |
| **MN** | 0.14 ± 0.05 ($R^2$=0.9676) | 0.63 ± 0.14 ($R^2$=0.9743) | 0.0025± 0.0044 ($R^2$=0.9021) | 0.26 ± 0.03 ($R^2$=0.9917) | 1.5 ± 0.9 ($R^2$=0.8906) | 0.053 ± 0.018 ($R^2$=0.9291) | **1** |
| **HXB2** | 3.4 ± 0.8 ($R^2$=0.9729) | 1.6 ± 0.6 ($R^2$=0.9192) | 0.020 ± 0.019 ($R^2$=0.8188) | 0.36 ± 0.13 ($R^2$=0.9409) | 8.7 ± 4.0 ($R^2$=0.9197) | 0.064 ± 0.028 ($R^2$=0.8955) | **2/3** |
| **CDC42** | 1.6 ± 0.7 ($R^2$=0.9046) | 2.2 ± 0.7 ($R^2$=0.9404) | 0.070 ± 0.058 ($R^2$=0.8350) | 0.39 ± 0.10 ($R^2$=0.9712) | 2.1 ± 1.5 ($R^2$=0.8579) | 0.11 ± 0.03 ($R^2$=0.9626) | **2/3** |
| **RF** | b.i. | b.i. | 0.62 ± 0.18 ($R^2$=0.9581) | 11 ± 1 ($R^2$=0.9956) | b.i. | 3.2 ± 0.6 ($R^2$=0.9812) | **4** |
| **SF33** | b.i. | b.i. | b.i. | n.b. | b.i. | n.b. | **5** |
| **$K_D$ (nM)** | | | | | | | |

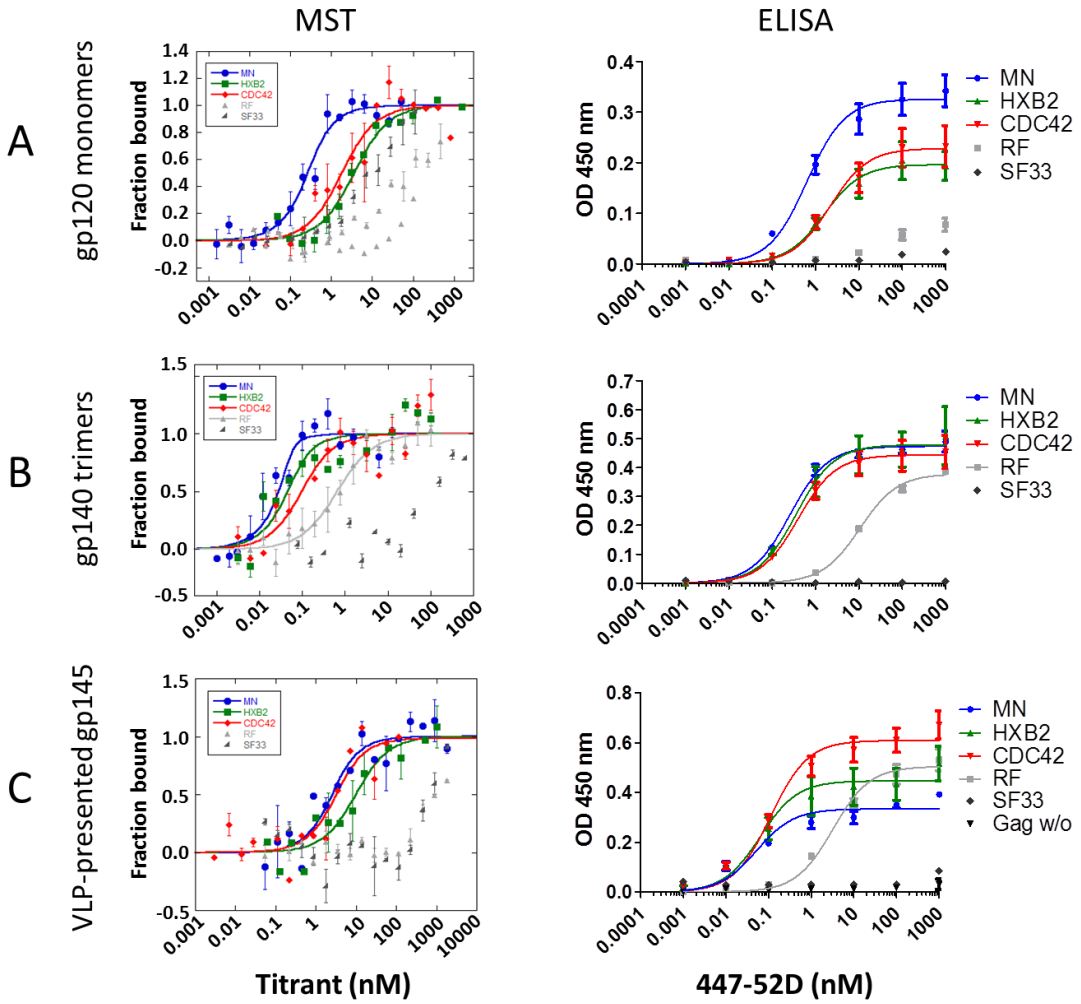



**Figure 23. Interaction of mAb 447-52D with Env/V3 chimeras in buffer, quantified by MST and ELISA.** MST **(left panel)** and ELISA analyses **(right panel)** of the interactions between 447-52D and monomeric gp120, trimeric gp140 and VLP-presented gp145 Env/V3 chimeras are shown in **(A)**, **(B)** and **(C)**, respectively. For MST analyses of gp120 and gp140 Env/V3 chimeras, Env proteins diluted in buffer were titrated to the Alexa647-labeled 447-52D which was kept at constant concentration. MST analysis of VLP-presented gp145 Env/V3 chimeras was performed in an inverse setup (447-52D was titrated to DY-647P1-labeled VLPs). Results of the MST measurements are presented as normalized binding curves of at least two independent experiments, error bars reflect the SEM of the independent experiments. Validation of MST by ELISA: Binding of 447-52D to gp120 and gp140 Env/V3 chimeras was assessed by a lectin (*Galanthus nivalis*) capture-ELISA (negative control: Wells coated with lectin, incubated with PBS). 447-52D, diluted in binding buffer, was titrated to immobilized Env molecules at indicated concentrations and a HRP-conjugated polyclonal rabbit anti-human Ab was used for readout. Env/V3-VLPs were directly coated to ELISA plates and analyzed in the same way (negative controls: Wells coated with blank VLPs (Gag w/o) and PBS, respectively). All ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

### 4.1.8 Interactions between mAb 447-52D and soluble, monomeric gp120 Env/V3 chimeras in 50 % human serum

Ideally, binding analysis of interaction partners should be performed under physiological conditions, since the environment of interaction potentially affects the behavior of interaction partners. Serum proteases, for example, are known to clip the V3 region of gp120 in subtype B Env proteins and thereby damaging protein structure (see **1.6**). In order to test if MST is in principle capable to determine affinities under physiological conditions, analysis of the interactions between 447-52D and Env/V3 chimeras were repeated in presence of human serum. Initial experiments showed that measurements 100 % human serum were not possible due to high background signals (data not shown). However, using 50 % serum resulted in evaluable data. Thus, the interactions between 447-52D and gp120, gp140 and VLP-presented Env/V3 chimeras were investigated under this "near-physiological" condition.

MST analyses of the interactions between gp120 Env/V3 chimeras and 447-52D in presence of 50 % serum (for measurement conditions see **3.5**, **Table 6**) showed that the binding profiles, see **Figure 24A**, were in agreement with the measurements for gp120 monomers in buffer (with MN depicting the best binding, followed by CDC42/HXB2 and only indicated binding of RF and SF33, see **Figure 20A**). However, the overall binding affinity of 447-52D seemed to be reduced in the presence of serum (compare $K_D$-values obtained from serum, **Table 9**, and buffer measurements, **Table 8**). In order to validate MST results, GNL capture-ELISAs were performed, in which biotinylated 447-52D was incubated in the presence of 50 % serum (see **3.4.9**, **Serum ELISA**). ELISA analysis in 50 % serum showed that the affinity ranking determined by MST was also preserved in ELISA analyses (**Figure 24B**), thereby validating MST derived data. Interestingly, and contrary to MST results described above, the presence of serum in ELISA analyses of gp120 Env/V3 chimeras did not result in a reduced affinity of 447-52D (compare $K_D$-values of gp120 Env/V3 chimeras in buffer, **Table 8,** and serum, **Table 9**).

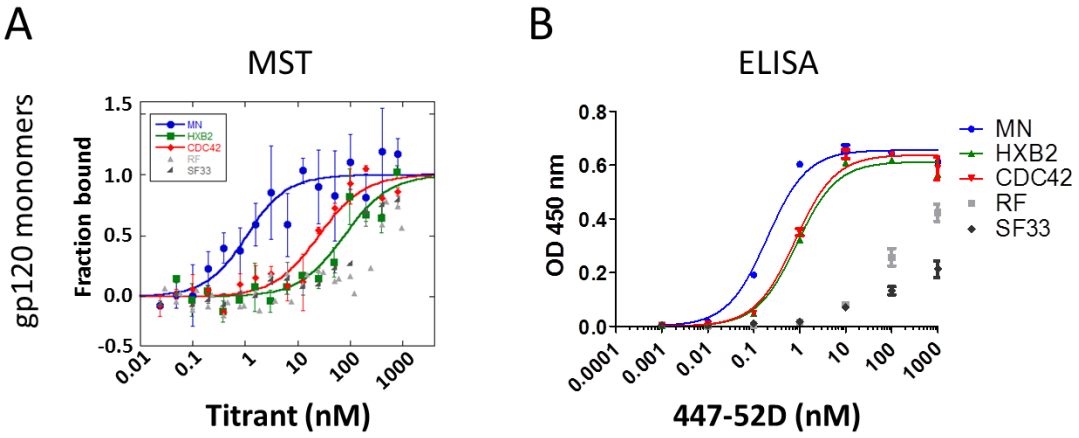



**Figure 24. Interaction of mAb 447-52D with monomeric gp120 Env/V3 chimeras in 50 % human serum, quantified by MST and ELISA.** For MST analysis **(A)**, gp120 Env/V3 chimeras diluted in 50 % human serum were titrated to the Alexa647-labeled 447-52D which was kept at constant concentration. Results of the MST measurements are presented as normalized binding curves of two independent experiments, error bars reflect the SEM of the independent experiments. **(B)** Validation of MST by ELISA: Binding of 447-52D to monomeric gp120 Env/V3 chimeras was assessed by a lectin (*Galanthus nivalis*) capture-ELISA (negative control: Wells coated with lectin, incubated with PBS). Biotinylated 447-52D, diluted in 50 % human serum, was titrated to immobilized Env proteins at indicated concentrations and streptavidin-POD conjugate was used for readout. ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

### 4.1.9 Interactions between mAb 447-52D and soluble, trimeric gp140 Env/V3 chimeras in 50 % human serum

Next, MST was used to analyze interactions between trimeric gp140 Env/V3 chimeras and 447-52D in the presence of 50 % serum (for experimental conditions see **3.5**, **Table 6**). The resulting binding curves demonstrated that variants MN, HXB2 and CDC42 bound with similar affinities to 447-52D, see **Figure 25A**. Thus, the binding profile of these three gp140 Env/V3 chimeras in serum is comparable to the binding profile of the same proteins analyzed in buffer, see **Figure 21A**. However, the overall binding affinity of 447-52D to gp140 Env/V3 chimeras was reduced in presence of serum, as it was already observed in the analyses of gp120 Env/V3 chimeras (compare MST-derived $K_D$-values of gp120/gp140 Env/V3 chimeras analyzed in buffer, **Table 8**, and in presence of serum, **Table 9**). In addition, while 447-52D showed saturated binding to gp140 Env/V3 chimera RF in buffer measurements, this was not the case in presence of serum. Here, as variant SF33, RF showed only indicated/no binding.

GNL capture-ELISAs with gp140 Env/V3 chimeras in serum demonstrated similar results. Within ELISA, variants MN, HXB2 and CDC42 bound with similar affinities to 447-52D whereas variants RF and SF33 showed reduced and no binding, respectively (**Figure 25B**). Thus, ELISA basically reflected and thereby validated MST data. Further, as previously observed for MST and ELISA analyses in buffer, 447-52D seems to bind with higher

affinities to trimeric gp140 Env/V3 chimeras than to the corresponding monomeric gp120 versions of the protein since this was also the case in serum measurements (compare $K_D$-values of gp120 and gp140 Env/V3 chimeras in **Table 9**). Contrary to MST analyses, and in accordance to the results obtained from the analyses of gp120 monomers, the presence of serum did not reduce the affinity of 447-52D to gp140 Env/V3 chimeras in ELISA (compare $K_D$-values of gp120/gp140 Env/V3 chimeras in buffer, **Table 8**, and serum, **Table 9**).

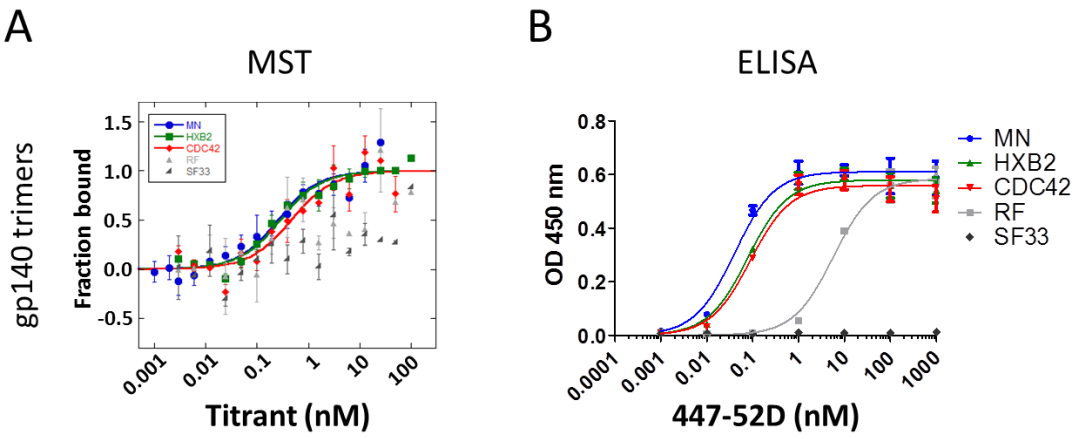



**Figure 25. Interaction of mAb 447-52D with trimeric gp140 Env/V3 chimeras in 50 % human serum, quantified by MST and ELISA.** For MST analysis **(A)**, gp140 Env/V3 chimeras diluted in 50 % human serum were titrated to the Alexa647-labeled 447-52D which was kept at constant concentration. Results of the MST measurements are presented as normalized binding curves of three independent experiments, error bars reflect the SEM of the independent experiments. **(B)** Validation of MST by ELISA: Binding of 447-52D to trimeric gp140 Env/V3 chimeras was assessed by a lectin (*Galanthus nivalis*) capture-ELISA (negative control: Wells coated with lectin, incubated with PBS). Biotinylated 447-52D, diluted in 50 % human serum, was titrated to immobilized Env proteins at indicated concentrations and streptavidin-POD conjugate was used for readout. ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

### 4.1.10 Interactions between mAb 447-52D and VLP-presented gp145 Env/V3 chimeras in 50 % human serum

The last experiments within the Env/V3 model system included MST and ELISA analyses of VLPs in presence of 50 % serum. For MST analyses in serum, 447-52D was titrated to fluorescent-labeled VLPs, according to the experimental setup for the analysis of VLPs in buffer. For measurement conditions of VLPs in serum see **3.5**, **Table 6**. MST analyses of the interactions between 447-52D and VLP-presented gp145 Env/V3 chimeras in serum revealed an overall binding profile (**Figure 26A**) that is comparable to the corresponding measurements of VLPs in buffer (**Figure 22A**).VLP-variants MN, HXB2 and CDC42 demonstrated similar affinities to 447-52D and whereas variants RF and SF33 showed only

indicated/no binding. In addition, there was also no binding of 447-52D to blank VLPs in presence of 50 % serum, see **Appendix Figure 2B**. Notably, $K_D$-values calculated from MST analyses of VLPs in buffer and serum were in the same range and did not drastically differ between buffer and serum measurements as $K_D$-values of gp120 and gp140 Env V3 chimeras did (compare MST $K_D$-values in buffer, **Table 8**, and serum, **Table 9**). VLP-ELISAs in presence of 50 % serum delivered comparable results. Within ELISA, VLP-variants MN, HXB2 and CDC42 bound with similar affinities to 447-52D while variant SF33 and blank VLPs (Gag w/o) demonstrated no binding (**Figure 26B**). However, VLP variant RF exhibited saturated binding to 447-52D in serum ELISA which was not observed in the corresponding MST analysis (**Figure 26A**). Of note, this was also observed in the analysis of VLPs in buffer, see (**Figure 22**). In addition, the results obtained from VLP-ELISAs in serum are in accordance to VLP-ELISAs in buffer, including comparable $K_D$-values (see **Table 8**, and **Table 9**) and the reduced absolute binding signal for variant MN, explained in **4.1.6**.

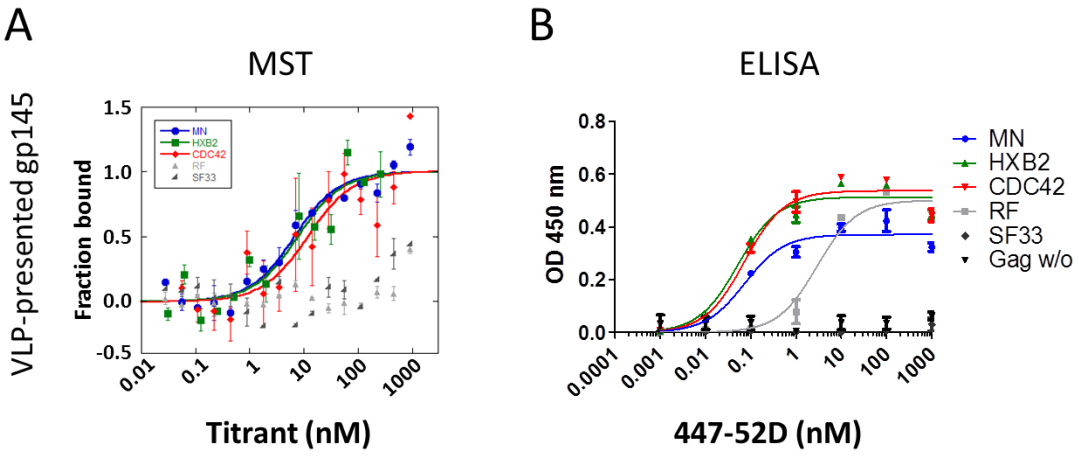



**Figure 26 Interaction of mAb 447-52D with VLP-presented Env/V3 chimeras in 50 % human serum, quantified by MST and ELISA. (A)** MST analysis of VLP-presented Env/V3 chimeras in 50 % human serum. 447-52D, diluted in 50 % human serum, was titrated to DY-647P1-labeled VLPs which were kept at constant concentration. Results of the MST measurements are presented as normalized binding curves of two independent experiments, error bars reflect the SEM of the independent experiments. **(B)** Validation of MST by ELISA: Binding of 447-52D to VLP-presented Env/V3 chimeras was assessed by direct-ELISA. VLPs were directly coated to ELISA plates (negative controls: Wells coated with blank VLPs (Gag w/o) and PBS, respectively). Biotinylated 447-52D, diluted in 50 % human serum, was titrated to immobilized Env proteins at indicated concentrations and streptavidin-POD conjugate was used for readout. ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments. VLP-ELISA data were generated during the doctoral thesis of Tobias Fischer under my experimental supervision.

### 4.1.11 Summary: Interactions between Env/V3 chimeras and mAb 447-52D in 50 % human serum

In summary, it has been demonstrated that MST provides consistent data within the overall expected affinity ranking (MN≥CDC42/HXB2>RF>SF33) of the interactions between gp120, gp140 and VLP-presented gp145 Env/V3 chimeras and 447-52D also in presence of 50 % human serum. Further, MST data have been successfully validated by ELISA, see **Figure 27**. However, the calculated $K_D$-values were only comparable within the respective measurement method used, see **Table 9**. Interestingly, the presence of serum in MST analyses seemed to reduce the overall binding affinity of 447-52D to gp120 and gp140 Env/V3 chimeras. However, this was not the case in MST analyses of VLPs were binding affinities between buffer and serum were comparable. In addition, serum effects in terms of reduced binding affinities were not observed in ELISA analyses. For comparison of calculated $K_D$-values in buffer and serum analyses see **Table 8** and **Table 9**, respectively. Furthermore, as already observed for MST and ELISA analyses in buffer, 447-52D bound with higher overall affinities to soluble trimeric gp140 Env/V3 chimeras compared to the corresponding soluble gp120 versions of the proteins. In conclusion, data underline that MST is capable to determine affinities under near-physiological conditions.

**Table 9. Apparent equilibrium dissociation constants of the 447-52D-Env/V3 chimera interactions in 50 % human serum.** $K_D$-values in nM represent the mean value, ± the SEM of n experiments. Detectable, but not saturated signals were excluded from $K_D$-fitting and are therefore designated as "binding indicated". Background-level binding is designated as "no binding" (n.b.). ELISA $K_D$-values were calculated by using non-linear least squares regression (one site binding, hyperbola; GraphPad Prism 5). MST $K_D$-values are computed by the $K_D$-fit model of KaleidaGraph. $R^2$-values of the respective fits are included in brackets.

| | **gp120 monomers** | | **gp140 trimers** | | **VLP-presented gp145** | | |
|---|---|---|---|---|---|---|---|
| **Env/V3 chimera** | MST (n=2) | ELISA (n=2) | MST (n=3) | ELISA (n=2) | MST (n=2) | ELISA (n=2) | **Affinity ranking** |
| **MN** | 0.80 ± 0.42 ($R^2$=0.8953) | 0.20 ± 0.03 ($R^2$=0.9901) | 0.21 ± 0.09 ($R^2$=0.9228) | 0.040 ± 0.009 ($R^2$=0.9742) | 5.7 ± 1.9 ($R^2$=0.9516) | 0.073 ± 0.021 ($R^2$=0.9532) | **1** |
| **HXB2** | 75 ± 37 ($R^2$=0.8917) | 0.86 ± 0.11 ($R^2$=0.9915) | 0.23 ± 0.06 ($R^2$=0.9667) | 0.079 ± 0.015 ($R^2$=0.9800) | 6.7 ± 4.7 ($R^2$=0.8701) | 0.052 ± 0.013 ($R^2$=0.9681) | **2/3** |
| **CDC42** | 25 ± 10 ($R^2$=0.9211) | 0.80 ± 0.12 ($R^2$=0.9875) | 0.42 ± 0.22 ($R^2$=0.8821) | 0.090 ± 0.020 ($R^2$=0.9729) | 11 ± 8 ($R^2$=0.8012) | 0.072 ± 0.018 ($R^2$=0.9647) | **2/3** |
| **RF** | b.i. | b.i. | b.i. | 5.7 ± 1.1 ($R^2$=0.9822) | b.i. | 2.9 ± 0.8 ($R^2$=0.9630) | **4** |
| **SF33** | b.i. | b.i. | b.i. | n.b. | b.i. | n.b. | **5** |
| **$K_D$ (nM)** | | | | | | | |

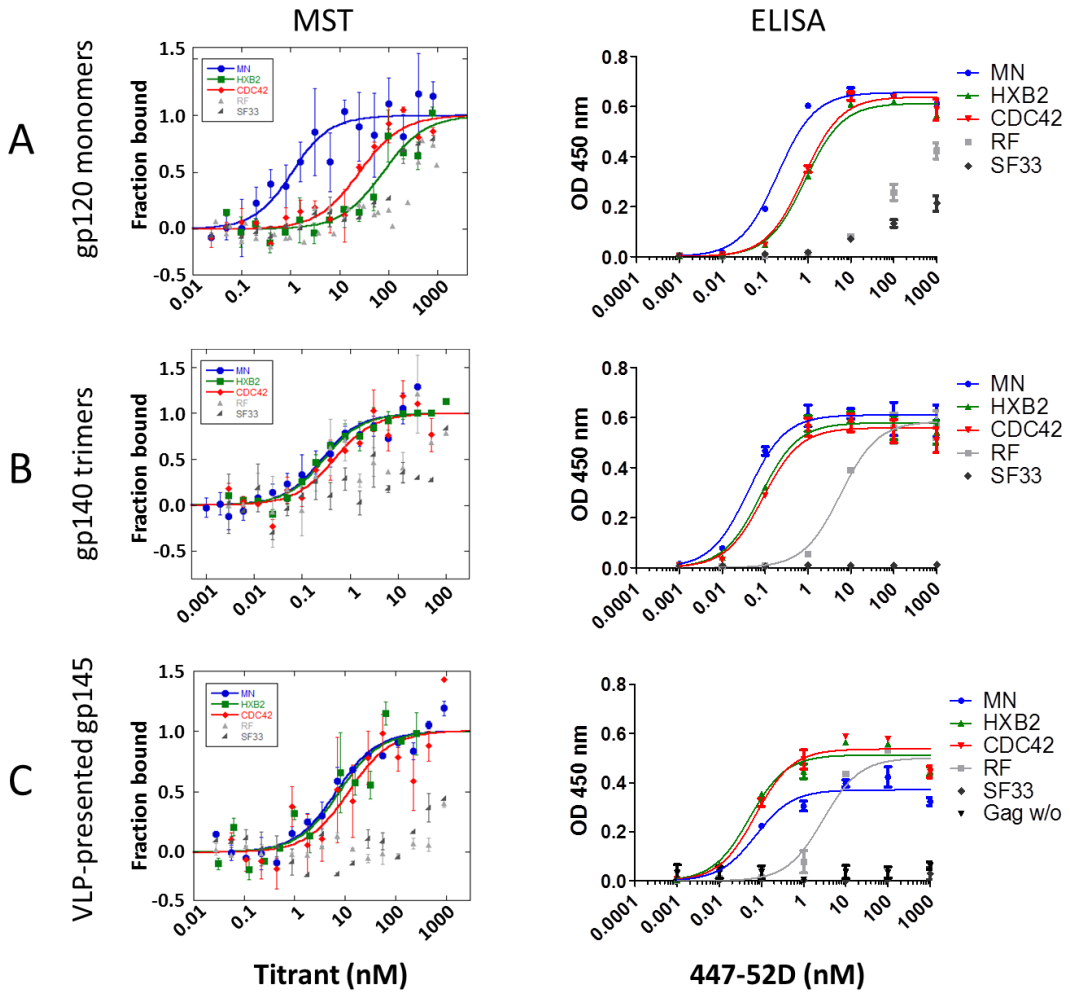



**Figure 27. Interaction of mAb 447-52D with Env/V3 chimeras in 50 % human serum, quantified by MST and ELISA.** MST **(left panel)** and ELISA analyses **(right panel)** of the interactions between 447-52D and monomeric gp120, trimeric gp140 and VLP-presented gp145 Env/V3 chimeras are shown in **(A)**, **(B)** and **(C)**, respectively. For MST analyses of gp120 and gp140 Env/V3 chimeras, Env proteins diluted in 50 % human serum were titrated to the Alexa647-labeled 447-52D which was kept at constant concentration. MST analysis of VLP-presented gp145 Env/V3 chimeras was performed in an inverse setup (447-52D in 50 % human serum was titrated to DY-647P1-labeled VLPs). Results of the MST measurements are presented as normalized binding curves of at least two independent experiments, error bars reflect the SEM of the independent experiments. Validation of MST by ELISA: Binding of 447-52D to gp120 and gp140 Env/V3 chimeras was assessed by a lectin (*Galanthus nivalis*) capture-ELISA (negative control: Wells coated with lectin, incubated with PBS). Biotinylated 447-52D, diluted in 50 % human serum, was titrated to immobilized Env molecules at indicated concentrations and streptavidin-POD conjugate was used for readout. Env/V3-VLPs were directly coated to ELISA plates and analyzed in the same way (negative controls: Wells coated with blank VLPs (Gag w/o) and PBS, respectively). All ELISAs were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

### 4.1.12 The BG505 SOSIP.664 model system

Contrary to other biophysical or biochemical methods like SPR, BLI and ELISA, MST does not require surface immobilization of one interaction partner, which has the potential to negatively influence the structure of the immobilized interaction partner and thereby the outcome of an experiment. This becomes particular relevant in the analysis of the antigenicity profile of native-like Env immunogens. Consequently, the next step was to examine if next-generation Env immunogens in MST analyses retain their native-like structure that is characterized by a favorable antigenicity profile. The previously used Env/V3 model system, although a valuable tool for assay establishment, cannot address this question since it includes uncleaved, artificial Env chimeras that very likely do not adopt a native-like conformation. Thus, the previously described and well-characterized native-like BG505 SOSIP.664 trimer[61], representing the prototype of next-generation Env immunogens, and a panel of quaternary structure-characterizing mAbs were chosen as an alternative model system. The antibody panel included bNAbs PGT145[178,179] and PGT151[107,108] targeting epitopes in V1/V2 (PGT145) and the gp120-g41 interface (PGT151). The special feature of these antibodies is that they only bind to Env trimers with an intact quaternary structure[179,108,119]. Further, the panel included the CD4-binding site directed non-neutralizing mAb F105[183,184] that binds to monomers, open and disordered trimers but not to well-folded native-like trimers[191,61]. Last but not least, the panel included bNAbs VRC01[181,182] and 2G12[101,105] that target epitopes in the CD4-binding site (VRC01) and the N332-glycan supersite (2G12). Binding of VRC01 and 2G12 does not depend on an intact quaternary structure, since it has been shown that both bNAbs similarly bind to BG505 SOSIP.664 trimers, BG505 gp120-gp41$_{ECTO}$ protomers and BG505 gp120 monomers[119]. Since the antibody panel described above represents a useful tool for structural characterization Env proteins it was used to analyze BG505 SOSIP.664 trimers in MST analyses.

### 4.1.13 Production and purification of BG505 SOSIP.664 trimers

BG505 SOSIP.664 trimers were produced by transient cotransfection of 293F suspension cells with plasmids encoding for Env and Furin and purified to trimer homogeneity using affinity and size exclusion chromatography, as described in **3.4.2**. For pooled protein fractions see SEC profile in **Appendix Figure 1**. Blue PAGE analysis, performed as described in **3.4.4**, revealed that BG505 SOSIP.664 trimers were successfully purified to high trimer homogeneity, see **Figure 28**.

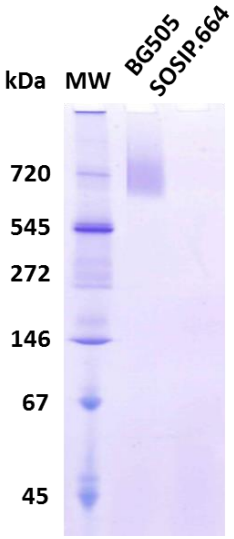



**Figure 28. BN-PAGE analysis of affinity/SEC purified BG505 SOSIP.664 trimers.** Molecular weight (MW) markers are indicated.

### 4.1.14 Interactions between BG505 SOSIP.664 trimers and quaternary structure-characterizing antibodies in buffer and 50 % human serum

In order to examine if native-like trimers retain their favorable antigenicity profile in MST measurements, BG505 SOSIP.664 trimers and a panel of quaternary structure-characterizing antibodies (described in **4.1.12**) were subjected to MST analysis in buffer. To classify results, two different ELISA setups were performed for comparison. These included a His-capture ELISA that allows directed presentation of BG505 SOSIP.664 trimers via their His-tag, and a direct-ELISA setup in which the proteins were directly coated to the surface ELISA plates. The resulting MST binding curves (**Figure 29A**, **left** with corresponding $K_D$-values in **Table 10**) showed that bNAbs VRC01 and 2G12 bound with high affinity to BG505 SOSIP.664 trimers which was expected since the binding of these antibodies is only marginally influenced by both, the oligomerization state and an intact quaternary structure. Quaternary

epitope-specific bNAbs PGT145 and PGT151 also bound with high affinities, whereas non-neutralizing antibody F105 which binds to monomers, open and disordered trimers but not well-folded native-like trimers, demonstrated only indicated binding at high concentrations. Taken together, MST results demonstrated that BG505 SOSIP.664 trimers retained their well-folded structure in MST measurements. His-capture ELISA represents a commonly used method for assessment of the antigenicity profile of Env vaccine candidates since this ELISA setup allows the directed presentation of His-tagged trimers, thereby retaining their native-like structure[191]. Thus, this ELISA format was used to further validate MST results. The results of the His-capture ELISA (see **Figure 29B**, **left** with corresponding $K_D$-values in **Table 10**) in general reflected MST-derived data. All bNAbs bound with high affinities to BG505 SOSIP.664 trimers, whereas non-neutralizing antibody F105 demonstrated only marginal/indicated binding. In contrast, when BG505 SOSIP.664 trimers were directly coated to the solid phase in the direct-ELISA setup, the binding of the quaternary structure-depending bNAbs PGT145 and PGT151 was lost and high affinity binding of F105 was detected, see **Figure 29C**, **left**, with corresponding $K_D$-values in **Table 10**. This indicates a disturbance of the native-like structure, most likely caused by hydrophobic and electrostatic interactions between the BG505 SOSIP.664 trimers and the solid phase of the ELISA plate. In summary, results demonstrate that the native-like protein structure of BG505 SOSIP.664 trimers is preserved during MST measurements since MST results reflect data obtained from His-capture ELISAs, and further that the choice of the immobilization strategy in ELISA analyses can strongly impact the antigenicity profile of immobilized Env proteins.

The antigenicity profile of Env vaccine candidates is mostly assessed in artificial buffer and not under near-physiological conditions. Since experiments within the Env/V3 system have shown that MST has the capacity to quantify interactions in presence of 50 % serum, MST was used to investigate the antigenicity profile of BG505.SOSIP.664 trimers, representing the prototype of next-generation Env immunogens, under near-physiological conditions. Again, His-capture and direct-ELISAs were performed for comparison. Results of MST analyses in 50 % serum, shown in **Figure 29A**, **right**, with corresponding $K_D$-values in **Table 10** were in general comparable to results obtained in buffer analyses (**Figure 29A**, **left**) with the exception that saturated binding of non-neutralizing antibody F105 was detected. This indicates that a least a part of the BG505.SOSIP.664 trimers turned into a more open conformation, which allows binding of F105. However, this was not observed in the His-capture ELISA, which demonstrated results comparable to measurements in buffer, as did direct-ELISA (see **Figure 29B**, **C)**. Surprisingly and unexpected, in the presence of 50 %

human serum binding of 2G12 was substantially reduced in both, His-capture and direct-ELISAs. This turned out to be an effect of the biotinylation of this antibody since the same effect was observed in a control experiment without serum (data not shown).

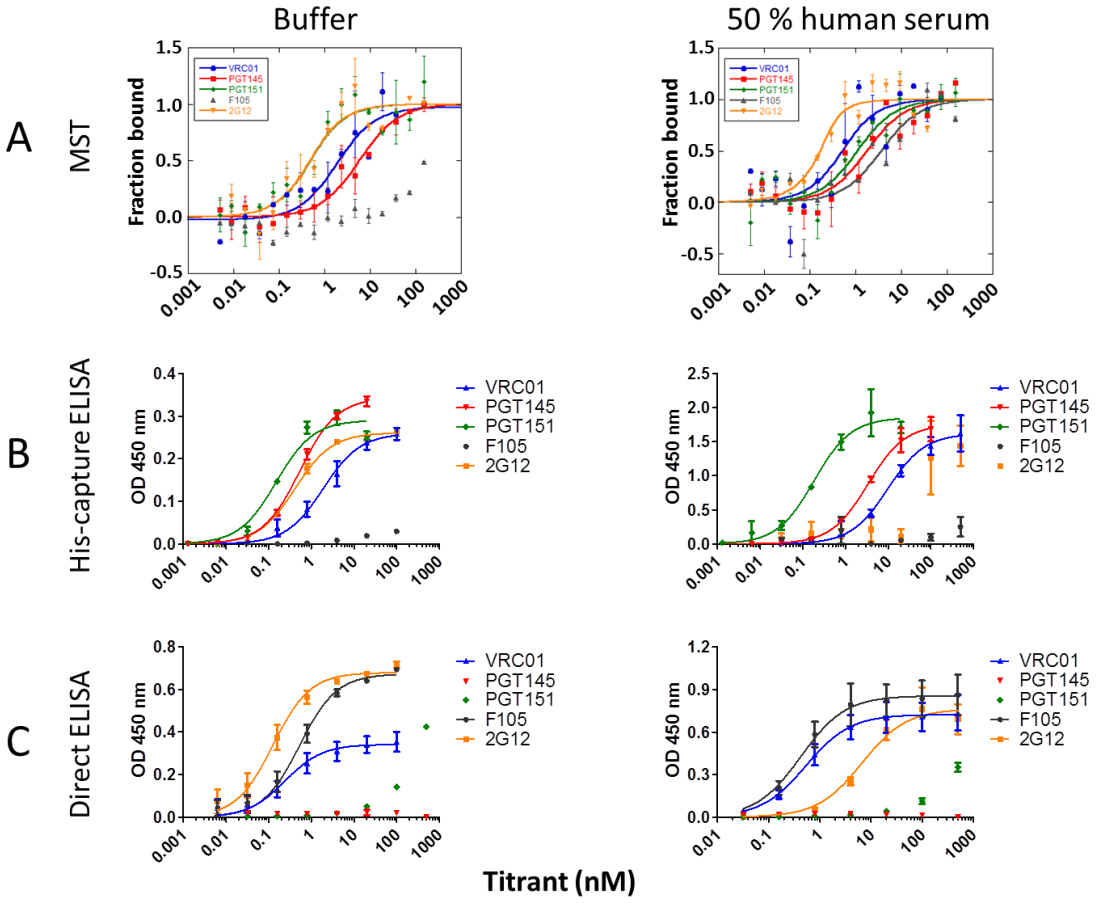



**Figure 29. Interactions of BG505 SOSIP.664 trimers with quaternary structure-characterizing mAbs in buffer and 50 % human serum, quantified by MST, His-capture and direct-ELISA.** For MST analyses, BG505 SOSIP.664 trimers diluted in buffer or 50 % human serum were titrated to Alexa647-labeled mAbs (VRC01, PGT145, PGT151, F105, 2G12) which were kept at constant concentration. Results in buffer **(A, left)** and 50 % human serum **(A, right)** are presented as normalized binding curves of two independent experiments, error bars reflect the SEM of the independent experiments. His-capture ELISA was performed by immobilizing His-tagged BG505 SOSIP.664 trimers via coated anti-His Ab. In case of direct-ELISA, BG505 SOSIP.664 trimers were directly coated to the surface of ELISA plates (negative controls for His-capture and direct-ELISA: Wells with captured/coated PBS). MAbs diluted in PBS were titrated to captured/coated BG505 SOSIP.664 trimers at indicated concentrations. A HRP-conjugated polyclonal rabbit anti-human Ab was used for readout. Results of His-capture and direct-ELISAs in buffer are shown in **(B, C, left)**. His-capture and direct-ELISAs in 50 % human serum **(B, C, right)** were performed the same way, with the differences that biotinylated mAbs were diluted in 50 % human serum and streptavidin-POD conjugate was used for readout. All ELISA experiments were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

**Table 10. Apparent equilibrium dissociation constants of BG505 SOSIP.664-antibody interactions in buffer and 50 % human serum.** $K_D$-values in nM represent the mean value, ± the SEM of n experiments. Detectable, but not saturated signals were excluded from $K_D$-fitting and are therefore designated as "binding indicated". Background-level binding is designated as "no binding" (n.b.). ELISA $K_D$-values were calculated by using non-linear least squares regression (one site binding, hyperbola; GraphPad Prism 5). MST $K_D$-values are computed by the $K_D$-fit model of KaleidaGraph. $R^2$-values of the respective fits are included in brackets.

| **BG505 SOSIP.664 trimers** | | | | | | |
|---|---|---|---|---|---|---|
| | **$K_D$ (nM) in buffer** | | | **$K_D$ (nM) in 50 % human serum** | | |
| **Antibody** | MST (n=2) | His-capture ELISA (n=2) | Direct-ELISA (n=2) | MST (n=2) | His-capture ELISA (n=2) | Direct-ELISA (n=2) |
| **VRC01** | 1.7 ± 1.0 ($R^2$=0.8758) | 1.9 ± 0.4 ($R^2$=0.9688) | 0.25 ± 0.08 ($R^2$=0.9068) | 0.39 ± 0.36 ($R^2$=0.7575) | 9.8 ± 2.3 ($R^2$=0.9623) | 0.54 ± 0.16 ($R^2$=0.9300) |
| **2G12** | 0.35 ± 0.20 ($R^2$=0.8931) | 0.42 ± 0.06 ($R^2$=0.9819) | 0.13 ± 0.02 ($R^2$=0.9740) | 0.042 ± 0.041 ($R^2$=0.9071) | b.i. | 6.7 ± 2.0 ($R^2$=0.9390) |
| **F105** | b.i. | b.i. | 0.53 ± 0.08 ($R^2$=0.9821) | 3.5 ± 2.4 ($R^2$=0.8108) | b.i. | 0.43 ± 0.14 ($R^2$=0.9206) |
| **PGT145** | 5.6 ± 1.3 ($R^2$=0.9703) | 0.54 ± 0.04 ($R^2$=0.9954) | n.b. | 1.6 ± 0.9 ($R^2$=0.8760) | 3.4 ± 0.5 ($R^2$=0.9818) | n.b. |
| **PGT151** | 0.36 ± 0.18 ($R^2$=0.9085) | 0.15 ± 0.03 ($R^2$=0.9615) | b.i. | 0.97 ± 0.48 ($R^2$=0.8935) | 0.18 ± 0.05 ($R^2$=0.9519) | b.i. |

In conclusion, MST analyses of the interaction between BG505 SOSIP.664 trimers and the panel of quaternary structure-characterizing antibody panel in buffer have shown that the native-like Env structure is retained in MST measurements. Further, results of His-capture ELISAs validated MST-derived data. In addition, it has been demonstrated that direct coating of BG505 SOSIP.664 trimers to the surface of ELISA plates destroys the native-like structure of the protein. Importantly, the presence of serum had a detectable influence on the antigenicity profile of BG505.SOSIP.664 trimers in terms of an enhanced binding of the non-neutralizing antibody F105, indicating that at least a proportion of BG505 SOSIP trimers turned into a more open conformation. However, this was only detected in MST analysis and not in the corresponding His-capture ELISA.

