STRUCTURE-GUIDED DESIGN AND BIOPHYSICAL CHARACTERIZATION OF NOVEL ANTI-HIV REAGENTS

Thesis by

Siduo Jiang

In Partial Fulfillment of the Requirements for the

degree of

Biology

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2014

ACKNOWLEDGEMENTS

Firstly, I would like to especially thank my thesis advisor Pamela for the amazing four years and two summers spent in her lab. Pamela gave the first seminar that I ever attended at Caltech on engineering antibody architectures. From that lecture, I discovered a new passion for protein structure and design, which under your close mentorship has evolved into an exciting new career path. I am greatly thankful for our conversations which are always educational for me, and especially the chance to prepare two theses and a draft of a manuscript under your guidance. The opportunity to join your lab and be your student has been one of highlights of my college career.

As the person that first interviewed me, I would especially like to thank Anthony for getting me even more excited about the lab. You were always there to keep me on track and instrumental in focusing my ideas into something worthwhile and tangible. It was great when we were initially making the gp120 mutants: my first constructs where one or two mutations actually made a huge difference in binding. Thank you again for being a phenomenal mentor and welcoming me to the lab.

And to Rachel, who has been an inspiration from the very start, from showing me how to clone, to enlightening my mind with brilliant ideas, and teaching me immunology. I could not thank you enough for being such an amazing mentor and friend. I always feel like I am constantly bugging you, but really, I would not be anywhere without all of your help and advice. It was always a blast working with you and hanging out both in and out of lab; I hope you will return to NYC before I leave there!

Of course, even though I have worked with Stuart for the least amount of time, your strict work style has rubbed off on me in the best way possible. I can no longer write more than 3 lines of code without annotating, and I cannot design primers without quadruple checking the sequence. Your knowledge of protein design is really inspiring, and I learned a tremendous amount from you and your excellent, constant feedback. Also, I hope I have not driven you crazy with sometimes trying to re-invent the wheel.

Finally, I am greatly thankful to all lab members for always lending a hand and making me feel welcome, even though I was the youngest and least experienced for the longest time.

ABSTRACT

Despite over 30 years of effort, an HIV-1 vaccine that elicits protective antibodies still does not exist. Recent clinical studies have identified that during natural infection about 20% of the population is capable of mounting a potent and protective antibody response. Closer inspection of these individuals reveal that a subset of these antibodies, recently termed potent VRC01-like (PVL), derive exclusively from a single human germline heavy chain gene. Induced clonal expansion of the B cell encoding this gene is the first step through which PVL antibodies may be elicited. Unfortunately, naturally occurring HIV gp120s fail to bind to this germline, and as a result cannot be used as the initial prime for a vaccine regimen. We have determined the crystal structure of an important germline antibody that is a promising target for vaccine design efforts, and have set out to engineer a more likely candidate using computationally-guided rational design.

In addition to prevention efforts on the side of vaccine design, recently characterized broadly neutralizing anti-HIV antibodies have excellent potential for use in gene therapy and passive immunotherapy. The separation distance between functional Fabs on an antibody is important due to the sparse distribution of envelop spikes on HIV compared to other viruses. We set out to build and characterize novel antibody architectures by incorporating structured linkers into the hinge region of an anti-HIV antibody b12. The goal was to observe whether these linkers increased the arm-span of the IgG dimer. When incorporated, flexible Gly4Ser repeats did not result in detectable extensions of the IgG antigen binding domains, by contrast to linkers including more rigid domains such as β2-microglobulin, Zn-α2-glycoprotein, and tetratricopeptide repeats (TPRs). This study adds an additional set of linkers with varying lengths and rigidities to the available linker repertoire, which may be useful for the modification and construction of antibodies and other fusion proteins.

TABLE OF CONTENTS

Chapter I

Structure-guided Engineering of Novel Vaccine Candidates Against HIV

In this study we report the crystallography of a germline anti-HIV antibody and preliminary design efforts for creating a novel vaccine candidate against HIV.

Background

The HIV pandemic continues to burden more than 30 million individuals worldwide with nearly 70% of all cases presenting in the developing countries of Sub-Saharan Africa1. While antiretroviral therapy (ART) has proven to be highly effective in controlling disease in much of the Western world, these African countries lack both the economic and clinical resources necessary to successfully deliver drugs to all patients in need. This is especially evident in South Africa, where HIV prevalence is as high as 50% in the regions of Kwa-Zulu Natal, with an overwhelming 21% of the population afflicted with severe AIDS. In spite of these statistics, only about 25% of the total positive population is receiving clinical attention.

Today, it is becoming evident that these expensive and unpleasant anti-retrovirals that provide at best transient protection is not a permanent solution[2]. Much work therefore has been devoted to the effort of developing a vaccine. To date, six HIV-1 vaccines have gone through Phase IIb/III efficacy trials and all have failed to significantly reduce viral transmission.

Table 1. Completed HIV-1 vaccine efficacy trials

Trial name	Vaccine candidate	Elicited Immunity	Reduction in new infections	Year completed
VAX004[5]	Single gp120	Antibody response	No reduction	2005
VAX003[6]	Bivalent gp120	Antibody response	No reduction	2006
HVTN502[7]	Adenovirus containing HIV-1 gag/pol/nef genes	Cellular response	No reduction	2007
RV144[9]	VAX003 plus canarypox virus containing HIV-1 env/gag/pol genes	Antibody and cellular response	31.2% efficacy	2009
HVTN503[9]	HVTN502; done in different country	Cellular response	No reduction	2011
HVTN505[10]	Adenovirus containing HIV-1 gag/pol/nef/env (gp120) genes	Antibody and cellular response	No reduction	2013

Table 1 summarizes these findings and compares the mechanism by which each candidate attempted to elicit immunity. Despite the importance of cell-mediated responses in preventing severe disease[3], it is considered only as a second line of defense after the antibody response. This is due to the fact that cell-mediated immunity can suppress viral production in already infected cells, but is incapable of preventing viral transmission and therefore new infections[4]. As a result, a potent antibody response will likely be a critical component of an effective vaccine.

Currently, many problems exist with eliciting an effective antibody response against HIV[3,11,12]. The most notable is the rapid mutational rate of the HIV Env gene encoding surface antigen gp120, which mediates viral entry, leading to massive diversification of the circulating viral strains within a single infected individual. Glycan shields and conformational masking of critical epitopes can also guard gp120 against immune recognition. No natural antigen to date (such as that used in VAX003/004) has been able to elicit broadly neutralizing antibodies (bNAbs) capable for tackling these diverse and structurally-protected strains[3]. The responses from past clinical trials have been weak overall[13], represented by the development strain-specific or non-neutralizing antibodies[14]. It is becoming evident therefore that a more targeted approach using rational design is necessary for generating an antigen suitable for use in an antibody-based vaccine.

In the past decade, clinical studies have identified that approximately 20% of HIV-1 infected individuals are capable of producing bNAbs, but the natural process takes about 2 years to be successful[15]. These antibodies, when delivered passively, are protective against infection in animal models. One of the most potent and broad class of these antibodies is the one which targets the natural gp120 ligand-binding site (the CD4-binding site)[16], and a subclass of these CD4-binding site antibodies, recently termed potent VRC01-like (PVL)[17], are a particularly promising class for vaccine design.

At a clinical scale, these PVL antibodies have been isolated independently from at least 6 individuals[10], which suggests that a broader population can develop this response. In addition, PVL heavy chains derive exclusively from the B cell germline IgV_H V_H1-2*02, and their light chains from three B cell germline IgV_Ks. This identification yields a promising starting point for vaccine design. As background, the process of antibody development begins with the successful coupling of a VJ-recombined[18] germline heavy chain gene with a VJ-recombined germline light chain gene forming the germline B cell receptor (receptor is a membrane-bound version of secreted antibody). Upon antigen-binding to this receptor, the germline B cell is activated and undergoes clonal expansion followed by somatic hypermutation[19], which diversifies the B cell receptor repertoire through genetic mutations. Positive selection for stronger antigen binders then takes place, thereby improving the average affinity of the B cell repertoire against the antigen. This process repeats a many times in the case of HIV, forcing the B cell receptor to evolve through several intermediate states before maturing into a potent broadly-neutralizing form.

Since all PVL antibodies originally derive from the same germline IgV_H, an effective vaccine would first induce activation and clonal expansion of that particular germline gene by utilizing a gp120 (antigen) that can strongly bind to that germline B cell receptor. Secondly, gp120s acting as booster shots should be administered in order to guide the process of somatic hypermutation. This would require antigens that bind to intermediates of the evolutionary pathway in order to put a selective advantage on B cells evolving in the correct direction. Since it is difficult to know with certainty the precise pathway of evolution, this task may be challenging. However, recent studies suggest that such a mechanism can be derived for specific cases with access to long-term patient data[10,20,21]. Furthermore, designed gp120s should all mimic the binding interface between natural gp120s and PVL antibodies. This would bias selection for intermediate antibodies preserving PVL structural features. If this can be achieved, there will be a better chance that the final product of these antibodies will be broadly neutralizing as the case with PVLs.

Crystallography of 3BNC60 germline

Presently, no natural gp120 reported in the literature appreciably binds to the V_H1-2*02 heavy chain coupled to any of its three germline light chain genes. A recently engineered gp120, named eOD-GT622, showed high nanomolar affinity for binding to the VH1-2*02 heavy chain coupled to the germline light chain precursor (IgKV3-20*01) of 3 potent PVLs, one of which was VRC01[10,17]. When multimerized onto nanoparticles, eOD-GT6 demonstrated moderate levels of B cell activation of the predicted germline in vitro, suggesting that this engineering method may be general for targeted activation of any desired germline receptor.

One major caveat with eOD-GT6 however is that the gp120 does not bind VH1-2*02 coupled to the germline light chain precursor of the PVL 3BNC60 (IgKV1-33*01), which also represents the precursor to 26 other PVLs[17]. This larger repertoire suggests that this precursor may be able to evolve potent antibodies using multiple pathways. Antibodies evolved from this germline also typically have fewer insertions in the CDRH3 and CDRL3 loops, likely indicating that they are faster to develop and easier to elicit.

