# A robust pipeline for rapid production of versatile nanobody repertoires

Peter C Fridy[1,8], Yinyin Li[2,8], Sarah Keegan[3], Mary K Thompson[1], Ilona Nudelman[1], Johannes F Scheid[4], Marlene Oeffinger[5,6], Michel C Nussenzweig[4,7], David Fenyö[3], Brian T Chait[2] & Michael P Rout[1]

**Nanobodies are single-domain antibodies derived from the variable regions of Camelidae atypical immunoglobulins. They show promise as high-affinity reagents for research, diagnostics and therapeutics owing to their high specificity, small size (~15 kDa) and straightforward bacterial expression. However, identification of repertoires with sufficiently high affinity has proven time consuming and difficult, hampering nanobody implementation. Our approach generates large repertoires of readily expressible recombinant nanobodies with high affinities and specificities against a given antigen. We demonstrate the efficacy of this approach through the production of large repertoires of nanobodies against two antigens, GFP and mCherry, with $K_d$ values into the subnanomolar range. After mapping diverse epitopes on GFP, we were also able to design ultrahigh-affinity dimeric nanobodies with $K_d$ values as low as ~30 pM. The approach presented here is well suited for the routine production of high-affinity capture reagents for various biomedical applications.**

There is a continual need in biomedicine for antibodies that recognize target molecules with high affinity and specificity. When high-affinity antibodies are not available for the molecule of interest, common protein tags such as GFP, mCherry and Flag, which do have known antibodies, are invaluable. Most studies demand high-quality antibodies against these protein tags, particularly when affinity isolation is required[1–4]. Monoclonal or polyclonal antibodies are the primary bait reagents available for these purposes. However, their large size, limited availability and batch-to-batch variation have often proven problematic for biochemical or proteomic studies[5].

Single-domain antibodies, referred to as nanobodies[6], have emerged as an alternative to traditional antibodies. Nanobodies are usually derived from camelids such as llamas, which make a unique subset of immunoglobulins consisting of heavy-chain homodimers devoid of light chains[7–9]; their variable region ($V_HH$) is the smallest antigen-binding single polypeptide chain found in any natural antibody[8–12]. Nanobodies are recombinant antigen-binding domains derived from these $V_HH$ regions. Unlike monoclonal antibodies, they can readily be produced in large amounts in bacterial expression systems[9,13]. Moreover, nanobodies are usually extremely stable, can bind antigens with affinities in the nanomolar range and are smaller in size (approximately 15 kDa) than other antibody constructs[11,14–18]. However, rapid and robust techniques for isolating extensive repertoires of high-affinity nanobodies have proven elusive: the labor-intensiveness and poor efficiency of current approaches are a major bottleneck for the widespread use of these reagents[8,12,14], explaining why demand for nanobodies greatly exceeds supply[8].

Here we present a highly optimized pipeline that allows the rapid production of large repertoires of high-affinity nanobodies against selected proteins. The approach is based on high-throughput DNA sequencing of a marrow lymphocyte $V_HH$ cDNA library from an immunized llama combined with mass spectrometric (MS) identification of high-affinity $V_HH$ regions derived from serum of the same animal.

## RESULTS

### Strategy for nanobody identification

Our approach to nanobody discovery centers on MS identification of affinity-purified heavy-chain antibodies isolated from an individual llama, using a DNA sequence database generated from the same animal (**Fig. 1**). This concept is inspired by our previous efforts to identify circulating neutralizing HIV antibodies in humans by MS in conjunction with patient-specific antibody cDNA databases[19]. Our approach represents a pipeline for nanobody production in which each stage has been highly optimized (**Supplementary Protocol**).

To generate nanobody repertoires of maximal utility, we chose the GFP and mCherry tags as our first target antigens owing to their central roles in cell biological studies. Furthermore, although these fluorescent proteins have broadly similar β-barrel structures, they are evolutionarily divergent and thus are distinct

[1]Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, New York, USA. [2]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York, USA. [3]Center for Health Informatics and Bioinformatics, New York University School of Medicine, New York, New York, USA. [4]Laboratory of Molecular Immunology, The Rockefeller University, New York, New York, USA. [5]Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada. [6]Département de Biochimie et Médecine Moléculaire, Faculté de Médecine, Université de Montréal, Montreal, Quebec, Canada. [7]Howard Hughes Medical Institute, The Rockefeller University, New York, New York, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to M.P.R. (rout@rockefeller.edu), B.T.C. (chait@rockefeller.edu) or D.F. (david@fenyolab.org).

immunogens[20]. After immunization of individual llamas with these antigens and confirmation of an immune response, we serially fractionated serum bleeds to obtain exclusively $V_H$H-containing heavy-chain antibodies, taking advantage of the different affinities of protein A and protein G for $V_H$H versus

Diverse target antigens



**Figure 1** | Overview of nanobody identification and production pipeline. The example nanobody structure shown was obtained from PDB 3K1K (ref. 29). LC-MS/MS, liquid chromatography–tandem mass spectrometry.

conventional antibodies[7] (**Supplementary Fig. 1a**). We affinity-purified the $V_H$H-containing fraction over antigen-coupled resin, washed with $MgCl_2$ at various stringencies (**Supplementary Fig. 1b**) and digested with papain on-resin to cleave away the constant regions and leave behind the desired minimal fragments of $V_H$H variable region. Finally, the antigen-bound $V_H$H fragments were eluted and separated by SDS-PAGE, allowing purification of ~15-kDa $V_H$H fragments free from residual conventional Fab fragments and Fc fragments (both ~25 kDa) and undigested antibodies (~50 kDa) (**Supplementary Fig. 1c**). The gel-purified bands were trypsin digested and analyzed by liquid chromatography–MS and tandem MS (MS/MS) (**Fig. 2a**). We recovered the highest-affinity $V_H$H fragments via the highest-stringency washes, which also decreased the complexity of the eluted sample, aiding MS analysis.

