Structural and thermodynamic analysis of the GFP:GFP-nanobody complex

Marta H. Kubala,[†] Oleksiy Kovtun,[†] Kirill Alexandrov,* and Brett M. Collins*

Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia

Received 11 April 2010; Revised 13 September 2010; Accepted 19 September 2010
DOI: 10.1002/pro.519
Published online 13 October 2010 proteinscience.org

Abstract: The green fluorescent protein (GFP)-nanobody is a single-chain V_HH antibody domain developed with specific binding activity against GFP and is emerging as a powerful tool for isolation and cellular engineering of fluorescent protein fusions in many different fields of biological research. Using X-ray crystallography and isothermal titration calorimetry, we determine the molecular details of GFP:GFP-nanobody complex formation and explain the basis of high affinity and at the same time high specificity of protein binding. Although the GFP-nanobody can also bind YFP, it cannot bind the closely related CFP or other fluorescent proteins from the mFruit series. CFP differs from GFP only within the central chromophore and at one surface amino acid position, which lies in the binding interface. Using this information, we have engineered a CFP variant (I146N) that is also able to bind the GFP-nanobody with high affinity, thus extending the toolbox of genetically encoded fluorescent probes that can be isolated using the GFP-nanobody.

Keywords: GFP; nanobody; ITC; thermodynamics; heat capacity; X-ray crystallography

Introduction

One of the main challenges of 21st century biology is to develop a comprehensive toolbox for selective detection, isolation, and analysis of proteins in or from their native environment and complex mixtures. Although much effort has been expended on the development of such probes based on small molecules, the requirement for their dedicated synthetic production and difficulties in designing of highly specific compounds complicates their general application. At the same time, the genetically encoded binders or reporter domain-based probes have emerged as a dominant theme in all segments of biological research. Fluorescent proteins have revolutionized cell biology and biochemistry by providing easy to use genetically encodable fluorescent protein tags.[1] In a parallel development generation and selection of protein scaffold libraries yielded a range of selective and tight binders to a range of research and pharmaceutical protein targets.[2] One such development is the advancement of single-chain antibodies displaying simpler selection and improved stability, solubility, and yield.[3]

The Camelid species possess unusual heavy-chain IgG antibodies, devoid of light-chains and containing a single antigen binding variable domain (V_HH).[3] This V_HH domain is also referred to as a nanobody, renamed by the company developing these domains as novel therapeutics, Ablynx. A nanobody is strictly monomeric, highly stable, and generally smaller than the V_H variable domain of a classical antibody and can be readily expressed in heterologous systems such as bacteria at high levels. Although the paratope of a classical heterodimeric antibody is expected to be larger than a nanobody due to the additional presence of three complementarity determining regions (CDRs) of the V_L light chain, longer CDR loops in the nanobody domains

[†]Marta H. Kubala and Oleksiy Kovtun contributed equally to this work.

Grant sponsors: University of Queensland, Australian Research Council (ARC) Future Fellowship; Australian National Health and Medical Research Council (NHMRC) R. Douglas Wright Career Development Award (401626); Grant sponsor: Australian Research Council; Grant number: DP0986496..

**Correspondence to:* Kirill Alexandrov, Institute for Molecular Bioscience, University of Queensland, Brisbane 4072, Australia. E-mail: k.alexandrov@imb.uq.edu.au (or) Brett M. Collins, Institute for Molecular Bioscience, University of Queensland, Brisbane 4072, Australia. E-mail: b.collins@imb.uq.edu.au

combined with greater structural diversity are expected to compensate for this to some extent.

Typically an antigen-specific nanobody is generated after immunization of a llama or dromedary, and the V_HH repertoire is cloned from a B lymphocyte RT-PCR library into a phage display vector. Several rounds of panning are then used to select the highest affinity binding nanobodies, which is far more efficient than for classical antibodies using heterodimeric V_H–V_L pairs. Nanobodies are now being used in many different applications, both for therapeutic and research purposes. Two interesting recent developments have been their usage as affinity tags for isolating and studying the localization of proteins in cells,[4–7] and as stabilizing agents in protein structural studies.[8–14]

Recently nanobodies have been developed to specifically recognize the green fluorescent protein (GFP).[5,6] In the first instance, this was fused to the red fluorescent protein (RFP) to target this second chromophore against GFP-tagged proteins expressed in HeLa cells.[6] Subsequent studies have utilized the GFP-targeted nanobody (herein referred to as GFP-nanobody) to isolate various GFP-tagged proteins from cell extracts and redirect them to alternative cellular locations,[5] and to alter plant phenotypes by trapping GFP-fusion proteins.[7] In this study, we have used X-ray crystallography and isothermal titration calorimetry (ITC) to determine the molecular details of GFP:GFP-nanobody complex formation and determine the possible routes for redirecting its specificity to other fluorescent proteins.

Results

Structure determination of the GFP:GFP-nanobody complex

The high specificity and affinity of the GFP-nanobody interaction with GFP allows quantitative isolation of the latter from crude cell and tissue lysates enabling a range of biochemical and proteomic applications. Although both GFP and YFP are recognized by the GFP-nanobody, the related fluorescent proteins such as CFP, mFruit series, and other red-shifted proteins are not.[5,7,15] To understand the basis for specificity of the GFP-nanobody for GFP, we decided to perform a structural and biophysical analysis of the GFP:GFP-nanobody interaction. The GFP:GFP-nanobody complex was assembled by mixing GFP with excess GFP-nanobody and subjecting the mixture to gel filtration. Gel filtration analysis indicated that a homogeneous 1:1 assembly of GFP:GFP-nanobody was formed (not shown), and the resulting complex was subjected to crystallization trials as described in materials and methods. Several conditions yielded crystals, and here, we determined the structure of the GFP:GFP-nanobody complex using crystals of spacegroup $P2_12_12$, and data collected in-house to a resolution of 2.8 Å. The structure was determined by molecular replacement using separate chains of GFP and an engineered antibody fragment as search models (Fig. 1, Table I).

Structure of the GFP:GFP-nanobody complex

The conformations of both GFP and the GFP-nanobody undergo only limited structural changes upon binding compared with structures of GFP alone and previously characterized nanobodies, outside of the expected differences at terminal residues and within the three CDRs of the nanobody (CDR1-3). Bound and unbound GFP have an r.m.s.d of 0.3 Å over 224 Cα atoms. Cys22 and Cys97 of V_HH domains typically form an intradomain disulphide bridge [Fig. 1(B)]. Our data provide evidence for the presence of a mixed species of GFP-nanobody with oxidized and reduced cysteines probably due to the presence of reducing agent DTT in the protein solution. Unusually stable toward proteolysis and denaturation, GFP is an 11-stranded β-barrel wrapped around a single central helix where each strand consists of 9–13 residues. The chromophore in classical GFP is formed from the central helix by spontaneous cyclization of residues Ser65, Tyr66, and Gly67.[16] For the purpose of this study, we used the so-called enhanced GFP (eGFP), a GFP mutant with increased brightness, which instead of Ser65, has Thr65 [Fig. 1(C)].

The typical fold of a nanobody contains a single antigen binding variable domain (V_HH) and exhibits the classical immunoglobulin fold with a scaffold of nine antiparallel β-strands forming two sandwiching β-sheets. Each nanobody has three hypervariable loops called CDRs, which are the primary regions that undergo genetic alteration to present an antigen-specific binding surface.[3] In the complex, all the three CDRs of the GFP-nanobody make specific contacts with GFP, although the main contacts are observed between CDR3 of the nanobody and β-strands 8–11 of GFP. [Fig. 1(B,C)]. An expanded view of the binding interface is presented in Figure 2. Hydrophilic, hydrophobic, and electrostatic interactions present at the binding interface determine high specificity and at the same time high affinity of complex formation. Hydrophobic interactions between Phe102 of GFP-nanobody and Leu221, Ala206, and Phe223 of GFP as well as between Trp47 of GFP-nanobody with Val176 and Ser175 of GFP influence high affinity of the complex. Simultaneously, two buried salt-bridges, between GFP Arg168 and GFP-nanobody Glu103 and GFP Glu142 and GFP-nanobody Arg35, are critical for the specificity of complex formation. Arg168 of GFP additionally forms a hydrogen bond with the hydroxyl of Tyr37 of GFP-nanobody and is, therefore, particularly important to the orientation of the binding interface.