In order to identify the bases for the lack of binding between the 3BNC60 germline and eOT-GT6, we determined the crystal structure of full 3BNC60 germline to 1.9Å resolution. Proteins for the construct were purified and concentrated down to 15mg/mL, and screened

using Crystal Screen, Index and PegRX, all from Hampton research. Phasing was performing by molecular replacement on a previous structure of the 3BNC60 germline, which lacked a critical “DISNY” insertion in the CDRL1.

As a result of this structure, we propose that a reason for the lack of binding between 3BNC60 germline and eOD-GT6 may be attributed to an extended CDRL1 loop in the light chain of the 3BNC60 precursor (unpublished), which sterically precludes binding in an aligned structure (Figure 1). This loop may also inhibit other VRC01 germline binders such as engineered 426c[23]. In addition, eOT-GT6 contains only the outer domain of gp120, which removes important contacts of the inner domain and the bridging sheets responsible for improved potency of PVLs[24]. In the mature PVL NIH45-46 for example, increased contacts with the bridging sheets and inner domain as measured by buried surface area constitutes a large portion of the interactions. Since eOD-GT6 does not encode a selective advantage for developing inner domain and bridging sheet contacts, these features would be neutral in the antibody developmental process. Figure 1 summarizes these points through comparison of the germline antibodies in complex with eOD-GT6 and a wildtype gp120, 93TH057.

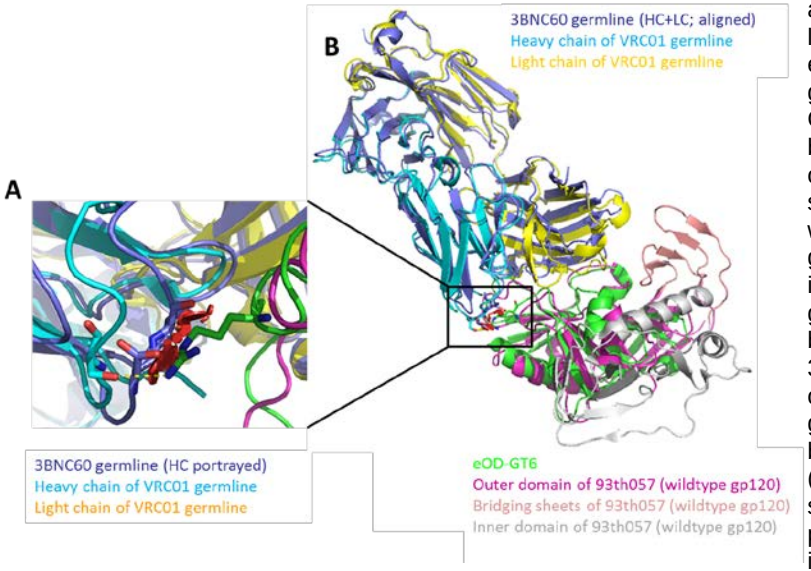


Fig. 1. Structural features of eOD-GT6 and 3BNC60 germline antibody. (*A*) A look at CDRL1 contacts. Currently, no engineered gp120 binds to the 3BNC60 germline. One possible reason is that the CDRL1 loop of 3BNC60 germline (dark blue) is more extended compared to that of VRC01 germline (light blue), which sterically precludes binding by clashing with loop residues (R81 on eOD-GT6 in green) on the gp120. Clashes are indicated in red. The CDRL1 of VRC01 germline makes a favorable hydrogen bond with the same Arg. (*B*) Alignment of 3BNC60 germline onto a crystal structure of eOD-GT6 in complex with VRC01 germline. eOD-GT6 does not have the bridging sheets (pink) or an inner domain (white), which in wildtype gp120 make significant contacts to increase the potency of mature PVLs. Such interactions should be engineered in future vaccine candidates.

Rational design of a 3BNC60 germline activator

In this project, I designed and tested a series of gp120s with the goal of improving their binding to the 3BNC60 germline (derived from V_H1-2*02 and IgKV1-33*01). The project utilized two primary tools: 1) computational design using the program Rosetta[25], 2) high-throughput biochemical techniques for validating and evolving potential binders.

We began by reconstructing the V5 loop which was not resolved in the original unliganded crystal structure of the entire core construct. This was done by manually inserting the residues in PyMOL, followed by a loop closure and refinement protocol implemented in Rosetta1. The refined NCSYU2 was then superimposed onto the crystal structure of eOD-GT6 (PDB: 4JPK)[2] replacing the eOD portion (engineered outer domain of NCS93TH057) of the co-crystal forming the VRC01GL+NSCYU2. This entire structure was then subjected to a high resolution refinement protocol in Rosetta3 in order to generate an energetically stable complex as the starting point for design. The RMSD of the entire refined structure was within 0.5Å of the original docked VRC01GL+NCSYU2. Though eventually we wish to work with variously starting gp120 cores (including NCSYU2) and 3BNC60GL, this structure was easy enough to obtain as a test structure for generating the appropriate downstream scripts needed for operation on the final complexes.

Before re-designing the complex for greater stabilization, we needed to first define the residues to be subjected to multiple rounds of mutations and minimization. This was done using a PyMOL script which identifies the interface residues of the gp120 with either the HC or LC of the germline Ab (interface defined as residues of the gp120 with atoms which are within 10Å of an atom of either the HC or LC). These residues were identified, and a script which exports these residues in a Rosetta executable “resfile” was generated. The “resfile” stipulates that the defined interface residues may be mutated to any other residue (or remain unmutated) except cysteine. The script for this entire process is reliably and quickly applied to any starting structure.

The first step to design in Rosetta is to generate a task manager that simultaneously couples side-chain mutagenesis with rotamer re-packing post mutagenesis. The function used is named “PackResidueMover,” which we manipulated to read instructions of the “resfile.” The key to design is the “score function” of the protein complex, which accounts for biophysical energy terms such as the net van der Waals forces, hydrogen bonding and

solvation energy as well as statistical terms such as Ramachandran propensities. The score function calculates the overall energy of a complex depending on the weights of these parameters specified in the score function object; for example, a complex with several potential residue clashes would score a much higher global energy if the net van der Waals repulsive term is increased in the score function.

A “standard” set of parameters for the score function was defined in one of the first successful applications of Rosetta[4] and became the basis for initial cycles of scripting testing and optimization. In addition to the design (mutagenesis) and repacking steps, we incorporated a series refinement movers which, after each repacking, makes slight adjustments to backbone torsion angles of the design residues in order to minimize energy. These movers accomplish this task by seeking out a steepest decent path toward a local minimum and accepts changes based off of the Metropolis Monte Carlo criterion[5]. Because of this, the global minimum may not always be achieved, but this may be accomplished with the high resolution refinement protocol previously used to refine the docked (a single application of this protocol takes over ~6 hours to run on the fastest lab machine so it is only used sparingly).

Eventually, we were able to generate a script which iterates over many cycles of design and repacking as opposed to a single round. Such a strategy suggests that there may be a more sensible approach to defining the score function used to evaluate each step of the design. A gradient-based score function was generated based off of the following logic. First, a decreased van der Waals repulsive terms was used while an increased attractive term was used so that the very initial refinement steps do not completely relax and “blow up” the structure. In other words, favorable interactions may be established initially without the worry that simultaneous generation of clashes prevents these favorable interactions to be accepted. This process is then reversed for later steps where after defining favorable interactions, the program attempts to correct unfavorable clashes. Statistical terms also follow a gradient path from having a low weight initially to a large weight in the later steps. This allows for sampling of residues at positions not generally employed by nature, which may be beneficial. The entire design, repacking, minimize procedure is interspaced with high resolution refinement, and acceptance of each intermediate structure is based off of Metropolis criterion. In order to gain a better understanding of the evolution of the gp120, the script commands that each generated structure, regardless of whether it is accepted,

be outputted as a PDB file. Lots of debugging and many overnights rounds of script testing took place with this entire protocol but a full script is ready to use for design and refinement of any starting structure.

One issue we have come across is that glycine interface residues, regardless of the number of iterations, are not subjected to mutations by the program. This may be that some residues are an absolute requirement at certain positions, but may also be that flexibility of these residues confuses the program by trapping the structure at a local minimum, and that the steps undertaken during each Monte Carlo design procedure is not large enough to escape this minimum. As a result, we are currently testing the potential of manually mutating all glycine residues to alanine initially, and increasing Ramachandran propensity weight term in the score function to ensure that positions which strictly require glycine readily mutate back.

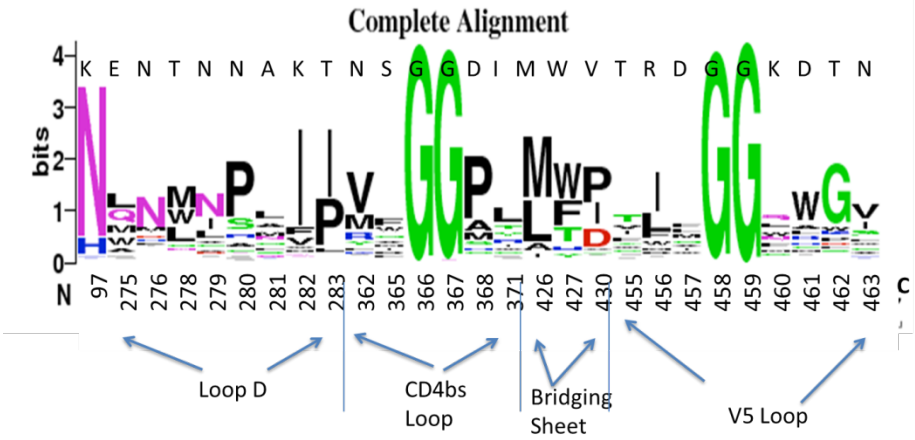
For our designs, Rosetta was used to narrow the search space of possible gp120s by computationally optimizing the germline-antibody interface. The basic computational procedure started with the recently solved crystal structure of the 3BNC60 germline as shown in Figure 1 docked to a wildtype gp120. This complex was modeled using an alignment of the germline antibody to crystal structure of a mature antibody-gp120 complex. An ensemble of such starting complexes was generated using rigid-body perturbations in Rosetta in order to represent some of the possible accessible states of an unstable starting complex. For each modeled complex, the residues which lie in close proximity to the antibody-gp120 interface was identified based on the residue's buried surface area. From this analysis, Rosetta identified 28 residues on which to perform design calculations. Rosetta proceeded to randomly mutate these residues one at a time while optimizing rotomer conformation using fixed-backbone design. Each round of mutation was coupled with a backrub mover which allows the protein backbone to make adjustments. The energy of the designed complex was calculated using supplied values as well as potential functions internal to Rosetta, and each mutation is accepted or rejected based on the Metropolis criterion. For each residue, about a thousand mutations and rotomer optimizations were attempted. The design then proceeded through each residue and the energy for the final complex was generated. In order to allow multiple design trajectories, each initial perturbed complex was cycled through the design procedure several times to generate greater diversity.