To create an animal-specific antibody sequence database, we obtained lymphocyte RNA samples from individual immunized llamas for high-throughput sequencing. We isolated mononuclear cells from bone marrow aspirates, enriching for long-lived antibody-secreting plasma cells, which transcribe elevated levels of immunoglobulin mRNA[19,21,22]. Notably, we did not create expression or display libraries, and thus we removed the need for efficient exogenous expression, folding and presentation of the clones.

We generated cDNA from total lymphocyte RNA and performed nested PCR to specifically amplify sequences encoding the $V_H$H variable regions[14]. Sequencing of this PCR product using a high-throughput Roche 454 (GFP) or Illumina MiSeq (mCherry) platform resulted in ~800,000 or ~3,000,000 unique reads, respectively. These reads were translated, filtered and trypsin digested *in silico* to create a searchable peptide database for MS analysis (**Fig. 1** and **Supplementary Fig. 2**).

The identification of specific $V_H$H sequences is more challenging than a typical MS-based protein identification, as they consist in large part of highly conserved framework regions that are less easily distinguished by MS. Moreover, whereas well-established databases are available for other proteins, for $V_H$H a cDNA database must be generated for each immunized animal. To deal with both challenges, we developed a bioinformatic pipeline that identifies the highest-probability matches from a pool of related $V_H$H sequences (Llama Magic; http://www.llamamagic.org/). In this pipeline, $V_H$H sequences are ranked by a metric based on MS sequence coverage of complementarity-determining region 3 (CDR3, the most diverse $V_H$H region) as well as CDR1 and CDR2 coverage, total $V_H$H coverage, sequencing counts, mass spectral counts and the expectation values of matched peptides (**Supplementary Figs. 2** and **3**). Preliminary attempts to identify $V_H$H sequences solely by their CDR3 regions revealed that identical CDR3 sequences are frequently shared between multiple $V_H$H regions that have different CDR1 and CDR2 sequences. This is probably a result of somatic gene conversion, in which, after V(D)J recombination, secondary recombination occurs between upstream V gene segments and already rearranged V(D)J genes[23,24].