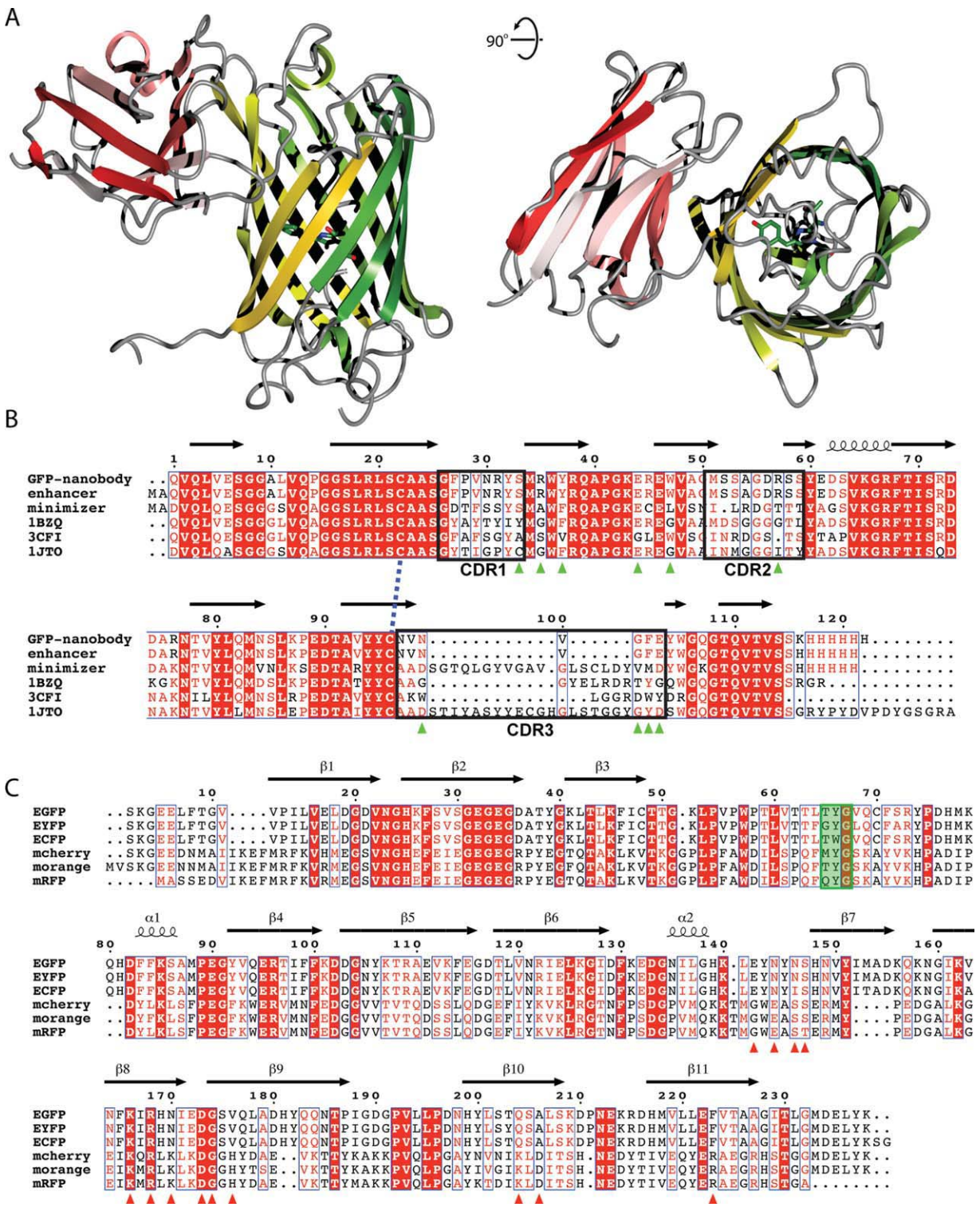

Figure 1. Structure of the GFP:GFP-nanobody heterodimeric complex determined by X-ray crystallography. (A) Ribbon diagram showing perpendicular views of the GFP:GFP-nanobody complex. GFP is shown in green to gold and GFP-nanobody is shown colored red to pink (N- to C-terminal). The GFP chromophore is shown in stick representation. (B) Sequence alignment of the GFP-nanobody with other camelid nanobodies of known structure. The CDR regions are boxed, and the characteristic disulfide bridge indicated. Side-chains that directly contact GFP are indicated by green triangles. (C) Sequence alignment of GFP with other fluorescent proteins. The three residues that are cyclized to form the chromophore are boxed in green. Side-chains that make direct contact with the GFP-nanobody are indicated by red triangles.

Table I. *Summary of Crystallographic Structure Determination Statistics*[a]

	GFP:GFP-nanobody complex
Data collection	
Space group	$P2_12_12$
Cell dimensions	
a, *b*, *c* (Å)	140.1, 147.6, 101.6
α, β, γ (°)	90, 90, 90
Resolution (Å)	42.42–2.80 (2.90–2.80)
R_{merge}	0.080 (0.267)
$I/\sigma I$	11.8 (2.9)
Completeness (%)	91.8 (72.6)
Redundancy	4.5 (2.6)
Wilson plot *B* (Å)	39.9
Refinement	
Resolution (Å)	42.4–2.80 (2.90–2.80)
No. reflections/No. R_{free}	45703/2300 (3085/159)
R_{work}/R_{free}	0.204/0.253 (0.268/0.337)
No. atoms	
Protein	10,900
Solvent	312
Average *B*-factor	34.6
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.18

[a] Highest resolution shell is shown in parentheses.

Comparison with other GFP:nanobody structures

A number of other structures have been reported of nanobodies in complex with different ligand proteins. In some cases, these structures have been determined to better understand the mode of binding used by different nanobody V_HH domains, and in others the nanobodies have been used as specific tools to assist in the crystallization process, by rigidifying ligand structures and providing alternative sites for crystal contact formation.[8–14] Several representative structures are shown in Figure 3 including the EpsI:EpsJ:nanobody complex[10] [Fig. 3(B)], the RNaseA:nanobody complex[14] [Fig. 3(C)], and the lysozyme:nanobody complex[17] [Fig. 3(D)]. Sequences of these nanobodies are provided in Figure 1(B). The most notable difference between the GFP-nanobody structure and these other nanobodies is the length and the conformation of the CDR3 loop region. In most other reported structures, this region is significantly extended and folds back over the framework 2 region of the nanobody. In classical antibodies, this region would provide the interface for the light chain VL. In the GFP-nanobody structure, the CDR3 loop is very short and makes significantly fewer contacts with the GFP ligand compared with other nanobodies. Furthermore, the shortness of this CDR3 loop exposes the framework 2 region, which in turn makes an unusually major contribution to the binding interface with GFP.

