Structure-guided SCHEMA recombination of VRC01-class antibodies for reduced polyreactivity

Thesis by

Sarah Knox Gillespie

In Partial Fulfillment of the Requirements for
the Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
(Defended October 4, 2024)

ACKNOWLEDGEMENTS

I am tremendously grateful for all of the people who have supported me throughout my journey at Caltech.

First and foremost, I would like to profoundly thank my advisor, Professor Steve Mayo, for his unwavering support and guidance throughout my training. I deeply appreciate his encouragement, patience, and immense knowledge. Words cannot express the depth of my gratitude for everything Steve has done for me over the years. I will be forever grateful to him.

I am also extremely thankful to my thesis committee members, Professors Pamela Bjorkman, Bruce Hay, and Lea Goentoro, for their remarkable support. I feel so exceptionally fortunate to have had many enriching scientific discussions with them. Thank you for your continuous encouragement and your confidence in my abilities as a scientist. I have learned so much during our time together.

I would like to express my deepest gratitude to Jennifer Keefe for her encouragement, guidance, advice, and suggestions. Your scientific insights are invaluable. Additionally, this project would not have been possible without you. More than anything, I am deeply grateful for your friendship and support.

I could not have undertaken this journey without Jost Vielmetter. Learning from you has been a highlight of my training. I have learned so much about assay development from you. I am so immensely grateful for your mentorship.

Moreover, I would like to thank Tim Wannier for teaching me about red fluorescent proteins. I learned so much about protein engineering from Tim. I feel so fortunate to have had the opportunity to collaborate with him.

Thank you to all current and former Mayo lab members, including, but not limited to Jingzhou Wang, Aiden Aceves, Shan Huang, Lucas Schaus and Nick Friesenhahn. I feel so fortunate to have been inspired by so many brilliant people. Also, I would like to especially thank Monica Breckow, our lab manager, for always being eager to lend a hand in any way possible.

I would also like to thank all of the Bjorkman lab members who have been so generous with their time. I would especially like to thank Anthony West for an insightful discussion about VRC-01 class antibodies that was critical for this project.

Furthermore, many thanks to Annie Lam in the Protein Expression Center. Your assistance with troubleshooting experiments has been indispensable over the years.

Additionally, I would like to thank Alistair M. at WordVice and Doreen at Scribbr for their helpful assistance with professional proofreading of this thesis.

I am so grateful to have met so many remarkable people and to have made lifelong friends during my time at Caltech. I would especially like to thank Shannon Esswein and Suzy Beeler for their friendship and support along the way.

And finally, I would like to thank my incredible family for their unconditional love and support. Thank you for everything. This would not have been possible without you.

ABSTRACT

The therapeutic administration of monoclonal antibodies (mAbs) has revolutionized treatment options for many diseases over the last decade. Recent findings from clinical trials have demonstrated that broadly neutralizing antibodies (bNAbs) could have a potential role in the future treatment and prevention of HIV-1. There is a group of broad and potent bNAbs that target the CD4-binding site (CD4bs) on the envelope glycoprotein gp120. These VRC01-class antibodies are notable for both their breadth and potency. The Bjorkman lab has designed a remarkably broad and potent bNAb, 45-46m2, that unfortunately cannot currently be used clinically due to its increased polyreactivity and short *in vivo* half-life.

In this study we designed a SCHEMA-guided recombination library composed of sequence fragments of the VRC01-class bNAbs 45-46m2 and 3BNC117, aiming to create a bNAb that binds strongly to gp120 but is not polyreactive. We endeavored to maintain the strong binding of 45-46m2 while gaining the low polyreactivity of 3BNC117.

Our analysis of this family shuffled library of chimeric antibodies revealed the sequence elements that led to strong binding to gp120 for the chimeras in our library. We also identify three key framework regions that can be modified to significantly reduce polyreactivity. Furthermore, we report three novel chimeras from the family shuffled library that bind as strongly to gp120 as 45-46m2 but are significantly reduced in polyreactivity.

PUBLISHED CONTENT AND CONTRIBUTIONS

Chapter 4 :

Keeffe, J.R., Gnanapragasam, P.N., Gillespie, S.K., Yong, J., Bjorkman, P.J., Mayo, S.L. (2011). Designed oligomers of cyanovirin-N show enhanced HIV neutralization. PNAS; 108(34): 14079-14084. doi: 10.1073/pnas.1108777108.

S.K.G. conducted the experiments and analyzed the data showing that CVN_2L0 is domain swapped in solution. S.K.G. created the original figure 4 (showing that the anti-HIV activity of CVN_2L0 correlates with the number of functional binding sites), as well as supplemental figure S3 (showing that the partial binding site knockout mutants do not significantly affect the neutralization potency). S.K.G. wrote a draft to describe the experiments and results that led to figure 4 and S3. This draft was later rewritten and improved by J.R.K. J.R.K. also improved S.K.G.'s original figures (changed the original circles to squares) for publication.

Chapter 5:

Wannier, T.M., Gillespie S.K., Hutchins N., McIsaac R.S., Wu S.Y., Shen Y., Campbell R.E., Brown K.S., Mayo S.L. (2018). Monomerization of far-red fluorescent proteins. PNAS; 115(48): E11294-E11301. doi: 10.1073/pnas.1807449115.

S.K.G. assisted with characterizing the important red fluorescent protein variants in the library. S.K.G. assisted with collecting absorbance scans, fluorescence excitation scans, fluorescence emission scans, measurements of quantum yield, measurements of extinction coefficients, and thermal stability measurements. S.K.G. also assisted with oligomeric determination by analytical ultracentrifugation.

Manuscripts in preparation (working titles):

Chapter 2 :

Gillespie, S.K., Sievers, S.A., West, A.P., Keeffe, J.R., Bjorkman, P.J., and Mayo, S.L. Structure-guided SCHEMA recombination of VRC01-class antibodies for reduced polyreactivity. [Manuscript in preparation].

S.A.S., A.P.W., and P.J.B. conceived of the project. J.R.K., P.J.B, and S.L.M. advised on experiments and analysis, and provided helpful discussion. S.A.S. used SCHEMA to design the family shuffled library in silico. S.A.S. used Gibson cloning to create chimeric V_H gene DNA from gene block DNA. S.A.S. also used Gibson cloning to insert the 128 chimeric V_H genes into the vector backbone.

S.K.G. conducted all other experiments, analyzed the data, prepared the figures, and wrote the paper.

Chapter 3 :

Gillespie, S.K., Vielmetter, J.G., Keeffe, J.R., and Mayo, S.L. Using high-throughput surface plasmon resonance (SPR-32) to determine the concentration of monoclonal antibodies in crude transfected cell supernatants. [Manuscript in preparation].

J.G.V., J.R.K., and S.K.G. designed the SPR-32 experiments. J.R.K., J.G.V., and S.L.M. advised on experiments and analysis, and provided helpful discussion.

S.K.G. conducted the experiments, analyzed the data, prepared the figures, and wrote the paper.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

It has been over 40 years since the discovery of HIV-1, the virus that causes AIDS. However, HIV/AIDS continues to be a major global public health challenge. By the end of 2023, there were an estimated 39.9 million people living with HIV-1 (UNAIDS, 2024). In 2023, an estimated 1.3 million people became newly infected with HIV-1 (UNAIDS, 2024).

The development of antiretroviral therapy has been a major success in medical research. However, treatment with antiretroviral therapy is lifelong and requires constant access to newer drug regimens (Zaidi, 2024). One reason for this is that antiretroviral therapy does not eliminate latent reservoirs of HIV-1 from the host cells, and the virus can become resistant to the medications (Sun et al., 2024). Another reason is that some patients develop long-term side effects from lifetime usage of these drugs (Chawla et al., 2018). While patients in developed countries have various treatment options, patients in the developing world have access to fewer treatment options (Zaidi, 2024). New methods to treat and prevent infection are urgently needed.

One possible solution might be the use of broadly neutralizing antibodies (bNAbs) for passive delivery to treat and/or prevent HIV-1 infection (Halper-Stromberg et al., 2016). Recent findings from clinical trials have demonstrated the potential role of broadly neutralizing antibodies (bNAbs) in the treatment and prevention HIV-1 (Thavarajah et al., 2024).

There is a group of broad and potent bNAbs that target the CD4-binding site on the envelope glycoprotein gp120. These VRC01-class antibodies are notable for both their breadth and potency, and they share a set of signature features (West et al., 2012). Diskin, Bjorkman and colleagues used structure-based design (Diskin et al., 2011; Diskin et al., 2013) to improve NIH45-46 (Scheid et al., 2011), a more potent clonal variant of VRC01 (Wu et al., 2010; Zhou et al., 2010). This work resulted in the design of NIH45-46^{G54W}, 45-46m2, and 45-

46m7, three remarkably broad and potent bNAbs (Diskin et al., 2011; Diskin et al., 2013). Also, the addition of 45-46m2 and 45-46m7 to a mixture of other bNAbs was shown to increase the efficacy of antibody-based therapeutics by restricting potential HIV-1 escape pathways and/or selecting mutant virions with reduced fitness (Diskin et al., 2013).

Unfortunately, NIH45-46^{G54W}, 45-46m2, and 45-46m7 cannot currently be used clinically due to their increased polyreactivity and short *in vivo* half-lives. Polyreactivity is correlated with short *in vivo* half-lives because antibodies that bind off-target are eliminated from the circulation by mechanisms involving antigen-mediated clearance (Mager et al., 2006; Wang et al., 2008). There are many unique hypotheses about the biophysical determinants of polyreactivity in antibodies. Some suggestions include: net hydrophobicity (Lecerf et al., 2019), net charge (Rabia et al., 2018), increased prevalence of arginine and tyrosine residues (Kelly et al., 2018; Birtalan et al., 2008), enrichment in glutamine residues (Lecerf et al., 2019), CDR3 length (Prigent et al., 2018), CDR flexibility (Prigent et al., 2018), a shift toward neutrality of the binding interface (Boughter et al., 2020), loop crosstalk in the heavy chain (Boughter et al., 2020), and high isoelectric points (Kelly et al., 2018; Birtalan et al., 2008; Lecerf et al., 2019; Rabia et al., 2018). However, the relative importance of these features is still inconclusive (Harvey et al., 2022; Birtalan et al., 2008). Harvey et al. (2022) recently developed a method for predicting the polyreactivity of a given nanobody sequence based on machine learning models (Harvey et al., 2022). This method also suggests mutations that can be implemented to reduce polyreactivity for the given nanobody (Harvey et al., 2022). However, this method was originally intended only for use in nanobodies (Harvey et al., 2022). A more comprehensive understanding of the sequence features of polyreactive antibodies is still needed. Currently, there is no method that can identify the sequence features of a polyreactive antibody so that the problematic sequences could be modified.

In order to create new variants of 45-46m2 with lower polyreactivity, Stuart Sievers, a former postdoctoral fellow in the Bjorkman lab, and Anthony West, a senior research scientist in the Bjorkman lab, proposed the use of directed evolution by DNA shuffling of a family of related

genes. Family shuffling permits a greater exploration of sequence space than random mutagenesis and makes changes already demonstrated to be compatible with the fold and function of the protein (Crameri et al., 1998). What makes family shuffling possible in this case is that VRC01-class antibodies all bind to gp120 with the same basic geometry despite large sequence diversity. In addition, this class of antibodies includes naturally occurring antibodies with low polyreactivity (Prigent et al., 2018).

In Chapter 2, we describe the construction of a SCHEMA-guided recombination library composed of sequence fragments of the VRC01-class bNAbs 45-46m2 and 3BNC117. Our goal was to create a bNAb that maintained the strong binding to gp120 of 45-46m2, while gaining the reduced polyreactivity of 3BNC117. In addition, we hoped to gain knowledge about the sequence features that result in decreased polyreactivity so that this knowledge could be applied to newly designed antibodies that are not amenable to family shuffling. We used SCHEMA, a computational approach, to accomplish this goal. SCHEMA uses structure-guided recombination to produce large families of protein chimeras that have high sequence and functional diversity (Otey et al., 2006). Chimeras are made by recombining a set of homologous parent proteins at crossover locations chosen to minimize structural disruption (Endelman, et al., 2004). SCHEMA uses the 3D structures of the parent proteins to identify pairs of amino acids that are interacting. Each chimera is characterized by the value E, the number of residue-residue contacts that are broken upon recombination. SCHEMA uses the recombination as a shortest path problem (RASPP) algorithm to minimize the number of contacts that are disrupted when blocks are swapped to generate new sequences (Endelman, et al., 2004). RASPP minimizes E, which improves the chances a chimera will be folded and functional (Meyer et al., 2006).

A key benefit of SCHEMA is the ability to use the “sample, model, and predict” approach developed by Arnold and colleagues (Li et al., 2007; Heinzelman et al., 2009; Romero et al., 2012; Heinzelman et al., 2013). A small sample set of chimeras is characterized, and then this data is used to construct models that accurately predict the properties of every family member in the library (Heinzelman et al., 2013). This approach has been used successfully

to construct libraries of novel chimeric proteins, including beta-lactamases, cytochrome p450s, arginases, fungal cellulases, channelrhodopsins, adeno-associated virus (AAV) capsids, and fungal laccases (Meyer et al., 2003; Otey et al., 2006; Li et al., 2007; Romero et al., 2012; Heinzelman et al., 2009; Bedbrook et al., 2017; Ojala et al., 2018; Mateljak et al., 2019).

In Chapter 2, we pinpoint the sequence elements that led to strong binding to gp120 for the chimeras in our family shuffled library. We also identify three key framework regions that can be modified to significantly reduce polyreactivity. Furthermore, we report three novel chimeras from the family shuffled library that bind as strongly to gp120 as 45-46m2, but are significantly reduced in polyreactivity.

In Chapter 3, we describe a new method to measure the concentration of monoclonal antibodies in crude transfected cell supernatants without the need for antibody purification. We used this method to measure the concentration of the engineered chimeric antibodies in our family shuffled library.

The traditional workflow for generating monoclonal antibodies for use in downstream assays is not practical for generating large libraries of hundreds of engineered antibody variants. A new method for measuring antibody concentration that could obviate the need for affinity-purification would be useful. Additionally, a method that would not require large volumes would be significantly advantageous, as our workflow for generating engineered family shuffled antibody variants would include small-scale transfections performed on a liquid handling robot. Furthermore, a method that could quantify only the folded and functional antibodies in a complex mixture of both active and inactive molecules would be useful for many downstream assays involving monoclonal antibodies.

We sought to test whether high-throughput surface plasmon resonance (SPR-32) could be used to measure the concentration of antibodies in crude transfected cell supernatant, a complex mixture of many proteins. SPR is a label-free optical technique that measures molecular interactions in real time (Jason-Moller et al., 2006).

The concept for this new method is as follows. First, we created a sensor chip with immobilized protein A on 16 of the 32 individual detection spots. The remaining 16 spots were reference spots. These spots were used for reference subtraction to compensate for bulk refractive index differences between the buffer and analyte samples. Next, we created a serial dilution of an affinity-purified and SEC-purified IgG of known concentration. We then performed a run on the SPR-32, where the analyte (IgG) was flowed over the 32 detection spots.

The sensorgrams for the known standard samples were used to create standard reference curves (calibration curves) for all 16 spots with immobilized protein A. We selected a report point, and then the resonance units at the report point for each known sample of the standard curve were plotted against the known concentration. This was done for each spot on the chip that had immobilized protein A. We then determined the equation that best fit the standard reference curves. Using this equation, the concentration of an unknown sample could be inferred from the resonance units at the report point for that unknown sample.

This method allowed us to accurately measure the concentration of expressed antibodies in crude transfected cell supernatants in high-throughput. These measurements were critical for the study described in Chapter 2 because it allowed us to normalize the engineered chimeric antibodies in our library to a standard concentration in preparation for the downstream assays.

In the final two chapters of this thesis, we explore the use of other protein engineering methods applied to unique problems. In Chapter 4, we describe the engineering of obligate cyanovirin-N (CV-N) oligomers to determine whether increasing the number of binding sites would have an effect on viral neutralization (Keeffe et al., 2011). A tandem repeat of two CV-N molecules (CVN2) increased HIV-1 neutralization activity by up to 18-fold compared to wild-type CV-N (Keeffe et al., 2011). In addition, the CVN2 variants showed extensive cross-clade reactivity (Keeffe et al., 2011). My involvement in this research was to create a series of binding site mutants of CVN_2L0 in order to show that CVN_2L0 is domain swapped

in solution. We showed that the anti-HIV activity of CVN_2L0 correlates with the number of functional binding sites, and that the partial binding site knockout mutants do not significantly affect the neutralization potency (Keeffe et al., 2011).

