Published in final edited form as:
*Biochemistry*. 2012 March 27; 51(12): 2471–2485. doi:10.1021/bi201422g.

# A Variable Light Domain Fluorogen Activating Protein Homodimerizes to Activate Dimethylindole Red†

**Nina Senutovitch**[1], **Robyn L. Stanfield**[3], **Shantanu Bhattacharyya**[1], **Gordon S. Rule**[*,1,2], **Ian A. Wilson**[3,4], **Bruce A. Armitage**[5,2], **Alan S. Waggoner**[1,2], and **Peter B. Berget**[§,1,2]

[1]The Department of Biological Sciences, Carnegie Mellon University Pittsburgh, PA 15213, USA

[2]The Molecular Biosensor and Imaging Center, Carnegie Mellon University Pittsburgh, PA 15213, USA

[3]The Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[4]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[5]The Department of Chemistry, Carnegie Mellon University Pittsburgh, PA 15213, USA

## Abstract

Novel fluorescent tools such as green fluorescent protein analogs and Fluorogen Activating Proteins (FAPs) are useful in biological imaging to track protein dynamics in real-time with low fluorescence background. FAPs are single-chain variable fragments (scFvs) selected from a yeast surface display library that produce fluorescence upon binding a specific dye or fluorogen that is normally not fluorescent when present in solution. FAPs generally consist of human immunoglobulin variable heavy ($V_H$) and variable light ($V_L$) domains covalently attached via a glycine and serine rich linker. Previously, we determined that the yeast surface clone, $V_H$-$V_L$ M8, could bind and activate the fluorogen dimethylindole red (DIR), but that the fluorogen activation properties were localized to the M8$V_L$ domain. We report here that both NMR and X-ray diffraction methods indicate the M8$V_L$ forms non-covalent, anti-parallel homodimers that are the fluorogen activating species. The M8$V_L$ homodimers activate DIR by restriction of internal rotation of the bound dye. These structural results, together with directed evolution experiments of both $V_H$-$V_L$ M8 and M8$V_L$, led us to rationally design tandem, covalent homodimers of M8$V_L$ domains joined by a flexible linker that have a high affinity for DIR and good quantum yield.

The development of fluorescent technologies has revolutionized cellular imaging and molecular biology, and the utility of genetically encoded fluorescent proteins, such as green fluorescent protein (GFP), for the detection of particular proteins of interest is well

†This work was supported by NIH grant U54 RR022241 and Howard Hughes Medical Institute grant 52005865. Portions of this work were carried out at GM/CA CAT at the Advanced Photon Source and the Molecular Biology Consortium at the Advanced Light Source. GM/CA CAT has been funded, in whole or in part, from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). Use of the Advanced Photon Source was supported by the U.S. Department of Energy (DE-AC02-06CH1135). The Advanced Light Source is supported by the U.S. Department of Energy (DE-AC02-05CH11231).

*Corresponding Author: 4400 Fifth Ave Carnegie Mellon University Mellon Institute Rm 751 Pittsburgh PA 15213 rule@andrew.cmu.edu Fax: 412-268-7129 Phone: 412-268-1839 .
§Peter B. Berget current address: 600 S. 43rd St University of the Sciences McNeil Science &Technology Center STC 342 Philadelphia PA 19104 p.berget@usciences.edu Fax: 215-596-8710 Phone: 215-596-7417 berget@andrew.cmu.edu

Accession Codes X-ray structure coordinates and structure factors for M8$V_L$, M8$V_L$-DIR, and M8$V_L$S$^{L55}$P-DIR have been deposited to the Protein Data Bank with codes 3T0V, 3T0W, and 3T0X, respectively.

Supporting Information Available Supporting Information available for sequence information, absorbance spectra and details of crystallography data. This information is free of charge from the ACS website via the Internet at http://pubs.acs.org.

documented (1). There is still a need for additional, well-characterized tools that provide real-time, high signal-to-noise fluorescence and demonstrate high fluorescence quantum yield ($\phi_f$), photo-stability, and a broad spectral range. Fluorogen Activating Proteins (FAPs) are part of novel, immunoglobulin-based, fluoromodule platforms that induce fluorescence emission of cognate fluorogenic dyes in solution (2). FAPs cause a dramatic increase in the $\phi_f$ or fluorescence enhancement of the fluorogenic dyes that they bind. These fluoromodules have enhanced photo-stability due to exchange of bleached dye. Their fluorescence emission spans the visible spectrum from blue to far red, comparable to other fluorescence proteins, and often a single FAP can activate multiple dyes (3) (4) (5). The fluorogenic dyes bound by FAPs have low fluorescence background in aqueous solution and increase in fluorescence as much as two-thousand fold upon interaction with a cognate FAP protein. Previous studies give some insight into the generation of fluorescence from fluorogens, showing that these compounds become fluorescent when the rotation of the aromatic functional groups are restricted by a binding partner, such as upon intercalation into DNA (6).

Several FAPs were isolated that activate the red (640nm) emitting fluorogenic dye, dimethylindole red (DIR) (3). These FAPs are isolated from a naïve human IgG single chain Fv (scFv) library created in a yeast surface display vector typically consisting of IgG variable heavy ($V_H$) and light ($V_L$) chain domains covalently connected by a flexible linker comprised of serine and glycine repeat sequences (7) (8) (9). Two of these FAPs, named $V_H$-$V_L$ M8 and $V_H$-$V_L$ K10, were unusual in that they both contained identical $V_L$ domains but different $V_H$ domains. Based on the sequence similarity between M8 and K10 it was proposed, and demonstrated experimentally, that the $V_L$ domain alone was sufficient in the yeast surface display format to bind and activate DIR (10).

The $V_L$ domain of M8 provides the opportunity to investigate the potential for optimization of DIR fluoromodules. Improvements in the $\phi_f$ and the dye affinity of fluoromodules is desirable, so that they will produce stronger fluorescence intensities for light microscopy at lower dye concentrations. Previously isolated FAPs have $\phi_f$ that compare favorably to other fluorescent proteins, yet the extent by which DIR-activating FAPs can be improved in their $\phi_f$ has not yet been determined (2). In order to improve the characteristics of FAP-fluorogen pairs, FAP genes can be subjected to directed evolution, utilizing PCR-based mutagenesis, transformation back into yeast and selection by a fluorescence activated cell sorter (FACS) (11) (12). Because the $V_H$-$V_L$ M8 FAP is active in both the original isolated $V_H$-$V_L$ (scFv) format, as well as an isolated $V_L$ domain ($M8V_L$), we were able to compare the directed evolution of this FAP in the two different formats. The aim of this study was to better understand the mechanism for DIR activation by the $M8V_L$ FAP, in particular determination of the conformational restraints placed on DIR. Following the directed evolution experiments, rational design of the linker region and structural analysis of the active form of the $M8V_L$ FAP were undertaken to more fully understand this unusual interaction and the mechanism by which the $M8V_L$ FAP can activate DIR. This study describes the first structural data for FAP-induced fluorescence activation of the environmentally-sensitive fluorogen DIR.

## Experimental Procedures

### Directed Evolution

Directed evolution of $V_H$-$V_L$ M8 and$M8V_L$ genes was accomplished through error-prone PCR, homologous recombination and FACS enrichment as previously described (2) (11). Three rounds of FACS enrichment of a library of mutants with a $10^6$ diversity was performed at 250 pM DIR for $V_H$-$V_L$ M8 and 1 nM DIR for the $M8V_L$ domain. At the final round of sorting, cells were autocloned onto induction plates and visually screened for fluorescence enhancement (3). Further detail regarding the directed evolution enrichments

for $M8V_HV_L$ and $M8V_L$ is available in the Supporting Information Figure S1. Plasmid DNA from individual clones was isolated using the Zymoprep Yeast Plasmid Miniprep Kit II (Zymogen Inc) and transformed into bacterial MachT1 cells (Invitrogen). Bacterial mini prep DNA was isolated (Fermentas) and the altered genes were commercially sequenced by Retrogen.

### Affinity, fluorescent enhancement and quantum yield determination

The dissociation constant ($K_D$) of purified soluble FAPs and yeast cell surface displayed FAPs for DIR were determined by titrations of DIR into samples as previously described (3). The $K_D$ values for soluble monomeric FAPS (e.g. $V_H$-$V_L$) was obtained by fitting the binding data to the standard quadratic binding equation (eq. 1) using in-house software that finds both the global and local minima in $\chi^2$.

$$F_{obs}=F_{Max}\frac{([P_T]+[L_T]+K_d])-\sqrt[2]{([P_T]+[L_T]+K_d])^2-4[P_T][L_T]}}{2[P_T]} \quad (1)$$

$[P_T]$ is the total concentration of the protein, and $[L_T]$ is the total concentration of the dye. Fitting synthetic data sets indicated that it is possible to distinguish $K_D$ values as low as 0.1 nM with the protein concentrations used in this study (see supporting information S11A and S11B). In the case of $V_L$ FAPs that dimerize in solution due to the addition of DIR ($M8V_L$, $M8V_LS^{L55}P$, Q9) a number of different binding models are possible: i) formation of $V_L$-$V_L$ dimers followed by dye binding, or ii) dye binding to one $V_L$ monomer, followed by the addition of the second $V_L$ domain. The second model was selected based on NMR data that showed that both $M8V_L$ and $M8V_LS^{L55}P$ are monomeric in solution at concentrations of ~500 μM. Consequently, the binding curves were fit to the model illustrated in Scheme 1, using the equations provided by Mack *et al* (13). In this case, the concentration of the intermediate species $V_L$-DIR was assumed to be small, based on NMR studies with both M8V and $M8V_LS^{L55}P$. Consequently α was set at $10^7$ to ensure that $[(V_L)_2DIR] > [V_L\text{-}DIR]$. The overall dissociation constant for the reaction, $(K_D)^2/\alpha$, was insensitive to the choice of α for values of $10^5$ or higher. Note that the overall dissociation constant for this reaction reports on both the dye-protein interaction ($K_D$) and the interaction of the free protein with the protein-dye complex ($K_D/\alpha$). Since we cannot determine α independently, we report the overall affinity, $(K_D)^2/\alpha$. For both models, the fitted parameters were the $K_D$, the fluorescence at saturation, $F_{Max}$, and the total protein concentration $[P_T]$. In the case of $M8V_L$, a plot of $\chi^2$ versus $[P_T]$ yielded a number of essentially equivalent local minima, in which case the minimum closest (within 10%) to the experimentally measured protein concentration was selected. Errors in $K_D$ were obtained by generating random data sets, assuming the error for each data point followed a normal distribution, and then determining the distribution of the resultant $K_D$ values. Because of the high affinity associated with the synthetic tandem dimers it was not possible to compare the effect of the serine to proline mutation ($M8V_L$ versus $M8V_LS^{L55}P$) on the affinity of DIR when these FAPs were in the dimer form. However, a quantitative comparison was possible using the isolated $V_L$ domains because the apparent affinity can be controlled by the judicious choice of the protein concentration.

Binding data for FAPs on the yeast surface were fit using as single-site binding model, assuming that the concentration of free DIR equaled the total amount of DIR, e.g.

$$F_{obs}=F_{Max}\frac{[L_T]}{[L_T]+K_d} \quad (2)$$

A preliminary FACS screen was run prior to all titration assays to determine levels of cell surface FAP expression and non-viable yeast cells on all FAP constructs, including uninduced control samples. Fluorescence enhancement measurements were performed as described previously (2).

Quantum yields were determined as previously described (5) utilizing two cross-calibrated reference standards Cy5.18 and Cresyl Violet based on the spectral overlap with DIR, and data analyzed with Origin or GraphPad (14).