### 4.1.15 Interactions between PLA-adsorbed BG505 SOSIP.664 trimers and quaternary structure-characterizing antibodies in buffer and 50 % human serum

As described in **1.5.3**, presentation of immunogens by particles conceptually offers several advantages over soluble proteins, like enhanced uptake of antigens into T-cells or enhanced B-cell activation, highlighting the potential of particles to contribute to stronger and potentially broader Ab responses. However, the antigenic profile of Env immunogens should be preserved when they are presented on particles, since this may contribute to direct antibody responses against vulnerable sites of the Env protein. MST has been proven useful for the characterization of the interactions between mAb 447-52D and VLPs in the Env/V3 model system, and further for the analysis of the antigenicity profile of soluble, native-like BG505 SOSIP.664 trimers. Consequently, it was considered reasonable to use MST to investigate the antigenicity profile of next-generation Env immunogens, presented on a state-of-the-art particle system. For this purpose, poly(lactic acid) particles (PLA particles) with adsorbed BG505 SOSIP.664 trimers were chosen as a model system, since, as described in **1.5.3**, PLA particles represent well-established, biodegradable particles that allow adsorption of proteins, and further offer the option to incorporate immunostimulatory molecules as further adjuvants. Thus, PLA particles represent promising vaccine carriers. PLA particles with adsorbed BG505 SOSIP.664 trimers were produced in cooperation with Bernard Verrier, as described in **3.4.8**. The resulting particles with passively adsorbed BG505 SOSIP.664 trimers were about 175 nm in size with a calculated amount of ~50 trimers per particle.

For MST experiments, PLA particles were titrated to fluorescent-labeled antibodies, starting with a calculated concentration of ~100 nM Env in the PLA-particle solution. This setup was chosen in order to avoid labeling of the limited amount of BG505 SOSIP trimers present on the particle surface since, as described in **4.1.4** , labeling may affect antibody epitopes. However, due to its setup, this experiment allowed only a comparison in binding of different antibodies to PLA-presented BG505 SOSIP.664 trimers, which is sufficient to investigate the antigenicity profile, but it is not suitable to calculate precise $K_D$-values since it is difficult to determine the exact concentration of PLA-adsorbed BG505 SOSIP trimers. However, results of MST measurements in principle demonstrated that all mAbs bound to PLA-adsorbed BG505 SOSIP.664 trimers in both, artificial buffer and 50 % human serum, see **Figure 30A and B**. However, it is important to mention that MST analyses, and in particular measurements in buffer, suffered from unspecific binding of mAbs to control particles

without adsorbed Env proteins, see **Appendix Figure 3**. Even pre-blocking of control particles with blocking buffers did not result in reduction of unspecific binding in buffer analysis (data not shown). Although it seemed that saturated binding of mAbs was only observed for PLA particles with adsorbed BG505 SOSIP.664 trimers, data were difficult to interpret. However, the presence of serum during MST analysis improved the quality of measurements since, despite of some unspecific binding, saturated binding curves were only observed for PLA particles with adsorbed BG505 SOSIP.664 trimers, see **Appendix Figure 4**. Thus, the binding of all mAbs was considered as specific. The resulting antigenicity profile obtained from MST analyses in serum (and also inferred from analyses in buffer) indicated that PLA-adsorbed BG505 SOSIP.664 trimers most likely represented a mixture of trimers with native-like and open conformation since both was detected, binding of bNAbs including quaternary structure-depending PGT145 and PGT151, and binding of non-neutralizing antibody F105. For comparison, the antigenicity profile of PLA-adsorbed BG505 SOSIP.664 trimers was analyzed in a direct-ELISA setup in which PLA particles were directly coated to the surface of ELISA plates and mAbs were used as titrants. This setup was chosen for two reasons: 1) it was considered that an antibody capturing approach, e.g. by 2G12 ELISA, will be unsuitable to capture particles with ~175 nm in size and a limited amount of protein on the surface; 2) it was assumed that immobilized PLA particles will serve as platform which avoids direct coupling of BG505 SOSIP.664 trimers to the surface of the ELISA plate. However, ELISA results, shown in **Figure 30C**, differed from the data obtained from MST analyses. Saturated binding was only detected for bNAbs 2G12 and VRC01 that do not depend on a native-like structure of the Env protein. Non-neutralizing antibody F105 demonstrated indicated binding, whereas binding of quaternary structure-depending bNAbs PGT145 and PGT151 was lost (PGT151 showed somewhat marginal binding, but limited highest antibody concentration used). In addition, there was no detectable background binding of mAbs to control particles in the direct-ELISA setup, see **Appendix Figure 5**. Comparing MST results with ELISA data indicates that direct coating of PLA particles to the surface of ELISA plates disrupted the particle structure together with the structure of BG505 SOSIP.664 trimers since in direct-ELISA, contrary to MST, there was no binding of quaternary structure-depending bNAbs PGT145 and PGT151 detectable.

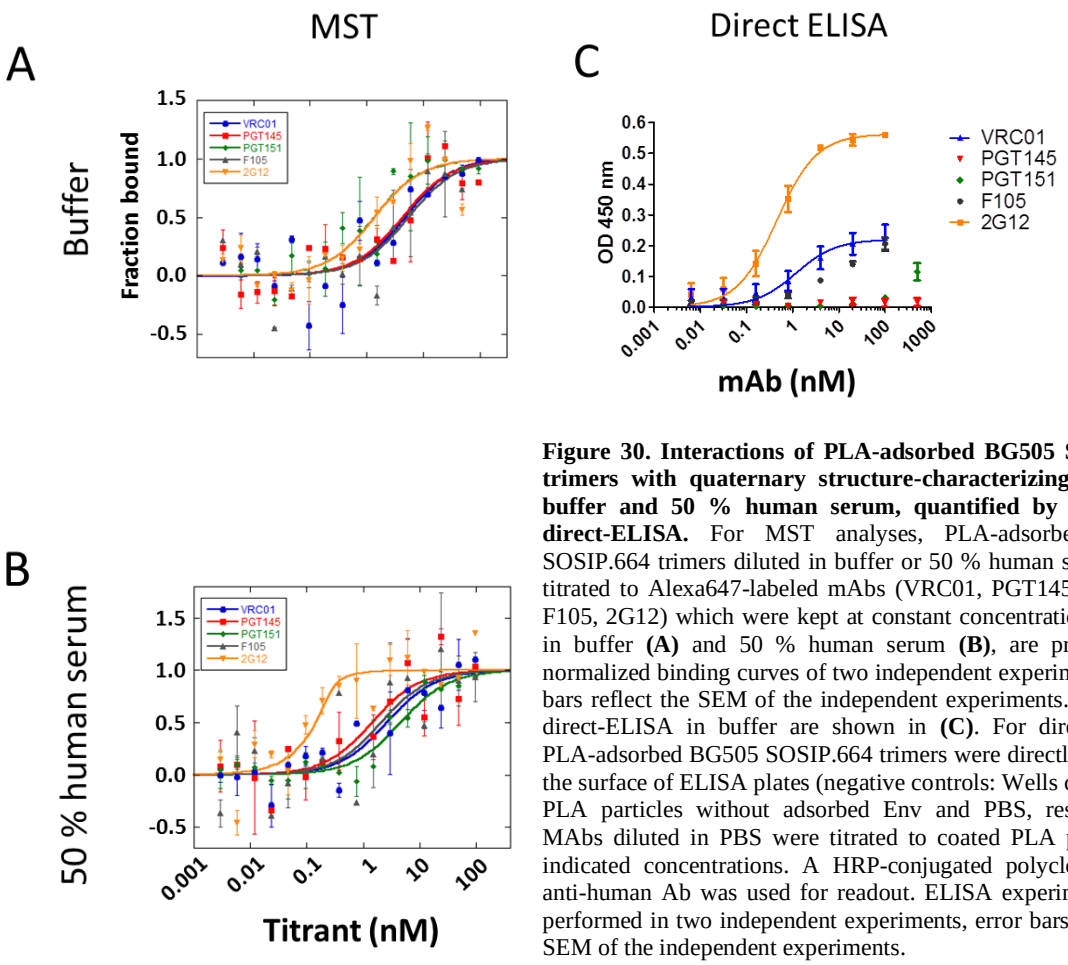



**Figure 30. Interactions of PLA-adsorbed BG505 SOSIP.664 trimers with quaternary structure-characterizing mAbs in buffer and 50 % human serum, quantified by MST and direct-ELISA.** For MST analyses, PLA-adsorbed BG505 SOSIP.664 trimers diluted in buffer or 50 % human serum were titrated to Alexa647-labeled mAbs (VRC01, PGT145, PGT151, F105, 2G12) which were kept at constant concentration. Results in buffer **(A)** and 50 % human serum **(B)**, are presented as normalized binding curves of two independent experiments, error bars reflect the SEM of the independent experiments. Results of direct-ELISA in buffer are shown in **(C)**. For direct-ELISA, PLA-adsorbed BG505 SOSIP.664 trimers were directly coated to the surface of ELISA plates (negative controls: Wells coated with PLA particles without adsorbed Env and PBS, respectively). MAbs diluted in PBS were titrated to coated PLA particles at indicated concentrations. A HRP-conjugated polyclonal rabbit anti-human Ab was used for readout. ELISA experiments were performed in two independent experiments, error bars reflect the SEM of the independent experiments.

In summary, MST analyses of PLA adsorbed BG505 SOSIP.664 trimers were challenging since control experiments revealed unspecific binding of mAbs to PLA particles without adsorbed Env proteins. Thus, it could be not excluded that unspecific binding partially contributed to the binding of mAbs in the MST analyses of PLA-adsorbed BG505 SOSIP.664 trimers. However, binding of mAbs to PLA-adsorbed BG505 SOSIP.664 trimers was considered as specific since saturated binding was only observed in this case. Interestingly, the presence of serum seemed to improve the quality of MST analyses. Taken together, MST analyses of the interactions between PLA-adsorbed BG505 SOSIP.664 trimers and quaternary structure-characterizing mAbs indicate that PLA-adsorbed BG505 SOSIP.664 trimers are present as a mixture of native-like trimers and open conformations. Notably, ELISA results indicate that the direct coating disrupted both, structure of PLA particles and BG505 SOSIP.664 trimers. Thus, the analysis of this kind of particles may benefit from measurements in solution.

## 4.2 Generation of a BG505 SOSIP alanine substitution library for cell-based antibody epitope mapping

The development of next-generation Env immunogens by rational immunogen design requires detailed information of the interaction between bNAbs and Env proteins. Analyses of Env-antibody co-crystals for example provide valuable structural information[60,78,192], whereas functional information about amino acid residues that critically contribute to antibody binding can be obtained from peptide screenings[193,194,] or mutant analyses[175,195,196]. Mammalian cell display in combination with flow cytometry offers the ability to analyze cell-membrane exposed Env mutant libraries in high throughput, and further provides advantages like mammalian glycosylation and presentation of Env proteins in their natural membrane-bound context. Exploiting these features, our group recently developed a mammalian cell display- and flow cytometry-based approach for functional mapping bNAb epitopes in high throughput. For this purpose, a clade C isolate 96ZM651-based alanine substitution library (designed to express gp145) was generated, displayed on the surface of HEK293T cells and subsequently screened for antibody binding via flow cytometry. Several successful mapping experiments highlighted the applicability of the method for functional screening (Kliche and Grassmann *et al*., manuscripts in preparation, PhD thesis Veronika Grassmann[197]). The experiments included the mapping of previously determined bNAb epitopes of HGN194[194], HJ16[194] (PDB ID: 4YE4, Chen *et al*., to be published), VRC01[181,182], PG9[175,176,177] and the epitope of soluble CD4[58]. Thereby it has been demonstrated that the library properly displayed linear (HGN194) and conformational gp120 (HJ16, VRC01, sCD4) epitopes as well as the quaternary epitope of PG9. However, the library was not suitable for mapping of newly identified, highly quaternary structure-depending bNAbs[197], such as PGT145[178,179] or PGT151[107,108]. The cleavage deficiency of the 96ZM651 gp145 library, likely resulting in the expression of Env trimers with perturbed quaternary structure, was considered mainly responsible for this outcome. Hence, one objective of this thesis was the development of an alanine substitution library that overcomes the limitations of the 96ZM651 gp145 library and allows the display of Env proteins with an intact quaternary structure, enabling mapping of bNAbs with quaternary epitopes.

### 4.2.1 Impact of stabilizing SOSIP modifications and cytoplasmic tail truncation on expression levels of cell-membrane expressed BG505 Env

The clade A transmitted/founder isolate BG505 (GenBank # DQ208458.1) was chosen a basis for the development of an alanine substitution library for several reasons: 1) this isolate is known to benefit from intrinsic trimer enhancing propensities in soluble proteins[127], thus BG505 may be also superior to other isolates in terms of forming a homogenous Env trimer population on the cell surface; 2) the soluble, stabilized version BG505 SOSIP.664 displays a favorable antigenicity profile, in terms of binding to nearly all known bNAbs[56], that is likely transferable to the membrane-exposed BG505; 3) The structure of the soluble, stabilized version BG505 SOSIP.664 is well characterized[56,60,78]. Cleavage competence was considered as a key feature of the new library since folding of soluble BG505 SOSIP.664 into native-like trimers depends on proper cleavage of the Env precursor into gp120-gp41 subunits[116]. Notably, cleavage competence is not only desirable for soluble proteins but also for membrane-exposed Env since it avoids binding of non-neutralizing antibodies[198], nowadays seen as evidence for a disturbed quaternary structure. Thus, the library was planned to contain an optimized Furin cleavage site for cleavage enhancement (substitution of the naturally occurring amino acid motif REKR by R6)[61]. However, cleavage involves the risk of the dissociation of non-covalently linked gp120 and gp41 subunits (gp120 shedding[87]). In the soluble, stabilized version of BG505 (BG505 SOSIP.664) and also VLP-exposed Env proteins, dissociation of gp120-gp41subunits was successfully prevented by the introduction of disulfide-bond forming cysteines, termed "SOS" [61,87]. Thus, the introduction of SOS modifications at amino acid positions 501 and 605 (corresponding to soluble BG505 SOSIP.664) was conceived as an option to prevent shedding in the new library approach. Further, it was considered to incorporate the helix-breaking I to P substitution at amino acid position 559 in gp41, since this substitution was included in BG505 SOSIP.664 to promote trimerization[61,199] and may do so in membrane-exposed BG505. However, it was unknown if it is in general possible to express a membrane-bound version of BG505, containing modifications that were originally designed to stabilize soluble proteins. Thus, the first step prior to the generation of the new library was to perform a comparative analysis in order to investigate how stabilizing modifications (and their combination) affect expression levels of membrane-bound BG505, presented on the surface of mammalian cells. In addition, the effect of Env cytoplasmic tail truncation was also investigated in this analysis, since the truncation

of the CT is associated with enhanced Env expression levels[145,200]. In summary, eight different constructs based on the codon-optimized BG505 sequence (GenBank # DQ208458.1) were generated and cloned into the mammalian expression vector pcDNA5/FRT/TO as described in **3.2.1**. The eight constructs, all sharing the optimized Furin cleavage site (R6) and the T332N substitution to restore the central glycan of the V3-glycan bNAb epitope[106,180], consisted of each, four variants with the full-length BG505 sequence (gp160 constructs, including CT), and four variants with truncated CT (gp145 constructs), with the following stabilizing modifications: 1) Wildtype (WT), no further modification; 2) SOS (disulfide-bond forming cysteines); 3) IP (helix-breaking amino acid substitution I559P); and 4) SOSIP (combination of both stabilizing modifications). To compare Env surface expression of the generated Env variants, HEK293T cells were transiently cotransfected with plasmids enconding for the eight Env variants and Furin (to promote cleavage of the proteins) as described in **3.3.2** and analyzed by flow cytometry, see **3.3.6**. The bNAb 2G12 was used as a marker for Env surface expression in flow cytometry since SPR analysis of soluble BG505 variants demonstrated that the binding of this antibody is not affected by stabilizing modifications[119]. The results, shown in **Figure 31**, demonstrated that the truncation of the CT, as expected, resulted in an enhanced Env surface expression since all gp145 variants exhibited a higher 2G12 signal than their respective gp160 counterpart. Notably, the I559P substitution seemed to be beneficial in terms Env expression levels since all variants containing this modification (IP and SOSIP) demonstrated an enhanced surface expression (~factor of 2) compared to the WT in both, gp145 and gp160 Env variants. However, the presence of the SOS modification seemed to slightly reduce the beneficial effect of the I559P substitution in gp145 and gp160 SOSIP variants. In addition, the SOS modification alone did not alter the overall expression levels since the 2G12 signal in gp145 and gp160 SOS variants was always comparable to the corresponding WT signal.

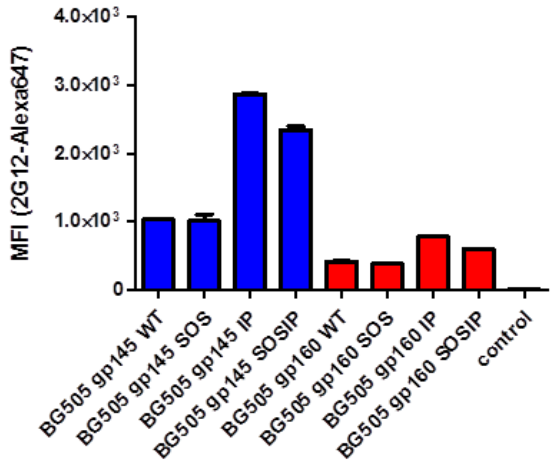



**Figure 31. Impact of stabilizing modifications and cytoplasmic tail truncation on BG505 Env cell-surface expression levels.** HEK293T cells were transiently cotransfected with plasmids encoding for Env and Furin (control: Mock transfected cells, no Env/Furin DNA). 48 h post transfection, cells were stained with 250 nM Alexa647-labeled bNAb 2G12 and analyzed by flow cytometry. Depicted are the mean fluorescence intensities (MFIs) of 2G12 binding to gp145 (blue bars) and corresponding gp160 Env variants (red bars). Results represent the mean value of two independent experiments with error bars reflecting the SEM of the independent experiments.

In conclusion, BG505 SOSIP was chosen as a basis for the development of a new alanine substitution library, mainly for two reasons: 1) Compared to WT, the IP modification resulted in enhanced surface expression, which is beneficial because antibody mapping requires robust antibody signals; and 2) the SOS modification, despite of slightly reducing expression-enhancing effects of IP, could prevent shedding of gp120. Further, there was the question if the library should express full-length gp160 or the CT-truncated gp145 versions of the proteins. There were good reasons to express library variants as CT-truncated gp145 Env proteins. These included the enhanced surface expression and good experiences with the 96ZM651 library that was also designed to express gp145 Env variants. However, Chen *et al*. recently reported that the truncation of the CT diminishes Env binding of trimer-specific bNAbs and results in the exposure non-neutralizing Env epitopes, indicating that the CT largely affects the antigenicity profile of the Env ectodomain[145]. Taking the results of this study in consideration, it was decided to design the library based on the full-length BG505 SOSIP gp160, see **4.2.2**.

### 4.2.2 Features of BG505 SOSIP alanine substitution library

The BG505 SOSIP gp160 alanine substitution library was generated as described in detail in **3.2.3**. Starting with amino acid position 34 (HXB2 numbering) every amino acid (with exception of SOSIP modifications, R6 cleavage site and naturally occurring alanines/cysteines) of the extracellular part of the protein, including the transmembrane domain (gp145, AA 34-712), was consecutively substituted for an alanine while the CT remained unmodified, see **Figure 32B**. In total, 615 alanine substitution variants were generated. However, it was considered that it at some point may be interesting to analyze gp145 or gp140 Env variants (for example to investigate the influence of alanine substitutions

in the context of CT-truncation or in soluble trimers). Thus, for being able to quickly transform the gp160 alanine substitution library into either a gp145 or a gp140 library, elements were implemented to allow this switching by simple molecular biology tools. These consisted of XhoI and SfiI restriction sites and 3´ parts of gp145 and gp140, respectively, see **Figure 32A**. Together these elements allow to convert the BG505 SOSIP gp160 alanine variants into the corresponding gp145 or gp140 constructs by simple restriction digest and re-ligation. The possibility to use the library for generation of Env-expressing stable cell lines represents another feature of the library. Since the library bases on a modified pcDNA5/FRT/TO plasmid vector (pLib1.1, see **3.2.3**) it can be used for generation of stable cell lines using the ThermoFisher Scientific Flp-In System. Here, Flp-In™ T-REx™ 293 cells are cotransfected with a plasmid vector containing a Flp recombination target (FRT) site (e.g. pLib1.1 with Env) and a plasmid encoding for the Flp recombinase (pOG44). Hence, the recombinase can mediate a homologous recombination event between the FRT site of the plasmid vector and the FRT site present in the genome of Flp-In™ T-REx™ 293 cells. This results in the integration of the FRT containing plasmid vector, and thus in the integration of a single copy *env* gene into the predefined FRT site of Flp-In™ T-REx™ 293 cells. Further, stable transfected Flp-In™ T-REx™ 293 cells offer the advantage of a doxycycline inducible gene expression that is regulated by the Tet operator/repressor system. This in turn enables cultivation of stable cell lines without potential negative effects that may be caused by Env cytotoxicity.

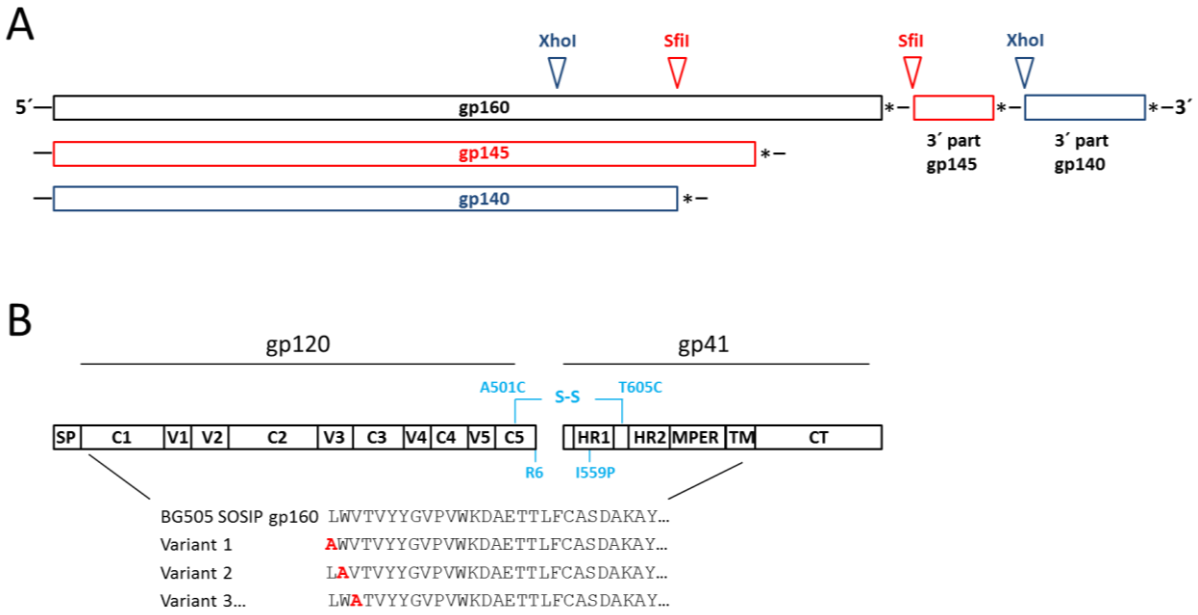



**Figure 32. Schematic overview of the BG505 alanine substitution library. (A)** Schematic representation of BG505 SOSIP gp160 in the library plasmid vector pLib1.1. The BG505 SOSIP alanine substitution library was generated based on the full-length gp160 sequence (black). In this configuration BG505 SOSIP is expressed as full-length, membrane-bound gp160 with gp41 cytoplasmic tail (CT). In order to easily convert the BG505 SOSIP gp160 alanine substitution library into a gp145 or gp140 library, elements were implemented that allow this switching by simple restriction digest and re-ligation. These elements consisted of XhoI and SfiI restriction sites, depicted as triangles in red and blue, and 3´ parts of gp145 and gp140 shown in red and blue, respectively. Stop codons are indicated as asterisks. Restriction digest of the gp160 constructs with SfiI and subsequent re-ligation results in gp145 expressing constructs (red), while XhoI digestion leads to gp140 constructs (blue). Of note, using XhoI digestion, gp145 constructs can be further converted into gp140 constructs. **(B)** Schematic overview of the BG505 SOSIP gp160 alanine substitution library. Starting with the first amino acid after the signal peptide (SP), every amino acid of the extracellular part of the protein, including the transmembrane domain (TM), was consecutively substituted for an alanine (AA 34-712, according to HXB2 numbering), as indicated in red for the first three amino acid positions. Importantly, SOSIP modifications, R6 cleavage site and naturally occurring alanines/cysteines as well as the CT remained unmodified. In summary, 615 alanine substitution variants were generated. Further indicated: Constant regions (C1-C5), variable loops (V1-V5), heptad repeats 1 and 2 (HR1, HR2), membrane-proximal external region (MPER), R6 cleavage site, I559P substitution, disulfide-bond (S-S) forming cysteine substitutions at positions 501 (A501C) and 605 (T605C).

### 4.2.3 Antibody epitope mapping with the BG505 SOSIP gp160 alanine substitution library

The in-depth characterization of the BG505 SOSIP gp160 alanine substitution library as a mapping tool was done by Iris Ganser in her master`s thesis[201], which was supervised by Dr. David Peterhoff. Since some results of the master` thesis were important for this thesis, relevant outcomes are shortly summarized in this section.

During the master´s thesis the BG505 SOSIP gp160 alanine substitution library was tested for its applicability as a mapping tool in several proof-of-principle screenings. In a typical mapping experiment, Env variants of the BG505 SOSIP gp160 alanine substitution library were expressed on the surface of HEK293T cells. To this end, HEK293T cells were transiently cotransfected with plasmids encoding for Env library variants and Furin in a high-throughput 96-well format (one variant per well). After 48 h of Env expression, cells were stained with the Alexa647-labeled screening antibody and analyzed by flow cytometry. To assess the impact of alanine substitutions on antibody binding, the MFI of the alanine substitution variants was normalized to the signal of the WT reference (WT=BG505 SOSIP gp160 without alanine substitution), which was present on each plate. Data that were evaluated this way are further referred to as “unreferenced measurements”. Since this kind of evaluation does not consider Env expression effects, potentially resulting from alanine substitutions themselves or varying transfection efficiencies, a second evaluation method was applied in parallel. In this so called “referenced measurements” the WT-normalized MFI of the screening antibody was divided by the WT-normalized MFI of a reference antibody to normalize for varying Env surface expression levels between different alanine substitution variants. Thereby, a ratio greater than 1 indicated a gain of binding, while a ratio smaller than 1 represented a loss of binding. To enable referenced measurements, cells were screened in a “multiplex” approach, which avoids co-staining of cells with two antibodies in parallel, involving the risk of binding interference between antibodies. For multiplexing, HEK293T cells in two 96-well plates were transfected with identical Env DNAs (one Env variant/well). After 48 h, cells expressing corresponding alanine substitution variants on each plate were mixed, split to equal parts and separately analyzed by screening and reference antibody, respectively. This procedure ensured that cells analyzed by screening and reference antibody were highly comparable regarding transfection efficiency and cell quality, and further avoided co-staining. Furthermore, the multiplexing approach allowed the direct comparison of unreferenced and referenced evaluation, since data for both analyses were generated within

one experiment. Alanine substitutions resulting in significant loss and gain of binding were calculated with t-tests using a false discovery rate (FDR) correction in R. Here, an FDR-corrected p-value < 0.05 was assumed as significant change compared to WT. The first proof-of-principle screening involved mapping of the epitope of bNAb HGN194[194]. This antibody targets a linear (conformational) epitope within the crown of the V3 loop of gp120, which was previously mapped by peptide screening[194,202]. In addition, the epitope of this antibody was successfully mapped using the 96ZM651 gp145 alanine substitution library (Kliche and Grassmann *et al*., manuscript in preparation). Thus, this antibody was considered useful for assay establishment. The HGN194 screening included all BG505 SOSIP gp160 library variants that encode for alanine substitutions in the V3 loop, while randomly selected library variants encoding for alanine substitutions in gp120 surface residues served as controls. The HGN194 screening was performed as described above, while MPER-targeting bNAb 4E10[101,102] was tested as a reference antibody. The unreferenced HGN194 screening identified four alanine substitutions that resulted significant changes of HGN194 binding, see **Figure 33A**. Three of the alanine substitutions reduced HGN194 binding, while one alanine substitution led to enhanced HGN194 binding. Importantly, all alanine substitutions that significantly altered binding of HGN194 were located within the V3 loop, whereas control residues demonstrated no significant changes in binding. However, when the WT-normalized HGN194 signals were referenced to WT-normalized 4E10 signals, almost all screened alanine substitutions demonstrated a significant loss of HGN194 binding, irrespective of V3 loop or control residues, see **Figure 33B**. This effect can be explained by the fact that 4E10 exhibited enhanced binding to nearly all alanine substitution variants compared to the WT (data not shown). Thus, referencing the WT-normalized HGN194 signals to the WT-normalized 4E10 signals resulted in a loss of HGN194 binding. Consequently, 4E10 was not useful as a reference antibody. Additional proof-of-principle screenings aiming to map more complex antibody epitopes (VRC01, PGT121[178,203]) ended up with similar results (data not shown). Overall, proof-of-principle screenings (unreferenced) identified alanine substitutions that mainly resulted in losses of binding in regions that had been described as the respective antibody epitopes, demonstrating that the BG505 SOSIP gp160 alanine substitution library is in principle a suitable tool for epitope mapping[201]. However, epitope-mapping of the quaternary structure-depending bNAb PGT145 delivered no reliable results, since no distinct binding site has been identified, see **Appendix Figure 6**. Overall, mapping conditions were suboptimal due to the lack of a suitable method for referencing and the low Env surface expression levels of BG505 SOSIP gp160. Env expression levels, next to other factors such as

brightness of the fluorescent dye, antibody labeling efficiency and antibody-Env binding affinity, essentially contribute to the flow cytometer`s limit of detection and thus to mapping feasibility[201]. In summary, results demonstrated that the BG505 SOSIP library is in principle applicable for epitope mapping, however further efforts are required to enhance the quality of mapping experiments.

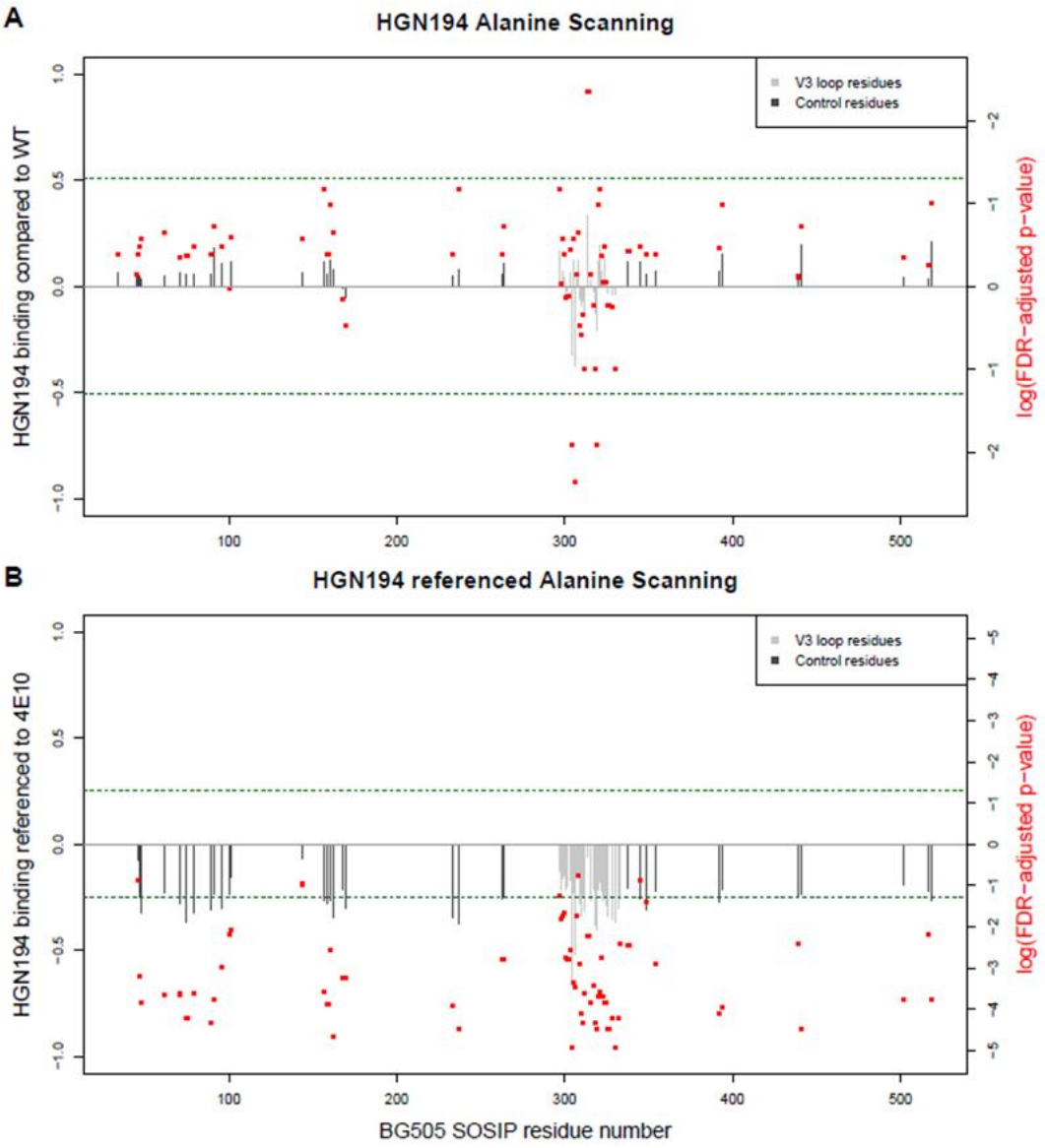



**Figure 33. HGN194 preliminary screening.** Gains (bars upwards) and losses of binding (bars downwards) are plotted for every screened variant. V3 loop residues are shown in light grey, control residues in dark grey. The logarithmically transformed false discovery rate (FDR)-corrected p-value of every library variant is indicated by red points. Points above the upper green dashed line (for gains of binding) or below the lower green dashed line (for losses of binding) correspond to a p-value < 0.5. **(A)** In the unreferenced screening, three significant losses and one significant gain of binding could be identified in the V3 loop. **(B)** As the reference antibody 4E10 showed gains of binding for almost all variants, the HGN194 signals were calculated as losses of binding in the referenced setup. The screening was performed in five replicates, data were generated during the master`s thesis of Iris Ganser under experimental supervision of Dr. David Peterhoff.

### 4.2.4 Summary

One objective of this thesis was the generation of an Env alanine substitution library that can be used for mapping of antibodies with quaternary epitopes in a mammalian cell display approach. As precursor cleavage essentially contributes to proper protein folding, cleavage competence was a prerequisite of the new library. Since SOSIP modifications were successfully used to express soluble, fully-cleaved, stabilized Env trimers, it was considered to implement these modifications in the library design in order to benefit from their stabilizing effects, regarding the expression of a homogeneous trimer population on the cell surface. BG505 was chosen as a library base because this isolate benefits from trimer enhancing propensities and was successfully used for the generation the well-characterized BG505 SOSIP.664 trimers, binding to nearly all known bNAbs. Prior to library generation, a comparative analysis was performed that aimed to investigate the impact of stabilizing modifications and CT truncation on expression levels of cell membrane-bound BG505, see **Figure 31**. Notably, all Env variants carrying the I559P substitution demonstrated enhanced surface expression compared to the WT (in gp145 and gp160), whereas the SOS modification did not alter overall expression levels. Further, as expected, truncation of the cytoplasmic tail increased Env surface expression. Since expression levels seemed to be sufficient and in consideration of a recent publication[145], BG505 SOSIP gp160 was chosen as basis for the new alanine substitution library. In summary, the extracellular part of BG505 SOSIP gp160 was consecutively substituted by an alanine, resulting in a total number of 615 generated alanine substitution variants, see **Figure 32B**. Further, elements were implemented that allow conversion of BG505 SOSIP gp160 alanine variants into the corresponding gp145 or gp140 constructs by simple restriction digest and re-ligation, see **Figure 32A**. In addition, the plasmid vector utilized for library generation offers the ability to generate stable cell lines with doxycycline inducible Env expression via the ThermoFisher Scientific Flp-In System. Mapping experiments performed by Iris Ganser during her master`s thesis demonstrated that BG505 SOSIP gp160 library is in principle a suitable tool for epitope mapping. Overall, proof-of-principle screenings (e.g. HGN194, see **Figure 33**) identified alanine substitutions that mainly resulted in losses of binding in regions that had been described as the respective antibody epitopes. However, mapping of the PGT145 epitope remained unsuccessful. Overall, measurements suffered from low expression levels of gp160 Env and the lack of a suitable method for referencing. In summary, epitope mapping using the BG505 SOSIP alanine substitution library requires further optimization efforts.

## 4.3 High-throughput analysis of a cellular BG505 SOSIP gp145 glycan-knockout library by FACS in combination with NGS

Screening of mutant libraries displaying a vast diversity of cell-surface exposed Env proteins by FACS offers great opportunities for Env immunogen design, including the identification of amino acid residues that are critical for protein structure or binding of bNAbs. Further, the method can be used to select Env variants with distinct phenotypes, e.g. enhanced binding to bNAbs or reduced binding to non-neutralizing antibodies. In addition, cell sorting offers the ability to re-cultivate cells with desired phenotypes, and thus allows iterative selection processes that can be used for the development of germline-targeting immunogens and intermediate Env variants for sequential immunization strategies (see **1.5.2**). Prior to this thesis, Dr. Tim-Henrik Bruun and Dr. Veronika Grassmann developed a mammalian cell display and FACS-based procedure that allows selection of Env variants according to their antibody affinity. The applicability of the method was demonstrated by enrichment analysis of a cell-surface expressed five-member Env/V3 chimeric model library (see **4.1.2**)[154]. One objective of this thesis was to apply the selection procedure to a cellular library expressing membrane-exposed, cleavage-competent BG505 SOSIP alanine substitution variants that result in the deletion of N-linked glycosylation sites.

### 4.3.1 Overview of the mammalian cell display-based FACS selection system

With the exception of minor adjustments, the mammalian cell display-based FACS selection used in this work was identical to the procedure that was previously reported by Bruun and Grassmann *et al.*[154]. Thus, the latter is shortly reviewed. The previously described mammalian cell display and FACS-based procedure consisted of five different steps, see **Figure 34**. The first step comprises the generation of a cellular library, consisting of stable HEK293T cell lines that carry a single copy of a given *env* variant at a defined locus in the cellular genome. Thereby, the stable cell lines with doxycycline inducible Env expression are generated using the ThermoFisher Scientific Flp-In System. To this end, Flp-In™ T-REx™ 293 cells are cotransfected with a plasmid vector containing a Flp recombination target (FRT) site (in this case pQL13 encoding for Env) and a plasmid encoding for the Flp recombinase (pOG44), see **Figure 34**, **step 1**. The Flp recombinase can mediate a homologous recombination event

between the FRT site of the plasmid vector and the FRT site present in the genome of Flp-In™ T-REx™ 293 cells. This leads to the integration of the plasmid vector into the predefined FRT site of Flp-In™ T-REx™ 293 cells. Since Flp-In™ T-REx™ 293 cells contain precisely one FRT site, integration of a single copy *env* gene results in linkage of genotype and phenotype. Further, by successful integration of the plasmid vector in to FRT site, Flp-In™ T-REx™ 293 cells acquire a hygromycin resistance that allows selection of stable transfected cells. After about 25 days of selection, stable cell lines are pooled to generate a cellular Env library, see **Figure 34**, **step 2**. Prior to FACS, the Env expression of stable cell lines is induced by doxycycline as the Env expression in Flp-In™ T-REx™ 293 cells is regulated by the Tet operator/repressor system. After staining with a fluorescent-labeled screening antibody (**Figure 34**, **step 3**), Env expressing cells are subjected to FACS to select for Env variants with the desired antibody binding profile, see **Figure 34**, **step 4**. Thereby, Env expression levels were normalized to green fluorescent protein (GFP) expression as Env and GFP expression were genetically coupled by a "self-cleaving" TaV 2A peptide[154,204] (feature of the pQL13 vector). To determine the genotypes of selected Env variants, and further to calculate the enrichment of Env variants by the FACS selection process, gDNA of input and output samples is recovered and analyzed by NGS, see **Figure 34**, **step 5**. Alternatively, selected cells can be re-expanded and applied to a second round of sorting to increase enrichment rates.