In order to supply Rosetta with desirable parameters for energy calculations, guidelines from past structural studies was employed. Characterizations of PVL antibodies, including 3BNC60, have identified several exposed signature residues preserved across all PVLs that are critical residues for gp120 binding and HIV neutralization[17]. These residues mostly form hydrogen bonds near and around the CD4-binding site of the gp120. In order to mimic natural binding, these interactions need to be built and maintained in the design procedure. This was one of computational screening criteria used to identify desirable final structures.

At the end of the design procedure, the primary set of information that can be extracted was the preferred amino acid usage, or type of amino acid, at each of the designed interface positions. This includes critical residues for PVL contact, as well as other residues which potentially favor packing or build hydrophobic interactions. This was analyzed using a home-written script that identifies correlations of preferred amino acid usage at the designed positions.

Figure 2 presents the calculated amino acid preference at each of the designed positions. Some positions are highly consistent in amino acid usage; for example residues 458 and 459 are preserved at Gly due to Ramachandran constraints at that position. Other positions are highly variable with a variety of hydrophobic, polar or charged residues.

Figure 2. Preferred amino acid usage at Rosetta designed positions using fixed-backbone design.



With this information, libraries of randomized gp120s containing all possible combinations of these preferred amino acids was generated using degenerate primer Gibson cloning[27].

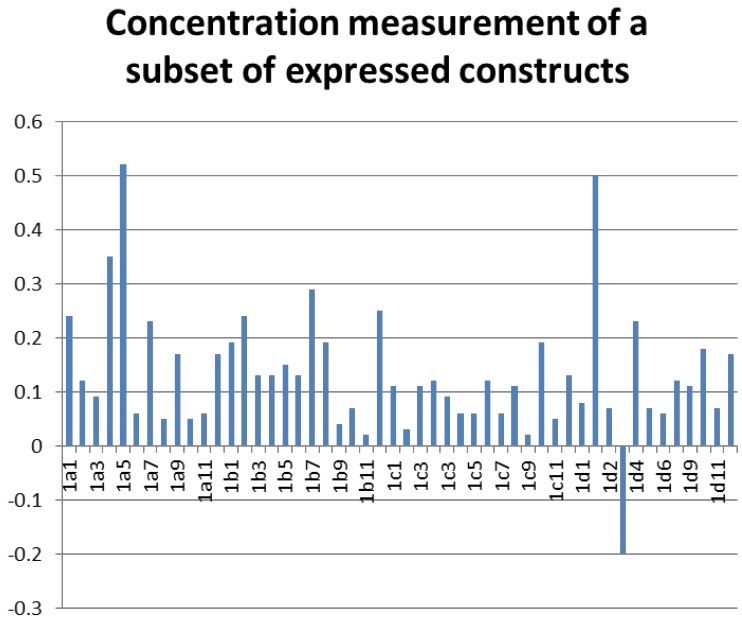
Table 2 presents all of the potential mutations each position could become in this artificial library. The greatest diversity was generated in the CDR3 region of the antibody, which correspond to residues 455-463. Residues 458 and 459 were Gly residues with constrained torsion angles, which were identified by Rosetta to be necessary to maintain.

Table 2. Potential mutations in engineered gp120 library following guideline from Rosetta calculations.

Position	Original amino acid	Potential amino acids in engineered gp120 library
275	Glu	Glu, Met, Leu, Trp, Gln
276	Asn	Asn, Met, Thr, Asp
278	Thr	Thr, Met, Trp
279	Asn	Asn, Asp
280	Asn	Asn, Pro, Ala
281	Ala	Ala, Gly, Leu, Met, Asn, Ser, Phe
282	Lys	Lys, Ile, Met, Val, Phe, Trp
283	Thr	Thr, Ile, Leu, Val, Pro
362	Asn	Asn, Val, Met, Arg, Tyr, Phe
365	Ser	Ser, Phe, Trp
368	Asp	Pro, Ala, Met, Ser, His
371	Ile	Ile, Tyr, His, Thr
426	Met	Met, Leu, Met, Ala, Gly,
427	Trp	Trp, His, Thr, Tyr, Phe
430	Val	Val, Pro, Ile, Asp
455	Thr	Thr, Val, Ile, Val, Ile, Phe, Tyr, Trp
456	Arg	Arg, Leu, Ile, Met, Trp, Phe, Tyr
457	Asp	Asp, Phe, Tyr, Ile, Leu, Met
460	Lys	Lys, Gln, His, Leu, Met, Trp, Tyr
461	Asp	Asp, Ala, Gly, Phe, Tyr, Trp, His, Asn
462	Thr	Thr, Arg, Lys, Trp, Ala, Gly, Leu, Met, Phe, Tyr
463	Asn	Asn, Asp, Leu, Met, Ile, Leu, Val, Gly

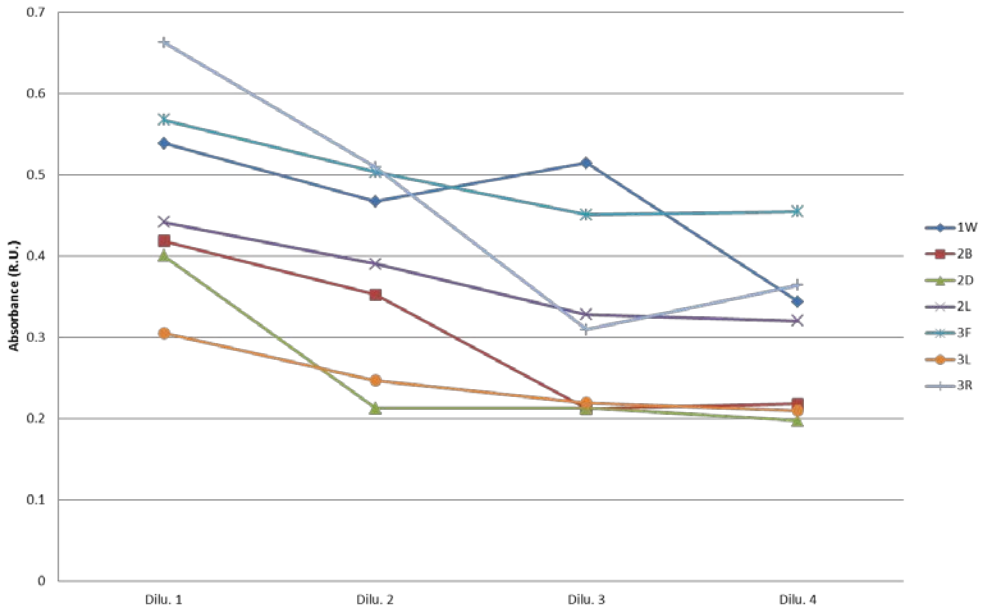
This gp120 library was screened recombinantly expressed using standard mammalian systems and tested for binding affinity using ELISA assays. The expression was performed in GnTI- cells for reduction of complex N-glycans. Expression was done in 4mL cultures in 24-block format and harvested by centrifuging at 2000xg for 30min. Due to the large number of potential mutations at each position, the entire library was very large and impractical to screen completely. A subset of 69 clones was randomly chosen in the first round of screening. Figure 3 shows OD measurements for a subset of expressed constructs.

Figure 3. A concentration measurement for a subset of expressed constructs.



The recent installment of a robotic system for performing large-scale expression and ELISA screening allowed automation in the entire procedure. ELISAs were performed on the panel of engineered constructs for binding to the 3BNC60 germline, and a sample binding curve is given in Figure 4. No significant binders were identified on this initial round of screening which suggested that we needed more stringent selection and design criteria in the computational step.

Figure 4. ELISAs for testing the binding of engineered gp120s to the 3BNC60 germline.



Design of an NIH45-46 germline chimera activator

NIH45-46 is another highly potent antibody against HIV and vaccine candidates that can induce production of similar antibodies would be a highly effective preventative medicine. The germline chimera is a version of the antibody in which mature light chain is paired with the germline heavy chain. This construct demonstrates better binding to most gp120s than the germline version alone, which allows us to generate a rank order of weak binders that otherwise would be undetectable in the complete germline background. In order to design a gp120 that can bind to this germline chimera, we started with a computational approach to approximate the positions of germline HC and LC residues in an antibody-gp120 co-crystal structure using existing structures from hypermutated antibodies-gp120 complexes. These include the co-crystal structures of NCS93TH057 (gp120) in complex with NIH45-46, VRC01, VRC03 and VRC-PVG09, the first three being highly potent broadly neutralizing antibodies. The method carried out in Rosetta applies a grafting technique in which the carboxyl-amino backbone of the protein complex is assumed to be fixed during residue reconstruction, and side-chain rotomers are chosen based off of the lowest energy conformation out of ~6000 possible conformations. Upon generating this initial structure, visual inspection allowed us to pinpoint regions in the structure where mutations might lead to a greater favorability in complex formation. These mutations were then generated

in silico and the new structure was scored based on a standard free energy score function. A new structure with the mutation was then generated and visual inspection applied to see if the desired effect was achieved. This process was repeated several times until a list of potential favorable mutants was generated. The complete list of mutations designed and constructed by this process is given in Table 3, along with rational for the mutation. Figure 5 shows two examples of residue replacement from the reconstructed 45-46GL-NCS93TH057 complex that may lead to a salt bridge formation and thus improved binding.

Table 3. Initial list of mutations and SPR results of the first set of mutations.