## Cloning and Protein Expression and Purification

$M8V_L$ tandem dimer genes were constructed using the wild-type gene and a synthetic, *E. coli* codon-optimized, $M8V_L$ gene (DNA 2.0) that altered the nucleotide sequence but not the aminoacid sequence, in order to avoid homologous recombination of the duplicated genes. The synthetic $M8V_L$ gene was ligated into the 5′ end of the wild type $M8V_L$ using the NheI and BamHI restriction sites in the pPNL6 plasmid containing the wild-type $M8V_L$ gene. Clones were sequenced to confirm presence of the 5′ synthetic gene and the $(G_4S)_3$ linker region followed by the 3′ wild type gene.

Tandem dimers of $M8V_L$ with longer $(G_4S)$ repeats in the linker were constructed by inserting oligomers of differing lengths into the BamHI restriction site at the beginning of the $(G_4S)$ linker of pPNL6-$M8V_L$ tandem dimer described above. The oligomer pair for adding one $(G_4S)$ repeat was (5′GATCAGGTGGCGGTGGCAGCA3′) and (5′GATCTGCTGCCACCGCCACCT3′). The oligomer pair for inserting three $(G_4S)$ repeats was
(5′GATCAGGTGGCGGTGGCAGCGGCGGTGGTGGTTCCGGAGGCGGCGGTTCTA3′ ) and
(5′GATCTAGAACCGCCGCCTCCGGAACCACCACCGCCGCTGCCACCGCCACCT3′). Complementary oligomers were annealed by incubating the oligomer for 5 minutes at 95°C and then cooling slowly to 23°C prior to ligation in buffer (50mM potassium acetate, 20mM Tris-acetate pH 7.9, 10mM magnesium acetate, 1mM dithiothreitol). DNA sequencing confirmed insertion of the extended linker oligomers.

All genes were PCR amplified with primers containing non-identical SfiI restriction sites that are compatible with the SfiI sites in the hexahistidine containing pAK400 *E. coli* periplasmic expression vector (gift from A. Plückthun). These primers are (5′GGCCCAGCCGGCCATGGCGGGTTCTGCTAGCCAGCCTGTGC3′) and (5′GGCCCCCCAGGCCGCTAGGACGGTGACCTTGGTCC3′). M8V , a mutant $M8V_LS^{L55}P$ and all tandem dimer SfiI-tailed PCR products were blunt cloned into the pJET plasmid (Fermentas), sequenced, and then inserted into pAK400 via SfiI sites. Selected genes were also cloned into pPNL9 yeast secretion plasmid by gap repair and transformation into the yeast strain YVH10 as previously described (2).

M8V and $M8V_LS^{L55}LP$ proteins were produced in milligram quantities for crystallographic studies from the pAK400 vector in *E. coli* MachT1 (Invitrogen) cells. Typically, 3 grams of cells from 0.5L of culture were lysed in 20mLs of buffer (50mM Tris-Cl pH 7.5, 750mM NaCl, 0.1% Triton X100 and 0.05% Tween-20) for 10 minutes in a cell homogenizer (Avestin EmulsiFlex-C3). The lysate was centrifuged for 30 minutes at 28,000g and the supernatant fraction batch bound to nickel agarose resin (ThermoFisher). The nickel resin was washed in ten column volumes of buffer (20mM imidazole, 10mM Tris-Cl pH 8.0, 50mM $KH_2PO_4$ and 300mM NaCl), followed by ten column volumes of buffer (20mM imidazole, 10mM Tris-Cl pH 8.0, 50mM $KH_2PO_4$ and 750mM NaCl) and two column volumes of buffer (50mM imidazole, 10mM Tris-Cl pH 8.0, 50mM $KH_2PO_4$ and 300mM NaCl). The protein was recovered in elution buffer (250mM imidazole, 10mM Tris-Cl pH

8.0, 50mM $KH_2PO_4$ and 300mM NaCl). Proteins were further purified by ion exchange chromatography on a HiTrap XP XL column (GE Healthcare) following dialysis into 50mM MES pH 6.0 and 50mM NaCl. A linear gradient of NaCl from 50mM to 1M was used to elute the protein from the ion exchange column. Protein preparations were analyzed by SDS-PAGE gel electrophoresis to determine purity. $M8V_L$ was further purified by size exclusion chromatography on a Superdex 200, 10/30 column in 0.2M Tris pH 7.5, 0.15M NaCl.

To allow labeling with $^{13}C$ for NMR experiments the $M8V_L$ gene was transferred from the pAK400-$M8V_L$ vector, using flanking NdeI and HindIII sites (enzymes from New England Biolabs), into pET22b+ plasmid (Novagen). Studier's minimal media was used to produce labeled $M8V_L$ (50 mM $Na_2HPO_4$, 50 mM $KH_2HPO_4$, 25 mM $(^{15}NH_4)_2SO_4$, 0.5% glucose, 1X Trace metals, 2mM $MgSO_4$ ) (15). The $^{15}N$ source was $(^{15}NH_4)_2SO_4$ and the $^{13}C$ source was glucose, both from Cambridge Isotope Laboratories. The labeled protein was purified essentially as described above.

## Crystallization, data collection, and structure determination

Crystal structures were determined for $M8V_L$, $M8V_L$ with DIR and $M8V_LS^{L55}P$ with DIR. Data collection and refinement statistics for the three structures are summarized in Table 3. For all three structures, data were processed with HKL-2000 (16), phases were determined by molecular replacement using Phaser (17), and model building and refinement were carried out with Coot (18), Refmac5 (19), and Phenix (20). The coordinates are numbered by the Kabat convention (21).

The $M8V_L$ was co-crystallized with DIR by mixing $M8V_L$ (15 mg/ml in 0.2M Tris, 0.15M NaCl, pH 7.5) with a 5:1 molar excess of DIR (10 mg/ml; in 20% DMSO, 0.2M Tris, 0.15M NaCl, pH 7.5). Initial crystallization conditions were identified by screening 384 different solutions with the Topaz System (Fluidigm) (22) and the crystal used for data collection was grown at 22°C in a sitting drop with a well solution of 35% PEG 1500. The crystal was cryoprotected with a mixture of 90% well solution and 10% PEG 200, and cryocooled by plunging into liquid nitrogen. Data were collected at the Advanced Photon Source, beamline 23-ID-B to a resolution of 1.50Å, and the structure was determined by molecular replacement using the lambda light chain variable domain from Fab 2219 (PDB code 2b0s) as a starting model. The final model (PDB code 3T0W) contains two $M8V_L$ molecules in the asymmetric unit (residues A1-108, B1-108), two half-occupancy DIR molecules, 2 chloride ions, 2 PEG molecules, and 319 waters, and has $R_{cryst}$ and $R_{free}$ values of 14.8% and 18.3%, respectively.

The $M8V_L$ (15 mg/ml in 0.2M Tris, 0.15M NaCl, pH 7.5) in the absence of DIR was crystallized at 22°C in sitting drops with a well solution of 2M ammonium sulfate, 2% PEG 400, 0.1M HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, that represented a condition originally identified using the JCSG/ IAVI CrystalMation robot (Rigaku) and screening 384 different conditions at two different temperatures. The crystals were cryoprotected with a mixture of 70% well solution, 30% ethylene glycol, and cryocooled by plunging into liquid nitrogen. Data were collected at the Advanced Light Source, beamline 4.2.2 to a resolution of 1.45Å. The structure was determined by molecular replacement with search model coordinates consisting of one $V_L$ domain from the previously determined $M8V_L$ structure. The final model (PDB code 3T0V) contains one unliganded $M8V_L$ molecule in the asymmetric unit (residues A1-110), 127 waters, and 1 Tris, 1 sulfate ion, 4 ethylene glycol, and 2 PEG molecules. The final $R_{cryst}$ and $R_{free}$ values are 17.9% and 20.8%, respectively.

The $M8V_LS^{L55}P$ co-crystallized with DIR was formed by mixing $M8V_LS^{L55}P$ (3 mg/ml) with a 10 fold molar excess of DIR (10 mg/ml in 20% DMSO, 0.2M Tris, 0.15M NaCl) and concentrating the mixture to 15mg/ml. Initial crystallization screening was carried out with the CrystalMation robot and the crystal used for data collection was grown in a sitting drop with a well solution of 0.25M ammonium sulfate, 30% PEG 4000. The crystal was cryoprotected with a mixture of 70% well solution and 30% ethylene glycol and cryocooled by plunging into liquid nitrogen. Data were collected at the Advanced Photon Source, beamline 23-ID-B to a resolution of 1.95Å and the structure was determined by molecular replacement using one $V_L$ domain from the $M8V_L$ structure as starting model. The final model (PDB code 3T0X) contains two $M8V_LS^{L55}P$ molecules (residues A1-108, B1-106), one DIR molecule with two alternate conformations for the indole moiety, 1 sulfate ion, 7 ethylene glycol molecules and 147 waters. Final $R_{cryst}$ and $R_{free}$ values are 20.9 and 24.0.

### NMR Analysis

Purified $M8V_L$ protein was dialyzed into buffer (10mM $Na_2HPO_4$ pH 6.3, 250 mM NaCl) with 0.75mM each of arginine and glutamate to reduce aggregation and 0.02% azide (23).The protein was concentrated to 0.9 mM for NMR relaxation experiments. The $M8V_L$ fluoromodule with DIR was formed using a 4:1 molar ratio of lyophilized DIR that was resuspended into the dialyzed protein sample and incubated overnight. NMR experiments were performed with a Bruker 600MHz NMR spectrometer fitted with a cryoprobe at 304K. The spectra were processed with NMRPipe (24). The $T_1$, $T_2$ and heteronuclear NOE experiments (25) on $^{15}N$ labeled material were analyzed using NMRViewJ (26). In-house scripts were used to further analyze the data and prepare the input files for Relax-NMR(27) (28) (29) (30) (31). The global correlation times, $\tau_m$, were determined using the 'model-free.py' script available in Relax-NMR (32) (33) (34).The theoretical values for global correlation time were obtained using HYDRONMR program (35). Standard multi-nuclear NMR sequences (36) (37) were used in conjunction with $^{15}N$ and $^{13}C$ labeled material to obtain resonance assignments for mainchain atoms in the unliganded complex utilizing an in-house Monte Carlo based assignment program (38). Inter-proton NOEs were obtained from $^{15}N$ separated NOESY experiments that detected either the $^{1}H$ frequency, or the $^{13}C$ frequency of the attached carbon, of protons that were dipolar coupled to NH protons (39).

### Calculation of protein-DIR van der Waals Energies

CHARMM version 33b2 (40) was used to compare van der Waals energies using the top aa22 parameter file. Bond lengths, angles, and planar torsional angles for DIR were obtained from the X-ray coordinate files described in this paper. Standard non-bonded atom parameters were used for aliphatic and aromatic carbons and hydrogens in DIR. The protein-DIR system was minimized with 100 steps of steepest decent and then adopted basis set Newton Raphson minimization was applied until the energy converged. During the minimization the protein mainchain N,$C_\alpha$ and carbonyl carbons were constrained to positions in the original X-ray structure. The difference between the van der Waals energy for the protein plus DIR minus the protein without DIR was used to determine the contribution of protein-DIR interactions to the van der Waals energy.