Our automatic ranking pipeline, coupled with careful manual inspection, overcame these issues and provided 44 high-probability hits for nanobodies against GFP. We classified these as llama antibody against GFP (LaG) 1–44 and subjected them to further screening (**Supplementary Fig. 4**). In the analysis of llama antibodies against mCherry (LaM), a smaller subset of eight clones was chosen for follow-up (LaM 1–8; **Supplementary Fig. 5**).
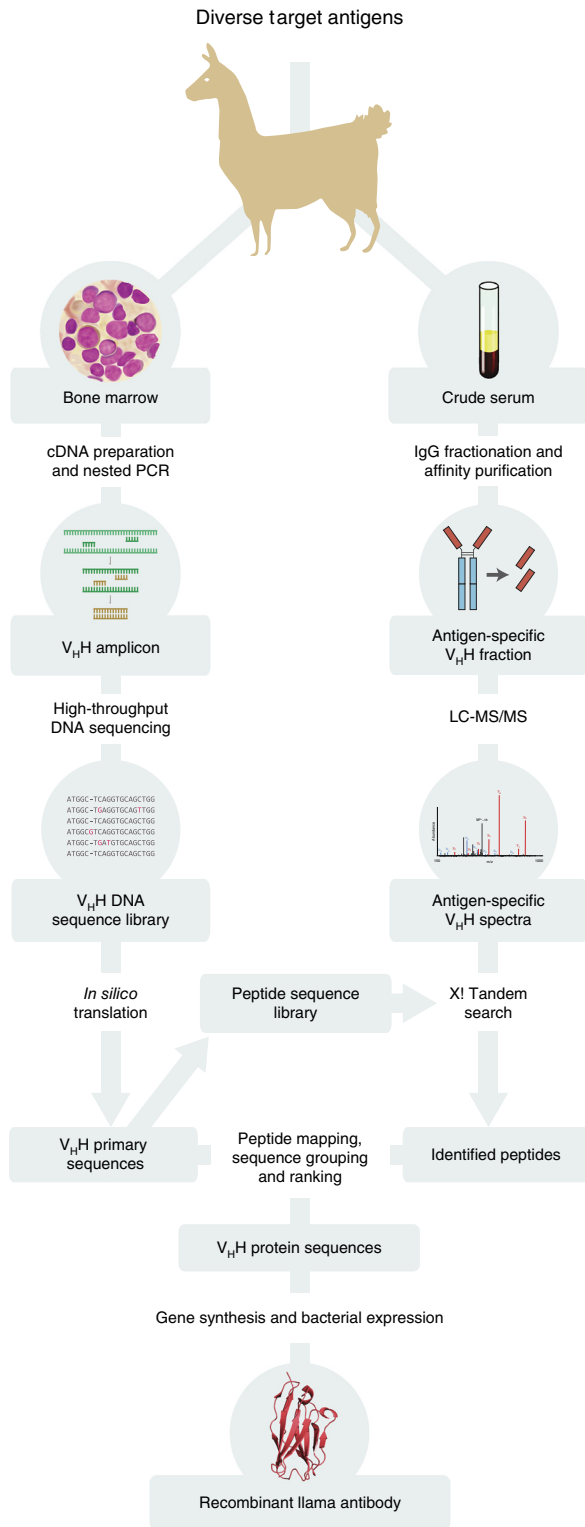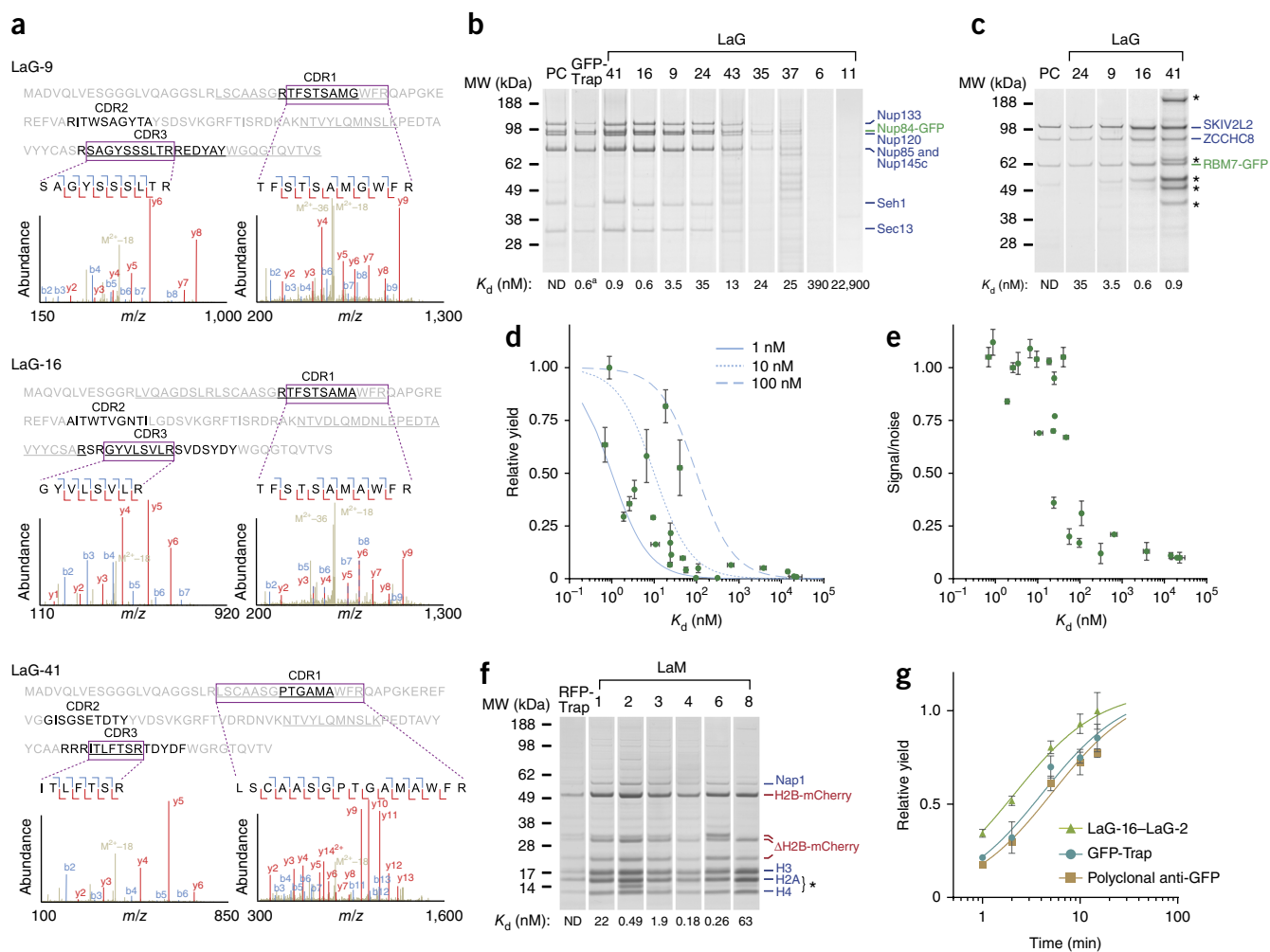
**Figure 2** | Characterization of V$_H$H IgG and recombinant nanobodies. (**a**) Tandem mass spectra of identified peptides (shown boxed), mapped to the CDR regions of three candidate V$_H$H sequences. The MS-covered regions of these sequences are underlined. Dashed lines indicate overlapping peaks. (**b,c**) Affinity isolations of Nup84-GFP from *S. cerevisiae* (**b**) or RBM7-GFP from HeLa cells (**c**) using LaGs, GFP-Trap or polyclonal anti-GFP llama antibody (PC). Eluted proteins were analyzed by SDS-PAGE followed by MS. $K_d$ values for GFP are listed. MW, molecular weight. ND, not determined. Contaminant bands are identified by asterisks. $^a$$K_d$ taken from published value (ref. 29). (**d**) Relative yields of affinity-isolated Nup84-GFP protein plotted against LaG *in vitro* affinities for GFP (green dots). Theoretical curves of the expected fraction of ligand bound to an immobilized binding partner at various $K_d$ values are shown for three hypothetical ligand concentrations (blue lines). (**e**) Signal-to-noise ratio of three Nup84-complex components plotted against each LaG's $K_d$. (**f**) Affinity isolation of mCherry-tagged histone H2B (HTB2) from *S. cerevisiae* by LaMs or RFP-Trap. Eluted proteins were analyzed by SDS-PAGE and identified by MS. Breakdown products of H2B are labeled ΔH2B-mCherry. Asterisk indicates LaM nanobody that leaked during the affinity-purification procedure. LaM lanes are labeled with the $K_d$ for mCherry. (**g**) Affinity isolations of yeast Nup84-GFP using the commercial nanobody GFP-Trap, polyclonal anti-GFP or a LaG-16–LaG-2 dimer with a glycine-rich peptide linker. The complex was isolated at various time points, and relative yield was determined by quantification of Coomassie-stained bands of known Nup84-complex components. Data are representative of two experiments (**a–c,f**) or are means from two experiments ± s.e.m. (**d,e,g**).