#	Mutation	Rational
1	S365Q	Increased salt bridge interaction with HC 65Q of 60GL
2	S365Y	Hydrophobic alternative to compare with S365Q
3	T455Q	Improved polar interactions with nearby residues
4	T455Y	Hydrophobic alternative to compare with T455Q
5	T460Q	Improved polar interactions with nearby residues
6	T467Q	Increased salt bridge interaction with HC 65Q of 60GL
7	A460Q, T463A, S464A, T467Q	Eliminate N-linked glycosylation sites at 461, 462 and 465, a primary interface between the gp120 and the GL antibody.
8	A460Q, N462Q, T463A, T467Q	Second mechanism to eliminate N-linked glycosylation sites at 461, 462 and 465.
9	Delete: 462, 464; T467Q	Decrease the size of the V5 loop to 7 residues. Most gp120s have a loop size around 5-6 aa residues, and shortened V5 length is associated with better binding. NCS93TH057 has 9 aa residues. Additionally, test T467Q in context of loop size reduction.
10	Delete: 462, 464, 465	Decrease the size of the V5 loop to 6 residues.
11	Delete: 462, 464; N279D, T467Q	Out of all of the existing gp120s, N and D account for nearly 100% of the residue identity at 279. Defective PVL antibodies tend to prefer D at 279, which eliminates a glycosylation site and suggest that this mutation may provide further assistance to binding. This was superimposed onto the V5 loop of mutant 9 after the first round of binding.
12	Delete: 462, 464; T455Q, T467Q	Combining mutants 3 and 9 which initially demonstrated improved binding to the GL and chimera precursors.

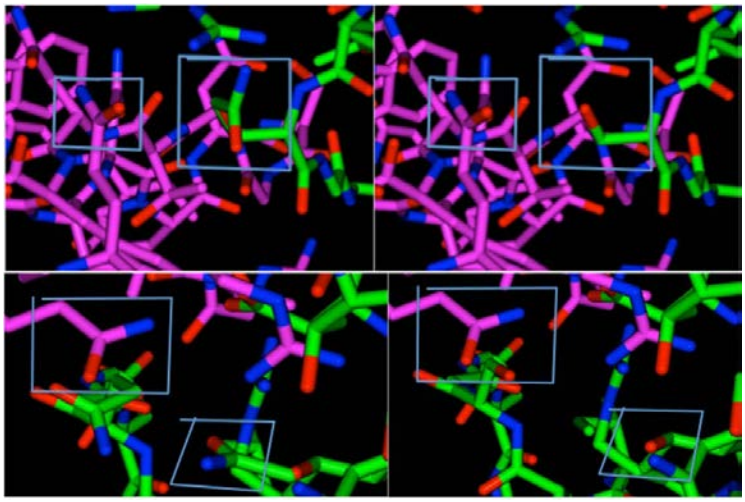


Figure 5. Comparison of reconstructed germline antibody binding to NCS93TH057 with S365Q (top) and T467Q (bottom) point mutants. Left images represent the mutants with added Q residue boxed; right images represent NCS93TH057 with S or T residues boxed. HC of the reconstructed germline is in purple. NCS93TH057 and the mutants are in green. In all four images, HC residue of the germline 64Q may interact favorably with the point mutants through salt bridge formation while having little interaction with the native gp120. The S365Q mutation was modeled after the reconstruction of VRC01, with an overall energy difference of -2.02. The T467Q mutant was a model of the reconstruction of the germline in the NIH45-46 scaffold, and an energy difference of -5.64.

In order to determine the binding affinity of the mutants to the germline and chimeric versions of the precursor antibody, two binding assays were used: a gp120 ELISA and SPR. Both protocols are detailed in the methods section. The mutants were initially characterized using a gp210 ELISA due to the ease and availability of the assay. This was performed in Figure 6. The results were not completely decisive however due to the overall low affinity of the constructs, but construct 9 (mutant 9) which later demonstrated greatest affinity for the germline and chimera percursors, also appears to be the best mutant here in the chimera ELISA. Due the overall weak affinity and insensitivity of this assay, SPR was then performed to further characterize binding. SPR is much more sensitive to changes in binding, and the binding constants generated represent binding constants relative to the wildtype (also known in Figure 6 as T278A, the initial mutant generated from a post-doc who pursued this project before leaving).

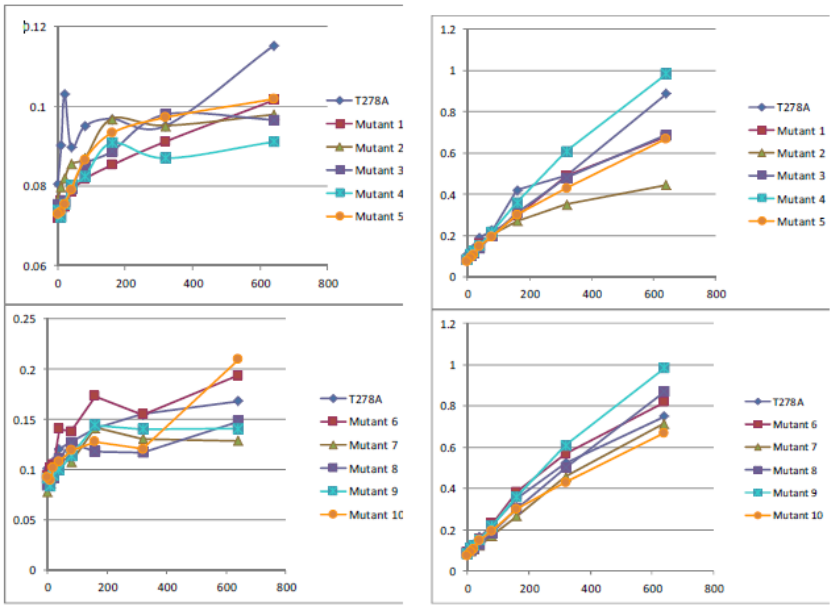
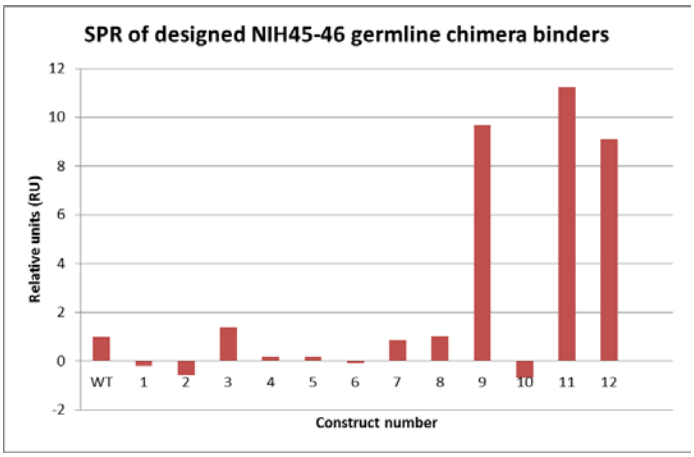


Figure 6. Binding curves of NCS93TH057 mutants using a gp120 ELISA assay. Top: Mutants 1-5 along with T278A control against pure germline (Left) and germline chimera (Right). Bottom: Mutants 6-10 against pure germline (Left) and germline chimera (Right). In all of the mutants along with the control, the pure GL antibody does not bind very well and binding coefficients appear to be below the sensitivity of this assay. In the case of the chimera GL, potential improvements are seen in mutants 9, 6, 8 and 4, though again due to the overall poor binding of the GL and GL chimera to these gp120s, it is difficult to tell for certain. As a result, more sensitive assays such as SPR are planned for the future. One potential immediate assay is to repeat the ELISAs but with a high concentration of GL and GL chimera antibodies applied before detection, as it appears that only about around 320ng of the antibody is there a significant difference observed between each of the mutants with respect to the control. A range from 125ng-2000ng might be more appropriate for this assay. Finally, depending on the result of SPR, double mutants may be generated, or new designs will have to be sought to achieve the ultimate goal of an antigen that potently binds to the germline.

Finally, SPR was performed on these designed constructs as shown in Figure 7. Construct 9 demonstrated the greatest promise, along with constructs 3. Combining the mutations from construct 9 and 3 did not yield significant improvement to the binding; however, construct 11, which was construct 9 but with a N276D mutation, increased binding to the NIH45-46 chimera by more than 10-fold.

Figure 7. SPR analysis of engineered gp120 binding to NIH45-46 germline



Design of a complete vaccine

The gp120 with the designed interface will likely be used only as the initial antigen to prime activation, induce clonal expansion, and direct the first rounds of somatic hypermutation of the 3BNC60 germline B cell receptor. This is due to the fact that broadly-neutralizing antibodies against HIV require drastic changes in sequence and CDR loop length, and a single antigen most likely cannot encode the structural features to drive all such changes. A complete vaccine strategy would likely require a series antigens optimized for binding to the intermediate states of the antibody. As discussed previously, the roadmap for how a PVL antibody, or any HIV antibody, evolves into its final form is only starting to be written. Generation of a complete vaccine strategy will likely require additional clinical data defining these natural pathways. Though outside the scope of this particular project, rational gp120 design as discussed for this project will likely be an essential tool in process, providing the platform for constructing molecular features into an antigen in order to direct the evolution of a potent protective descendent.

We have also run a second designed and have begun screening the results with stringent selection criteria. The primary difference in this round of design is the relaxation step. Rather than performing a high resolution large-scale relaxation on the mutated gp120-3BNC60 germline complex, we utilized shear and small movers which screen for small torsion angle adjustments and accept or reject them based on the Metropolis criterion. This is much more suitable since we already have a good idea of how the structure should look from the initial alignment. The selection criterion also looks for complexes in which the PVL signature residues[17] are key players in mediating contact with

the engineered gp120. The structures that emphasize binding with these residues are more likely to be good candidates both in terms of computational accuracy as well as a final candidate for a vaccine.

Improvements can also be made in the future for engineering gp120s. Currently, the design protocol allows only for mutations in the existing structure. However, given the variability in length of the loops in the gp120s and their drastic effects on binding (which we saw when we truncated the V5 loop of 93TH057), an ideal program should be able to generate and score insertions and deletions in the loop regions of the gp120. We're currently working on a protocol that can limit these changes to the loop region, and effectively re-pack the side-chain rotomers.

If such a protocol can be generated, we may be able to further expand the script to simultaneously evolve both the germline antibody and the gp120 (potentially using experimental guidelines from [6]). Since maturation toward a potent neutralizer will likely take place over several cycles of B cell selection, we would like to search for shortest paths of this evolution by identifying potential intermediate forms of the maturing antibody, and then designing the ideal immunogen candidate to target each stage. The process will hopefully be identifiable with somatic hypermutations in that out of an ensemble of mutated germline antibodies, a few will be selected for further maturation with evolving gp120.

Citations

1. WHO, U. AIDS Epidemic Update. (2010)

2. Andrews, Jason R., et al. "Projecting the benefits of antiretroviral therapy for HIV prevention: the impact of population mobility and linkage to care." Journal of Infectious Diseases 206.4 (2012): 543-551.