## Results

### Directed evolution of the $V_H$-$V_L$M8 FAP

In order to determine if changes in affinity and quantum yield can increase fluorogenicity of the $V_H$-$V_L$ M8, directed evolution was performed using FACS selection for increased fluorescence of yeast surface displayed FAP at low DIR concentration (8). The FACS enrichment selections were carried out at 250 pM DIR after determining the yeast cell surface affinity of $V_H$-$V_L$ M8 is 1.2nM (Supporting Information S1, Panels A-C). Each

round of enrichment and selection monitored both the amount of yeast surface expression and intensity of DIR fluorescence signal of cells in the mutagenized population. Single cells were isolated during the third cycle of enrichment as described previously (9). Individual clones of interest were identified based on visual inspection for increased fluorescence on induction plates compared to the wild type $V_H$-$V_L$ M8. Amino-acid changes were identified by DNA sequencing. DIR $\phi_f$ and fluorescence enhancement of cell surface displayed FAPs were then determined by FACS analysis of individual clones. The equilibrium dissociation constants for isolated clones from directed evolution were initially determined in order to identify mutants with enhanced binding to DIR (Supporting Information Figure S2). These measurements identified three classes, as determined from sequence length and alignment to the original $V_H$-$V_L$ M8, from a total of twelve single clones (Supporting Information Figure S3 and S4).

The first class of isolated clones, which have both $V_H$ and $V_L$ domains, is represented by L9 in Table 1 (see S2 for binding curves of FAPs on the yeast surface, and S5 for the binding curves of soluble Q9 and J8), the yeast surface display fluorescence enhancement of this clone was increased five-fold over the parent M8. Sequence analysis revealed several amino-acid changes in the $V_H$ domain of the clone $V_H$-$V_L$ L9: $F^{H29}S$, $W^{H36}C$, $A^{H93}V$, $I^{H113i}T$ according to Kabat numbering system (H113i is an extension of Kabat numbering based on extra residues on the C-terminus of the M8 heavy chain) (21).

A second class of clones, which have only the $V_L$ domain, is represented by Q9 in Table 1. This clone displayed an increased fluorescence enhancement of four-fold above the $V_H$-$V_L$ M8 parent. Sequence analysis revealed that this clone contained 7 mutations in the $V_L$: $Q^{L1}R$,$T^{L14}I$, $D^{L30}G$, $Q^{L37}R$, $S^{L55}P$, $F^{L62}L$, $S^{L80}P$, in addition to suffering deletion of the $V_H$ domain (Supporting Information Figure S4).

A third class represented by one clone, J8, had a six fold increase in fluorescence enhancement compared to the parent $V_H$-$V_L$ M8 (Table 1). This clone is remarkable in that it is a “pseudodimer” gene containing two tandem-linked $V_L$ domains, presumably generated by recombination during the yeast gap repair process used in the creation of the directed evolution library. Each of the two $V_L$ domains retains approximately 95% homology to the $V_L$ domain of parent $V_H$-$V_L$ M8 (Supporting Information Figure S6). The $V_L$-$V_L$ J8 pseudodimer displayed six total amino-acid substitutions in both light domains. Four changes were present in the N-terminal light chain, $D^{L85}N$, $K^{L103}T$, $L^{L107}S$, $I^{L108i}T$ and an additional two changes were present in the C-terminal light chain, $S^{L9}P$, $S^{L32}P$. The pseudodimer contains 12 residues that do not align to the M8$V_L$ sequence and are located prior to the flexible linker. These residues align with the region of the $V_H$ M8 that is N-terminal to the linker. They are described as $A^{L108a}$,$S^{L108b}$, $T^{L108c}$ , $K^{L108d}$, $G^{L108e}$, $P^{L108f}$,$S^{L108g}$,$G^{L108h}$,$T^{L108i}$,$L^{L108j}$, $G^{L108k}$ because they are not found in conventional antibodies and are an addition to the typical Kabat numbering scheme (Supporting Information Figure S7). Homologous recombination is likely to have generated these additional residues as well.

## Directed evolution of the M8$V_L$ FAP

Directed evolution of M8$V_L$ was performed as described above using a DIR concentration of 1 nM during FACS sorting (Supporting Information Figure S1, Panels D-F). This selection generated a single class of clones, represented by clone A4 in Table 1, with a greater than 8 fold increase in fluorescence enhancement compared to the parent M8 $V_L$-DIR (Table 1). All 16 clones from this class contained only one change of serine to proline at position 55 ($S^{L55}P$). This clone is subsequently referred to as M8$V_L$$S^{L55}P$, in accordance with Kabat nomenclature.

### Characterization of tandem homodimers of M8$V_L$ FAPs on the yeast cell surface

The isolation of a tandem linked $V_L$-$V_L$ J8 pseudodimer from the directed evolution of $V_H$-$V_L$ M8 led us to engineer tandem homodimer genes of the M8$V_L$ domain and investigate the effect of different ($G_4S$) linker lengths on fluorescence enhancement. Yeast cell surface expression and fluorescence enhancement measurements of clones with different linker lengths were performed. There was no significant difference in DIR affinity of tandem homodimers with extended ($G_4S$) linkers based on cell surface affinity measurements (Supporting Information S8). None of the linker lengths were less than $(G_4S)_3$ (15 amino acids) and thus not predicted to form diabodies or triabodies (41) (42). Fluorescence enhancement was increased two-fold for homodimers containing the $(G_4S)_4$ or $(G_4S)_6$ linkers. The tandem homodimer gene made from the M8$V_L$S$^{L55}$P domain produced a FAP with a three-and-a-half fold increased fluorescence enhancement compared to the similar tandem homodimer made with the M8$V_L$ domain (Table 2).

### Characterization of $V_H$-$V_L$ M8, M8$V_L$ and tandem $V_L$ homodimer FAPs as soluble proteins

To verify that the affinity of the FAPs for DIR is not influenced by the yeast surface display of the protein, we attempted to express soluble versions of these FAPs. The parent $V_H$-$V_L$ M8 FAP showed a remarkable affinity for DIR on the yeast cell surface ($K_D$ = 1.2nM); however, a $K_D$ for the soluble protein produced by yeast secretion could not be determined by fluorescence measurements (Supporting Information Figure S9); the purified $V_H$-$V_L$ M8 protein showed a complete absence of fluorogen activation. To date, no other FAP proteins that we have characterized have shown strong affinity or activity by cell surface titration measurements but a complete lack of activity when assayed as a soluble protein. The yield of yeast secreted protein from $V_H$-$V_L$ M8 was unexpectedly low, based on the typically strong correlation of yeast secretion yield and yeast surface display (43) (44). Thus, further biochemical characterization of soluble $V_H$-$V_L$ M8 protein was not possible due to low protein expression levels by either yeast secretion or bacterial expression (data not shown). Similarly, no characterization was possible for purified proteins from any of the clones from directed evolution that contained the $V_H$ domain of $V_H$-$V_L$ M8 (such as $V_H$-$V_L$ L9), due to poor protein expression levels (data not shown). However, the purified soluble $V_L$-$V_L$ J8 pseudodimer protein has a $K_D$ for DIR of less than 0.1nM (Table 1). This dissociation constant is similar to that found for the synthetic $V_L$ dimer (M8$V_L$(G4S)$_3$), however the low $K_D$ values for these two proteins precludes an accurate comparison of their $K_D$ values. Determination of $\phi_f$ for J8 FAP was not possible due to low protein yield.

The solution binding properties of three different monomer $V_L$ FAPs were studied, Q9, M8$V_L$, and M8$V_L$S$^{L55}$P. Purified M8$V_L$ protein showed an overall dissociation constant of $2.5\times10^{-15}$ M$^2$ (Figure 1) and a robust $\phi_f$ of 71%. The affinity of the M8$V_L$S$^{L55}$P mutant increased greater than ten-fold to $1.0\times10^{-16}$ M$^2$ (Figure 1). Due to the nature of the binding reaction, the amount of DIR required to half-saturate the protein (apparent $K_D$) depends on the protein concentration; at a protein concentration of 1nM, the apparent $K_D$ values would be 2.5μM for M8$V_L$ and 0.1μM for M8$V_L$S$^{L55}$P, i.e. a ~25 fold increase in affinity. Although the M8$V_L$S$^{L55}$P mutant protein showed an increase in affinity, the $\phi_f$ decreased to 58% (Table 1). The increase in affinity of M8$V_L$S$^{L55}$P is greater than that observed for the $V_L$ domain isolated during affinity maturation of the original $V_L$-$V_H$ construct (Q9); the overall dissociation constant of the soluble Q9 protein for DIR is $1.3\times10^{-15}$ M$^2$. Although the affinity of Q9 for DIR is about 2-fold higher than M8$V_L$, the $\phi_f$ is relatively low at 19% (Table 1). There were no significant differences observed in the DIR absorbance spectra of fluoromodules M8$V_L$ or Q9$V_L$, despite differences in $\phi_f$ (Supporting Information Figure S10).

We also analyzed the affinities for DIR and $\phi_f$ of the soluble proteins made by $M8V_L$ tandem dimers with modified linker lengths. Altering the $(G_4S)$ repeats produced soluble proteins with differing affinities and $\phi_f$. The $K_D$ values of the $M8V_L$ dimer with the $(G_4S)_3$, $(G_4S)_4$, or $(G_4S)_6$ linker was less than 0.1nM and thus it is not possible to compare the $K_D$ values for these different constructs. The quantum yield was somewhat dependent on the length of the linker; a value of 38% was found for the $(G_4S)_3$ linker, 64% for the $(G_4S)_4$ linker, and 55% for the $(G_4S)_6$ linker (Table 2 and Supporting Information Figure S11B). A slight drop in the quantum yield for the longer linker may be due to the formation of intermolecular complexes, similar to diabodies or triabodies, but this was not indicated by size exclusion chromatography (Supporting Information S12). The $M8V_LS^{L55}P$ dimer with a $(G_4S)_3$ linker has a $K_D$ of less than 0.1 nM and the lowest $\phi_f$ of 14% (Table 2).