In Chapter 5, we explore protein variant libraries targeted at monomerizing far-red red fluorescent protein (RFP) variants and describe a method to monomerize RFPs of interest (Wannier et al., 2018). This method preserved the fluorescence of the molecule throughout its monomerization, which is in contrast to older break-fix methods (Wannier et al., 2018). This new method also resulted in selective enrichment of bright, far-red monomeric variants (Wannier et al., 2018). In addition, we identified four bright monomeric RFPs, which were among the most red-shifted of any monomeric fluorescent proteins (Wannier et al., 2018). My involvement with this research was to assist in the characterization of the important red fluorescent protein variants in our library. I assisted with collecting absorbance scans, fluorescence excitation scans, fluorescence emission scans, measurements of quantum yield, measurements of extinction coefficients, and thermal stability measurements. I also assisted with the oligomeric determination of the important red fluorescent proteins in our library by analytical ultracentrifugation.

Chapter 1 References

Bedbrook, C. N., Rice, A. J., Yang, K. K., Ding, X., Chen, S., LeProust, E. M., Gradinaru, V., & Arnold, F. H. (2017). Structure-guided SCHEMA recombination generates diverse chimeric channelrhodopsins. *Proceedings of the National Academy of Sciences*, *114*(13), E2624–E2633. https://doi.org/10.1073/pnas.1700269114

Birtalan, S., Zhang, Y., Fellouse, F. A., Shao, L., Schaefer, G., & Sidhu, S. S. (2008). The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *Journal of Molecular Biology*, *377*(5), 1518–1528. https://doi.org/10.1016/j.jmb.2008.01.093

Boughter, C. T., Borowska, M. T., Guthmiller, J. J., Bendelac, A., Wilson, P. C., Roux, B., & Adams, E. J. (2020). Biochemical patterns of antibody polyreactivity revealed through a bioinformatics-based analysis of CDR loops. *ELife*, *9*, e61393. https://doi.org/10.7554/eLife.61393

Chawla, A., Wang, C., Patton, C., Murray, M., Punekar, Y., De Ruiter, A., & Steinhart, C. (2018). A review of Long-Term toxicity of Antiretroviral treatment regimens and Implications for an aging population. *Infectious Diseases and Therapy*, *7*(2), 183–195. https://doi.org/10.1007/s40121-018-0201-6

Crameri, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391, 288-291.

Diskin, R., Scheid, J. F., Marcovecchio, P. M., West, A. P., Klein, F., Gao, H., Gnanapragasam, P. N. P., Abadir, A., Seaman, M. S., Nussenzweig, M. C., & Bjorkman, P. J. (2011). Increasing the Potency and Breadth of an HIV Antibody by Using Structure-Based Rational Design. *Science*, *334*(6060), 1289–1293. https://doi.org/10.1126/science.1213782

Diskin, R., Klein, F., Horwitz, J. A., Halper-Stromberg, A., Sather, D. N., Marcovecchio, P. M., Lee, T., West, A. P., Gao, H., Seaman, M. S., Stamatatos, L., Nussenzweig, M. C., & Bjorkman, P. J. (2013). Restricting HIV-1 pathways for escape using rationally designed anti–HIV-1 antibodies. *Journal of Experimental Medicine*, *210*(6), 1235–1249. https://doi.org/10.1084/jem.20130221

Endelman, J. B., Silberg, J. J., Wang, Z.-G., & Arnold, F. H. (2004). Site-directed protein

recombination as a shortest-path problem. Protein Engineering, Design & Selection, 17, 589–594.

Harvey, E. P., Shin, J., Skiba, M. A., Nemeth, G. R., Hurley, J. D., Wellner, A., Shaw, A. Y., Miranda, V. G., Min, J. K., Liu, C. C., Marks, D. S., & Kruse, A. C. (2022). An in silico method to assess antibody fragment polyreactivity. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-35276-4

Halper-Stromberg, A., & Nussenzweig, M. C. (2016). Towards HIV-1 remission: potential roles for broadly neutralizing antibodies. *Journal of Clinical Investigation*, *126*(2), 415–423. https://doi.org/10.1172/jci80561

Heinzelman, P., Romero, P. A., & Arnold, F. H. (2013). Efficient Sampling of SCHEMA Chimera Families to Identify Useful Sequence Elements. In *Methods in Enzymology* (Vol. 523, pp. 351–368). Elsevier. https://doi.org/10.1016/B978-0-12-394292-0.00016-3

Heinzelman, P., Snow, C. D., Smith, M. A., Yu, X., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., & Arnold, F. H. (2009). SCHEMA Recombination of a Fungal Cellulase Uncovers a Single Mutation That Contributes Markedly to Stability. *Journal of Biological Chemistry*, *284*(39), 26229–26233. https://doi.org/10.1074/jbc.C109.034058

Jason-Moller, L., Murphy, M., & Bruno, J. (2006). Overview of Biacore systems and their applications. *Current Protocols in Protein Science*, *45*(1). https://doi.org/10.1002/0471140864.ps1913s45

Keeffe, J. R., Gnanapragasam, P. N. P., Gillespie, S. K., Yong, J., Bjorkman, P. J., & Mayo, S. L. (2011). Designed oligomers of cyanovirin-N show enhanced HIV neutralization. *Proceedings of the National Academy of Sciences*, *108*(34), 14079–14084. https://doi.org/10.1073/pnas.1108777108

Kelly RL, Le D, Zhao J, Wittrup KD. 2018. Reduction of nonspecificity motifs in synthetic antibody libraries. Journal of Molecular Biology 430:119–130.
https://doi.org/10.1016/j.jmb.2017.11.008

Lecerf, M., Kanyavuz, A., Lacroix-Desmazes, S., & Dimitrov, J. D. (2019). Sequence features of variable region determining physicochemical properties and polyreactivity of therapeutic

antibodies. *Molecular Immunology*, *112*, 338–346. https://doi.org/10.1016/j.molimm.2019.06.012

Li, Y., Drummond, D. A., Sawayama, A. M., Snow, C. D., Bloom, J. D., & Arnold, F. H. (2007). A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nature Biotechnology*, *25*(9), 1051–1056. https://doi.org/10.1038/nbt1333

Mager, D. E. (2006). Target-mediated drug disposition and dynamics. *Biochemical Pharmacology*, *72*(1), 1–10. https://doi.org/10.1016/j.bcp.2005.12.041

Mateljak, I., Rice, A., Yang, K., Tron, T., & Alcalde, M. (2019). The generation of thermostable fungal laccase chimeras by SCHEMA-RASPP Structure-Guided Recombination in vivo. *ACS Synthetic Biology*, *8*(4), 833–843. https://doi.org/10.1021/acssynbio.8b00509

Meyer, M. M., Silberg, J. J., Voigt, C. A., Endelman, J. B., Mayo, S. L., Wang, Z.-G., & Arnold, F. H. (2003). Library analysis of SCHEMA-guided protein recombination. *Protein Science*, *12*(8), 1686–1693. https://doi.org/10.1110/ps.0306603

Meyer, M. M., Hochrein, L., & Arnold, F. H. (2006). Structure-guided SCHEMA recombination of distantly related beta-lactamases. Protein Engineering, Design & Selection, 19, 563–570.

Ojala, D. S., Sun, S., Santiago-Ortiz, J. L., Shapiro, M. G., Romero, P. A., & Schaffer, D. V. (2018). In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. *Molecular Therapy*, *26*(1), 304–319. https://doi.org/10.1016/j.ymthe.2017.09.006

Otey, C. R., Landwehr, M., Endelman, J. B., Hiraga, K., Bloom, J. D., & Arnold, F. H. (2006). Structure-Guided Recombination Creates an Artificial Family of Cytochromes P450. *PLoS Biology*, *4*(5), e112. https://doi.org/10.1371/journal.pbio.0040112

Prigent, J., Jarossay, A., Planchais, C., Eden, C., Dufloo, J., Kök, A., Lorin, V., Vratskikh, O., Couderc, T., Bruel, T., Schwartz, O., Seaman, M. S., Ohlenschläger, O., Dimitrov, J. D., & Mouquet, H. (2018). Conformational Plasticity in Broadly Neutralizing HIV-1 Antibodies Triggers Polyreactivity. *Cell Reports*, *23*(9), 2568–2581. https://doi.org/10.1016/j.celrep.2018.04.101

Rabia, L. A., Zhang, Y., Ludwig, S. D., Julian, M. C., & Tessier, P. M. (2018). Net charge of antibody complementarity-determining regions is a key predictor of specificity. *Protein*

Engineering, Design and Selection, *31*(11), 409–418. https://doi.org/10.1093/protein/gzz002

Romero, P. A., Stone, E., Lamb, C., Chantranupong, L., Krause, A., Miklos, A. E., Hughes, R. A., Fechtel, B., Ellington, A. D., Arnold, F. H., & Georgiou, G. (2012). SCHEMA-Designed Variants of Human Arginase I and II Reveal Sequence Elements Important to Stability and Catalysis. *ACS Synthetic Biology*, *1*(6), 221–228. https://doi.org/10.1021/sb300014t

Scheid, J. F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T. Y. K., Pietzsch, J., Fenyo, D., Abadir, A., Velinzon, K., Hurley, A., Myung, S., Boulad, F., Poignard, P., Burton, D. R., Pereyra, F., Ho, D. D., Walker, B. D., Seaman, M. S., … Nussenzweig, M. C. (2011). Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. *Science*, *333*(6049), 1633–1637. https://doi.org/10.1126/science.1207227

Sun, Y., & Wang, L. (2024). Development of Anti-HIV therapeutics: From conventional drug discovery to Cutting-Edge Technology. *Pharmaceuticals*, *17*(7), 887. https://doi.org/10.3390/ph17070887

Thavarajah, J. J., Hønge, B. L., & Wejse, C. M. (2024). The use of broadly neutralizing antibodies (BNABs) in HIV-1 treatment and prevention. *Viruses*, *16*(6), 911. https://doi.org/10.3390/v16060911

UNAIDS. (2024, July 22). *Fact sheet - Latest global and regional statistics on the status of the AIDS epidemic.* Retrieved from https://www.unaids.org/en/resources/documents/2024/UNAIDS_FactSheet

Wang, W., Wang, E., & Balthasar, J. (2008). Monoclonal Antibody Pharmacokinetics and Pharmacodynamics. *Clinical Pharmacology & Therapeutics*, *84*(5), 548–558. https://doi.org/10.1038/clpt.2008.170

Wannier, T. M., Gillespie, S. K., Hutchins, N., McIsaac, R. S., Wu, S., Shen, Y., Campbell, R. E., Brown, K. S., & Mayo, S. L. (2018). Monomerization of far-red fluorescent proteins. *Proceedings of the National Academy of Sciences*, *115*(48). https://doi.org/10.1073/pnas.1807449115

West, A. P., Diskin, R., Nussenzweig, M. C., & Bjorkman, P. J. (2012). 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), E2083–E2090. https://doi.org/10.1073/pnas.1208984109

Wu, X., Yang, Z., Li, Y., Hogerkorp, C., Schief, W. R., Seaman, M. S., Zhou, T., Schmidt, S. D., Wu, L., Xu, L., Longo, N. S., McKee, K., O'Dell, S., Louder, M. K., Wycuff, D. L., Feng, Y., Nason, M., Doria-Rose, N., Connors, M., . . . Mascola, J. R. (2010). Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*, *329*(5993), 856–861. https://doi.org/10.1126/science.1187659

Zaidi, S. (2024). *Differential treatment: restricted access to newer antiretrovirals. United Nations*. https://www.un.org/en/chronicle/article/differential-treatment-restricted-access-newer-antiretrovirals

Zhou, T., Georgiev, I., Wu, X., Yang, Z.-Y., Dai, K., Finzi, A., Do Kwon, Y., Scheid, J. F., Shi, W., Xu, L., Yang, Y., Zhu, J., Nussenzweig, M. C., Sodroski, J., Shapiro, L., Nabel, G. J., Mascola, J. R., & Kwong, P. D. (2010). Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. *Science*, *329*(5993), 811–817. https://doi.org/10.1126/science.1192819

C h a p t e r 2

STRUCTURE-GUIDED SCHEMA RECOMBINATION OF VRC01-CLASS ANTIBODIES FOR REDUCED POLYREACTIVITY

[This chapter is temporarily embargoed.]

C h a p t e r 3

USING HIGH-THROUGHPUT SURFACE PLASMON RESONANCE (SPR-32) TO DETERMINE THE CONCENTRATION OF MONOCLONAL ANTIBODIES IN CRUDE TRANSFECTED CELL SUPERNATANTS: ASSAY DEVELOPMENT AND VALIDATION

[This chapter is temporarily embargoed.]

C h a p t e r 4

DESIGNED OLIGOMERS OF CYANOVIRIN-N SHOW ENHANCED HIV NEUTRALIZATION

Adapted from :

Keeffe, J.R., Gnanapragasam, P.N., **Gillespie, S.K.**, Yong, J., Bjorkman, P.J., Mayo, S.L. (2011). Designed oligomers of cyanovirin-N show enhanced HIV neutralization. PNAS; 108(34): 14079-14084. doi: 10.1073/pnas.1108777108.

S.K.G conducted the experiments and analyzed the data showing that CVN_2L0 is domain swapped in solution. S.K.G. created the original figure 4 (showing that the anti-HIV activity of CVN_2L0 correlates with the number of functional binding sites), as well as supplemental figure S3 (showing that the partial binding site knockout mutants do not significantly affect the neutralization potency). S.K.G. wrote a draft to describe the experiments and results that led to figure 4 and S3. This draft was later rewritten and improved by J.R.K. J.R.K. also improved S.K.G.'s original figures (changed the original circles to squares) for publication.

4.1 Abstract

Cyanovirin-N (CV-N) is a small, cyanobacterial lectin that neutralizes many enveloped viruses, including human immunodeficiency virus type I (HIV-1). This antiviral activity is attributed to two homologous carbohydrate binding sites that specifically bind high mannose glycosylation present on envelope glycoproteins such as HIV-1 gp120. We created obligate CV-N oligomers to determine whether increasing the number of binding sites has an effect on viral neutralization. A tandem repeat of two CV-N molecules (CVN_2) increased HIV-1 neutralization activity by up to 18-fold compared to wild-type

CV-N. In addition, the CVN_2 variants showed extensive cross-clade reactivity and were often more potent than broadly neutralizing anti-HIV antibodies. The improvement in activity and broad cross-strain HIV neutralization exhibited by these molecules holds promise for the future therapeutic utility of these and other engineered CV-N variants.

4.2 Introduction

Cyanovirin-N (CV-N), a cyanobacterial lectin, is uniquely positioned to become a novel therapeutic and prophylactic for diseases caused by enveloped viruses. CV-N is a small, two-domain protein that neutralizes HIV by specifically binding to high mannose glycans on the envelope glycoprotein gp120, thereby preventing interaction of the virus with a host cell (1, 2). In addition to its potent activity against HIV, CV-N is also active against a number of other enveloped viruses including influenza (3, 4), Ebola (5, 6), hepatitis C (7), and herpesvirus 6 (2).

The two domains of CV-N are homologous in both their sequence (32% sequence identity and 58% sequence similarity (8)) and their three-dimensional structure (9, 10). Wild-type (WT) CV-N exists mainly as a monomer in solution and a domain-swapped dimer in crystals (Fig. S1). NMR structures of the monomer show that the protein is an ellipsoid with ten β-strands and four 3_{10}-helical turns, approximately 55 Å in length and 25 Å wide (Fig. S1*A*). Each CV-N monomer contains two symmetrically-related structural domains (A and B), with each domain containing a carbohydrate binding site that specifically interacts with α (1-2) linked oligomannose moieties within Man-8 or Man-9 glycans (9, 11-13). Domain A contains the N- and C-termini and includes residues 1–39 and 90–101, and domain B contains residues 40–89. Although purified as a monomer, a trapped, metastable domain-swapped dimer can be formed during folding and crystallization, and WT CV-N crystallizes exclusively as a domain-swapped dimer (10, 14, 15). In the domain-swapped dimer structure, domain A interacts with B′ to form a “monomer-like unit,” whereas domain A′ and B interact to form the second “monomer-like unit” (Fig. S1*B* and *C*). The domain swapping does not result in additional intramolecular interactions, but instead results from the extension of residues 50 to 53

across the interface. Carbohydrate binding is similar in the domain-swapped crystal structures and monomeric NMR structures, although binding in the B and B′ domains are not seen in crystal structures due to potential steric constraints or crystal packing artifacts (15) (Fig. S1*A* and *B*).

The two binding sites in monomeric CV-N exhibit distinct affinities for carbohydrate in solution: the binding site in domain B, located distal from the N- and C-termini, has an equilibrium dissociation constant (K_D) of approximately 140 nM for Manα1 $\rightarrow$ 2Man disaccharide, which is about 10-fold higher affinity than the binding site in domain A, located near the termini, which binds to Manα1 $\rightarrow$ 2Man disaccharide with a K_D of about 1.5 μM (9). Numerous studies have shown, however, that both sites are necessary for viral neutralization and that destruction of either site renders the CV-N variant inactive (16, 17). However, a recent study showed that in the context of a CV-N dimer that was covalently crosslinked using disulfide bonds, two out of the four possible binding sites are sufficient to maintain neutralization activity, indicating that it is the number and not the identity of sites that is important for neutralization (18, 19). These results point toward a key role for avidity in the viral neutralization activity of CV-N.