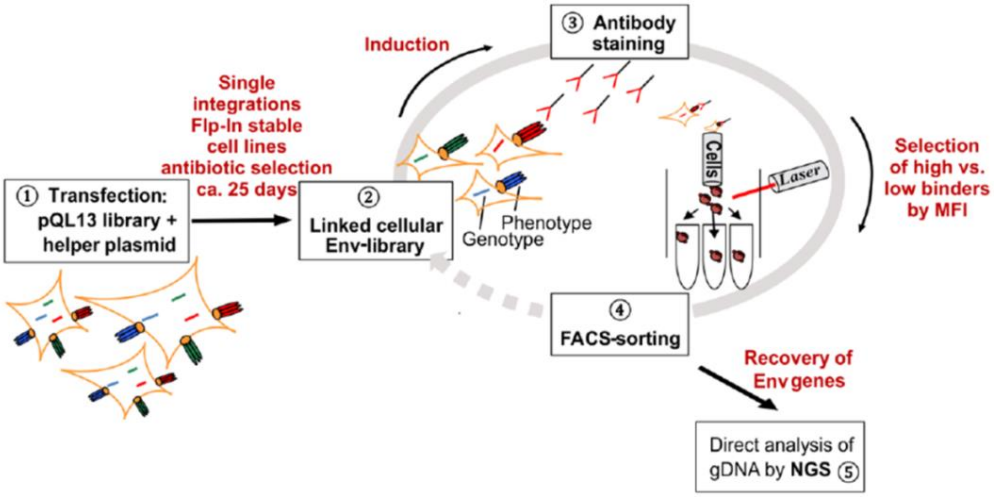



**Figure 34. Schematic overview of the FACS-based selection procedure. (1)** Generation of stable cell lines using the ThermoFisher Scientific Flp-In System. Flp-In™ T-REx™ 293 cells are cotransfected with an Env encoding plasmid vector (pQL13) containing a Flp recombination target (FRT) site and the helper plasmid pOG44 encoding for the Flp recombinase. The Flp recombinase mediates the integration of the FRT containing plasmid vector into the predefined FRT site of Flp-In™ T-REx™ 293 cells via homologous recombination. As Flp-In™ T-REx™ 293 cells contain precisely one FRT site, linkage of genotype and phenotype is achieved. Successful integration of the plasmid vector into the FRT site of Flp-In™ T-REx™ 293 cells results in the acquisition of a hygromycin resistance, allowing antibiotic selection of stable cell lines. **(2)** Stable cell lines are pooled to generate a cellular Env library. After induction of Env expression with doxycycline **(**Env expression in Flp-In™ T-REx™ 293 cells is regulated by the Tet operator/repressor system), cells are stained with a fluorescent-labeled screening antibody **(3)**. FACS is used to select for cells that display Env variants with the desired antibody binding profile **(4)**, thereby Env expression levels are normalized to genetically coupled GFP expression. Genomic DNA (gDNA) of input and output samples is recovered and analyzed by next generation sequencing (NGS) to determine genotypes of selected Env variants, and further to calculate the enrichment of Env variants by the FACS selection process **(5)**. Alternatively, to enhance enrichment rates, selected cells can be re-expanded and applied to a second round of sorting (indicated by dashed line). Figure adapted from Bruun, T.-H. *et al*. Mammalian cell surface display for monoclonal antibody-based FACS selection of viral envelope proteins. *MAbs* 9, 1052–1064 (2017)[154].

### 4.3.2 Comparative analysis of stable cell lines and transiently transfected cells regarding Env expression and antigenicity

The initial step in the mammalian cell display-based FACS selection system comprises the generation of a cellular library based on Env expressing stable cell lines. Based on the experiences made with the BG505 SOSIP gp160 alanine substitution library in transient mapping experiments there were obvious reasons to use BG505 SOSIP gp145 as basis for the cellular library since it was expected that CT truncation likely enhances Env expression. However, keeping in mind that the CT might largely affect the antigenicity profile of the Env ectodomain[145], there was again the question if the cellular library should express full-length gp160 or the CT-truncated gp145 version of the Env protein. In order to answer this question and find an optimal solution two aspects were investigated in detail, including: 1) Env expression levels in stable cell lines; and 2) the influence of the CT on antibody binding to stabilized, cell-surface presented BG505 SOSIP Env proteins. The first step consisted of the

generation of stable cell lines for expression BG505 SOSIP gp160 full-length and CT-truncated gp145 Env proteins. Thereby, stable cell lines were generated based on a constitutive Furin overexpressing cell line to promote Env precursor cleavage. This cell line, termed Flp-In™ T-REx™ 293 CMV_Furin+ and further referred to as CMV F+, constitutively overexpresses Furin under a control of CMV promotor and was previously established by our group during the master`s thesis of Dominik Fiegle[165]. Next to constitutive Furin overexpression, CMV F+ cells offer all advantages of normal Flp-In™ T-REx™ 293 cells (generation of stable cell lines with inducible Env expression, see **4.3.1**) since the latter were used as basis for their generation. Stable cell lines for expression of BG505 SOSIP gp160 and gp145 Env were generated by stable transfection of CMV F+ cells with plasmids (pLib1.1, see **3.2.3**) encoding for BG505 SOSIP gp160 or gp145, see **3.3.3**. After successful selection, CMV F+ cell lines with stably integrated BG505 SOSIP gp145 or gp160 *env* genes were analyzed for their Env surface expression levels. To this end, stable cell lines were cultured in presence of doxycycline to induce Env expression (see **3.3.4**) and analyzed via flow cytometry (see **3.3.6**) using bNAb 2G12 as a marker for Env surface expression. For comparison, HEK293T cells transiently cotransfected with plasmids encoding for Furin and BG505 SOSIP gp145 or gp160 were also included in this analysis. The results, shown in **Figure 35**, demonstrated that stable cell lines were superior to transiently transfected cells regarding expression levels of BG505 SOSIP gp145 (~factor of 2) and gp160 Env (~factor of 5). Further, as expected, the presence of the CT in gp160 full-length Env resulted in reduced expression levels also in stable cell lines.

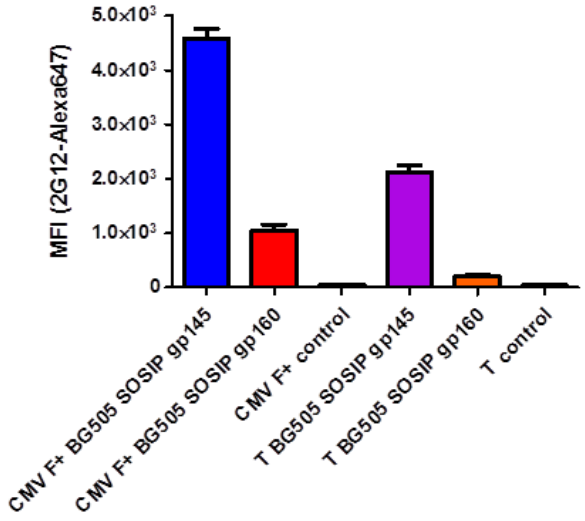



**Figure 35. Comparative analysis of the expression levels of cell-surface presented BG505 SOSIP gp145 and gp160 Env, expressed by stable cell lines and transiently transfected cells.** HEK293T cells were transiently co-transfected with plasmids encoding for Furin and BG505 SOSIP gp145 or gp160 Env (transfection in T175 flasks). Cell lines, based on constitutive Furin overexpressing CMV F+ cells, with stably integrated BG505 SOSIP gp145 or gp160 *env* genes, were cultured in T175 flasks and Env expression was induced with doxycycline. 48 h post transfection/induction, cells were stained with 250 nM Alexa647-labeled bNAb 2G12 and analyzed by flow cytometry. Depicted are MFIs of 2G12 binding to BG505 SOSIP gp145 and gp160 Env expressed by stable cell lines (CMV F+, blue and red bars) and transiently transfected HEK293T cells (T, purple and orange bars). CMV F+ cells (doxycycline induced, without integrated *env*) and mock transfected HEK293T cells (plasmid vector without *env* together with a Furin encoding vector) were used as negative controls (black bars) for stable cell lines (CMV F+ control) and transiently transfected cells (T control), respectively. Results represent the mean value of four independent experiments with error bars reflecting the SEM of the independent experiments.

Next, stable cell lines were analyzed in flow cytometry-based antibody titration experiments to investigate the influence of the CT on the BG505 SOSIP antigenicity profile. In order to compare the antigenicity profile between Env expressing stable cell lines and transiently transfected cells, the latter were also included in the experiment. The antibody panel used for assessment of Env antigenicity profiles consisted of the quaternary structure-characterizing mAbs described in **4.1.12** and bNAb PG9[175,176,177] recognizing a quaternary epitope in V1/V2 as well as mAb 17b[185,186] binding to Env with transiently sampled- or CD4-induced (CD4i) open conformations[205]. In the experiment, transiently transfected cells and stable CMV F+ cells expressing BG505 SOSIP gp145 or gp160 were subjected to antibody titration and analyzed by flow cytometry as described in **3.3.7**. Thereby, binding signals of titration antibodies were normalized for varying Env expression levels and referred to as "relative MFI". This was achieved by calculating the ratio between binding signals of titration antibodies and 2G12, which was used as a marker for Env surface expression (cells were analyzed in separate wells to avoid co-staining and potential antibody-binding interference, see **3.3.7**. Results of the titration experiments are shown in **Figure 36**. With regard to BG505 SOSIP expressing stable cell lines, all bNAbs, including quaternary-structure depending bNAbs PG9, PGT145 and PGT151 demonstrated comparable binding to CT-truncated gp145 and full-length gp160, while non-neutralizing antibodies F105 and 17b exhibited no binding. This indicates that membrane-exposed BG505 SOSIP displays a favorable antigenicity

profile, which is not affected by the presence or absence of the CT. Similar was observed within transiently transfected cells, here BG505 SOSIP gp145 and gp160 were comparably bound by bNAbs and, however, also by non-neutralizing antibodies F105 and 17b. Interestingly, comparing antigenicity profiles of BG505 SOSIP gp145/gp160 expressed by stable cell lines and transiently transfected cells revealed that quaternary-structure depending bNAbs PG9, PGT145 and PGT151 by trend exhibited better binding to proteins expressed by stable cell lines. In turn, BG505 SOSIP gp145 and gp160 expressed by transiently transfected cells demonstrated binding to non-neutralizing antibodies F105 and 17b, which was not the case for corresponding proteins expressed by stable cell lines. Taken together, these observations indicate that stable cell lines display homogeneous BG505 SOSIP gp145 and gp160 trimer populations whereas transiently transfected cells likely present a mixture of native-like trimers and trimers with an open conformation (or Env in other oligomerization states e.g. dimers/monomers). BNAb 2G12, as expected, bound similar to BG505 SOSIP gp145 and gp160, irrespective if expressed by stable cell lines or transiently transfected cells. BNAb VRC01 demonstrated better binding to BG505 SOSIP gp145 and gp160 expressed by transiently transfected cells compared to stable cell lines. Interestingly, a previously reported comparative analysis of soluble BG505 SOSIP.664 and less stabilized/uncleaved versions of BG505 demonstrated that VRC01, although binding to all versions of BG505, exhibited the weakest binding to the most stabilized, native-like form BG505 SOSIP.664[119]. Consequently, more open BG505 SOSIP conformations or presence of other oligomerization states may have contributed to enhanced VRC01 binding in case of transiently transfected cells. The same reasons are likely responsible for binding of 17b (not CD4-induced) to BG505 SOSIP g145 and gp160 expressed by transiently transfected cells. Irrespective if expressed by stable cell lines or transiently transfected cells, BG505 SOSIP gp145 and gp160 demonstrated binding to 17b after incubation with soluble CD4 (sCD4) that induces the (open) CD4i Env conformation resulting in the exposure of the 17b epitope. This result was expected since it was previously reported that sCD4 induces the CD4i conformation in soluble BG505 SOSIP.664 trimers[61]. However, 17b demonstrated enhanced binding to CD4-induced BG505 SOSIP gp145 and gp160 expressed by transiently transfected cells, compared to the same proteins expressed by stable cell lines.

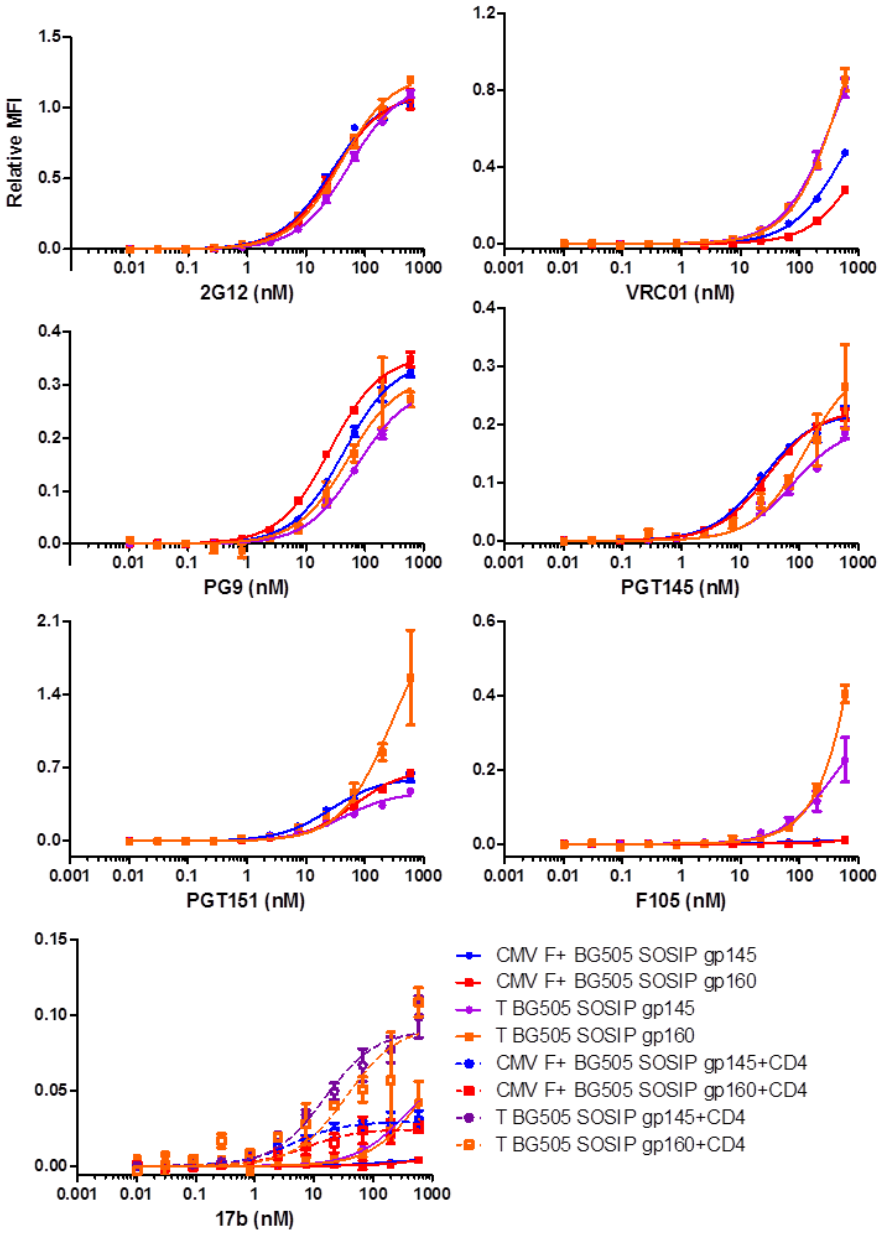



**Figure 36. Comparative analysis of the antigenicity profile of cell-surface presented BG505 SOSIP gp145 and gp160 Env, expressed by stable cell lines and transiently transfected cells.** HEK293T cells were transiently co-transfected with plasmids encoding for Furin and BG505 SOSIP gp145 or gp160 Env (transfection in T175 flasks). Cell lines, based on constitutive Furin overexpressing CMV F+ cells, with stably integrated BG505 SOSIP gp145 or gp160 *env* genes, were cultured in T175 flasks and Env expression was induced with doxycycline. 48 h post transfection/induction, cells were split into two parts and analyzed by flow cytometry. Thereby, one part of cells was stained with 250 nM Alexa647-labeled bNAb 2G12 to determine Env surface expression levels, while the other part was used for titration of Alexa647-labled mAbs. Of note, binding of mAb 17b (recognizes the CD4-induced Env configuration) was additionally analyzed in presence of soluble CD4.To this end, cells were pre-incubated with sCD4, followed by 17b titration. Doxycycline induced CMV F+ cells without integrated *env* and mock transfected HEK293T cells (plasmid vector without *env* together with a Furin encoding vector) were used as negative controls. For evaluation, MFIs of Env expressing cells were corrected for antibody background binding by subtracting the respective negative control signals and are further referred to as referenced MFIs. To normalize binding signals of titration antibodies for varying Env expression levels, the respective referenced MFIs obtained from titration experiments were divided by the corresponding referenced 2G12 signals and are further referred to as relative MFIs. Shown are binding curves (relative MFI) of antibody titrations to BG505 SOSIP gp145 and gp160 Env expressed by stable cell lines (CMV F+, in blue and red) and transiently transfected HEK293T cells (T, in purple and orange). Binding curves of 17b in presence of sCD4 (+CD4) are depicted as dashed curves. Results represent the mean value of two independent experiments with error bars reflecting the SEM of the independent experiments.

In summary, the constitutive Furin overexpressing cell line CMV F+ was used for the generation of stable cell lines with integrated BG505 SOSIP gp145 or gp160 *env* genes. A comparative analysis between the generated stable cell lines and transiently transfected cells revealed that stable cell lines are superior to transiently transfected cells regarding BG505 SOSIP gp145 and gp160 expression levels. Further, the antigenicity profiles of BG505 SOSIP gp145 and gp160 expressed by stable cell lines and transiently transfected cells were analyzed in titration experiments, using a panel of quaternary structure-characterizing mAbs. Here, it has been demonstrated that the favorable antigenicity profile of membrane-exposed BG505 SOSIP is not affected by presence or absence of the CT in both, stable cell lines and transiently transfected cells. Notably, stable cell lines were superior to transiently transfected cells regarding the Env antigenicity profile. Whereas binding of bNAbs and some binding of non-neutralizing antibodies was detected for BG505 SOSIP gp145 and gp160 expressed by transiently transfected cells, corresponding Env proteins expressed by stable cell lines were exclusively bound by bNAbs and not by non-neutralizing antibodies. This indicates that stable cell lines display homogeneous BG505 SOSIP gp145 and gp160 trimer populations.

### 4.3.3 Generation and quality control of a cellular BG505 SOSIP gp145 glycan-knockout library

The extensive glycosylation of Env, acting as a "shield" to prevent binding of neutralizing antibodies, essentially contributes to HIV immune evasion, see **1.3.3**. However, structural analyses revealed that a proportion of bNAbs does not only bind to gp120 protein residues but also recognizes glycans attached to gp120[206,176,207,180,208,209]. Based on these findings, the analysis of the Env glycan shield became increasingly important in HIV-1 Env vaccine design. Thus, this thesis focused on the generation of a cellular BG505 SOSIP glycan-knockout library that can be used to analyze the impact of single glycan deletions on antibody binding via a mammalian cell display and FACS-based selection procedure, see **4.3.1**. Since antibody titration experiments (see **4.3.2**) demonstrated that the truncation of the CT does not negatively affect the antigenicity profile of membrane-exposed BG505 SOSIP, the cellular library was generated based on gp145 to benefit from high expression levels.

The already existing BG505 SOSIP gp160 alanine substitution library was used as a template for the generation of the cellular BG505 SOSIP gp145 glycan-knockout library (for details see **3.3.5**). Briefly, the following pLib1.1 plasmid vectors of the BG505 SOSIP gp160 alanine substitution library were cloned into gp145 expressing constructs: 1) all plasmid vectors that encode for alanine substitutions in the 28 NXT/S glycosylation sequons (alanine substitutions

of N and T/S, since it was expected that both alanine substitutions equally prevent the glycosylation of the asparagine in the sequon); 2) two plasmid vectors that encode for alanine substitutions at two potential O-linked glycosylation sites[50,51] (positions T499 and T606, HXB2 numbering). The resulting 58 pLib1.1 plasmids were used for stable transfection of CMV F+ cells to generate 58 individual stable cell lines. Finally, 56 out 58 stable transfections resulted in the successful generation and selection of stable cell lines. Next, stable cell lines were tested for their Env expression levels. For this purpose, stable cell lines with induced Env expression were analyzed via flow cytometry (see **3.3.6**) using bNAbs 2G12 and VRC01 as markers for Env surface expression. VRC01 was chosen a second marker since binding this bNAb does not depend on the presence of glycans[210]. At a first glance, all stable cell lines demonstrated Env expression when analyzed by 2G12 and VRC01, see **Figure 37A**. Further, there was a good correlation between expression levels quantified by 2G12 and VRC01, see **Figure 37B**, indicating that both antibodies are suitable for the quantification of Env surface expression. Overall, expression levels between stable cell lines and the WT (no glycan deletion) varied by approximately a factor of 2 or less. Thereby most variants demonstrated a slightly reduced expression compared to the WT. However, some stable cell lines including variants N262A, N448A and T450A exhibited very low Env expression levels. Repeating the analysis for these particular cell lines confirmed low expression levels, except for variant T450A which demonstrated negligible expression (data not shown). In addition, N and T/S alanine substitutions in a particular sequon tended to result in comparable expression levels. Taking a closer look to glycans in the proximity of VRC01 and 2G12 epitopes lead to interesting results. VRC01 for example demonstrated enhanced binding (compared to WT) to stable cell lines expressing Env variants with a glycan deletion at AA position N276 (N276A and T278A), which was not observed for 2G12.This observation can be explained by the critical role of the N276 glycan regarding the accessibility of the CD4bs. It has been previously reported that the presence of this glycan likely prevents interaction between germline VRC01 and Env due to a steric clash[135]. Further, as demonstrated in the PhD thesis of Veronika Grassmann, the removal of this glycan also improves binding of the mature VRC01[197]. Interestingly, the comparison of VRC01 and 2G12 expression levels revealed the glycan deletion at AA position N332 (N332A, S334A) resulted in WT-like 2G12 binding, despite the fact that the N332 glycan is part of the 2G12[105] epitope.

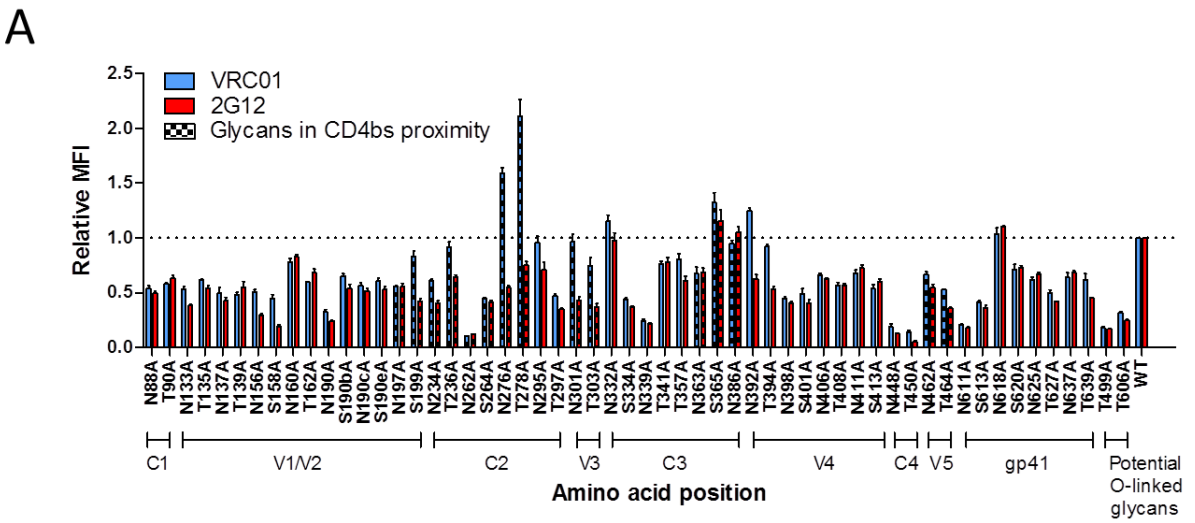



**Figure 37. Expression analysis of stable cell lines displaying BG505 SOSIP gp145 Env variants with single glycan deletions.** Stable cell lines were cultured in 96-well plates (all 56 variants in one plate, six plates in total) and Env expression was induced with doxycycline. Stable cell lines expressing BG505 SOSIP gp145 WT (WT=no glycan knockout) and CMV F+ cells without integrated *env* were present on each plate as positive and negative controls, respectively. 48 h post induction, cells were stained with 250 nM Alexa647-labeled bNAbs 2G12 or VRC01 (three plates for each antibody) and analyzed by flow cytometry. To compare expression levels between library variants and WT and between different antibodies used for their quantification, data were processed as follows. MFIs of Env expressing cells were corrected for the respective antibody background binding by subtracting negative control signals (resulting in referenced MFIs) and set into relation to the WT signal (relative MFIs). Relative MFIs were separately calculated for each plate (n=3 for each antibody). **(A)** Env expression levels of library variants compared to the WT (WT=1, indicated by the dashed line) measured by 2G12 (red) and VRC01 (blue). Shown are mean values and SEM of the three replicates. Further illustrated: Positions of glycans within Env regions and proximity to the CD4 binding site (CD4bs). **(B)** Linear regression analysis of Env expression levels, quantified by 2G12 and VRC01. Alanine substitution variants N276A and T278A resulting in enhanced binding of VRC01 compared to WT (not observed for 2G12) are highlighted. Linear regression analysis was performed using GraphPad Prism 5.

In summary, the BG505 SOSIP gp160 alanine substitution library was used as a template to generate stable cell lines expressing gp145 Env variants with alanine substitutions in NXT/S glycosylation sequons. Thereby, stable cell lines were individually generated and tested for Env expression. Finally, individual stable cell lines were pooled to equal parts to generate the cellular BG505 SOSIP gp145 glycan-knockout library.

### 4.3.4 FACS of the BG505 SOSIP gp145 glycan-knockout library with bNAb PGT135

The bNAb PGT135[178,180,106] recognizes the V3-glycan epitope of Env (see **1.4.2**) and its binding depends on the presence of the glycan at AA position N332. Due to its glycan-dependency PGT135 was chosen as a model antibody for an initial FACS analysis with the BG505 SOSIP gp145 glycan-knockout library. Thereby, to get a first glance, the FACS-based methodology focused on the identification of glycans that result in a loss of PGT135 binding.

FACS was performed as described in **3.3.8**. In the first step, a freshly thawed aliquot of the BG505 SOSIP gp145 glycan knockout library (generated by pooling of independently generated stable cell lines) was grown to sufficient cell numbers (18 days). After induction with doxycycline, cells were grown for further 48 h and co-stained with PGT135 (100 nM, linear binding range) and VRC01 (250 nM) which was used as expression marker. VRC01 was chosen as a reference antibody because a control experiment demonstrated that VRC01 does not compete with PGT135 binding as 2G12 did (data not shown). Finally, approximately 2 x $10^7$ cells were subjected to FACS selection. To select for cells that exhibit a loss of PGT135 binding in relation to their Env expression level, cells were sorted for the lowest PGT135 signal in relation to the VRC01 signal (gate P3), see **Figure 38**. In summary, from approximately 1.5 x $10^7$ analyzed cells about 3 x $10^5$ cells were sorted in to the final gate P4 (living cells).

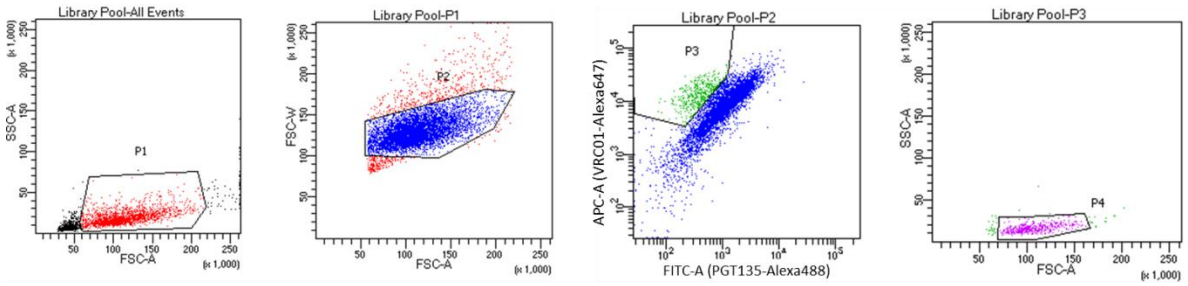
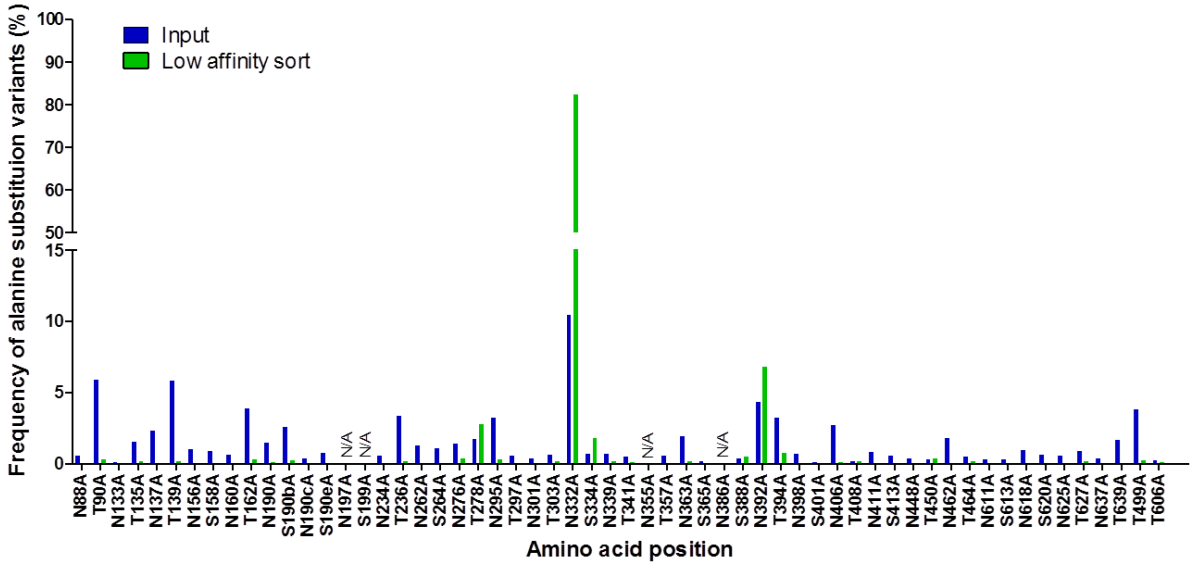



**Figure 38. Gating strategy of the FACS-based selection procedure.** The sorting procedure of 1 x $10^4$ cells is exemplarily shown. Cells were gated for living, single cells (P1 and P2). Cells separated into uninduced (low FITC-A / low APC-A) and induced populations (higher APC-A). Induced cells were gated for lowest (P3) FITC-A signals (PGT135) in relation to their APC-A signal (VRC01, surface expression marker). Finally, sorted cells were again gated for living cells (P4). Gating hierarchy: All events (black) → P1 (red) → P2 (blue) → P3 (green) → P4 (purple).

Subsequently, gDNA samples of sorted cells and input material were analyzed by NGS, see **3.2.5**. This enabled the identification of Env variants present in the input material and sorted cells, and further allowed to calculate the fold-enrichment of variants displaying a PGT135 low-affinity phenotype. The frequency of alanine substitution variants in sorted cells compared to the input material is shown in **Figure 39A**. Variant N332A, knocking out the essential glycan required for PGT135 binding, accounted for ~82 % of the sorted population, corresponding to a 35-fold enrichment of this variant in the sorted cells, see **Figure 39B**. Env variant S334A, also expected to knockout the glycosylation of N332, however, was only enriched by a factor of 2. All other Env variants demonstrated no enrichment in the sorted population.

A




B

| Fold enrichment[a] low affinity gate | |
|---|---|
| N332A | 35 |
| S334A | 2 |
| All other variants | No enrichment |

[a] $\left[\frac{n\,(N332A)sort}{n\,(N332A)input}\right] / \left[\frac{n\,(Non-N332A)sort}{n\,(Non-N332A)input}\right]$

(Exemplarily shown for N332A enrichment)

**Figure 39. Single-round FACS selection of the BG505 SOSIP gp145 glycan-knockout library with bNAb PGT135.** 48 h post induction, the BG505 SOSIP gp145 glycan-knockout library was subjected to one round of FACS selection with PGT135 as screening antibody. Cells were sorted for the lowest PGT135 signal in relation to Env surface expression. After sorting, gDNA of input control (drawn prior to sorting) and sorted cells was recovered and analyzed by NGS. The frequencies of alanine substitution variants in input control and sorted cells in % are shown in **(A)**. All variants which were not present in the library are designated as not available (N/A). **(B)** Enrichment of stable cell lines in the sorted population. Enrichment is represented as $\log_2$-fold change in the number (n) of alanine substitutions for each variant compared to the input control. Calculations were made according to the formula from Bruun et al.[154].

Besides, NGS was used to analyze the variability of the cellular library in the initial stage (frozen library stock), after cultivation for 18 days (time to get sufficient cell numbers for sorting), and further, after cultivation for additional 2 days in presence of doxycycline (induction of Env expression). The results shown in **Appendix Figure 7** demonstrated a varying distribution between library stock and cultured/induced cells. Importantly, however, all cell lines present in the library stock were also present in the input material which was subjected to sorting.

Taken together, the previously described FACS selection system in combination with NGS was successfully used to enrich and identify PGT135 low-affinity variants out of 54-member BG505 SOSIP gp145 glycan knockout library.

### 4.3.5 Summary

Recently, our group reported a mammalian cell display and FACS-based methodology that allows selection of Env variants according to their antibody affinity. Thereby, the applicability of the method was demonstrated using a cell-surface expressed five-member Env/V3 chimeric model library[154]. One objective of this thesis was to apply the selection procedure to a cellular library that can be used to investigate the impact of single glycan deletions on antibody binding. To this end, the BG505 SOSIP gp160 alanine substitution library, described in **4.2.2**, was used as template for the generation of a glycan-knockout library, comprising plasmids that encode for alanine substitutions in N-linked and O-linked glycosylation sites of BG505 SOSIP, see **4.3.3**. Since comparative analyses demonstrated that truncation of the CT enhanced Env expression levels without negative effects on the favorable antigenicity profile of membrane-exposed BG505 SOSIP (see **4.3.2**), variants of the BG505 SOSIP glycan-knockout library were cloned to express gp145 Env. Subsequently, constructs were stably transfected in constitutive Furin overexpressing CMV F+ cells to generate stable cell lines. Stable cell lines were individually tested for Env expression and pooled to generate the cellular BG505 SOSIP gp145 glycan-knockout library, see **4.3.3**. After successful generation, the cellular library was applied to FACS selection using the N332 glycan-dependent bNAb PGT135 as screening antibody. Thereby, FACS in combination with NGS was successfully used to enrich PGT135 low-affinity variants, see **4.3.4**.

# 5. Discussion

A protective HIV vaccine could essentially contribute to end the HIV-1 pandemic. During the last years, the field focused on rational Env vaccine design with the aim to elicit bNAbs, targeting vulnerable sites of the HIV Env protein. Although induction of bNAbs by active immunization remained unsuccessful so far, major progress has been made. Since 2009, new screening procedures and technologies led to the identification of a huge number of bNAbs (see **1.4.2**) with some of them defining new vulnerable sites on the highly variable and heavily glycosylated Env protein (see **1.3**). A hallmark in HIV-vaccine design was the development of the soluble, stabilized native-like BG505 SOSIP.664 trimers, displaying a favorable antigenicity profile in terms of binding to bNAbs but no or reduced binding to non-neutralizing antibodies, as determined by several biochemical and biophysical assays (see **1.5.1**). These trimers essentially contributed to the identification/characterization of many of the newly identified bNAbs (see **1.4.2**), and further gave valuable insights into the highly dynamic, metastable nature of the Env protein (see **1.3**). Within the last few years, binding characteristics of bNAbs and structural knowledge were used to generate a multitude of stabilized Env immunogens (see **1.5.1**). When tested in animal models, several stabilized immunogens (including BG505 SOSIP.664) induced autologous neutralizing antibody responses, considered as first success on the road towards broader neutralizing antibody responses. Based on these promising results, efforts have been made to improve the quality and breadth of antibody responses, including particle-based delivery of (stabilized) Env immunogens (see **1.5.3**). However, as particle immobilization of Env immunogens may affect their antigenicity profile, the latter has to be investigated by suitable methods. Furthermore, analysis of the antigenicity profile of Env immunogens, soluble or particle-presented, should be ideally performed under physiological conditions as the environment may affect their structural integrity. MicroScale Thermophoresis represents a relatively new and powerful technique which offers the ability to measure affinities in solution and under physiological conditions (see **1.6**). Thereby, the method is in principle not limited in respect to molecule size. As MST offers great opportunities for the investigation of the Env antigenicity profile, one objective of this thesis was to test whether MST is a useful technique for the characterization of soluble and particle-presented Env proteins in buffer and under physiological conditions. In addition, this thesis focused on the development of a BG505 SOSIP-based alanine substitution library for functional mapping of bNAb epitopes using

mammalian cell display. A third objective of this thesis was the generation of a BG505 SOSIP gp145 glycan knockout library to investigate antibody-glycan interactions via a previously described cell sorting based approach[154].

## 5.1 MicroScale Thermophoresis and ELISA analysis of Env proteins

One objective of this thesis was to investigate if MicroScale Thermophoresis is a useful technique for the characterization of soluble and particle-presented Env proteins in buffer and under physiological conditions. A model system consisting of five chimeric Env variants (Env/V3 chimeras) with distinct binding capacities to mAb 447-52D was used to address this question step-by-step. Experiments with soluble (gp120 monomers/gp140 trimers) and VLP-presented Env/V3 chimeras have shown that MST in general can meet the criteria mentioned above, albeit with the minor restrictions that measurement in presence of 100 % serum was not possible. However, MST tolerated presence of 50 % serum, and thus enabled the investigation Env-antibody interactions under near-physiological conditions. Noteworthy, MST results were in general agreement with ELISA data that were collected in parallel, although measured affinities were only comparable within the respective method used. Further, MST analyses of the interactions between BG505 SOSIP.664 trimers and a panel of quaternary structure-characterizing mAbs demonstrated that the favorable antigenicity profile of BG505 SOSIP.664 trimers was preserved in MST measurements in buffer, as validated by His-capture ELISA. From this starting point and with the ability to measure affinities in presence of 50 % serum, MST was successfully used to investigate the antigenicity profile of BG505 SOSIP.664 trimers under near-physiological conditions. Further, MST was useful to assess the antigenicity profile of PLA particle-adsorbed BG505.SOSIP.664 trimers.

### 5.1.1 Interactions between Env/V3 chimeras and mAb 447-52D in buffer and 50 % human serum

MST analyses within the Env/V3 model system in buffer and 50 % serum demonstrated that MST provided consistent data within the overall expected affinity ranking (MN≥CDC42/HXB2>RF>SF33) of the interactions between gp120, gp140 and VLP-presented gp145 Env/V3 chimeras and 447-52D, as confirmed by ELISA, see **Figure 23** and **Figure 27**. Interestingly, it was observed that mAb 447-52D bound with higher overall affinities to trimeric gp140 Env/V3 chimeras than to respective gp120 versions of the

proteins. This was for the case for both, ELISA and MST analyses, in buffer as well as in presence of 50 % serum, see **Table 8** and **Table 9**. 447-52D targets a linear conformational epitope in the crown of the V3 loop of gp120[202,160,188] meaning that the epitope this antibody is three-times present within an Env trimer. Thus, one could consider that avidity effects, mediated by bivalent binding of the antibody within an Env trimer, were responsible for the enhanced affinity of 447-52D towards trimeric gp140 Env/V3 chimeras. However, this is an unlikely explanation since the architecture of Env trimers in general impedes intra-spike crosslinking[211,212]. A possible reason for the enhanced binding to trimeric Env/V3 chimeras may be that gp140 trimers seemed to be more stable than their monomeric counterparts. The instability of gp120 monomers was already observed during the protein purification process. Here, gp120 monomers did not tolerate addition of a His-tag and were highly sensitive to further concentration after SEC purification (both resulted in aggregation of the proteins), which was not observed for gp140 trimers. Thus, the instability of gp120 monomers may have negatively affected the presentation of the linear, but conformationally influenced 447-52D epitope, which may explain the reduced overall binding affinity of 447-52D to gp120 Env/V3 monomers.

As mentioned in the summary of the analysis of Env/V3 chimeras in buffer, it was waived to directly compare affinity data of VLPs measurements with the corresponding affinity data of soluble gp120 and gp140 Env/V3 chimeras since different assay setups were used for their analysis in both, MST and ELISA. For example, whereas 447-52D was labeled for the analysis of soluble Env/V3 chimeras, an inversed setup was used for the analysis of VLPs. Since fluorescent labeling has the potential to influence either the binding affinity of 447-52D or the 447-52D epitope present on labeled VLPs, a valid comparison of overall binding affinities is only possible within the respective assay setup. This holds also true for ELISA were soluble Env/V3 chimeras were analyzed by lectin capture and VLPs by direct-ELISA. In conclusion, comparing absolute affinities between VLP-presented and soluble proteins would require that the affinities are determined in within one assay setup. For ELISA, one could try if VLPs can be captured by lectin. In case of MST, a conceivable scenario would be to use Env proteins (soluble and VLP-presented) as fluorescent-labeled interaction partners and use antibodies as titrants. The fluorescent labeling of soluble and VLP-presented Env could be achieved by adding suitable tags (e.g. Flag- or Avi-tag) that are accessible in both, VLP-presented and soluble Env proteins. This would allow site-directed labeling of Env proteins, and thus avoid disturbance of antibody epitopes as it might be the case in unspecific labeling via NHS-chemistry. However, a C-terminal tag, like it is often used for the purification of

soluble proteins, is not an option for VLP-presented Env proteins since the C-terminus in this case is buried within the particle, and thus no accessible for antibody-directed labeling. Interestingly, flow cytometry analyses previously performed in our group have shown that is possible to express BG505 SOSIP gp145 Env proteins with a N-terminal Flag-tag on the surface of mammalian cells. Further, analyses of the antigenicity profile, assessed by flow cytometry-based titration experiments with a panel of quaternary structure-characterizing antibodies revealed that the N-terminal Flag-tag had no negative effects on the protein structure. In conclusion, expression of Env proteins with N-terminal tags would allow site directed labeling of soluble and VLP-presented Env proteins and thus, offers an option to analyze soluble and VLP-presented Env proteins within one assay setup.

Comparison of MST-derived $K_D$-values demonstrated that the presence of serum during MST analyses seemed to reduce the overall binding affinity of 447-52D to soluble monomeric gp120 and trimeric gp140 Env/V3 chimeras. However, this was not the case in MST analyses with VLPs, in which binding affinities between buffer and serum analyses were comparable. Interestingly, serum effects were not observed in ELISA analyses (compare $K_D$-values of MST and ELISA analyses in buffer and serum in **Table 8** and **Table 9**, respectively). As described in **1.6**, Env proteins and in particular Env proteins of subtype B, are susceptible to serum proteases-mediated proteolytical damage of the V3 region (V3-clipping). Since all V3-loops of Env/V3 chimeras were derived from subytype B-isolates[213,214,215,216,217], V3 clipping has to be considered as a factor that may have contributed to the reduced binding of 447-52D, because this antibody binds a V3 epitope. However, V3-clipping of soluble Env/V3 chimeras by serum proteases seems to be an unlikely reason for the MST-measured reduced binding affinity of 447-52D for several reasons: 1) the reduced affinity of 447-52D was only detected for soluble proteins and not for VLP-presented Env/V3 chimeras, although one would expect clipping also for VLPs; 2) it is assumed that V3-clipping would have also reduced the binding affinities of 447-52D in ELISA analyses. A possible explanation for the reduced affinity of 447-52D could be that soluble Env/V3 chimeras, gp120 monomers as well as gp140 trimers, are less stable in the presence of serum. Similar, as discussed above for gp120 monomers, this may have negatively influenced the linear conformational epitope, and thus the binding affinity of 447-52D. Interestingly, the reduced binding affinity was only observed in MST analysis and there may be (a combination of) two reasons why this was the case: 1) MST, based on the thermophorectic effect which is influenced by multiple parameters (size, charge, hydration shell or conformation of a molecule[150,149]) is exceptional sensitive; and 2) that serum was present during the measurement itself, which was not the case in ELISA analyses.