3. Walker, Bruce D., and Dennis R. Burton. "Toward an AIDS vaccine." science 320.5877 (2008): 760-764.

4. Burton, Dennis R., et al. "A blueprint for HIV vaccine discovery." Cell host & microbe 12.4 (2012): 396-407.

5. Flynn, N. M., et al. "Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection." The Journal of infectious diseases 191.5 (2005): 654-665.

6. Pitisuttithum, Punnee, 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." Journal of Infectious Diseases 194.12 (2006): 1661-1671.
7. Barouch, Dan H., and Bette Korber. "HIV-1 vaccine development after STEP." Annual review of medicine 61 (2010): 153.
8. Rerks-Ngarm, Supachai, et al. "Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand." New England Journal of Medicine 361.23 (2009): 2209-2220.
9. Gray, Glenda E., et al. "Safety and efficacy of the HVTN 503/Phambili study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study." The Lancet infectious diseases 11.7 (2011): 507-515.
10. Hammer, Scott M., et al. "Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine." New England Journal of Medicine 369.22 (2013): 2083-2092.
11. Kulp, Daniel W., and William R. Schief. "Advances in structure-based vaccine design." Current opinion in virology 3.3 (2013): 322-331.
12. Walker, Bruce D., and Dennis R. Burton. "Toward an AIDS vaccine." science 320.5877 (2008): 760-764.
13. Moody, M. Anthony, et al. "HIV-1 gp120 vaccine induces affinity maturation in both new and persistent antibody clonal lineages." Journal of virology 86.14 (2012): 7496-7507.
14. McElrath, M. J. & Haynes, B. F. Induction of immunity to human immunodeficiency virus type-1 by vaccination. Immunity 33, 542–554 (2010).
15. Kwong, Peter D., John R. Mascola, and Gary J. Nabel. "Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning." Nature Reviews Immunology 13.9 (2013): 693-701.
16. Haynes, Barton F., et al. "B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study." Nature biotechnology 30.5 (2012): 423-433.
17. West, Anthony P., et al. "Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120." Proceedings of the National Academy of Sciences 109.30 (2012): E2083-E2090.
18. Dreyer, W. J., and J. C. Bennett. "The molecular basis of antibody formation." Proc. Natl. Acad. Sci. USA 54 (1965): 864-869.
19. Haynes, Barton F., et al. "B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study." Nature biotechnology 30.5 (2012): 423-433.

20. Sok, Devin, et al. "The Effects of Somatic Hypermutation on Neutralization and Binding in the PGT121 Family of Broadly Neutralizing HIV Antibodies." PLoS pathogens 9.11 (2013): e1003754.
21. Liao, Hua-Xin, et al. "Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus." Nature 496.7446 (2013): 469-476.
22. Jardine, Joseph, et al. "Rational HIV immunogen design to target specific germline B cell receptors." Science 340.6133 (2013): 711-716.
23. McGuire, Andrew T., et al. "Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies." The Journal of experimental medicine 210.4 (2013): 655-663.
24. Jardine, Joseph, et al. "Rational HIV immunogen design to target specific germline B cell receptors." Science 340.6133 (2013): 711-716.
25. Scharf, Louise, et al. "Structural basis for HIV-1 gp120 recognition by a germ-line version of a broadly neutralizing antibody." Proceedings of the National Academy of Sciences 110.15 (2013): 6049-6054.
26. Kuhlman, Brian, et al. "Design of a novel globular protein fold with atomic-level accuracy." Science 302.5649 (2003): 1364-1368.
27. Gibson, Daniel G., et al. "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nature methods 6.5 (2009): 343-345.

Chapter II

Design and characterization of polypeptide linkers for extending the arm span of an IgG antibody

In this study we report the design and characterization of a variety of structured linkers in the context of an anti-HIV antibody b12.

Introduction

Fusion proteins are engineered biomolecules containing parts from two or more genes synthesized as a single multi-functional construct. These have been critical in many areas of biological research including affinity purification (Lichty *et al.*, 2005) and protein stabilization for structure determination (Zou *et al.*, 2012). Biopharmaceuticals also employ bi-specific reagents in which the active drug domain is fused to a carrier domain allowing for the drug's proper transport (Chen *et al.*, 2013). Such proteins have been designed to penetrate epithelial membranes such as the blood brain barrier, as well as to target a specific cell population (Pardridge, 2010). Due to the modularity of protein domains in the generation of functional constructs, fusion proteins will likely have increasing importance in research and drug design.

The successful construction of fusion proteins relies on the proper choice of a protein linker as direct fusion of two domains can lead to compromised biological activity (Bai *et al.*, 2005, Zhang *et al.*, 2009). Several studies have utilized existing databases to compile and characterize linkers in naturally occurring multi-domain proteins (Argos, 1990, George and Heringa, 2002). These sstudies have yielded amino acid sequence propensities for natural linkers of various sizes and lengths, as well as information on rigidity and secondary structure. This information has helped the design of empirical linkers that are customized for particular applications.

Empirical linkers can be classified into three groups; flexible linkers, rigid linkers and cleavable linkers (Chen, Zaro and Shen, 2013). Flexible linkers are generally composed of small, non-polar or polar residues such as Gly, Ser and Thr. The most common is the $(Gly_4Ser)_n$ linker (Gly-Gly-Gly-Gly-Ser)$_n$, where n indicates the number of repeats of the GGGGS motif. Poly-glycine linkers have also been generated, but the addition of a polar residue such as serine can reduce linker-protein interactions and alleviate disturbance of protein function. Due to their flexibility, these linkers are unstructured and thus provided limited domain separation in a previous study (Evers *et al.*, 2006). As a result, more rigid linkers including poly-proline motifs (Schuler *et al.*, 2005) and an all alpha-helical linker $A(EAAAK)_nA$ (Arai *et al.*, 2001) have been developed.

We are interested in using relatively rigid protein linkers to separate anti-HIV binding proteins at distances that would permit bi- or multivalent binding to HIV Env glycoproteins.

The idea is to create reagents capable of cross-linking within a single Env trimer (intra-spike crosslinking). Such reagents would take advantage of avidity effects to minimize HIV's ability to evade neutralizing antibodies by rapidly mutating to lower the affinity between the HIV epitopes and the antigen recognition fragment (Fab) of the antibody (Klein *et al.*, 2009). Though the architecture of the HIV spike trimer does not permit intra-spike cross-linking by most natural antibodies (Klein and Bjorkman, 2010, Zhu *et al.*, 2006), it should be possible to create reagents separated by linkers of the appropriate length that would permit bivalent binding; i.e., binding to two sites on an HIV Env trimer, by either two identical reagents or two different reagents. Here we report the design, construction and characterization of a series of structured protein linkers incorporating both rigid and flexible domains that can be used to achieve a variety of different desired separations. The linkers were incorporated into the hinge region of an intact IgG antibody and then evaluated for their relative lengths and rigidities by dynamic light scattering.

Methods

Plasmid construction and protein purification

Genes encoding designed linkers were synthesized (Blue Heron) with restriction sites for the enzymes NheI and either NgoMIV or HindIII. These sites were also introduced into the gene encoding the heavy chain of the HIV-neutralizing antibody b12 (Roben *et al.*, 1994) such that the insert would be located between hinge region residues His235 and Thr236. Linker genes were subcloned into the b12 heavy chain gene, and the genes encoding the b12-linker fusion proteins were subcloned into the pTT5 mammalian expression vector. b12-linker IgGs were expressed transiently in HEK-6E cells by co-transfecting the b12-linker heavy chain genes with the b12 light chain gene as described (Diskin *et al.*, 2011).

IgG-linker fusion constructs were purified by protein A affinity chromatography (GE Healthcare) followed by purification and analysis over size exclusion chromatography (SEC) using a Superdex 200 10/300 GL column (GE Healthcare) in phosphate-buffered saline, 0.05% w/v sodium azide, pH 7.4.

Dynamic light scattering (DLS)

Fractions corresponding to the center of the SEC elution peak were concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore) with a molecular weight cutoff of 100kDa

to a volume of 80-400uL and concentrations of 0.5-1mg/mL. Concentration differences within this range did not appear to affect the hydrodynamic radius values determined by DLS (data not shown). Sample sizes ranging from 80-350µL were loaded into a disposable cuvette, and measurements were performed on a DynaPro® NanoStar™ (Wyatt Technology) using manufacturer's suggested settings. A fit of the second order autocorrelation function to a globular protein model was performed in order to derive the hydrodynamic radius.

Results and Discussion

Design and identity of designed linkers

In order to design potential structured linkers, we surveyed the Protein Data Bank (PDB) to find structures that were relatively elongated and rigid, or represented small globular proteins, with the idea of joining these in various combinations with short flexible linkers. For a relatively rigid, elongated structure, we chose Zn-α2-glycoprotein (ZAG; PDB code: 1ZAG) (Sanchez *et al.*, 1999), and as examples of small globular proteins, we chose β2-microgloblin (β2m) and ubiquitin (Ub; PDB code: 1UBQ). β2m is a stable 12kDa protein with an immunoglobulin constant region-like fold that forms a rigid structure with a separation distance between the N- and C-terminus of approximately 35Å (Trinh *et al.*, 2002). Likewise, Ub is a compact, stable 8.5kDa protein with an N-terminal and C-terminal separation distance of about 37Å (Vijay-Kumar *et al.*, 1987). Proline-rich linkers were also generated using the hinge sequence from IgA1, which is a glycosylated peptide that is able to confer rotational flexibility on the Fab relative to the Fc in the context of wildtype dimeric IgA1(Bonner *et al.*, 2008). ZAG, β2m and Ub proteins were joined in various combinations with short linker regions, either $(Gly_2Ser)_n$ repeats or proline-rich sequences, to create linkers L1 – L12 (Table 1).

We also created linkers (L13 – L16; Table 1) using tetratricopeptide repeat domains (TPRs; PDB code: 2AVP) (Kajander *et al.*, 2007) that are found in natural proteins such as HSP70/90 (Scheufler *et al.*, 2000). These domains are optimal for use as potential structured linkers because the length of a set of tandem TPR domains corresponds predictably with the number of repeats. Each repeat consists of 34 amino acids with a defined sequence motif that forms two α-helices (D'Andrea and Regan, 2003). 7-8 TPRs form a complete superhelical turn with a pitch of about 72Å. For our TPR constructs, we

used a consensus sequence defined by the amino acid of the greatest global propensity in the natural database of the TPR domains at each position, which was shown to form a stable superhelix and was therefore named the consensus TPR sequence or cTPR (Main *et al.*, 2003).