## Structure of unliganded $M8V_L$ and co-crystals of $M8V_L$ and $M8V_LS^{L55}P$ with DIR

To investigate both the mechanism of DIR fluorescence activation by the $M8V_L$ FAPs and to understand the increased affinity and altered fluorescence enhancement of the $M8V_LS^{L55}P$ mutant we determined the crystal structure of two of these proteins (Table 3). Crystal structures were obtained for both $M8V_L$ and $M8V_LS^{L55}P$ in complex with DIR (1.5Å and 1.96Å, resolutions, respectively), as well as the $M8V_L$ in the absence of DIR (1.45Å). In all three crystal structures, the $M8V_L$ or $M8V_LS^{L55}P$ domains adopt a typical Ig-fold. CDR loops from $V_L$ domains from all three structures adopt the expected canonical structures, and are classified as: L1=5λ/13A, L2=1/7A, and L3=5λ/11A (45) (46). Both $M8V_L$ and $M8V_LS^{L55}P$ crystallize as a non-covalent homodimer of $V_L$ domains, with the DIR molecule packed tightly at the homodimer interface (Figure 2). In both structures, the $V_L$ domains are related by an approximate non-crystallographic two-fold axis, with an angle of ~178° relating the two $V_L$ domains in each structure, and the DIR molecules bound on the non-crystallographic symmetry (NCS) axis (Figure 3 and Table 4). In the $M8V_L$ structure, the DIR occupies two conformations, related by the approximate NCS. In the $M8V_LS^{L55}P$ structure, only the indole ring adopts alternate conformations, also related by the NCS two-fold axis (Figure 4). The alternate conformations of the DIR are a crystallization artifact, where the $V_L$ dimer binds DIR in one single conformation but the complex then crystallizes in two orientations around NCS 2-fold axis. In the absence of DIR, the $M8V_L$ crystallizes with one monomer in the asymmetric unit. This monomer does not closely associate with any symmetry-related neighbors to form a dimer in the crystal (Figure 7a). The homodimeric arrangement seen for $M8V_L$ and $M8V_LS^{L55}P$ differs strikingly from that seen for a $V_L$-$V_H$ dimer in an Fab molecule, and also differs from any $V_L$-$V_L$ dimer structures found in the PDB in that the two $V_L$ domains are anti-parallel (Supporting Information Figure S13), while still burying the hydrophobic interface that would normally be buried in a $V_L$-$V_H$ interaction in a typical antibody. The DIR is buried in a deep pocket and forms contacts primarily with the CDR loops for both $M8V_L$ and $M8V_LS^{L55}P$ (Figure 2). DIR contact residues are contributed by all CDR loops: $Tyr^{L34}$, $Asn^{L50}$, $Arg^{L53}$, $Ser/Pro^{L55}$, $Ser^{L56}$, $Leu^{L89}$and $Trp^{L96}$ (Table 4). $Tyr^{L34}$ participates in π-stacking with the conjugated DIR polymethine bridge for both $M8V_L$ and $M8V_LS^{L55}P$ (Figures 3 and 4), with the centroid of the $Tyr^{L34}$ rings located 5.6-6.1 Å from the plane of the DIR, and angles between $Tyr^{L34}$ and DIR between 31.7°-36.7° (Supporting Information S14). $Tyr^{L49}$ also participates in parallel π-stacking with the DIR quinoline ring for both $M8V_L$ and $M8V_LS^{L55}P$ (Figures 3 and 4), with the centroid of the $Tyr^{L49}$ rings located 3.6Å −4.0Å from the plane of the DIR, and angles between Tyr and DIR ranging between 5.0°-10.7° (Supporting Information S14). $Trp^{L96}$ packs perpendicular to the indole ring for both $M8V_L$ and $M8V_LS^{L55}P$ (Figure 3 and 4), with the centroid of the $Trp^{L96}$ L ring located 5.5Å-5.7 Å from the plane of the DIR with angles between the $Trp^{L96}$ and indole ring ranging between 70.3°-78.6° (Supporting Information S14). The DIR molecules are bound with nearly co-planar quinoline and indole rings, with angles between the two ring systems (for the two DIR

conformers) in M8V$_L$-DIR of 8.5° and 4.4° and in M8V$_L$S$^{L55}$P-DIR of 4.2° and 3.0° (Supporting Information S14 and S15).

M8V$_L$ and M8V$_L$S$^{L55}$P complexes with DIR bury a similar amount of molecular surface upon binding DIR, with 573Å$^2$ of protein and 482Å$^2$ of DIR surface buried in the M8V$_L$ complex, and 581Å$^2$ of protein and 456Å$^2$ of DIR surface buried in the M8V$_L$S$^{L55}$P binding pocket (Figure 3). Contacts between DIR and protein are all van der Waals except for hydrogen bonds between protein and the sulfonate groups of the DIR (Figure 4). The M8V$_L$S$^{L55}$P V$_L$ domains do not associate as closely with DIR, with only 53/75 total contacts (van der Waals and hydrogen bonds, to each alternate conformation of DIR) to DIR as opposed to 67/83 contacts seen for the M8V$_L$-DIR complex (Table 4). In particular, the M8V$_L$-DIR Tyr$^{L49}$ has slightly more van der Waals (10) contacts to the DIR quinoline ring than the M8V$_L$S$^{L55}$P-DIR Tyr$^{L49}$. The effect of the reduced number of van der Waals contacts in the M8V$_L$S$^{L55}$P-DIR structure compared to the M8V$_L$-DIR on the binding energy was investigated by calculate the differences in DIR binding energies between the two structures using CHARMM. In spite of the reduced number of van der Waals contacts in the M8V$_L$S$^{L55}$P-DIR structure, the overall contribution of van der Waals interactions to the DIR binding energy is more negative in the M8V$_L$S$^{L55}$P-DIR at −48.8 kcal/mol; while the van der Waals energy is less negative and thus less favorable at −43.5 kcal/mol for the M8V$_L$-DIR. Consequently, the replacement of serine by proline leads to an additional stabilization of the complex by 5.3 kcal/mol. Superimposition of the M8V$_L$ and M8V$_L$S$^{L55}$P structures reveals no significant conformational changes between the structures, despite a greater than ten-fold decrease in the $K_D$ of the M8V$_L$S$^{L55}$P FAP protein for DIR compared to M8V$_L$. RMSDs for the Cα atoms from residues 1-108 are 0.43Å for A chains, and 0.46Å for the B chains, and 0.52Å for both A and B chains (A and B chains designate the two light chains). When the A chains from each structure are superimposed, it takes only a 1.6° rotation to overlap the B chains. Thus the M8V$_L$ and M8V$_L$S$^{L55}$P structure have similar tertiary and quaternary structures. Three of the four aromatic residues that surround the DIR are slightly closer in the M8V$_L$ complex than in the M8V$_L$S$^{L55}$P complex, with distances between the ring centers of Tyr$^{L34A\text{-}34B}$, Tyr$^{L49A\text{-}49B}$, Trp$^{L96A\text{-}96B}$, and Phe$^{L98A\text{-}98B}$of 8.6Å, 7.2Å, 8.8Å, and 7.3Å for M8VL and 8.8Å, 7.4Å, 8.6Å, and 7.5Å for M8V$_L$S$^{L55}$P. This observation is in good agreement with the slightly smaller number of van der Waals contacts to DIR for the M8V$_L$S$^{L55}$P. Pro$^{L55}$ is close to residue Tyr$^{L49}$ that packs next to the DIR quinoline ring system. The Pro$^{L55}$ residue in the M8V$_L$S$^{L55}$P packs against Tyr$^{L49}$ in an edge-to face manner, with the plane of the proline approximately 4Å from the edge of Tyr$^{L49}$ (Figure 5) and engages in stronger van der Waals interactions with DIR than the corresponding Ser$^{L55}$ in M8V$_L$.

The structure of the unliganded M8V$_L$ is very similar to the M8V$_L$-DIR complex. The RMSD for backbone atoms is 0.95 Å, with the largest difference in backbone configuration in the region of P$^{L8}$, which shows a 4.1 Å displacement between the $C_\alpha$ carbons. Two of the aromatic sidechains that π-stack on the bound DIR, Tyr$^{L34}$ and Tyr$^{L49}$, are superimposable in both structures (Figure 6). In contrast, Tyr$^{L36}$ and Phe$^{L98}$, which also interact with the bound DIR, show large changes in sidechain orientation between the unliganded and M8V$_L$-DIR complex (Figure 6). The X-ray derived structure of the unliganded M8V$_L$ appears to reflect the solution conformation of monomeric M8V$_L$. A comparison of 108 $H_N$-$H_N$ and $H_N$-aliphatic NOE derived distances show only three violations greater than 0.1 Å and no violations greater than 0.5 Å.

### NMR data concludes that dimerization of two monomeric M8V$_L$ is dependent on DIR

Attempts to measure the oligomeric state of M8V$_L$ by size exclusion were unsuccessful due to interaction of the protein with the column (Supporting Information Figure S16). Consequently, solution NMR experiments were performed to determine the $^{15}$N $T_2$

relaxation times in the presence and absence of DIR to confirm that $M8V_L$ dimerization is not an artifact of crystallization (Figure 7 b). The mean transverse relaxation time ($T_2$) for $M8V_L$ in the absence of DIR was found to be ~95.8 ms. Similar values were obtained for $M8V_LS^{L55}P$ (data not shown). The mean $T_2$ value for the $M8V_L$ in the presence of DIR was 57.15 ms (Supporting Information Figure S17) and a similar decrease was observed for $M8V_LS^{L55}P$ (data not shown). To verify the oligomeric state as suggested by $T_2$ relaxation time, we determined the theoretical and experimental global correlation time ($\tau_m$). The $\tau_m$ for the monomer and homodimer were theoretically determined by HYDRONMR using the appropriate PDB file. The $\tau_m$ for the $M8V_L$ in the absence of DIR was 9.36ns compared to the theoretical $\tau_m$ for $M8V_L$ in the presence of DIR at 16.38ns. The $\tau_m$ values determined from the experimental relaxation data were 9.90 ns for the $M8V_L$ in the absence of DIR and 15.87 ns for the $M8V_L$complexed with DIR. These experimental data are consistent with the formation of a dimer in solution when DIR is present. We performed further calculations to confirm the oligomeric state indicated from the mean $T_2$ values using the $T_1$, $T_2$ and heteronuclear NOE data. The theoretically predicted $\tau_m$ for monomeric $M8V_L$ of 9.36 ns closely matches the experimental $\tau_m$value of 9.90 ns. The experimental $\tau_m$ value for dimeric $M8V_L$ using the model-free approach is 12.37 ns. The model-free spectral density function that was used assumes a spherical shape for the molecule, while the actual $M8V_L$ structure is more consistent with a prolate ellipsoid with 2:1 axial ratio. The experimental $\tau_m$ value for $M8V_L$-DIR, when adjusted for this shape, is 15.87 ns which is consistent with the predicted value of 16.38ns for the homodimer (47) (48).

## Discussion

We have described the structure-function relationship of a novel fluoromodule based on a scFv that selectively binds the environmentally sensitive fluorogen DIR. We were interested in the dramatic fluorescence activation of DIR by the $V_L$ domain when separated from the $V_H$ domain and how it related to the fluorescence-generating function of these domains. Affinities and $\phi_f$ were determined for several tandem homodimers designed to covalently link two $V_L$ domains containing differing serine-glycine linker repeats.

### Comparison of directed evolution of $V_H$-$V_L$M8 and $M8V_L$ FAPs

The directed evolution of the $V_H$-$V_L$ M8 produced three distinct families of clones as opposed to the directed evolution of the $M8V_L$ that yielded a single clone. The $V_H$-$V_L$ M8 provided more genetic material for directed evolution and more variants were isolated, underscoring the utility of the directed evolution approach to select desired protein properties without previous structural information (8). These results also suggest that there are limitations to the usefulness of rational design approaches, which may fail to account for unusual protein conformations, protein-protein interactions or ligand interactions.

The finding that the $V_H$-$V_L$ M8 did not activate DIR when purified as a soluble protein led us to speculate that interactions may be occurring between neighboring FAPs on the surface of the yeast cell. We suggest that the tight affinity of yeast surface displayed $V_H$-$V_L$ M8 for DIR is due primarily to the proximity of two nearby $V_L$ domains as a result of the pPNL6 yeast surface display system. This proximity would allow two nearby $V_L$ domains from two $V_H$-$V_L$ chains to readily dimerize after addition of DIR, as the yeast cell surface is highly studded with expressed scFv (49) (50). Homodimerization of $V_L$ domains occurs in additional $V_L$ domains that bind the fluorogen Malachite green (personal communication Christopher Szent-Gyorgyi).

The two directed evolution approaches indicate that yeast surface display favors the dimerization of the $M8V_L$, regardless of the starting FAP, as both monomeric $V_L$ domain and covalent $V_L$-$V_L$ FAP clones were isolated from the directed evolution of the $V_H$-$V_L$

M8. The strong selective pressure of 250pM DIR is likely responsible for the isolation of covalently linked M8$V_L$ domains from $V_H$-$V_L$ M8 parent. Additionally, three $V_L$ clones isolated without associated $V_H$ domains from the $V_H$-$V_L$ M8 directed evolution contained cysteine substitutions in the flexible linker normally between the $V_H$ and $V_L$ domains. These cysteine residues might enhance dimerization on the yeast cell surface through the formation of disulfide bonds between the linker regions of nearby $V_L$ domains. This hypothesis is difficult to test without disruption of disulfide bonds that anchor the FAP to the yeast cell surface via the linage to the Aga1p and Aga2p proteins on the yeast cell.