Here, serum was only present during the incubation of the primary Ab (447-52D), while the final detection was done after several washing steps. However, the question why VLPs in MST analyses were not affected by presence of serum remains elusive. It is known that the transmembrane domain and the cytoplasmic tail contribute to the stability of membrane-exposed Env proteins which may explain why truncated soluble Env proteins without stabilizing modifications are unstable and conformationally heterogeneous[218,219]. In line with the arguments that instability of soluble Env/V3 chimeras probably contributed to the reduced affinity of 447-52D in MST serum analyses, one could consider that Env/V3 chimeras were more stable when they are presented in the membrane exposed context, and thus not affected by presence of serum. However, to what extent this may have contributed to the experimental outcome of MST analyses is unknown since VLPs that expose C-terminal tail truncated Env/V3 chimeras are a highly artificial system. In summary, the Env/V3 system represents a valuable tool for assay establishment, but since it is a highly artificial system it is difficult to draw final conclusions how presence of serum affected antibody binding in this case. This was one reason why MST experiments were expanded to the well-characterized, native-like BG505 SOSIP.664 trimers as their structure can be nicely analyzed by quaternary structure-characterizing antibodies.

### 5.1.2 Interactions between BG505 SOSIP.664 trimers and quaternary structure-characterizing antibodies in buffer and 50 % human serum

With the intention to investigate a relevant next-generation Env immunogen, MST was used to analyze the interaction of native-like BG505 SOSIP.664 trimers and set of quaternary structure-characterizing mAbs in buffer and 50 % human serum. Thereby, MST-derived data were compared to results from direct and His-capture ELISAs. Measurements in buffer demonstrated that the native-like structure of the BG505 SOSIP.664 trimers was preserved in MST analysis since all bNAbs (PGT145, PGT151, VRC01, 2G12) bound with high affinities, whereas non-neutralizing mAb F105 exhibited only marginal binding. His-capture ELISAs produced comparable results and thereby validated MST-derived data, see **Figure 29**. Importantly, comparing results of MST and His-capture ELISA with data obtained from direct-ELISA nicely demonstrated that direct coating of BG505 SOSIP.664 trimers destroys the native-like structure of the protein, which has been previously reported by Schiffner *et al.*[125]. Noteworthy, His- and direct ELISA results measured within this thesis were general in good agreement with ELISA data reported in this publication. Further, MST was used to

investigate if the presence of serum affects the antigenicity profile of BG505 SOSIP.664 trimers. Results showed that the antigenicity profile of BG505 SOSIP.664 trimers was in general agreement to measurements in buffer, with the exception that saturated binding of non-neutralizing antibody F105 was detected, see **Figure 29**. This indicated that in presence of serum at least a small proportion of the trimers turned into a more open conformation. However, this was not detected in His-capture ELISA, likely for reasons already described in **5.1.1** (high sensitivity of MST, presence of serum during analysis). In addition, the results obtained from the MST analyses of BG505 SOSIP.664 trimers nicely demonstrated that presence of serum can affect the stability of Env proteins as it was discussed for soluble Env/V3 chimeras. The reasons why BG505 SOSIP.664 trimers partially turned into an open conformation, as indicated by binding of mAb F105, remain elusive. V3-clipping can be excluded since the V3 sequence of BG505 is not particularly vulnerable to this kind of proteolytical damage[220].

### 5.1.3 Interactions between PLA-adsorbed BG505 SOSIP.664 trimers and quaternary structure-characterizing antibodies in buffer and 50 % human serum

PLA particles with adsorbed BG505 SOSIP.664 trimers were generated in collaboration with Bernard Verrier and analyzed by MST and ELISA in order to test if MST is useful to assess the antigenicity profile of native-like Env trimers presented in the context of a state-of-the art particle-based vaccine delivery system. Contrary to MST analyses with VLPs, measurements of PLA particles suffered from background binding of the mAbs to control particles (without adsorbed Env), see **Appendix Figure 3**. This can be explained by passive adsorption of mAbs to the positively charged surface of PLA particles, which also enabled adsorption of BG505 SOSIP.664 trimers. Notably, MST analyses of control PLA particles did benefit from the presence of 50 % serum, since serum components reduced unspecific binding of mAbs, most likely by blocking the particle surface, see **Appendix Figure 4**. In summary, it could not be excluded that unspecific binding, at least partially, contributed to the binding of mAbs to PLA particles with adsorbed BG505 SOSIP.664 trimers in MST analyses, since the calculated amount of ~50 trimers per particle (175 nm in size) is likely not sufficient block the complete particle surface from unspecific adsorption of mAbs. However, binding of all mAbs was considered as specific since saturated binding curves were only observed for PLA particles with adsorbed BG505 SOSIP.664 trimers, see **Appendix Figure 3**, **Appendix Figure 4**. MST analyses of PLA-adsorbed BG505 SOSIP.664 trimers in buffer and 50 % serum revealed

binding of bNAbs including quaternary structure-depending PGT145/PGT151, and binding of non-neutralizing antibody F105, see **Figure 30**. Consequently, MST results suggested that PLA-adsorbed BG505 SOSIP.664 trimers were present on particles surfaces as a mixture of native-like trimers and trimers with disturbed quaternary structure. It is likely that the mixed trimer population resulted from the immobilization strategy used for particle decoration (passive adsorption) since direct-ELISA experiments with BG505 SOSIP.664 trimers have shown that passive adsorption negatively influenced the protein structure, see **Figure 29**. In order to validate MST data, direct-ELISAs with PLA-adsorbed BG505 SOSIP trimers were performed with the assumption that PLA particles would serve as a platform for the presentation of BG505 SOSIP.664 trimers. Thus, it was expected to detect an antigenicity profile comparable to MST results, including binding of all bNAbs and non-neutralizing antibody F105. However, results of the direct-ELISA, contrary to MST results, indicated that the native-like structure of the BG505 SOSIP.664 trimers was completely disturbed. Whereas binding of bNAbs VRC01 and 2G12 (no quaternary structure dependency) and also non-neutralizing F105 was detected, binding of quaternary structure-depending PGT145 was completely lost and PGT151 demonstrated some marginal signal limited to the highest antibody concentration used, see **Figure 30**. Taken together, results indicated the disruption of both, PLA particles and BG505 SOSIP.664 trimers due to coating to ELISA plates.

### 5.1.4 Conclusion and potential applications for MST

In summary, experiments have demonstrated that MST represents useful tool for the analysis of (stabilized) Env proteins for various reasons:

1) MST is capable to determine affinities of interactions, starting from very simple to complex ligand presentations, as it was demonstrated in an exemplary way for soluble and VLP-presented Env/V3 chimeras and also in measurements with BG505 SOSIP.664 trimers in soluble or PLA-adsorbed state. This, in principle, allows the antigenic characterization of soluble Env vaccines or particle-presented Env vaccines, using a single technology. This is from particular interest for particle-presented Env vaccines, as it became clear that they are more effective in B-cell stimulation compared to soluble versions of the protein[221,222,143]. However inherent properties of particles and the experimental design strongly influence if it is possible to determine binding affinities (like in case of VLPs) or if the measurement is restricted to allow characterization of the antigenicity profile by comparing antibody binding profiles

rather than by precise determination of binding affinities, like in case of PLA-adsorbed BG505 SOSIP.664 trimers.

2) MST is in principle applicable to nearly any kind of buffer, and thus offers the option to investigate Env-antibody interactions under near-physiological conditions as it was shown for experiments in the presence of 50 % human serum. This is of particular interest for HIV vaccine design, since the antigenic profile of Env vaccine candidates is, as state-of-the-art, mostly assessed in artificial buffer and not under near-physiological conditions that could affect the antigenicity profile as demonstrated in the analysis of the antigenicity profile of BG505 SOSIP.664 trimers in the presence of serum. MST thereby offers the advantage that binding of antibodies can be directly detected in presence of serum and not indirectly as it was the case in ELISA experiments. A nice experiment would be to incubate BG505 SOSIP.664 trimers or other stabilized Env immunogens in 20 % human serum as surrogate for lymphatic fluid (lymphatic fluid generally exhibits lower protein content than complete serum[222]) and analyze the antigenicity profile over different periods of time. This would maybe give a hint to what extent the antigenicity profile is already affected by serum components when proteins reach B-cells in the draining lymph nodes (parentally injected proteins are typically detectable in lymph nodes within minutes to a few hours[223]). Another possible application for MST would be to analyze how different kinds of immunostimulatory adjuvants influence the antigenic profile of next generation Env immunogens since only few studies addressed this issue in detail[191,140].

3) The main advantage of MST over ELISA and other methods which depend on surface-immobilization of one interaction partner is that measurement takes place in solution. This might be important since immobilization of one interaction partner might interfere with the binding event or negatively influence the structure of the immobilized protein, like it was demonstrated in direct-ELISA experiments with BG505 SOSIP.664 trimers. Further, as it was assumed for direct-ELISA experiments with PLA-adsorbed BG505 SOSIP.664 trimers, immobilization of particles may have resulted in particle-disruption. Since analysis takes place in solution, MST avoids these kinds of problems from the outset. However, it is worth mentioning that also other methods can be used for analyses of Env-carrying particles. BLI for example was successfully used to characterize the Env-carrying liposomes[143] mentioned in **1.5.3**. In general, surface immobilization of particles by direct coating or capturing

approaches may be not applicable to every type of particle, thus MST could represent a more universal approach to analyze interactions between particle-presented proteins and interaction partners.

4) Other benefits of MST are fast assay development, short hands-on time as well as rapid measurements. Measurement of monomers, trimers VLPs and PLAs in buffer and serum were performed under comparable conditions with only minor adjustments (except for VLPs, were an inverse assay setup was used), see **3.5**. Furthermore, in comparison to ELISA, MST consumed only small amounts of sample material and was faster, see **Appendix Table 1**.

In conclusion, MST combines the advantages of low sample consumption, fast assay development and fast readout with high sensitivity and the possibility to analyze interactions in artificial as well as near-physiological environments with only minor adjustments in the assay conditions. Therefore, MST is considered as a valuable tool, not only for the analysis of soluble or particle-presented Env immunogens, but as a universal method to investigate the antigenicity profiles of also other immunogens like influenza hemagglutinin or ebola glycoprotein.

## 5.2 Generation of a BG505 SOSIP alanine substitution library for cell-based antibody epitope mapping

Since 2009, new screening procedures and technologies, including high-throughput neutralization screenings of single B-cell cultures and antibody isolation by antigen-specific single B-cell FACS, led to the identification of a huge number of bNAbs[44]. Such antibodies provide insight into the development of HIV specific B cell responses and at the same time represent valuable tools for rational immunogen design. For both, the precise mapping of bNAb binding specificities is mandatory. This thesis focused on the development of an alanine substitution library that should enable mapping of bNAbs with quaternary epitopes, as this was only possible to a limited extent with the already existing 96ZM651 gp145 alanine library (see **4.2**). The cleavage deficiency of the 96ZM651 gp145 library, likely resulting in the expression of trimers with perturbed quaternary structure, was considered mainly responsible that mapping of highly quaternary structure-depending bNAbs remained unsuccessful[197]. Hence, cleavage competence was a prerequisite of the new library which was generated based on the cleavage-competent BG505 SOSIP gp160, see **4.2**.

### 5.2.1 Antibody epitope mapping with the BG505 SOSIP gp160 alanine substitution library

The in-depth characterization of the BG505 SOSIP gp160 alanine substitution library, including proof-of-principle screenings, was performed by Iris Ganser during her master`s thesis[201]. Here, proof-of-principle screenings with bNAbs HGN194, VRC01 and PGT121 identified alanine substitutions that mainly resulted in losses of binding in regions that had been described as their respective epitopes (exemplarily shown for HGN194, see **Figure 33**) demonstrating that the library in principle is a suitable tool for epitope mapping. However, epitope mapping using the BG505 SOSIP gp160 alanine substitution library faced several challenges including the low surface expression levels of BG505 SOSIP gp160. Env expression levels, next to other factors such as brightness of the fluorescent dye, antibody labeling efficiency and antibody-Env binding affinity, essentially contribute to the flow cytometer`s limit of detection. The interplay of low Env expression levels with the factors described above could have been responsible that in some cases signals were too low to distinguish small differences in binding affinities, thereby impeding mapping quality and reliability. However, there are several options to improve signals and thereby the quality and reliability of future mapping experiments as discussed in detail in the master`s thesis[201].

Noteworthy, the truncation of the Env CT offers a very promising approach for signal enhancement by improving Env expression levels, see **Figure 31**. The truncation of the CT was initially avoided in the library design when a study reported that CT truncation can negatively influence the Env structure[145]. However, during this thesis it turned out that this seemed to be not the case in the context of membrane-exposed BG505 SOSIP, see **Figure 36**. A reason for this could be that SOSIP modifications can provide stabilizing effects that counteract potential destabilizing effects resulting from CT truncation. However, if this theory holds true has to be tested. The comparison of the antigenicity profile of BG505 WT gp145 and gp160 without stabilizing modifications would be a likely starting point to address this question. Taken together, in case of the BG505 SOSIP alanine substitution library, the truncation of the Env CT represents a very likely starting point for future improvements since experiments with the previously described 96ZM651 gp145 library demonstrated that Env CT truncation provides sufficient expression levels for robust epitope mapping[197]. Notably, due to the library design (see **4.2.2**), the conversion of the gp160 library into a gp145 library can be achieved rather quickly. The lack of suitable method for referencing Env surface expression levels represented another major issue in mapping experiments with the BG505 SOSIP gp160 alanine scanning library. Signal referencing is considered crucial to obtain reliable and high-quality mapping results as it offers the advantage to normalize binding signals of the screening antibody for varying expression levels between different library variants. Proof-of-principle screenings (e.g. HGN194, see **Figure 33**) with the BG505 SOSIP gp160 alanine library were conducted in multiplex approach, allowing for a direct comparison between referenced and unreferenced measurements, see **4.2.3**. Thereby, the MPER targeting bNAb 4E10 was chosen as reference antibody as the analyzed subset of library variants did not contain alanine substitutions within this region. Hence, it was assumed that the antibody binds comparably to all variants of the subset. At this point it has to be mentioned that it was initially planned to apply a co-staining strategy as soon as suitable reference was identified since this approach was successfully used in mapping experiments with the 96ZM651 gp145 library[197], (Kliche and Grassmann *et al*., manuscript in preparation). Therefore, gp41 antibodies were preferred over gp120 targeting reference antibodies, simply for the reason that binding interference is more likely when the reference antibody targets the same Env subunit as the screening antibody. However, while unreferenced HGN194 measurements detected losses of binding in the V3 loop, representing the HGN194 epitope, 4E10 referenced measurements resulted in loss of HGN194 binding for almost all screened variants, irrespective of V3 loop or control residues. This resulted from the fact that 4E10 exhibited

enhanced binding to nearly all alanine substitution variants compared to the WT, meaning that 4E10 is not suitable for surface referencing. Notably, bNAbs 2G12 and VRC01 have been proven useful for quantifying Env surface expression levels in case of the cellular BG505 SOSIP gp145 glycan-knockout library, see **Figure 37**. Thus, these antibodies could be considered as an option for referencing Env expression levels in future mapping experiments. However, when used in a co-staining approach (e.g. Alexa488-labeled screening Ab and Alexa647-labeled reference Ab) binding interference has to be tested. Otherwise, library screening has to be conducted in a multiplex approach, see **4.2.3**. Another option that may help to solve the referencing problem would be the addition of an N-terminal Flag-tag to the library variants. Experiments previously performed in group already demonstrated the applicability of this strategy, see **5.1.1**. The expression of Flag-tagged Env would allow quantification of Env expression levels via an anti-Flag antibody. This setup would be beneficial since the Flag-tag is not affected by alanine substitutions, which might be the case when an Env-targeting antibody is used for referencing. Further, this setup likely excludes binding interferences with the screening antibody. However this strategy would require re-design and re-cloning of the library, but has to be considered for future library approaches.

### 5.2.2 Conclusion

In summary, stabilizing SOSIP modifications were successfully implemented into membrane-exposed BG505 and a BG505 SOSIP gp160 alanine substitution library was generated (see **4.2.2**). Initial proof-of-principle mapping experiments with antibodies HGN194, VRC01 and PGT121 demonstrated that the library is in principle a suitable tool for epitope mapping. However, the quality of mapping experiments suffered from low Env expression levels and the lack of a suitable method for referencing (see **4.2.3)**. Notably, there are options to overcome these limitations in future mapping experiments. These include the possibility to convert the gp160 into a gp145 expressing library using the cloning features of the library (see **4.2.2**). This strategy is recommended as titration experiments with BG505 SOSIP trimers demonstrated that their structural integrity does not depend on the presence of the Env CT (see **4.3.2**). In addition, bNAbs 2G12 and VRC01 represent likely candidates to address the problem of surface referencing since they were successfully used to quantify Env expression levels in the cellular BG505 SOSIP gp145 glycan-knockout library approach (see **4.3.3**). In conclusion, enhanced Env surface expression levels together with a robust method for surface referencing have the potential to greatly improve the overall mapping quality. Such improvements also represent first steps towards the mapping of quaternary structure

depending antibodies (e.g. PGT145), which remained unsuccessful under the given conditions. In conclusion, future experiments under optimized conditions are required to explore the full potential of the library.

## 5.3 High-throughput analysis of a cellular BG505 SOSIP gp145 glycan-knockout library by FACS in combination with NGS

Recently, our group reported a mammalian cell surface display and FACS based methodology that enables the selection of Env variants with improved or reduced affinity towards a screening antibody[154]. The methodology involves a cellular library of stable cell lines with inducible Env expression, providing advantages such as mammalian glycosylation and presentation of Env proteins in their natural membrane-exposed context. For selection the library is applied to FACS, thereby, Env expressing cells are selected towards their affinity to a screening antibody and subsequently identified using NGS. This procedure was successfully utilized to enrich low and high affinity Env variants from a five-member Env/V3 chimeric model library[154], demonstrating its opportunities for selection and analysis of Env immunogens. Within last years the analysis of the Env glycan shield (see **1.3.3**) became increasingly important in HIV-1 Env vaccine design since structural analyses revealed that a proportion of bNAbs (see **1.4.2**) does not only bind to gp120 protein residues but also recognizes glycans attached to gp120[206,176,207,180,208,209]. Hence, the third objective of this thesis focused to apply the procedure described above to a library that can be used to analyze the impact of single glycan deletions on antibody binding. For this purpose a 54-member cellular BG505 SOSIP gp145 glycan-knockout library was generated and subjected to screening with bNAb PGT135[178,180,106] recognizing a V3-glycan epitope. The initial screening, aiming to select for Env variants with low PGT135 affinity, demonstrated the N332 glycan essentially contributes to the binding of this antibody, as previously described[180,106].

### 5.3.1 Advantages of the applied stable cell line based mammalian cell display system and comparison to the previously reported approach

The mammalian cell surface display which was previously reported by our group provides several advantages over other currently used methods. Contrary to yeast[224,225] or phage[226] display approaches, it enables the presentation Env trimers in their natural membrane-exposed environment, and further provides mammalian glycosylation[154]. In addition, the Flp-In™ T-REx™ system supports the integration of a single copy *env* gene at a pre-defined locus. This provides an advantage over stable cell lines that were generated by random integration

methods (e.g. lentiviral transduction) since it reduces the risk of varying expression levels between different cell lines that may result from the genomic context of the integration locus. In addition, Flp-In™ T-REx™ 293 cells provide an inducible Env expression via the Tet operator/repressor system, eliminating Env cytotoxicity effects during stable cell line generation and cultivation. The stable cell line based BG505 SOSIP glycan-knockout library that was developed during this thesis benefits from all advantages described above. However, there were two major differences between the previously described and the newly developed library system. First, although proven useful to normalize Env expression levels in the proof-of concept study with the five-member Env/V3 library, GFP expression was excluded as normalization method in the BG505 glycan-knockout library system. Here, a reference antibody was used to normalize Env expression levels (see **4.3.4**). This option was chosen for two reasons: 1) although genetically coupled to GFP expression, the amount of GFP remaining in the cytoplasm has not necessarily to correlate with the Env level on the cell surface as both proteins may deviate in their turnover; and 2) it was considered to avoid the risk of a potential insufficient TaV 2A "cleavage" which would result in GFP-Env fusion proteins with likely disturbed Env quaternary structure. The second major difference regards the cell line that was used for the generation of stable cell lines. In the proof-of principle study with the Env/V3 chimera library commercially available Flp-In™ T-REx™ 293 cells were used for generation of stable cell lines. However, the endogenous Furin expression of mammalian cells including HEK293T cells is not sufficient to confer proper protein cleavage when Env proteins are expressed at high levels[227,228]. This was no issue in the proof-of-principle study with Env/V3 chimeras but considered as a major obstacle for a library that was designed to express fully-cleaved Env with intact quaternary structure. To avoid limitations in Env cleavage that could potentially arise from insufficient endogenous Furin expression a Furin overexpressing cell line, termed CMV F+, was used for stable cell line generation in case of the BG505 SOSIP gp145 glycan knockout library. This cell line was generated based on Flp-In™ T-REx™ 293 cells and constitutively expresses Furin under a control of CMV promotor, thereby providing all advantages of normal Flp-In™ T-REx™ 293 cells. The advantages of this cell line over normal Flp-In™ T-REx™ 293 cells was impressively demonstrated in the master`s thesis of Dominik Fiegle during which CMV F+ cells were developed. Here, CMV F+ cells demonstrated expression of fully-cleaved BG505 SOSIP.664 while parental Flp-In™ T-REx™ 293 cells achieved only partial protein cleavage[165]. In a comparative analysis BG505 SOSIP gp145 and gp160 expressed by CMV F+ cells were exclusively bound by bNAbs and not by non-neutralizing antibodies, see **Figure 36**. This

demonstrates the capacity of the CMV F+ cells to express fully-cleaved Env with intact quaternary structure, assuming that partial cleavage would have resulted in detectable binding of non-neutralizing antibodies.

### 5.3.2 FACS of the BG505 SOSIP gp145 glycan-knockout library with bNAb PGT135

An initial FACS experiment with the BG505 SOSIP glycan-knockout library and bNAb PGT135 demonstrated that the previously described mammalian cell display and FACS-based selection procedure[154] can be successfully used for the analysis of bNAb glycan-dependencies. The screening procedure successfully identified Env variant N332A as crucial for PGT135 binding, since this variant accounted for approximately for 82 % of the cells in the sorted population, which corresponds to 35-fold enrichment by the sorting procedure (see **Figure 39**). Interestingly, Env variant T334A which also prevents the glycosylation of the asparagine at position 332 was only 2-fold enriched in the sorted population. This is an interesting observation since it indicates that not only N332 glycan residues essentially contribute to PGT135 binding, as previously described[180], but also the asparagine itself. However, this observation has to be investigated in detail by further experiments, including PGT135 titration experiments with the respective cell lines. Notably, next to several protein contact residues, three additional glycans at AA positions 295, 386 and 392 are associated with binding of PGT135[180]. However, the FACS selection did not result in a considerable fold-enrichment for the respective glycan-knockout variants. The stringent gating strategy might be an explanation for this observation since there were still several events protruding from the main population towards the low affinity gate (see **Figure 38**) which were not sorted in this first analysis. It may be that this "intermediate" population included glycan-knockout variants for the glycan positions described above. This theory is supported by a recent publication in which capture-ELISA was used to analyze the interaction between PGT135 and soluble glycan-knockout variants[229]. Here, it has been demonstrated that the alanine substitution of the asparagine at AA position 392 also reduces binding of PGT135, however, not as drastically as the N332A variant. Further, it has been shown that the N386A glycan deletion only slightly decreased PGT135 binding, whereas the N295A variant exhibited WT-like binding. Taken together, the results that were reported in this publication may partially explain the outcomes of the PGT135 sorting experiment which successfully demonstrated the potential of the library for the analysis of bNAb glycan-dependencies. However, for future

experiments a less stringent gating strategy would be desirable to enable detection of Env-glycan-knockouts with less pronounced effects on antibody binding.

### 5.3.3 Future directions

The first screening experiment with the bNAb PGT135 demonstrated that the BG505 SOSIP gp145 glycan-knockout library is a useful tool for the investigation of antibody-glycan interactions, offering the option to identify glycan-dependencies of already known or newly identified bNAbs. However, there are remaining questions and potential applications that could be addressed by future experiments. These include the in-depth characterization of the relationship between glycan deletions and Env cell surface expression levels. When single cell lines were tested for Env expression most cell lines varied within a factor of 2 compared to the WT. However, some cell lines exhibited very low Env expression levels, see **Figure 37**. The deletion of the glycan at AA position 448 for example reduced Env expression close to the background level. Of course one could consider that this could be an artefact but it is interesting that both glycan knockout variants, N448A and T450A, exhibited the same phenotype. This has to be further investigated since it was reported that glycan knockouts can affect incorporation of Env into virions[230]. A first step to validate observations for the 448 glycan deletion could be the re-sequencing of genomic DNA samples to exclude mutations in the *env* sequences. In addition, quantitative real-time PCR could be used to analyze the relation between mRNA and Env surface expression levels. This would provide first insights if the reduced surface expression is associated with the transcription level or later processes in protein expression. Notably, this kind of analysis would be highly desirable not only for variants mentioned above , but for the whole library to better understand varying Env expression levels between stable cell lines.

Future applications of this library could include screenings to identify glycan-dependencies of antibodies as described for PGT135. However there are also other considerable applications. Env immunogens, including BG505 and other SOSIP.664 stabilized Env trimers react only very weakly with germline-reverted bNAbs[121,231,134,232]. For this reason they likely fail to engage B-cells that carry receptors of bNAbs precursors. The Env glycan shield seems to play an important role in this context since it masks conserved epitopes from recognition by bNAb germline precursors[233]. In conclusion, the library could be used to screen for glycan deletions that result in enhanced binding of germline bNAbs to gain knowledge for germline-targeting immunogen design. However, also worth a try, it is likely that a single glycan deletion is not sufficient to enable detectable binding of germline bNAbs. Alternatively, mature bNAbs

could be screened for enhanced binding since enhanced binding in such a case would likely result from removal of a steric clash by the respective glycan deletion. This could be an indirect hint that the removal of the glycan may contribute to access binding of the germline version of the particular antibody. There are observations that support the likelihood of such a scenario. By applying bNAb VRC01to a screening with the 96ZM651 alanine substitution library, Veronika Grassmann identified the N276A alanine substitution as variant with enhanced binding towards VRC01[197]. Similar was observed in the context of the BG505 SOSIP gp145 glycan-knockout library when stable cell lines were analyzed for their Env expression via VRC01 and 2G12. Here, VRC01 demonstrated enhanced binding to the variants that result in the knockout of the glycan at position N276 (T276A, T278A), whereas 2G12 indicated an expression below the WT reference for the respective variants, see **Figure 37**. Interestingly, the deletion of the N276 glycan restored binding of germline reverted versions of VRC01-class antibodies to the clade C Env 426c[234] and is nowadays included in germline-targeting Env immunogens[138,135]. Hence, taking observations from the VRC01 screening with the 96ZM951libary, this thesis, and results from literature in consideration, mature bNAbs can be used for sorting of glycan deletion variants with enhanced binding since potentially identified variants may be interesting for germline-targeting Env vaccine design.

In addition, the experiences and results that were generated during this thesis could be used to develop more advanced cellular libraries, including for example permutation libraries in which every amino acid is substituted by every possible natural amino acid. The feasibility of such a library approach was recently demonstrated by our group which developed and characterized a permutation library that based on the clade C 16055 Env (PhD thesis Julia Koop[164]). However, similar to the 96ZM651 library, the permutation library expressed cleavage-defective Env, excluding experiments with highly quaternary structure-depending bNAbs. Such limitations could be avoided using a stabilized, fully-cleaved Env and the Furin overexpressing CMV F+ cell line as bases for future libraries. Potential applications of such a library could for example involve the selection and development of germline-targeting immunogens. Recently, Steichen *et al*. demonstrated that cell-surface presented Env libraries in combination with FACS can be used for the design of germline targeting Env immunogens, highlighting the applicability of a FACS-based selection strategy. Here, different libraries, based on BG505 SOSIP gp140 fused to a platelet-derived growth factor receptor (PDGFR), were generated by lentiviral transgene integration and used for engineering of Env trimers with affinity to PGT121-class antibodies[210]. Contrary to the strategy that was applied by Steichen *et al*. the here described approach would also allow screenings with MPER-directed

(germline) bNAbs and does not require Furin transfection prior to the induction of the Env expression. In addition, it allows the integration of a single copy env gene at a predefined locus, minimizing effects that may result from random integration events or multiple infections that may occur when transgenes are introduced via lentiviral transduction.

### 5.3.4 Conclusion

In summary, CMV F+ cells allowing for presentation of fully-cleaved Env were used to generate a BG505 SOSIP gp145 glycan-knockout library (see **4.3.3**). Appling the library to a screening with glycan-dependent bNAb PGT135 led to the enrichment and identification of PGT135 low-affinity variants (see **4.3.4**). This demonstrated the applicability of a previously described FACS-based selection procedure that was developed with a five-member Env/V3 chimera library[154]. Although further experiments are required to characterize the glycan-knockout library in more detail it offers several options for future applications. These include mapping of bNAb glycan dependencies, identification of glycans that may play a critical role in binding of bNAb germline precursors, or to investigate the influence of glycans on the Env surface expression. In addition, the successful presentation of fully-cleaved Env with intact quaternary structure via CMV F+ cells laid the foundation for the generation of advanced cellular libraries e.g. permutation libraries of membrane-exposed stabilized trimers.

# 6. References


1. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **1,** a006841 (2011).

2. Waheed, A. A. & Freed, E. O. HIV type 1 Gag as a target for antiviral therapy. *AIDS Res. Hum. Retroviruses* **28,** 54–75 (2012).

3. Okoye, A. A. & Picker, L. J. CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunol. Rev.* **254,** 54–64 (2013).

4. D'arc, M. *et al.* Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E1343-52 (2015).

5. HIV Circulating Recombinant Forms (CRFs). Available at: https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html. (Accessed: 10th August 2018)

6. Becerra, J. C., Bildstein, L. S. & Gach, J. S. Recent Insights into the HIV/AIDS Pandemic. *Microb. cell (Graz, Austria)* **3,** 451–475 (2016).

7. Hemelaar, J. The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* **18,** 182–92 (2012).

8. UNAIDS. Global HIV & AIDS statistics — 2018 fact sheet | UNAIDS. Available at: http://www.unaids.org/en/resources/fact-sheet. (Accessed: 10th August 2018)

9. UNAIDS. Miles to go—closing gaps, breaking barriers, righting injustices | UNAIDS. Available at: http://www.unaids.org/en/resources/documents/2018/global-aids-update. (Accessed: 10th August 2018)

10. De Béthune, M.-P. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: A review of the last 20 years (1989–2009). *Antiviral Res.* **85,** 75–90 (2010).

11. Ghosn, J., Taiwo, B., Seedat, S., Autran, B. & Katlama, C. HIV. *Lancet (London, England)* **0,** (2018).

12. AIDSinfo. FDA-Approved HIV Medicines | Understanding HIV/AIDS | AIDSinfo. Available at: https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/21/58/fda-approved-hiv-medicines. (Accessed: 18th August 2018)

13. INSIGHT START Study Group, T. I. S. S. *et al.* Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *N. Engl. J. Med.* **373,** 795–807 (2015).

14. O'Connor, J. *et al.* Effect of immediate initiation of antiretroviral therapy on risk of severe bacterial infections in HIV-positive people with CD4 cell counts of more than 500 cells per μL: secondary outcome results from a randomised controlled trial. *lancet. HIV* **4,** e105–e112 (2017).

15. World Health Organization. *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach. June 2013*. (2013).

16. World Health Organization. *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach. Second edition 2016*. (2016).

17. UNAIDS. 90–90–90 - An ambitious treatment target to help end the AIDS epidemic | UNAIDS. Available at: http://www.unaids.org/en/resources/documents/2017/90-90-90. (Accessed: 18th August 2018)

18. UNAIDS. Fast-Track - Ending the AIDS epidemic by 2030 | UNAIDS. Available at: http://www.unaids.org/en/resources/documents/2014/JC2686_WAD2014report. (Accessed: 18th August 2018)

19. Medlock, J. *et al.* Effectiveness of UNAIDS targets and HIV vaccination across 127 countries. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 4017–4022 (2017).

20. Esparza, J. A brief history of the global effort to develop a preventive HIV vaccine. *Vaccine* **31,** 3502–3518 (2013).

21. Trovato, M. *et al.* HIV Vaccination: A Roadmap among Advancements and Concerns. *Int. J. Mol. Sci.* **19,** 1241 (2018).

22. Gray, G. E., Laher, F., Lazarus, E., Ensoli, B. & Corey, L. Approaches to preventative and therapeutic HIV vaccines. *Curr. Opin. Virol.* **17,** 104–109 (2016).

23. Flynn, N. M. *et al.* Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J. Infect. Dis.* **191,** 654–65 (2005).

24. Pitisuttithum, P. *et al.* 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,** 1661–71 (2006).

25. Buchbinder, S. P. *et al.* Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet (London, England)* **372,** 1881–1893 (2008).

26. Gray, G. E. *et al.* Recombinant adenovirus type 5 HIV gag/pol/nef vaccine in South Africa: unblinded, long-term follow-up of the phase 2b HVTN 503/Phambili study. *Lancet Infect. Dis.* **14,** 388–396 (2014).

27. Priddy, F. H. *et al.* Safety and Immunogenicity of a Replication-Incompetent Adenovirus Type 5 HIV-1 Clade B *gag/pol/nef* Vaccine in Healthy Adults. *Clin. Infect. Dis.* **46,** 1769–1781 (2008).

28. Day, T. A. & Kublin, J. G. Lessons learned from HIV vaccine clinical efficacy trials. *Curr. HIV Res.* **11,** 441–9 (2013).

29. Excler, J.-L., Robb, M. L. & Kim, J. H. HIV-1 vaccines: challenges and new perspectives. *Hum. Vaccin. Immunother.* **10,** 1734–46 (2014).

30. Hsu, D. C. & O'Connell, R. J. Progress in HIV vaccine development. *Hum. Vaccin. Immunother.* **13,** 1018–1030 (2017).

31. Hammer, S. M. *et al.* Efficacy Trial of a DNA/rAd5 HIV-1 Preventive Vaccine. *N. Engl. J. Med.* **369,** 2083–2092 (2013).

32. Churchyard, G. J. *et al.* A phase IIA randomized clinical trial of a multiclade HIV-1 DNA prime followed by a multiclade rAd5 HIV-1 vaccine boost in healthy adults (HVTN204). *PLoS One* **6,** e21225 (2011).

33. Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* **361,** 2209–20 (2009).

34. Nitayaphan, S. *et al.* Safety and immunogenicity of an HIV subtype B and E prime-boost vaccine combination in HIV-negative Thai adults. *J. Infect. Dis.* **190,** 702–6 (2004).

35. Robb, M. L. *et al.* Risk behaviour and time as covariates for efficacy of the HIV vaccine regimen ALVAC-HIV (vCP1521) and AIDSVAX B/E: a post-hoc analysis of the Thai phase 3 efficacy trial RV 144. *Lancet. Infect. Dis.* **12,** 531–7 (2012).

36. Tomaras, G. D. & Plotkin, S. A. Complex immune correlates of protection in HIV-1 vaccine efficacy trials. *Immunol. Rev.* **275,** 245–261 (2017).

37. Corey, L. *et al.* Immune correlates of vaccine protection against HIV-1 acquisition. *Sci. Transl. Med.* **7,** 310rv7 (2015).

38. Li, S. S. *et al.* FCGR2C polymorphisms associate with HIV-1 vaccine protection in RV144 trial. *J. Clin. Invest.* **124,** 3879–90 (2014).

39. Tay, M. Z. *et al.* Antibody-Mediated Internalization of Infectious HIV-1 Virions Differs among Antibody Isotypes and Subclasses. *PLoS Pathog.* **12,** e1005817 (2016).

40. Lin, L. *et al.* COMPASS identifies T-cell subsets correlated with clinical outcomes. *Nat. Biotechnol.* **33,** 610–6 (2015).

41. Haynes, B. F. *et al.* Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N. Engl. J. Med.* **366,** 1275–86 (2012).

42. Tomaras, G. D. *et al.* Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 9019–24 (2013).

43. Harper, K. N. HVTN100 phase 1/2 vaccine trial results promising; phase 2b/3 trial to commence. *AIDS* **31,** N1–N2 (2017).

44. McCoy, L. E. & Burton, D. R. Identification and specificity of broadly neutralizing antibodies against HIV. *Immunol. Rev.* **275,** 11–20 (2017).

45. Guttman, M. & Lee, K. K. A functional interaction between gp41 and gp120 is observed for monomeric but not oligomeric, uncleaved HIV-1 Env gp140. *J. Virol.* **87,** 11462–75 (2013).

46. Ward, A. B. & Wilson, I. A. Insights into the trimeric HIV-1 envelope glycoprotein structure. *Trends Biochem. Sci.* **40,** 101–7 (2015).

47. Lobritz, M. A., Ratcliff, A. N. & Arts, E. J. HIV-1 Entry, Inhibitors, and Resistance. *Viruses* **2,** 1069–105 (2010).

48. Pritchard, L. K. *et al.* Structural Constraints Determine the Glycosylation of HIV-1 Envelope Trimers. *Cell Rep.* **11,** 1604–1613 (2015).

49. Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu. Rev. Biochem.* **54,** 631–664 (1985).

50. Behrens, A.-J. *et al.* Molecular Architecture of the Cleavage-Dependent Mannose Patch on a Soluble HIV-1 Envelope Glycoprotein Trimer. *J. Virol.* **91,** (2017).

51. Go, E. P. *et al.* Comparative Analysis of the Glycosylation Profiles of Membrane-Anchored HIV-1 Envelope Glycoprotein Trimers and Soluble gp140. *J. Virol.* **89,** 8245–57 (2015).

52. Stansell, E. *et al.* Gp120 on HIV-1 Virions Lacks O-Linked Carbohydrate. *PLoS One* **10,** e0124784 (2015).

53. Thomas, G. Furin at the cutting edge: From protein traffic to embryogenesis and disease. *Nat. Rev. Mol. Cell Biol.* **3,** 753–766 (2002).

54. Moulard, M. & Decroly, E. Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. *Biochim. Biophys. Acta* **1469,** 121–32 (2000).

55. Checkley, M. A., Luttge, B. G. & Freed, E. O. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J. Mol. Biol.* **410,** 582–608 (2011).

56. Sanders, R. W. & Moore, J. P. Native-like Env trimers as a platform for HIV-1 vaccine design. *Immunol. Rev.* **275,** 161–182 (2017).

57. Tian, J. *et al.* Effect of Glycosylation on an Immunodominant Region in the V1V2 Variable Domain of the HIV-1 Envelope gp120 Protein. *PLoS Comput. Biol.* **12,** e1005094 (2016).

58. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–659 (1998).

59. Garg, H. & Blumenthal, R. Role of HIV Gp41 mediated fusion/hemifusion in bystander apoptosis. *Cell. Mol. Life Sci.* **65,** 3134–44 (2008).

60. Julien, J.-P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science (80-. ).* **342,** (2013).

61. Sanders, R. W. *et al.* A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog.* **9,** e1003618 (2013).

62. Clayton, K. L. *et al.* Resistance of HIV-infected macrophages to CD8+ T lymphocyte–mediated killing drives activation of the immune system. *Nat. Immunol.* **19,** 475–486 (2018).

63. Kong, L. *et al.* Uncleaved prefusion-optimized gp140 trimers derived from analysis of HIV-1 envelope metastability. *Nat. Commun.* **7,** 12040 (2016).

64. Rao, M. *et al.* HIV-1 variable loop 2 and its importance in HIV-1 infection and vaccine development. *Curr. HIV Res.* **11,** 427–38 (2013).

65. Markosyan, R. M., Leung, M. Y. & Cohen, F. S. The six-helix bundle of human immunodeficiency virus Env controls pore formation and enlargement and is initiated at residues proximal to the hairpin turn. *J. Virol.* **83,** 10048–57 (2009).

66. Scherer, L., Rossi, J. J. & Weinberg, M. S. Progress and prospects: RNA-based therapies for treatment of HIV infection. *Gene Ther.* **14,** 1057–1064 (2007).

67. Santoro, M. M. & Perno, C. F. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol.* **2013,** 481314 (2013).

68. Araújo, L., Almeida, S., Araújo, L. A. L. & Almeida, S. E. M. HIV-1 Diversity in the Envelope Glycoproteins: Implications for Viral Entry Inhibition. *Viruses* **5,** 595–604 (2013).

69. Lynch, R. M., Shen, T., Gnanakaran, S. & Derdeyn, C. A. Appreciating HIV type 1 diversity: subtype differences in Env. *AIDS Res. Hum. Retroviruses* **25,** 237–48 (2009).

70. Korber, B. *et al.* Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* **58,** 19–42 (2001).

71. van den Kerkhof, T. L. G. M. *et al.* HIV-1 escapes from N332-directed antibody neutralization in an elite neutralizer by envelope glycoprotein elongation and introduction of unusual disulfide bonds. *Retrovirology* **13,** 48 (2016).

72. Curlin, M. E. *et al.* HIV-1 Envelope Subregion Length Variation during Disease Progression. *PLoS Pathog.* **6,** e1001228 (2010).

73. Zolla-Pazner, S. & Cardozo, T. Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. *Nat. Rev. Immunol.* **10,** 527–35 (2010).

74. Cao, L. *et al.* Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nat. Commun.* **8,** 14954 (2017).

75. Stewart-Jones, G. B. E. *et al.* Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell* **165,** 813–26 (2016).

76. Wei, X. *et al.* Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373,** 117–122 (1995).

77. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422,** 307–312 (2003).

78. Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514,** 455–61 (2014).