In our structure, four copies of the heterodimeric complex are present in the asymmetric unit. As shown in Figure 4(A), there is no significant difference in the structures of the four copies of the complex in the asymmetric unit, with less than 0.52 Å r.m.s.d. over 338 Cα atoms, confirming that the interface is relatively rigid in its orientation. Of greatest direct relevance to the structure reported here are the recent structures of GFP in complex with camelid V_HH nanobodies that have been shown to modulate the fluorescence of the GFP molecule.[4] In this study by Kirchhofer *et al.*, several GFP-binding nanobodies were identified and screened for their ability to affect GFP fluorescence. One nanobody was found to increase GFP fluorescence intensity by ~50% and was named the GFP-enhancer, whereas a second nanobody was found to reduce GFP fluorescence by ~75% and was named the GFP-minimizer.[4] The

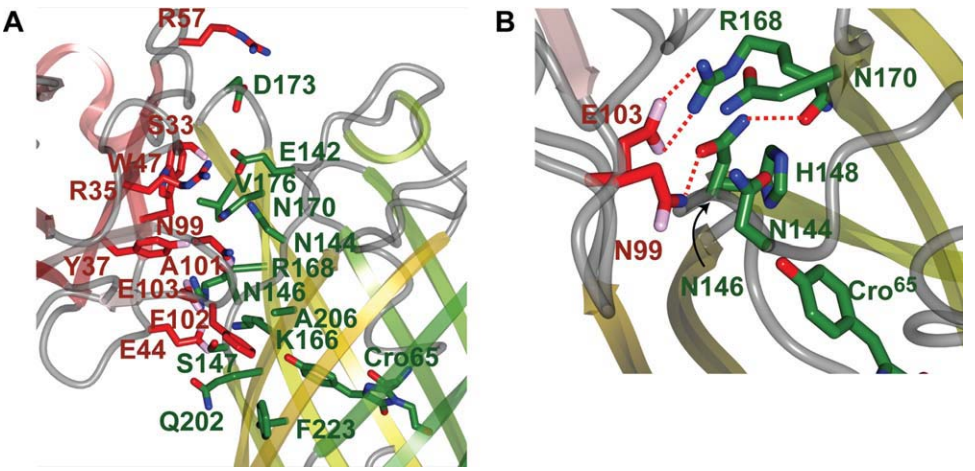

Figure 2. Details of the GFP:GFP-nanobody interface. (A) View of the binding interface. Residues participating in the interface are shown as sticks. Proteins are colored as in Figure 1. (B) Close up showing details of the environment surrounding GFP Asn146. Asn146 is the only side-chain in the interface that varies between GFP and CFP (Ile146 in CFP) [Fig. 1(D)] and, therefore, determines the nanobody specificity. Hydrogen bonds are indicated with dashed red lines.

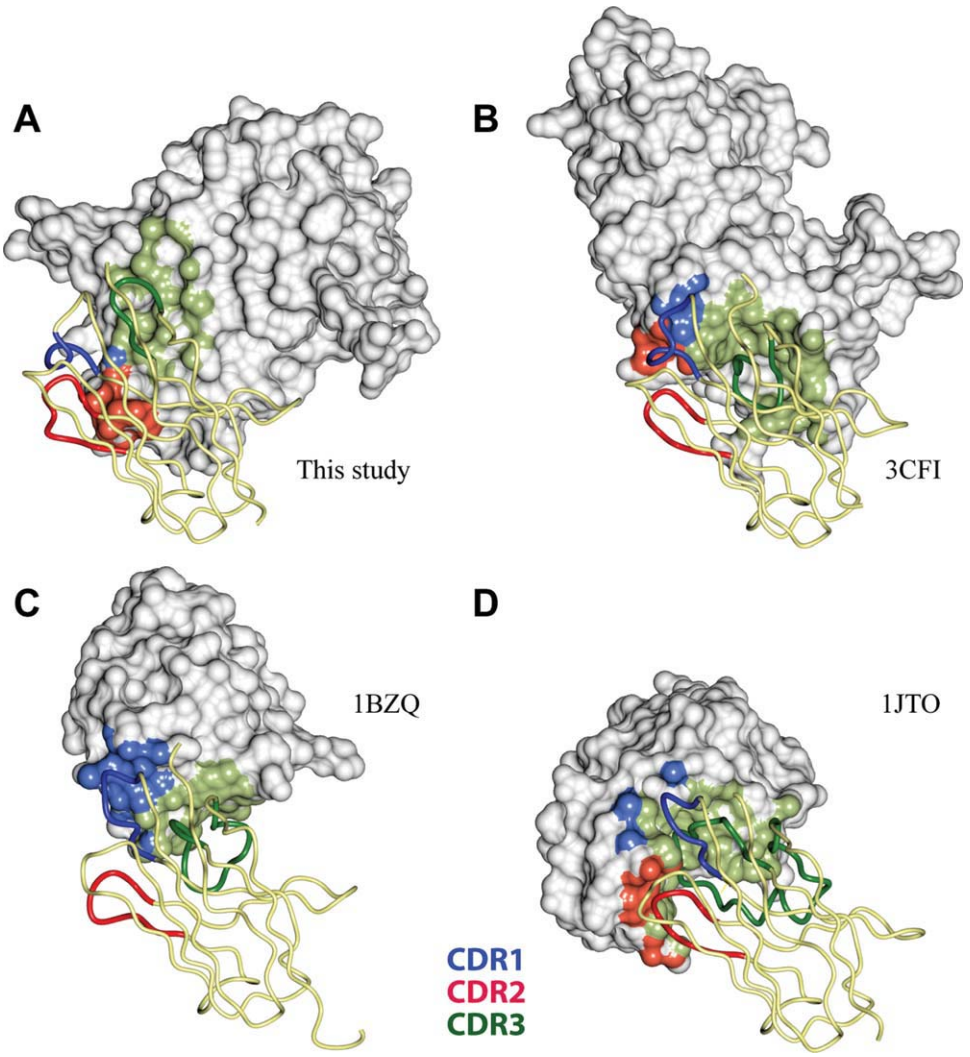

Figure 3. Comparison of the GFP:GFP-nanobody complex with other representative camelid nanobody protein complexes. The representative sequences of each nanobody are given in Figure 1(B). In each structure the nanobody is shown as a yellow worm with CDR1 colored blue, CDR2 colored red, and CDR3 colored green. The bound proteins are shown as surface representations with atoms that contact each CDR loop colored, respectively. (A) The GFP:GFP-nanobody complex. (B) The EpsI:EpsJ:nanobody complex (PDB ID 3CFI[10]). (C) The RNaseA:nanobody complex (PDB ID 1BXQ[14]). (D) The lysozyme:nanobody complex (PDB ID 1JTO[17]).

structures of the GFP:GFP-enhancer and GFP:GFP-minimizer complexes are shown in Figure 4(B,C) respectively, and compared with the GFP:GFP-nanobody complex. As can be seen in Figure 4(B), the GFP-enhancer V_HH domain adopts the same binding orientation to the GFP-nanobody reported here, and in fact when the sequences are compared the two proteins are identical [Fig. 1(B)] (although notably the crystal forms are different). As discussed by Kirchhofer *et al.*, the binding of GFP by the GFP-enhancer/GFP-nanobody molecule acts to stabilize GFP Arg168 (in particular through a direct salt bridge to GFP-nanobody/GFP-enhancer side-chain Glu103) in an orientation that positions GFP His148 in very close proximity to the chromophore hydroxyl. This was proposed to improve proton extraction, which results in brighter fluorescence. In contrast the GFP-minimizer V_HH domain binds in a completely different orientation to the GFP-nanobody/GFP-enhancer, although it does utilize an overlapping GFP surface that results in competitive binding between the two nanobodies[4] [Fig. 4(C)].