For comparisons, we constructed a series of $(Gly_4Ser)_n$ linkers (L17 – L24; Table 1) in order to determine the effect of increasing the number of flexible Gly_4Ser repeats on the hydrodynamic radius of the IgG. Table 1 summarizes the sequence identity of our designed set of linkers. The complete sequence of each linker is given in Table S1.

As a scaffold for comparing the designed structured linkers, we inserted each into the hinge region of an intact IgG antibody (the anti-HIV antibody b12) (Roben et al. citation). The hinge region of an IgG, which encompasses the amino acids between the C-terminus of the heavy chain portion of the antigen-binding fragment (Fab) and the N-terminus of the Fc, can tolerate large protein insertions (Figure 1).

Characterization of the IgGs containing structured linkers

b12 IgG proteins containing linkers L1 – L24 were expressed by transient transfection in mammalian cells and purified by affinity and size exclusion chromatography. Visualization by SDS-PAGE for IgGs containing the L1 – L8 linkers showed that all proteins were purified to >95% homogeneity (Figure 2). Under reducing conditions, two heavy chain bands were observed for b12-L1, which included a linker containing three potential N-linked glycosylation sites, indicating the presence of multiple glycosylated isoforms. An overlay of the chromatograms derived from SEC showed that the IgGs containing the L1 – L8 structured linkers all exhibited a decrease in retention volume relative to wild type IgG, consistent with the expected increases in the radius of gyration (R_g) of each of the constructs due to the addition of a structured linker.

We next derived the hydrodynamic radii using dynamic light scattering (DLS) for wildtype b12 and the b12 proteins containing designed linkers. DLS measures fluctuations in the intensity of scattered light of a protein solution over time, which can be used to calculate an autocorrelation function of intensity (Nobbmann *et al.*, 2007). Typical monodisperse samples (including our hinge-linked antibodies) generate an exponential decay in the autocorrelation. A least squares fit can be performed to calculate the decay constant, which directly relates to the diffusion coefficient. The diffusion coefficient is then inversely related

to the characteristic hydrodynamic radius R_H, which reflects the radius of a hypothetical solid sphere that would diffuse at the same rate as the protein. The R_H value is not a direct measurement of the length that the linker contributes to the size of the IgG. However, comparative analysis can yield rank order differences for the relative lengths and rigidity of the various linkers. For example, if the separation between the IgG Fc and Fab domains were increased by the addition of a designed hinge linker, we would expect an observable increase in the R_H of the fusion construct compared to the parental b12 IgG due to increased size of the diffusion sphere.

The b12-linker proteins were separated into three groups, and the hydrodynamic radii were measured for each of the IgG-linker fusion proteins and compared with an internal wildtype b12 IgG control. The first group consisted of L1-L12, which were designed fusion constructs containing elongated or small protein domains separated by internal flexible linking regions. The second group consisted of L13-L16, which contained linkers constructed using different numbers of cTPRs. The final group comprised b12 IgGs that included flexible $(Gly_4Ser)_n$ linkers of various lengths (L16 – L24), which allowed us to directly compare the effects of incorporating different lengths of flexible versus structured proteins linkers.

Figure 4 displays hydrodynamic radius data for IgGs incorporating the designed linkers. We observed a consistent trend for the R_H values between glycosylated and non-glycosylated linkers. For proline-rich linkers derived from the hinge region of IgA1, incorporation of three potential N-linked glycosylation sites appeared to increase the length of the linker, possibly through stabilization of the folded state leading the linker to adopt a more extended conformation (Shental-Bechor and Levy, 2008). While the addition of a single potential N-linked glycosylation site did not seem to affect the length of proline-rich linkers, potential N-linked glycosylation in the Gly_2Ser region of a combination proline-rich and Gly_2Ser linker (L3) increased its length beyond the length of a larger glycosylated proline-rich linker (L1). These observations were consistent with the observation that potential N-linked glycosylation confers rigidity in the backbone of a flexible linker resulting in a more extended conformation (Liu *et al.*, 2000). Thus incorporating potential N-linked glycosylation sites within flexible linkers may be a method to increase linker rigidity.

A single β2m domain appeared to increase the hydrodynamic radius of the IgG more than the proline-rich linker derived from the hinge region of IgA1 (comparing L4 and L5). However, two tandem β2m domains separated by a non-glycosylated $(Gly_2Ser)_4$ peptide

provided only a moderate increase in the apparent size of the linker compared to a single β2m domain (comparing L5 and L8). This is consistent with the observation that an additional β2m domain, when joined with a $(Gly_2Ser)_3GAS$ peptide, did not significantly increase the hydrodynamic radius of the hinge-linked IgG compared to the wildtype IgG (comparing L6 and L7). These results demonstrated that coupling a flexible GlySer linker with the rigid β2m domain partially abrogated the separation between Fc and Fab regions provided by β2m alone. A similar observation was made for hinge constructs containing ZAG. A linker containing ZAG alone appeared to increase the hydrodynamic radius of the IgG more than the proline-rich linker (comparing L10 and L4). However, replacement of a proline-rich domain with ZAG that is flanked at both termini by a $(G_2S)_4$ peptide resulted in a decrease in hydrodynamic radius (comparing L10 and L11).

cTPR linker series

cTPR constructs were generated with 3, 6, 9 and 12 tandem repeats. All cTPR linkers were flanked by $(Gly_4Ser)_3$ sequences. By SEC, we saw a consistent decrease in elution volume as a function of the repeat length (Figure 3). These constructs also predictably increased the R_H of the linked IgG with increased number of tandem repeats (Figure 4). The hydrodynamic radius of the cTPR12 construct corresponded to approximately the size of L4, a proline-rich linker. These data indicated that the increase in separation between the Fab and Fc provided by the linker correlated predictably with the number of cTPRs, despite the presence of Gly_4Ser peptides flanking the N and C-termini.

$(Gly_4Ser)_n$ linker series

Unlike the SEC profiles for designed linker constructs, there were only small differences in elution volume for the IgGs including Gly_4Ser linkers (L17 – L24). These small differences did not necessarily correlate with molecular mass of the linkers as GS9, the IgG with the largest linker, eluted at approximately the same volume as wildtype IgG, which eluted after some of the constructs with shorter linkers (Figure 3).

Unlike proline-rich linkers and rigid linkers consisting of natural protein domains such as β2m, Gly_4Ser linkers that did not contain a potential N-linked glycosylation site did not detectably increase the hydrodynamic radius of the IgG, suggesting that these linkers did not provide increased separation between the Fab and Fc domains. An internal control of

$(Gly_4Ser)_9$ from two preparations showed a slight variation between the two measurements by about 0.1nm, a small difference compared with the others in Figure 4. These results are consistent with the observation that Gly_4Ser linkers did not provide significant separation between the joined domains in the context of other fusion proteins (Arai, Ueda, Kitayama, Kamiya and Nagamune, 2001).

Proper linker choice is important for the construction of multifunctional fusion proteins, and linkers come in many varieties with regard to effective length and rigidity. In this study, we used size exclusion chromatography and dynamic light scattering to characterize the behavior of designed linkers in the context of an IgG in order to determine whether these linkers could be used to increase the distance between the antigen binding fragments. We found that flexible Gly_4Ser linkers did not increase the R_H of fused reagents, and as a result did not provide adequate separation between the Fab and Fc domains even with up to 9 Gly_4Ser repeats. By contrast, the structured helical cTPR linkers, even when fused to more flexible Gly_4Ser domains, provided consistent increases in separation distances as a function of repeat number. Various other designed linkers including those containing naturally occurring proteins such as β2m and ZAG were able to increase separation distances, but in some cases such as the β2m and ZAG-containing linkers, fusion with a flexible Gly_4Ser linker tended to alleviate this effect. The systematic characterization of the lengths and rigidity properties of the structured protein linkers and a range of $(Gly_4Ser)_n$ linkers reported here should provide a new set of tools to the available linker repertoire.

Conflict of Interest

The authors declare that they have no competing financial interests.

Funding information

This work was supported by the Director's Pioneer Award [1DP1OD006961-01 to P.J.B.] and National Institutes of Health HIVRAD [P01 Al100148 to P.J.B.].

Acknowledgement

We thank the Caltech Protein Expression Center for producing engineered antibodies.

Figure legends

Fig. 1. *A.* Schematic of wildtype IgG (left) and IgG including designed linkers in its hinge region (middle). IgG domains are shown individually on the right. *B.* Ribbon diagrams for structured domains and their relative sizes. The consensus tetratricopeptide repeat contains 8 tandem repeats. The N- and C-terminal residues are shown as sticks, color-coded blue for the N-terminus and red for the C-terminus. Protein data bank codes: ZAG (1ZAG), β2m (1LDS), cTPR (2FO7), ubiquitin (1UBQ).

Fig. 2. SDS-PAGE analysis of IgG-structured linker proteins run under reducing (left) or non-reducing (right) conditions.

Fig. 3. Overlay of size exclusion chromatograms for IgGs containing flexible and structured protein linkers. Structured linkers (L1-L8) exhibited larger decreases in retention volume with respect to wildtype compared to Gly_4Ser linkers, which exhibited little to no decrease depending on the number of repeats. Structured cTPR linkers also exhibited consistent decreases in retention volume as a function of the number of repeats.

Fig. 4. Comparative analysis by DLS of the hydrodynamic radii (R_H) of designed linkers in the context of the b12 IgG.

References

Arai R., Ueda H., Kitayama A., Kamiya N. and Nagamune T. (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng,* **14**, 529-532. First published on.

Argos P. (1990) An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *Journal of molecular biology,* **211**, 943-958. First published on, doi: 10.1016/0022-2836(90)90085-Z.

Bai Y., Ann D.K. and Shen W.-C. (2005) Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. *Proceedings of the National Academy of Sciences of the United States of America,* **102**, 7292-7296. First published on, doi: 10.1073/pnas.0500062102.

Bonner A., Furtado P.B., Almogren A., Kerr M.A. and Perkins S.J. (2008) Implications of the near-planar solution structure of human myeloma dimeric IgA1 for mucosal immunity and IgA nephropathy. *Journal of immunology,* **180**, 1008-1018. First published on.