79. West, A. P. *et al.* Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell* **156,** 633–48 (2014).

80. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420,** 678–682 (2002).

81. Frey, G. *et al.* A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3739–44 (2008).

82. de Taeye, S. W., Moore, J. P. & Sanders, R. W. HIV-1 Envelope Trimer Design and Immunization Strategies To Induce Broadly Neutralizing Antibodies. *Trends Immunol.* **37,** 221–232 (2016).

83. Rutten, L. *et al.* A Universal Approach to Optimize the Folding and Stability of Prefusion-Closed HIV-1 Envelope Trimers. *Cell Rep* **23,** 584–595 (2018).

84. Horwitz, J. A. *et al.* Non-neutralizing Antibodies Alter the Course of HIV-1 Infection In Vivo. *Cell* **170,** 637–648.e10 (2017).

85. Ozorowski, G. *et al.* Open and Closed Structures Reveal Allostery and Pliability in the HIV-1 Envelope Spike. *Nature* **547,** 360–363 (2017).

86. Galimidi, R. P. *et al.* Intra-spike crosslinking overcomes antibody evasion by HIV-1. *Cell* **160,** 433–46 (2015).

87. Moore, P. L. *et al.* Nature of nonfunctional envelope proteins on the surface of human immunodeficiency virus type 1. *J. Virol.* **80,** 2515–28 (2006).

88. Ringel, O. *et al.* The Hard Way towards an Antibody-Based HIV-1 Env Vaccine: Lessons from Other Viruses. *Viruses* **10,** 197 (2018).

89. Burton, D. R. & Mascola, J. R. Antibody responses to envelope glycoproteins in HIV-1 infection. *Nat Immunol* **16,** 571–576 (2015).

90. Tomaras, G. D. *et al.* Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J. Virol.* **82,** 12449–63 (2008).

91. Overbaugh, J. & Morris, L. The Antibody Response against HIV-1. *Cold Spring Harb. Perspect. Med.* **2,** a007039 (2012).

92. Burton, D. R. & Hangartner, L. Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. *Annu. Rev. Immunol.* **34,** 635–59 (2016).

93. Klein, F. *et al.* Antibodies in HIV-1 vaccine development and therapy. *Science* **341,** 1199–204 (2013).

94. Lynch, R. M. *et al.* HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *J. Virol.* **89,** 4201–13 (2015).

95. Moore, P. L., Williamson, C. & Morris, L. Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends Microbiol.* **23,** 204–11 (2015).

96. Corti, D. & Lanzavecchia, A. Broadly Neutralizing Antiviral Antibodies. *Annu. Rev. Immunol.* **31,** 705–742 (2013).

97. Klein, F. *et al.* Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* **153,** 126–138 (2013).

98. Yu, L. & Guan, Y. Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. *Front. Immunol.* **5,** 250 (2014).

99. Verkoczy, L. & Diaz, M. Autoreactivity in HIV-1 broadly neutralizing antibodies: implications for their function and induction by vaccination. *Curr. Opin. HIV AIDS* **9,** 224–34 (2014).

100. Burton, D. R. *et al.* Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266,** 1024–7 (1994).

101. Buchacher, A. *et al.* Generation of Human Monoclonal-Antibodies Against Hiv-1 Proteins - Electrofusion and Epstein-Barr-Virus Transformation for Peripheral-Blood Lymphocyte Immortalization. *AIDS Res. Hum. Retroviruses* **10,** 359–369 (1994).

102. Irimia, A., Sarkar, A., Stanfield, R. L. & Wilson, I. A. Crystallographic Identification of Lipid as an Integral Component of the Epitope of HIV Broadly Neutralizing Antibody 4E10. *Immunity* **44,** 21–31 (2016).

103. Bryson, S., Julien, J.-P., Hynes, R. C. & Pai, E. F. Crystallographic Definition of the Epitope Promiscuity of the Broadly Neutralizing Anti-Human Immunodeficiency Virus Type 1 Antibody 2F5: Vaccine Design Implications. *J. Virol.* **83,** 11862–11875 (2009).

104. Purtscher, M. *et al.* A broadly neutralizing human monoclonal antibody against gp41 of human immunodeficiency virus type 1. *AIDS Res. Hum. Retroviruses* **10,** 1651–1658 (1994).

105. Murin, C. D. *et al.* Structure of 2G12 Fab2 in Complex with Soluble and Fully Glycosylated HIV-1 Env by Negative-Stain Single-Particle Electron Microscopy. *J. Virol.* **88,** 10177–10188 (2014).

106. Longo, N. S. *et al.* Multiple Antibody Lineages in One Donor Target the Glycan-V3 Supersite of the HIV-1 Envelope Glycoprotein and Display a Preference for Quaternary Binding. *J. Virol.* **90,** 10574–10586 (2016).

107. Falkowska, E. *et al.* Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity* **40,** 657–668 (2014).

108. Blattner, C. *et al.* Structural delineation of a quaternary, cleavage-dependent epitope at the gp41-gp120 interface on intact HIV-1 Env trimers. *Immunity* **40,** 669–680 (2014).

109. Zhang, Z. *et al.* Antiviral Therapy by HIV-1 Broadly Neutralizing and Inhibitory Antibodies. *Int. J. Mol. Sci.* **17,** 1901 (2016).

110. Moldt, B. *et al.* Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 18921–5 (2012).

111. Hessell, A. J. *et al.* Broadly Neutralizing Human Anti-HIV Antibody 2G12 Is Effective in Protection against Mucosal SHIV Challenge Even at Low Serum Neutralizing Titers. *PLoS Pathog.* **5,** e1000433 (2009).

112. Watkins, J. D. *et al.* An Anti-HIV-1 V3 Loop Antibody Fully Protects Cross-Clade and Elicits T-Cell Immunity in Macaques Mucosally Challenged with an R5 Clade C SHIV. *PLoS One* **6,** e18207 (2011).

113. Hurwitz, J. L. *et al.* First phase I clinical trial of an HIV-1 subtype D gp140 envelope protein vaccine: immune activity induced in all study participants. *AIDS* **22,** 149–151 (2008).

114. Lewis, D. J. *et al.* Phase I Randomised Clinical Trial of an HIV-1 CN54, Clade C, Trimeric Envelope Vaccine Candidate Delivered Vaginally. *PLoS One* **6,** e25165 (2011).

115. Spearman, P. *et al.* A Trimeric, V2-Deleted HIV-1 Envelope Glycoprotein Vaccine Elicits Potent Neutralizing Antibodies but Limited Breadth of Neutralization in Human Volunteers. *J. Infect. Dis.* **203,** 1165–1173 (2011).

116. Ringe, R. P. *et al.* Cleavage strongly influences whether soluble HIV-1 envelope glycoprotein trimers adopt a native-like conformation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 18256–61 (2013).

117. Ringe, R. P. *et al.* Influences on the Design and Purification of Soluble, Recombinant Native-Like HIV-1 Envelope Glycoprotein Trimers. *J. Virol.* **89,** 12189–210 (2015).

118. Wu, X. *et al.* Neutralization Escape Variants of Human Immunodeficiency Virus Type 1 Are Transmitted from Mother to Infant. *J. Virol.* **80,** 835–844 (2006).

119. Yasmeen, A. *et al.* Differential binding of neutralizing and non-neutralizing antibodies to native-like soluble HIV-1 Env trimers, uncleaved Env proteins, and monomeric subunits. *Retrovirology* **11,** 41 (2014).

120. Hu, J. K. *et al.* Murine Antibody Responses to Cleaved Soluble HIV-1 Envelope Trimers Are Highly Restricted in Specificity. *J. Virol.* **89,** 10383–98 (2015).

121. Medina-Ramírez, M., Sanders, R. W. & Sattentau, Q. J. Stabilized HIV-1 envelope glycoprotein trimers for vaccine use. *Curr. Opin. HIV AIDS* **12,** 241–249 (2017).

122. Sanders, R. W. *et al.* HIV-1 VACCINES. HIV-1 neutralizing antibodies induced by native-like envelope trimers. *Science* **349,** aac4223 (2015).

123. Klasse, P. J. *et al.* Epitopes for neutralizing antibodies induced by HIV-1 envelope glycoprotein BG505 SOSIP trimers in rabbits and macaques. *PLOS Pathog.* **14,** e1006913 (2018).

124. Klasse, P. J. *et al.* Sequential and Simultaneous Immunization of Rabbits with HIV-1 Envelope Glycoprotein SOSIP.664 Trimers from Clades A, B and C. *PLOS Pathog.* **12,** e1005864 (2016).

125. Schiffner, T. *et al.* Chemical Cross-Linking Stabilizes Native-Like HIV-1 Envelope Glycoprotein Trimer Antigens. *J. Virol.* **90,** 813–28 (2016).

126. Sharma, S. K. *et al.* Cleavage-Independent HIV-1 Env Trimers Engineered as Soluble Native Spike Mimetics for Vaccine Design. *Cell Rep.* **11,** 539–550 (2015).

127. Guenaga, J. *et al.* Structure-Guided Redesign Increases the Propensity of HIV Env To Generate Highly Stable Soluble Trimers. *J. Virol.* **90,** 2806–17 (2015).

128. Pauthner, M. *et al.* Elicitation of Robust Tier 2 Neutralizing Antibody Responses in Nonhuman Primates by HIV Envelope Trimer Immunization Using Optimized Approaches. *Immunity* **46,** 1073–1088.e6 (2017).

129. Kulp, D. W. *et al.* Structure-based design of native-like HIV-1 envelope trimers to silence non-neutralizing epitopes and eliminate CD4 binding. *Nat. Commun.* **8,** 1655 (2017).

130. de Taeye, S. W. *et al.* Immunogenicity of Stabilized HIV-1 Envelope Trimers with Reduced Exposure of Non-neutralizing Epitopes. *Cell* **163,** 1702–15 (2015).

131. Torrents de la Peña, A. *et al.* Improving the Immunogenicity of Native-like HIV-1 Envelope Trimers by Hyperstabilization. *Cell Rep.* **20,** 1805–1817 (2017).

132. Dosenovic, P. *et al.* Immunization for HIV-1 Broadly Neutralizing Antibodies in Human Ig Knockin Mice. *Cell* **161,** 1505–15 (2015).

133. Escolano, A. *et al.* Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1 Antibodies in Ig Knockin Mice. *Cell* **166,** 1445–1458.e12 (2016).

134. Sliepen, K. *et al.* Binding of inferred germline precursors of broadly neutralizing HIV-1 antibodies to native-like envelope trimers. *Virology* **486,** 116–20 (2015).

135. Jardine, J. *et al.* Rational HIV immunogen design to target specific germline B cell receptors. *Science* **340,** 711–6 (2013).

136. Jardine, J. G. *et al.* HIV-1 VACCINES. Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. *Science* **349,** 156–61 (2015).

137. Briney, B. *et al.* Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing Antibodies. *Cell* **166,** 1459–1470.e11 (2016).

138. Medina-Ramírez, M. *et al.* Design and crystal structure of a native-like HIV-1 envelope trimer that engages multiple broadly neutralizing antibody precursors in vivo. *J. Exp. Med.* **214,** 2573–2590 (2017).

139. Lee, S. & Nguyen, M. T. Recent advances of vaccine adjuvants for infectious diseases. *Immune Netw.* **15,** 51–7 (2015).

140. Ozorowski, G. *et al.* Effects of Adjuvants on HIV-1 Envelope Glycoprotein SOSIP Trimers In Vitro. *J. Virol.* **92,** e00381-18 (2018).

141. Asbach, B. & Wagner, R. Particle-based delivery of the HIV envelope protein. *Curr. Opin. HIV AIDS* **12,** 265–271 (2017).

142. Frietze, K. M., Peabody, D. S. & Chackerian, B. Engineering virus-like particles as vaccine platforms. *Curr. Opin. Virol.* **18,** 44–9 (2016).

143. Martinez-Murillo, P. *et al.* Particulate Array of Well-Ordered HIV Clade C Env Trimers Elicits Neutralizing Antibodies that Display a Unique V2 Cap Approach. *Immunity* **46,** 804–817.e7 (2017).

144. Gutjahr, A. *et al.* Biodegradable Polymeric Nanoparticles-Based Vaccine Adjuvants for Lymph Nodes Targeting. *Vaccines* **4,** (2016).

145. Chen, J. *et al.* Effect of the cytoplasmic domain on antigenic characteristics of HIV-1 envelope glycoprotein. *Science* **349,** 191–5 (2015).

146. Pavot, V. *et al.* Poly(lactic acid) and poly(lactic- *co* -glycolic acid) particles as versatile carrier platforms for vaccine delivery. *Nanomedicine* **9,** 2703–2718 (2014).

147. Ataman-Onal, Y. *et al.* Surfactant-free anionic PLA nanoparticles coated with HIV-1 p24 protein induced enhanced cellular and humoral immune responses in various animal models. *J. Control. Release* **112,** 175–85 (2006).

148. Guillon, C. *et al.* Formulation of HIV-1 Tat and p24 antigens by PLA nanoparticles or MF59 impacts the breadth, but not the magnitude, of serum and faecal antibody responses in rabbits. *Vaccine* **25,** 7491–7501 (2007).

149. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9,** 342–53 (2011).

150. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077,** 101–113 (2014).

151. Pugach, P. *et al.* A native-like SOSIP.664 trimer based on an HIV-1 subtype B env gene. *J. Virol.* **89,** 3380–95 (2015).

152. Moore, P. L. & Williamson, C. Approaches to the induction of HIV broadly neutralizing antibodies. *Curr. Opin. HIV AIDS* **11,** 569–575 (2016).

153. Lamalle-Bernard, D. *et al.* Coadsorption of HIV-1 p24 and gp120 proteins to surfactant-free anionic PLA nanoparticles preserves antigenicity and immunogenicity. *J. Control. Release* **115,** 57–67 (2006).

154. Bruun, T.-H. *et al.* Mammalian cell surface display for monoclonal antibody-based FACS selection of viral envelope proteins. *MAbs* **9,** 1052–1064 (2017).

155. Graf, M. *et al.* Concerted action of multiple cis-acting sequences is required for Rev dependence of late human immunodeficiency virus type 1 gene expression. *J. Virol.* **74,** 10822–6 (2000).

156. Fath, S. *et al.* Multiparameter RNA and Codon Optimization: A Standardized Tool to Assess and Enhance Autologous Mammalian Gene Expression. *PLoS One* **6,** e17596 (2011).

157. Green, M. R. & Sambrook, J. *Molecular Cloning: A Laboratory Manual*. (John Inglis, Cold Spring Harbor Laboratory Press, 2012).

158. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166,** 557–580 (1983).

159. Bernard, P., Gabarit, P., Bahassi, E. M. & Couturier, M. Positive-selection vectors using the F plasmid ccdB killer gene. *Gene* **148,** 71–74 (1994).

160. Bruun, T.-H., Mühlbauer, K., Benen, T., Kliche, A. & Wagner, R. A mammalian cell based FACS-panning platform for the selection of HIV-1 envelopes for vaccine development. *PLoS One* **9,** e109196 (2014).

161. Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4645–9 (1990).

162. Quail, M. *et al.* A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics* **13,** 341 (2012).

163. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).

164. Koop, J. Biological Engineering and Characterization of an HIV-1 Envelope-based Genomic Library. (2018).

165. Fiegle, D. Provision of a cell line stably overexpressing the furin protease and enabling efficient expression of quantitatively processed and stabilized HIV-1 Env trimers. (2017).

166. Artimo, P. *et al.* ExPASy : SIB bioinformatics resource portal. *Nucleic Acids Res.* **40,** 597–603 (2012).

167. Asbach, B., Meier, J. P., Pfeifer, M., Köstler, J. & Wagner, R. Computational Design of Epitope-Enriched HIV-1 Gag Antigens with Preserved Structure and Function for Induction of Broad CD8+ T Cell Responses. *Sci. Rep.* **8,** 11264 (2018).

168. Miller, L. Quantifying western blots without expensive commercial quantification software. (2007). Available at: http://www.lukemiller.org/journal/2007/08/quantifying-western-blots-without.html. (Accessed: 10th April 2018)

169. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685 (1970).

170. Glaser, R. W. & Hausdorf, G. Binding kinetics of an antibody against HIV p24 core protein measured with real-time biomolecular interaction analysis suggest a slow conformational change in antigen p24. *J. Immunol. Methods* **189,** 1–14 (1996).

171. Kuttner, G. *et al.* Immunoglobulin V regions and epitope mapping of a murine monoclonal antibody against p24 core protein of HIV-1. *Mol. Immunol.* **29,** 561–564 (1992).

172. Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491,** 406–412 (2012).

173. Irimia, A. *et al.* Lipid interactions and angle of approach to the HIV-1 viral membrane of broadly neutralizing antibody 10E8: Insights for vaccine and therapeutic design. *PLoS Pathog.* **13,** 1–20 (2017).

174. Fiebig, U., Schmolke, M., Eschricht, M., Kurth, R. & Denner, J. Mode of interaction between the HIV-1-neutralizing monoclonal antibody 2F5 and its epitope. *Aids* **23,** 887–895 (2009).

175. Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326,** 285–9 (2009).

176. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480,** 336–43 (2011).

177. Julien, J.-P. *et al.* Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. *Proc. Natl. Acad. Sci.* **110,** 4351–4356 (2013).

178. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477,** 466–470 (2011).

179. Lee, J. H. *et al.* A Broadly Neutralizing Antibody Targets the Dynamic HIV Envelope Trimer Apex via a Long, Rigidified, and Anionic β-Hairpin Structure. *Immunity* **46,** 690–702 (2017).

180. Kong, L. *et al.* Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nat. Struct. Mol. Biol.* **20,** 796–803 (2013).

181. Wu, X. *et al.* Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. *Science* **329,** 853–856 (2010).

182. Zhou, T. *et al.* Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. *Science (80-. ).* **329,** 811–817 (2010).

183. Posner, M. R. *et al.* An IgG human monoclonal antibody that reacts with HIV-1/GP120, inhibits virus binding to cells, and neutralizes infection. *J. Immunol.* **146,** 4325–32 (1991).

184. Chen, L. *et al.* Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science* **326,** 1123–7 (2009).

185. Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* **67,** 3978–88 (1993).

186. Kwong, P. D. *et al.* Structure of an HIV gp 120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–659 (1998).

187. Gorny, M. K. *et al.* Repertoire of Neutralizing Human Monoclonal Antibodies Specific for the V3 Domain of HIV-1 gp120. *J Immunol* **150,** 635–643 (1993).

188. Dutta, M., Liu, J., Roux, K. H. & Taylor, K. A. Visualization of retroviral envelope spikes in complex with the V3 loop antibody 447-52D on intact viruses by cryo-electron tomography. *J. Virol.* **88,** 12265–75 (2014).

189. Zillner, K. *et al.* Microscale Thermophoresis as a Sensitive Method to Quantify Protein: Nucleic Acid Interactions in Solution. in *Methods in molecular biology (Clifton, N.J.)* **815,** 241–252 (2012).

190. Mueller, A. M. *et al.* MicroScale Thermophoresis: A Rapid and Precise Method to Quantify Protein–Nucleic Acid Interactions in Solution. in *Methods in molecular biology (Clifton, N.J.)* **1654,** 151–164 (2017).

191. Feng, Y. *et al.* Thermostability of Well-Ordered HIV Spikes Correlates with the Elicitation of Autologous Tier 2 Neutralizing Antibodies. *PLoS Pathog.* **12,** e1005767 (2016).

192. Scharf, L. *et al.* Broadly Neutralizing Antibody 8ANC195 Recognizes Closed and Open States of HIV-1 Env. *Cell* **162,** 1379–90 (2015).

193. Langedijk, J. P. M., Zekveld, M. J., Ruiter, M., Corti, D. & Back, J. W. Helical peptide arrays for lead identification and interaction site mapping. *Anal. Biochem.* **417,** 149–155 (2011).

194. Corti, D. *et al.* Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS One* **5,** e8805 (2010).

195. Pantophlet, R. *et al.* Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. *J. Virol.* **77,** 642–58 (2003).

196. Olshevsky, U. *et al. Identification of Individual Human Immunodeficiency Virus Type 1 gp120 Amino Acids Important for CD4 Receptor Binding*. *JOURNAL OF VIROLOGY* **64,** (1990).

197. Grassmann, V. Development and application of high- throughput screening methods to generate novel HIV-1 envelope immunogens. (2017).

198. Pancera, M. & Wyatt, R. Selective recognition of oligomeric HIV-1 primary isolate envelope glycoproteins by potently neutralizing ligands requires efficient precursor cleavage. *Virology* **332,** 145–156 (2005).

199. Sanders, R. W. *et al.* Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J. Virol.* **76,** 8875–89 (2002).

200. Wyss, S. *et al.* Regulation of human immunodeficiency virus type 1 envelope glycoprotein fusion by a membrane-interactive domain in the gp41 cytoplasmic tail. *J. Virol.* **79,** 12231–41 (2005).

201. Ganser, I. Optimization of Flow Cytometry based HIV Envelope Library Screening Methods and Characterization of Novel Stabilized Envelope Trimers. (2017).

202. Zolla-Pazner, S. & Cardozo, T. STRUCTURE-FUNCTION RELATIONSHIPS OF HIV-1 ENVELOPE SEQUENCE-VARIABLE REGIONS PROVIDE A PARADIGM FOR VACCINE DESIGN. *Nat Rev Immunol* **10,** 527–535 (2010).

203. Julien, J.-P. *et al.* 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 (2013).

204. Szymczak, A. L. *et al.* Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide–based retroviral vector. *Nat. Biotechnol.* **22,** 589–594 (2004).

205. Guttman, M. *et al.* Antibody potency relates to the ability to recognize the closed, pre-fusion form of HIV Env. *Nat. Commun.* **6,** 6144 (2015).

206. Qi, Y., Jo, S. & Im, W. Roles of glycans in interactions between gp120 and HIV broadly neutralizing antibodies. *Glycobiology* **26,** 251–60 (2016).

207. Pejchal, R. *et al.* A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* **334,** 1097–103 (2011).

208. Pancera, M. *et al.* Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. *Nat. Struct. Mol. Biol.* **20,** 804–13 (2013).

209. Garces, F. *et al.* Structural evolution of glycan recognition by a family of potent HIV antibodies. *Cell* **159,** 69–79 (2014).

210. Steichen, J. M. *et al.* HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. *Immunity* **45,** 483–496 (2016).

211. Galimidi, R. P. *et al.* Intra-spike crosslinking overcomes antibody evasion by HIV-1. *Cell* **160,** 433–46 (2015).

212. Klein, J. S. & Bjorkman, P. J. Few and Far Between: How HIV May Be Evading Antibody Avidity. *PLoS Pathog.* **6,** e1000908 (2010).

213. Samikkannu, T. *et al.* Human immunodeficiency virus type 1 clade B and C gp120 differentially induce neurotoxin arachidonic acid in human astrocytes: implications for neuroAIDS. *J. Neurovirol.* **17,** 230–8 (2011).

214. Kong, W.-P. *et al.* Immunogenicity of multiple gene and clade human immunodeficiency virus type 1 DNA vaccines. *J. Virol.* **77,** 12764–72 (2003).

215. Korber, B. T. M., Macinnes, K., Smith, R. F. & Myers', G. *Mutational Trends in V3 Loop Protein Sequences Observed in Different Genetic Lineages of Human Immunodeficiency Virus Type 1*. *JOURNAL OF VIROLOGY* (1994).

216. Li, Y. *et al.* Explosive HIV-1 subtype B' epidemics in Asia driven by geographic and risk group founder events. *Virology* **402,** 223–227 (2010).

217. Cecilia, D. *et al. Neutralization Profiles of Primary Human Immunodeficiency Virus Type 1 Isolates in the Context of Coreceptor Usage*. *JOURNAL OF VIROLOGY* (1998).

218. Dev, J. *et al.* Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* **353,** 172–175 (2016).

219. Vzorov, A. N., Wang, L., Chen, J., Wang, B.-Z. & Compans, R. W. Effects of modification of the HIV-1 Env cytoplasmic tail on immunogenicity of VLP vaccines. *Virology* **489,** 141–150 (2016).

220. Chung, N. P. *et al.* Stable 293 T and CHO cell lines expressing cleaved, stable HIV-1 envelope glycoprotein trimers for structural and vaccine studies. *Retrovirology* **11,** 33 (2014).

221. Ingale, J. *et al.* High-Density Array of Well-Ordered HIV-1 Spikes on Synthetic Liposomal Nanoparticles Efficiently Activate B Cells. *Cell Rep.* **15,** 1986–99 (2016).

222. Bale, S. *et al.* Covalent Linkage of HIV-1 Trimers to Synthetic Liposomes Elicits Improved B Cell and Antibody Responses. *J. Virol.* **91,** (2017).

223. Tam, H. H. *et al.* Sustained antigen availability during germinal center initiation enhances antibody responses to vaccination. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E6639–E6648 (2016).

224. Mata-Fink, J. *et al.* Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. *J. Mol. Biol.* **425,** 444–56 (2013).

225. Grimm, S. K., Battles, M. B. & Ackerman, M. E. Directed Evolution of a Yeast-Displayed HIV-1 SOSIP gp140 Spike Protein toward Improved Expression and Affinity for Conformational Antibodies. *PLoS One* **10,** e0117227 (2015).

226. Delhalle, S. *et al.* Phages and HIV-1: From Display to Interplay. *Int. J. Mol. Sci.* **13,** 4727–4794 (2012).

227. Binley, J. M. *et al.* Enhancing the proteolytic maturation of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* **76,** 2606–16 (2002).

228. Binley, J. M. *et al.* A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J. Virol.* **74,** 627–43 (2000).

229. Pritchard, L. K. *et al.* Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat. Commun.* **6,** 7479 (2015).

230. Wang, W. *et al.* A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* **10,** 14 (2013).

231. LaBranche, C. C. *et al.* HIV-1 envelope glycan modifications that permit neutralization by germline-reverted VRC01-class broadly neutralizing antibodies. *PLoS Pathog.* **14,** e1007431 (2018).

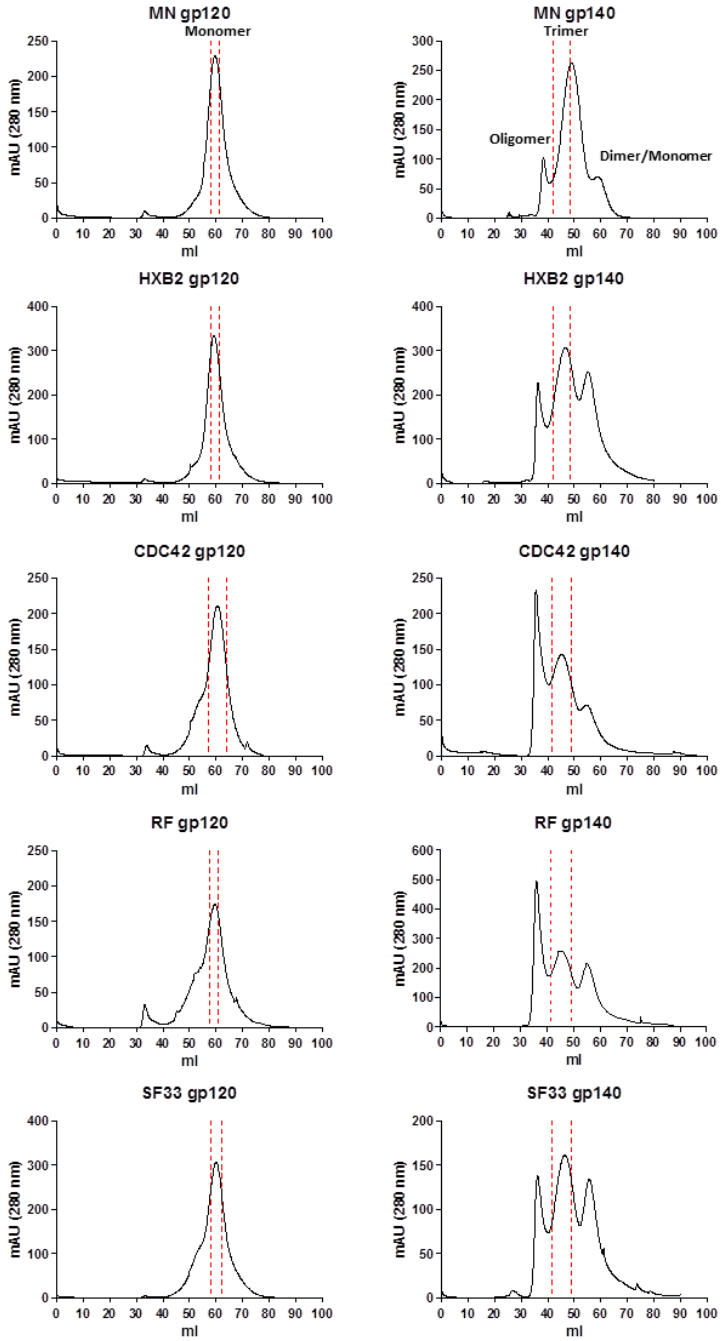
232. Xiao, X. *et al.* Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochem. Biophys. Res. Commun.* **390,** 404–9 (2009).

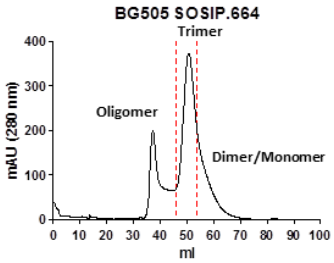
233. Horiya, S., MacPherson, I. S. & Krauss, I. J. Recent strategies targeting HIV glycans in vaccine design. *Nat. Chem. Biol.* **10,** 990–9 (2014).

234. McGuire, A. T. *et al.* Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies. *J. Exp. Med.* **210,** 655–63 (2013).

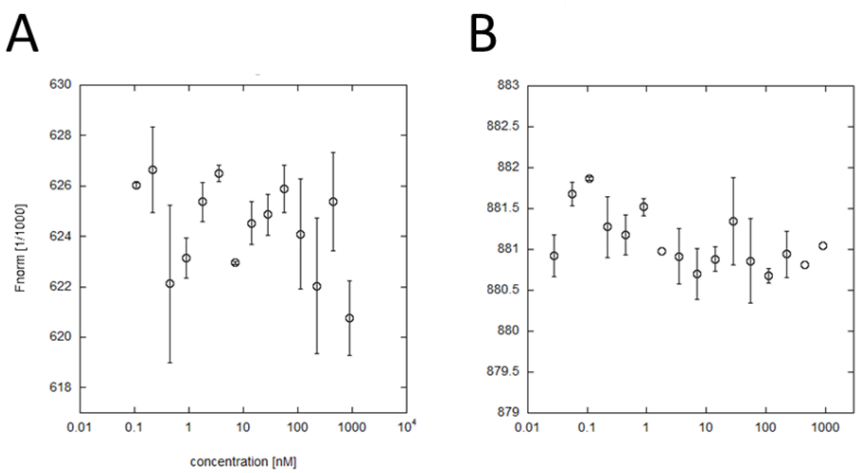
235. O'Gorman, S., Fox, D. T. & Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251,** 1351–5 (1991).

# 7. Appendix

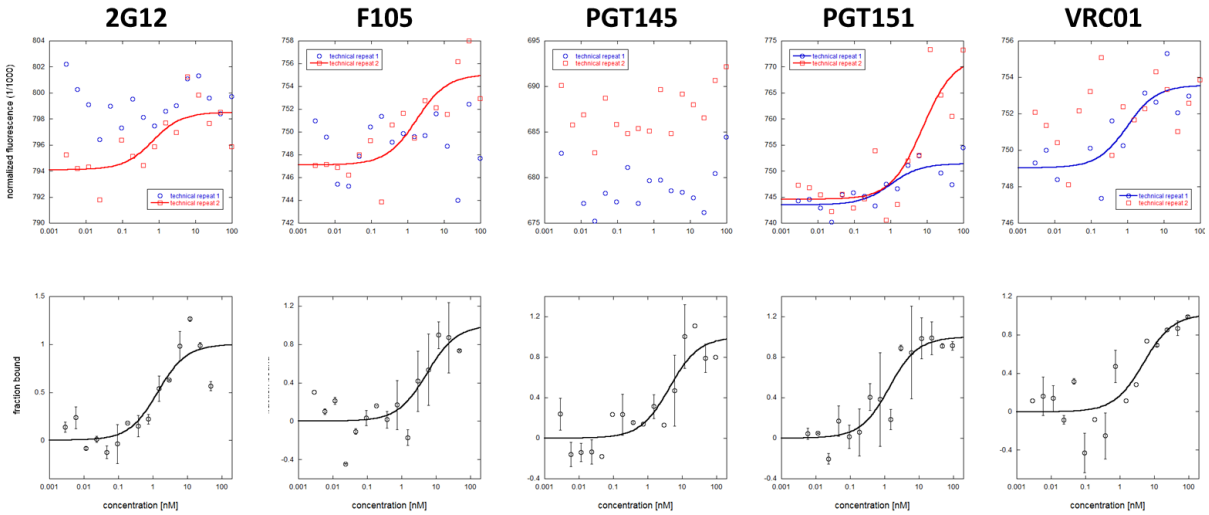


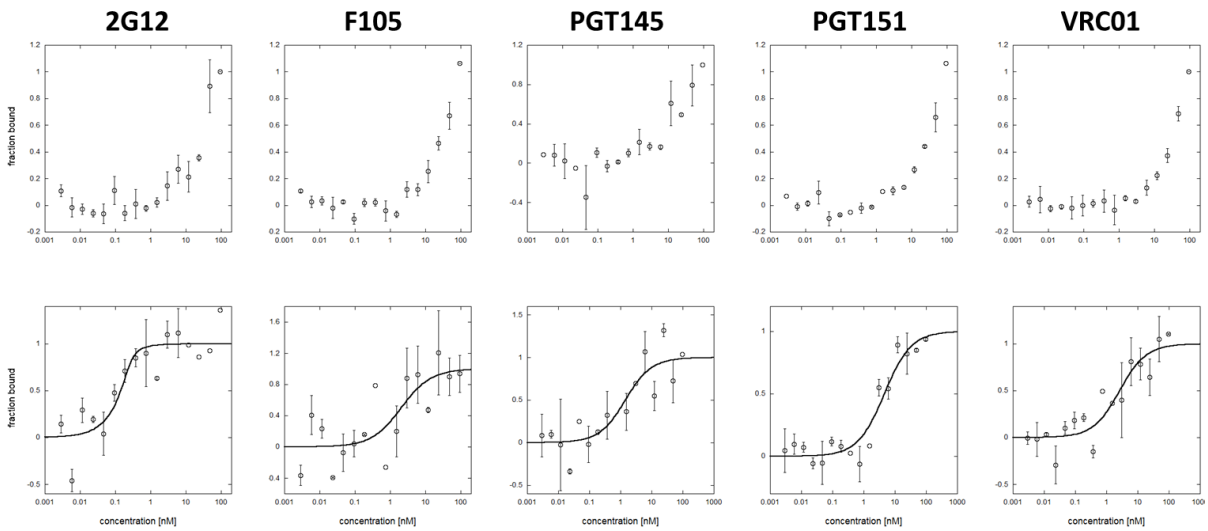

**Appendix Figure 1.** Preparative size exclusion profiles of affinity purified gp120 and gp140 Env/V3 chimeras and BG505 SOSIP.664. Proteins were expressed by transiently transfected 293-F suspension cells. In case of BG505 SOSIP.664 expression, plasmids encoding for Env and Furin were cotransfected in a 3:1 (w/w) ratio to promote cleavage of the protein. Fractions between red lines represent pooled protein fractions.



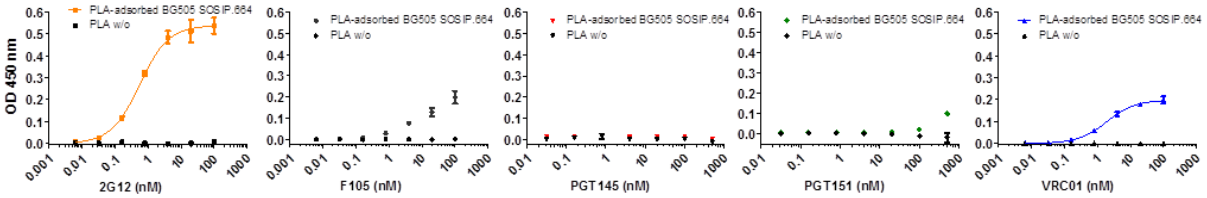

**Appendix Figure 2. MST: Binding of mAb 447-52D to blank VLPs (Gag w/o) in buffer and 50 % serum.** To determine potential background binding of 447-52D to blank VLPs (without surface-presented Env), 447-52D was titrated to DY-647P1-labeled VLPs, which were kept at constant concentration. **(A)** and **(B)** show the normalized fluorescence (FNorm) of two independent experiments in buffer and 50 % human serum, respectively. Error bars reflect the SEM of the independent experiments. In summary, 447-52D exhibited no binding to blank VLPs.



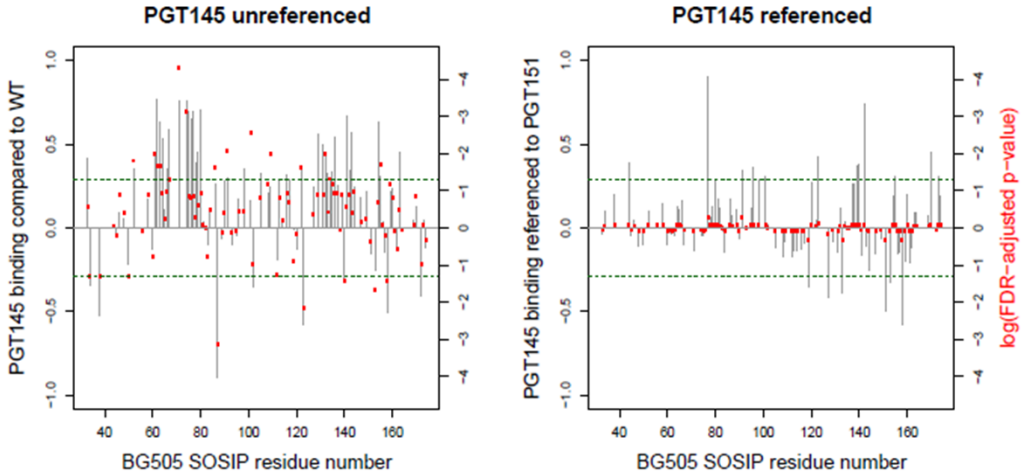

**Appendix Figure 3. MST: Binding of quaternary structure-characterizing mAbs to PLA particles without adsorbed Env proteins in buffer.** To determine background binding, PLA particles without adsorbed Env proteins (control particles) diluted in buffer were titrated to Alexa647-labeled mAbs (VRC01, PGT145, PGT151, F105, 2G12) which were kept at constant concentration. The **upper panel** shows the normalized fluorescence (FNorm) of control particle titrations to mAbs. Depicted are technical repeats (blue and red, respectively) within one MST analysis. Titration curves of PLA-adsorbed BG505 SOSIP.664 trimers in buffer are shown for comparison **(lower panel)**. Results are presented as normalized binding curves of two independent experiments with error bars reflecting the SEM of the independent experiments. In summary, there was background binding detectable.



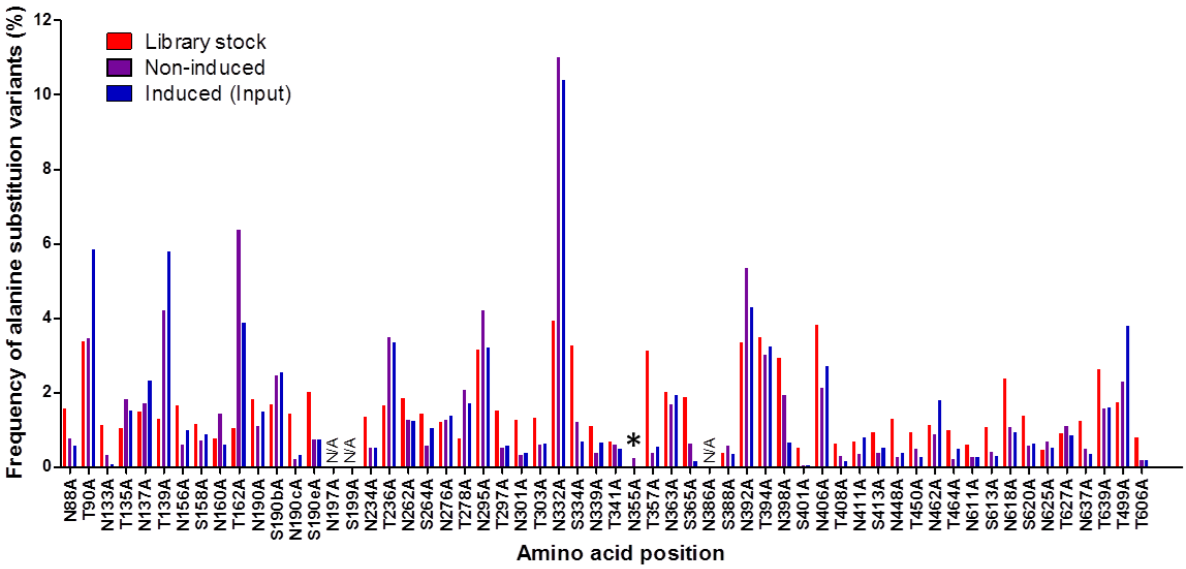

**Appendix Figure 4. MST: Binding of quaternary structure-characterizing mAbs to PLA particles without adsorbed Env proteins in 50 % human serum.** To determine background binding, PLA particles without adsorbed Env proteins (control particles) diluted in 50 % human serum were titrated to Alexa647-labeled mAbs (VRC01, PGT145, PGT151, F105, 2G12) which were kept at constant concentration. The **upper panel** shows the normalized binding curves of control particle titrations to mAbs. Error bars reflect the SEM of two technical repeats within one experiment. Titration curves of PLA-adsorbed BG505 SOSIP.664 trimers in 50 % serum are shown for comparison **(lower panel)**. Results are presented as normalized binding curves of two independent experiments with error bars reflecting the SEM of the independent experiments. In summary, there was also background binding detectable. However, binding of PLA-adsorbed BG505 SOSIP.664 trimers can be considered as specific since saturation was only reached when particle-adsorbed BG505 SOSIP.664 trimers were present.