Kinetics of the GFP:GFP-nanobody interaction measured by biolayer interferometry

To estimate kinetic parameters of GFP-GFP-nanobody interactions, interferometry using the Octet RED analyzer was used. GFP-nanobody-coated sensors were exposed to GFP dilutions for 6 min to measure association of the complex and then to the buffer without GFP for 18 min to acquire a dissociation time-course [Fig. 5(A)]. The measured K_d of

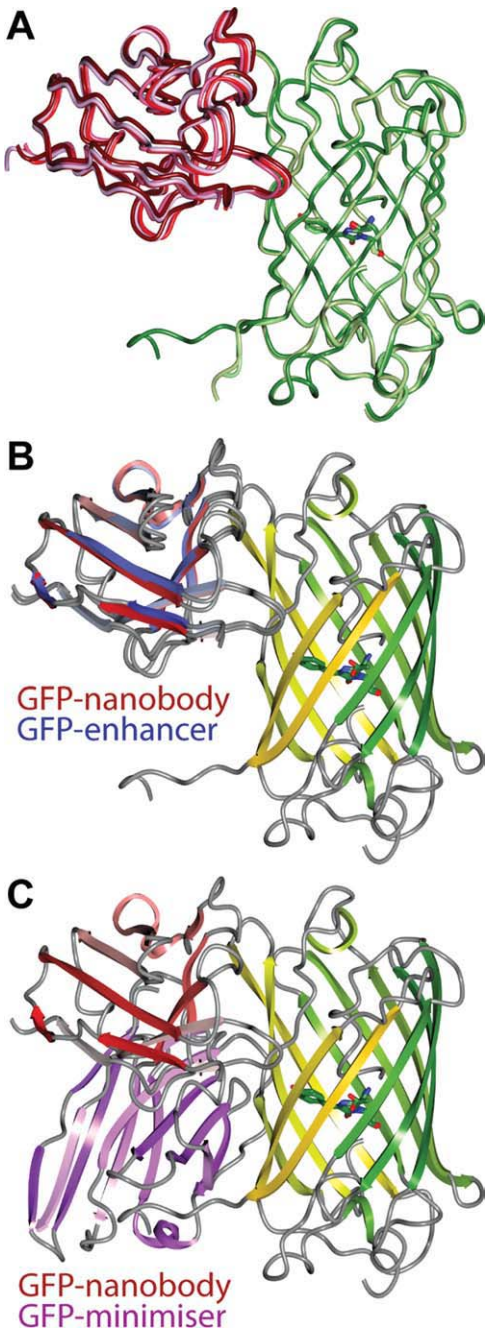

Figure 4. Comparison of the GFP:GFP-nanobody complex with GFP-enhancer and GFP-minimizer structures.[4] (A) In the GFP:GFP-nanobody crystals four copies of the complex are present in the asymmetric unit. An overlay is shown of the four copies as Cα traces, with GFP in different shades of green and GFP-nanobody in shades of red. There is no significant difference between the four subunits. (B) Comparison of the GFP:GFP-nanobody structure (this study) colored as in Figure 1(A) with the GFP:GFP-enhancer structure (PDB ID 3K1K[4]). Only GFP from this study is shown for clarity, whereas the GFP-enhancer is shown in blue. (C) Comparison of the GFP:GFP-nanobody structure (this study) colored as in Figure 1(A) with the GFP:GFP-minimizer structure (PDB ID 3G9A[4]). Only GFP from this study is shown for clarity, while the GFP-minimizer is shown in purple.

1.4 n*M* is very similar to those reported previously (0.23–0.59 n*M*) (Table II). The small difference is mainly related to a slower apparent association rate and could be due to a number of reasons, for example, interference caused by the biotin-labeling or differences caused by the experimental setups using either microfluidics or tips inserted into a stationary solution followed by mixing.

Thermodynamics of GFP interaction with GFP-nanobody

To examine the thermodynamics of GFP:GFP-nanobody complex formation, we have used ITC [Fig. 5(B)]. Because of the high affinity of association (K_d ~1 n*M*), it is difficult to use ITC to measure the equilibrium binding constant with precision, however the enthalpy ΔH and stoichiometry of binding are very well determined. Measurements performed at different temperatures can also be used to determine the heat capacity of binding ΔC_p ($\delta\Delta H/\delta T$) to better understand the thermodynamic basis of the high affinity and specificity of complex formation. Thus, thermodynamic binding parameters that provide important information about the nature of the binding process can still be determined by ITC even when accurate affinity determination is not possible. Binding of GFP to the GFP-nanobody yields an exothermic binding isotherm in the temperature ranges examined here, and a stoichiometry of association of 1 as expected from the gel filtration and crystallographic analyses. The enthalpic contribution to binding showed a linear dependency on temperature [Fig. 5(C); Table II], ranging from -6.5 kcal mol^{-1} at 10°C to -13.2 kcal mol^{-1} at 37°C. The heat capacity change of the binding process was calculated from the slope of the graph of ΔH versus temperature and has a value of -239 cal mol^{-1} K^{-1}. A negative ΔC_p is indicative of a significant degree of hydrophobic association and is typical of most protein–protein interactions.

ΔC_p is a thermodynamic parameter that is being increasingly used in structure-thermodynamics studies to probe the nature and extent of biomolecular binding interfaces.[18,19] Paramaterization of binding interfaces has lead to several empirical formulations for calculating the expected binding ΔC_p based on the proportion of buried hydrophobic and polar surfaces.[19] In relation to such studies it is critical that improved datasets of high-resolution structures of protein complexes with accurately determined thermodynamic binding parameters are constructed to develop and improve on the current predictive models.[20] The program NACCESS was used to calculate the solvent accessible surface areas (SASA) of GFP, GFP-nanobody, and the GFP:GFP-nanobody complex to obtain changes in SASA (ΔSASA) that occur upon complex formation (Table III). A surface representation of the complex is shown in Figure