A number of groups have attempted to study the oligomerization of CV-N to determine whether the domain swapping is a crystallization artifact or a biologically relevant state. However, because the domain-swapped dimer of WT CV-N is not stable at physiological temperatures, a significant amount of purified dimer may revert to monomer during the course of a viral neutralization assay (14). Therefore, mutations have been used to stabilize either the monomer (14) or the domain-swapped dimer (14, 20, 21). However, the effect of dimerization remains unclear, as some groups have concluded that the dimeric state is more active than monomeric WT CV-N (21), whereas others find that monomeric and dimeric variants have similar antiviral activities (20).

In this study, we show that by linking two CV-N molecules together in a head-to-tail fashion, we can stabilize the domain-swapped dimeric form of the protein in solution.

These linked dimers show enhanced HIV neutralization compared to WT CV-N against 33 strains from 3 clades. In addition, we show that although two carbohydrate binding sites are sufficient for activity as previously reported (18, 19), variants with more binding sites (three or four) have increased neutralization activity.

4.3 Results

Design and Construction of CV-N Oligomers. To directly assay the effects of multimerization on the activity of CV-N, we generated CV-N dimers (CVN_2s) containing tandem repeats of CV-N in which the C-terminus of one copy of CV-N was linked to the N-terminus of the next copy through a flexible polypeptide linker. Because WT CV-N has the ability to domain swap, we hypothesized that the oligomeric molecules would adopt either a monomeric-like linked structure in which the two CV-N repeats are folded as monomers and connected through the linker (Fig. 1*A* and *B*) or a domain-swapped linked structure that takes a form similar to the domain-swapped dimeric crystal structures of WT CV-N but contains a direct linkage between the two CV-N repeats (Fig. 1*C* and *D*). We generated different versions of CVN_2 proteins using 10 different linker lengths ranging from 0 to 20 amino acids (Table S1). We also constructed trimers (CVN_3) with linkers containing 0, 5, or 10 amino acids and a tetramer (CVN_4) in which two CVN_2L0 molecules were linked through a 20 amino acid linker.

CV-N Oligomers Show Enhanced HIV Neutralization. CV-N and CV-N oligomers were assayed for their ability to neutralize HIV in an in vitro luminescence-based neutralization assay (Fig. 2*A* and Table S2) (22). Initial characterizations of the CV-N proteins were conducted by assaying activity against the clade B HIV strain SC422661.8. WT CV-N showed half-maximal neutralization at concentrations (IC_{50}s) between 1.0 and 9.4 nM (0.012 to 0.12 μg/mL) over 30 independent trials, with an average of 4.4 ± 2.6 nM (0.054 ± 0.032 μg/mL), consistent with published values (20, 23-25).

Most dimeric variants (CVN_2s) exhibited more potent HIV neutralization than WT CV-N, with enhancements of approximately 3- to 6-fold when correcting for the

increased molecular weight of the dimers (Fig. 2*B*). The single exception was CVN_2L20, which displayed an IC_{50} similar to that seen for WT CV-N. In this case, the long linker allows a different domain-swapping state relative to variants with shorter linkers. Additionally, CVN_2L20 was only assayed on a single occasion (in triplicate) and therefore the true behavior of the molecule may not be accurately represented. Although there were significant increases in potency for most of the dimers, the addition of the third CV-N repeat to make the CVN_3 proteins did not improve HIV neutralization compared to the dimeric variants; in fact the trimers showed decreased potency compared to the lower molecular weight dimeric molecules (Fig. 2*C*). A tetrameric CV-N protein (CVN_4) showed similar activity to the CVN_3 molecules (data not shown).

CV-N Dimers Exhibit Broad Cross-Clade Reactivity and Are Comparable or More Potent Than Anti-HIV Antibodies. We then assayed WT CV-N and two of our more potent dimers (CVN_2L0 and CVN_2L10) for their ability to neutralize other strains of HIV across different clades. The proteins were tested against a total of 33 strains from three clades. WT CV-N and the dimeric mutants effectively neutralized all 33 HIV pseudoviruses (IC_{50}s less than 300 nM) (Table S3). Consistent with earlier results, the dimeric variants were more potent than WT CV-N against all strains tested (Fig. 3*A*). In 32 out of 33 cases, CVN_2L0 neutralized virus with greater potency than CVN_2L10, whereas CVN_2L10 was the most potent against one clade C strain.

We were also interested in the overall potency of the CV-N proteins compared to known broadly neutralizing anti-HIV antibodies (NAbs): 4E10 (26), 2G12 (26, 27), 2F5 (26, 28), b12 (29), PG9 (30), PG16 (30), and VRC01 (31). IC_{50} values for these NAbs were taken from the literature (22, 32, 33) or determined for this study (PG9, PG16, and VRC01) and converted to molarity for comparisons with the smaller CV-N proteins. When compared to each of the seven broadly neutralizing anti-HIV antibodies, CVN_2L0 showed similar or greater potency against most of the strains tested (Fig. 3*B*). Since 2-fold differences in IC_{50} values are generally not significant, we consider IC_{50}s within 2-fold of the IC_{50} of the antibody to be similar. CVN_2L0 has similar or greater potency against 100% of the

viruses compared to 4E10, 2G12, and 2F5. Our molecule also fairs very well when compared to the newly discovered broadly neutralizing antibodies PG9 (72%), PG16 (66%), and VRC01 (77%).

CV-N Dimers Are Domain-Swapped in Crystal Structures. To elucidate a mechanism for the enhanced neutralization activity of the dimeric proteins, we solved crystal structures of CVN_2L0 and CVN_2L10 (*SI text*, Table S4, and Fig. S2). Both structures were very similar to each other and were intramolecularly domain-swapped with no major deviations relative to WT CV-N domain-swapped structures. Small differences between the CVN_2 structures and previously published WT CV-N structures were noted near the termini, as expected from the linkage, and also in the region of the domain swap. These differences, however, are minor and are unlikely to be responsible for the observed differences in antiviral activity.

Binding Site Mutants of CVN_2L0 Show That It Is Domain-Swapped in Solution. Although crystallographic studies definitively showed that the CVN_2 molecules are domain-swapped in crystals, it was still possible that the dimeric proteins were folded as two linked monomers (Fig. 1*A*) rather than as a domain-swapped dimer (Fig. 1*B*) in solution. To address this issue, we generated variants of CVN_2L0 that contained previously described CV-N carbohydrate binding site knockouts (16, 34) (Fig. 4*A*). Each of the binding sites in a domain-swapped dimer (A, B, A′, and B′) are formed from residues in both CV-N repeats, so we created variants in which a binding site in either the A or B domain was knocked out completely in the context of the linked monomer dimeric model ($CVN_2L0_{\Delta Amm}$, $CVN_2L0_{\Delta Bmm}$) or the domain-swapped dimeric model ($CVN_2L0_{\Delta A}$, $CVN_2L0_{\Delta B}$) (Table S5). In each case, if the model that the mutations were based on is correct, the mutations would form a single full binding site knockout (black squares in Fig. 4*A*). However, if the model is incorrect, the mutations would form two half-site knockouts (black triangles in Fig. 4*A*). We hypothesized that CVN_2L0 mutants with a complete binding site knockout in solution would show decreased ability to neutralize HIV, whereas mutants with two half-site knockouts would be less severely

affected (34). Control variants with only a single half-site knockout showed only modest decreases in potency, verifying that half-site knockouts could be distinguished from complete binding site knockouts (Table S5, Fig. S3).

Variants with binding site knockouts made in the context of the domain-swapped dimeric model were significantly less active against HIV than those designed based on the monomeric model (Fig. 4*B*), indicating that the linked dimers are domain-swapped in solution as well as in crystals.

The Number of Functional Binding Sites in CVN_2L0 Is Directly Proportional to Its Anti-HIV Activity. To determine whether CVN_2L0 has enhanced neutralization activity relative to WT CV-N due to its increased number of binding sites, we created a series of binding site knockout mutants in which 1, 2, or 3 of the sites were fully knocked out (Fig. 4*C*). Variants with only a single binding site knockout showed approximately 20- to 35-fold decreases in potency relative to CVN_2L0, whereas those with two binding site knockouts exhibited decreases of 80- to almost 2000-fold (Fig. 4*D*). Variants with three binding sites knocked out were unable to neutralize HIV at the concentrations tested (data not shown). These data are consistent with previously published accounts, which showed that at least two functional binding sites are required for activity (16, 19), and are also consistent with the hypothesis that avidity is an important factor for viral neutralization by CV-N. Interestingly, CVN_2L0 variants that contain one functional A and one functional B binding site did not neutralize with the same potency as WT CV-N. $CVN_2L0_{\Delta A||B}$ and $CVN_2L0_{\Delta AxB}$ in which the active A and B binding sites are on the same pseudo-monomer or opposite pseudo-monomer (Fig. 4*C*), respectively, were 15- to 75-fold less potent than monomeric WT CV-N, which also contains only a single A and single B binding site. This observation indicates that factors in addition to avidity, including possible steric occlusion of binding sites and/or the relative orientation of binding sites, contribute to the potency of an oligomeric CV-N molecule.

4.4 Discussion

By covalently linking two or more copies of CV-N together, we generated CV-N oligomers with significantly enhanced HIV neutralization activity compared to WT CV-N. CVN_2L0 not only exhibited broadly neutralizing activity across three clades of HIV-1, but was also able to neutralize many HIV strains with potency similar to that of seven well-studied broadly neutralizing antibodies (4E10, 2G12, 2F5, b12, PG9, PG16, and VRC01). In addition, our crystallographic and mutagenesis studies revealed that the CVN_2s form domain-swapped dimers in solution as well as in crystal form. By increasing the local concentration of CV-N through linking two molecules together, we have stabilized the domain-swapped dimeric form of the protein, allowing it to be stable under physiological conditions.

Previous studies were divided about whether the domain-swapped dimer of WT CV-N is more active than the monomeric form (20, 21). However, because the CV-N domain-swapped dimer is meta-stable under physiological temperatures and significant amounts can revert to monomer over the course of an assay (14), the domain-swapped form has been difficult to evaluate with current HIV neutralization assays. In contrast, our variants are covalently linked at their termini and are thereby effectively forced into the domain-swapped dimeric form by the effective increase in local concentration. Because there were no major structural differences between the linked dimers and WT CV-N, our results suggest that the dimeric species of WT CV-N would also be a more potent neutralization agent if it were stable during the assay. Therefore, other methods in which the dimer is stabilized may also result in increased neutralization activity.

In addition to the potential benefit of domain swapping, the simple increase in avidity in the CVN_2s significantly improves the neutralization activity. WT CV-N itself has a high affinity for gp120 (1, 23), but by doubling the number of carbohydrate binding sites in the CVN_2 variants, the increase in avidity may prevent possible dissociation and escape of the virus. As shown by our knockout studies, CVN_2L0 variants with more functional binding sites were significantly more potent at neutralizing HIV, indicating that higher

avidity translates to greater potency. Interestingly, we also found that deletion of a binding site in the B domain had a greater effect on the ability to neutralize HIV than deletion of a binding site in the A domain. $CVN_2L0_{\Delta BB}$, which contains only two functional A sites, is approximately 15-fold worse at neutralizing HIV than $CVN_2L0_{\Delta AA}$, which contains only two functional B sites. This is consistent with earlier studies that showed that binding site B has an approximately 10-fold lower K_D for Manα1-2Man than binding site A (9) and may indicate that the overall activity of CV-N could be improved by improving the affinity of site A.

An alternate mechanism for increased neutralization could result from the fact that the binding sites in CVN_2s can potentially sample distances farther apart than the binding sites in monomeric WT CV-N. The wider spacing could allow CVN_2s to crosslink glycosylation sites within a single gp120, across multiple gp120 subunits on an envelope spike or, less likely, across multiple spikes. This crosslinking would prevent a larger number of gp120 subunits from binding to CD4, the primary receptor for HIV, than would be blocked by WT CV-N, thus decreasing the IC_{50}. An interesting note is that in one conformation of the domain-swapped structure of WT CV-N (Fig. S1*B*), every pair of carbohydrate binding sites is approximately 30 to 35 Å apart (Fig. 5*A*). The neutralizing antibody 2G12, which is also domain-swapped and binds carbohydrates on gp120, has carbohydrate binding sites that are also approximately 30 to 35 Å apart (35) (Fig. 5*B*). Perhaps by stabilizing the domain-swapped structure of CV-N, the carbohydrate binding sites of the CVN_2 variants are optimally positioned to interact with gp120.

While addition of a second CV-N molecule increases the potency of HIV neutralization significantly, the addition of a third or fourth CV-N repeat (CVN_3, CVN_4) does not increase it further. Although the mechanism for enhanced activity is not fully understood, perhaps non-domain-swapped CV-N repeats do not have a significant impact on the activity; for example, one of the three repeats in trimeric CVN_3 molecules may have little effect on activity. Alternatively, due to the close proximity of the N- and C-termini in the

WT structure and their proximity to the lower affinity carbohydrate binding site (site A), the additional CV-N molecule(s) may sterically occlude access to some of the carbohydrate binding sites in the molecule, rendering those sites nonfunctional and therefore inhibiting any additional effect. Longer linkers, including structured linkers, may be necessary to prevent steric occlusion of the binding sites.

WT CV-N and the CVN_2 molecules show excellent cross-clade and cross-strain reactivity. This property is promising for the development of these or other variants for therapeutic use, as they can potentially be used throughout the world. Because of the increase in potency relative to WT CV-N, CVN_2L0 could be more effective in any prophylactic treatment protocol that WT CV-N is currently being investigated for, including gels, suppositories, and in vivo *Lactobacillus* delivery (36, 37). In addition to the increase in potency of CVN_2L0, the lack of a proteolytically sensitive linker between the CV-N repeats suggests that this variant will probably have similar stability in vivo as WT CV-N. CVN_2L0 shows similar potency to many of the broadly neutralizing antibodies that have recently been reported but is easier to express than intact antibodies and therefore could be used for a range of therapeutics that are intractable for antibodies. CV-N variants could also theoretically be used in combination therapy with anti-gp120 antibodies to direct gp120 evolution toward decreased glycosylation. Glycosylation itself has been shown to be important in the folding and function of viral glycoproteins (38), and in the case of HIV, deglycosylation of gp120 diminishes its binding to CD4, making the virus less infective (39, 40). Alternatively, deglycosylation of gp120 could reveal protein epitopes that can be recognized by the adaptive immune system, allowing the immune system to fight off infection more effectively.

4.5 Materials and Methods

Construct Generation. The gene for WT CV-N was constructed using a recursive PCR method with 40-mer synthesized oligos (41), then sub-cloned into the NdeI and BamHI sites of pET11a. The protein contained an N-terminal 6-histidine purification tag followed by a Factor Xa protease cleavage site. CVN_2L5 and CVN_2L10 were constructed

using PCR-based cloning to insert a tandem repeat of the WT CV-N gene and sequence encoding the flexible polypeptide linker into the WT plasmid. The CVN_3L5 gene was created by inserting an *Escherichia coli*-optimized WT CV-N DNA sequence between the two existing copies of the WT gene in CVN_2L5. Other dimeric and trimeric genes of varying linker lengths were constructed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) to insert or delete codons corresponding to linker amino acids. All constructs were verified through DNA sequencing and restriction analysis to ensure the correct sequence and number of CV-N repeats.

Binding site knockout mutant constructs were generated in the background of a CVN_2L0 template gene containing two distinct DNA sequences for each CV-N repeat. Mutations were made using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene).

Protein Expression and Purification. The expression of WT CV-N and all oligomeric variants was induced with IPTG in BL21(DE3) *E. coli* cells in LB including ampicillin. The harvested cells were lysed using an EmulsiFlex-C5 (Avestin, Inc.), and the insoluble fraction was resuspended in buffer containing 6 M GnHCl and 10 mM imidazole and centrifuged to remove debris. The solubilized CV-N was then purified under denaturing conditions using a Ni-NTA gravity column (Qiagen) and refolded by dialyzing the Ni-NTA eluate against native buffer overnight at room temperature (42). Following refolding, proteins were additionally purified on a Superdex-75 column and eluted in 25 mM sodium phosphate pH 7.4, 150 mM NaCl. The N-terminal 6-histadine purification tag was not removed prior to functional or structural assays. Pure protein was concentrated or stored as eluted at 4˚C.

Amino acid analysis was performed on WT CV-N, CVN_2L5, CVN_2L10, CVN_3L5, and CVN_3L10 to determine extinction coefficients at 280 nm (Texas A&M University). These experimentally determined extinction coefficients (WT: 10471 $M^{-1}cm^{-1}$; CVN_2s: 20800 $M^{-1}cm^{-1}$; CVN_3s: 32000 $M^{-1}cm^{-1}$) were used to calculate the protein concentration.