**Appendix Figure 5. ELISA: Binding of quaternary structure-characterizing mAbs to PLA particles without adsorbed Env proteins in buffer.** For direct-ELISA, PLA-adsorbed BG505 SOSIP.664 trimers and PLA particles without adsorbed Env (control particles, indicated as PLA w/o) were directly coated to the surface of ELISA plates (negative control: Wells coated PBS). MAbs diluted in PBS were titrated to coated PLA particles at indicated concentrations. A HRP-conjugated polyclonal rabbit anti-human Ab was used for readout. ELISA experiments were performed in two independent experiments, error bars reflect the SEM of the independent experiments. This figure presents the titration curves of mAbs which are summarized in **Figure 30C** with the difference that the background signal of control particles (PLA w/o) is not subtracted. In summary, there was no detectable background binding of mAbs to control particles in direct ELISA.




**Appendix Figure 6. PGT145 alanine library screening under optimized conditions. (Left)** The second PGT145 mapping was carried out with changed conditions such as a screening concentration of 5 nM PGT145, controlled quality of all transfected DNAs and screening of surface-exposed gp120 residues. The gains and losses of binding of every residue normalized on BG505 SOSIP gp160 WT and the respective false discovery rate (FDR)-corrected p-values are shown for amino acid residues 33-174. Eight significant losses and 15 significant gains of binding were recognized, yet no distinct patches of losses could be identified. **(Right)** WT-normalized PGT145 signals were referenced on the WT-normalized PGT151 signals and plotted as gains or losses of binding. Former significant gains or losses of binding from the unreferenced screening were calculated as non-significant in the referenced setup. The screening was performed in five replicates, data were generated during the master`s thesis of Iris Ganser under experimental supervision of Dr. David Peterhoff.




**Appendix Figure 7. Frequency of alanine substitution variants in the BG505 SOSIP gp145 glycan-knockout library.** NGS was used to investigate the variability of the cellular library over time. To this end, the frequency of alanine substitution variants in the library was analyzed at three different points of time: 1) directly after pooling of stable cell lines (frozen library stock); 2) non-induced cells, cultured for 18 days; and 3) induced cells (input material for sort, cultured for further 48 h in presence of doxycycline). The NGS analysis revealed that four stable cell lines were missing in the library stock (N197A, S199A, N355A, N386A). Despite a varying distribution, all variants detected in the library stock were present in the input sample. Variant N355A (indicated by an asterisk) seemed to be only apparent in the non-induced sample. However, as N355A was not present in the library stock, the variant was considered as not available (N/A) as were N197A, S199A and N386A.

**Appendix Table 1. Comparison of MST and ELISA regarding consumables and experimental effort.** MST calculations are based on 16 dilutions (1:2 dilution series, starting with 1000 nM) with a volume of 6 µl/capillary. ELISA calculations are based on 8 dilutions (1:10 dilution series, starting with 1000 nM) and a volume of 50 µl/well.

| | **MST** | **ELISA** |
|---|---|---|
| Amount titrant | Antibody: ~2 µg<br>Env: ~3 µg | Antibody: ~9 µg |
| Amount labeled interaction partner MST/solid phase ELISA | Antibody: ~0.7 – 7 ng<br>VLPs: < 100 ng Env | Env: ~400 ng<br>VLPs: ~200 ng Env |
| Hands on time | ~15 min | ~ 60 min |
| Duration experiment | ~ 20 min for measurement of 16 capillaries | ~ 5 h |

**Appendix Table 2. Oligonucleotide sequences used for cloning/sequencing of envelope variants in pQL13 and pcDNA5/FRT/TO.**

| **Name** | **Sequence 5´→3´** |
|---|---|
| Primer for cloning envelope variants into pQL13 (addition of Esp3I restriction sites) | |
| THB-5I3-fwd-QL-ZM96 | ATAATACGTCTCGCTAGCATGGGAGTGCGGGAGATCCTGCGGAACTG GCAGCGGTGGTGG |
| RevZM96gp140TagQL | ATATTCGTCTCCTCGAGCTATCAGTGGTGATGGTGGTGGTGCTTGTCGT CGTCGTCCTTGTAGTCGCTGCCGCCTCCGCCGCTCTTGATGTACCACAG CCACTTGG |
| Rev-96zm651-gp120-QL | ATATTCGTCTCCTCGAGCTACTACTAACTCTTCTCCCGCTCCACCACCC GCCGCTTGGCCTCGG |
| Primer for cloning envelope variants into pcDNA5/FRT/TO (addition of HindIII and XhoI restriction sites) | |
| p X BG505_Ala fw | GATCAAGCTTGCCACCATGAGAGTGATGGGCATCCAGCGG |
| BG505 gp145 rev | GATCCTCGAGTTATCAGTAGCCCTGCCGCACTCTGTGG |
| BG505 gp160 rev | GATCCTCGAGTTATCACAGCAGTGCTCTTTCCAGGCCC |
| Primer for overlap extension PCR (REKR to REKS substitution) for cloning of gp140 96ZM651 Env/V3 chimeras | |
| FusPrim-synZM96-cleav-mut-For | GGGTGGTGGAGCGGGAGAAGAGCGCCGTGGGCATCGG |
| FusPrim-synZM96-cleav-mut-Rev | CCGATGCCCACGGCGCTCTTCTCCCGCTCCACCACCC |
| Primer for overlap extension PCR (insertion of I to P substitution at AA position 559 and R6 site in the BG505 envelope sequence, stop codon deletion for cloning of gp160 constructs) | |
| BG505 6x Arginin mut for | GGCCAAGAGAAGAGTCGTGGGACGGCGGAGAAGGCGGAGAGCCGTG GGCATCGGAGCCGTGTTTCTGG |
| BG505 6x Arginin mut rev | CCAGAAACACGGCTCCGATGCCCACGGCTCTCCGCCTTCTCCGCCGTC CCACGACTCTTCTCTTGGCC |
| BG505 I559P for | GCAATCTGCTCAGAGCCCCTGAGGCCCAGCAG |
| BG505 I559P rev | CTGCTGGGCCTCAGGGGCTCTGAGCAGATTGC |
| BG505 gp145 Stop deletion | GTAGCCCTGCCGCACTCTGTGGATC |
| Sequencing primer for 96ZM651 Env/V3 chimeras in pQL13 | |
| pQL13_seq_for | CTGAGCACCCAGTCCGCCCTGAGC |
| THB ZM96 seq-1-fwd | ATCCCCATCCACTACTGCG |
| THB ZM96 seq-2-fwd | AGCACCAACGACAGCACC |
| pcDNA3.1seq r3191 BGHpA rev | CCCCCAGAATAGAATGACACC |
| Sequencing primer for BG505 envelope variants in pcDNA5/FRT/TO | |
| pC-CMV-For | GTAGGCGTGTACGGTGGGAGG |

| | |
|---|---|
| BG505 SOSIP Seq1 | GGACAAGAAGTTCAACGG |
| BG505 SOSIP Seq2 | AGAGAAGAGTCGTGGGACG |
| pcDNA3.1seq r3191 BGHpA rev | CCCCCAGAATAGAATGACACC |

**Appendix Table 3. Plasmid vectors used in this thesis.**

| Description | Selectable marker | Specification |
|---|---|---|
| pcDNA3.1(+) | $Amp^R$, $Neo^R$ | Mammalian expression vector (ThermoFisher Scientific, Waltham, USA, # V79020); expression of TeeGag1 for VLP production. |
| pcDNA3.1(+) QL | $Amp^R$, $Neo^R$ | Derivative of pcDNA3.1(+), including a CcdB cloning cassette at the MCS; expression of Furin; cotransfected with BG505 R6 envelope variants to promote protein cleavage. |
| pcDNA5/FRT/TO | $Amp^R$, $Hygro^R$ (no ATG) | Tetracycline inducible mammalian expression vector. Designed for generation of stable cell lines using Flp-In™ T-REx™ system (ThermoFisher Scientific, # V652020). |
| pLib1.1 | $Amp^R$, $Hygro^R$ (no ATG) | Derivative of pcDNA5/FRT/TO, for details see **3.2.3**; used for cloning of the BG505 alanine substitution library. |
| pQL13 | $Amp^R$, $Hygro^R$ (no ATG) | Derivative of pcDNA5/FRT/TO, including a CcdB cloning cassette and eGFP[154]; used for expression of gp120, gp140 and gp145 Env/V3 chimeras. |
| pOG44 | $Amp^R$ | Mammalian expression vector for transient expression of the Flp recombinase[235]; (ThermoFisher Scientific, # V600520); used for generation of stable cell lines. |
| pSEVA silent | $Kan^R$ | Derivative of pSEVA241 (GenBank # JX560329), including a CcdB cloning cassette at the MCS; used for generation of exchange cassettes, see **3.2.3**. |
| CMVR VRC01 mAb heavy/light chain expression vectors | $Kan^R$ | Mammalian expression vector (NIH AIDS Reagent Program, Germantown, Maryland, USA, # 12035, # 12036); used for antibody expression. |

**Appendix Table 4. Oligonucleotide, DNA linker and gBlock sequences used for cloning of the BG505 SOSIP alanine substitution library.**

| **Oligonucleotides** | |
|---|---|
| **Name** | **Sequence 5´→3´** |
| Primer for cloning of BG505 gp145/gp160 constructs into pcDNA5/FRT/TO (addition of HindIII restriction site) | |
| p X BG505_Ala fw | GATCAAGCTTGCCACCATGAGAGTGATGGGCATCCAGCGG |
| Primer for cloning of BG505 gp145/gp160 constructs into pcDNA5/FRT/TO (addition of stop codon and XhoI restriction site) | |
| BG505 gp145 rev | GATCCTCGAGTTATCAGTAGCCCTGCCGCACTCTGTGG |
| BG505 gp160 rev | GATCCTCGAGTTATCACAGCAGTGCTCTTTCCAGGCCC |
| Primer for adding NotI, NheI and AarI restriction sites as well as a stop codon to BG505 gp145 | |
| BG505-gp145+X-rev | GTGCGCTAGCGCGGCCGCTTATCACCTGCCTCAGTAGCCCTGCCGCACTCTGTGGATC |
| Primer for gBlockCT amplification, attachment of a glycine-serine linker and a hexa histidine-tag | |
| gBlock_CT+X_ampli_fw | AGGGCTACTGAGGCAGGTGATAAGC |
| BG505_Ala_His_rev | GGGCCCTTATCAGTGGTGATGGTGGTGGTGGCTGCCATCCAGAGCCAGCAGATCCTG |
| Primer for gBlockXhoI amplification | |
| gBlock-Xho_fw | GGCAAAGCTTTCTAGACCATGGGCGCTG |
| gBlock-Xho_rev | GTGCGCTAGCGCGGCCGCTTATCACC |

| | |
|---|---|
| Primer to remove a BsaI site in the ampicillin resistance of pcDNA5/FRT/TO by silent mutation | |
| pcDNA5_no_BsaI_fw | GCGGGTCTCTACCGCGAGAgCCACGCTCACCGGCTCCAGATTTATC |
| pC-BGH-Rev | GCAACTAGAAGGCACAGTCGAGG |
| Sequencing primer for pcDNA5/FRT/TO | |
| pcDNA5/FRT/TO seq 1 fw | TCTCCCGATCCCCTATGGT |
| pcDNA5/FRT/TO seq 2 fw | GTACGGTGGGAGGTCTATAT |
| pcDNA5/FRT/TO seq 3 fw | ACTTGATTAGGGTGATGGTT |
| pcDNA5/FRT/TO seq 4 fw | CAACATCTTCTTCTGGAGGC |
| pcDNA5/FRT/TO seq 5 fw | GTGCCAGCTGCATTAATGA |
| pcDNA5/FRT/TO seq 6 fw | GCGGTGGTTTTTTTGTTTGC |
| pcDNA5/FRT/TO seq 7 fw | GAGAATAGTGTATGCGGCGA |
| Sequencing primer pLib1.1 | |
| BG505_gp160rev | AGAAAGCCGGACACCAGT |
| **DNA linker (annealed oligonucleotide pairs)** | |
| **Name** | **Sequence 5´→3´** |
| Cloning of 3´gp145 part with SfiI restriction site in pLib1.1 (via KpnI and XhoI) | |
| lig-p145_fw | CGGCCTGATCGGCCTCCGCATTGTCTTCGCTGTCCTCTCCGTCATTCATAGCGTACGCC AAGGATATTAATGAGC |
| lig-p145_rev | TCGAGCTCATTAATATCCTTGGCGTACGCTATGAATGACGGAGAGGACAGCGAAGAC AATGCGGAGGCCGATCAGGCCGGTAC |
| Linker pair to remove a KpnI restriction site in the MCS of pcDNA5/FRT/TO (via BamHI and HindIII) | |
| pcDNA5-Kpn-del-li-fw | AGCTTGGTAGCGAGCTCG |
| pcDNA5-Kpn-del-li-rev | GATCCGAGCTCGCTACCA |
| Linker pair to remove a KpnI restriction site in the MCS of pSEVA silent (via BamHI and EcoRI) | |
| pSEVA-KpnI_del_li_fw | AATTCGAGCTCAGTACCCGGG |
| pSEVA-KpnI_del_li_rev | GATCCCCGGGTACTGAGCTCG |
| **gBlock sequences** | |
| **Name** | **Sequence 5´→3´** |
| gBlockCT | |
| DP_BG505_CT-X | AGGGCTACTGAGGCAGGTGATAAGCGGCCGCGCTAGCGCACACCTGCTAGGCTACAG CCCCCTGAGCTTTCAGACCCACACCCCCAACCCCAGAGGCCTGGACAGACCCGAGAG AATCGAGGAAGAGGACGGCGAGCAGGACCGGGGCAGATCTACCAGACTGGTGTCCG GCTTTCTGGCCCTGGCCTGGGACGATCTGAGAAGCCTGTGCCTGTTCTGCTACCACCG GCTGCGGGACTTCATCCTGATCGCCGCCAGAATCGTGGAACTGCTGGGCCACAGCTCC CTGAAGGGCCTGAGACTGGGATGGGAGGGCCTGAAGTACCTGTGGAACCTGCTGGCC TACTGGGGCAGAGAGCTGAAGATCAGCGCCATCAACCTGTTCGACACAATCGCCATT GCCGTGGCCGAGTGGACCGACAGAGTGATCGAGATCGGCCAGCGGCTGTGCAGAGCC TTCCTGCACATCCCCAGACGCATCCGGCAGGGCCTGGAAAGAGCACTGCTGTAATGA GGTACCCTCGAGGAGAGTCAGAATCAGCAGGAGAAAAATGAACAGGATCTGCTGGCT CTGGATTGATAAGGGCCCGATCCGATC |

| gBlockXhoI | |
|---|---|
| BG505_SOSIP_R6_145-3'part+exobox | GGCAAAGCTTTCTAGACCATGGGCGCTGCCAGCATGACACTGACCGTGCAGGCCCGG<br>AACCTGCTGAGCGGCATCGTGCAGCAGCAGAGCAATCTGCTCAGAGCCCCTGAGGCC<br>CAGCAGCATCTGCTGAAACTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCAGAGTG<br>CTGGCCGTGGAAAGATACCTGCGGGATCAGCAGCTCCTGGGCATCTGGGGATGCAGC<br>GGCAAGCTGATCTGCTGCACCAATGTGCCCTGGAACAGCAGCTGGTCCAACCGGAAC<br>CTGAGCGAGATCTGGGACAACATGACCTGGCTGCAGTGGGACAAAGAGATCAGCAAC<br>TACACCCAGATCATCTACGGCCTGCTCGAGGAGAGCCAGAACCAGCAGGAAAAGAAC<br>GAGCAGGACCTGCTGGCCCTGGACAAGTGGGCCAGCCTGTGGAATTGGTTCGACATC<br>AGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGCCTGATCGGCC<br>TGCGGATCGTGTTTGCCGTGCTGAGCGTGATCCACAGAGTGCGGCAGGGCTACTGAGG<br>CAGGTGATAAGCGGCCGCGCTAGCGCAC |

**Appendix Table 5. Oligonucleotide pairs for cloning of sequence segment exchange cassettes in pSEVA silent containing the BG505 SOSIP R6 library template.**

| Exchange cassette | Forward primer 5´→3´ | Reverse primer 5´→3´ |
|---|---|---|
| 1 | GTAGGTACCGGTCTCACTGTTCTGCGCCA GCGACGC | GTAGGTACCGGTCTCGGTTCTCGGCGGCA GAGCAGA |
| 2 | GTAGGTACCGGTCTCAGCCACCCACGCCT GCGTGCC | GTAGGTACCGGTCTCGGGTTGTCTCGGCGT CCTTCC |
| 3 | GTAGGTACCGGTCTCAAACGTGACCGAAG AGTTCAA | GTAGGTACCGGTCTCGCCACACGTTGTGC TTCTCTG |
| 4 | GTAGGTACCGGTCTCAACCGACATCATCA GCCTGTG | GTAGGTACCGGTCTCGTTCCAGGTGGATTT CCTGGG |
| 5 | GTAGGTACCGGTCTCACCCCTGTGCGTGA CCCTGCA | GTAGGTACCGGTCTCGGTGCATCTGTTCCA CCATGT |
| 6 | GTAGGTACCGGTCTCAATGCGGGGCGAGC TGAAGAA | GTAGGTACCGGTCTCGGGTCAGCTTCACG CAGGGCT |
| 7 | GTAGGTACCGGTCTCAAAGAAACAGAAGG TGTACAG | GTAGGTACCGGTCTCGGTCGTCGGTGATG TTGTTGG |
| 8 | GTAGGTACCGGTCTCAGAGAACCAGGGCA ACAGAAG | GTAGGTACCGGTCTCGGTCCCGCAGCTCG GTGGTCA |
| 9 | GTAGGTACCGGTCTCATGCAACACCAGCG CCATCAC | GTAGGTACCGGTCTCGGTTGATCTGCACC ACGTCCA |
| 10 | GTAGGTACCGGTCTCACCCATCCACTACT GCGCCCC | GTAGGTACCGGTCTCGGTTGATCAGCCGG TACTCTT |
| 11 | GTAGGTACCGGTCTCAAAGTTCAACGGCA CCGGCCC | GTAGGTACCGGTCTCGGATGGGCTCGAAG GACACCT |
| 12 | GTAGGTACCGGTCTCAGGCATCAAGCCCG TGGTGTC | GTAGGTACCGGTCTCGCTTGTCCTTGCACT TCAGGA |
| 13 | GTAGGTACCGGTCTCAGAGGAAGTGATGA TCAGAAG | GTAGGTACCGGTCTCGGTGGGTACACTGC ACGGTGG |
| 14 | GTAGGTACCGGTCTCAGTGCAGTTCAACA CCCCCGT | GTAGGTACCGGTCTCGTTCGGCGAGGCTG CCGTTCA |
| 15 | GTAGGTACCGGTCTCAAGAAAGAGCATCC GGATCGG | GTAGGTACCGGTCTCGCAGGATGTTCTTG GCGTTGT |
| 16 | GTAGGTACCGGTCTCAATCGGCGACATCC GGCAGGC | GTAGGTACCGGTCTCGGGTGTTGTTGTTGG GCCGGG |
| 17 | GTAGGTACCGGTCTCAACACTGGGCAAGG TGGTGAA | GTAGGTACCGGTCTCGGATGTCGCCGGTG GCGTAGA |
| 18 | GTAGGTACCGGTCTCAATCAGATTCGCCA ACAGCTC | GTAGGTACCGGTCTCGCTCGTTCCAGGTG GCCTTGG |
| 19 | GTAGGTACCGGTCTCAAACTGTGGCGGCG AGTTCTT | GTAGGTACCGGTCTCGGATGGTGTTGTTCC CGAAGT |
| 20 | GTAGGTACCGGTCTCATGGATCAGCAATA CCAGCGT | GTAGGTACCGGTCTCGGAAGCTGTGGGTG GTCACTT |
| 21 | GTAGGTACCGGTCTCAATCACCCTGCCCT GCCGGAT | GTAGGTACCGGTCTCGGGTGCTGTTGAAC AGGCCGG |
| 22 | GTAGGTACCGGTCTCACAGGCTATGTACG CCCCACC | GTAGGTACCGGTCTCGGCTGTCGTTGCTGC CGGTGC |
| 23 | GTAGGTACCGGTCTCAACCGGCCTGATCC TGACCCG | GTAGGTACCGGTCTCGGCCGATCCGCTGC CACATAT |

| | | |
|---|---|---|
| 24 | GTAGGTACCGGTCTCATTCAGACCCGGCG GAGGCGA | GTAGGTACCGGTCTCGGATATTGGACACG CATCTGA |
| 25 | GTAGGTACCGGTCTCATACAAAGTGGTGA AAATCGA | GTAGGTACCGGTCTCGGGTTTCGGTGGTG CTGTTGG |
| 26 | GTAGGTACCGGTCTCAAGAGTCGTGGGAC GGCGGAG | GTAGGTACCGGTCTCGCTTGTACAGCTCG CTCCGCC |
| 27 | GTAGGTACCGGTCTCACTGGGCTTTCTGGG AGCCGC | GTAGGTACCGGTCTCGTCTCTTGCACCGG GTGGGGG |
| 28 | GTAGGTACCGGTCTCAACCGTGCAGGCCC GGAACCT | GTAGGTACCGGTCTCGAAACACGGCTCCG ATGCCCA |
| 29 | GTAGGTACCGGTCTCACTCAGAGCCCCTG AGGCCCA | GTAGGTACCGGTCTCGCAGTGTCATGCTG GCAGCGC |
| 30 | GTAGGTACCGGTCTCAAAGCAGCTGCAGG CCAGAGT | GTAGGTACCGGTCTCGCAGATTGCTCTGCT GCTGCA |
| 31 | GTAGGTACCGGTCTCACTCCTGGGCATCT GGGGATG | GTAGGTACCGGTCTCGGATGCCCCACACG GTCAGTT |
| 32 | GTAGGTACCGGTCTCATGGAACAGCAGCT GGTCCAA | GTAGGTACCGGTCTCGCTGCTGATCCCGC AGGTATC |
| 33 | GTAGGTACCGGTCTCATGGCTGCAGTGGG ACAAAGA | GTAGGTACCGGTCTCGGGGCACATTGGTG CAGCAGA |
| 34 | GTAGGTACCGGTCTCACTCGAGGAGAGCC AGAACCA | GTAGGTACCGGTCTCGGGTCATGTTGTCCC AGATCT |
| 35 | GTAGGTACCGGTCTCAGACAAGTGGGCCA GCCTGTG | GTAGGTACCGGTCTCGCAGGCCGTAGATG ATCTGGG |
| 36 | GTAGGTACCGGTCTCAATCAAGATCTTCAT CATGAT | GTAGGTACCGGTCTCGCAGGGCCAGCAGG TCCTGCT |
| 37 | GTAGGTACCGGTCTCAGCCGTGCTGAGCG TGATCCA | GTAGGTACCGGTCTCGGTACCACAGCCAG TTGCTGA |
| 38 | GTAGGTACCGGTCTCATGAGGCAGGTGAT AAGCGGC | GTAGGTACCGGTCTCGAAACACGATCCGC AGGCCGA |

**Appendix Table 6. DNA linker sequences encoding for alanine substitutions at indicated positions.**

| Residue HXB2 | Amino acid BG505 | Linker with alanine substitution (forward sequence 5´→3´) | Linker with alanine substitution (reverse sequence 5´→3´) |
|---|---|---|---|
| 34 | L | GAACGCCTGGGTGACAGTGTACTACGGCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGCCGTAGTACACTGTCACCCAGGC |
| 35 | W | GAACCTGGCCGTGACAGTGTACTACGGCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGCCGTAGTACACTGTCACGGCCAG |
| 36 | V | GAACCTGTGGGCCACAGTGTACTACGGCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGCCGTAGTACACTGTGGCCCACAG |
| 37 | T | GAACCTGTGGGTGGCCGTGTACTACGGCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGCCGTAGTACACGGCCACCCACAG |
| 38 | V | GAACCTGTGGGTGACAGCCTACTACGGCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGCCGTAGTAGGCTGTCACCCACAG |
| 39 | Y | GAACCTGTGGGTGACAGTGGCCTACGGCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGCCGTAGGCCACTGTCACCCACAG |
| 40 | Y | GAACCTGTGGGTGACAGTGTACGCCGGCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGCCGGCGTACACTGTCACCCACAG |
| 41 | G | GAACCTGTGGGTGACAGTGTACTACGCCGT GCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GCACGGCGTAGTACACTGTCACCCACAG |
| 42 | V | GAACCTGTGGGTGACAGTGTACTACGGCGC CCCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG GGGCGCCGTAGTACACTGTCACCCACAG |
| 43 | P | GAACCTGTGGGTGACAGTGTACTACGGCGT GGCCGTGTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCACACGG CCACGCCGTAGTACACTGTCACCCACAG |
| 44 | V | GAACCTGTGGGTGACAGTGTACTACGGCGT GCCCGCCTGGAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTCCAGGCGG GCACGCCGTAGTACACTGTCACCCACAG |
| 45 | W | GAACCTGTGGGTGACAGTGTACTACGGCGT GCCCGTGGCCAAGGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCCTTGGCCACGG GCACGCCGTAGTACACTGTCACCCACAG |
| 46 | K | GAACCTGTGGGTGACAGTGTACTACGGCGT GCCCGTGTGGGCCGACGCCGAGACAACC | ACAGGGTTGTCTCGGCGTCGGCCCACACGG GCACGCCGTAGTACACTGTCACCCACAG |
| 47 | D | GAACCTGTGGGTGACAGTGTACTACGGCGT GCCCGTGTGGAAGGCCGCCGAGACAACC | ACAGGGTTGTCTCGGCGGCCTTCCACACGG GCACGCCGTAGTACACTGTCACCCACAG |
| 49 | E | GAACCTGTGGGTGACAGTGTACTACGGCGT GCCCGTGTGGAAGGACGCCGCCACAACC | ACAGGGTTGTGGCGGCGTCCTTCCACACGG GCACGCCGTAGTACACTGTCACCCACAG |
| 50 | T | GAACCTGTGGGTGACAGTGTACTACGGCGT GCCCGTGTGGAAGGACGCCGAGGCCACC | ACAGGGTGGCCTCGGCGTCCTTCCACACGG GCACGCCGTAGTACACTGTCACCCACAG |
| 51 | T | GAACCTGTGGGTGACAGTGTACTACGGCGT GCCCGTGTGGAAGGACGCCGAGACAGCC | ACAGGGCTGTCTCGGCGTCCTTCCACACGG GCACGCCGTAGTACACTGTCACCCACAG |

| | | | |
|---|---|---|---|
| 52 | L | AACCGCCTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAAGGC |
| 53 | F | AACCCTGGCCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGGCCAG |
| 56 | S | AACCCTGTTCTGCGCCGCCGACGCCAAGGC<br>CTACGAGACAGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCTGTCTCGTA<br>GGCCTTGGCGTCGGCGGCGCAGAACAG |
| 57 | D | AACCCTGTTCTGCGCCAGCGCCGCCAAGGC<br>CTACGAGACAGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCTGTCTCGTA<br>GGCCTTGGCGGCGCTGGCGCAGAACAG |
| 59 | K | AACCCTGTTCTGCGCCAGCGACGCCGCCGC<br>CTACGAGACAGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCTGTCTCGTA<br>GGCGGCGGCGTCGCTGGCGCAGAACAG |
| 61 | Y | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CGCCGAGACAGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCTGTCTCGGC<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 62 | E | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGCCACAGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCTGTGGCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 63 | T | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGGCCGAGAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTCTCGGCCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 64 | E | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGCCAAGCACAACGTGTGG | TGGCCCACACGTTGTGCTTGGCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 65 | K | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGAGGCCCACAACGTGTGG | TGGCCCACACGTTGTGGGCCTCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 66 | H | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGAGAAGGCCAACGTGTGG | TGGCCCACACGTTGGCCTTCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 67 | N | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGAGAAGCACGCCGTGTGG | TGGCCCACACGGCGTGCTTCTCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 68 | V | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGAGAAGCACAACGCCTGG | TGGCCCAGGCGTTGTGCTTCTCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 69 | W | AACCCTGTTCTGCGCCAGCGACGCCAAGGC<br>CTACGAGACAGAGAAGCACAACGTGGCC | TGGCGGCCACGTTGTGCTTCTCTGTCTCGTA<br>GGCCTTGGCGTCGCTGGCGCAGAACAG |
| 71 | T | GTGGGCCGCCCACGCCTGCGTGCCCACCGA<br>CCCCAACCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGTTGGG<br>GTCGGTGGGCACGCAGGCGTGGGCGGC |
| 72 | H | GTGGGCCACCGCCGCCTGCGTGCCCACCGA<br>CCCCAACCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGTTGGG<br>GTCGGTGGGCACGCAGGCGGCGGTGGC |
| 75 | V | GTGGGCCACCCACGCCTGCGCCCCCACCGA<br>CCCCAACCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGTTGGG<br>GTCGGTGGGGGCGCAGGCGTGGGTGGC |
| 76 | P | GTGGGCCACCCACGCCTGCGTGGCCACCGA<br>CCCCAACCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGTTGGG<br>GTCGGTGGCCACGCAGGCGTGGGTGGC |
| 77 | T | GTGGGCCACCCACGCCTGCGTGCCCGCCGA<br>CCCCAACCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGTTGGG<br>GTCGGCGGGCACGCAGGCGTGGGTGGC |
| 78 | D | GTGGGCCACCCACGCCTGCGTGCCCACCGC<br>CCCCAACCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGTTGGG<br>GGCGGTGGGCACGCAGGCGTGGGTGGC |
| 79 | P | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CGCCAACCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGTTGGC<br>GTCGGTGGGCACGCAGGCGTGGGTGGC |
| 80 | N | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCGCCCCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGGGGCGG<br>GGTCGGTGGGCACGCAGGCGTGGGTGGC |
| 81 | P | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCAACGCCCAGGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCCTGGGCGTTGGG<br>GTCGGTGGGCACGCAGGCGTGGGTGGC |
| 82 | Q | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCAACCCCGCCGAAATCCACCTGGAA | CGTTTTCCAGGTGGATTTCGGCGGGGTTGG<br>GGTCGGTGGGCACGCAGGCGTGGGTGGC |
| 83 | E | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCAACCCCCAGGCCATCCACCTGGAA | CGTTTTCCAGGTGGATGGCCTGGGGGTTGG<br>GGTCGGTGGGCACGCAGGCGTGGGTGGC |
| 84 | I | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCAACCCCCAGGAAGCCCACCTGGAA | CGTTTTCCAGGTGGGCTTCCTGGGGGTTGGG<br>GTCGGTGGGCACGCAGGCGTGGGTGGC |
| 85 | H | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCAACCCCCAGGAAATCGCCCTGGAA | CGTTTTCCAGGGCGATTTCCTGGGGGTTGGG<br>GTCGGTGGGCACGCAGGCGTGGGTGGC |
| 86 | L | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCAACCCCCAGGAAATCCACGCCGAA | CGTTTTCGGCGTGGATTTCCTGGGGGTTGGG<br>GTCGGTGGGCACGCAGGCGTGGGTGGC |
| 87 | E | GTGGGCCACCCACGCCTGCGTGCCCACCGA<br>CCCCAACCCCCAGGAAATCCACCTGGCC | CGTTGGCCAGGTGGATTTCCTGGGGGTTGG<br>GGTCGGTGGGCACGCAGGCGTGGGTGGC |
| 88 | N | GGAAGCCGTGACCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGGC |
| 89 | V | GGAAAACGCCACCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTGGCGTT |
| 90 | T | GGAAAACGTGGCCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGCCACGTT |
| 91 | E | GGAAAACGTGACCGCCGAGTTCAACATGTG<br>GAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGAACTCGGCGGTCACGTT |
| 92 | E | GGAAAACGTGACCGAAGCCTTCAACATGTG<br>GAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGAAGGCTTCGGTCACGTT |
| 93 | F | GGAAAACGTGACCGAAGAGGCCAACATGT<br>GGAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGGCCTCTTCGGTCACGTT |
| 94 | N | GGAAAACGTGACCGAAGAGTTCGCCATGTG<br>GAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCACATGGCGAACTCTTCGGTCACGTT |
| 95 | M | GGAAAACGTGACCGAAGAGTTCAACGCCTG<br>GAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>CCAGGCGTTGAACTCTTCGGTCACGTT |

| | | | |
|---|---|---|---|
| 96 | W | GGAAAACGTGACCGAAGAGTTCAACATGGC<br>CAAGAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTCTT<br>GGCCATGTTGAACTCTTCGGTCACGTT |
| 97 | K | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GGCCAACAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGTTGGC<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 98 | N | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGGCCAACATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGTTGGCCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 99 | N | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGAACGCCATGGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACCATGGCGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 100 | M | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGAACAACGCCGTGGAACAGATGCAC | CGGTGTGCATCTGTTCCACGGCGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 101 | V | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGCCGAACAGATGCAC | CGGTGTGCATCTGTTCGGCCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 102 | E | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGTGGCCCAGATGCAC | CGGTGTGCATCTGGGCCACCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 103 | Q | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGTGGAAGCCATGCAC | CGGTGTGCATGGCTTCCACCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 104 | M | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGTGGAACAGGCCCAC | CGGTGTGGGCCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 105 | H | GGAAAACGTGACCGAAGAGTTCAACATGTG<br>GAAGAACAACATGGTGGAACAGATGGCC | CGGTGGCCATCTGTTCCACCATGTTGTTCTT<br>CCACATGTTGAACTCTTCGGTCACGTT |
| 106 | T | GCACGCCGACATCATCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGATGATGTCGGC |
| 107 | D | GCACACCGCCATCATCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGATGATGGCGGT |
| 108 | I | GCACACCGACGCCATCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGATGGCGTCGGT |
| 109 | I | GCACACCGACATCGCCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGGCGATGTCGGT |
| 110 | S | GCACACCGACATCATCGCCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGGCGATGATGTCGGT |
| 111 | L | GCACACCGACATCATCAGCGCCTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCAGGCGCTGATGATGTCGGT |
| 112 | W | GCACACCGACATCATCAGCCTGGCCGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCGGCCAGGCTGATGATGTCGGT |
| 113 | D | GCACACCGACATCATCAGCCTGTGGGCCCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGGCCCACAGGCTGATGATGTCGGT |
| 114 | Q | GCACACCGACATCATCAGCCTGTGGGACGC<br>CAGCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGC<br>TGGCGTCCCACAGGCTGATGATGTCGGT |
| 115 | S | GCACACCGACATCATCAGCCTGTGGGACCA<br>GGCCCTGAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTCAGGG<br>CCTGGTCCCACAGGCTGATGATGTCGGT |
| 116 | L | GCACACCGACATCATCAGCCTGTGGGACCA<br>GAGCGCCAAGCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGCTTGGCGC<br>TCTGGTCCCACAGGCTGATGATGTCGGT |
| 117 | K | GCACACCGACATCATCAGCCTGTGGGACCA<br>GAGCCTGGCCCCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGGGGCCAGGC<br>TCTGGTCCCACAGGCTGATGATGTCGGT |
| 118 | P | GCACACCGACATCATCAGCCTGTGGGACCA<br>GAGCCTGAAGGCCTGCGTGAAGCTGACC | GGGGGGTCAGCTTCACGCAGGCCTTCAGGC<br>TCTGGTCCCACAGGCTGATGATGTCGGT |
| 120 | V | GCACACCGACATCATCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGCCAAGCTGACC | GGGGGGTCAGCTTGGCGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGATGATGTCGGT |
| 121 | K | GCACACCGACATCATCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGGCCCTGACC | GGGGGGTCAGGGCCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGATGATGTCGGT |
| 122 | L | GCACACCGACATCATCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGGCCACC | GGGGGGTGGCCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGATGATGTCGGT |
| 123 | T | GCACACCGACATCATCAGCCTGTGGGACCA<br>GAGCCTGAAGCCCTGCGTGAAGCTGGCC | GGGGGGCCAGCTTCACGCAGGGCTTCAGGC<br>TCTGGTCCCACAGGCTGATGATGTCGGT |
| 124 | P | GACCGCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGC |
| 125 | L | GACCCCCGCCTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGTCACGCAGGCGGG |
| 127 | V | GACCCCCCTGTGCGCCACCCTGCAGTGCAC<br>CAACGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGTGGCGCACAGGGG |
| 128 | T | GACCCCCCTGTGCGTGGCCCTGCAGTGCAC<br>CAACGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGCCACGCACAGGGG |
| 129 | L | GACCCCCCTGTGCGTGACCGCCCAGTGCAC<br>CAACGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGTT<br>GGTGCACTGGGCGGTCACGCACAGGGG |
| 130 | Q | GACCCCCCTGTGCGTGACCCTGGCCTGCAC<br>CAACGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGTT<br>GGTGCAGGCCAGGGTCACGCACAGGGG |
| 132 | T | GACCCCCCTGTGCGTGACCCTGCAGTGCGC<br>CAACGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGTT<br>GGCGCACTGCAGGGTCACGCACAGGGG |
| 133 | N | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CGCCGTGACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTCACGGC<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 134 | V | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGCCACCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGTGGCGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 135 | T | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGGCCAACAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGTTGGCCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |

| | | | |
|---|---|---|---|
| 136 | N | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCGCCAACATCACCGACGAC | GCATGTCGTCGGTGATGTTGGCGGTCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 137 | N | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCAACGCCATCACCGACGAC | GCATGTCGTCGGTGATGGCGTTGGTCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 138 | I | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCAACAACGCCACCGACGAC | GCATGTCGTCGGTGGCGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 139 | T | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCAACAACATCGCCGACGAC | GCATGTCGTCGGCGATGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 140 | D | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCAACAACATCACCGCCGAC | GCATGTCGGCGGTGATGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 141 | D | GACCCCCCTGTGCGTGACCCTGCAGTGCAC<br>CAACGTGACCAACAACATCACCGACGCC | GCATGGCGTCGGTGATGTTGTTGGTCACGTT<br>GGTGCACTGCAGGGTCACGCACAGGGG |
| 150 | M | CGACGCCCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCCCGGGC |
| 151 | R | CGACATGGCCGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCGGCCAT |
| 152 | G | CGACATGCGGGCCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGGCCCGCAT |
| 153 | E | CGACATGCGGGGCGCCCTGAAGAACTGCAG<br>CTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTCAGGGCGCCCCGCAT |
| 154 | L | CGACATGCGGGGCGAGGCCAAGAACTGCA<br>GCTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTGGCCTCGCCCCGCAT |
| 155 | K | CGACATGCGGGGCGAGCTGGCCAACTGCAG<br>CTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTGGCCAGCTCGCCCCGCAT |
| 156 | N | CGACATGCGGGGCGAGCTGAAGGCCTGCAG<br>CTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGGCCTTCAGCTCGCCCCGCAT |
| 158 | S | CGACATGCGGGGCGAGCTGAAGAACTGCGC<br>CTTCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGAA<br>GGCGCAGTTCTTCAGCTCGCCCCGCAT |
| 159 | F | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CGCCAACATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGTTGGC<br>GCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 160 | N | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCGCCATGACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTCATGGCGA<br>AGCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 161 | M | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACGCCACCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGTGGCGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 162 | T | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGGCCACCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGTGGCCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 163 | T | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCGCCGAGCTGCGGGAC | TCTTGTCCCGCAGCTCGGCGGTCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 164 | E | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCACCGCCCTGCGGGAC | TCTTGTCCCGCAGGGCGGTGGTCATGTTGA<br>AGCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 165 | L | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCACCGAGGCCCGGGAC | TCTTGTCCCGGGCCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 166 | R | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCACCGAGCTGGCCGAC | TCTTGTCGGCCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 167 | D | CGACATGCGGGGCGAGCTGAAGAACTGCAG<br>CTTCAACATGACCACCGAGCTGCGGGCC | TCTTGGCCCGCAGCTCGGTGGTCATGTTGAA<br>GCTGCAGTTCTTCAGCTCGCCCCGCAT |
| 168 | K | GGACGCCAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAACAGGCTGTACACCTTCTGTTTGGC |
| 169 | K | GGACAAGGCCCAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAACAGGCTGTACACCTTCTGGGCCTT |
| 170 | Q | GGACAAGAAAGCCAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAACAGGCTGTACACCTTGGCTTTCTT |
| 171 | K | GGACAAGAAACAGGCCGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAACAGGCTGTACACGGCCTGTTTCTT |
| 172 | V | GGACAAGAAACAGAAGGCCTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAACAGGCTGTAGGCCTTCTGTTTCTT |
| 173 | Y | GGACAAGAAACAGAAGGTGGCCAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAACAGGCTGGCCACCTTCTGTTTCTT |
| 174 | S | GGACAAGAAACAGAAGGTGTACGCCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAACAGGGCGTACACCTTCTGTTTCTT |
| 175 | L | GGACAAGAAACAGAAGGTGTACAGCGCCTT<br>CTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GAAGGCGCTGTACACCTTCTGTTTCTT |
| 176 | F | GGACAAGAAACAGAAGGTGTACAGCCTGG<br>CCTACCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGTA<br>GGCCAGGCTGTACACCTTCTGTTTCTT |
| 177 | Y | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CGCCCGGCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGCCGGGC<br>GAACAGGCTGTACACCTTCTGTTTCTT |
| 178 | R | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACGCCCTGGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCCAGGGCGTA<br>GAACAGGCTGTACACCTTCTGTTTCTT |
| 179 | L | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGGCCGACGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGTCGGCCCGGTA<br>GAACAGGCTGTACACCTTCTGTTTCTT |
| 180 | D | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGCCGTGGTGCAGATCAAC | TCTCGTTGATCTGCACCACGGCCAGCCGGT<br>AGAACAGGCTGTACACCTTCTGTTTCTT |
| 181 | V | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGCCGTGCAGATCAAC | TCTCGTTGATCTGCACGGCGTCCAGCCGGTA<br>GAACAGGCTGTACACCTTCTGTTTCTT |