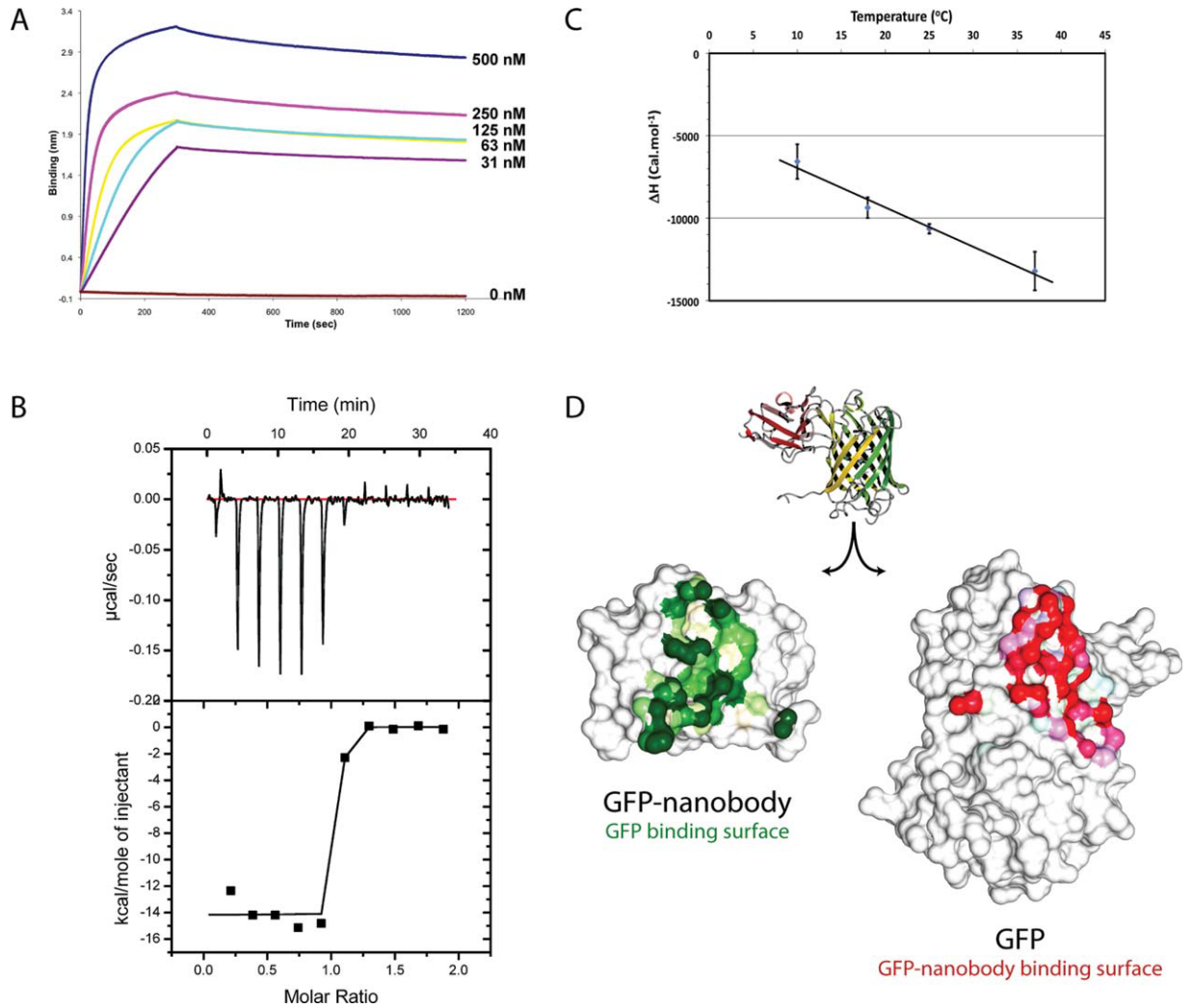

Figure 5. Interaction analysis of the GFP:GFP-nanobody complex. (A) Binding of GFP to biotinylated GFP-nanobody measured using an Octet RED interferometer. Depicted sensorgrams represent complex formation (first 300 s) at 500, 250, 125, 62.5, 31.25, and 0 n*M* of GFP and subsequent dissociation of the complex in binding buffer without GFP (300–1200 s). (B) ITC binding thermogram of GFP binding to GFP-nanobody. GFP-nanobody (47 µ*M*) was titrated into GFP (5.2 µ*M*) at 37°C. The top panel shows the raw data, whereas the bottom panel shows the integrated normalized data for binding. (C) Graph of Δ*H* vs. temperature for the interaction of GFP with the GFP-nanobody. (D) Cut-away surface representations show the binding interfaces of GFP and GFP-nanobody. GFP is colored with red to highlight contact with GFP-nanobody, whereas GFP-nanobody is conversely colored green to indicate contact with GFP.

5(D), with the surfaces of each molecule that are buried upon complex formation indicated. This analysis revealed a total surface area of 1374 Å^2 is desolvated upon complex formation, composed of ~50% apolar (696 Å^2) and 50% polar surface area (677 Å^2). A significant hydrophobic interface is typical of most protein–protein interactions and is consistent with a negative measured ΔC_p of −239 cal mol^{-1} K^{-1}. The relatively large extent of the polar surface proportional to the overall binding interface however is unusual but is well correlated with the relatively low magnitude of the measured ΔC_p value. The high proportion of polar and electrostatic buried surface area is highly unusual for a protein–protein binding interface, where polar contacts typically constitute on the order of one-third of the binding surface area.[21] Using previous empirically derived formulations for determining ΔC_p from the extent of buried hydrophobic and polar surfaces, we calculate a predicted ΔC_p in the range of −127 to −213 cal mol^{-1} K^{-1}, which is lower in magnitude than the measured value from ITC.

One possible explanation for this discrepancy is burial of ordered waters, which are expected to contribute on the order of −18 cal mol^{-1} K^{-1} per buried water molecule.[22] In our model, six water molecules are present at the complex interface, which is close to the average for an interface of ~690 Å^2 (Ref. 23) but may be an underestimate taking into consideration the moderate resolution of our structure (2.8 Å). These could certainly influence the measured ΔC_p. A second possibility is that regions of the bound

Table II. *Thermodynamic and Kinetic Binding Data for the GFP-GFP-Nanobody Complex*

ITC[a]			
Temperature (K)	N	ΔH (kcal mol^{-1})	ΔC_p (cal mol^{-1} K^{-1})
283	0.98 ± 0.11	−6.6 ± 1.0	−239
291	0.96 ± 0.04	−9.4 ± 0.6	
298	0.96 ± 0.03	−10.6 ± 0.3	
310	0.96 ± 0.05	−13.2 ± 1.2	
Interferometry using Octet Biosensor			
k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_d (k_{off}/k_{on}) (nM)	
8.84×10^4	1.24×10^{-4}	1.40	
Surface plasmon resonance			
k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_d (k_{off}/k_{on}) (nM)	
No data	No data	0.32[b]	
7.68×10^5	1.74×10^{-4}	0.23[c]	
Quartz crystal microbalance[d]			
k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_d (k_{off}/k_{on}) (nM)	
2.45×10^5	1.45×10^{-4}	0.59 ± 0.11	

[a] All ITC measurements are quoted as the average of three independent experiments ± standard deviation.
[b] Ref. 30.
[c] Ref. 6.
[d] Ref. 4.

molecules undergo localized rigidification upon complex formation,[24,25] which is an attractive possibility given the nature of the flexible CDR binding regions of the GFP-nanobody.

Engineering of the CFP protein for binding to the GFP-nanobody

Previous experiments have determined that the GFP-nanobody can also bind to YFP but not CFP, RFP, or mCherry fluorescent proteins, a specificity that is important in many different applications.[5,7] The sequence of RFP, mCherry, and other mFruit proteins are substantially different to that of GFP, CFP, and YFP, and it is clear that the GFP-nanobody is unable to associate with these molecules because residues critical for association are poorly conserved [Fig. 1(C)]. YFP and GFP are conserved in all residues involved in the GFP-nanobody interface explaining the dual specificity of the GFP-nanobody. Of all specificity-determining side-chains in the GFP structure, the only substitution in CFP is the alteration of Asn146 to Ile [Fig. 1(C)]. In Figure 2(B), a close-up of this side-chain is shown in its local environment. The GFP Asn146 Oδ atom participates in a hydrogen bond with the GFP-nanobody Asn99 Nδ atom. Critically GFP Asn146 is closely packed within a buried hydrophilic environment, involving the salt bridge between GFP Arg168 and GFP-nanobody Glu103, as well as GFP Asn170 and GFP Asn144 side-chains. In addition, GFP Asn146 is stabilized by a hydrogen bond of its Nδ with the carbonyl oxygen of GFP Arg168. This explains why substitution of Asn146 to hydrophobic Ile disrupts formation of the hydrophilic patch and, therefore, causes loss of the specificity of GFP-nanobody to bind CFP. We predict therefore that mutation of CFP Ile146 to Asn will allow binding of the GFP-nanobody. To test this we used a novel *in vitro* translation system based on lysate derived from the eukaryotic protozoan *Leishmania tarantolae*.[26] Initially, we tested whether the mutation of either GFP or CFP at position 146 altered their spectral properties. The normalized emission spectra of *in vitro* translated GFP, GFP(N146I), CFP, and CFP(I146N) are shown in Figure 6(A), whereas the overall fluorescence intensity of each protein is shown in Figure 6(B). We find that mutation of the GFP protein at position 146 has limited impact on the spectral properties of GFP but does decrease the overall fluorescence intensity. This mutant showed as expected a dramatic reduction in binding to immobilized GFP-nanobody in pulldown experiments [Fig. 6(C,D)]. Conversely, we found that mutation of CFP(I146N) reversed its inability to associate with the GFP-nanobody, confirming this side chain as being critical for the specificity of association [Fig. 6(C,D)]. Although CFP(I146N) displayed a slightly altered emission spectrum, with ablation of the typical double maxima observed in the wild-type CFP protein,[27] and also a partially reduced fluorescence intensity, which is expected considering this mutation was originally isolated based on its improved emission intensity,[27] its spectral properties indicate it should still be a suitable construct for CFP fluorescence experiments.