It was unexpected that the directed evolution of the M8$V_L$ would produce a single clonal population. This result is an indication that the M8$V_L$ either required less optimization to increase affinity and $\phi_f$ than $V_H$-$V_L$ M8, or when displayed as a single light domain, M8$V_L$ has more stringent structural requirements. In the latter case, single amino-acid changes in the $V_L$ domain have to be accommodated across the $V_L$-$V_L$ homodimer to allow the tight interaction of both the protein and dye component of the fluoromodule.

## Yeast surface display analysis of two domain FAPs

Two clones isolated from the directed evolution of $V_H$-$V_L$ M8 underscore the novel sequence variations that these FAPs can undergo to cause fluorescence enhancement of DIR. The clone $V_H$-$V_L$ L9 that resulted from the $V_H$-$V_L$ M8 enrichment retained both $V_H$-$V_L$ domains, but contained several mutations only in the $V_H$ ($F^{H29}S$, $W^{H36}C$, $A^{H93}V$, $I^{H113i}T$). The increased fluorescence enhancement of this clone could be a result of $V_L$ dimer formation that is enhanced by the $W^{H36}C$ mutation. This mutation might allow the formation of a disulfide bridge between neighboring $V_H$ domains on the yeast cell surface, facilitating a closer proximity of their connected $V_L$ domains. The $V_L$-$V_L$ J8 tandem $V_L$ dimer isolated from directed evolution of $V_H$-$V_L$ M8 covalently links two M8$V_L$ domains, resulting in a substantial decrease in $K_D$ that is comparable to the synthetically constructed homodimeric M8$V_L(G_4S)_3$. It is noteworthy that the $V_L$-$V_L$ J8 also contains residue insertions prior to the $G_4S$ linker region that may impart increased affinity or $\phi_f$. We hypothesize that residue insertions that originated from the $V_H$ domain act to extend the linker length and provide a more suitable dimer structure for the activation of DIR. However, the additional amino-acid mutations present in both light domains of $V_L$-$V_L$ J8 might have an additional effect on the affinity of $V_L$-$V_L$ J8 for DIR.

The isolation of the $V_L$-$V_L$ J8 tandem $V_L$ homodimer was novel and unexpected, which led to the investigation of engineered tandem M8$V_L$ dimers with different linker lengths. The additional amino acids prior to the linker region in $V_L$-$V_L$ J8, as well as measurements from the crystal structure, suggested that a six repeat $G_4S$ linker would have the most relaxed structural constraints for DIR. The yeast surface displayed tandem dimers showed no significant difference in DIR affinity regardless of linker length or inclusion of the M8$V_L S^{L55}P$ mutation. Thus the linker length between two tandem dimers on yeast surface display does not alter the affinity of the tandem dimers for DIR, however it does alter the affinity of the same soluble proteins. These data indicate that the yeast surface display platform may allow $V_L$-$V_L$ interactions of independently displayed $V_L$ domains.

## Differences in fluorescence activity for yeast surface display versus solution

A consistent observation is that affinities and fluorescence enhancement of fluorgen activating proteins on the surface of yeast can be markedly different than obtained for the same protein free in solution. In particular, $K_D$ values for the homodimeric $V_L$-$V_L$ proteins in solution were consistently lower than the affinity found on the surface of the yeast cell. For example, the $K_D$ values for M8$V_L$-$(G_4S)_3$ was 1.7nM on the surface, but substantially less than 1 nM in solution. Similarly, J8 showed a surface $K_D$ of ~10 nM, yet the $K_D$ in

solution was also less than 1 nM. One possible explanation for this effect is that the high density of FAPs on the surface of the protein can result in novel protein-protein interactions. An extreme example of this is the discovery that the soluble $V_H$-$V_L$ M8 protein failed to activate DIR while showing a strong fluorescence enhancement on the yeast surface. This result suggests that in solution the $V_H$ domain may have a high affinity for the attached $V_L$ domain; thus preventing $V_L$-$V_L$ homodimerization that likely occurs on the surface of the yeast cell. Alternatively, the soluble $V_H$-$V_L$ M8 may contain a favorable affinity for DIR but fails to restrain DIR in an appropriate conformation for fluorescence emission, similar to what is predicted for other fluorogenic cyanine dyes (51).

### Relationship between mutations and dimerization on affinity and quantum yield

The soluble Q9 protein has strong affinity for DIR, but a low $\phi_f$ of 19%. The numerous mutations isolated in Q9 ($Q^{L1}R$, $T^{L14}I$,$D^{L30}G$,$Q^{L37}R$, $S^{L55}P$, $F^{L62}L$, $S^{L80}P$) may be responsible for the decrease in $\phi$ , although this effect may be combinatorial. Both the M8V L55 f LS P and Q9 contain the $S^{L55}P$ mutation that may cause a decrease in $\phi_f$ and potentially improve the respective affinities. However, the Q9 $\phi_f$ is significantly lower than either the $M8V_L$ or $M8V_LS^{L55}P$ $\phi_f$. Thus Q9 and $M8V_LS^{L55}P$ suggest that directed evolution of FAP complexes with DIR does not necessarily yield simultaneous increases in both affinity and $\phi_f$. Further characterization of FAPs selected for other fluorogens will determine if decreased $\phi_f$ coupled with enhanced affinity can occur among FAPs subjected to directed evolution.

The linking of $M8V_L$ and $M8V_LS^{L55}P$ by tandem $(G_4S)_3$ linkers generated FAPs with high affinity for DIR, as expected by reducing the entropic penalty for bringing two protein domains together with a covalent linkage. Although the affinity increased, the quantum yield decreased for both $M8V_L$ and $M8V_LS^{L55}P$. Alternatively, the quantum yields of the extend dimers are larger and thus more efficient at emitting fluorescence than the tandem $(G_4S)_3$ $M8V_L$. Thus, the affinity and $\phi_f$ do not simultaneously increase for the different linker constructions of the $M8V_L$ tandem homodimers. These data taken together indicate that the conformational requirements for increased affinity of the fluoromodule do not produce a robust $\phi_f$.

### X-ray crystallography and NMR data support the homodimerization of two $M8V_L$ domains in the presence of DIR and elucidate the mechanism for fluoromodule fluorescence generation

Both the genetic and structural data support the conclusion that the $M8V_L$ forms a homodimer in the presence of DIR. The crystal structures of both M8V and $M8V_LS^{L55}P$ reveal a dimer of $V_L$ domains sandwiching the DIR and constraining the rotation of the two DIR heterocycle rings, so that the quinoline and indole rings are essentially planar. This nearly planar orientation may be optimal for fluorescence decay as it allows DIR to emit fluorescence as it relaxes from the excited to the ground state while the heterocycles are fixed around the conjugated polymethine bridge (52).

NMR experiments are consistent with both $M8V_L$ and $M8V_LS^{L55}P$ forming a homodimer only in the presence of DIR. Based on the canonical scFv interaction between $V_H$ and $V_L$ domains, it would be expected that the domains would have a low micromolar $K_D$ for each other (53); however, even at ~1 millimolar concentrations of $M8V_L$ and $M8V_LS^{L55}P$ in the NMR experiments, there is no evidence of a strong interaction between light domains. The average transverse relaxation time, $T_2$ for the monomeric $M8V_L$ was approximately double that of the $M8V_L$ homodimer in the presence of DIR. This result is consistent with the homodimerization of $M8V_L$ only in the presence of DIR.

The isolation of functional $V_L$ domains that possess strong affinity for small molecule fluorogens is intriguing and shows similarities to previous reports describing Bence-Jones proteins isolated from multiple myeloma patients (54). Bence-Jones proteins are immunoglobulin variable light chains that form homodimers. It has been demonstrated that small molecule haptens in solution penetrate into preformed crystals of Bence-Jones proteins and the haptens interact with the hydrophobic regions formed between two variable light domains (55) (56). Although the FAPs described here are distinct from Bence-Jones proteins as they lack a constant light chain domain, they display low nM affinity for fluorogenic compounds.

The $M8V_LS^{L55}P$ displays a greater than ten fold enhancement in affinity for DIR compared to wild type $M8V_L$, despite having fewer van der Waals contacts with DIR (Tables 1 and 4). Calculations indicate that the complex between $M8V_LS^{L55}P$ and DIR is stabilized by stronger van der Waals contacts. In the absence of detailed solution phase structural and dynamic data, we cannot conclude precisely how the $S^{L55}P$ mutation increases the affinity for DIR. One hypothesis is that the serine to proline mutation imparts more rigidity to the $M8V_LS^{L55}P$ protein and ultimately creates a less flexible dimer binding pocket to retain DIR. This increased protein rigidity may reduce entropic changes during binding, thus modulating the affinity. These interactions will be the focus of future studies to determine the energetics and solution dynamics that underlie the affinity changes of these FAPs with DIR.

Both $M8V_L$ and $M8V_LS^{L55}P$ significantly enhance the fluorescence of DIR, likely by constraining the angle between the planes of the indole and quinoline groups to less than 10 degrees. Previous theoretical predictions of fluorescence generation for torsionally responsive fluorogens such as thiazole orange indicate that optimal radiative decay occurs when the heterocyclic groups are restricted to a 0-60 degree interplanar angle (51). There is no corresponding model for DIR. However, our data indicate that the $\phi_f$ is robust when DIR is held in a nearly planar conformation. The π-stacking interactions between tyrosines in $M8V_L$ and $M8V_LS^{L55}P$ with DIR may also stabilize the fluorescence signal from DIR based on studies of unsymmetrical cyanine dyes that π-stack during DNA intercalation (6) (57). The structural data suggest that conserved π-stacking interactions with $Tyr^{L34}$ and $Tyr^{L49}$ are important for fluorescence activation of both the $M8V_L$ and $M8V_LS^{L55}P$ homodimers and will be the focus of future studies.

The $\phi_f$ of $M8V_LS^{L55}P$ is lower than that of M8V . The $M8V_LS^{L55}P$ mutation may contribute additional effects to the fluorescence of DIR that are unrelated to a global increase in protein rigidity. For example, direct interactions between the proline and DIR in solution may only weakly constrain the conformational of DIR, thereby reducing the $\phi_f$. The reduced quantum yield associated with other FAPs, Q9 for example, may be due to restriction of the planer groups in DIR at a greater angle than optimal for fluorescence emission. Finally, FAPs may show reduced quantum yields due to quenching of the DIR in the excited state caused by electron transfer involving nearby redox-active side chains. Further structural studies of other DIR binding FAPs will help to clarify the interactions required for $\phi_f$.

## Conclusions

Fluoromodules consisting of a specific fluorescence activating protein and an environmentally sensitive fluorogen demonstrate unique properties such as homodimerization induced by the fluorogen DIR. Experimental data show that DIR fluoresces when it is rigidly held between two immunoglobulin variable light domains that dimerize in the presence of DIR. The structural data suggest that conserved π-stacking interactions with tyrosine residues are important for fluorescence activation of both the

$M8V_L$ and $M8V_LS^{L55}P$ homodimers. $M8V_LS^{L55}P$ holds DIR in a potentially less flexible binding pocket compared to that from the $M8V_L$. This change is caused by a single serine to proline mutation that also alters $\phi_f$ and affinity. Our results demonstrate the challenge of predicting enhanced FAP fluorescence via mutagenic approaches in the absence of directed evolution. Linker scanning, for example, may provide information on tolerated structural changes to a particular protein, but fail to successfully constrain a fluorogen to produce fluorescence. Furthermore, it is reasonable that even with high resolution crystal structures and subsequent rational design we may not correctly predict the mutations that increase both DIR affinity and $\phi_f$, given that a single mutation in $M8V_LS^{L55}P$ enhanced binding but not the $\phi_f$. However, combining detailed structural information with computational or rational design approaches may yield significant improvements in fluorescence quantum yield and affinity. For example, performing targeted mutagenesis on the residues that surround the fluorogen in the binding pocket prior to enrichment procedures might increase the probability of obtaining mutants with improved fluorescence. Such an approach may also include computational modeling of the altered protein in addition to random mutagenesis to enhance the FAP complex.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Yehuda Creeger for assistance with FACS, Andreas Plückthun for the gift of pAK-400 plasmid, Gloria Silva and Nathaniel Shank for DIR, along with Josef Franke and Cheryl Telmer for critical review of the manuscript, students in the HHMI supported summer research institute for preliminary NMR data, Paul Barton for help in designing the synthetic $M8V_L$ gene, V. Simplaceanu for help in acquiring the NMR relaxation data and Dr. P.R. Gooley with help with the analysis of the relaxation data.