| | | | |
|---|---|---|---|
| 182 | V | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGCCCAGATCAAC | TCTCGTTGATCTGGGCCACGTCCAGCCGGTA<br>GAACAGGCTGTACACCTTCTGTTTCTT |
| 183 | Q | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGGCCATCAAC | TCTCGTTGATGGCCACCACGTCCAGCCGGT<br>AGAACAGGCTGTACACCTTCTGTTTCTT |
| 184 | I | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGGCCAAC | TCTCGTTGGCCTGCACCACGTCCAGCCGGTA<br>GAACAGGCTGTACACCTTCTGTTTCTT |
| 185 | N | GGACAAGAAACAGAAGGTGTACAGCCTGTT<br>CTACCGGCTGGACGTGGTGCAGATCGCC | TCTCGGCGATCTGCACCACGTCCAGCCGGT<br>AGAACAGGCTGTACACCTTCTGTTTCTT |
| 186 | E | CAACGCCAACCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTGGC |
| N/A | N | CAACGAGGCCCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGGCCTC |
| N/A | Q | CAACGAGAACGCCGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCGGCGTTCTC |
| N/A | G | CAACGAGAACCAGGCCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGGCCTGGTTCTC |
| N/A | N | CAACGAGAACCAGGGCGCCAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGGCGCCCTGGTTCTC |
| N/A | R | CAACGAGAACCAGGGCAACGCCAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTGGCGTTGCCCTGGTTCTC |
| N/A | S | CAACGAGAACCAGGGCAACAGAGCCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGGCTCTGTTGCCCTGGTTCTC |
| N/A | N | CAACGAGAACCAGGGCAACAGAAGCGCCA<br>ACAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GTTGGCGCTTCTGTTGCCCTGGTTCTC |
| N/A | N | CAACGAGAACCAGGGCAACAGAAGCAACG<br>CCAGCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGCT<br>GGCGTTGCTTCTGTTGCCCTGGTTCTC |
| 187 | S | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACGCCAACAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGTTGGC<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 188 | N | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCGCCAAAGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCTTTGGCGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 189 | K | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCAACGCCGAGTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTACTCGGCGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 190 | E | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGCCTACCGGCTGATCAAC | TGCAGTTGATCAGCCGGTAGGCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 191 | Y | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGGCCCGGCTGATCAAC | TGCAGTTGATCAGCCGGGCCTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 192 | R | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACGCCCTGATCAAC | TGCAGTTGATCAGGGCGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 193 | L | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGGCCATCAAC | TGCAGTTGATGGCCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 194 | I | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGGCCAAC | TGCAGTTGGCCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 195 | N | CAACGAGAACCAGGGCAACAGAAGCAACA<br>ACAGCAACAAAGAGTACCGGCTGATCGCC | TGCAGGCGATCAGCCGGTACTCTTTGTTGCT<br>GTTGTTGCTTCTGTTGCCCTGGTTCTC |
| 197 | N | CAACTGCGCCACCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGTGGCGCA |
| 198 | T | CAACTGCAACGCCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGCGTTGCA |
| 199 | S | CAACTGCAACACCGCCGCCATCACCCAGGC<br>CTGCCCCAAGGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACCTTGGGGC<br>AGGCCTGGGTGATGGCGGCGGTGTTGCA |
| 201 | I | CAACTGCAACACCAGCGCCGCCACCCAGGC<br>CTGCCCCAAGGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACCTTGGGGC<br>AGGCCTGGGTGGCGGCGCTGGTGTTGCA |
| 202 | T | CAACTGCAACACCAGCGCCATCGCCCAGGC<br>CTGCCCCAAGGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACCTTGGGGC<br>AGGCCTGGGCGATGGCGCTGGTGTTGCA |
| 203 | Q | CAACTGCAACACCAGCGCCATCACCGCCGC<br>CTGCCCCAAGGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACCTTGGGGC<br>AGGCGGCGGTGATGGCGCTGGTGTTGCA |
| 206 | P | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCGCCAAGGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACCTTGGCGC<br>AGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 207 | K | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCCCCGCCGTGTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGACACGGCGGGG<br>CAGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 208 | V | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGCCTCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGAGGCCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 209 | S | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGTGGCCTTCGAGCCCATC | TGGGGATGGGCTCGAAGGCCACCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 210 | F | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGTGTCCGCCGAGCCCATC | TGGGGATGGGCTCGGCGGACACCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 211 | E | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGTGTCCTTCGCCCCCATC | TGGGGATGGGGGCGAAGGACACCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 212 | P | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGTGTCCTTCGAGGCCATC | TGGGGATGGCCTCGAAGGACACCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 213 | I | CAACTGCAACACCAGCGCCATCACCCAGGC<br>CTGCCCCAAGGTGTCCTTCGAGCCCGCC | TGGGGGCGGGCTCGAAGGACACCTTGGGGC<br>AGGCCTGGGTGATGGCGCTGGTGTTGCA |
| 214 | P | CATCGCCATCCACTACTGCGCCCCTGCCGGC<br>TTCGCCATCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGATGGCGAA<br>GCCGGCAGGGGCGCAGTAGTGGATGGC |

| | | | |
|---|---|---|---|
| 215 | I | CATCCCCGCCCACTACTGCGCCCCTGCCGGC<br>TTCGCCATCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGATGGCGAA<br>GCCGGCAGGGGCGCAGTAGTGGGCGGG |
| 216 | H | CATCCCCATCGCCTACTGCGCCCCTGCCGGC<br>TTCGCCATCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGATGGCGAA<br>GCCGGCAGGGGCGCAGTAGGCGATGGG |
| 217 | Y | CATCCCCATCCACGCCTGCGCCCCTGCCGGC<br>TTCGCCATCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGATGGCGAA<br>GCCGGCAGGGGCGCAGGCGTGGATGGG |
| 220 | P | CATCCCCATCCACTACTGCGCCGCCGCCGG<br>CTTCGCCATCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGATGGCGAA<br>GCCGGCGGCGGCGCAGTAGTGGATGGG |
| 222 | G | CATCCCCATCCACTACTGCGCCCCTGCCGCC<br>TTCGCCATCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGATGGCGAA<br>GGCGGCAGGGGCGCAGTAGTGGATGGG |
| 223 | F | CATCCCCATCCACTACTGCGCCCCTGCCGGC<br>GCCGCCATCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGATGGCGGC<br>GCCGGCAGGGGCGCAGTAGTGGATGGG |
| 225 | I | CATCCCCATCCACTACTGCGCCCCTGCCGGC<br>TTCGCCGCCCTGAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTCAGGGCGGCGAA<br>GCCGGCAGGGGCGCAGTAGTGGATGGG |
| 226 | L | CATCCCCATCCACTACTGCGCCCCTGCCGGC<br>TTCGCCATCGCCAAGTGCAAGGACAAG | ACTTCTTGTCCTTGCACTTGGCGATGGCGAA<br>GCCGGCAGGGGCGCAGTAGTGGATGGG |
| 227 | K | CATCCCCATCCACTACTGCGCCCCTGCCGGC<br>TTCGCCATCCTGGCCTGCAAGGACAAG | ACTTCTTGTCCTTGCAGGCCAGGATGGCGA<br>AGCCGGCAGGGGCGCAGTAGTGGATGGG |
| 229 | K | CATCCCCATCCACTACTGCGCCCCTGCCGGC<br>TTCGCCATCCTGAAGTGCGCCGACAAG | ACTTCTTGTCGGCGCACTTCAGGATGGCGA<br>AGCCGGCAGGGGCGCAGTAGTGGATGGG |
| 230 | D | CATCCCCATCCACTACTGCGCCCCTGCCGGC<br>TTCGCCATCCTGAAGTGCAAGGCCAAG | ACTTCTTGGCCTTGCACTTCAGGATGGCGAA<br>GCCGGCAGGGGCGCAGTAGTGGATGGG |
| 231 | K | CATCCCCATCCACTACTGCGCCCCTGCCGGC<br>TTCGCCATCCTGAAGTGCAAGGACGCC | ACTTGGCGTCCTTGCACTTCAGGATGGCGA<br>AGCCGGCAGGGGCGCAGTAGTGGATGGG |
| 232 | K | CAAGGCCTTCAACGGCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGGGCCGGTGCCGTTGAAGGC |
| 233 | F | CAAGAAGGCCAACGGCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGGGCCGGTGCCGTTGGCCTT |
| 234 | N | CAAGAAGTTCGCCGGCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGGGCCGGTGCCGGCGAACTT |
| 235 | G | CAAGAAGTTCAACGCCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGGGCCGGTGGCGTTGAACTT |
| 236 | T | CAAGAAGTTCAACGGCGCCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGGGCCGGCGCCGTTGAACTT |
| 237 | G | CAAGAAGTTCAACGGCACCGCCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGGGGCGGTGCCGTTGAACTT |
| 238 | P | CAAGAAGTTCAACGGCACCGGCGCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGCGCCGGTGCCGTTGAACTT |
| 240 | P | CAAGAAGTTCAACGGCACCGGCCCCTGCGC<br>CAGCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGC<br>TGGCGCAGGGGCCGGTGCCGTTGAACTT |
| 241 | S | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CGCCGTGTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGACACGG<br>CGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 242 | V | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CAGCGCCTCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGAGGCGC<br>TGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 243 | S | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CAGCGTGGCCACCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGTGGCCACGC<br>TGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 244 | T | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CAGCGTGTCCGCCGTGCAGTGTACCCAC | TGCCGTGGGTACACTGCACGGCGGACACGC<br>TGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 245 | V | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGCCCAGTGTACCCAC | TGCCGTGGGTACACTGGGCGGTGGACACGC<br>TGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 246 | Q | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGGCCTGTACCCAC | TGCCGTGGGTACAGGCCACGGTGGACACGC<br>TGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 248 | T | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTGCCCAC | TGCCGTGGGCACACTGCACGGTGGACACGC<br>TGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 249 | H | CAAGAAGTTCAACGGCACCGGCCCCTGCCC<br>CAGCGTGTCCACCGTGCAGTGTACCGCC | TGCCGGCGGTACACTGCACGGTGGACACGC<br>TGGGGCAGGGGCCGGTGCCGTTGAACTT |
| 250 | G | CCACGCCATCAAGCCCGTGGTGTCCACCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGTGGACACCACGGGCTTGATGGC |
| 251 | I | CCACGGCGCCAAGCCCGTGGTGTCCACCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGTGGACACCACGGGCTTGGCGCC |
| 252 | K | CCACGGCATCGCCCCCGTGGTGTCCACCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGTGGACACCACGGGGGCGATGCC |
| 253 | P | CCACGGCATCAAGGCCGTGGTGTCCACCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGTGGACACCACGGCCTTGATGCC |
| 254 | V | CCACGGCATCAAGCCCGCCGTGTCCACCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGTGGACACGGCGGGCTTGATGCC |
| 255 | V | CCACGGCATCAAGCCCGTGGCCTCCACCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGTGGAGGCCACGGGCTTGATGCC |
| 256 | S | CCACGGCATCAAGCCCGTGGTGGCCACCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGTGGCCACCACGGGCTTGATGCC |
| 257 | T | CCACGGCATCAAGCCCGTGGTGTCCGCCCA<br>GCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GCTGGGCGGACACCACGGGCTTGATGCC |
| 258 | Q | CCACGGCATCAAGCCCGTGGTGTCCACCGC<br>CCTGCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGCA<br>GGGCGGTGGACACCACGGGCTTGATGCC |

| | | | |
|---|---|---|---|
| 259 | L | CCACGGCATCAAGCCCGTGGTGTCCACCCA GGCCCTGCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGCAGGG CCTGGGTGGACACCACGGGCTTGATGCC |
| 260 | L | CCACGGCATCAAGCCCGTGGTGTCCACCCA GCTGGCCCTGAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTCAGGGCCA GCTGGGTGGACACCACGGGCTTGATGCC |
| 261 | L | CCACGGCATCAAGCCCGTGGTGTCCACCCA GCTGCTGGCCAACGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGTTGGCCAGCA GCTGGGTGGACACCACGGGCTTGATGCC |
| 262 | N | CCACGGCATCAAGCCCGTGGTGTCCACCCA GCTGCTGCTGGCCGGCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGCCGGCCAGCAGCA GCTGGGTGGACACCACGGGCTTGATGCC |
| 263 | G | CCACGGCATCAAGCCCGTGGTGTCCACCCA GCTGCTGCTGAACGCCAGCCTCGCCGAA | CCTCTTCGGCGAGGCTGGCGTTCAGCAGCA GCTGGGTGGACACCACGGGCTTGATGCC |
| 264 | S | CCACGGCATCAAGCCCGTGGTGTCCACCCA GCTGCTGCTGAACGGCGCCCTCGCCGAA | CCTCTTCGGCGAGGGCGCCGTTCAGCAGCA GCTGGGTGGACACCACGGGCTTGATGCC |
| 265 | L | CCACGGCATCAAGCCCGTGGTGTCCACCCA GCTGCTGCTGAACGGCAGCGCCGCCGAA | CCTCTTCGGCGGCGCTGCCGTTCAGCAGCA GCTGGGTGGACACCACGGGCTTGATGCC |
| 267 | E | CCACGGCATCAAGCCCGTGGTGTCCACCCA GCTGCTGCTGAACGGCAGCCTCGCCGCC | CCTCGGCGGCGAGGCTGCCGTTCAGCAGCA GCTGGGTGGACACCACGGGCTTGATGCC |
| 268 | E | CGAAGCCGAAGTGATGATCAGAAGCGAGA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTTCTGATCATCACTTCGGC |
| 269 | E | CGAAGAGGCCGTGATGATCAGAAGCGAGA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTTCTGATCATCACGGCCTC |
| 270 | V | CGAAGAGGAAGCCATGATCAGAAGCGAGA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTTCTGATCATGGCTTCCTC |
| 271 | M | CGAAGAGGAAGTGGCCATCAGAAGCGAGA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTTCTGATGGCCACTTCCTC |
| 272 | I | CGAAGAGGAAGTGATGGCCAGAAGCGAGA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTTCTGGCCATCACTTCCTC |
| 273 | R | CGAAGAGGAAGTGATGATCGCCAGCGAGA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTGGCGATCATCACTTCCTC |
| 274 | S | CGAAGAGGAAGTGATGATCAGAGCCGAGA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGGCTCTGATCATCACTTCCTC |
| 275 | E | CGAAGAGGAAGTGATGATCAGAAGCGCCA ACATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GTTGGCGCTTCTGATCATCACTTCCTC |
| 276 | N | CGAAGAGGAAGTGATGATCAGAAGCGAGG CCATCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGAT GGCCTCGCTTCTGATCATCACTTCCTC |
| 277 | I | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACGCCACCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGTGGC GTTCTCGCTTCTGATCATCACTTCCTC |
| 278 | T | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACATCGCCAACAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGTTGGCGAT GTTCTCGCTTCTGATCATCACTTCCTC |
| 279 | N | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACATCACCGCCAACGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGTTGGCGGTGA TGTTCTCGCTTCTGATCATCACTTCCTC |
| 280 | N | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACATCACCAACGCCGCCAAGAACATCCTG | GCACCAGGATGTTCTTGGCGGCGTTGGTGA TGTTCTCGCTTCTGATCATCACTTCCTC |
| 282 | K | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACATCACCAACAACGCCGCCAACATCCTG | GCACCAGGATGTTGGCGGCGTTGTTGGTGA TGTTCTCGCTTCTGATCATCACTTCCTC |
| 283 | N | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACATCACCAACAACGCCAAGGCCATCCTG | GCACCAGGATGGCCTTGGCGTTGTTGGTGA TGTTCTCGCTTCTGATCATCACTTCCTC |
| 284 | I | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACATCACCAACAACGCCAAGAACGCCCTG | GCACCAGGGCGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTTCTGATCATCACTTCCTC |
| 285 | L | CGAAGAGGAAGTGATGATCAGAAGCGAGA ACATCACCAACAACGCCAAGAACATCGCC | GCACGGCGATGTTCTTGGCGTTGTTGGTGAT GTTCTCGCTTCTGATCATCACTTCCTC |
| 286 | V | CCTGGCCCAGTTCAACACCCCCGTGCAGAT TAACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATCTGCACGGGGGTGTTGAACTGGGC |
| 287 | Q | CCTGGTGGCCTTCAACACCCCCGTGCAGATT AACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATCTGCACGGGGGTGTTGAAGGCCAC |
| 288 | F | CCTGGTGCAGGCCAACACCCCCGTGCAGAT TAACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATCTGCACGGGGGTGTTGGCCTGCAC |
| 289 | N | CCTGGTGCAGTTCGCCACCCCCGTGCAGATT AACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATCTGCACGGGGGTGGCGAACTGCAC |
| 290 | T | CCTGGTGCAGTTCAACGCCCCCGTGCAGAT TAACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATCTGCACGGGGGCGTTGAACTGCAC |
| 291 | P | CCTGGTGCAGTTCAACACCGCCGTGCAGAT TAACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATCTGCACGGCGGTGTTGAACTGCAC |
| 292 | V | CCTGGTGCAGTTCAACACCCCCGCCCAGAT TAACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATCTGGGCGGGGGTGTTGAACTGCAC |
| 293 | Q | CCTGGTGCAGTTCAACACCCCCGTGGCCATT AACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT AATGGCCACGGGGGTGTTGAACTGCAC |
| 294 | I | CCTGGTGCAGTTCAACACCCCCGTGCAGGC CAACTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGTT GGCCTGCACGGGGGTGTTGAACTGCAC |
| 295 | N | CCTGGTGCAGTTCAACACCCCCGTGCAGAT TGCCTGCACCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGTGCAGGC AATCTGCACGGGGGTGTTGAACTGCAC |
| 297 | T | CCTGGTGCAGTTCAACACCCCCGTGCAGAT TAACTGCGCCCGGCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGCCGGGCGCAGTT AATCTGCACGGGGGTGTTGAACTGCAC |
| 298 | R | CCTGGTGCAGTTCAACACCCCCGTGCAGAT TAACTGCACCGCCCCCAACAACAACACC | TTCTGGTGTTGTTGTTGGGGGCGGTGCAGTT AATCTGCACGGGGGTGTTGAACTGCAC |

| | | | |
|---|---|---|---|
| 299 | P | CCTGGTGCAGTTCAACACCCCCGTGCAGAT<br>TAACTGCACCCGGGCCAACAACAACACC | TTCTGGTGTTGTTGTTGGCCCGGGTGCAGTT<br>AATCTGCACGGGGGTGTTGAACTGCAC |
| 300 | N | CCTGGTGCAGTTCAACACCCCCGTGCAGAT<br>TAACTGCACCCGGCCCGCCAACAACACC | TTCTGGTGTTGTTGGCGGGCCGGGTGCAGTT<br>AATCTGCACGGGGGTGTTGAACTGCAC |
| 301 | N | CCTGGTGCAGTTCAACACCCCCGTGCAGAT<br>TAACTGCACCCGGCCCAACGCCAACACC | TTCTGGTGTTGGCGTTGGGCCGGGTGCAGTT<br>AATCTGCACGGGGGTGTTGAACTGCAC |
| 302 | N | CCTGGTGCAGTTCAACACCCCCGTGCAGAT<br>TAACTGCACCCGGCCCAACAACGCCACC | TTCTGGTGGCGTTGTTGGGCCGGGTGCAGTT<br>AATCTGCACGGGGGTGTTGAACTGCAC |
| 303 | T | CCTGGTGCAGTTCAACACCCCCGTGCAGAT<br>TAACTGCACCCGGCCCAACAACAACGCC | TTCTGGCGTTGTTGTTGGGCCGGGTGCAGTT<br>AATCTGCACGGGGGTGTTGAACTGCAC |
| 304 | R | CACCGCCAAGAGCATCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGATGCTCTTGGC |
| 305 | K | CACCAGAGCCAGCATCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGATGCTGGCTCT |
| 306 | S | CACCAGAAAGGCCATCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGATGGCCTTTCT |
| 307 | I | CACCAGAAAGAGCGCCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGGCGCTCTTTCT |
| 308 | R | CACCAGAAAGAGCATCGCCATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATGGCGATGCTCTTTCT |
| 309 | I | CACCAGAAAGAGCATCCGGGCCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGGCCCGGATGCTCTTTCT |
| 312 | G | CACCAGAAAGAGCATCCGGATCGCCCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGGGCGATCCGGATGCTCTTTCT |
| 313 | P | CACCAGAAAGAGCATCCGGATCGGGGCCGG<br>GCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCGGCCCCGATCCGGATGCTCTTTCT |
| 314 | G | CACCAGAAAGAGCATCCGGATCGGGCCTGC<br>CCAGGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCCT<br>GGGCAGGCCCGATCCGGATGCTCTTTCT |
| 315 | Q | CACCAGAAAGAGCATCCGGATCGGGCCTGG<br>GGCCGCCTTCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGAAGGCGG<br>CCCCAGGCCCGATCCGGATGCTCTTTCT |
| 317 | F | CACCAGAAAGAGCATCCGGATCGGGCCTGG<br>GCAGGCCGCCTACGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGTAGGCGGCCT<br>GCCCAGGCCCGATCCGGATGCTCTTTCT |
| 318 | Y | CACCAGAAAGAGCATCCGGATCGGGCCTGG<br>GCAGGCCTTCGCCGCCACCGGCGACATC | CGATGATGTCGCCGGTGGCGGCGAAGGCCT<br>GCCCAGGCCCGATCCGGATGCTCTTTCT |
| 320 | T | CACCAGAAAGAGCATCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCGCCGGCGACATC | CGATGATGTCGCCGGCGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGATGCTCTTTCT |
| 321 | G | CACCAGAAAGAGCATCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGCCGACATC | CGATGATGTCGGCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGATGCTCTTTCT |
| N/A | D | CACCAGAAAGAGCATCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGCCATC | CGATGATGGCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGATGCTCTTTCT |
| 322 | I | CACCAGAAAGAGCATCCGGATCGGGCCTGG<br>GCAGGCCTTCTACGCCACCGGCGACGCC | CGATGGCGTCGCCGGTGGCGTAGAAGGCCT<br>GCCCAGGCCCGATCCGGATGCTCTTTCT |
| 323 | I | CATCGCCGGCGACATCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGTT<br>GCAGTGGGCCTGCCGGATGTCGCCGGC |
| 324 | G | CATCATCGCCGACATCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGTT<br>GCAGTGGGCCTGCCGGATGTCGGCGAT |
| 325 | D | CATCATCGGCGCCATCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGTT<br>GCAGTGGGCCTGCCGGATGGCGCCGAT |
| 326 | I | CATCATCGGCGACGCCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGTT<br>GCAGTGGGCCTGCCGGGCGTCGCCGAT |
| 327 | R | CATCATCGGCGACATCGCCCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGTT<br>GCAGTGGGCCTGGGCGATGTCGCCGAT |
| 328 | Q | CATCATCGGCGACATCCGGGCCGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGTT<br>GCAGTGGGCGGCCCGGATGTCGCCGAT |
| 330 | H | CATCATCGGCGACATCCGGCAGGCCGCCTG<br>CAACGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGTT<br>GCAGGCGGCCTGCCGGATGTCGCCGAT |
| 332 | N | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CGCCGTGTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGACACGG<br>CGCAGTGGGCCTGCCGGATGTCGCCGAT |
| 333 | V | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CAACGCCTCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGAGGCGTT<br>GCAGTGGGCCTGCCGGATGTCGCCGAT |
| 334 | S | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CAACGTGGCCAAGGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCCTTGGCCACGTT<br>GCAGTGGGCCTGCCGGATGTCGCCGAT |
| 335 | K | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CAACGTGTCCGCCGCCACCTGGAACGAG | GTGTCTCGTTCCAGGTGGCGGCGGACACGT<br>TGCAGTGGGCCTGCCGGATGTCGCCGAT |
| 337 | T | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCGCCTGGAACGAG | GTGTCTCGTTCCAGGCGGCCTTGGACACGTT<br>GCAGTGGGCCTGCCGGATGTCGCCGAT |
| 338 | W | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCGCCAACGAG | GTGTCTCGTTGGCGGTGGCCTTGGACACGTT<br>GCAGTGGGCCTGCCGGATGTCGCCGAT |
| 339 | N | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGGCCGAG | GTGTCTCGGCCCAGGTGGCCTTGGACACGT<br>TGCAGTGGGCCTGCCGGATGTCGCCGAT |
| 340 | E | CATCATCGGCGACATCCGGCAGGCCCACTG<br>CAACGTGTCCAAGGCCACCTGGAACGCC | GTGTGGCGTTCCAGGTGGCCTTGGACACGT<br>TGCAGTGGGCCTGCCGGATGTCGCCGAT |
| 341 | T | CGAGGCCCTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTCACCACCTTGCCCAGGGC |

| | | | |
|---|---|---|---|
| 342 | L | CGAGACAGCCGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTCACCACCTTGCCGGCTGT |
| 343 | G | CGAGACACTGGCCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTCACCACCTTGGCCAGTGT |
| 344 | K | CGAGACACTGGGCGCCGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTCACCACGGCGCCCAGTGT |
| 345 | V | CGAGACACTGGGCAAGGCCGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTCACGGCCTTGCCCAGTGT |
| 346 | V | CGAGACACTGGGCAAGGTGGCCAAACAGCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTGGCCACCTTGCCCAGTGT |
| 347 | K | CGAGACACTGGGCAAGGTGGTGGCCCAGCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGGGCCACCACCTTGCCCAGTGT |
| 348 | Q | CGAGACACTGGGCAAGGTGGTGAAAGCCCT<br>GCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGGGCTTTCACCACCTTGCCCAGTGT |
| 349 | L | CGAGACACTGGGCAAGGTGGTGAAACAGG<br>CCCGGAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTCCG<br>GGCCTGTTTCACCACCTTGCCCAGTGT |
| 350 | R | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GGCCAAGCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGCTTGGC<br>CAGCTGTTTCACCACCTTGCCCAGTGT |
| 351 | K | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGGCCCACTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGTGGGCCC<br>GCAGCTGTTTCACCACCTTGCCCAGTGT |
| 352 | H | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGGCCTTCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGAAGGCCTTCCG<br>CAGCTGTTTCACCACCTTGCCCAGTGT |
| 353 | F | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACGCCGGGAACAACACCATC | TGATGATGGTGTTGTTCCCGGCGTGCTTCCG<br>CAGCTGTTTCACCACCTTGCCCAGTGT |
| 354 | G | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGCCAACAACACCATC | TGATGATGGTGTTGTTGGCGAAGTGCTTCCG<br>CAGCTGTTTCACCACCTTGCCCAGTGT |
| 355 | N | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGGCCAACACCATC | TGATGATGGTGTTGGCCCCGAAGTGCTTCC<br>GCAGCTGTTTCACCACCTTGCCCAGTGT |
| 356 | N | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACGCCACCATC | TGATGATGGTGGCGTTCCCGAAGTGCTTCC<br>GCAGCTGTTTCACCACCTTGCCCAGTGT |
| 357 | T | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACAACGCCATC | TGATGATGGCGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTCACCACCTTGCCCAGTGT |
| 358 | I | CGAGACACTGGGCAAGGTGGTGAAACAGCT<br>GCGGAAGCACTTCGGGAACAACACCGCC | TGATGGCGGTGTTGTTCCCGAAGTGCTTCCG<br>CAGCTGTTTCACCACCTTGCCCAGTGT |
| 359 | I | CATCGCCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGGC |
| 360 | R | CATCATCGCCTTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAAGGCGAT |
| 361 | F | CATCATCAGAGCCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGGCTCTGAT |
| 363 | N | CATCATCAGATTCGCCGCCAGCTCTGGCGG<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGGCGGCGAATCTGAT |
| 364 | S | CATCATCAGATTCGCCAACGCCTCTGGCGG<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGGCGTTGGCGAATCTGAT |
| 365 | S | CATCATCAGATTCGCCAACAGCGCCGGCGG<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGCCGGCGCTGTTGGCGAATCTGAT |
| 366 | G | CATCATCAGATTCGCCAACAGCTCTGCCGG<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGGCAGAGCTGTTGGCGAATCTGAT |
| 367 | G | CATCATCAGATTCGCCAACAGCTCTGGCGC<br>CGACCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGT<br>CGGCGCCAGAGCTGTTGGCGAATCTGAT |
| 368 | D | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGCCCTGGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCCAGGG<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 369 | L | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACGCCGAAGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACTTCGGCGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 370 | E | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGCCGTGACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTCACGGCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 371 | V | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGCCACCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGTGGCTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 372 | T | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGGCCACCCACAGCTTC | AGTTGAAGCTGTGGGTGGCCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 373 | T | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGACCGCCCACAGCTTC | AGTTGAAGCTGTGGGCGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 374 | H | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGACCACCGCCAGCTTC | AGTTGAAGCTGGCGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 375 | S | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGACCACCCACGCCTTC | AGTTGAAGGCGTGGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 376 | F | CATCATCAGATTCGCCAACAGCTCTGGCGG<br>CGACCTGGAAGTGACCACCCACAGCGCC | AGTTGGCGCTGTGGGTGGTCACTTCCAGGT<br>CGCCGCCAGAGCTGTTGGCGAATCTGAT |
| 377 | N | CTTCGCCTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGGC |
| 379 | G | CTTCAACTGTGCCGGCGAGTTCTTCTACTGC<br>AATACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGGCACAGTT |
| 380 | G | CTTCAACTGTGGCGCCGAGTTCTTCTACTGC<br>AATACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGGCGCCACAGTT |

| | | | |
|---|---|---|---|
| 381 | E | CTTCAACTGTGGCGGCGCCTTCTTCTACTGC<br>AATACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGAAGGCGCCGCCACAGTT |
| 382 | F | CTTCAACTGTGGCGGCGAGGCCTTCTACTGC<br>AATACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGGCCTCGCCGCCACAGTT |
| 383 | F | CTTCAACTGTGGCGGCGAGTTCGCCTACTGC<br>AATACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGGCGAACTCGCCGCCACAGTT |
| 384 | Y | CTTCAACTGTGGCGGCGAGTTCTTCGCCTGC<br>AATACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGGCGAAGAACTCGCCGCCACAGTT |
| 386 | N | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>GCCACCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGTGG<br>CGCAGTAGAAGAACTCGCCGCCACAGTT |
| 387 | T | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATGCCTCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGAGGCAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 388 | S | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCGCCGGCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGCCGGCGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 389 | G | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCTCCGCCCTGTTCAACAGCACC | TCCAGGTGCTGTTGAACAGGGCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 390 | L | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCTCCGGCGCCTTCAACAGCACC | TCCAGGTGCTGTTGAAGGCGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 391 | F | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCTCCGGCCTGGCCAACAGCACC | TCCAGGTGCTGTTGGCCAGGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 392 | N | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCTCCGGCCTGTTCGCCAGCACC | TCCAGGTGCTGGCGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 393 | S | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCTCCGGCCTGTTCAACGCCACC | TCCAGGTGGCGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 394 | T | CTTCAACTGTGGCGGCGAGTTCTTCTACTGC<br>AATACCTCCGGCCTGTTCAACAGCGCC | TCCAGGCGCTGTTGAACAGGCCGGAGGTAT<br>TGCAGTAGAAGAACTCGCCGCCACAGTT |
| 395 | W | CACCGCCATCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATGGC |
| 396 | I | CACCTGGGCCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGGCCCA |
| 397 | S | CACCTGGATCGCCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGGCGATCCA |
| 398 | N | CACCTGGATCAGCGCCACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTGGCGCTGATCCA |
| 399 | T | CACCTGGATCAGCAATGCCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGCATTGCTGATCCA |
| 401 | S | CACCTGGATCAGCAATACCGCCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGGCGGTATTGCTGATCCA |
| 402 | V | CACCTGGATCAGCAATACCAGCGCCCAGGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGGGCGCTGGTATTGCTGATCCA |
| 403 | Q | CACCTGGATCAGCAATACCAGCGTGGCCGG<br>GAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCGGCCACGCTGGTATTGCTGATCCA |
| 404 | G | CACCTGGATCAGCAATACCAGCGTGCAGGC<br>CAGCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGCT<br>GGCCTGCACGCTGGTATTGCTGATCCA |
| 405 | S | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GGCCAACAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGTTGGC<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 406 | N | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCGCCAGCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGCTGGCGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 407 | S | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCAACGCCACCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGTGGCGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 408 | T | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCGCCGGCAGCAACGACAGC | TGATGCTGTCGTTGCTGCCGGCGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 409 | G | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGCCAGCAACGACAGC | TGATGCTGTCGTTGCTGGCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 410 | S | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCGCCAACGACAGC | TGATGCTGTCGTTGGCGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 411 | N | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCGCCGACAGC | TGATGCTGTCGGCGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 412 | D | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGCCAGC | TGATGCTGGCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 413 | S | CACCTGGATCAGCAATACCAGCGTGCAGGG<br>GAGCAACAGCACCGGCAGCAACGACGCC | TGATGGCGTCGTTGCTGCCGGTGCTGTTGCT<br>CCCCTGCACGCTGGTATTGCTGATCCA |
| 414 | I | CAGCGCCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>CTGCTTGATCCGGCAGGGCAGGGTGGC |
| 415 | T | CAGCATCGCCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>CTGCTTGATCCGGCAGGGCAGGGCGAT |
| 416 | L | CAGCATCACCGCCCCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>CTGCTTGATCCGGCAGGGGGCGGTGAT |
| 417 | P | CAGCATCACCCTGGCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>CTGCTTGATCCGGCAGGCCAGGGTGAT |
| 419 | R | CAGCATCACCCTGCCCTGCGCCATCAAGCA<br>GATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>CTGCTTGATGGCGCAGGGCAGGGTGAT |
| 420 | I | CAGCATCACCCTGCCCTGCCGGGCCAAGCA<br>GATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>CTGCTTGGCCCGGCAGGGCAGGGTGAT |

| | | | |
|---|---|---|---|
| 421 | K | CAGCATCACCCTGCCCTGCCGGATCGCCCA<br>GATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>CTGGGCGATCCGGCAGGGCAGGGTGAT |
| 422 | Q | CAGCATCACCCTGCCCTGCCGGATCAAGGC<br>CATCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGAT<br>GGCCTTGATCCGGCAGGGCAGGGTGAT |
| 423 | I | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GGCCATCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGATGG<br>CCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 424 | I | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCGCCAATATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATATTGGCGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 425 | N | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCGCCATGTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCACATGGCGATGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 426 | M | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATGCCTGGCAGCGGATCGGC | CCTGGCCGATCCGCTGCCAGGCATTGATGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 427 | W | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATATGGCCCAGCGGATCGGC | CCTGGCCGATCCGCTGGGCCATATTGATGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 428 | Q | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGGCCCGGATCGGC | CCTGGCCGATCCGGGCCCACATATTGATGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 429 | R | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGCAGGCCATCGGC | CCTGGCCGATGGCCTGCCACATATTGATGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 430 | I | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGCAGCGGGCCGGC | CCTGGCCGGCCCGCTGCCACATATTGATGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 431 | G | CAGCATCACCCTGCCCTGCCGGATCAAGCA<br>GATCATCAATATGTGGCAGCGGATCGCC | CCTGGGCGATCCGCTGCCACATATTGATGA<br>TCTGCTTGATCCGGCAGGGCAGGGTGAT |
| 432 | Q | CGGCGCCGCTATGTACGCCCCACCCATCCA<br>GGGCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGC<br>CCTGGATGGGTGGGGCGTACATAGCGGC |
| 434 | M | CGGCCAGGCTGCCTACGCCCCACCCATCCA<br>GGGCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGC<br>CCTGGATGGGTGGGGCGTAGGCAGCCTG |
| 435 | Y | CGGCCAGGCTATGGCCGCCCCACCCATCCA<br>GGGCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGC<br>CCTGGATGGGTGGGGCGGCCATAGCCTG |
| 437 | P | CGGCCAGGCTATGTACGCCGCCCCCATCCA<br>GGGCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGC<br>CCTGGATGGGGGCGGCGTACATAGCCTG |
| 438 | P | CGGCCAGGCTATGTACGCCCCAGCCATCCA<br>GGGCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGC<br>CCTGGATGGCTGGGGCGTACATAGCCTG |
| 439 | I | CGGCCAGGCTATGTACGCCCCACCCGCCCA<br>GGGCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGC<br>CCTGGGCGGGTGGGGCGTACATAGCCTG |
| 440 | Q | CGGCCAGGCTATGTACGCCCCACCCATCGC<br>CGGCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGC<br>CGGCGATGGGTGGGGCGTACATAGCCTG |
| 441 | G | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGCCGTGATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATCACGG<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 442 | V | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGGCGCCATCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGATGGCGC<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 443 | I | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGGCGTGGCCAGATGCGTGTCCAATATC | CGGTGATATTGGACACGCATCTGGCCACGC<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 444 | R | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGGCGTGATCGCCTGCGTGTCCAATATC | CGGTGATATTGGACACGCAGGCGATCACGC<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 446 | V | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGGCGTGATCAGATGCGCCTCCAATATC | CGGTGATATTGGAGGCGCATCTGATCACGC<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 447 | S | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGGCGTGATCAGATGCGTGGCCAATATC | CGGTGATATTGGCCACGCATCTGATCACGC<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 448 | N | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGGCGTGATCAGATGCGTGTCCGCCATC | CGGTGATGGCGGACACGCATCTGATCACGC<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 449 | I | CGGCCAGGCTATGTACGCCCCACCCATCCA<br>GGGCGTGATCAGATGCGTGTCCAATGCC | CGGTGGCATTGGACACGCATCTGATCACGC<br>CCTGGATGGGTGGGGCGTACATAGCCTG |
| 450 | T | TATCGCCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTCAGGATCAGGCCGGC |
| 451 | G | TATCACCGCCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTCAGGATCAGGGCGGT |
| 452 | L | TATCACCGGCGCCATCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTCAGGATGGCGCCGGT |
| 453 | I | TATCACCGGCCTGGCCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTCAGGGCCAGGCCGGT |
| 454 | L | TATCACCGGCCTGATCGCCACCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTGGCGATCAGGCCGGT |
| 455 | T | TATCACCGGCCTGATCCTGGCCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGCCAGGATCAGGCCGGT |
| 456 | R | TATCACCGGCCTGATCCTGACCGCCGACGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCGGCGGTCAGGATCAGGCCGGT |
| 457 | D | TATCACCGGCCTGATCCTGACCCGGGCCGG<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGGCCCGGGTCAGGATCAGGCCGGT |
| 458 | G | TATCACCGGCCTGATCCTGACCCGGGACGC<br>CGGCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGCC<br>GGCGTCCCGGGTCAGGATCAGGCCGGT |
| 459 | G | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGCCAGCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGCTGGC<br>GCCGTCCCGGGTCAGGATCAGGCCGGT |
| 460 | S | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCGCCACCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGTGGCGC<br>CGCCGTCCCGGGTCAGGATCAGGCCGGT |

| | | | |
|---|---|---|---|
| 461 | T | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCGCCAACAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGTTGGCGCTGCC<br>GCCGTCCCGGGTCAGGATCAGGCCGGT |
| 462 | N | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCGCCAGCACCACCGAAACC | TGAAGGTTTCGGTGGTGCTGGCGGTGCTGC<br>CGCCGTCCCGGGTCAGGATCAGGCCGGT |
| 463 | S | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCAACGCCACCACCGAAACC | TGAAGGTTTCGGTGGTGGCGTTGGTGCTGC<br>CGCCGTCCCGGGTCAGGATCAGGCCGGT |
| 464 | T | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCGCCACCGAAACC | TGAAGGTTTCGGTGGCGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTCAGGATCAGGCCGGT |
| 465 | T | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCACCGCCGAAACC | TGAAGGTTTCGGCGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTCAGGATCAGGCCGGT |
| 466 | E | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGCCACC | TGAAGGTGGCGGTGGTGCTGTTGGTGCTGC<br>CGCCGTCCCGGGTCAGGATCAGGCCGGT |
| 467 | T | TATCACCGGCCTGATCCTGACCCGGGACGG<br>CGGCAGCACCAACAGCACCACCGAAGCC | TGAAGGCTTCGGTGGTGCTGTTGGTGCTGCC<br>GCCGTCCCGGGTCAGGATCAGGCCGGT |
| 468 | F | AACCGCCAGACCCGGCGGAGGCGACATGA<br>GAGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGGC |
| 469 | R | AACCTTCGCCCCCGGCGGAGGCGACATGAG<br>AGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGGGCGAA |
| 470 | P | AACCTTCAGAGCCGGCGGAGGCGACATGAG<br>AGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGCTCTGAA |
| 471 | G | AACCTTCAGACCCGCCGGAGGCGACATGAG<br>AGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGGCGGGTCTGAA |
| 472 | G | AACCTTCAGACCCGGCGCCGGCGACATGAG<br>AGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCGGCGCCGGGTCTGAA |
| 473 | G | AACCTTCAGACCCGGCGGAGCCGACATGAG<br>AGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGGCTCCGCCGGGTCTGAA |
| 474 | D | AACCTTCAGACCCGGCGGAGGCGCCATGAG<br>AGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGGCGCCTCCGCCGGGTCTGAA |
| 475 | M | AACCTTCAGACCCGGCGGAGGCGACGCCAG<br>AGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>TCTGGCGTCGCCTCCGCCGGGTCTGAA |
| 476 | R | AACCTTCAGACCCGGCGGAGGCGACATGGC<br>CGACAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGTC<br>GGCCATGTCGCCTCCGCCGGGTCTGAA |
| 477 | D | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGCCAACTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGTTGGC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 478 | N | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACGCCTGGCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGCCAGGCGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 479 | W | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACAACGCCCGGAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTCCGGGCGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 480 | R | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACAACTGGGCCAGCGAGCTGTACAAG | TGTACTTGTACAGCTCGCTGGCCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 481 | S | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACAACTGGCGGGCCGAGCTGTACAAG | TGTACTTGTACAGCTCGGCCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 482 | E | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACAACTGGCGGAGCGCCCTGTACAAG | TGTACTTGTACAGGGCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 483 | L | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACAACTGGCGGAGCGAGGCCTACAAG | TGTACTTGTAGGCCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 484 | Y | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACAACTGGCGGAGCGAGCTGGCCAAG | TGTACTTGGCCAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 485 | K | AACCTTCAGACCCGGCGGAGGCGACATGAG<br>AGACAACTGGCGGAGCGAGCTGTACGCC | TGTAGGCGTACAGCTCGCTCCGCCAGTTGTC<br>TCTCATGTCGCCTCCGCCGGGTCTGAA |
| 486 | Y | CAAGGCCAAAGTGGTGAAAATCGAGCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGGTGGGGGCCACGCC<br>CAGGGGCTCGATTTTCACCACTTTGGC |
| 487 | K | CAAGTACGCCGTGGTGAAAATCGAGCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>CAGGGGCTCGATTTTCACCACGGCGTA |
| 488 | V | CAAGTACAAAGCCGTGAAAATCGAGCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>CAGGGGCTCGATTTTCACGGCTTTGTA |
| 489 | V | CAAGTACAAAGTGGCCAAAATCGAGCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>CAGGGGCTCGATTTTGGCCACTTTGTA |
| 490 | K | CAAGTACAAAGTGGTGGCCATCGAGCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>CAGGGGCTCGATGGCCACCACTTTGTA |
| 491 | I | CAAGTACAAAGTGGTGAAAGCCGAGCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>CAGGGGCTCGGCTTTCACCACTTTGTA |
| 492 | E | CAAGTACAAAGTGGTGAAAATCGCCCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>CAGGGGGGCGATTTTCACCACTTTGTA |
| 493 | P | CAAGTACAAAGTGGTGAAAATCGAGGCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>CAGGGCCTCGATTTTCACCACTTTGTA |
| 494 | L | CAAGTACAAAGTGGTGAAAATCGAGCCCGC<br>CGGCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGCC<br>GGCGGGCTCGATTTTCACCACTTTGTA |
| 495 | G | CAAGTACAAAGTGGTGAAAATCGAGCCCCT<br>GGCCGTGGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCCACGG<br>CCAGGGGCTCGATTTTCACCACTTTGTA |
| 496 | V | CAAGTACAAAGTGGTGAAAATCGAGCCCCT<br>GGGCGCCGCCCCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGGGGCGGCGC<br>CCAGGGGCTCGATTTTCACCACTTTGTA |
| 498 | P | CAAGTACAAAGTGGTGAAAATCGAGCCCCT<br>GGGCGTGGCCGCCACCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGTGGCGGCCACGCC<br>CAGGGGCTCGATTTTCACCACTTTGTA |