HIV Neutralization Assays. HIV-1 pseuodovirus particles from pseudotyped primary virus strains were prepared as described (22, 43). The SC422661.8 strain (clade B) was used for all assays unless otherwise noted. HIV neutralization assays were performed either in-house (Fig. 2 and Table S2) or by the Collaboration for AIDS Vaccine Discovery (CAVD) core neutralization facility (Fig. 3 and Table S3) as previously described (22). Briefly, 250 infectious viral units of virus per well were incubated with three-fold dilutions of CV-N or a CV-N variant in triplicate (our assays) or duplicate (CAVD assays) for one hour at 37˚C after which approximately 10,000 Tzm-Bl cells were added to each well and incubated for 48 h. The cells were then lysed using Bright Glo Luciferase Assay Buffer (Promega), and luciferase expression was assayed using a Victor^3 Multilabel Counter (PerkinElmer).

To determine the IC_{50} of neutralization, the luminescence was first averaged across the replicates, then the percent neutralization *(%Neutralization*) was calculated using Equation 1, where *RLU* is the average relative luminescence for a given concentration, *CC* is the average luminescence from the cell control wells, and *VC* is the average luminescence from the viral control wells. The percent of virus neutralized was then plotted as a function of neutralizing protein in Kaleidagraph (Synergy Software) and fitted to Equation 2, where IC_{50} is the concentration of CV-N at which half of the virus is neutralized, *C* is the concentration of CV-N, and *m* is a Hill coefficient. IC_{50}s are reported as the average of a minimum of four independent trials and the error reported is the standard deviation of the IC_{50}s from those trials. CVN_2L20, CVN_3L5, CVN_3L10, and the binding site mutants were tested on only one occasion and therefore a standard deviation is not reported.

$$\%Neutralization = \left(1 - \frac{RLU - CC}{VC - CC}\right) * 100 \qquad \textbf{[1]}$$

$$\%Neutralization = \frac{100}{1 + \left(\frac{IC_{50}}{C}\right)^{m}} \quad \textbf{[2]}$$

Crystallization, Crystallographic Data Collection, and Refinement. See *SI text*.

ACKNOWLEDGMENTS. We thank Leonard Thomas, Pavle Nikolovski, and the Molecular Observatory at Caltech, which is supported by the Gordon and Betty Moore Foundation, for assistance in setting up crystal trays, collecting and processing diffraction data, and refining crystal structures. We also would like to thank the Collaboration for AIDS Vaccine Discovery Neutralizing Antibody Core Laboratories for performing HIV neutralization assays and Marie Ary for critical review of the manuscript. This work was funded by the National Security Science and Engineering Faculty Fellowship program and the Defense Advanced Research Projects Agency Protein Design Processes program (to S.L.M.) and by the Bill and Melinda Gates Foundation Grant 38660 through the Grand Challenges in Global Health Initiative (to P.J.B.). X-ray data for CVN_2L10 were collected at the Stanford Synchrotron Radiation Lightsource (Beam Line 12-2). Operations at Stanford Synchotron Radiation Lightsource are supported by the Department of Energy and the National Institutes of Health.

4.6 References

1. Boyd MR, et al. (1997) Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* 41(7):1521-1530.

2. Dey B, et al. (2000) Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *J Virol* 74(10):4562-4569.

3. O'Keefe BR, et al. (2003) Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* 47(8):2518-2525.

4. Smee DF, et al. (2007) Influenza A (H1N1) virus resistance to cyanovirin-N arises naturally during adaptation to mice and by passage in cell culture in the presence of the inhibitor. *Antivir Chem Chemother* 18(6):317-327.

5. Barrientos LG, Gronenborn AM (2005) The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini Rev Med Chem* 5(1):21-31.

6. Barrientos LG, et al. (2003) Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* 58(1):47-56.

7. Helle F, et al. (2006) Cyanovirin-N inhibits hepatitis C virus entry by binding to envelope protein glycans. *J Biol Chem* 281(35):25177-25183.

8. Gustafson KR, et al. (1997) Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-N, an anti-HIV (human immunodeficiency virus) protein from the cyanobacterium Nostoc ellipsosporum. *Biochem Biophys Res Commun* 238(1):223-228.

9. Bewley CA, Otero-Quintero S (2001) The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *J Am Chem Soc* 123(17):3892-3902.

10. Yang F, et al. (1999) Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol* 288(3):403-412.

11. Bolmstedt AJ, O'Keefe BR, Shenoy SR, McMahon JB, Boyd MR (2001) Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. *Mol Pharmacol* 59(5):949-954.

12. Shenoy SR, et al. (2002) Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. *Chem Biol* 9(10):1109-1118.

13. Shenoy SR, O'Keefe BR, Bolmstedt AJ, Cartner LK, Boyd MR (2001) Selective interactions of the human immunodeficiency virus-inactivating protein cyanovirin-N with high-mannose oligosaccharides on gp120 and other glycoproteins. *J Pharmacol Exp Ther* 297(2):704-710.

14. Barrientos LG, et al. (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* 10(5):673-686.

15. Botos I, et al. (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* 277(37):34336-34342.

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

17. Fromme R, et al. (2007) A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry* 46(32):9199-9207.

18. Liu Y, et al. (2009) Multivalent interactions with gp120 are required for the anti-HIV activity of Cyanovirin. *Biopolymers* 92(3):194-200.

19. Matei E, et al. (2010) Anti-HIV activity of defective cyanovirin-N mutants is restored by dimerization. *J Biol Chem* 285:13057-13065.

20. Barrientos LG, Lasala F, Delgado R, Sanchez A, Gronenborn AM (2004) Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity. *Structure* 12(10):1799-1807.

21. Kelley BS, Chang LC, Bewley CA (2002) Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. *J Am Chem Soc* 124(13):3210-3211.

22. Li M, et al. (2005) Human immunodeficiency virus type 1 *env* clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 79(16):10108-10125.

23. Colleluori DM, et al. (2005) Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. *Protein Expr Purif* 39(2):229-236.

24. Mori T, et al. (2002) Functional homologs of cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. *Protein Expr Purif* 26(1):42-49.

25. Mori T, et al. (1997) Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. *Biochem Biophys Res Commun* 238(1):218-222.

26. Buchacher A, et al. (1994) 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(4):359-369.

27. Trkola A, et al. (1996) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* 70(2):1100-1108.

28. Purtscher M, et al. (1994) A broadly neutralizing human monoclonal antibody against gp41 of human immunodeficiency virus type 1. *AIDS Res Hum Retroviruses* 10(12):1651-1658.

29. Burton DR, et al. (1991) A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. *Proc Natl Acad Sci U S A* 88(22):10134-10137.

30. Walker LM, et al. (2009) Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326(5950):285-289.

31. Wu X, et al. (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 329(5993):856-861.

32. Li M, et al. (2006) Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular *env* clones from acute and early heterosexually acquired infections in Southern Africa. *J Virol* 80(23):11776-11790.

33. West AP, Jr., et al. (2010) Evaluation of CD4-CD4i antibody architectures yields potent, broadly cross-reactive anti-human immunodeficiency virus reagents. *J Virol* 84(1):261-269.

34. Chang LC, Bewley CA (2002) Potent inhibition of HIV-1 fusion by cyanovirin-N requires only a single high affinity carbohydrate binding site: characterization of low affinity carbohydrate binding site knockout mutants. *J Mol Biol* 318(1):1-8.

35. Calarese DA, et al. (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* 300(5628):2065-2071.

36. Liu X, et al. (2006) Engineered vaginal lactobacillus strain for mucosal delivery of the human immunodeficiency virus inhibitor cyanovirin-N. *Antimicrob Agents Chemother* 50(10):3250-3259.

37. Ndesendo VM, et al. (2008) A review of current intravaginal drug delivery approaches employed for the prophylaxis of HIV/AIDS and prevention of sexually transmitted infections. *AAPS PharmSciTech* 9(2):505-520.

38. Vigerust DJ, Shepherd VL (2007) Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol* 15(5):211-218.

39. Fenouillet E, Gluckman JC, Bahraoui E (1990) Role of N-linked glycans of envelope glycoproteins in infectivity of human immunodeficiency virus type 1. *J Virol* 64(6):2841-2848.

40. Montefiori DC, Robinson WE, Jr., Mitchell WM (1988) Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* 85(23):9248-9252.

41. Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164(1):49-53.

42. Barrientos LG, et al. (2002) Design and initial characterization of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N. *Proteins* 46(2):153-160.

43. Wei XP, et al. (2003) Antibody neutralization and escape by HIV-1. *Nature* 422(6929):307-312.

44. Bewley CA (2001) Solution structure of a cyanovirin-N:Manα1-2Manα complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* 9(10):931-940.

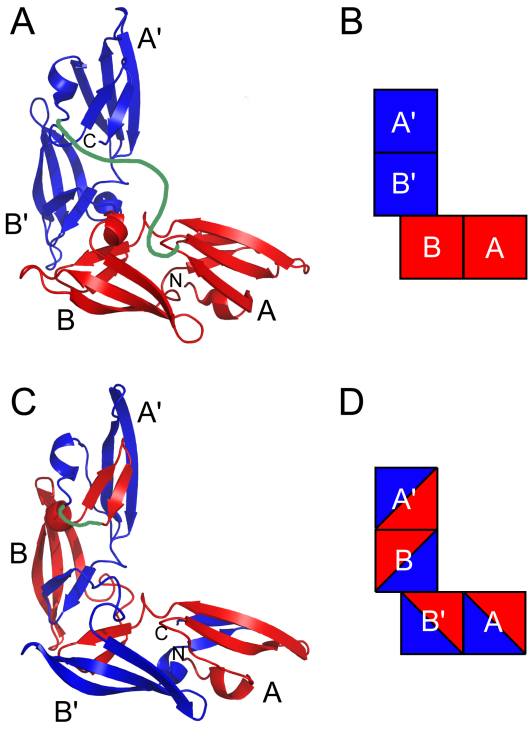

Fig. 1. Two models of CVN_2 proteins. (*A, B*) The linked monomer dimeric model for CVN_2s. (*C, D*) The linked domain-swapped dimeric model for CVN_2s. The CV-N repeats are shown in red and blue, and the flexible polypeptide linker is modeled in green. The N- and C-termini are labeled N and C. Each of the four carbohydrate binding domains are labeled A, B, A′, or B′ in each model. Models (*A*) and (*C*) are based on solved WT CV-N structures (10, 51) and (*B*) and (*D*) are block representations of the models in (*A*) and (*C*), respectively.

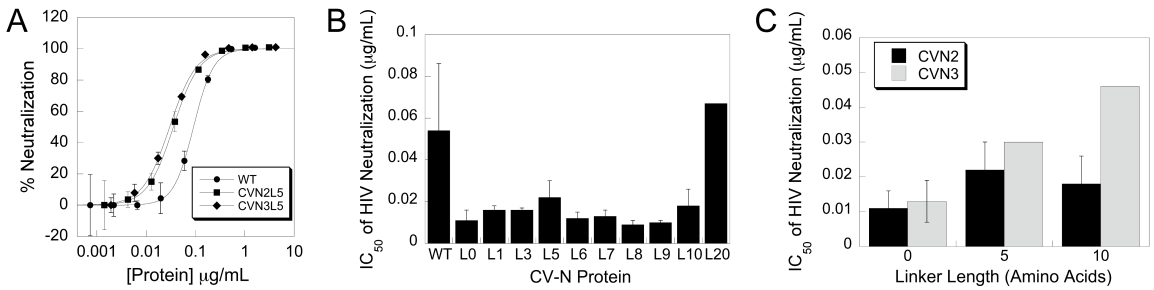

Fig. 2. CV-N oligomers exhibit enhanced HIV neutralization. (*A*) Typical neutralization data and curve fits for WT CV-N and two variants run in triplicate on the same 96-well plate. (*B*) The IC_{50}s of HIV neutralization for WT CV-N and CVN_2 dimers containing varying linker lengths. All linked dimers show significant enhancements in their HIV neutralization compared to WT CV-N, and dimers containing 0 to 10 amino acid linkers are more potent than CVN_2L20. $N \geq 4$; CVN_2L20, $N = 1$. (*C*) IC_{50}s of HIV neutralization for CVN_2s and CVN_3s of the same linker length. $N \geq 3$; CVN_3L5 and CVN_3L10, $N = 1$. (*A-C*) Error bars = SD.

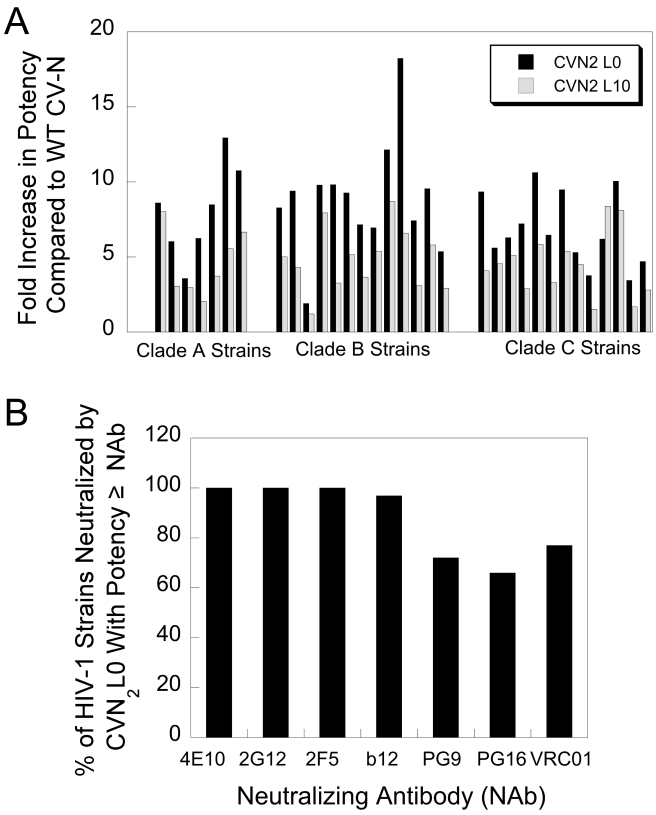

Fig. 3. CV-N dimers neutralize HIV broadly and potency is similar to broadly neutralizing anti-HIV antibodies. (*A*) The designed dimers show enhanced neutralization activity relative to WT CV-N across all 33 strains tested from 3 clades. CVN_2L0 is more potent than CVN_2L10 in 32 of 33 cases. (*B*) When the IC_{50}s of CVN_2L0 neutralization against a panel of HIV-1 strains were compared to the IC_{50}s of seven broadly neutralizing antibodies (Table S3), we saw that most strains were as sensitive to CVN_2L0 as they were to the broadly neutralizing antibodies. Since 2-fold differences in potency are generally not significant, similar potency (≥) is defined as a potency for CVN_2L0 that is within 2-fold of the potency of the antibody or higher.

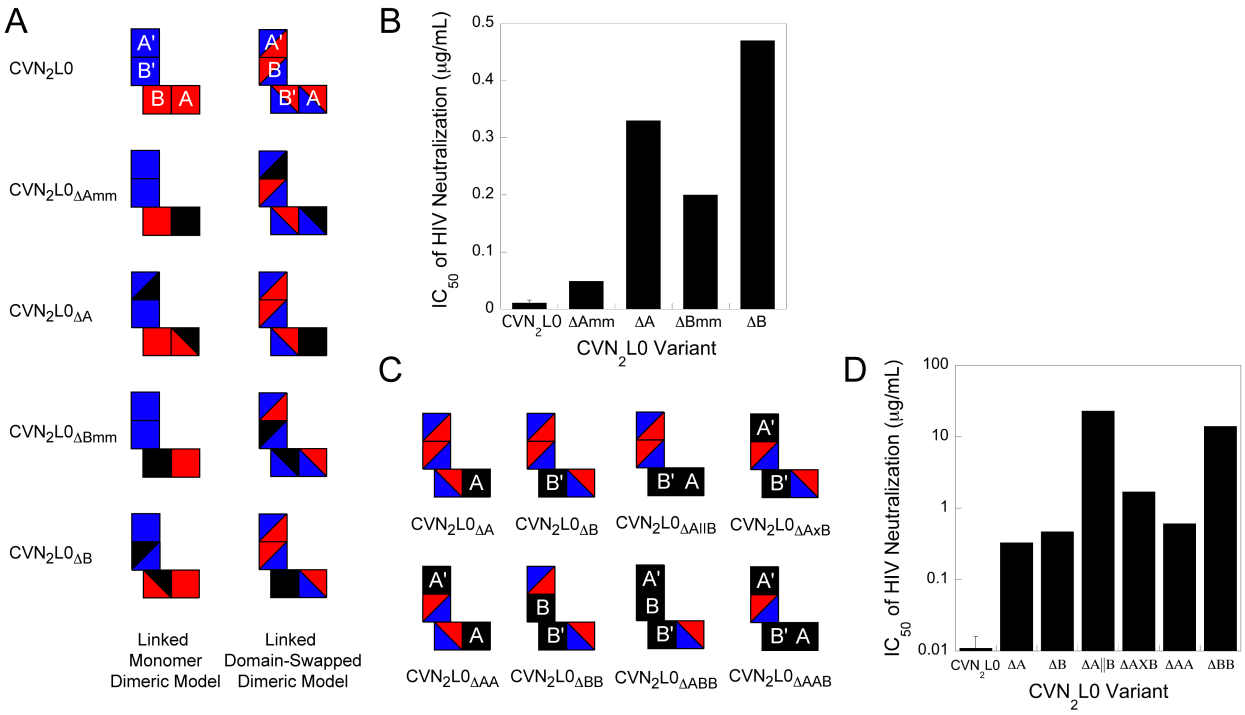

Fig. 4. Anti-HIV activity of CVN_2L0 correlates with number of functional binding sites. (*A*) Schematic representation of variants to determine whether CVN_2L0 is in a linked monomer dimeric structure (mm) or domain-swapped dimeric structure. The two CV-N repeats are represented in red and blue, as in Fig. 1*B* and *D*. Black triangles represent partial carbohydrate binding site knockouts and red squares represent complete binding site knockouts. (*B*) HIV neutralization results for mutants in panel *A*. Mutants with full binding site deletions in the context of the domain-swapped dimer model have more significant increases in their HIV neutralization IC_{50}s compared to mutants with full binding site deletions in the context of the monomer model. (*C*) Schematic representations of multiple binding site mutants. All variants contain one or more complete binding site knockouts according to the CVN_2L0 domain-swapped dimer model. Black squares represent binding sites that have been knocked out, and squares

containing red and blue triangles represent WT (functional) binding sites. (*D*) HIV neutralization results for mutants in panel *C*. The number of functional binding sites in CVN_2L0 is proportional to its ability to neutralize HIV. Mutants with two functional binding sites are less active than those with three sites. Additionally, the deletion of a B binding site has a greater effect on activity than the deletion of an A binding site.