Chen X., Zaro J.L. and Shen W.C. (2013) Fusion protein linkers: property, design and functionality. *Advanced drug delivery reviews,* **65**, 1357-1369. First published on, doi: 10.1016/j.addr.2012.09.039.

D'Andrea L.D. and Regan L. (2003) TPR proteins: the versatile helix. *Trends in biochemical sciences,* **28**, 655-662. First published on, doi: 10.1016/j.tibs.2003.10.007.

Diskin R., Scheid J.F., Marcovecchio P.M., West A.P., Jr., Klein F., Gao H., Gnanapragasam P.N., Abadir A., Seaman M.S., Nussenzweig M.C. *et al.* (2011) Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science,* **334**, 1289-1293. First published on, doi: 10.1126/science.1213782.
Evers T.H., van Dongen E.M., Faesen A.C., Meijer E.W. and Merkx M. (2006) Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers. *Biochemistry,* **45**, 13183-13192. First published on, doi: 10.1021/bi061288t.
George R.A. and Heringa J. (2002) An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng,* **15**, 871-879. First published on.
Kajander T., Cortajarena A.L., Mochrie S. and Regan L. (2007) Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta crystallographica Section D, Biological crystallography,* **63**, 800-811. First published on, doi: 10.1107/S0907444907024353.
Klein J.S. and Bjorkman P.J. (2010) Few and far between: how HIV may be evading antibody avidity. *PLoS pathogens,* **6**, e1000908. First published on, doi: 10.1371/journal.ppat.1000908.
Klein J.S., Gnanapragasam P.N., Galimidi R.P., Foglesong C.P., West A.P., Jr. and Bjorkman P.J. (2009) Examination of the contributions of size and avidity to the neutralization mechanisms of the anti-HIV antibodies b12 and 4E10. *Proceedings of the National Academy of Sciences of the United States of America,* **106**, 7385-7390. First published on, doi: 10.1073/pnas.0811427106.
Lichty J.J., Malecki J.L., Agnew H.D., Michelson-Horowitz D.J. and Tan S. (2005) Comparison of affinity tags for protein purification. *Protein expression and purification,* **41**, 98-105. First published on, doi: 10.1016/j.pep.2005.01.019.
Liu H.L., Doleyres Y., Coutinho P.M., Ford C. and Reilly P.J. (2000) Replacement and deletion mutations in the catalytic domain and belt region of Aspergillus awamori glucoamylase to enhance thermostability. *Protein Eng,* **13**, 655-659. First published on.
Main E.R., Jackson S.E. and Regan L. (2003) The folding and design of repeat proteins: reaching a consensus. *Current opinion in structural biology,* **13**, 482-489. First published on.
Nobbmann U., Connah M., Fish B., Varley P., Gee C., Mulot S., Chen J., Zhou L., Lu Y., Shen F. *et al.* (2007) Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. *Biotechnology & genetic engineering reviews,* **24**, 117-128. First published on.
Pardridge W.M. (2010) Biopharmaceutical drug targeting to the brain. *Journal of drug targeting,* **18**, 157-167. First published on, doi: 10.3109/10611860903548354.
Roben P., Moore J.P., Thali M., Sodroski J., Barbas C.F., 3rd and Burton D.R. (1994) Recognition properties of a panel of human recombinant Fab fragments to

the CD4 binding site of gp120 that show differing abilities to neutralize human immunodeficiency virus type 1. *Journal of virology,* **68**, 4821-4828. First published on.
Sanchez L.M., Chirino A.J. and Bjorkman P. (1999) Crystal structure of human ZAG, a fat-depleting factor related to MHC molecules. *Science,* **283**, 1914-1919. First published on.
Scheufler C., Brinker A., Bourenkov G., Pegoraro S., Moroder L., Bartunik H., Hartl F.U. and Moarefi I. (2000) Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell,* **101**, 199-210. First published on, doi: 10.1016/S0092-8674(00)80830-2.
Schuler B., Lipman E.A., Steinbach P.J., Kumke M. and Eaton W.A. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proceedings of the National Academy of Sciences of the United States of America,* **102**, 2754-2759. First published on, doi: 10.1073/pnas.0408164102.
Shental-Bechor D. and Levy Y. (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America,* **105**, 8256-8261. First published on, doi: 10.1073/pnas.0801340105.
Trinh C.H., Smith D.P., Kalverda A.P., Phillips S.E. and Radford S.E. (2002) Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. *Proceedings of the National Academy of Sciences of the United States of America,* **99**, 9771-9776. First published on, doi: 10.1073/pnas.152337399.
Vijay-Kumar S., Bugg C.E. and Cook W.J. (1987) Structure of ubiquitin refined at 1.8 A resolution. *Journal of molecular biology,* **194**, 531-544. First published on.
Zhang J., Yun J., Shang Z., Zhang X. and Pan B. (2009) Design and optimization of a linker for fusion protein construction. *Progress in Natural Science,* **19**, 1197-1200. First published on, doi: http://dx.doi.org/10.1016/j.pnsc.2008.12.007.
Zhu P., Liu J., Bess J., Jr., Chertova E., Lifson J.D., Grise H., Ofek G.A., Taylor K.A. and Roux K.H. (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature,* **441**, 847-852. First published on, doi: 10.1038/nature04817.
Zou Y., Weis W.I. and Kobilka B.K. (2012) N-terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor. *PloS one,* **7**, e46039. First published on, doi: 10.1371/journal.pone.0046039.

Figures

(Figure 1)

Figure 1.

(Figure 2)

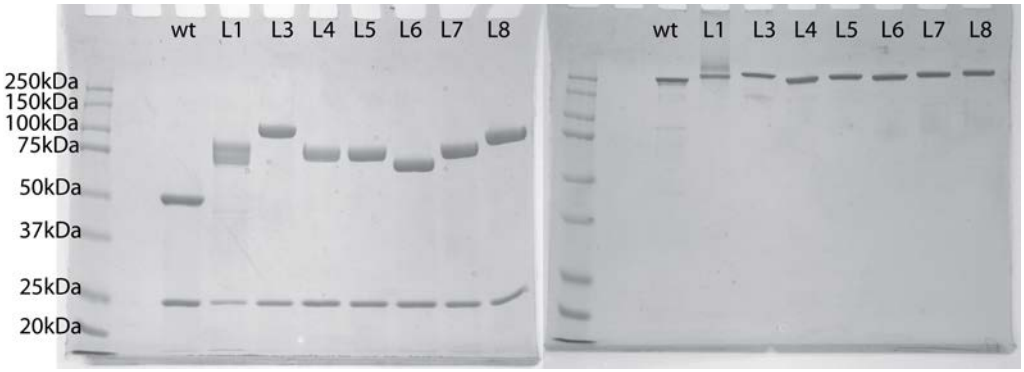


Figure 2.

(Figure 3**)**

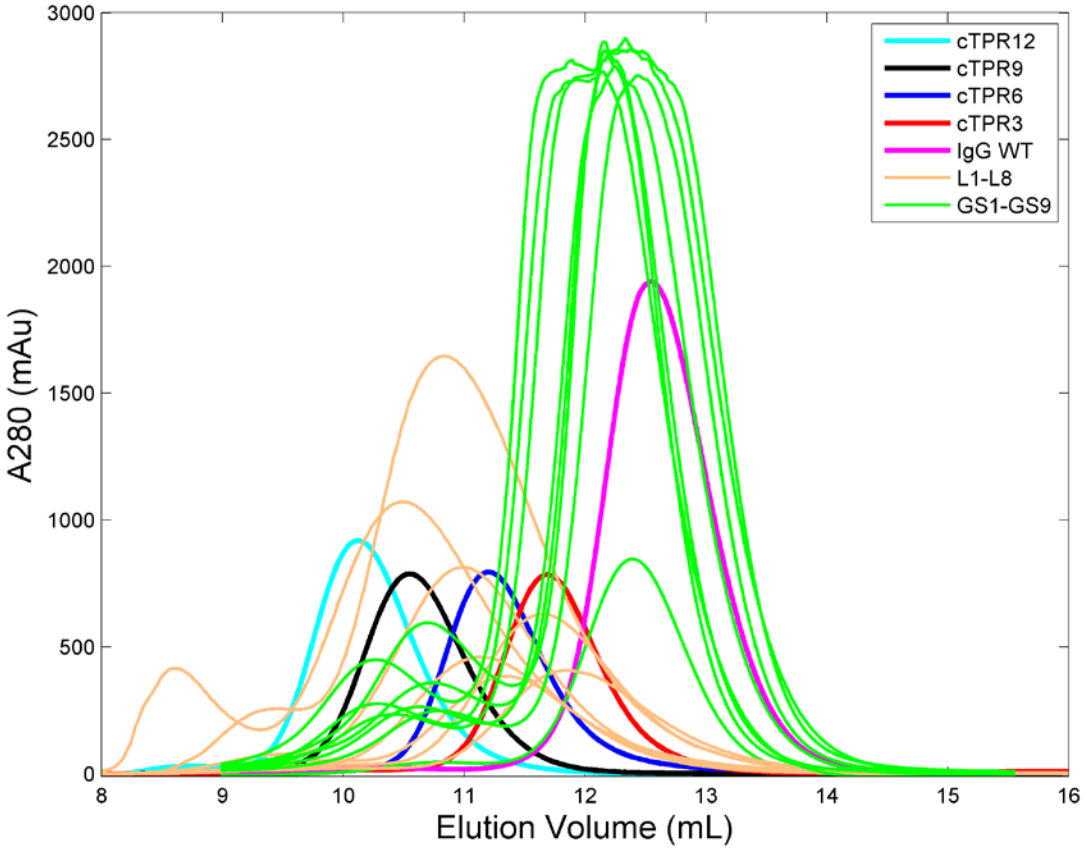


Figure 3.