Codon-optimized genes for these hits were synthesized and cloned into a bacterial expression vector. After expression, we passed lysates over antigen-coupled resin to identify nanobodies that displayed robust expression as well as high and specific affinity (**Supplementary Fig. 6**). From these screens, we found 25 specific nanobodies against GFP and 6 against mCherry. Phylogenetic analysis of the verified nanobodies revealed substantial sequence diversity among clones (**Supplementary Fig. 7**). Although our protocol is not directly analogous to phage display, the high success rate of this single screening step (57–75%) is favorable in comparison to the final panning and selection steps of phage display, in which up to 10[7] clones are screened to identify even a few positive clones[12,14,25,26]. We assessed the affinity of the nanobodies further by either surface plasmon resonance

(SPR) or *in vitro* binding assays with immobilized nanobodies (**Supplementary Figs. 8–10**). For the LaG repertoire these experiments revealed a wide range of affinities, with dissociation constants ($K_d$) from 0.5 nM to over 20 μM (**Table 1**), and identified 16 nanobodies with high-affinity binding (≤50 nM; **Supplementary Fig. 8**). The affinities of the six LaMs were consistently strong, with $K_d$ values ranging from 0.18 nM to 63 nM (**Table 1** and **Supplementary Fig. 9**).

## Specificity and efficacy of recombinant nanobodies

We performed a variety of experiments to assess our nanobodies. We first affinity-isolated endogenous GFP- and mCherry-tagged proteins in yeast and human cells. All 25 positive LaGs were used for the isolation of GFP-tagged Nup84, a structural nuclear-pore-complex

**Table 1** | Characteristics of LaG, LaG dimer, and LaM proteins

| Clone ID | MW (Da) | $K_d$ (nM) | Nup84-GFP S:N | RBM7-GFP S:N | Binds *A. macrodactyla* CFP (LaG) or DsRed (LaM) | GFP epitope group | No. of binding-site residues | ASA of binding-site residues (Å²) |
|---|---|---|---|---|---|---|---|---|
| LaG-2 | 15,919 | 19[a], 16 | 1.03 | 0.42 | – | III | 55 | 2,204 |
| LaG-3 | 15,329 | 25 | 0.77 | 1.13 | + | ND | ND | ND |
| LaG-6 | 15,700 | 310 | 0.12 | ND | + | ND | ND | ND |
| LaG-9 | 16,062 | 3.5 | 1.02 | 1.04 | + | I | 62 | 2,551 |
| LaG-10 | 15,748 | 97 | 0.17 | ND | + | ND | ND | ND |
| LaG-12 | 16,090 | 56 | 0.20 | ND | + | ND | ND | ND |
| LaG-14 | 16,002 | 1.9 | 0.84 | 0.58 | + | I | 66 | 2,519 |
| LaG-16 | 16,306 | 0.7 | 1.05 | 0.92 | + | I | 60 | 2,605 |
| LaG-17 | 15,823 | 50 | 0.67 | ND | + | I | 60 | 2,543 |
| LaG-19 | 15,528 | 24.6[a] | 0.95 | 1.06 | + | II | 54 | 2,404 |
| LaG-21 | 15,452 | 7 | 1.09 | ND | + | II | 56 | 2,340 |
| LaG-24 | 14,763 | 41 | 1.05 | 1.09 | – | ND | ND | ND |
| LaG-26 | 16,221 | 2.6 | 1.00 | ND | + | II | 53 | 2,070 |
| LaG-27 | 15,565 | 9.5 | 1.04 | ND | + | II | 57 | 2,216 |
| LaG-29 | 15,449 | 110 | 0.31 | ND | + | ND | ND | ND |
| LaG-30 | 16,159 | 0.5 | 1.04 | ND | + | ND | ND | ND |
| LaG-35 | 16,010 | 23.5[a] | 0.70 | ND | + | ND | ND | ND |
| LaG-37 | 16,329 | 24 | 0.36 | ND | + | ND | ND | ND |
| LaG-41 | 15,471 | 0.9 | 1.12 | 0.41 | + | II | 53 | 2,091 |
| LaG-42 | 15,490 | 600 | 0.21 | ND | + | ND | ND | ND |
| LaG-43 | 16,167 | 11 | 0.69 | ND | + | I | 55 | 2,381 |
| LaG-5 | 15,589 | 14,200[a] | 0.11 | ND | ND | ND | ND | ND |
| LaG-8 | 15,953 | 20,000[a] | 0.10 | ND | ND | ND | ND | ND |
| LaG-11 | 16,221 | 22,900[a] | 0.10 | ND | ND | ND | ND | ND |
| LaG-18 | 16,459 | 3,800[a] | 0.13 | ND | ND | ND | ND | ND |
| LaG-16–G₄S–LaG-2 | 30,791 | 0.036 | ND | ND | ND | ND | ND | ND |
| LaG-16–3×Flag–LaG-2 | 32,972 | 0.268 | ND | ND | ND | ND | ND | ND |
| LaG-41–G₄S–LaG-2 | 29,956 | 0.150 | ND | ND | ND | ND | ND | ND |
| LaM-1 | 15,380 | 22 | N/A | N/A | – | N/A | N/A | N/A |
| LaM-2 | 15,151 | 0.49 | N/A | N/A | – | N/A | N/A | N/A |
| LaM-3 | 15,196 | 1.9 | N/A | N/A | + | N/A | N/A | N/A |
| LaM-4 | 14,866 | 0.18 | N/A | N/A | + | N/A | N/A | N/A |
| LaM-6 | 14,428 | 0.26 | N/A | N/A | – | N/A | N/A | N/A |
| LaM-8 | 14,666 | 63 | N/A | N/A | – | N/A | N/A | N/A |