| | | | |
|---|---|---|---|
| 499 | T | CAAGTACAAAGTGGTGAAAATCGAGCCCCT<br>GGGCGTGGCCCCCGCCCGGTGCAAGAGA | CTCTTCTCTTGCACCGGGCGGGGGCCACGC<br>CCAGGGGCTCGATTTTCACCACTTTGTA |
| 500 | R | CAAGTACAAAGTGGTGAAAATCGAGCCCCT<br>GGGCGTGGCCCCCACCGCCTGCAAGAGA | CTCTTCTCTTGCAGGCGGTGGGGGCCACGC<br>CCAGGGGCTCGATTTTCACCACTTTGTA |
| 502 | K | CAAGTACAAAGTGGTGAAAATCGAGCCCCT<br>GGGCGTGGCCCCCCACCCGGTGCGCCAGA | CTCTTCTGGCGCACCGGGTGGGGGCCACGC<br>CCAGGGGCTCGATTTTCACCACTTTGTA |
| 503 | R | CAAGTACAAAGTGGTGAAAATCGAGCCCCT<br>GGGCGTGGCCCCCACCCGGTGCAAGGCC | CTCTGGCCTTGCACCGGGTGGGGGCCACGC<br>CCAGGGGCTCGATTTTCACCACTTTGTA |
| 504 | R | GAGAGCCGTCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGGCATCGGAGCCGTGTTT | CCAGAAACACGGCTCCGATGCCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGACGGC |
| 505 | V | GAGAAGAGCCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGGCATCGGAGCCGTGTTT | CCAGAAACACGGCTCCGATGCCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGGCTCT |
| 506 | V | GAGAAGAGTCGCCGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGGCATCGGAGCCGTGTTT | CCAGAAACACGGCTCCGATGCCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCGGCGACTCT |
| 507 | G | GAGAAGAGTCGTGGCCCGGCGGAGAAGGC<br>GGAGAGCCGTGGGCATCGGAGCCGTGTTT | CCAGAAACACGGCTCCGATGCCCACGGCTC<br>TCCGCCTTCTCCGCCGGGCCACGACTCT |
| 513 | V | GAGAAGAGTCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGCCGGCATCGGAGCCGTGTTT | CCAGAAACACGGCTCCGATGCCGGCGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGACTCT |
| 514 | G | GAGAAGAGTCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGCCATCGGAGCCGTGTTT | CCAGAAACACGGCTCCGATGGCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGACTCT |
| 515 | I | GAGAAGAGTCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGGCGCCGGAGCCGTGTTT | CCAGAAACACGGCTCCGGCGCCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGACTCT |
| 516 | G | GAGAAGAGTCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGGCATCGCCGCCGTGTTT | CCAGAAACACGGCGGCGATGCCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGACTCT |
| 518 | V | GAGAAGAGTCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGGCATCGGAGCCGCCTTT | CCAGAAAGGCGGCTCCGATGCCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGACTCT |
| 519 | F | GAGAAGAGTCGTGGGACGGCGGAGAAGGC<br>GGAGAGCCGTGGGCATCGGAGCCGTGGCC | CCAGGGCCACGGCTCCGATGCCCACGGCTC<br>TCCGCCTTCTCCGCCGTCCCACGACTCT |
| 520 | L | GTTTGCCGGCTTTCTGGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCCAGAAAGCCGGC |
| 521 | G | GTTTCTGGCCTTTCTGGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCCAGAAAGGCCAG |
| 522 | F | GTTTCTGGGCGCCCTGGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCCAGGGCGCCCAG |
| 523 | L | GTTTCTGGGCTTTGCCGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCGGCAAAGCCCAG |
| 524 | G | GTTTCTGGGCTTTCTGGCCGCCGCCGGAAGC<br>ACCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCGGCCAGAAAGCCCAG |
| 527 | G | GTTTCTGGGCTTTCTGGGAGCCGCCGCCAGC<br>ACCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>TGCTGGCGGCGGCTCCCAGAAAGCCCAG |
| 528 | S | GTTTCTGGGCTTTCTGGGAGCCGCCGGAGC<br>CACCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>TGGCTCCGGCGGCTCCCAGAAAGCCCAG |
| 529 | T | GTTTCTGGGCTTTCTGGGAGCCGCCGGAAG<br>CGCCATGGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCCATGG<br>CGCTTCCGGCGGCTCCCAGAAAGCCCAG |
| 530 | M | GTTTCTGGGCTTTCTGGGAGCCGCCGGAAG<br>CACCGCCGGCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGCCGGCGG<br>TGCTTCCGGCGGCTCCCAGAAAGCCCAG |
| 531 | G | GTTTCTGGGCTTTCTGGGAGCCGCCGGAAG<br>CACCATGGCCGCTGCCAGCATGACACTG | CGGTCAGTGTCATGCTGGCAGCGGCCATGG<br>TGCTTCCGGCGGCTCCCAGAAAGCCCAG |
| 534 | S | GTTTCTGGGCTTTCTGGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCGCCATGACACTG | CGGTCAGTGTCATGGCGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCCAGAAAGCCCAG |
| 535 | M | GTTTCTGGGCTTTCTGGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCAGCGCCACACTG | CGGTCAGTGTGGCGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCCAGAAAGCCCAG |
| 536 | T | GTTTCTGGGCTTTCTGGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCAGCATGGCCCTG | CGGTCAGGGCCATGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCCAGAAAGCCCAG |
| 537 | L | GTTTCTGGGCTTTCTGGGAGCCGCCGGAAG<br>CACCATGGGCGCTGCCAGCATGACAGCC | CGGTGGCTGTCATGCTGGCAGCGCCCATGG<br>TGCTTCCGGCGGCTCCCAGAAAGCCCAG |
| 538 | T | ACTGGCCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGC |
| 539 | V | ACTGACCGCCCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGGGCGGT |
| 540 | Q | ACTGACCGTGGCCGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCGGCCACGGT |
| 542 | R | ACTGACCGTGCAGGCCGCCAACCTGCTGAG<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGCTCAGCAGGTTGGCGGCCTGCACGGT |
| 543 | N | ACTGACCGTGCAGGCCCGGGCCCTGCTGAG<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGCTCAGCAGGGCCCGGGCCTGCACGGT |
| 544 | L | ACTGACCGTGCAGGCCCGGAACGCCCTGAG<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGCTCAGGGCGTTCCGGGCCTGCACGGT |
| 545 | L | ACTGACCGTGCAGGCCCGGAACCTGGCCAG<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGCTGGCCAGGTTCCGGGCCTGCACGGT |
| 546 | S | ACTGACCGTGCAGGCCCGGAACCTGCTGGC<br>CGGCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGC<br>CGGCCAGCAGGTTCCGGGCCTGCACGGT |
| 547 | G | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGCCATCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGATGG<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |

| | | | |
|---|---|---|---|
| 548 | I | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCGCCGTGCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGCACGGCGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 549 | V | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGCCCAGCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGCTGGGCGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 550 | Q | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGGCCCAGCAGAGCAATCTG | TGAGCAGATTGCTCTGCTGGGCCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 551 | Q | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGGCCCAGAGCAATCTG | TGAGCAGATTGCTCTGGGCCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 552 | Q | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGCAGGCCAGCAATCTG | TGAGCAGATTGCTGGCCTGCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 553 | S | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGCAGCAGGCCAATCTG | TGAGCAGATTGGCCTGCTGCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 554 | N | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGCAGCAGAGCGCCCTG | TGAGCAGGGCGCTCTGCTGCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 555 | L | ACTGACCGTGCAGGCCCGGAACCTGCTGAG<br>CGGCATCGTGCAGCAGCAGAGCAATGCC | TGAGGGCATTGCTCTGCTGCTGCACGATGC<br>CGCTCAGCAGGTTCCGGGCCTGCACGGT |
| 556 | L | TCTGGCCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGGC |
| 557 | R | TCTGCTCGCCGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCGGCGAG |
| 560 | E | TCTGCTCAGAGCCCCTGCCGCCCAGCAGCA<br>TCTGCTGAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTCAGCA<br>GATGCTGCTGGGCGGCAGGGGCTCTGAG |
| 562 | Q | TCTGCTCAGAGCCCCTGAGGCCGCCCAGCA<br>TCTGCTGAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTCAGCA<br>GATGCTGGGCGGCCTCAGGGGCTCTGAG |
| 563 | Q | TCTGCTCAGAGCCCCTGAGGCCCAGGCCCA<br>TCTGCTGAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTCAGCA<br>GATGGGCCTGGGCCTCAGGGGCTCTGAG |
| 564 | H | TCTGCTCAGAGCCCCTGAGGCCCAGCAGGC<br>CCTGCTGAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTCAGCA<br>GGGCCTGCTGGGCCTCAGGGGCTCTGAG |
| 565 | L | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TGCCCTGAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTCAGGG<br>CATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 566 | L | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGGCCAAACTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGTTTGGCCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 567 | K | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGGCCCTGACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTCAGGGCCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 568 | L | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAAGCCACCGTGTGGGGCATC | GCTTGATGCCCCACACGGTGGCTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 569 | T | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAACTGGCCGTGTGGGGCATC | GCTTGATGCCCCACACGGCCAGTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 570 | V | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAACTGACCGCCTGGGGCATC | GCTTGATGCCCCAGGCGGTCAGTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 571 | W | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAACTGACCGTGGCCGGCATC | GCTTGATGCCGGCCACGGTCAGTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 572 | G | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAACTGACCGTGTGGGCCATC | GCTTGATGGCCCACACGGTCAGTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 573 | I | TCTGCTCAGAGCCCCTGAGGCCCAGCAGCA<br>TCTGCTGAAACTGACCGTGTGGGGCGCC | GCTTGGCGCCCCACACGGTCAGTTTCAGCA<br>GATGCTGCTGGGCCTCAGGGGCTCTGAG |
| 574 | K | CATCGCCCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCCAC<br>GGCCAGCACTCTGGCCTGCAGCTGGGC |
| 575 | Q | CATCAAGGCCCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCCAC<br>GGCCAGCACTCTGGCCTGCAGGGCCTT |
| 576 | L | CATCAAGCAGGCCCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCCAC<br>GGCCAGCACTCTGGCCTGGGCCTGCTT |
| 577 | Q | CATCAAGCAGCTGGCCGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCCAC<br>GGCCAGCACTCTGGCGGCCAGCTGCTT |
| 579 | R | CATCAAGCAGCTGCAGGCCGCCGTGCTGGC<br>CGTGGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCCAC<br>GGCCAGCACGGCGGCCTGCAGCTGCTT |
| 580 | V | CATCAAGCAGCTGCAGGCCAGAGCCCTGGC<br>CGTGGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCCAC<br>GGCCAGGGCTCTGGCCTGCAGCTGCTT |
| 581 | L | CATCAAGCAGCTGCAGGCCAGAGTGGCCGC<br>CGTGGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCCAC<br>GGCGGCCACTCTGGCCTGCAGCTGCTT |
| 583 | V | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGCCGAAAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTTTCGG<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |
| 584 | E | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGCCAGATACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTATCTGGCCA<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |
| 585 | R | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAGCCTACCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGTAGGCTTCCA<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |
| 586 | Y | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGAGCCCTGCGGGATCAGCAG | GGAGCTGCTGATCCCGCAGGGCTCTTTCCA<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |
| 587 | L | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACGCCCGGGATCAGCAG | GGAGCTGCTGATCCCGGGCGTATCTTTCCAC<br>GGCCAGCACTCTGGCCTGCAGCTGCTT |
| 588 | R | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGGCCGATCAGCAG | GGAGCTGCTGATCGGCCAGGTATCTTTCCA<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |
| 589 | D | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGCGGGCCCAGCAG | GGAGCTGCTGGGCCCGCAGGTATCTTTCCA<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |

| | | | |
|---|---|---|---|
| 590 | Q | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGCGGGATGCCCAG | GGAGCTGGGCATCCCGCAGGTATCTTTCCA<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |
| 591 | Q | CATCAAGCAGCTGCAGGCCAGAGTGCTGGC<br>CGTGGAAAGATACCTGCGGGATCAGGCC | GGAGGGCCTGATCCCGCAGGTATCTTTCCA<br>CGGCCAGCACTCTGGCCTGCAGCTGCTT |
| 592 | L | GCAGGCCCTGGGCATCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGATGCCCAGGGC |
| 593 | L | GCAGCTCGCCGGCATCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGATGCCGGCGAG |
| 594 | G | GCAGCTCCTGGCCATCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGATGGCCAGGAG |
| 595 | I | GCAGCTCCTGGGCGCCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGGCGCCCAGGAG |
| 596 | W | GCAGCTCCTGGGCATCGCCGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCGGCGATGCCCAGGAG |
| 597 | G | GCAGCTCCTGGGCATCTGGGCCTGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCAGGCCCAGATGCCCAGGAG |
| 599 | S | GCAGCTCCTGGGCATCTGGGGATGCGCCGG<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGCCGGCGCATCCCCAGATGCCCAGGAG |
| 600 | G | GCAGCTCCTGGGCATCTGGGGATGCAGCGC<br>CAAGCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGCT<br>TGGCGCTGCATCCCCAGATGCCCAGGAG |
| 601 | K | GCAGCTCCTGGGCATCTGGGGATGCAGCGG<br>CGCCCTGATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATCAGGG<br>CGCCGCTGCATCCCCAGATGCCCAGGAG |
| 602 | L | GCAGCTCCTGGGCATCTGGGGATGCAGCGG<br>CAAGGCCATCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGATGGCCT<br>TGCCGCTGCATCCCCAGATGCCCAGGAG |
| 603 | I | GCAGCTCCTGGGCATCTGGGGATGCAGCGG<br>CAAGCTGGCCTGCTGCACCAATGTGCCC | TCCAGGGCACATTGGTGCAGCAGGCCAGCT<br>TGCCGCTGCATCCCCAGATGCCCAGGAG |
| 606 | T | GCAGCTCCTGGGCATCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCGCCAATGTGCCC | TCCAGGGCACATTGGCGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGATGCCCAGGAG |
| 607 | N | GCAGCTCCTGGGCATCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCGCCGTGCCC | TCCAGGGCACGGCGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGATGCCCAGGAG |
| 608 | V | GCAGCTCCTGGGCATCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGCCCCC | TCCAGGGGGCATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGATGCCCAGGAG |
| 609 | P | GCAGCTCCTGGGCATCTGGGGATGCAGCGG<br>CAAGCTGATCTGCTGCACCAATGTGGCC | TCCAGGCCACATTGGTGCAGCAGATCAGCT<br>TGCCGCTGCATCCCCAGATGCCCAGGAG |
| 610 | W | GCCCGCCAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTCCGGTTGGACCAGCTGCTGTTGGC |
| 611 | N | GCCCTGGGCCAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTCCGGTTGGACCAGCTGCTGGCCCA |
| 612 | S | GCCCTGGAACGCCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTCCGGTTGGACCAGCTGGCGTTCCA |
| 613 | S | GCCCTGGAACAGCGCCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTCCGGTTGGACCAGGCGCTGTTCCA |
| 614 | W | GCCCTGGAACAGCAGCGCCTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTCCGGTTGGAGGCGCTGCTGTTCCA |
| 615 | S | GCCCTGGAACAGCAGCTGGGCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTCCGGTTGGCCCAGCTGCTGTTCCA |
| 616 | N | GCCCTGGAACAGCAGCTGGTCCGCCCGGAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTCCGGGCGGACCAGCTGCTGTTCCA |
| 617 | R | GCCCTGGAACAGCAGCTGGTCCAACGCCAA<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GTTGGCGTTGGACCAGCTGCTGTTCCA |
| 618 | N | GCCCTGGAACAGCAGCTGGTCCAACCGGGC<br>CCTGAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTCAG<br>GGCCCGGTTGGACCAGCTGCTGTTCCA |
| 619 | L | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CGCCAGCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGCTGGC<br>GTTCCGGTTGGACCAGCTGCTGTTCCA |
| 620 | S | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGGCCGAGATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATCTCGGCCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 621 | E | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGCCATCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGATGGCGCTCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 622 | I | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGGCCTGGGACAACATGACC | GCCAGGTCATGTTGTCCCAGGCCTCGCTCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 623 | W | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCGCCGACAACATGACC | GCCAGGTCATGTTGTCGGCGATCTCGCTCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 624 | D | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGCCAACATGACC | GCCAGGTCATGTTGGCCCAGATCTCGCTCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 625 | N | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACGCCATGACC | GCCAGGTCATGGCGTCCCAGATCTCGCTCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 626 | M | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACGCCACC | GCCAGGTGGCGTTGTCCCAGATCTCGCTCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 627 | T | GCCCTGGAACAGCAGCTGGTCCAACCGGAA<br>CCTGAGCGAGATCTGGGACAACATGGCC | GCCAGGCCATGTTGTCCCAGATCTCGCTCA<br>GGTTCCGGTTGGACCAGCTGCTGTTCCA |
| 628 | W | GACCGCCCTGCAGTGGGACAAAGAGATCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGGGC |
| 629 | L | GACCTGGGCCCAGTGGGACAAAGAGATCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGGGCCCA |

| | | | |
|---|---|---|---|
| 630 | Q | GACCTGGCTGGCCTGGGACAAAGAGATCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCAGGCCAGCCA |
| 631 | W | GACCTGGCTGCAGGCCGACAAAGAGATCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCGGCCTGCAGCCA |
| 632 | D | GACCTGGCTGCAGTGGGCCAAAGAGATCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGGCCCACTGCAGCCA |
| 633 | K | GACCTGGCTGCAGTGGGACGCCGAGATCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCGGCGTCCCACTGCAGCCA |
| 634 | E | GACCTGGCTGCAGTGGGACAAAGCCATCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATGGCTTTGTCCCACTGCAGCCA |
| 635 | I | GACCTGGCTGCAGTGGGACAAAGAGGCCAG<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGCTGGCCTCTTTGTCCCACTGCAGCCA |
| 636 | S | GACCTGGCTGCAGTGGGACAAAGAGATCGC<br>CAACTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGT<br>TGGCGATCTCTTTGTCCCACTGCAGCCA |
| 637 | N | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CGCCTACACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGTAGG<br>CGCTGATCTCTTTGTCCCACTGCAGCCA |
| 638 | Y | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACGCCACCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGTGGCGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 639 | T | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACTACGCCCAGATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATCTGGGCGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 640 | Q | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACTACACCGCCATCATCTACGGCCTG | CGAGCAGGCCGTAGATGATGGCGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 641 | I | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACTACACCCAGGCCATCTACGGCCTG | CGAGCAGGCCGTAGATGGCCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 642 | I | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACTACACCCAGATCGCCTACGGCCTG | CGAGCAGGCCGTAGGCGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 643 | Y | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACTACACCCAGATCATCGCCGGCCTG | CGAGCAGGCCGGCGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 644 | G | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACTACACCCAGATCATCTACGCCCTG | CGAGCAGGGCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 645 | L | GACCTGGCTGCAGTGGGACAAAGAGATCAG<br>CAACTACACCCAGATCATCTACGGCGCC | CGAGGGCGCCGTAGATGATCTGGGTGTAGT<br>TGCTGATCTCTTTGTCCCACTGCAGCCA |
| 646 | L | CCTGGCCGAGGAGAGCCAGAACCAGCAGG<br>AAAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTCTCCTCGGC |
| 647 | E | CCTGCTCGCCGAGAGCCAGAACCAGCAGGA<br>AAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTCTCGGCGAG |
| 648 | E | CCTGCTCGAGGCCAGCCAGAACCAGCAGGA<br>AAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTGGCCTCGAG |
| 649 | S | CCTGCTCGAGGAGGCCCAGAACCAGCAGGA<br>AAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGGCCTCCTCGAG |
| 650 | Q | CCTGCTCGAGGAGAGCGCCAACCAGCAGGA<br>AAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTGGCGCTCTCCTCGAG |
| 651 | N | CCTGCTCGAGGAGAGCCAGGCCCAGCAGGA<br>AAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGGCCTGGCTCTCCTCGAG |
| 652 | Q | CCTGCTCGAGGAGAGCCAGAACGCCCAGGA<br>AAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGGGCGTTCTGGCTCTCCTCGAG |
| 653 | Q | CCTGCTCGAGGAGAGCCAGAACCAGGCCGA<br>AAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCGGCCTGGTTCTGGCTCTCCTCGAG |
| 654 | E | CCTGCTCGAGGAGAGCCAGAACCAGCAGGC<br>CAAGAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTCTT<br>GGCCTGCTGGTTCTGGCTCTCCTCGAG |
| 655 | K | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AGCCAACGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGTTGG<br>CTTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 656 | N | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AAAGGCCGAGCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGCTCGGCCT<br>TTTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 657 | E | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AAAGAACGCCCAGGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCCTGGGCGTTCT<br>TTTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 658 | Q | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AAAGAACGAGGCCGACCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGTCGGCCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 659 | D | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AAAGAACGAGCAGGCCCTGCTGGCCCTG | TGTCCAGGGCCAGCAGGGCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 660 | L | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AAAGAACGAGCAGGACGCCCTGGCCCTG | TGTCCAGGGCCAGGGCGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 661 | L | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AAAGAACGAGCAGGACCTGGCCGCCCTG | TGTCCAGGGCGGCCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 663 | L | CCTGCTCGAGGAGAGCCAGAACCAGCAGGA<br>AAAGAACGAGCAGGACCTGCTGGCCGCC | TGTCGGCGGCCAGCAGGTCCTGCTCGTTCTT<br>TTCCTGCTGGTTCTGGCTCTCCTCGAG |
| 664 | D | CCTGGCCAAGTGGGCCAGCCTGTGGAATTG<br>GTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA<br>ACCAATTCCACAGGCTGGCCCACTTGGC |
| 665 | K | CCTGGACGCCTGGGCCAGCCTGTGGAATTG<br>GTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA<br>ACCAATTCCACAGGCTGGCCCAGGCGTC |
| 666 | W | CCTGGACAAGGCCGCCAGCCTGTGGAATTG<br>GTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA<br>ACCAATTCCACAGGCTGGCGGCCTTGTC |
| 668 | S | CCTGGACAAGTGGGCCGCCCTGTGGAATTG<br>GTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA<br>ACCAATTCCACAGGGCGGCCCACTTGTC |

| | | | |
|---|---|---|---|
| 669 | L | CCTGGACAAGTGGGCCAGCGCCTGGAATTG GTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA ACCAATTCCAGGCGCTGGCCCACTTGTC |
| 670 | W | CCTGGACAAGTGGGCCAGCCTGGCCAATTG GTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA ACCAATTGGCCAGGCTGGCCCACTTGTC |
| 671 | N | CCTGGACAAGTGGGCCAGCCTGTGGGCCTG GTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA ACCAGGCCCACAGGCTGGCCCACTTGTC |
| 672 | W | CCTGGACAAGTGGGCCAGCCTGTGGAATGC CTTCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGA AGGCATTCCACAGGCTGGCCCACTTGTC |
| 673 | F | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GGCCGACATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGTCGG CCCAATTCCACAGGCTGGCCCACTTGTC |
| 674 | D | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGCCATCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGATGGCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 675 | I | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGACGCCAGCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGCTGGCGTCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 676 | S | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGACATCGCCAACTGGCTGTGGTAC | TGATGTACCACAGCCAGTTGGCGATGTCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 677 | N | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGACATCAGCGCCTGGCTGTGGTAC | TGATGTACCACAGCCAGGCGCTGATGTCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 678 | W | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGACATCAGCAACGCCCTGTGGTAC | TGATGTACCACAGGGCGTTGCTGATGTCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 679 | L | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGACATCAGCAACTGGGCCTGGTAC | TGATGTACCAGGCCCAGTTGCTGATGTCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 680 | W | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGACATCAGCAACTGGCTGGCCTAC | TGATGTAGGCCAGCCAGTTGCTGATGTCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 681 | Y | CCTGGACAAGTGGGCCAGCCTGTGGAATTG GTTCGACATCAGCAACTGGCTGTGGGCC | TGATGGCCCACAGCCAGTTGCTGATGTCGA ACCAATTCCACAGGCTGGCCCACTTGTC |
| 682 | I | GTACGCCAAGATCTTCATCATGATCGTGGG CGGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCCACGATCATGATGAAGATCTTGGC |
| 683 | K | GTACATCGCCATCTTCATCATGATCGTGGGC GGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCCACGATCATGATGAAGATGGCGAT |
| 684 | I | GTACATCAAGGCCTTCATCATGATCGTGGG CGGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCCACGATCATGATGAAGGCCTTGAT |
| 685 | F | GTACATCAAGATCGCCATCATGATCGTGGG CGGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCCACGATCATGATGGCGATCTTGAT |
| 686 | I | GTACATCAAGATCTTCGCCATGATCGTGGG CGGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCCACGATCATGGCGAAGATCTTGAT |
| 687 | M | GTACATCAAGATCTTCATCGCCATCGTGGG CGGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCCACGATGGCGATGAAGATCTTGAT |
| 688 | I | GTACATCAAGATCTTCATCATGGCCGTGGG CGGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCCACGGCCATGATGAAGATCTTGAT |
| 689 | V | GTACATCAAGATCTTCATCATGATCGCCGG CGGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGCCGGCGATCATGATGAAGATCTTGAT |
| 690 | G | GTACATCAAGATCTTCATCATGATCGTGGCC GGCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGC CGGCCACGATCATGATGAAGATCTTGAT |
| 691 | G | GTACATCAAGATCTTCATCATGATCGTGGG CGCCCTGATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATCAGGG CGCCCACGATCATGATGAAGATCTTGAT |
| 692 | L | GTACATCAAGATCTTCATCATGATCGTGGG CGGCGCCATCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGATGGCGC CGCCCACGATCATGATGAAGATCTTGAT |
| 693 | I | GTACATCAAGATCTTCATCATGATCGTGGG CGGCCTGGCCGGCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGCCGGCCAGGC CGCCCACGATCATGATGAAGATCTTGAT |
| 694 | G | GTACATCAAGATCTTCATCATGATCGTGGG CGGCCTGATCGCCCTGCGGATCGTGTTT | CGGCAAACACGATCCGCAGGGCGATCAGGC CGCCCACGATCATGATGAAGATCTTGAT |
| 695 | L | GTACATCAAGATCTTCATCATGATCGTGGG CGGCCTGATCGGCGCCCGGATCGTGTTT | CGGCAAACACGATCCGGGCGCCGATCAGGC CGCCCACGATCATGATGAAGATCTTGAT |
| 696 | R | GTACATCAAGATCTTCATCATGATCGTGGG CGGCCTGATCGGCCTGGCCATCGTGTTT | CGGCAAACACGATGGCCAGGCCGATCAGGC CGCCCACGATCATGATGAAGATCTTGAT |
| 697 | I | GTACATCAAGATCTTCATCATGATCGTGGG CGGCCTGATCGGCCTGCGGGCCGTGTTT | CGGCAAACACGGCCCGCAGGCCGATCAGGC CGCCCACGATCATGATGAAGATCTTGAT |
| 698 | V | GTACATCAAGATCTTCATCATGATCGTGGG CGGCCTGATCGGCCTGCGGATCGCCTTT | CGGCAAAGGCGATCCGCAGGCCGATCAGGC CGCCCACGATCATGATGAAGATCTTGAT |
| 699 | F | GTACATCAAGATCTTCATCATGATCGTGGG CGGCCTGATCGGCCTGCGGATCGTGGCC | CGGCGGCCACGATCCGCAGGCCGATCAGGC CGCCCACGATCATGATGAAGATCTTGAT |
| 701 | V | GTTTGCCGCCCTGAGCGTGATCCACAGAGT GCGGCAGGGCTAC | CTCAGTAGCCCTGCCGCACTCTGTGGATCAC GCTCAGGGCGGC |
| 702 | L | GTTTGCCGTGGCCAGCGTGATCCACAGAGT GCGGCAGGGCTAC | CTCAGTAGCCCTGCCGCACTCTGTGGATCAC GCTGGCCACGGC |
| 703 | S | GTTTGCCGTGCTGGCCGTGATCCACAGAGT GCGGCAGGGCTAC | CTCAGTAGCCCTGCCGCACTCTGTGGATCAC GGCCAGCACGGC |
| 704 | V | GTTTGCCGTGCTGAGCGCCATCCACAGAGT GCGGCAGGGCTAC | CTCAGTAGCCCTGCCGCACTCTGTGGATGG CGCTCAGCACGGC |
| 705 | I | GTTTGCCGTGCTGAGCGTGGCCCACAGAGT GCGGCAGGGCTAC | CTCAGTAGCCCTGCCGCACTCTGTGGGCCA CGCTCAGCACGGC |
| 706 | H | GTTTGCCGTGCTGAGCGTGATCGCCAGAGT GCGGCAGGGCTAC | CTCAGTAGCCCTGCCGCACTCTGGCGATCA CGCTCAGCACGGC |

| | | | |
|---|---|---|---|
| 707 | R | GTTTGCCGTGCTGAGCGTGATCCACGCCGTGCGGCAGGGCTAC | CTCAGTAGCCCTGCCGCACGGCGTGGATCACGCTCAGCACGGC |
| 708 | V | GTTTGCCGTGCTGAGCGTGATCCACAGAGCCCGGCAGGGCTAC | CTCAGTAGCCCTGCCGGGCTCTGTGGATCACGCTCAGCACGGC |
| 709 | R | GTTTGCCGTGCTGAGCGTGATCCACAGAGTGGCCCAGGGCTAC | CTCAGTAGCCCTGGGCCACTCTGTGGATCACGCTCAGCACGGC |
| 710 | Q | GTTTGCCGTGCTGAGCGTGATCCACAGAGTGCGGGCCGGCTAC | CTCAGTAGCCGGCCCGCACTCTGTGGATCACGCTCAGCACGGC |
| 711 | G | GTTTGCCGTGCTGAGCGTGATCCACAGAGTGCGGCAGGCCTAC | CTCAGTAGGCCTGCCGCACTCTGTGGATCACGCTCAGCACGGC |
| 712 | Y | GTTTGCCGTGCTGAGCGTGATCCACAGAGTGCGGCAGGGCGCC | CTCAGGCGCCCTGCCGCACTCTGTGGATCACGCTCAGCACGGC |

**Appendix Table 7. Oligonucleotides for NGS library preparation (PCR1).**

| Name | Forward primer 5´→3´ | Name | Reverse primer 5´→3´ |
|---|---|---|---|
| IL-fw1-GL | AGTTCTACAGTCCGACGATCCTGTGGGTGACAGTGTACTACG | IL-rev1-GL | CTTGGCACCCGAGAATTCCAGATGATGTCGGTGTGCATC |
| IL-fw2-GL | AGTTCTACAGTCCGACGATCGTGACCGAAGAGTTCAACATG | IL-rev2-GL | CTTGGCACCCGAGAATTCCAGAACAGGCTGTACACCTTCTG |
| IL-fw3-GL | AGTTCTACAGTCCGACGATCCTTCAACATGACCACCGAG | IL-rev3-GL | CTTGGCACCCGAGAATTCCATGTCCTTGCACTTCAGGATG |
| IL-fw4-GL | AGTTCTACAGTCCGACGATCCAAGGTGTCCTTCGAGCC | IL-rev4-GL | CTTGGCACCCGAGAATTCCAACTGCACCAGGATGTTCTTG |
| IL-fw5-GL | AGTTCTACAGTCCGACGATCGCGAGAACATCACCAACAAC | IL-rev5-GL | CTTGGCACCCGAGAATTCCAAGCTGTTTCACCACCTTGC |
| IL-fw6-GL | AGTTCTACAGTCCGACGATCCACCTGGAACGAGACACTG | IL-rev6-GL | CTTGGCACCCGAGAATTCCAGTGATGCTGTCGTTGCTG |
| IL-fw7-GL | AGTTCTACAGTCCGACGATCGATCAGCAATACCAGCGTG | IL-rev7-GL | CTTGGCACCCGAGAATTCCAGTTGTCTCTCATGTCGCCTC |
| IL-fw8-GL | AGTTCTACAGTCCGACGATCCAGCACCAACAGCACCAC | IL-rev8-GL | CTTGGCACCCGAGAATTCCATGCACGGTCAGTGTCATG |
| IL-fw9-GL | AGTTCTACAGTCCGACGATCATCGGAGCCGTGTTTCTG | IL-rev9-GL | CTTGGCACCCGAGAATTCCAGCTGATCCCGCAGGTATC |
| IL-fw10-GL | AGTTCTACAGTCCGACGATCGAGCAATCTGCTCAGAGCC | IL-rev10-GL | CTTGGCACCCGAGAATTCCAGCAGCCAGGTCATGTTGTC |
| IL-fw11-GL | AGTTCTACAGTCCGACGATCGATCTGCTGCACCAATGTG | IL-rev11-GL | CTTGGCACCCGAGAATTCCATGATGTCGAACCAATTCCAC |
| IL-fw12-GL | AGTTCTACAGTCCGACGATCATCTACGGCCTGCTCGAG | IL-rev12-GL | CTTGGCACCCGAGAATTCCAATATCCTTGGCGTACGCTATG |

**Appendix Table 8. Oligonucleotides for NGS library preparation (PCR2).**

| Name | Sequence 5´→3´ |
|---|---|
| ILLUMINAseq_fwd | AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATC |
| ILLUMINAseq_rev_Index2 | CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA |
| ILLUMINAseq_rev_Index9 | CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA |
| ILLUMINAseq_rev_Index10 | CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA |
| ILLUMINAseq_rev_Index11 | CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA |

# Abbreviations

| | |
|---|---|
| **AA** | Amino acid |
| **Ad5** | Adenovirus type 5 |
| **ADCC** | Antibody-dependent cellular cytotoxicity |
| **ADCVI** | Antibody-dependent cell-mediated virus inhibition |
| **AIDS** | Acquired immunodeficiency syndrome |
| **BLI** | Biolayer interferometry |
| **BN-PAGE** | Blue native polyacrylamide gel electrophoresis |
| **bNAb** | Broadly neutralizing antibody |
| **BSA** | Bovine serum albumin |
| **C1-C5** | Constant regions 1-5 |
| **CcdB** | Control of cell death toxin B |
| **CCR5** | C-C chemokine receptor type 5 |
| **CD4** | Cluster of differentiation 4 |
| **CD4bs** | CD4 binding site |
| **CD4i** | CD4-induced |
| **CD40L** | CD40 ligand |
| **CDR** | Complementarity determining region |
| **CMV F+** | Flp-In™ T-REx™ 293 CMV_Furin+ |
| **CRF** | Circulating recombinant form |
| **CTL** | Cytotoxic T-lymphocyte |
| **CT** | Cytoplasmic tail |
| **CV** | Column volume |
| **CXCR4** | C-X-C chemokine receptor type 4 |
| **DLS** | Dynamic light scattering |
| **DMEM** | Dulbecco`s Modified Eagle`s Medium |
| **DPBS** | Dulbecco`s phosphate-buffered saline |
| **DSL** | Disulfide loop |
| **E. coli** | Escherichia coli |
| **ELISA** | Enzyme-linked immunosorbent assay |
| **Env** | Envelope glycoprotein |
| **eOD** | Engineered outer domain |
| **ER** | Endoplasmatic reticulum |
| **FACS** | Fluorescence-activated cell sorting |
| **FBS** | Fetal bovine serum |
| **FDA** | Food and Drug Administration |
| **FDR** | False discovery rate |
| **FNorm** | Normalized fluorescence |
| **FP** | Fusion peptide |
| **FRT** | Flp recombination target |
| **FWR** | Framework region |
| **Gag** | Group-specific antigen |
| **GC** | Germinal center |
| **gDNA** | Genomic DNA |
| **GFP** | Green fluorescent protein |
| **GNL** | Galanthus nivalis lectin |
| **gp** | Glycoprotein |
| **GS** | Glycine-serine |
| **HCDR3** | Heavy chain complementarity-determining region 3 |
| **HEK** | Human Embryonic Kidney |
| **HIV-1** | Human immunodeficiency virus-1 |
| **HR1/HR2** | Heptad repeat 1 and 2 |
| **HRP** | Horseradish peroxidase |
| **HVTN** | HIV Vaccine Trials Network |
| **IFN** | Interferon |
| **IgG** | Immunoglobuline G |
| **IL** | Interleukin |
| **ITC** | Isothermal titration calorimetry |
| **KD** | Equilibrium dissociation constant |
| **LB** | Lysogeny broth |
| **mAb** | Monoclonal antibody |
| **MCS** | Multiple cloning site |
| **MFI** | Mean fluorescence intensity |
| **MPER** | Membrane-proximal external region |
| **MST** | MicroScale Thermophoresis |
| **MW** | Molecular weight |
| **NAb** | Neutralizing antibody |
| **nef** | Negative regulatory factor |
| **NFL** | Native-flexible linker |
| **NGS** | Next generation sequencing |
| **NIH** | National Institutes of Health |
| **OD** | Outer domain |
| **OD 450 nm** | Optical density at 450 nm |
| **PBS** | Phosphate buffered saline |
| **PCR** | Polymerase chain reaction |
| **PEI** | Polyethyleneimine |
| **Pen/Strep** | Penicillin/Streptomycin |
| **PLA particle** | Poly(lactic acid) particle |
| **PDGFR** | Platelet-derived growth factor receptor |
| **Pol** | Polymerase |
| **QL** | QuickLigation |
| **R6** | RRRRRR |
| **RCI** | Regensburg Center for Interventional Immunology |

| | |
|---|---|
| **RT** | Room temperature |
| **sCD4** | Soluble CD4 receptor |
| **SDS-PAGE** | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| **SEC** | Size exclusion chromatography |
| **SEM** | Standard error of the mean |
| **SHIV** | Simian-human immunodeficiency virus |
| **SHM** | Somatic hypermutation |
| **SIV** | Simian immunodeficiency viruses |
| **SOSIP** | SOS: Disulfide-bond forming cysteines, connecting gp120 and gp41; IP: I559P substitution in gp41 |
| **SP** | Signal peptide |
| **SPR** | Surface plasmon resonance spectroscopy |
| **TBS** | Tris-buffered saline |
| **Tfh** | T-follicular helper |
| **Th** | T-helper |
| **TM** | Transmembrane domain |
| **TMB** | 3,3′,5,5′-tetramethylbenzidine |
| **TNF** | Tumor necrosis factor |
| **UNAIDS** | Joint United Nations Programme on HIV/AIDS |
| **URF** | Unique recombinant form |
| **V1-V5** | Variable loops 1-5 |
| **v/v** | Volume per volume |
| **VLP** | Virus-like particle |
| **WHO** | World Health Organisation |
| **WT** | Wildtype |
| **w/v** | Weight per volume |
| **293-F cells** | FreeStyle™ 293-F suspension cells |
| **6HB** | Six-helix bundle |
| **ΔFNorm** | Baseline corrected normalized fluorescence |

# Danksagung

An erster Stelle möchte ich mich ganz herzlich bei meinem Doktorvater Prof. Dr. Ralf Wagner bedanken, sowohl für die exzellente fachliche Betreuung als auch für die vielfältigen, spannenden Aufgaben und die stets angenehme Arbeitsatmosphäre.

Des Weiteren möchte ich mich bei Prof. Dr. Gernot Längst und Prof. Dr. Rainer Merkl dafür bedanken, dass sie diese Arbeit als Mentoren begleitet haben und mir mit Ratschlägen und Ideen zur Seite standen.

Bei Dr. Thomas Schubert und Corinna Reglin möchte ich mich ganz herzlich für die tolle Zusammenarbeit im MST Projekt bedanken.

Mein herzlicher Dank gilt auch allen aktuellen und ehemaligen Kollegen an der AG Wagner. Besonders erwähnt seien hier Dr. David Peterhoff und Dr. Benedikt Asbach, die immer ein offenes Ohr bei allen sich ergebenden Fragestellungen hatten und mir stets mit Rat und Tat zur Seite standen. Bei Vroni, Anja, Julia, Jogi, Richi, Tom und David möchte ich mich ganz herzlich für gemeinsame Zeit innerhalb und außerhalb der Arbeit bedanken. Besonders erwähnt seien hier unsere Faschings- und (leider viel zu seltenen) Angelausflüge. Meinen Laborkolleginnen Vroni und Christina danke ich für die tolle und lustige Zeit im 65er.

Zuletzt möchte ich mich von ganzem Herzen bei meiner Familie bedanken. Meiner Oma und meinem Bruder Fabian danke ich für die tollen gemeinsamen Wochenenden auf dem Hof und die mehr als gelungene Ablenkung von der Schreibarbeit…Coz we got nothing else to do! Bei meinen Eltern Johanna und Erwin möchte ich mich von ganzem Herzen für die Unterstützung, Geduld und Motivation auf meinem gesamten Lebensweg bedanken. Danke, dass ihr stets an mich geglaubt habt und es mir ermöglicht habt der Mensch zu werden der ich heute bin. Mein besonderer Dank gilt meiner Frau Kristina und meinem Sohn Adam dafür, dass ihr alle Höhen und Tiefen während der Entstehung dieser Arbeit mitmachen musstet und mir auch an frustreichen Tagen stets ein Lächeln ins Gesicht gezaubert habt. Ohne euch hätte ich das nicht hinbekommen!