(Figure 4**)**

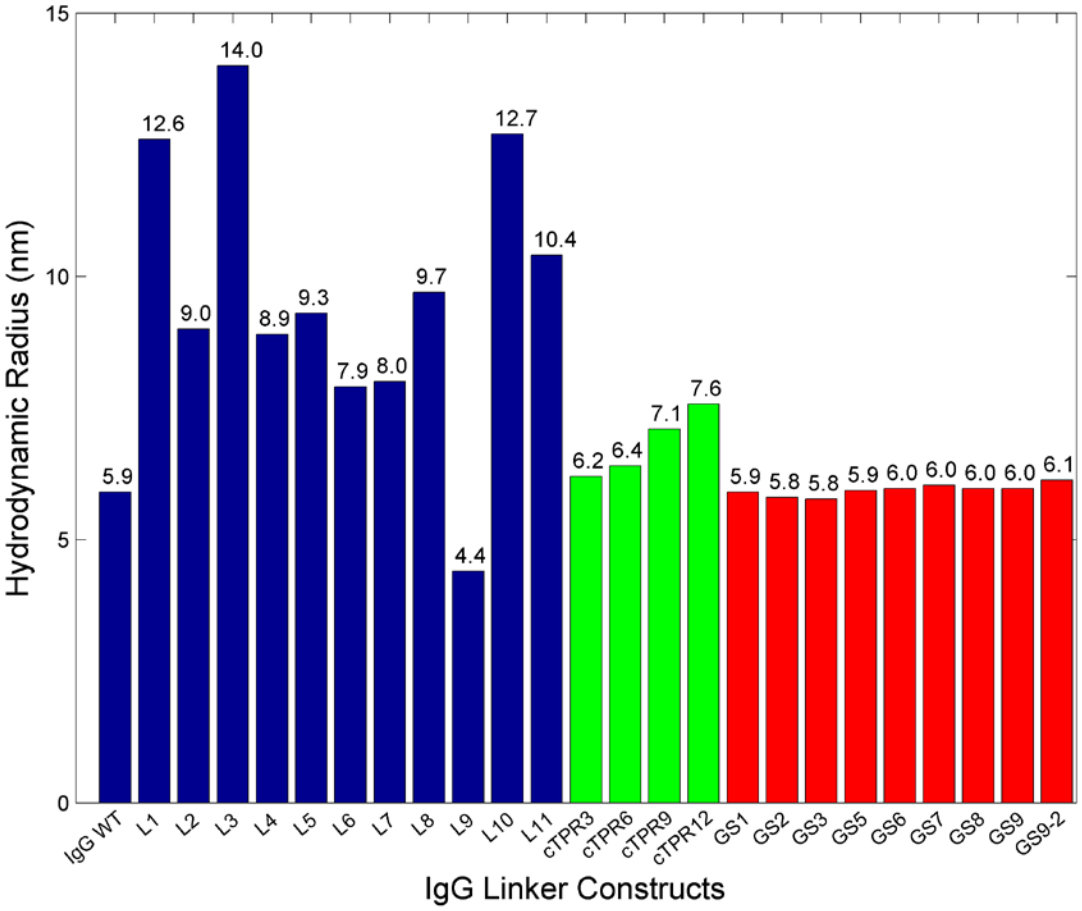


Figure 4.

Tables

(Table 1)

Table 1. Description of structured linker designs. $(Gly_4Ser)_n$=Gly-Gly-Gly-Gly-Ser sequence with n number of repeats; GlySer = (N-term: $AGS(GGS)_3$; Middle: $(GGS)_4$; C-term: $(GGS)_3GAS]_2S$); GlySer(Glyc)=Gly-Gly-Ser sequence with an embedded potential N-linked glycosylation site (Asn-Ser-Ser); polyPro=proline-rich hinge sequence from IgA1;

polyPro(Glyc)=proline-rich hinge sequence from IgA1 with an embedded potential N-linked glycosylation site (Asn-Ser-Ser); β2m=β-2-microglobulin; Ub=ubiquitin; ZAG=Zn-α2-glycoprotein; cTPRX=consensus tetratricopeptide repeat sequence with X number of repeats.

Linker	Name	Description
L1	GPcPcPc	GlySer-polyPro(Glyc)-polyPro(Glyc)-polyPro(Glyc)
L2	GPPcP	GlySer-polyPro-polyPro(Glyc)-polyPro
L3	GPGcP	GlySer-polyPro-GlySer(Glyc)-polyPro
L4	GPPP	GlySer-polyPro-polyPro-polyPro
L5	GPbP	GlySer-polyPro-β2m-polyPro
L6	GPbG	GlySer-polyPro-β2m-GlySer
L7	PbGbG	polyPro-β2m-GlySer-β2m-GlySer
L8	GPbGbP	GlySer-polyPro-β2m-GlySer-β2m-polyPro
L9	GPUG	GlySer-polyPro-Ub-GlySer
L10	GPZP	GlySer-polyPro-ZAG-polyPro
L11	GGZGZP	GlySer-GlySer-ZAG-GlySer-ZAG-polyPro
L12	GcGcP	GlySer(Glyc)- GlySer(Glyc)-polyPro
L13	cTPR3	$(G_4S)_3$-cTPR3-$(G_4S)_3$
L14	cTPR6	$(G_4S)_3$-cTPR6-$(G_4S)_3$
L15	cTPR9	$(G_4S)_3$-cTPR9-$(G_4S)_3$
L16	cTPR12	$(G_4S)_3$-cTPR12-$(G_4S)_3$
L17	GS1	$(G_4S)_1$
L18	GS2	$(G_4S)_2$
L19	GS3	$(G_4S)_3$
L20	GS5	$(G_4S)_5$
L21	GS6	$(G_4S)_6$
L22	GS7	$(G_4S)_7$
L23	GS8	$(G_4S)_8$
L24	GS9	$(G_4S)_9$

(Table 2)

Table 2. Complete sequences of designed linkers. Linkers L1-L21 were inserted into the hinge region of b12 IgG between residues His235 and Thr236. Linkers L22-L24 were inserted into the same hinge between residues Cys231 and Asp232.

Linker	Name	Complete Sequence
L1	GPcPcPc	AGSGGSGGSGGSPVPSTPPTNSSSTPPTPSPSPVPSTPPTNSSSTPPTPSPSPVPSTPPTNSSSTPPTPSPSAS
L2	GPPcP	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSPVPSTPPTNSSSTPPTPSPSPVPSTPPTPSPSTPPTPSPSAS
L3	GPGcP	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSGGSGNSSGSGGSPVPSTPPTPSPSTPPTPSPSAS
L4	GPPP	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSPVPSTPPTPSPSTPPTPSPSPVPSTPPTPSPSTPPTPSPSAS
L5	GPbP	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDPVPSTPPTPSPSTPPTPSPSAS
L6	GPbG	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDGGSGGSGGSGGSAS
L7	PbGbG	AGPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDGGSGGSGGSGGSGGSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDGGSGGSGGSGGSGAS
L8	GPbGbP	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDGGSGGSGGSGGSGGSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDPVPSTPPTPSPSTPPTPSPSAS
L9	GPUG	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGGGSGGSGGSGGSAS

L10	GPZP	AGSGGSGGSGGSPVPSTPPTPSPSTPPTPSPSDGRYSLTYIY TGLSKHVEDVPAFQALGSLNDLQFFRYNSKDRKSQPMGLWR QVEGMEDWKQDSQLQKAREDIFMETLKDIVEYYNDSNGSHVL QGRFGCEIENNRSSGAFWKYYYDGKDYIEFNKEIPAWVPFDP AAQITKQKWEAEPVYVQRAKAYLEEECPATLRKYLKYSKNILD RQDPPSVVVTSHQAPGEKKKLKCLAYDFYPGKIDVHWTRAGE VQEPELRGDVLHNGNGTYQSWVVVAVPPQDTAPYSCHVQHS SLAQPLVVPWEASPVPSTPPTPSPSTPPTPSAS
L11	GGZGZP	AGSGGSGGSGGSGGSGGSGGSGGSDGRYSLTYIYTGLSKHV EDVPAFQALGSLNDLQFFRYNSKDRKSQPMGLWRQVEGMED WKQDSQLQKAREDIFMETLKDIVEYYNDSNGSHVLQGRFGCE IENNRSSGAFWKYYYDGKDYIEFNKEIPAWVPFDPAAQITKQK WEAEPVYVQRAKAYLEEECPATLRKYLKYSKNILDRQDPPSVV VTSHQAPGEKKKLKCLAYDFYPGKIDVHWTRAGEVQEPELRG DVLHNGNGTYQSWVVVAVPPQDTAPYSCHVQHSSLAQPLVV PWEASGGSGGSGGSGGSDGRYSLTYIYTGLSKHVEDVPAFQ ALGSLNDLQFFRYNSKDRKSQPMGLWRQVEGMEDWKQDSQ LQKAREDIFMETLKDIVEYYNDSNGSHVLQGRFGCEIENNRSS GAFWKYYYDGKDYIEFNKEIPAWVPFDPAAQITKQKWEAEPV YVQRAKAYLEEECPATLRKYLKYSKNILDRQDPPSVVVTSHQA PGEKKKLKCLAYDFYPGKIDVHWTRAGEVQEPELRGDVLHNG NGTYQSWVVVAVPPQDTAPYSCHVQHSSLAQPLVVPWEASP VPSTPPTPSPSTPPTPSPSAS
L12	GcGcP	AGSGNSSGSGGSGGSGNSSGSGGSPVPSTPPTPSPSTPPTP SPSAS
L13	cTPR3	KLSGGGGSGGGGSGGGGSAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQR SAGGGGSGGGGSGGGGAS
L14	cTPR6	KLSGGGGSGGGGSGGGGSAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQAE AWKNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGD YQKAIEDYQKALELDPNNRSRSAGGGGSGGGGSGGGGAS
L15	cTPR9	KLSGGGGSGGGGSGGGGSAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQAE AWKNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGD YQKAIEDYQKALELDPNNRSAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL

		DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQR SAGGGGSGGGGSGGGGAS
L16	cTPR12	KLSGGGGSGGGGSGGGGSAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQAE AWKNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGD YQKAIEDYQKALELDPNNRSAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQAE AWKNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGD YQKAIEDYQKALELDPNNRSAGGGGSGGGGSGGGGAS
L17	GS1	GGGGSAS
L18	GS2	GGGGSGGGGSAS
L19	GS3	GGGGSGGGGSGGGGSAS
L20	GS5	GGGGSGGGGSGGGGSGGGGSGGGGSAS
L21	GS6	GGGGSGGGGSGGGGSGGGGSGGGGSGGGGSAS
L22	GS7	AGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSAS
L23	GS8	AGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGG GGSAS
L24	GS9	AGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGG GGSGGGGSAS