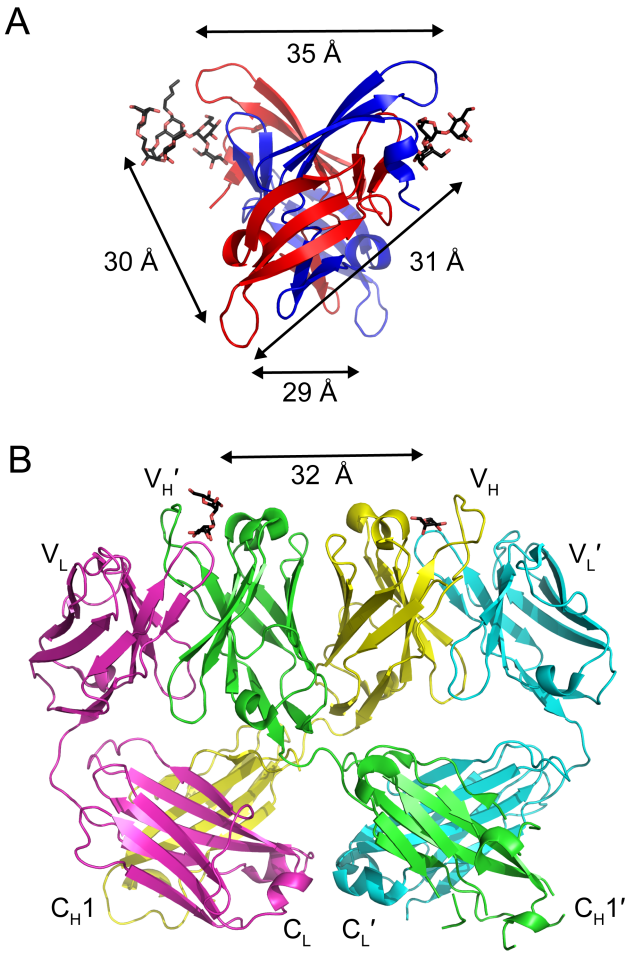

Fig. 5. Similarity in carbohydrate binding site spacing in CV-N and the 2G12 anti-HIV $(Fab)_2$. (*A*) Each of the four carbohydrate binding sites in one WT CV-N crystal structure (15) ($P4_12_12$ space group) is approximately 30 to 35 Å from the other sites (structure is viewed from the bottom with respect to Fig. 1). Carbohydrates (shown as sticks with black carbons and red oxygens) were only resolved in the A binding sites in the crystal structure. (*B*) Ribbon diagram of the domain-swapped $(Fab)_2$ from IgG 2G12, a broadly neutralizing antibody specific for carbohydrates on gp120 (34). The domain swapping creates a rigid $(Fab)_2$ dimer in which the carbohydrate binding sites at the antigen

combining sites are spaced approximately 30 to 35 Å apart. Carbohydrates are shown as sticks with black carbons and red oxygens and antibody domains are labeled.

C h a p t e r 5

MONOMERIZATION OF FAR-RED FLUORESCENT PROTEINS

Adapted from :

Wannier, T.M., **Gillespie S.K.**, Hutchins N., McIsaac R.S., Wu S.Y., Shen Y., Campbell R.E., Brown K.S., Mayo S.L. (2018). Monomerization of far-red fluorescent proteins. PNAS; 115(48): E11294-E11301. doi: 10.1073/pnas.1807449115.

S.K.G. assisted with characterizing the important red fluorescent protein variants in the library. S.K.G. assisted with collecting absorbance scans, fluorescence excitation scans, fluorescence emission scans, measurements of quantum yield, measurements of extinction coefficients, and thermal stability measurements. S.K.G. also assisted with oligomeric determination by analytical ultracentrifugation.

5.1 Abstract

Anthozoa class red fluorescent proteins (RFPs) are frequently used as biological markers, with far-red emitting variants (λ_{em} ~ 600 – 900 nm) sought for whole animal imaging because biological tissues are permeable to light in this range. A barrier to the use of the diversity of naturally occurring RFP variants as molecular markers is that all are tetrameric, which is not ideal for cell biological applications. Efforts to engineer monomeric RFPs have usually produced dimmer and blue-shifted variants, as the chromophore is sensitive to small structural perturbations. In fact, despite much effort, only four native RFPs have been successfully monomerized, leaving the vast majority of RFP biodiversity untapped in biomarker development. Here we report the generation of monomeric variants of HcRed and mCardinal, both far-red dimers, and describe a comprehensive methodology for the rapid monomerization of novel, red-shifted

oligomeric RFPs. Among the resultant variants, is mKelly1 (emission maximum: λ_{em} = 656 nm), which along with the recently reported mGarnet2, forms a new class of bright, monomeric, far-red FPs.

5.2 Introduction

The development of red fluorescent proteins (RFPs) as tags for molecular imaging has long focused on monomerization, increased brightness, and pushing excitation and emission to ever-longer wavelengths. These traits are desired for live animal imaging, as far-red to near infrared light penetrates tissue with minimal absorption in what is known as the near infrared window (~625–1300 nm) (1, 2). Monomericity is important because oligomerization of an FP tag can artificially aggregate its linked protein target, altering diffusion rates and interfering with target transport, trafficking, and activity (3, 4). Recently a new class of infrared fluorescent proteins (iRFPs) was developed from the bacterial phytochrome, but these require the covalent linkage of a small molecule chromophore, biliverdin, limiting their use to cells and organisms that make this molecule in sufficient quantity. *Anthozoa* class RFPs (such as mCherry and mKate) have the advantage that the chromophore is created via a self-processing reaction, necessitating only molecular O_2 for chromophore formation.

To our knowledge, ~50 native RFPs and ~40 chromoproteins (CPs) with peak absorbance in the red or far-red (absorbance maximum: λ_{abs} >550 nm) have been described to date, but most have not been extensively characterized because they are as a class tetrameric, and thus are less useful as biological markers (5, 6). An underlying biological context for the obligate tetramerization of native RFPs has been hinted at, but is not well understood (7-10). Oligomerization does seem to play an important structural role, however, as breaking tetramerization without abrogating fluorescence has proved difficult, and successful monomerization has always led to either a hypsochromic shift to λ_{em} or a decrease in brightness (11-14). Previous efforts to monomerize native RFP tetramers have relied on lengthy engineering trajectories, and have been successful in only four

cases (Table 1). Generally, mutations are first introduced into tight interfaces to weaken oligomerization, an inefficient process that impairs fluorescence, and then random mutagenesis and screening isolate partially recovered variants. After many such cycles monomeric variants have been found, but protein core and chromophore-proximal mutations are invariably introduced, making it difficult to exert any significant degree of control over the fluorescent properties of the resultant monomer. It is thus difficult to know whether the poor spectroscopic characteristics of engineered monomers are an unavoidable consequence of monomerization or only the manifestation of a suboptimal evolutionary path.

Here we present a comprehensive engineering strategy for the monomerization of novel RFPs that differentiates itself by treating separately the problems of protein stabilization, core optimization, and surface design. We sample mutational space both stochastically, through error-prone mutagenesis, and rationally, by analysis of multiple sequence alignments (MSAs) and computational protein design (CPD). Two far-red oligomeric proteins were targeted for monomerization: HcRed (λ_{em} = 633 nm), a dimer/tetramer (15), and mCardinal (λ_{em} = 658 nm), a reported monomer that we have confirmed to in fact be dimeric. The monomeric RFPs reported here include two monomeric HcRed variants: mGinger1 (λ_{em} = 637 nm) and mGinger2 (λ_{em} = 631), and two monomeric mCardinal variants: mKelly1 (λ_{em} = 648 nm) and mKelly2 (λ_{em} = 643 nm), which are among the brightest far-red monomeric FPs to have been reported.

5.3 Results

Step-wise monomerization of HcRed. We first chose HcRed, a far-red FP that has been engineered but never successfully monomerized (15, 16). As we have previously demonstrated that oligomericity and brightness can be treated as separate protein design problems (17), we devised a workflow that separately targets the chromophore environment—to engineer a protein core that maintains structural integrity absent stabilizing oligomeric interactions—and the protein surface—to drive monomerization.

Anthozoa class RFPs have two oligomeric interfaces, named AB and AC (18), with the AC interface being the more stable of the two and burying a large hydrophobic surface (19). Early engineering to HcRed partially disrupted oligomerization at the AB interface, but all mutations to the AC interface were found to vitiate fluorescence. To test the integrity of the AC interface, we made successive deletions to HcRed's C-terminal tail (residues 219-227), which plays an integral role in the AC interaction (Figure 1A). HcRed lost significant brightness with the deletion of just one C-terminal residue, and was non-fluorescent after any further deletion, demonstrating that optimization would be necessary prior to monomerization.

So, we endeavored to engineer a more stable core, identifying two mutational hotspots from an alignment of far-red RFPs: (A) a group of residues that surrounds alternative conformations of the chromophore's phenolate ring and (B) a region above the plane of the chromophore, between the central α-helix and the unbroken AC oligomeric interface (Figure S1). Generally in RFPs, the *cis* chromophore—the phenolate ring sits *cis* to the proximal nitrogen on the imidazolinone ring rather than *trans* to it—is the fluorescent species (20). In engineering HcRed from its chromoprotein parent HcCP, the *cis* chromophore was stabilized over the non-fluorescent *trans* chromophore by way of a cysteine to serine mutation at position 143, which provides a hydrogen bond to the *cis* phenolate oxygen (Figure 1C). We reasoned that further stabilization of the *cis* chromophore would increase brightness, and so designed a first core library (cLibA) to target hotspot A, mutating *trans*-stabilizing amino acids, placing bulkier side chains into the *trans* pocket, and allowing varied hydrogen bonding geometries to the *cis* chromophore. A second core library (cLibB) targeted hotspot B along with two chromophore-backing positions (Gly28 and Met41 are implicated in maturation and color) (19, 21, 22). Two key features of this hotspot are a channel populated by structural water molecules that stretches to the protein surface, and Arg67, a key catalytic residue. Mutations to this region may serve to occlude access to the chromophore by bulk solvent upon monomerization, and to allow room for chromophore processing. Small libraries of < 1,000 protein variants were guided by the far-red RFP algignment (Table S1), and after screening each library to > 95% coverage on large LB agar plates supplemented with

IPTG, we fully characterized 16 cLibA variants and 21 cLibB variants. The variants showed brightness increases of up to ten-fold and displayed an incredible range of emission profiles, with λ_{em} between 606 and 647 nm. To determine which if any variants would be amenable to monomerization, we tested a five-residue tail deletion. Eight variants showed detectable fluorescence after the tail deletion, with a double mutant (HcRed7: R67K/I196Y) being the most red-shifted (λ_{em} = 642 nm). The core mutations in HcRed7 bathochromically shift its emission by 9 nm, improve its quantum yield (Φ) by 60%, and thermostabilize the protein by 6 °C. HcRed7, however, loses significant brightness with the deletion of a sixth tail residue (Figure S5 – HcRed7Δ6) and becomes 16°C less thermostable, indicating that the protein is not wholly optimized for monomerization (Table 2).

To further optimize HcRed7Δ6 for monomerization, we took aim at improving the thermo-stability of the protein Thermostability has been shown to increase a protein's evolvability (23) and consensus design is one of the best tools for improving thermostability (24). We constructed a large MSA that consists of every *Aequorea victoria* class FP; a total of 741 sequences (see supplemental Methods), and then built a library to sample all 105 non-consensus positions in HcRed with the consensus amino acid, and compared this to a strategy of error-prone mutagenesis. We screened the consensus (~1.2 mutations per variant) and error-prone (~1.8 mutations per variant) libraries at 675 nm to allow maximal differentiation between far-red variants whose λ_{em} was between 630-640 nm and a large population of near-red variants whose emission peaked between 605-620 nm, but which were often brighter. The consensus library was screened to 40x coverage (~4,300 clones) and ~8,600 clones were screened from the error-prone library. Consensus library variants significantly outperformed error-prone library variants (Figure S2), and so we combined seven of the top consensus variants together into a chimeric protein, HcRed77, which recovered all of HcRed7Δ6's lost brightness and much of its thermostability.

Finally, to monomerize HcRed77 we targeted the AC interface with a CPD procedure that we describe in previous work (17). We focused on a set of five hydrophobic residues

(Val146, Val159, Ile170, Phe191, and Phe193) at the heart of the AC interface that make extensive intermolecular contacts (Figure S3), and built a small combinatorial library guided by the design. We isolated a first-generation monomer: HcRedm1 and verified it to be monomeric by fast protein liquid chromatography (FPLC) and analytical ultracentrifugation (AUC) (Figure 2), but the protein was dim and expressed poorly. We attributed these poor attributes to incomplete thermo-stabilization of HcRed77, and so we screened mutations from the error-prone library via DNA shuffling and then increased the temperature of screening from 30°C to 37°C for a final round of error-prone mutagenesis and isolated two bright variants with improved brightness and thermostability higher than the parent HcRed7: mGinger1 and mGinger2 (Table 2).

Two-step monomerization of mCardinal. The monomerization of HcRed required three design elements: core optimization, protein thermo-stabilization, and surface design; and pulled mutational diversity from three sources: a large MSA, CPD, and error-prone mutagenesis. While the engineering process for the development of the mGingers was rational and involved a shorter trajectory than past procedures, we felt that it could be further improved by integrating the three design objectives into one large library. We targeted mCardinal, a recently reported variant of mNeptune that was reported to be monomeric, but which we have shown to be dimeric by both FPLC and AUC (Figure 2). In fact, the crystal structure of mCardinal (4OQW) shows the protein adopting a classic tetrameric RFP conformation, similar to DsRed and mCardinal's wild-type progenitor, eqFP578 (25).

First, as with HcRed, we probed tail deletion variants of mCardinal, which was engineered to have a long, 20-AA C-terminal tail. The first 15 residues were easily removed (equivalent to HcRedΔ4), but as with HcRed, mCardinalΔ16 is significantly dimmer and mCardinalΔ18 is essentially non-fluorescent (Figure S5). To isolate error-prone mutations for the combined library approach, we targeted mCardinalΔ19, a near-total tail deletion, with random mutagenesis and isolated six mutations that restored measureable fluorescence and did not hypsochromically shift the emission curve. The six identified error-prone hits together (mCardinal-mut6) restored fluorescence and

thermostability nearly back to that of mCardinal. We then built a monomerization library that included the six stabilizing error-prone mutations and a complete tail deletion (Δ20), and that sampled a CPD-generated AC interface and the nine highest-scoring consensus mutations (Table S2). Because the first generation HcRed monomer needed further optimization for improved brightness, we chose to sample a larger surface design landscape than we did in the case of HcRed, again designing the five-residue core of the AC interface, but also allowing diversity in eight other nearby surface positions. The total theoretical library size was 5.7 x 10^7. After screening 1.1 x 10^5 variants by fluorescence activated cell sorting (FACS), we isolated two variants that were bright, monomeric, and retained a far-red emission: mKelly1 and mKelly2 (Table 2).