## Abbreviations

| | |
|---|---|
| **FAP** | (fluorogen activating protein) |
| **DIR** | (dimethylindole red) |
| **$K_D$** | (dissociation constant) |
| **$\phi_f$** | (fluorescence quantum yield) |
| **$V_H$** | (immunoglobulin variable heavy domain) |
| **$V_L$** | (immunoglobulin variable light domain) |
| **GFP** | (Green Fluorescence Protein) |
| **FACS** | (fluorescence activated cell sorting) |
| **CDR** | (complementarity determining regions) |
| **NOE** | (nuclear Overhauser effect) |
| **$\tau_m$** | (global rotational correlation time) |
| **$T_2$** | (transverse relaxation time) |

## References

1. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, WW.; Prasher, DC. Science. Vol. 263. New York, N.Y: 1994. Green fluorescent protein as a marker for gene expression; p. 802-805.

2. Szent-Gyorgyi C, Schmidt BF, Creeger Y, Fisher GW, Zakel KL, Adler S, Fitzpatrick JA, Woolford CA, Yan Q, Vasilev KV, Berget PB, Bruchez MP, Jarvik JW, Waggoner A. Fluorogen-activating single-chain antibodies for imaging cell surface proteins. Nature biotechnology. 2008; 26:235–240.
3. Ozhalici-Unal H, Pow CL, Marks SA, Jesper LD, Silva GL, Shank NI, Jones EW, Burnette JM 3rd, Berget PB, Armitage BA. A rainbow of fluoromodules: a promiscuous scFv protein binds to and activates a diverse set of fluorogenic cyanine dyes. Journal of the American Chemical Society. 2008; 130:12620–12621. [PubMed: 18761447]
4. Zanotti KJ, Silva GL, Creeger Y, Robertson KL, Waggoner AS, Berget PB, Armitage BA. Blue fluorescent dye-protein complexes based on fluorogenic cyanine dyes and single chain antibody fragments. Org. Biomol. Chem. 2011; 9:1012–1020. [PubMed: 21180706]
5. Shank NI, Zanotti KJ, Lanni F, Berget PB, Armitage BA. Enhanced photostability of genetically encodable fluoromodules based on fluorogenic cyanine dyes and a promiscuous protein partner. Journal of the American Chemical Society. 2009; 131:12960–12969. [PubMed: 19737016]
6. Lee LG, Chen CH, Chiu LA. Thiazole orange: a new dye for reticulocyte analysis. Cytometry. 1986; 7:508–517. [PubMed: 2430763]
7. Alfthan K, Takkinen K, Sizmann D, Soderlund H, Teeri TT. Properties of a single-chain antibody containing different linker peptides. Protein engineering. 1995; 8:725–731. [PubMed: 8577701]
8. Colby DW, Kellogg BA, Graff CP, Yeung YA, Swers JS, Wittrup KD. Engineering antibody affinity by yeast surface display. Methods in enzymology. 2004; 388:348–358. [PubMed: 15289082]
9. Feldhaus MJ, Siegel RW, Opresko LK, Coleman JR, Feldhaus JM, Yeung YA, Cochran JR, Heinzelman P, Colby D, Swers J, Graff C, Wiley HS, Wittrup KD. Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. Nature biotechnology. 2003; 21:163–170.
10. Falco CN, Dykstra KM, Yates BP, Berget PB. scFv-based fluorogen activating proteins and variable domain inhibitors as fluorescent biosensor platforms. Biotechnology journal. 2009; 4:1328–1336. [PubMed: 19606431]
11. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. Nature protocols. 2006; 1:755–768.
12. Boder ET, Midelfort KS, Wittrup KD. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:10701–10705. [PubMed: 10984501]
13. Mack ET, Perez-Castillejos R, Suo Z, Whitesides GM. Exact analysis of ligand-induced dimerization of monomeric receptors. Analytical chemistry. 2008; 80:5550–5555. [PubMed: 18543951]
14. Southwick PL, Ernst LA, Tauriello EW, Parker SR, Mujumdar RB, Mujumdar SR, Clever HA, Waggoner AS. Cyanine dye labeling reagents--carboxymethylindocyanine succinimidyl esters. Cytometry. 1990; 11:418–430. [PubMed: 2340776]
15. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein expression and purification. 2005; 41:207–234. [PubMed: 15915565]
16. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. 1997; 276A:307–326.
17. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
18. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica. 2010; 66:486–501.
19. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods in enzymology. 2003; 374:300–321. [PubMed: 14696379]
20. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. 2010; 66:213–221.
21. Kabat, EA.; Wu, TT.; Perry, HM.; Gottesman, KS.; Foeller, C. Sequences of proteins of immunological interest. 5 ed. Vol. Vol. 1. U.S. Department of health and human services; 1991.

22. Hansen CL, Skordalakes E, Berger JM, Quake SR. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:16531–16536. [PubMed: 12486223]

23. Golovanov AP, Hautbergue GM, Wilson SA, Lian LY. A simple method for improving protein solubility and long-term stability. Journal of the American Chemical Society. 2004; 126:8933–8939. [PubMed: 15264823]

24. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of biomolecular NMR. 1995; 6:277–293. [PubMed: 8520220]

25. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. Biochemistry. 1997; 36:2390–2402. [PubMed: 9054544]

26. Johnson, BA. Methods in molecular biology. Vol. 278. Clifton, N.J: 2004. Using NMRView to visualize and analyze the NMR spectra of macromolecules; p. 313-352.

27. d'Auvergne EJ, Gooley PR. The use of model selection in the model-free analysis of protein dynamics. Journal of biomolecular NMR. 2003; 25:25–39. [PubMed: 12566997]

28. d'Auvergne EJ, Gooley PR. Model-free model elimination: a new step in the model-free dynamic analysis of NMR relaxation data. Journal of biomolecular NMR. 2006; 35:117–135. [PubMed: 16791734]

29. d'Auvergne EJ, Gooley PR. Set theory formulation of the model-free problem and the diffusion seeded model-free paradigm. Molecular bioSystems. 2007; 3:483–494. [PubMed: 17579774]

30. d'Auvergne EJ, Gooley PR. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. Journal of biomolecular NMR. 2008; 40:121–133. [PubMed: 18085411]

31. d'Auvergne EJ, Gooley PR. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. Journal of biomolecular NMR. 2008; 40:107–119. [PubMed: 18085410]

32. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules II Analysis of experimental results. Journal of the American Chemical Society. 1982b; 104:4559–4570.

33. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules I Theory and range of validity. Journal of the American Chemical Society. 1982a; 104:4546–4559.

34. Clore G. Deviations from the simple 2-parameter model-free approach to the interpretation of N-15 nuclear magnetic-relaxation of proteins. Journal of the American Chemical Society. 1990b al, e.

35. de la Torre, J. Garcia; Navarro, S.; Martinez, M. C. Lopez; Diaz, FG.; Cascales, J. J. Lopez HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. Biophysical journal. 1994; 67:530–531. [PubMed: 7948671]

36. Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry. 1990; 29:4659–4667. [PubMed: 2372549]

37. Clubb RT, Thanabal V, Wagner G. A new 3D HN(CA)HA experiment for obtaining fingerprint HN-Halpha peaks in 15N- and 13C-labeled proteins. Journal of biomolecular NMR. 1992; 2:203–210. [PubMed: 1422153]

38. Hitchens TK, Lukin JA, Zhan Y, McCallum SA, Rule GS. MONTE: An automated Monte Carlo based approach to nuclear magnetic resonance assignment of proteins. Journal of biomolecular NMR. 2003; 25:1–9. [PubMed: 12566995]

39. Briercheck DM, Wood TC, Allison TJ, Richardson JP, Rule GS. The NMR structure of the RNA binding domain of E. coli rho factor suggests possible RNA-protein interactions. Nature structural biology. 1998; 5:393–399.

40. Brooks BR, Brooks CL 3rd, Mackerell AD Jr. Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB,

Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. Journal of computational chemistry. 2009; 30:1545–1614. [PubMed: 19444816]

41. Atwell JL, Breheney KA, Lawrence LJ, McCoy AJ, Kortt AA, Hudson PJ. scFv multimers of the anti-neuraminidase antibody NC10: length of the linker between VH and VL domains dictates precisely the transition between diabodies and triabodies. Protein engineering. 1999; 12:597–604. [PubMed: 10436086]
42. Albrecht H, Denardo GL, Denardo SJ. Monospecific bivalent scFv-SH: effects of linker length and location of an engineered cysteine on production, antigen binding activity and free SH accessibility. Journal of immunological methods. 2006; 310:100–116. [PubMed: 16499921]
43. Rakestraw JA, Sazinsky SL, Piatesi A, Antipov E, Wittrup KD. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. Biotechnology and bioengineering. 2009; 103:1192–1201. [PubMed: 19459139]
44. Shusta EV, Kieke MC, Parke E, Kranz DM, Wittrup KD. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. Journal of molecular biology. 1999; 292:949–956. [PubMed: 10512694]
45. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. Journal of molecular biology. 1997; 273:927–948. [PubMed: 9367782]
46. Martin AC, Thornton JM. Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. Journal of molecular biology. 1996; 263:800–815. [PubMed: 8947577]
47. Weber G. Polarization of the fluorescence of macromolecules. II. Fluorescent conjugates of ovalbumin and bovine serum albumin. The Biochemical journal. 1952; 51:155–167. [PubMed: 14944567]
48. Lakowicz J. Principles of Fluorescence Spectroscopy (2 ed). 1999
49. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. Nature biotechnology. 1997; 15:553–557.
50. Boder ET, Wittrup KD. Yeast surface display for directed evolution of protein expression, affinity, and stability. Methods in enzymology. 2000; 328:430–444. [PubMed: 11075358]
51. Silva GL, Ediz V, Yaron D, Armitage BA. Experimental and computational investigation of unsymmetrical cyanine dyes: understanding torsionally responsive fluorogenic dyes. Journal of the American Chemical Society. 2007; 129:5710–5718. [PubMed: 17411048]
52. Constantin TP, Silva GL, Robertson KL, Hamilton TP, Fague K, Waggoner AS, Armitage BA. Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules. Organic letters. 2008; 10:1561–1564. [PubMed: 18338898]
53. Mallender WD, Carrero J, Voss EW Jr. Comparative properties of the single chain antibody and Fv derivatives of mAb 4-4-20. Relationship between interdomain interactions and the high affinity for fluorescein ligand. The Journal of biological chemistry. 1996; 271:5338–5346. [PubMed: 8621386]
54. Epp O, Lattman EE, Schiffer M, Huber R, Palm W. The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0-A resolution. Biochemistry. 1975; 14:4943–4952. [PubMed: 1182131]
55. Edmundson AB, Ely KR, Girling RL, Abola EE, Schiffer M, Westholm FA, Fausch MD, Deutsch HF. Binding of 2,4-dinitrophenyl compounds and other small molecules to a crystalline lambda-type Bence-Jones dimer. Biochemistry. 1974; 13:3816–3827. [PubMed: 4850760]
56. Schiffer M, Girling RL, Ely KR, Edmundson AB. Structure of a lambda-type Bence-Jones protein at 3.5-A resolution. Biochemistry. 1973; 12:4620–4631. [PubMed: 4773847]
57. Rye HS, Yue S, Wemmer DE, Quesada MA, Haugland RP, Mathies RA, Glazer AN. Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: properties and applications. Nucleic acids research. 1992; 20:2803–2812. [PubMed: 1614866]
58. Sheriff S, Hendrickson WA, Smith JL. Structure of myohemerythrin in the azidomet state at 1.7/1.3 Å resolution. Journal of molecular biology. 1987; 197:273–296. [PubMed: 3681996]