Discussion

Given the emerging importance of the GFP-nanobody as a research tool, we have solved the structure of the GFP:GFP-nanobody complex and characterized the thermodynamics of complex formation, to better understand the basis of the high affinity and specificity of this interaction. Hydrophobic interactions at the binding interface of two proteins generally contribute strongly to stabilization and thus the affinity of complex formation, whereas polar interactions also help to stabilize association while directing recognition specificity.[28,29] Hydrophobic interactions are significantly responsible for the high affinity of the GFP:GFP-nanobody complex, in particular involving the burial of the GFP-nanobody Trp47 side-chain. But the burial of hydrophilic side chains and salt bridges responsible for the spatial alignment of the two molecules almost certainly explains the high degree of binding specificity. The examination of the surface of the GFP:GFP-nanobody complex revealed that 50.7% of the binding surface is composed of hydrophobic interactions and 49.3% made up of polar interactions. The negative but low magnitude ΔC_p calculated from the ΔH vs. temperature data obtained from ITC experiments is well correlated

Table III. *Area Coefficients and Heat Capacity Calculations for the Interaction of GFP with GFP-Nanobody*[a]

Source references	Data set	Δc_{ap} (cal/mol/K/Å^2)	Δc_p (Cal/mol/K/Å^2)	Calculated heat capacity change (cal/mol/K)
Spolar *et al*.	12 proteins	0.32	−0.14	−127
Murphy and Friere	Cyclic dipetides	0.45	−0.26	−136
Myers *et al*.	26 proteins	0.28	−0.09	−133
Makhatadze and Privalov	20 proteins	0.51	−0.21	−213
Robertson and Murphy	49 proteins	0.15	0.12	−194
Experimentally determined this study				−238

[a] Using NACCESS (http://www.bioinf.manchester.ac.uk/naccess/) we calculated the ΔSASA upon GFP-nanobody binding to GFP overall and for apolar and polar atoms as −1374 Å^2, −696 Å^2, and −677 Å^2, respectively. These values were used to calculate expected $\Delta C_p = \Delta c_{ap}.\ \Delta SASA_{ap} + \Delta c_p.\ \Delta SASA_p$ where Δc_{ap} and Δc_p represent empirically determined area coefficients as provided in Ref. 19.

with the unusually high proportion of polar and electrostatic contacts between the two proteins. Of interest to the empirical paramaterization of binding surfaces to calculate thermodynamic binding quantities,[19] we find that previous empirical formulations significantly underestimate the binding ΔC_p. This discrepancy may be explained by burial of ordered water, conformational rigidification of GFP-nanobody loops or a combination of both and, additionally, suggests that larger and improved datasets of carefully correlated thermodynamic binding data and high-resolution structural information are required to improve the accuracy of such predictions.

Taking into consideration, the possible applications of the GFP-nanobody in the context of protein expression and purification strategies of fluorescently tagged proteins we were particularly interested if other fluorescent proteins that do not bind the GFP-nanobody could be rationally engineered to create additional nanobody-binding molecules. Such proteins would provide additional tools for the

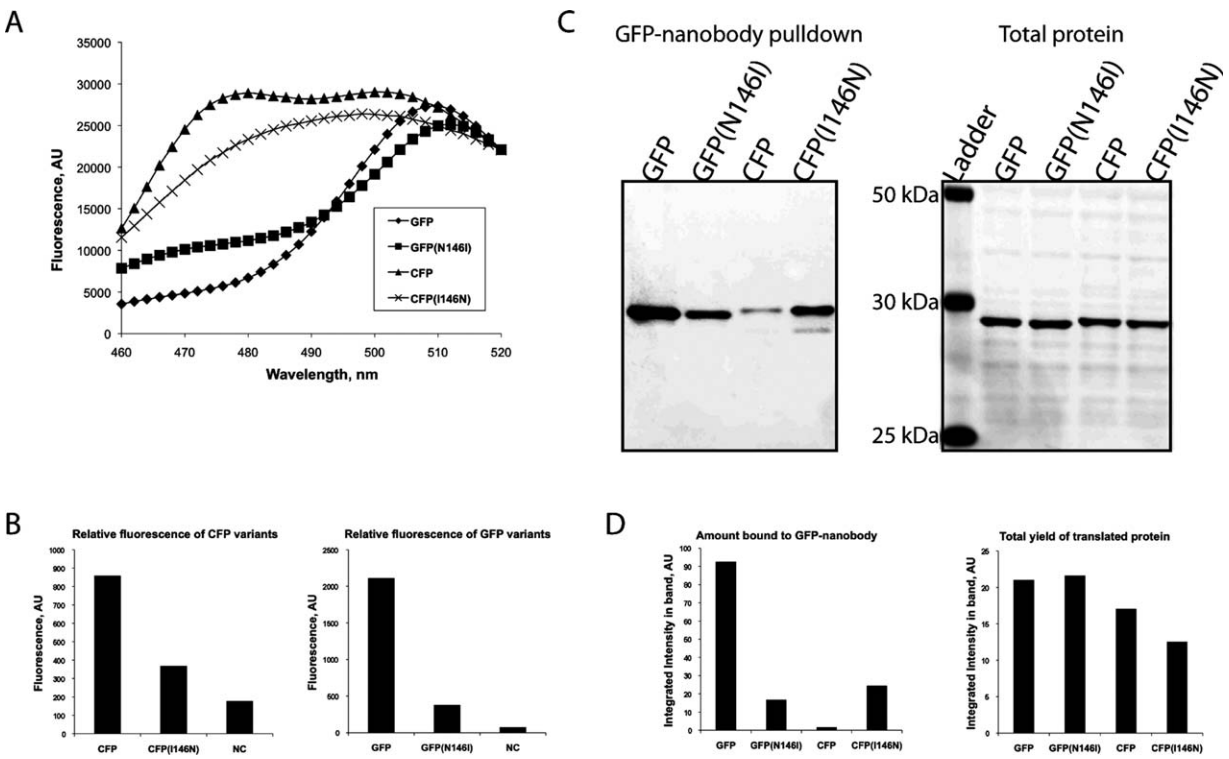

Figure 6. Engineering of CFP for GFP-nanobody binding. (A) Normalized fluorescence emission spectra for GFP, CFP, and mutant proteins after excitation at 430 nm. (B) Fluorescence intensity of GFP, CFP, and mutants measured using 430Ex/485Em and 485Ex/528Em filter sets, respectively. (C) Affinity precipitation of *in vitro* translated GFP, CFP, and their mutants were performed using sepharose-coupled GFP-nanobody, and the eluates were analyzed by SDS-PAGE followed by anti-GFP/CFP Western blotting (D) Quantification of gels shown in Figure 6(C) by IR densitometry (Odyssey Imaging System).

isolation and characterization of fluorescent fusion proteins. The structures of fluorescent proteins including GFP, YFP, CFP, mRFP, mOrange, and mCherry are characterized by highly similar backbone folds, but sequence comparison shows less than 22% identity for the mFruit derivatives compared with GFP-related proteins. The GFP-nanobody does not bind the mFruit derivatives, due to the significant sequence divergence of these molecules, nor does it bind CFP[5,7] (Fig. 5), which differs from GFP and YFP at only one amino acid position out of all those that take part in formation of the binding interface. The question arises as to whether it will be possible to create novel binding variants of either the GFP-nanobody or other fluorescent proteins using structure-based engineering strategies. Rational mutagenesis of the GFP-nanobody surface to create different specificities or to alter its affinity is certainly possible, but it is not entirely clear what mutations should be pursued to alter its binding properties, and most importantly whether this would be preferable to the alternative approach of generating novel nanobodies by established immunization and panning procedures. Mutation of mCherry and related proteins to interact with GFP-nanobody instead of going through the selection process to obtain anti-mCherry nanobodies would certainly be desirable for the generation of novel tools for protein localization and isolation. However, out of 12 GFP residues interacting with the nanobody, seven are different in mCherry. Thus, converting mCherry (or other mFruit proteins) into a form compatible with the GFP-nanobody variant will be complicated and may not be possible without significantly altering its properties.

However, we find that in the case of the nonbinding fluorescent protein CFP, a single substitution of Ile146 for Asn is sufficient for promoting a GFP-nanobody:CFP interaction. Despite some alterations of the spectral properties of this mutant our analysis indicates it should still be suitable for most fluorescent applications that would normally utilize the canonical CFP molecule. In summary, we have defined the structural details leading to the specificity of the GFP-nanobody for GFP and YFP, and this has led to the creation of a modified CFP as an additional new tool that may also be used in conjunction with the GFP-nanobody for the targeted isolation and characterization of genetically encoded fluorescent molecules.