$K_d$ values for GFP and mCherry binding were determined by SPR unless otherwise noted. $K_d$ values are also listed for LaG dimers fused by a glycine-rich peptide linker (three repeats of GGGGS, or G₄S) or by a 3×Flag linker. For yeast Nup84-GFP and mammalian RBM7-GFP affinity isolations using LaGs, Coomassie-stained bands from elutions separated by SDS-PAGE were quantified, and known specific and nonspecific bands were used to calculate signal-to-noise (S:N) ratios. Bead binding assays were used to determine affinity for fluorescent proteins, and the table highlights differences in specificity for *A. macrodactyla* CFP among LaGs and for DsRed among LaMs. GFP epitopes for LaGs were determined by NMR and classified into three groups according to their location (I–III). Also shown are the number of residues identified in the binding site and the site's calculated accessible surface area (ASA). MW, molecular weight; ND, not determined; N/A, not applicable.
[a]$K_d$ determined by bead binding assay.

component, in budding yeast[27,28] (**Fig. 2b**). We plotted each LaG's observed $K_d$ against a quantification of either signal-to-background ratio or yield from a Nup84-GFP affinity capture (**Fig. 2d**,**e** and **Table 1**). Almost all LaGs pulled down detectable amounts of Nup84-GFP and its associated proteins, and many performed as well as or better than either our best affinity-purified polyclonal antibodies[1] or the single commercially available anti-GFP nanobody, GFP-Trap (ChromoTek), which has a $K_d$ of 0.59 nM (**Fig. 2b**,**g**)[29]. For instance, when we determined depletion of Nup84-GFP by western blot, LaG-16 displayed slightly higher yields than GFP-Trap (**Supplementary Fig. 11**).

Generally speaking, there were strong correlations between low $K_d$ and both high signal-to-background ratio and high yield, which is consistent with the relationship theoretically predicted for the percentage of low-abundance yeast target proteins bound in solution[30] (**Fig. 2d**). Our ability to compare structurally similar nanobodies raised against a single antigen provided an opportunity to demonstrate the importance of very low $K_d$ to antibody performance in this type of application. Nanobodies showed a

precipitous decline in affinity-purification performance as their $K_d$ values rose, even at values as low as 10 nM—typically considered high affinity for an antibody. These findings highlight the need for ultrahigh-affinity reagents, such as the nanobodies described here, for proteome and interactome studies.

We also performed affinity-capture experiments with GFP-tagged RBM7, a component of the human nuclear exosome, from HeLa cells[4] (**Fig. 2c**). We observed yields and purities comparable to those seen with Nup84-GFP. However, for certain LaGs, notably LaG-41, the amount of contaminants differed between purifications in yeast and purifications in HeLa cells (**Fig. 2b**,**c**). These results underscore that even high-affinity reagents can produce unpredictable background in certain cell types. It is therefore useful to obtain and test large repertoires of such affinity reagents to improve the chances that at least one will be optimal for any particular application.

Similarly, we isolated mCherry-tagged histone H2B from yeast with LaMs conjugated to Dynabeads (Life Technologies; **Fig. 2f**). All six LaMs efficiently isolated the core nucleosome complex,

**Figure 3** | Efficacy of LaG and LaM nanobodies in immunofluorescence microscopy. (**a**,**b**) HeLa cells transiently transfected with tubulin-EmGFP or an EmGFP-tagged mitochondrial marker (in green) were fixed and immunostained with LaG-16 conjugated to Alexa Fluor 568 (AF568, in red). Nuclei were counterstained with DAPI (blue). (**c**) *T. brucei* cells expressing EGFP-tagged Sec13 were mixed 1:1 with wild-type cells, fixed and stained with LaG-16–AF568, with DAPI counterstaining. (**d**) An *S. cerevisiae* strain with mCherry-tagged histone H2B was fixed, permeabilized and then directly stained with LaM-4 conjugated to Alexa Fluor 488 (AF488). Scale bars, 10 μm. Images are representative of at least three experiments.
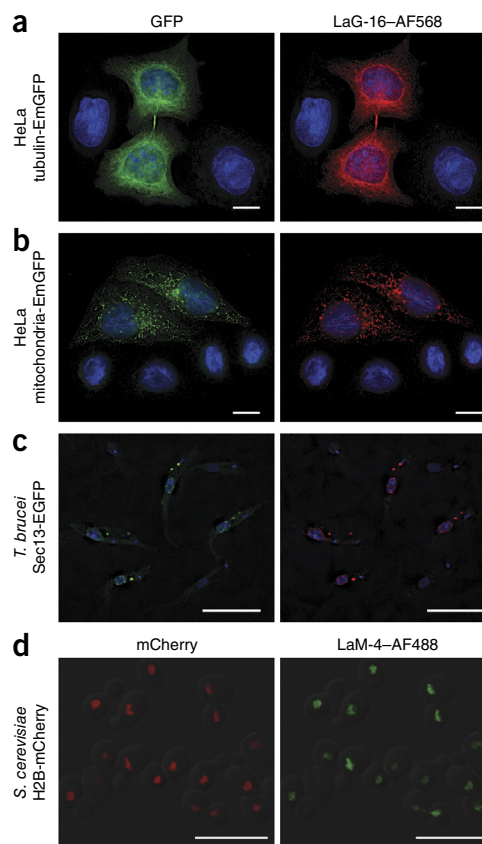
demonstrating the affinity and specificity of this second group of nanobodies. Consistent with the low $K_d$ values of all the identified LaMs, the yield and specificity of all affinity isolations were high. The commercial nanobody RFP-Trap (ChromoTek) was tested in parallel and had lower yields.