59. Kraulis P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein strucutres. J Appl Cryst. 1991; 24:846–950.

60. Esnouf RM. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta crystallographica. 1999; 55:938–940.

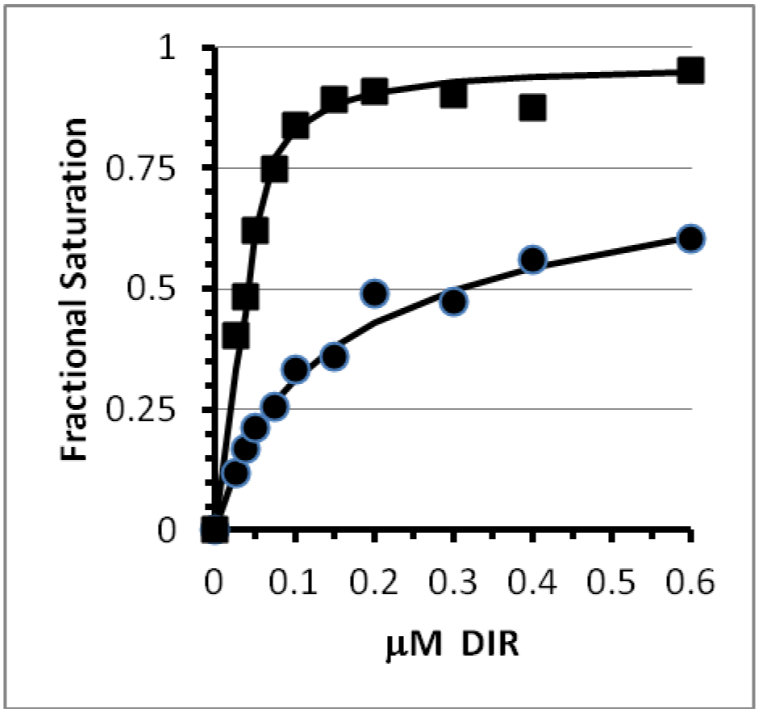



**Figure 1.**

Determination of equilibrium dissociation constants for soluble $M8V_L$ and $M8V_LS^{L55}P$. Fluorescence data for $M8V_L$ (circles) and $M8V_LS^{L55}P$ (squares) were fit to scheme 1. Solid lines show the best fit to the data. Error bars for the individual data points are within the size of plotted points. Serial dilutions of DIR was added to 11 nM soluble $M8V_L$ or 135nM soluble $M8V_LS^{L55}P$ protein in buffer (PBS pH 7.4 with 2mM EDTA, 0.1% w/v Pluronic F-127).

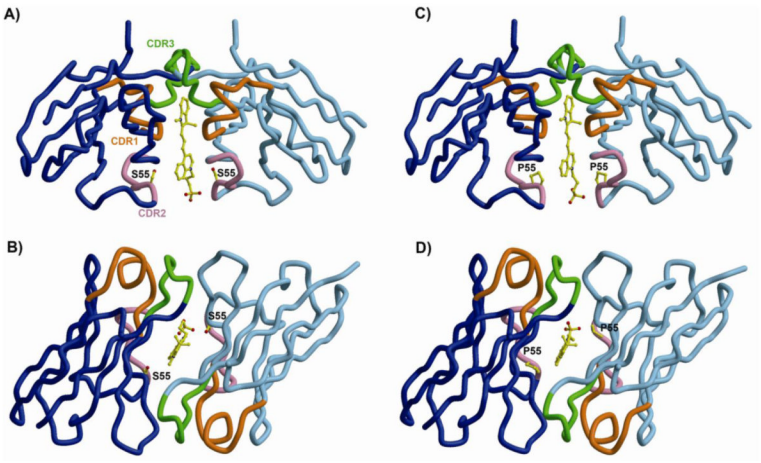



**Figure 2.**

Overall topology of M8V$_L$ and M8V$_L$S$^{L55}$P bound to DIR. (A,B) M8V$_L$ binds DIR (only one DIR conformation is shown for clarity) sandwiched between two identical V$_L$ domains (blue and light blue). The CDR1, CDR2, and CDR3 loops are colored orange, pink, and green, respectively, and are labeled in panel A. The view in (B) is rotated 90° about a horizontal axis. (C,D) M8V$_L$S$^{L55}$P binds DIR in an almost identical way as M8V$_L$. The S$^{L55}$P mutation is the only sequence difference between the two structures. The structure figures were generated with Molscript (59) and Bobscript (60).

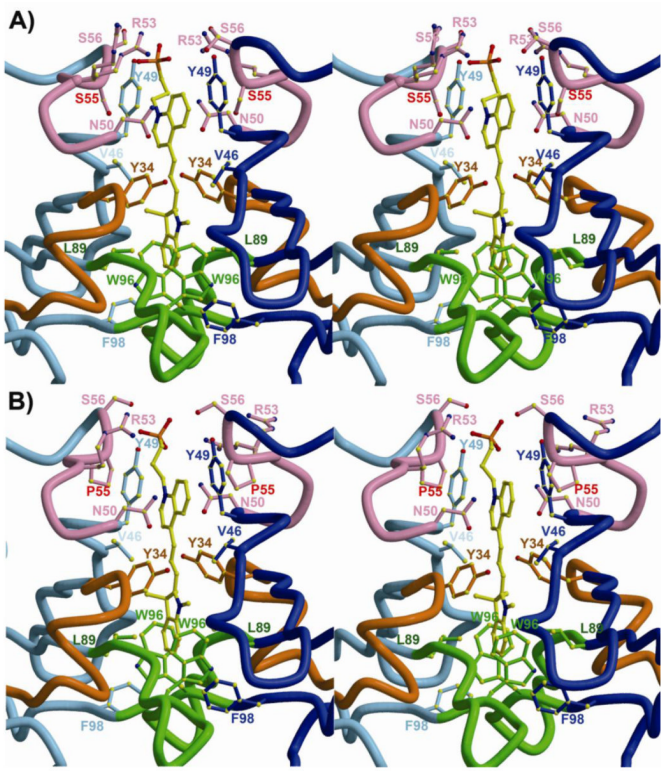



**Figure 3.**

Environment around DIR ligand in M8V$_L$ and M8V$_L$S$^{L55}$P. The two identical V$_L$ domains are shown in blue and light blue. The CDR1, CDR2, and CDR3 loops are colored orange, pink, and green, respectively. (A) Stereoview of the M8V$_L$ binding site with contacting residues labeled. Serine 55 is labeled in red. (B) Stereoview of the M8V$_L$S$^{L55}$P binding site. Proline 55 is labeled in red. Only one DIR conformation is shown for clarity.

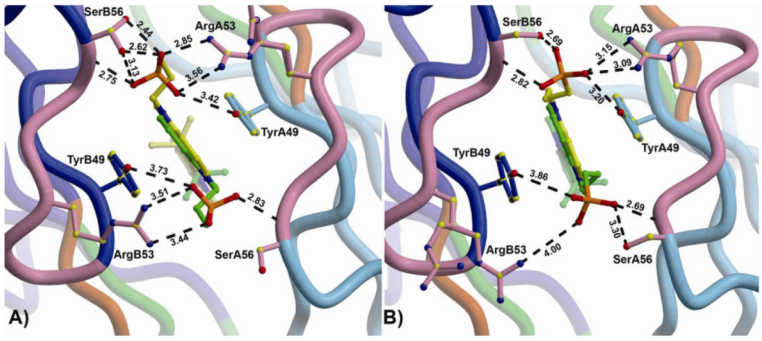



**Figure 4.**

Sulfonate-protein interactions in $M8V_L$ and $M8V_LS^{L55}P$. (A) The interactions between the DIR sulfonate and $M8V_L$ show the two orientations of the bound DIR in the crystal structure (green and yellow), and the A (light blue) and B (blue) $V_L$ chains. CDR2 is shown in pink. $Ser^{B56}$ and $Arg^{A53}$ also sample two alternate conformations. (B) The interactions between DIR sulfonates and $M8V_LS^{L55}P$. The indole ring of the DIR is bound in two alternate conformations, and $Arg^{B53}$ has two alternate conformations. In both $M8V_L$ and $M8V_LS^{L55}P$, one sulfonate (top in both of these views) has more interactions with protein than the alternate (bottom) sulfonate. Note that $S^{L55}$ or $P^{L55}$ is not shown in either panel.

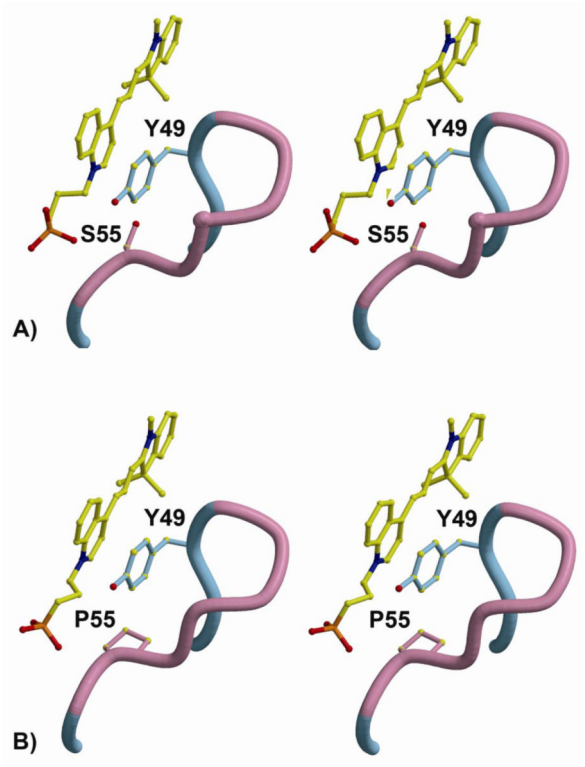



**Figure 5.**

Close-up view of the M8V$_L$S$^{L55}$ L P mutation. (A) Stereoview of M8V$_L$ and the interaction between DIR, Tyr$^{L49}$ and Ser$^{L55}$.Ser$^{L55}$ in the M8V$_L$ has 6 van der Waals contacts with DIR. (B) Same view of the M8V$_L$S$^{L55}$P mutant. The proline ring is approximately 4Å from the edge of the Tyr$^{L49}$ ring and makes 10 van der Waals contacts with DIR.