Materials and Methods

Protein expression and purification

The amino acid sequence of GFP-nanobody (V_HH domain cAbGFP4[30]) was reverse translated into genetic code optimized for *E. coli*-specific codon usage and cloned into pOPINE vector using NcoI/PmeI nucleases, giving the nanobody sequence with a C-terminal his-tag (Epochbiolab, USA). The GFP sequence (Clontech) was inserted into pOPINE backbone by using infusion cloning, which resulted in addition of $MA(H)_6SSGGS$ peptide to the N-terminus of GFP. Protein expression was conducted in *E. coli* strain BL21-GOLD (Stratagene) in LB medium. The cell culture was propagated to OD_{600} 0.5 at 37°C, and then, protein synthesis was induced by 0.5 mM of IPTG. Further fermentation was carried out at 20°C for 20 h. Resultant cell mass was harvested by centrifugation, disrupted using a fluidizer (Constant Systems, UK), and subjected to centrifugation to remove cell debris. The cleared cell lysate was subjected to IMAC chromatography (HisTrap FF column) followed by size-exclusion fractionation (Superdex 75) using an Akta Purifier FPLC system (GE Healthcare). The purified proteins were more than 95% pure according to electrophoretical analysis.

To isolate the complex, purified GFP was mixed with twofold molar excess of GFP-nanobody, incubated for 15 min at ambient temperature and loaded onto a gel-filtration column (Superdex 75 16/60) preequilibrated with 20 m*M* HEPES-KOH (pH 7.5), 150 m*M* NaCl. The complex was eluted as a symmetrical peak corresponding to the molecular weight of a 1:1 heterodimer and was well separated from the peak of the unbound GFP-Nanobody. The solution of purified complex was concentrated by ultrafiltration (Amicon) to 24 mg/mL before crystallization.

Protein crystallization

For crystallization experiments, the purified complex of GFP and GFP-nanobody proteins was concentrated by ultrafiltration to 24 mg/mL in 20 m*M* Hepes pH 7.5, 0.15*M* NaCl. Protein solution was flash-frozen in liquid nitrogen and stored at 193 K. Particulates were removed by centrifugation (13,200 rpm for 20 min at 277 K) immediately before setting up crystallization experiments. Crystals were grown by the sitting-drop vapor diffusion method in Cryschem 24 well plates (Hampton Research). The reservoir solution (400 μL) consisted of 2% (w/v) PEG 4000, 2% (w/v) isopropanol and 0.01*M* trisodium citrate dihydrate pH 5.6. The crystallization solution contained 1 μL of protein solution [24 mg/mL in 20 m*M* HEPES pH 7.5, 0.15*M* NaCl, and 1 m*M* dithiothreitol (DTT)] and 1 μL of 2% (w/v) PEG 4000, 2% (w/v) isopropanol and 0.01*M* trisodium citrate dihydrate pH 5.6. Crystals grew at 289 K in 1 week.

Crystallographic data collection and processing

Crystals were flash-cooled in a nitrogen gas stream at 100 K after soaking at 293 K in cryoprotectant solution containing the crystallization solutions supplemented with 10% glycerol. Diffraction data were collected to 2.8 Å resolution at the UQ ROCX diffraction facility on a Rigaku FR-E Superbright generator with Osmic Vari-Max HF optics and Rigaku Saturn

944 CCD detector. For complete data collection, 80° of rotation was measured with 0.5° oscillations using the X-ray wavelength of Cu Kα (1.5419 Å), an exposure duration of 30 s and a crystal-to-detector distance of 70 mm at 100 K. Data were integrated and scaled using d*Trek.[31]

Structure determination

Molecular replacement with PHASER[32] was used to solve the structure of the GFP:GFP-nanobody complex using as input models the GFP structure (PDB ID 2QU1) and the LAC031 anti-VLA1-engineered antibody fragment (PDB ID 3EOT). Refinement was performed in PHENIX with noncrystallographic symmetry restraints applied for each chain,[33] and model building done with COOT[34] yielded the final models. Illustrations were made in CCP4mg.[35] The final model contains four copies of the GFP:GFP-nanobody heterodimeric complex in the asymmetric unit. The crystal parameters and data collection statistics are summarized in Table I.

Surface interferometry-based analysis of GFP:GFP-Nanobody

Streptavidin-coated sensors (ForteBio) were used to analyze GFP–GFP-nanobody interaction with the OctetRED interaction analyzer (ForteBio). To biotinylate GFP-nanobody, the protein was rebuffered on a NAP-5 desalting column (GE Healthcare) pre-equilibrated with binding buffer and mixed with two times molar excess of Sulfo-NHS-Biotin (Pierce). The reaction mixture was incubated at 37°C for 30 min, and the biotinylated protein was rebuffered on a NAP-5 column pre-equlibrated with binding buffer. The sensors were exposed to 50 μg/mL solution of biotinylated nanobody and washed in PBS. GFP:GFP-nanobody interactions were measured in binding buffer with addition of 50 μg/mL of bovine serum albumin. The exceptionally high stability of GFP-nanobody allowed multiple regenerations of the sensors by short (3 times for 5 s) exposure to acidic buffer (0.1*M* Gly-HCl pH 2.2) without a noticeable decrease in GFP binding capacity. Octet readout data was analyzed using ForteBio Data Analysis software (version 6.2.0.21).

Isothermal titration calorimetry

GFP and GFP-nanobody proteins were purified as above and then dialyzed into 20 m*M* phosphate buffer (pH 7.4) and 150 m*M* NaCl (binding buffer). Experiments were performed using a Microcal iTC200 instrument at different temperatures. Typically, experiments were performed by titrating 11 × 3.8 μL aliquots of 50 μ*M* solution of GFP-nanobody into 5 μ*M* solution of GFP. Data was processed using ORIGIN to derive the thermodynamic parameter ΔH and the stoichiometry N. Concentrations of the proteins were determined by absorbance measurement at 280 nm using theoretical extinction coefficients. Because of the high affinity of binding ($K_d < 1$ n*M*), it was not possible to accurately determine the equilibrium dissociation constant K_d by ITC, however, the binding heat capacity could be determined from the relation $\Delta C_p = \delta\Delta H/\delta T$.

In vitro *translation template assembly and mutagenesis*

DNA templates encoding different fluorescent protein variants were synthesized using overlap extension PCR (OE PCR) as previously described.[26] First genes coding for GFP and CFP were split into two fragments with compatible ends by PCR with the plasmid pTUB-NEO[26] as a template. To introduce N146I mutation into GFP, primers 511/9221 and 9223/2511 were used to synthesize 5′ and 3′ fragments, respectively. Similarly, fragments for CFP(I146N) assembly were obtained using primer pairs 511/9220 and 9222/2511. In addition, PCR fragments for assembling of wild-type GFP and CFP were obtained using primers 511/9220, 9222/2511 and 511/9221, 9223/2511, respectively. All PCR fragments were gel purified and subjected to OE PCR yielding full-length templates for *in vitro* translation. Resultant OE PCR products were precipitated with ethanol and dissolved in distilled water. Primer sequences: 511 CTACAACACGACCCTCTCCG, 2511 ACGCGTACACAACACACGGAC, 9220 GTGGCTGT TGTAGTTGTAC, 9221 GTGGCTGATGTAGTTGTAC, 9222 GTACAACTACAACAGCCAC, and 9223 GTAC AACTACATCAGCCAC.

Preparation of the supplemented Leishmania tarentolae *Extract*

Leishmania cell culture propagation and cell lysate preparation was performed as previously described.[26] The obtained rebuffered lysate was supplemented to 40 % (v/v) of its volume with feeding solution containing 8.5 m*M* ATP, 3.17 m*M* GTP, 1.25 m*M* spermidine, 10 m*M* DTT, 200 m*M* creatine phosphate, 12.5 m*M* $Mg(OAc)_2$, 32 m*M* HepesKOH pH 7.6, 5% (v/v) PEG 3000, 5.25x protease inhibitor (Complete™ EDTA-free, Roche), 0.68 m*M* of each amino acid, 2.5 m*M* of each UTP and CTP, 0.05 m*M* antisplice leader DNA oligonucleotide, 0.5 mg/mL of T7 RNA polymerase, and 200 U/mL of creatine phosphokinase. The supplemented lysate was aliquoted into cryogenic tubes, snap-frozen in liquid nitrogen and stored at −80°C.