Nanobodies are powerful new tools for fluorescence microscopy, both standard and super-resolution[31]. We therefore tested the effectiveness of selected members of the LaG and LaM repertoires for immunofluorescence microscopy (**Fig. 3** and **Supplementary Fig. 12**). As target proteins, we first used Emerald GFP (EmGFP)-tagged tubulin and mitochondria-targeted EmGFP, transiently transfected into HeLa cells[32]. We stained fixed cells with LaG-16 conjugated to Alexa Fluor 568 (Life Technologies), producing specific and strong staining of EmGFP-tagged microtubules or mitochondrial structures, with negligible nonspecific staining of untransfected cells (**Fig. 3a,b**). To demonstrate the versatility of these reagents, we also used them for immunofluorescence in a *Trypanosoma brucei* strain with EGFP-tagged Sec13. This protein localizes to both the nuclear pore complex and COPII-coated vesicles, and indeed, the Alexa Fluor 568–nanobody signal colocalized with EGFP to produce the expected staining of the nuclear rim and endoplasmic reticulum[33] (**Fig. 3c**). To determine whether our anti-mCherry nanobodies were similarly well suited for immunofluorescence microscopy, we conjugated LaM-4 to Alexa Fluor 488 and stained *Saccharomyces cerevisiae* expressing mCherry-tagged H2B, resulting in specific, colocalized nuclear staining (**Fig. 3d**).

As previous studies have demonstrated potentially useful modulation of GFP fluorescence activity by nanobodies[29], we compared the fluorescence spectra of GFP in the presence or absence of various LaGs to look for spectral shifts upon binding. We observed moderate increases in fluorescence for several LaGs, with a maximum increase in fluorescence intensity of approximately 60% (**Supplementary Fig. 13**).

One additional question of specificity we sought to address was the ability of our nanobodies to recognize other fluorescent homologs of *Aequorea victoria* GFP and *Discosoma* sp. mCherry. We tested the 13 highest-affinity LaGs against a variety of fluorescent proteins: EGFP, two yellow fluorescent protein variants, two cyan fluorescent protein variants, BFP, mCherry and DsRed (**Fig. 4a**). None of these nanobodies bound DsRed or mCherry, two *Discosoma*-derived proteins with low sequence identity to EGFP (<30%), and none bound TurboYFP (Evrogen), derived from *Phialidium* sp., which has 53% sequence identity to EGFP[20,34,35]. All bound standard *Aequorea victoria*–derived CFP, YFP and BFP variants (>96% EGFP identity). Two LaGs did not bind a moderately divergent (78% EGFP identity) cyan fluorescent protein sequence from *Aequorea macrodactyla*, whereas all others

did[36]. These results indicate that although the identified LaGs bind specifically to fluorescent proteins that have high identity to EGFP, differential binding activities can be obtained through use of variants from other species. Our anti-mCherry LaM nanobodies bound to mCherry but not to any of the green, yellow or cyan fluorescent proteins we tested (**Fig. 4b**). Two LaMs (LaM-3 and LaM-4) bound to standard DsRed, which has approximately 80% sequence identity to mCherry. Given the different affinities of LaG and LaM nanobodies for fluorescent proteins, including variation in specificity for *A. macrodactyla* cyan fluorescent protein and DsRed, these reagents have diverse potential uses in differential labeling and affinity-capture experiments from cells simultaneously expressing multiple fluorescently tagged proteins.

## Mapping of the nanobody epitopes on GFP

We identified the epitopes on GFP recognized by the 12 highest-affinity LaGs using chemical shift perturbation, a well-established NMR technique. This method allows researchers to map binding sites on a protein by following changes in its characteristic 'fingerprint' spectrum (typically the $^{15}$N-$^{1}$H HSQC spectrum) that result from adding an unlabeled ligand to a $^{15}$N-labeled protein sample[37].

Because previous studies have already made backbone $^{15}$N-$^{1}$H chemical shift assignments for the GFPuv variant[38,39] (closely related to standard EGFP, with 97% sequence identity), we prepared $^{15}$N-labeled GFPuv, measured its $^{15}$N-$^{1}$H HSQC spectrum and obtained chemical shift assignments on the basis of those published[38,39] (**Supplementary Fig. 14a**). We then tested complexes between 12 high-affinity LaGs and $^{15}$N-labeled GFPuv and measured their $^{15}$N-$^{1}$H HSQC spectra. For 11 of the 12 cases,

**Figure 4** | Nanobody fluorescent protein binding. (**a,b**) SDS-PAGE analysis of high-affinity LaGs (**a**) or LaMs (**b**) that were conjugated to magnetic beads and incubated with various recombinant fluorescent proteins: *A. victoria* (*Av*) GFP and its variants CFP, BFP and YFP; a cyan fluorescent protein derived from *A. macrodactyla* (*Am* CFP); a yellow fluorescent protein from *Phialidium* (*Phi* YFP); and mCherry and DsRed from *Discosoma* (*Ds*). Structural models were obtained from PDB 1EMA (*Av*)[52], PDB 4HE4 (*Phi*)[53] and PDB 1GGX (*Ds*)[54]; the *Am* CFP model is a Phyre server prediction[36,55]. Gels are representative of at least two experiments.
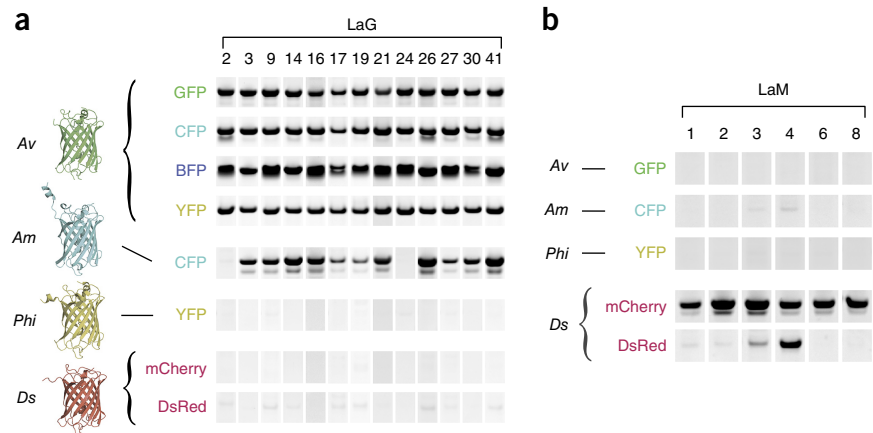
we observed clear and specific changes in chemical shifts of a large percentage of cross-peaks compared to the $^{15}$N-$^1$H HSQC spectrum of GFPuv alone (**Supplementary Figs. 14b,c** and **15**). In the twelfth case, LaG-24, the nanobody did not bind the GFPuv variant.