5.4 Discussion

Clear Design Objectives Speed Protein Development. We demonstrate that an engineering process that makes use of varied protein engineering tools can hasten the isolation of optimized protein variants. Small, focused libraries enriched for diverse but functional HcRed variants addressed separately the problems of brightness, stability, and oligomericity. Beginning with oligomers partially destabilized by the deletion of their C-terminal tails, we quickly moved through functional sequence space, incorporating 38 and 42 mutations over five rounds of design into mGinger1 and mGinger2 respectively (Figures 3 and S4). DNA shuffling (26, 27) enabled us to screen large numbers of candidate mutations in HcRed, but noting that high-value mutations were enriched during this process, we monomerized mCardinal with just one large library, incorporating 39 and 42 mutations respectively into mKelly1 and mKelly2. Importantly, and unlike previous RFP monomerization efforts, we maintained fluorescence at every design stage, allowing us to be stringent in our selections and to maintain far-red emission. The mutations in the final RFP variants were found by employing complementary but divergent engineering processes. Consensus design was used to improve thermo-stability, which has been shown to improve proteins' evolvability (23, 24), while error-prone mutagenesis added diversity to this pool of stabilizing mutations. Notably, consensus design significantly

outperformed random mutagenesis in improving the brightness of HcRed7 (Figure S2). Finally, to build stable and soluble β-sheet surfaces, an application suited neither to consensus design nor error-prone mutagenesis, we used CPD, which we had previously shown to be well suited to this purpose.

Mutations Accumulate in Key Structural Regions. A total of 45 mutations in mGinger1 and 52 mutations in mKelly1 separate them from their wild-type progenitors HcCP and eqFP578. These mutations cluster structurally, occurring at the designed AC interface, at chromophore-proximal positions, and near pockets of exposed hydrophobic residues on the protein surface. One region of particular note in which mutations cluster is an apparent channel populated by large numbers of structural water molecules that runs from a wide cleft in the β-barrel between β-strands 7 and 10 to a smaller deformation of β-strands 3 and 11, passing through the chromophore pocket (Figure 2B). These deformations of the β-barrel are bisected by the attachment site of the C-terminal tail, and appear to be stabilized by intermolecular interactions between monomers across the AC interface. A break of the AC interface may destabilize the water channel, putting the chromophore environment into contact with bulk solvent, which would in turn interfere with chromophore maturation and quench fluorescence (21, 28). Indeed, mGinger1 and mKelly1 have eleven and six mutations respectively to residues that are in close proximity (4 Å) to structural waters in this channel and that are not a part of the AC interface (Figure S5). Elsewhere, mGinger1 and mKelly1 have eleven and fifteen mutations respectively to their AC interfaces and two and three to their AB interfaces, breaking oligomerization. In mGinger1 we see eight mutations to patches of exposed hydrophobic surface residues not located at the oligomeric interfaces, as mapped by spatial aggregation propensity (SAP) (29, 30), whereas with mKelly1, we relatively fewer new surface mutations, as we expect that the intense selection and engineering that mCardinal had been subjected to had previously optimized its non-interface surfaces for solubility. Outside of these structural clusters, we introduced relatively few new mutations to mGinger1 and mKelly1, five in each case. mKelly1 does inherit eleven other uncharacterized mutations from mCardinal, to both its surface and core.

Protein Stability is Linked to Function. Past efforts to monomerize RFPs have ignored the role that scaffold stability may play in engineering a functional monomer. We suggest that as oligomericity is broken, a loss of structural integrity (approximated here by apparent T_m) can leave single monomers unstable and non-functional. As we monomerized HcRed, we measured the thermal stability of each important intermediate, and found a positive correlation between apparent T_m and quantum yield (Figure 4). This relationship may be related to scaffold rigidity, as in a more rigid excited-state chromophore there is less non-radiative decay of fluorescent energy via thermal motion or other atomic interactions. (31, 32). In small molecule fluorophores this is readily seen, as quantum yield increases with decreased temperature (33), and rational design of a chromophore-proximal β-strand was used improve quantum yield of a cyan FP to 0.93 (31). The correlation between quantum yield and apparent T_m, however, appears to divide into two distinct groups, with dimers having higher quantum yields than monomers. mGinger0.1 and mGinger0.2, for instance, despite being thermostabilized by 5 °C over the parental protein HcRed7, are less bright. It is unclear why a less-thermally stable dimer should be brighter than its thermostabilized monomeric derivative, despite sharing an almost identical protein core.

HcRed7's Structure Explains Brightness and Bathochromic Emission. We solved an x-ray crystal structure of HcRed7, which shows that the mutation from histidine to tyrosine at position 196 serves to add a π-stacking interaction with the chromophore phenolate ring (Figure 1C). Tyr196 π-stacks with the fluorescent *cis* orientation of the phenolate, serving to both stabilize the fluorescent *cis* phenolate over the non-fluorescent *trans* phenolate—HcRed's chromophore occupies both *cis* and *trans* conformations—and to red-shift the λ_{em}, as a π-stacking phenolate interaction has been shown to reduce the energy of the excited state of the chromophore (34-36). In turn, position 67 is a key catalytic residue that functions as a base, abstracting a proton from the bridging carbon of the phenolate side chain during cyclization (37, 38). This residue is almost invariably a lysine or arginine among RFPs, and we propose that the mutation from arginine to lysine here allows room for the π-stacking interaction and the bulkier tyrosine side chain. It has been previously noted that this π-stacking interaction can induce a bathochromic shift in

λ_{em}, but here we note that these two mutations also conveyed a 6 °C improvement to apparent Tm and a 60% improvement in quantum yield.

5.5 Conclusion

We engineered four new monomeric RFPs: mGinger1/2 and mKelly1/2, monomeric variants of the far-red fluorescent proteins HcRed and mCardinal, both dimeric RFPs that had been the targets of unsuccessful monomerization attempts. mKelly1 and mKelly2 join mGarnet1 and mGarnet2 as part of a new class of bright monomeric RFPs with emission peaking near to or longer than 650 nm (Figure 5). Previously we monomerized DsRed using a pre-stabilized core borrowed from mCherry, and showed that monomerization is possible with little to no change to an RFP's spectroscopic properties (17). Here we show that stabilization of the entire protein scaffold is important for monomerization. Despite the mGingers and mKellys being slightly dimmer and hypsochromically shifted from HcRed7 and mCardinal, they move the needle toward longer wavelengths and brighter emission. Past monomerization efforts have been beset by similar loss of brightness, but because they necessitated significant mutation to the core of the protein and the chromophore environment (11, 13, 39-41), it has been difficult to separate the effects of potentially suboptimal core mutation from the inescapable externalities of monomerization. The rational approach that we lay out in monomerizing HcRed and mCardinal includes elements of rational design, computational design, and directed evolution, and represents a marked improvement in both the speed and efficiency of RFP monomerization. Further exploration of stable RFP cores will be necessary to determine how to significantly improve brightness post-monomerization.

5.6 Materials and Methods

Plasmids and Bacterial Strains. The HcRed sequence was taken and modified from the HcCP Genbank entry (accession number AF363776). Ten amino acids were added to the

N-terminus, consisting of a Methionine followed by a 6x Histidine tag for protein purification, followed by a Gly-Ser-Gly linker sequence. All genes were constructed by overlap extension PCR from oligonucleotides designed by DNAworks and ordered from Integrated DNA Technologies (IDT). Assembled genes were PCR-amplified and cloned into the pET-53-DEST expression plasmid (EMD Millipore). Constructs were sequence-verified and transformed into BL21-Gold(DE3) competent cells for protein expression (Agilent).

Construction of Designed Libraries. The HcRed core variants were designed with DNAworks as "mutant runs" of the wild-type gene assembly. The HcRed AC surface library used the triplet codon "VRN" to replace the five design positions, which allows for the possible amino acids D/E/G/H/K/N/Q/R/S. The mCardinal monomer library was designed by hand, as it was too complex for DNAworks; we used degenerate bases where possible. For all libraries, oligonucleotides were ordered from IDT and cloning was carried out as described above.

Error Prone Mutagenesis. Error prone mutagenesis of HcRed variants was performed by addition of manganese chloride to Taq DNA polymerase PCR reactions. 10μM, 15μM, and 20μM $MnCl_2$ were tested and cloned with PIPE cloning into pET-53-DEST for sequencing. Twelve colonies from each library were picked and sequenced, and the library with a mutation rate closest to but not more than 2.0 mutations per gene was selected for further screening.

DNA Shuffling. The variants that were to be shuffled together were PCR-amplified and purified by gel electrophoresis with a standard spin-column gel purification kit (Qiagen). 5 μg of the purified DNA was then digested with 0.5 U of DNAseI (NEB) in a 50 μl reaction. The reaction was allowed to sit for 7.5 minutes at room temperature and then quenched with 5 μl of 100 mM EDTA (4x the concentration of $MgCl_2$ in the reaction buffer). The reaction was further heat-inactivated for 10 minutes at 90ºC in a thermocycler and electrophoresed. Bands of ~30 bp, as compared to standards [30 bp oligo (IDT) / 100 bp DNA ladder (NEB)], were excised, frozen, and then purified using a Freeze 'N Squeeze gel purification kit (BioRad) because the small band size precluded

spin column purification. Purified digested fragments were mixed together at a 1:1 ratio and assembled via overlap-extension PCR.

Protein Expression and Library Screening. Single bacterial colonies were picked with sterile toothpicks and inoculated into 300 µl of Super Optimal Broth (SOB) supplemented with 100 µg/ml ampicillin in 2 ml deep-well 96-well plates (Seahorse Biosciences). The plates were sealed with microporous film (Denville Scientific) to facilitate gas exchange during growth. Cultures were grown overnight at 37 °C / 300 RPM. The next morning 800 µl of fresh SOB with 100 µg/ml ampicillin and 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a total volume of 1 ml (evaporation losses overnight are approximately 100 µl). Plates were then shaken 12 hours at either 30ºC or 37 °C and 400 RPM. After overnight expression, plates were screened with a liquid handling robot (Tecan Freedom Evo) linked to a platereader (Tecan Saffire 2). 200 µl of each culture was added to Greiner UV-Star 96-well plates and imaged for fluorescence emission at 675 nm after excitation at 600 nm. Controls were included on each plate to account for plate-to-plate variation. Potential hits were streaked out onto a fresh LB-Amp plate, grown overnight at 37ºC, and four colonies were picked for each potential hit. These were then grown again and screened as detailed above, with hits then ranked on their significant variation from the parent or control.

Protein Purification. To further characterize important variants, 1 L of SOB in Fernbach flasks was inoculated 1:100 with overnight cultures, grown to an OD of ~0.5 and induced at 37ºC for 12 hours with 1mM IPTG. The broth was then transferred to centrifuge flasks and spun at 5,000 x g in a fixed angle rotor for 10 min and the supernatant decanted. Bacterial pellets were resuspended in 25 ml of lysis buffer (50 mM sodium phosphate, 150 mM NaCl, 0.1% v/v Triton-X, pH 7.4) supplemented with 50 Units/ml Benzonase (Sigma) and 0.05 mg/ml Hen Egg Lysozyme (Sigma). Resuspended pellets were then run over a microfluidizer to fully lyse the bacteria. To pellet down the cellular debris, the lysed cultures were again centrifuged for 10 minutes at 15,000 x g in a fixed angle rotor. The colored supernatant was then poured through a column of His-Select resin (Sigma), washed twice (50 mM sodium phosphate, 150 mM NaCl, 15 mM Imidazole, pH 7.4), and eluted with 500 µl elution buffer (50 mM sodium phosphate, 150 mM NaCl, 250 mM

Imidazole, pH 7.4). Proteins were further purified by FPLC (AKTA) with a Superdex 75 10/300 column, and in the process buffer exchanged into PBS.

Fluorescent Protein Characterization. Purified protein variants were assayed in triplicate in Greiner UV-Star 96-well plates with a Tecan Saffire 2. An absorbance scan (260 – 650 nm), a fluorescence excitation scan (500 – 640 nm excitation / 675 nm emission), and a fluorescence emission scan (550 nm excitation / 575 – 800 nm emission) were run on 100 µl of eluted protein to determine spectral peaks.

To measure the quantum yield we diluted each protein so that the absorbance for 200 µl of protein at 540 nm was between 0.1 and 0.5. We then measured the A_{550} in triplicate (or duplicate if it was a poorly expressed protein), diluted the sample to an A_{550} of 0.04 and took an emission scan (540 nm excitation / 550 – 800 nm emission). The area under the emission curve was calculated after fitting it to a 4th order Gaussian, and the quantum yield was calculated with the following formula:

$$\Phi_X = (A_S / A_X)(F_X / F_S)(n_X / n_S)^2 \Phi_S$$

Where Φ is quantum yield, A is absorbance, F is total fluorescent emission (area under the curve), and n is the refractive index of the solvents used. Subscript X refers to the queried substance and subscript S refers to a standard of known quantum yield. It is important that the standard be excited with the same wavelength of light as the unknown sample. We use DsRed, which has a known quantum yield of 0.79 as the protein standard.

To measure extinction coefficient we took 100 µl of the protein solution that had been diluted to an A_{550} of between 0.1 and 0.5 and measured absorbance between 400 nm and 700 nm in triplicate. We then added 100 µl of 2M NaOH to each well and remeasured absorbance between 400 nm and 700 nm. The base-denatured chromophore, which peaks at approximately 450 nm has a known extinction coefficient of 44,000 $M^{-1}cm^{-1}$. Then to calculate the extinction coefficient is calculated with the following formula:

$$\varepsilon = A_{Chromophore} * 44{,}000 M^{-1} cm^{-1} / A_{450}$$

Thermal Stability. Purified proteins were diluted to an absorbance of 0.2 at the wavelength of maximum absorbance (λ_{abs}) so that their fluorescence would not saturate the rtPCR detector. 50 µl of each purified protein was then loaded into a 96-well PCR plate and covered with clear optical tape. The proteins were incubated at 37°C for 10 minutes and then the temperature was ramped at 0.5°C every 30 seconds up to 99°C, with fluorescence measured every ramp step in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). We refer to this as a thermal melt. The derivative curve of the thermal melt finds the inflection point of the slope, which is the apparent temperature at which fluorescence is irrevocably lost (apparent Tm).

Oligomeric Determination. (A) Size exclusion chromatography. 100 µl of each purified protein analyzed was run over a Superdex 75 10/300 size exclusion column with 25 ml bed volume on an AKTA from GE Life Sciences. Absorbance was measured after passage through the column at 575 nm, where the red chromophore absorbs. (B) Analytical ultracentrifugation. Purified protein samples were diluted to an A_{575} of 0.5 for a path-length of 1.25 cm. These samples were put into two-channel sedimentation velocity cuvettes with the blank channel containing PBS. Sedimentation velocity was run at 40,000 RPM overnight with full A_{575} scans collected with no pause between reads. Data was loaded into Sedfit and a c(m) distribution was run with default assumptions made for PBS buffer viscosity. After integration, the c(m) curve was exported to Excel. (C) Homo-FRET. 200 µl of each purified protein was diluted to an Absorbance of 0.1 to 0.5 at 530 nm in 96-well Greiner UV-Star plates. Polarization scans were then taken with excitation at 530 nm and emission at 610 nm in a Tecan Safire2 plate-reader. Rose Bengal was used as a standard to calculate the instrument G factor (mP = 349).

Crystallography. Rectangular plate crystals of HcRed7 grew in 7 days by the sitting-drop vapor diffusion method in 100 mM Bis-Tris pH 6.5 with 200 mM ammonium sulfate and 25% w/v PEG 3350. Crystals were flash frozen in 2-Methyl-2,4-pentanediol (MPD) and shipped to beamline 12-2 at the Stanford Synchrotron Radiation Lightsource, where a 1.63 Å data set was collected. Phases were obtained through molecular replacement using the crystal structure of HcRed (PDB ID 1YZW).

Following molecular replacement, model building and refinement were run with COOT and PHENIX.(42, 43) NCS restraints were applied to early refinement steps and removed at the final stages of refinement. TLS parameters were used throughout. The chromophore was initially left out of the refinement and added at a later stage when clear density became evident for it. Coordinates were deposited in the Protein Data Bank. Data collection and refinement statistics are listed in Table S1.

5.7 References

1. Tromberg BJ*, et al.* (1999) Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy. *Neoplasia (New York, N.Y.)* 2(1-2):26-40.
2. Bashkatov AN, Genina EA, Kochubey VI, & Tuchin VV (2005) Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *Journal of Physics D: Applied Physics* 38(15):2543.
3. Wang S, Moffitt JR, Dempsey GT, Xie XS, & Zhuang X (2014) Characterization and development of photoactivatable fluorescent proteins for single-molecule–based superresolution imaging. *Proceedings of the National Academy of Sciences of the United States of America*.
4. Wenqin W, Gene-Wei L, Chongyi C, Xie XS, & Xiaowei Z (2011) Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria. *Science* 333(6048):1445-1449.
5. Alieva NO*, et al.* (2007) Diversity and evolution of coral fluorescent proteins. *PloS ONE* 3(7).
6. Shagin DA*, et al.* (2004) GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Molecular Biology and Evolution* 21(5):841-850.
7. Bou-Abdallah F, Chasteen ND, & Lesser MP (2006) Quenching of superoxide radicals by green fluorescent protein. *Biochimica et Biophysica Acta (…*.
8. Palmer CV, Modi CK, & Mydlarz LD (2009) Coral Fluorescent Proteins as Antioxidants. *PLoS ONE* 4(10).
9. Salih A, Larkum A, Cox G, Kühl M, & Hoegh-Guldberg O (2000) Fluorescent pigments in corals are photoprotective. *Nature* 408(6814):850-853.
10. Smith EG, D'Angelo C, Salih A, & Wiedenmann J (2013) Screening by coral green fluorescent protein (GFP)-like chromoproteins supports a role in photoprotection of zooxanthellae. *Coral reefs*.
11. Campbell RE*, et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*.