Cell-free protein translation and pull-down experiments

The *in vitro* translation was performed in 50 μL of translation reaction. In a typical setup, 35 μL of LTE supplemented lysate were mixed with water-dissolved PCR template to final concentrations of 20 n*M* and adjusted to the final volume with mQ

water. The reaction was then transferred into a 96-well plate and placed into a Synergy 4 plate reader (BioTek), thermostated at 26°C. The accumulation of CFP and GFP fluorescence was monitored using 430Ex/485Em and 485Ex/528Em filter sets, respectively. The reactions were monitored for ~2 h after which no increase of fluorescence was detectable. For pull-down experiments, the reaction mixtures were transferred into plastic tubes and mixed with 10 μL of GFP-nanobody-coated beads. Antibody-coated beads were prepared by coupling GFP-nanobody to NHS-Sepharose (GE Healthcare) according to manufacturer recommendations. After 20 min incubation at room temperature with moderate shaking, the beads were washed two times with 250 μL of buffered NaCl solution [25 m*M* Tris HCl (pH 7.6) and 500 m*M* NaCl] and retained proteins were eluted with 20 μL of hot SDS-PAGE loading buffer (Invitrogen). The translation reaction mixtures and eluates were resolved on 4–12% SDS-PAGE gels (NuPage, Invitrogen), transferred to nitrocellulose membrane (Pall) and blotted with anti-GFP antibodies (Cell Signalling Technologies) according to standard protocol for IR Odyssey Imaging System (Li-Cor). The membranes were then scanned on Odyssey Imaging System (Li-Cor) and the data was analyzed using Odyssey Application Software V3.0 (Li-Cor).

Accession numbers

Coordinates and structure factors for the GFP:GFP-nanobody complex have been deposited in the RCSB Protein Data Bank with accession number 3OGO. Raw diffraction data is available on the Diffraction Images Repository (DIMER) http://xr-diffraction.imb.uq.edu.au.

Acknowledgments

The authors would like to acknowledge support received from the University of Queensland Remote Op Crystallization and X-ray (UQ ROCX) facility as well as UQ protein expression facility.

References

1. Tsien RY (2009) Constructing and exploiting the fluorescent protein paintbox (Nobel Lecture). Angew Chem Int Ed Engl 48:5612–5626.
2. Skerra A (2007) Alternative non-antibody scaffolds for molecular recognition. Curr Opin Biotechnol 18: 295–304.
3. Muyldermans S, Baral TN, Retamozzo VC, De Baetselier P, De Genst E, Kinne J, Leonhardt H, Magez S, Nguyen VK, Revets H, Rothbauer U, Stijlemans B, Tillib S, Wernery U, Wyns L, Hassanzadeh-Ghassabeh G, Saerens D (2009) Camelid immunoglobulins and nanobody technology. Vet Immunol Immunopathol 128: 178–183.
4. Kirchhofer A, Helma J, Schmidthals K, Frauer C, Cui S, Karcher A, Pellis M, Muyldermans S, Casas-Delucchi CS, Cardoso MC, Leonhardt H, Hopfner KP, Rothbauer U (2010) Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol 17: 133–138.
5. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics 7:282–289.
6. Rothbauer U, Zolghadr K, Tillib S, Nowak D, Schermelleh L, Gahl A, Backmann N, Conrath K, Muyldermans S, Cardoso MC, Leonhardt H (2006) Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat Methods 3:887–889.
7. Schornack S, Fuchs R, Huitema E, Rothbauer U, Lipka V, Kamoun S (2009) Protein mislocalization in plant cells using a GFP-binding chromobody. Plant J 60: 744–754.
8. De Genst E, Silence K, Decanniere K, Conrath K, Loris R, Kinne J, Muyldermans S, Wyns L (2006) Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc Natl Acad Sci USA 103:4586–4591.
9. Korotkov KV, Pardon E, Steyaert J, Hol WG (2009) Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. Structure 17:255–265.
10. Lam AY, Pardon E, Korotkov KV, Hol WG, Steyaert J (2009) Nanobody-aided structure determination of the EpsI:EpsJ pseudopilin heterodimer from *Vibrio vulnificus*. J Struct Biol 166:8–15.
11. Tereshko V, Uysal S, Koide A, Margalef K, Koide S, Kossiakoff AA (2008) Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. Protein Sci 17: 1175–1187.
12. De Genst E, Silence K, Ghahroudi MA, Decanniere K, Loris R, Kinne J, Wyns L, Muyldermans S (2005) Strong *in vivo* maturation compensates for structurally restricted H3 loops in antibody repertoires. J Biol Chem 280:14114–14121.
13. Loris R, Marianovsky I, Lah J, Laeremans T, Engelberg-Kulka H, Glaser G, Muyldermans S, Wyns L (2003) Crystal structure of the intrinsically flexible addiction antidote MazE. J Biol Chem 278:28252–28257.
14. Decanniere K, Desmyter A, Lauwereys M, Ghahroudi MA, Muyldermans S, Wyns L (1999) A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. Structure 7:361–370.
15. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. Nat Biotechnol 22:1567–1572.
16. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. Science 273: 1392–1395.
17. Decanniere K, Transue TR, Desmyter A, Maes D, Muyldermans S, Wyns L (2001) Degenerate interfaces in antigen-antibody complexes. J Mol Biol 313:473–478.
18. Perozzo R, Folkers G, Scapozza L (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. J Recept Signal Transduct Res 24: 1–52.
19. Prabhu NV, Sharp KA (2005) Heat capacity in proteins. Annu Rev Phys Chem 56:521–548.
20. Falconer RJ, Penkova A, Jelesarov I, Collins BM (2010) Survey of the year 2008: applications of isothermal titration calorimetry. J Mol Recognit 23:395–413.

21. Xu D, Lin SL, Nussinov R (1997) Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. J Mol Biol 265:68–84.
22. Cooper A (2005) Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. Biophys Chem 115:89–97.
23. Rodier F, Bahadur RP, Chakrabarti P, Janin J (2005) Hydration of protein-protein interfaces. Proteins 60: 36–45.
24. James LC, Roversi P, Tawfik DS (2003) Antibody multispecificity mediated by conformational diversity. Science 299:1362–1367.
25. Krogsgaard M, Prado N, Adams EJ, He XL, Chow DC, Wilson DB, Garcia KC, Davis MM (2003) Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation. Mol Cell 12:1367–1378.
26. Mureev S, Kovtun O, Nguyen UT, Alexandrov K (2009) Species-independent translational leaders facilitate cell-free expression. Nat Biotechnol 27:747–752.
27. Heim R, Tsien RY (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Curr Biol 6:178–182.
28. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R (1996) Protein-protein interfaces: architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences. Crit Rev Biochem Mol Biol 31:127–152.
29. Veselovsky AV, Ivanov YD, Ivanov AS, Archakov AI, Lewi P, Janssen P (2002) Protein-protein interactions: mechanisms and modification by drugs. J Mol Recognit 15:405–422.
30. Saerens D, Pellis M, Loris R, Pardon E, Dumoulin M, Matagne A, Wyns L, Muyldermans S, Conrath K (2005) Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. J Mol Biol 352:597–607.
31. Pflugrath JW (1999) The finer things in X-ray diffraction data collection. Acta Crystallogr D 55: 1718–1725.
32. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Crystallogr 40:658–674.
33. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D 58: 1948–1954.
34. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D 60: 2126–2132.
35. Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A, Noble M (2004) Developments in the CCP4 molecular-graphics project. Acta Crystallogr D 60:2288–2294.