A chemical shift difference was calculated for all spectra, and residues exhibiting a difference greater than 0.03 p.p.m. were judged to be in the binding interface (**Supplementary Fig. 14b,c**)[37,40]. All the identified epitopes corresponded to large interfaces comprising more than 50 amino acids, consistent with the high-affinity binding we observed (**Fig. 5** and **Table 1**). The binding epitopes of the nanobodies can be divided into three distinct groups. The binding site of group I, containing five nanobodies (LaG-16, LaG-9, LaG-14, LaG-43 and LaG-17), overlaps with the binding site of group II, also containing five nanobodies (LaG-19, LaG-21, LaG-26, LaG-27 and LaG-41), whereas the group III nanobody (LaG-2) has a binding epitope on the opposite

side of the GFP molecule. As a control, we also used this NMR approach to determine the GFPuv binding site of the commercial GFP-Trap nanobody, the structure of whose complex with GFP has been crystallographically determined (PDB 3K1K)[29], and showed that the NMR-mapped epitope matched the published results[29,41] (**Supplementary Fig. 16**). Comparing the binding epitopes of our nanobodies with that of GFP-Trap, groups I and II show little or no overlap with the GFP-Trap binding site, whereas group III, which binds on the same face of GFP, shows substantial overlap (**Fig. 5**).

### Dimerized nanobodies as ultrahigh-affinity reagents

Because NMR identified multiple epitopes for these 12 LaGs, we engineered heterodimers of LaGs with nonoverlapping binding sites on GFP that could potentially bind with higher affinity, an approach that has been used successfully in various applications to develop high-avidity reagents[42,43]. A LaG-16-LaG-2 fusion with a flexible glycine-rich peptide linker (encoding three repeats of GGGGS) had the highest affinity as measured by SPR, with a $K_d$ of 36 pM. Dimers of other LaGs and those with a different linker (a 3×Flag tag) displayed $K_d$ values in the range of 100–200 pM. We also sought to determine whether the higher affinity of these dimers yielded faster affinity isolations after conjugation
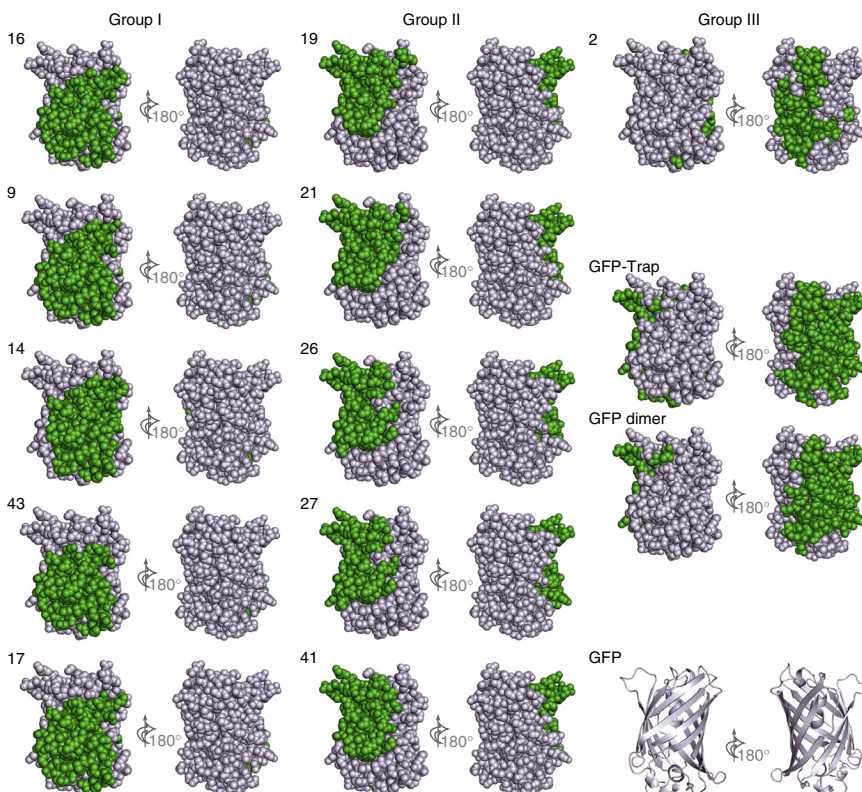


**Figure 5** | Mapping of nanobody binding epitopes on GFP by NMR. Binding epitopes of the 11 highest-affinity nanobodies on GFPuv are shown in three groups according to their location. For each nanobody, two opposite sides of GFPuv are shown (via a 180° rotation along a vertical axis), with the binding site of the respective nanobody colored green. All GFPuv molecules are represented in space-filling mode and have the same orientation in all panels. Maps below Group III in the right column show the GFP-Trap nanobody's binding epitope (top) and GFPuv's homodimerization interface (center). For reference, the ribbon diagram at bottom right depicts secondary structure elements of GFPuv, in the same orientation as other panels.

to magnetic beads, compared to single nanobodies or polyclonal anti-GFP. We therefore analyzed time courses of yeast Nup84-GFP isolations and compared the relative yields of known Nup84-complex components. The LaG-16-LaG-2 dimer showed higher yields at earlier time points, reaching approximately 80% of maximum yield after only 5 min and 90% after 10 min (**Fig. 2g** and **Supplementary Fig. 17**). These picomolar-affinity reagents open the door for increasingly rapid affinity isolations, potentially allowing the capture of weakly or transiently associated complex components for interactome studies. In addition, their high avidity would allow for the detection of low-abundance antigens, as is required for many diagnostic applications.