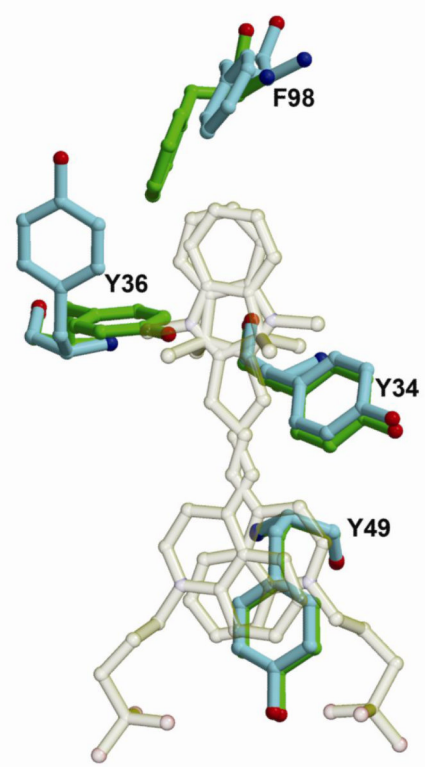



**Figure 6.**

Comparison of the DIR binding site in unliganded $M8V_L$ and the $M8V_L$-DIR complex. The bound DIR is transparent to allow for a better view of the protein side chains. Sidechains from $Tyr^{L34}$, $Tyr^{L36}$, $Tyr^{L49}$ and $Phe^{L98}$ are shown for the unliganded (green) and liganded (light blue) complexes.

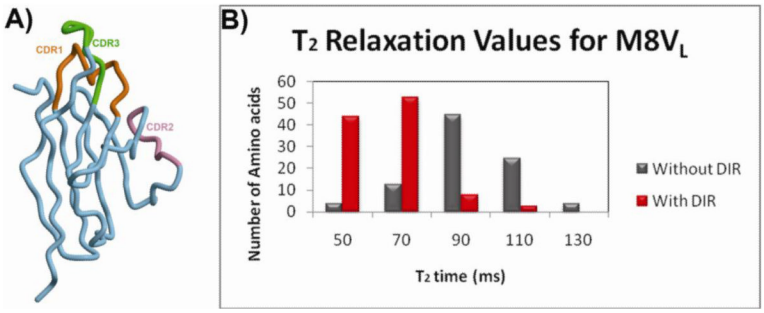
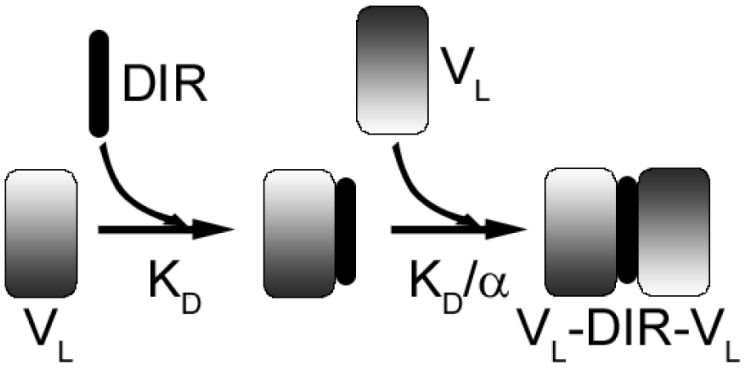



**Figure 7.**

Crystal structure and solution phase $^{15}N$ relaxation data for monomeric $M8V_{L.}$ (A) Overall crystal structure of the $M8V_L$ in the absence of DIR. This $V_L$ domain is monomeric in the crystal structure. (B) The mean NMR $T_2$ relaxation times in the absence (grey bars) or presence (red bars) of DIR.




**Scheme I.**

DIR binds to $V_L$ domain with a dissociation constant $K_D$, forming $V_L$-DIR. The binding of a second $V_L$ then occurs with a dissociation constant $K_D/\alpha$ where α represents a cooperativity parameter.

**Table 1**

Characteristics of clones isolated from directed evolution of $V_H$-$V_L$ M8 and M8$V_L$ FAPs.

| Clone name | Format | Fluorescence enhancement (YSD)* | Soluble $K_D$ or $K_D^2/\alpha$ | $\phi_f$ | Mutations in sequence different from wild type ($V_H$-$V_L$ M8 or M8$V_L$) |
|---|---|---|---|---|---|
| M8 | $V_H$-$V_L$ | 1.0 | ND | | |
| L9 | $V_H$-$V_L$ | 5.7X | ND | | $F^{H29}S$, $W^{H36}C$, $A^{H93}V$, $I^{H113i}T$ |
| Q9 | $V_L$ | 3.8X | $1.3\pm0.3\times10^{-15}$ $M^2$ | 0.19 | $Q^{L1}R$, $T^{L14}I$,$D^{L30}G$,$Q^{L37}R$, $S^{L55}P$, $F^{L62}L$, $S^{L80}P$ |
| J8 | $V_L$-$V_L$ | 6.4X | <0.1 nM† | | VL#1: $D^{L85}N$, $K^{L103}T$, $L^{L107}S$, $I^{L108i}T$<br>VL#2: $S^{L9}P$, $S^{L32}P$ |
| M8$V_L$ | $V_L$ | 3.7X | $2.5\pm0.7\times10^{15}$ $M^2$ | 0.71 | |
| M8$V_L$ $S^{L55}P$ | $V_L$ | 8.2X | $1.0\pm0.5\times10^{-16}$ $M^2$ | 0.58 | $S^{L55}P$ |

*Fluorescence enhancement is calculated as DIR fluorescence signal normalized for cell surface expression for each individual clone, divided by the equivalent value for the wild type parent ($V_H$-$V_L$ M8). The concentration of Q9 was 33 nM, J8 was 9 nM, M8$V_L$ was 10 nM, M8$V_L S^{L55}P$ was 135 nM.

†Data were equally well fit to a wide range of $K_d$ values less than 0.11 nM.

## Table 2

Results of altering glycine-serine rich linker length in tandem homodimers created from the $M8V_L$.

| Clone | $(G_4S)$ repeats in linker | Fluorescence Enhancement (YSD)* | Soluble $K_D$ | $\phi_f$ |
|---|---|---|---|---|
| $dM8V_L(G_4S)_3$ | 3 | 1.0 | <0.1† nM | 0.38 |
| $dM8V_L(G_4S)_4$ | 4 | 2.0 | <0.1† nM | 0.64 |
| $dM8V_LS^{L55}P(G_4S)_6$ | 6 | 2.0 | <0.1† nM | 0.55 |
| $dM8V_LS^{L55}P(G_4S)_3$ | 3 | 3.4 | <0.1† nM | 0.14 |

*Fluorescence enhancement is calculated as DIR fluorescence signal normalized for cell surface expression for each individual clone, divided by the equivalent value for the $M8V_L$. The concentration of $dM8V_L(G_4S)_3$ was 3 nM, $dM8V_L(G_4S)_4$ was 3 nM, $dM8V_L(G_4S)_6$ was 3nM, and $dM8V_LS^{L55}P$ was 12 nM.

†Data were equally well fit to $K_d$ values less than 0.1 nM.

## Table 3

Data collection and refinement statistics for $M8V_L$ and $M8V_LS^{L55}P$

| **Data collection** | **$M8V_L$(no DIR)** | **$M8V_L$+DIR** | **$M8V_LS^{L55}P$+DIR** |
|---|---|---|---|
| Beamline | ALS 4.2.2 | APS 23-ID-B | APS 23-ID-B |
| Wavelength (Å) | 1.000 | 1.033 | 0.722 |
| Resolution (Å)[a] | 1.45 (1.48-1.45) | 1.50 (1.53-1.50) | 1.96 (1.98-1.96) |
| Space group | $C222_1$ | $P2_12_12$ | $P4_32_12$ |
| a,b,c (Å) | 83.66, 93.64, 30.57 | 73.90, 83.53, 35.14 | 83.49, 83.49, 76.35 |
| # of observations | 80,786 (2233) | 250,067 (11,950) | 240,775 (12,015) |
| # of unique reflections | 21,290 (830) | 35,526 (1758) | 20,018 (977) |
| Completeness (%) | 97.7 (79.2) | 99.9 (100.0) | 100.0 (100.0) |
| $R_{sym}$ (%)[b] | 4.3 (31.7) | 8.7 (53.7) | 8.7 (54.9) |
| average I/σ | 30.1 (2.4) | 28.2 (4.2) | 35.5 (4.9) |
| **Refinement statistics all refl. > 0.0σF** | | | |
| Resolution (Å) | 29.2-1.45 | 41.8-1.50 | 37.3-1.96 |
| # reflections (working set) | 20,190 | 33,690 | 18,888 |
| # reflections (test set) | 1090 | 1779 | 1048 |
| $R_{cryst}$ (%)[c] | 17.9 | 14.8 | 20.9 |
| $R_{free}$ (%)[d] | 20.8 | 18.3 | 24.2 |
| # of protein atoms | 845 (29 in alternate conformations) | 1670 (70 in alternate conformations) | 1662 (14 in alternate conformations) |
| # of DIR atoms | 0 | 32 (32 in alternate conformations) | 32 (18 in alternate conformations) |
| # of waters | 127 | 319 (8 in alternate conformations) | 138 |
| # of solvent atoms | 51 | 12 | 33 |
| **Average B-values (Å$^2$)** | | | |
| Chain A | 18.9 | 11.7 | 32.5 |
| Chain B | NA | 15.5 | 46.6 |
| DIR | NA | 16.0 | 27.6 |
| Wilson B-value (Å$^2$) | 15.1 | 11.7 | 26.6 |
| **Ramachandran Plot** | | | |
| Most favored | 90.3 | 90.7 | 85.4 |
| Additionally allowed | 8.6 | 8.2 | 12.9 |
| Generously allowed | 0.0 | 0.0 | 0.6 |
| Disallowed[e] | 1.1 | 1.1 | 1.1 |

| Data collection | M8V$_L$(no DIR) | M8V$_L$+DIR | M8V$_L$S$^{L55}$P+DIR |
|---|---|---|---|
| **R.M.S. deviations** | | | |
| Bond lengths (Å) | .006 | .005 | .011 |
| Angles (Å) | 1.12 | 1.30 | 1.45 |

[a] Numbers in parentheses are for the highest resolution shell of data.

[b] $R_{sym} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$

[c] $R_{cryst} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$

[d] $R_{free}$ is the same as $R_{cryst}$, but for 5% of the data excluded from the refinement

[e] Residue Asn$^{51}$, which is in a conserved γ turn and is almost always found in this region in antibody structures.

## Table 4

Total contacts (van der Waals and hydrogen bonds) from $M8V_L$ and $M8V_LS^{L55}P$ to DIR (alternate conformation 1/2). Residues in the CDR loops are demoted with asterisks

| Residue | $M8V_L$ | | $M8V_LS^{L55}P$ | |
|---|---|---|---|---|
| | Protein Chain A | Protein Chain B | Protein Chain A | Protein Chain B |
| $^{*}Tyr^{L34}$ | 5/10 | 8/5 | 6/6 | 7/7 |
| $Val^{L46}$ | 3/1 | 1/2 | 0/0 | 0/4 |
| $Tyr^{L49}$ | 8/21 | 16/3 | 10/20 | 6/2 |
| $^{*}Asn^{L50}$ | 0/2 | 1/0 | 0/0 | 0/0 |
| $^{*}Arg^{L53}$ | 0/8 | 2/0 | 0/3 | 0/0 |
| $^{*}Ser/Pro^{L55}$ | 3/0 | 0/3 | 3/0 | 0/7 |
| $^{*}Ser^{L56}$ | 5/0 | 0/12 | 4/0 | 0/9 |
| $^{*}Leu^{L89}$ | 2/2 | 2/4 | 4/4 | 2/2 |
| $^{*}Trp^{L96}$ | 4/5 | 5/3 | 5/5 | 5/5 |
| $Phe^{L98}$ | 2/2 | 0/0 | 0/0 | 1/7 |
| Total | 32/51 | 35/32 | 32/38 | 21/37 |

$^{(*)}$Contacts calculated with Contacsym (58).