12. Kredel S*, et al.* (2008) mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PloS ONE* 4(2).
13. Shemiakina II*, et al.* (2012) A monomeric red fluorescent protein with low cytotoxicity. *Nature Communications*.
14. Takako K*, et al.* (2006) A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. *Nature Biotechnology* 24(5):577-581.
15. Gurskaya NG*, et al.* (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins1. *FEBS Letters* 507(1):16-20.
16. Subach OM*, et al.* (2008) Conversion of Red Fluorescent Protein into a Bright Blue Probe. *Chemistry & Biology* 15(10):1116-1124.
17. Wannier TM, Moore MM, Mou Y, & Mayo SL (2015) Computational Design of the β-Sheet Surface of a Red Fluorescent Protein Allows Control of Protein Oligomerization. *PloS one* 10(6).
18. Wall MA, Socolich M, & Ranganathan R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature Structural Biology* 7(12):1133-1138.
19. Wannier TM & Mayo SL (2014) The structure of a far-red fluorescent protein, AQ143, shows evidence in support of reported red-shifting chromophore interactions. *Protein Science* 23(8):11481153.
20. Mudalige K*, et al.* (2010) Photophysics of the red chromophore of HcRed: evidence for cis-trans isomerization and protonation-state changes. *The journal of physical chemistry. B* 114(13):4678-4685.
21. Moore MM, Oteng-Pabi SK, Pandelieva AT, Mayo SL, & Chica RA (2012) Recovery of Red Fluorescent Protein Chromophore Maturation Deficiency through Rational Design. *PLoS ONE* 7(12).
22. Piatkevich KD*, et al.* (2012) Extended Stokes shift in fluorescent proteins: chromophore-protein interactions in a near-infrared TagRFP675 variant. *Scientific reports* 3:1847.

23. Bloom JD, Labthavikul ST, Otey CR, & Arnold FH (2006) Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America* 103(15):5869-5874.

24. Jäckel C, Bloom JD, Kast P, Arnold FH, & Hilvert D (2010) Consensus protein design without phylogenetic bias. *Journal of molecular biology* 399(4):541-546.

25. Pletneva NV*, et al.* (2011) Crystallographic study of red fluorescent protein eqFP578 and its far-red variant Katushka reveals opposite pH-induced isomerization of chromophore. *Protein Science* 20(7):1265-1274.

26. Crameri A, Whitehorn EA, Tate E, & Stemmer WP (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature biotechnology* 14(3):315-319.

27. Zhao H & Arnold FH (1997) Optimization of DNA Shuffling for High Fidelity Recombination. *Nucleic Acids Research*.

28. Regmi C, K. , Bhandari YR, Gerstman BS, & Chapagain P, P. (2013) Exploring the Diffusion of Molecular Oxygen in the Red Fluorescent Protein mCherry Using Explicit Oxygen Molecular Dynamics Simulations. *The journal of physical chemistry. B*.

29. Black SD & Mould DR (1991) Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Analytical Biochemistry* 193(1):72-82.

30. Chennamsetty N, Voynov V, Kayser V, Helk B, & Trout BL (2010) Prediction of aggregation prone regions of therapeutic proteins. *The journal of physical chemistry. B* 114(19):6614-6624.

31. Goedhart J*, et al.* (2011) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature communications* 3:751.

32. Lelimousin M*, et al.* (2009) Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield. *Biochemistry* 48(42):10038-10046.

33. Rurack K & Spieles M (2011) Fluorescence quantum yields of a series of red and near-infrared dyes emitting at 600-1000 nm. *Analytical chemistry* 83(4):1232-1242.

34. Chica RA, Moore MM, Allen BD, & Mayo SL (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences of the United States of America* 107(47):20257-20262.

35. Morozova KS*, et al.* (2010) Far-red fluorescent protein excitable with red lasers for flow cytometry and superresolution STED nanoscopy. *Biophysical journal* 99(2):5.

36. Grigorenko BL, Nemukhin AV, Polyakov IV, & Krylov AI (2013) Triple-decker motif for red-shifted fluorescent protein mutants. *The Journal of Physical Chemistry Letters* 4(10):1743-1747.

37. Strack RL, Strongin DE, Mets L, Glick BS, & Keenan RJ (2010) Chromophore formation in DsRed occurs by a branched pathway. *Journal of the American Chemical Society* 132(24):8496-8505.

38. Tubbs JL, Tainer JA, & Getzoff ED (2005) Crystallographic Structures ofDiscosomaRed Fluorescent Protein with Immature and Mature Chromophores: Linking Peptide BondTrans−CisIsomerization and Acylimine Formation in Chromophore Maturation†,‡. *Biochemistry* 44(29):9833-9840.

39. Strongin DE*, et al.* (2007) Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering, Design & Selection* 20(11):525-534.

40. Henderson JN*, et al.* (2009) Excited State Proton Transfer in the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* 131(37):13212-13213.

41. Kredel S*, et al.* (2009) mRuby, a Bright Monomeric Red Fluorescent Protein for Labeling of Subcellular Structures. *PLoS ONE* 4(2).

42. Adams P*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):213-221.

43. Emsley P, Lohkamp B, Scott W, & Cowtan K (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 4):486-501.

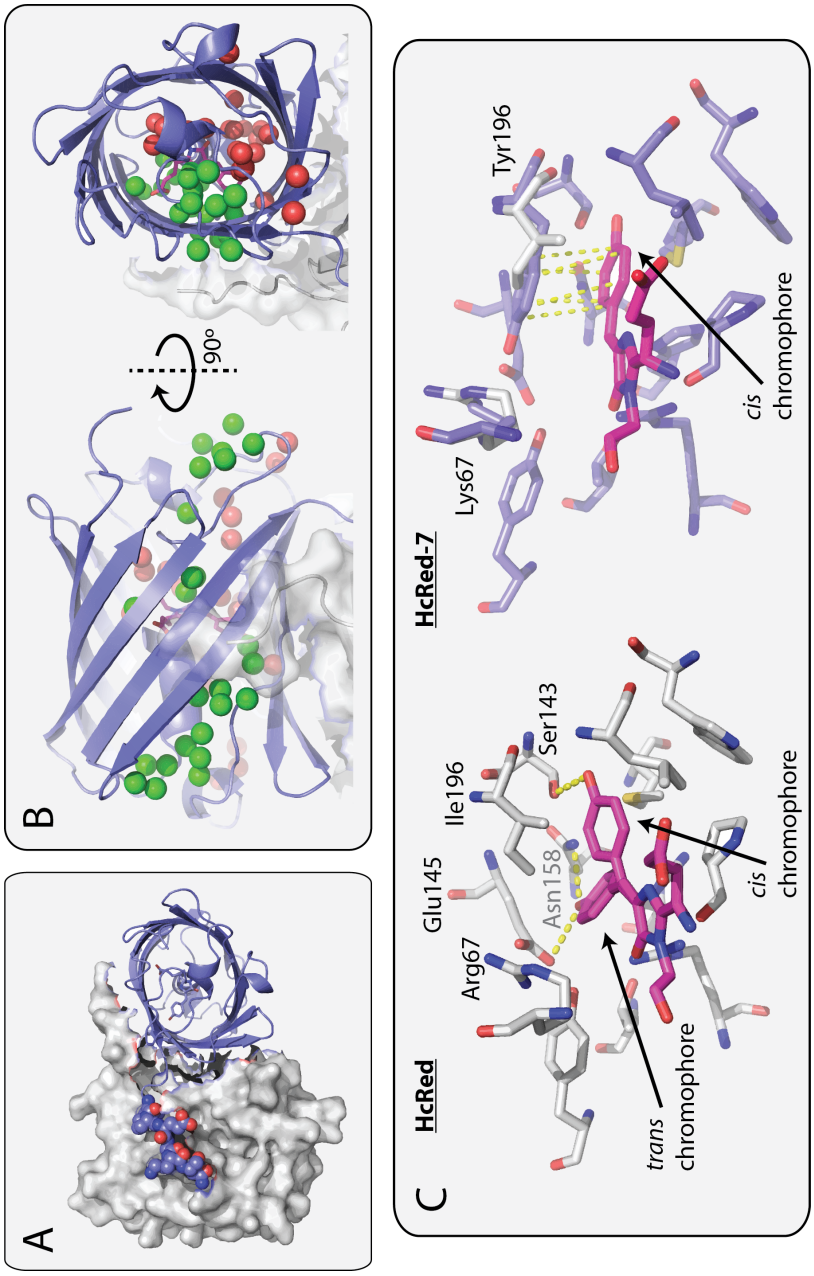

Figure 5.1. The structure of HcRed7, a far-red dimer.

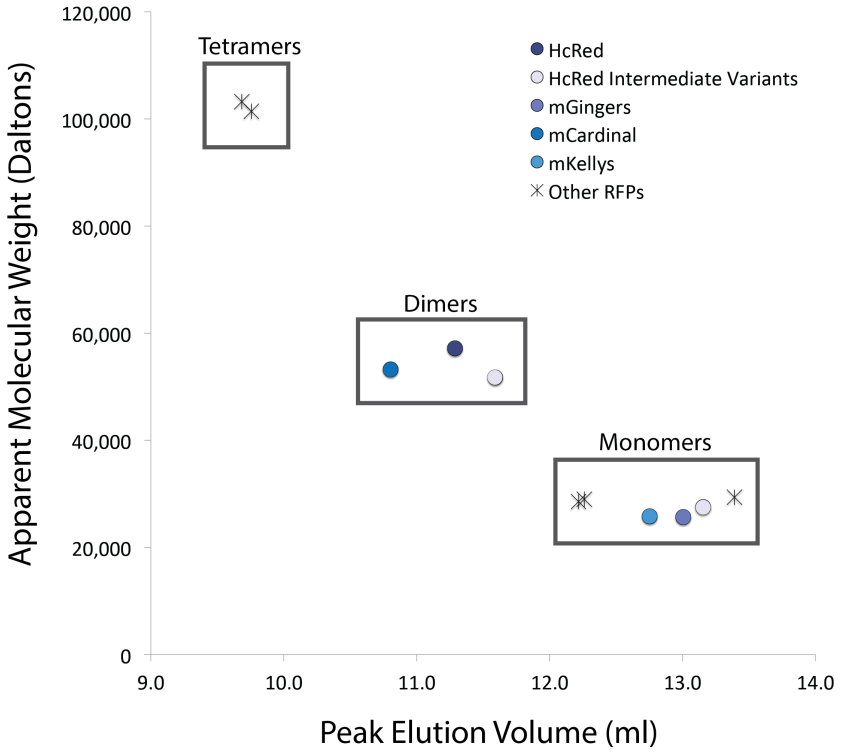

Figure 5.2. Oligomeric analysis of RFPs. The apparent molecular weight as calculated from a c(M) distribution of sedimentation velocity data from an analytical ultracentrifuge run is plotted on the Y-axis. The X-axis shows the peak elution volume as measured at 590 nm absorbance by size exclusion chromatography. Clear groupings are boxed as monomers, dimers, and tetramers.

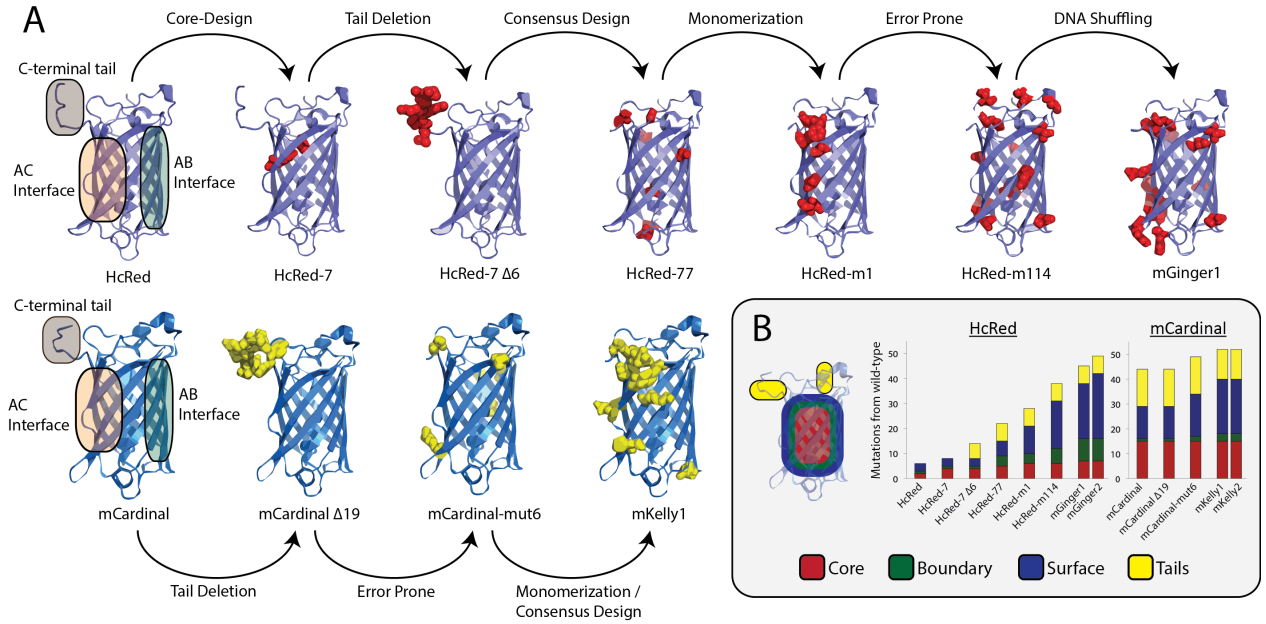

Figure 5.3. The engineering of the mGingers and mKellys. (A) Each engineering step is shown with the mutations made shown in red spheres in the HcRed background and yellow spheres in the mCardinal background. The crystal structures are of HcRed7 and mCardinal (B) Solvent-exposed surface area (SASA) was used to categorize the mutations into three buckets, with tails treated separately. The number of mutations separating each designed variant from its wild-type ancestral progenitor.

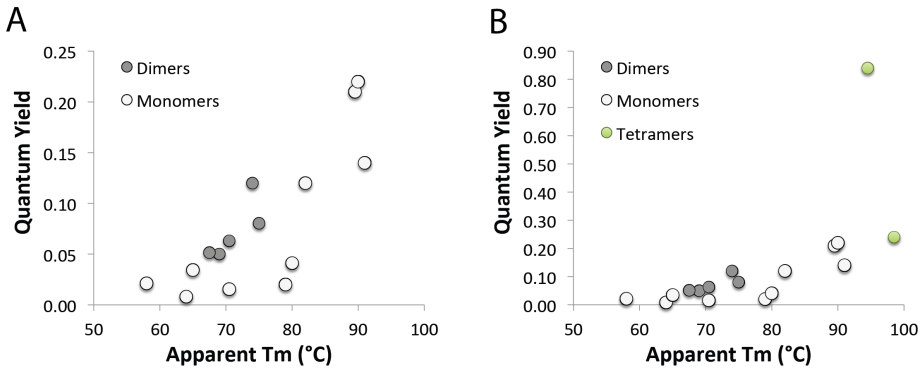

Figure 5.4. Thermal stability of HcRed, mCardinal, and DsRed variants are plotted against their quantum yields. The correlation can be clearly seen with monomers exhibiting lower quantum yield than dimers at equivalent wavelengths (A) and tetrameric proteins clearly exhibiting the highest quantum yields (B).

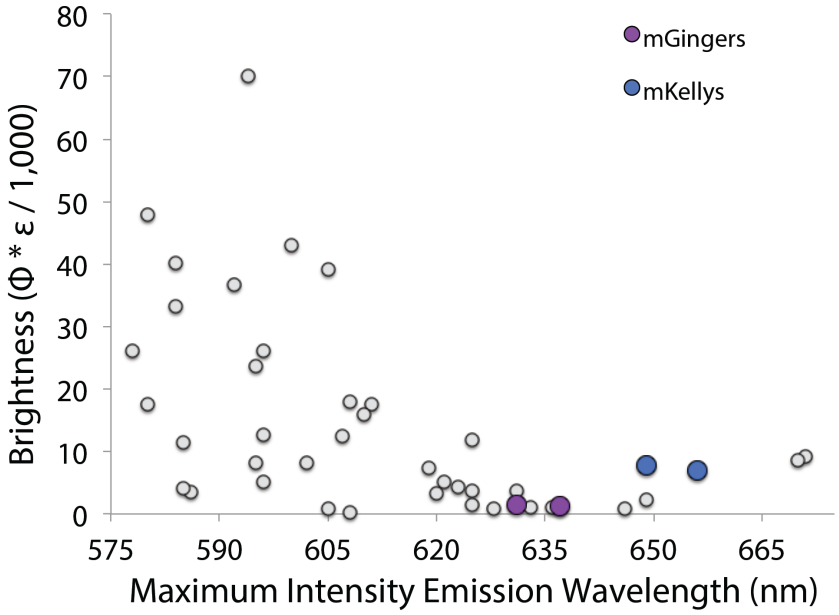

Figure 5.5. Brightness versus emission maximum for monomer RFPs. There is a negative correlation between brightness and λ_{em} among RFPs. All known monomeric RFPs whose brightness and λ_{em} have been measured are plotted. The mGingers and mKellys are among the furthest red-shifted monomeric proteins described to date.

Monomeric RFP	Brightness (Φ x ε) / 1000	λ_{em} (nm)	Mutations to Core	Total Mutations	Immediate Parent (dimer/tetramer)	Brightness (Φ x ε) / 1000	λ_{em} (nm)	Total Mutations
mRFP1	12.5	607	13	33	DsRed (T)	59.3	583	n/a
DsRed.M1	3.5	586	10	45	"	"	"	"
FusionRed	18.0	608	9	45	mKate2 (D)	18.0	630	27
mRuby	39.2	605	6	40	eqFP611 (T)	35.1	611	n/a
mKeima	3.5	620	7	17	dKeima (D)	7.6	616	13
mGinger1	1.2	637	7	45	HcRed-7 (D)	4.8	643	8
mGinger2	1.5	631	7	49	"	"	"	"
mKelly1	5.7	656	15	52	mCardinal (D)	9.6	658	44
mKelly2	6.5	649	15	52	"	"	"	"

Ancestral Parent (dimer/tetramer)	Brightness (Φ x ε) / 1000	λ_{em} (nm)
DsRed (T)	59.3	583
"	"	"
eqFP578 (T)	55.1	578
eqFP611 (T)	35.1	611
COCP (T)	--	--
hCriCP	--	--
"	--	--
eqFP578 (T)	55.1	578
"	"	"

Table 5.1. Previously monomerized first generation RFPs.

RFP	Φ	ε (M^{-1} cm^{-1}) / 1000	Brightness (Φ x ε) / 1000	λ_{ex} (nm)	λ_{em} (nm)	Apparent T_m (ºC)
HcRed	0.05	70	6.0	585	633	69.0
HcRed-7	0.08	75	8.4	592	645	75.0
HcRed-7 Δ5	0.06	69	4.3	592	643	70.5
HcRed-7 Δ6				582	635	65.0
HcRed-77	0.05	†	†			67.5
HcRed-m1	0.01	†	†			64.0
HcRed-m114	0.02	†	†			58.0
mGinger1	0.02	58	1.2	587	637	79.0
mGinger2	0.04	36	1.5	578	631	80.0
mCardinal	0.12	80	9.6	601	658	
mCardinalΔ20						
mCardinal-mut6	0.13	60	7.8			
mKelly1	0.13	44	5.7	596	656	
mKelly2	0.15	43	6.5	598	649	

† - Extinction coefficient (and therefore brightness) could not be measured because of multiple chromophore species present.

Table 5.2. Photophysical properties of protein variants derived from HcRed and mCardinal.

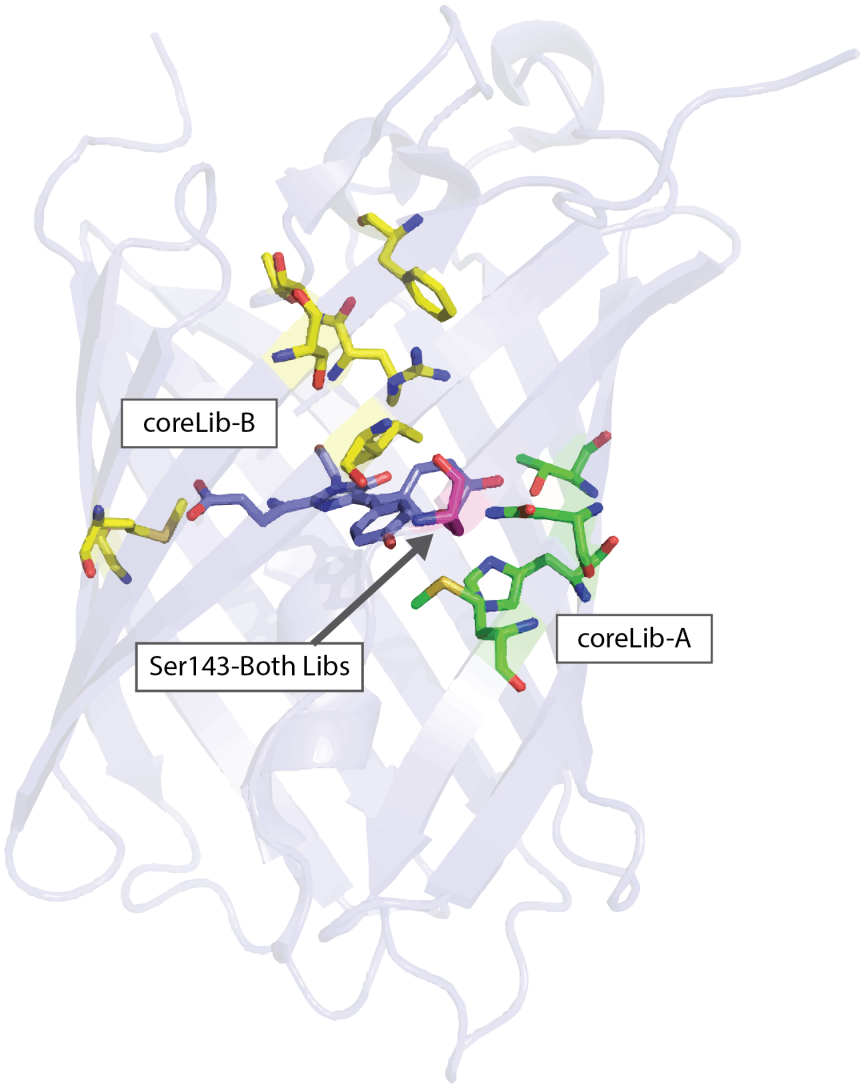

Figure S5.1. The design of HcRed7. Two core libraries targeted unique structural regions of the protein core. The first region (green) surrounds the phenolate side chain of the chromophore. The second (yellow), is a very highly mutated region in RFP monomer evolution. This region holds an internal water channel, key catalytic residues, and abuts the AC interface.

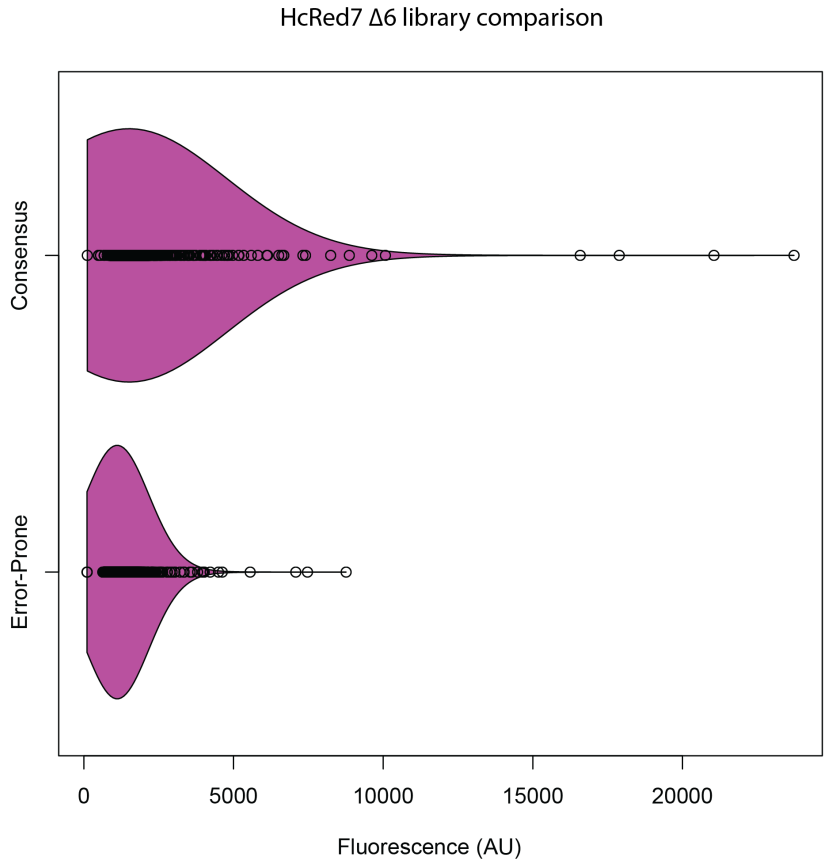

Figure S5.2. Consensus design compares favorably to error-prone mutagenesis in the improvement of fluorescence in HcRed7 Δ6. The two libraries were compared by screening in 96-well plates with ~4,000 screened variants from the consensus design library and ~8,000 variants screened from the error-prone library. Both individual variants and the population from the consensus library outperform the error-prone library, although both contributed valuable variants to the engineering process.

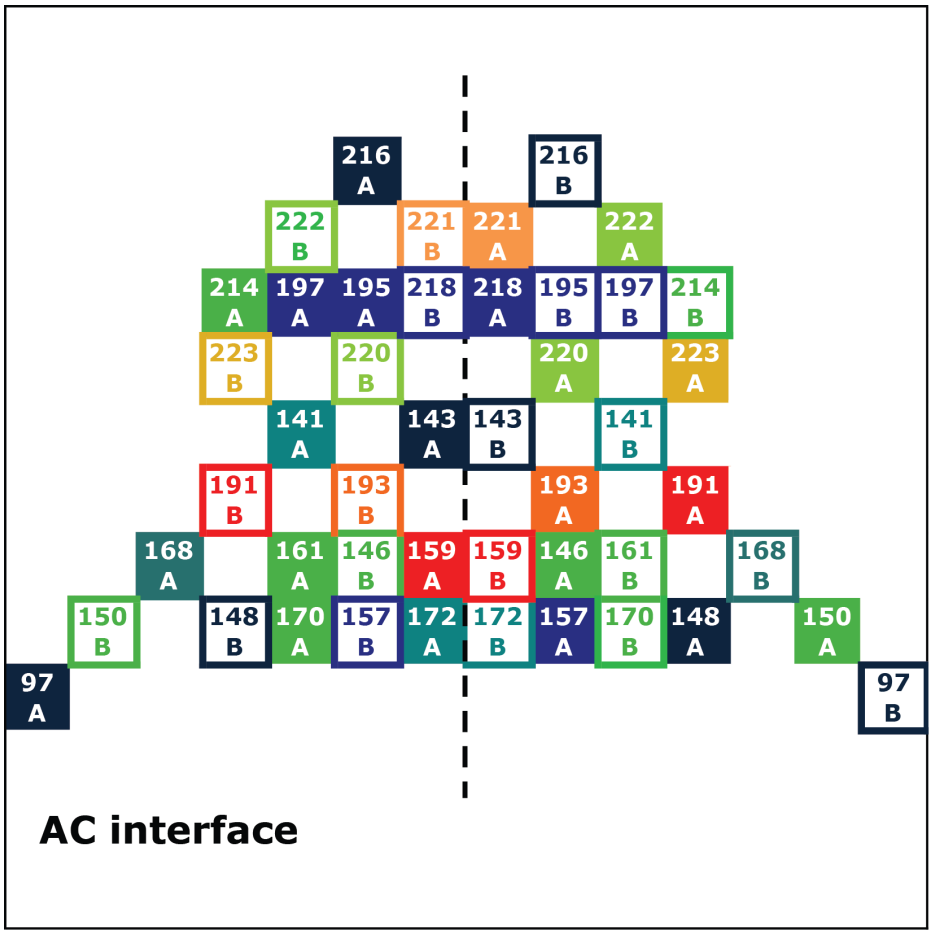

Figure S5.3. A map displaying the intermolecular contacts made by each residue of HcRed in the AC interface. The map is color coded by the frequency with which the residues are mutated during past instances of fluorescent protein monomerization. Residues 146, 159, 167, 168, 170, 174, 191, 193, 197, 201, and 214 were mutated during mGinger engineering. Residues 222-223 were deleted as part of the C-terminal tail.

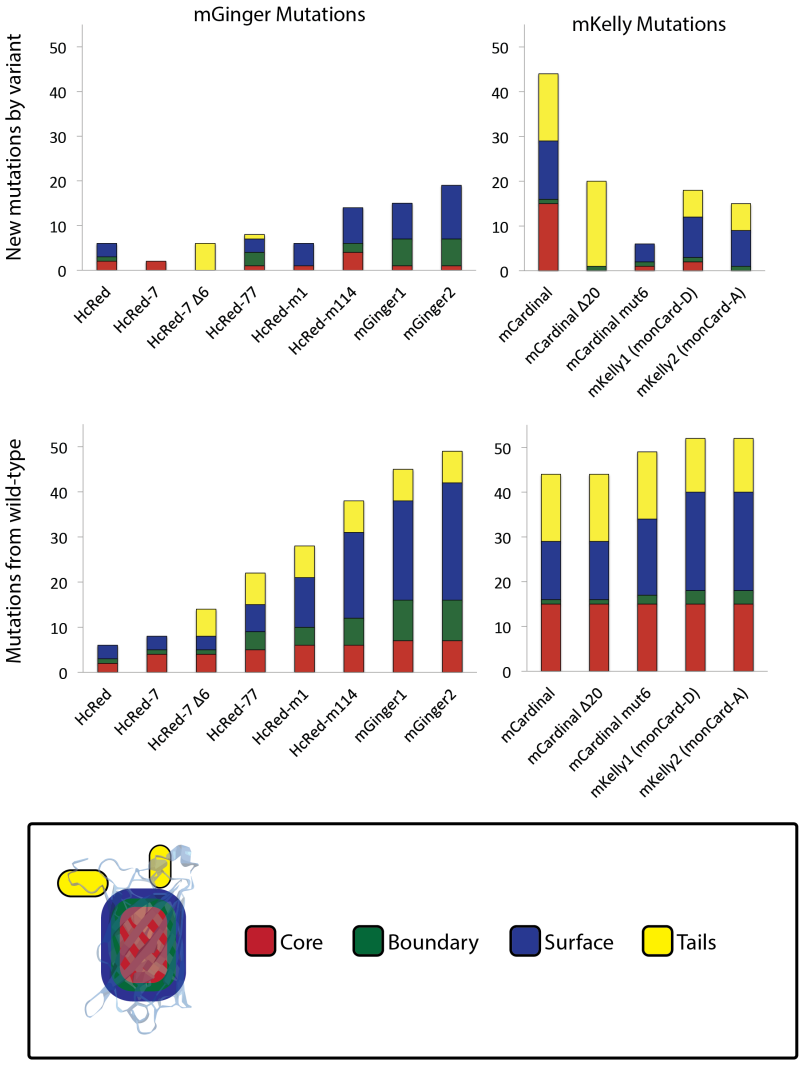

Figure S5.4. Mutations during mGinger and mKelly engineering categorized by solvent-accessible surface area (SASA). (A) The number of new mutations introduced in each step of protein engineering. (B) The cumulative number of mutations that separate each variant from the wild-type progenitor (HcCP in the case of HcRed variants and eqFP578 in the case of mCardinal variants).

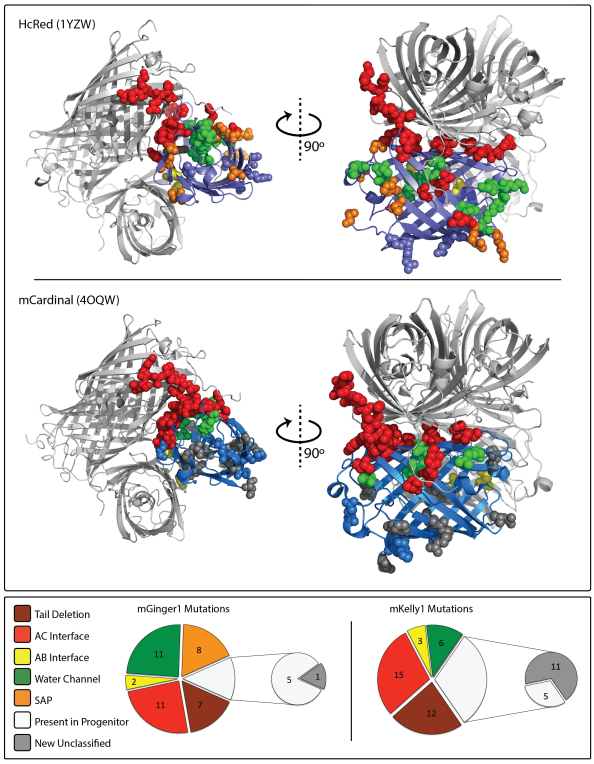

Figure S5.5. Mutations during mGinger and mKelly engineering categorized by solvent-accessible surface area (SASA). (A) The number of new mutations introduced in each step of protein engineering. (B) The cumulative number of mutations that separate each variant from the wild-type progenitor (HcCP in the case of HcRed variants and eqFP578 in the case of mCardinal variants).