## DISCUSSION

Our optimized pipeline for nanobody production allows for rapid generation of a large antibody repertoire against multiple epitopes in a chosen antigen. Notably, this approach identifies high-affinity nanobody sequences directly from animal serum, taking advantage of the complex selection and maturation processes occurring in the animal's immune system and avoiding intermediary expression systems. The pipeline can quickly produce a comprehensive repertoire of specific, high-affinity nanobodies for use in the characterization of target macromolecules, such as the GFP- and mCherry-tagged proteins used in this study. Only modest laboratory effort is required after the collection of samples from llamas (50–70 d after initial immunization, once an immune response is generated). The direct work required, including IgG purification (2 d), MS (2 d), cDNA generation and PCR (2 d), and final cloning and screening (3–6 d), can be performed over approximately 10 d. Animal handling, high-throughput sequencing, MS and gene synthesis can readily be outsourced, and depending on turnaround times, each can typically be carried out in 1–2 weeks. The entire process can take as little as 4–6 weeks after an immune response is generated, with only standard techniques required in the primary laboratory. This is faster and more direct than other approaches available, which often require specialized high-throughput capability. Our LaM and LaG reagents, generated against the widely used GFP and mCherry tags, will be of general use for the affinity isolation and enhanced visualization of these tags.

Our approach is well suited to the development of nanobody reagents against various types of protein targets, including proteins that are difficult to tag. The versatility and potential of nanobodies is huge, as reflected by the interest of the research community[8,10,44,45]. Nanobodies are much smaller than antibodies, resist aggregation and can be readily 'humanized'[8,46,47]. They have great potential in drug development, as they can bind with great specificity and efficacy to disease targets such as tumor cells, either independently (as a monomer or an ultrahigh-affinity nanobody dimer) or as a fusion with other protein domains, molecules or drugs[48–51]. Application of our method for nanobody generation has the potential to significantly advance a field that otherwise can take years to generate such reagents.

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**AUTHOR CONTRIBUTIONS**
The project was conceived of by M.P.R., B.T.C., P.C.F., Y.L., D.F. and M.C.N. Experiments relating to immunization, sample collection and processing were performed by P.C.F., M.K.T. and M.O.; J.F.S. assisted with bone marrow processing. MS was carried out by Y.L. Experimental work was supervised by M.P.R. and B.T.C. Bioinformatic analysis was performed by Y.L., S.K. and D.F. Production and characterization of recombinant nanobodies was performed by P.C.F. and M.K.T. NMR analyses were performed by I.N. The manuscript was cowritten by P.C.F., Y.L., I.N., M.P.R. and B.T.C., with contributions from all authors. M.P.R. communed with llamas in the Atacama desert.

1. Cristea, I.M., Williams, R., Chait, B.T. & Rout, M.P. Fluorescent proteins as proteomic probes. *Mol. Cell. Proteomics* **4**, 1933–1941 (2005).
2. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032 (1999).
3. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
4. Domanski, M. *et al.* Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. *Biotechniques* doi:10.2144/000113864 (May 2012)
5. Gingras, A.C., Aebersold, R. & Raught, B. Advances in protein complex analysis using mass spectrometry. *J. Physiol. (Lond.)* **563**, 11–21 (2005).
6. Cortez-Retamozo, V. *et al.* Efficient cancer therapy with a nanobody-based conjugate. *Cancer Res.* **64**, 2853–2857 (2004).
7. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
8. Muyldermans, S. Nanobodies: natural single-domain antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013).
9. Harmsen, M.M. & De Haard, H.J. Properties, production, and applications of camelid single-domain antibody fragments. *Appl. Microbiol. Biotechnol.* **77**, 13–22 (2007).
10. Romer, T., Leonhardt, H. & Rothbauer, U. Engineering antibodies and proteins for molecular *in vivo* imaging. *Curr. Opin. Biotechnol.* **22**, 882–887 (2011).
11. Dumoulin, M. *et al.* Single-domain antibody fragments with high conformational stability. *Protein Sci.* **11**, 500–515 (2002).
12. Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R. & Muyldermans, S. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett.* **414**, 521–526 (1997).
13. Arbabi-Ghahroudi, M., Tanha, J. & MacKenzie, R. Prokaryotic expression of antibodies. *Cancer Metastasis Rev.* **24**, 501–519 (2005).
14. Rothbauer, U. *et al.* Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat. Methods* **3**, 887–889 (2006).
15. Muyldermans, S. *et al.* Camelid immunoglobulins and nanobody technology. *Vet. Immunol. Immunopathol.* **128**, 178–183 (2009).
16. Bird, R.E. *et al.* Single-chain antigen-binding proteins. *Science* **242**, 423–426 (1988).
17. Skerra, A. & Pluckthun, A. Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science* **240**, 1038–1041 (1988).

18. Wörn, A. & Pluckthun, A. Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989–1010 (2001).

19. Scheid, J.F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).

20. Shagin, D.A. *et al.* GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol. Biol. Evol.* **21**, 841–850 (2004).

21. Dörner, T. & Radbruch, A. Antibodies and B cell memory in viral immunity. *Immunity* **27**, 384–392 (2007).

22. Benner, R., Hijmans, W. & Haaijman, J.J. The bone marrow: the major source of serum immunoglobulins, but still a neglected site of antibody formation. *Clin. Exp. Immunol.* **46**, 1–8 (1981).

23. Becker, R.S. & Knight, K.L. Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell* **63**, 987–997 (1990).

24. Knight, K.L. Restricted VH gene usage and generation of antibody diversity in rabbit. *Annu. Rev. Immunol.* **10**, 593–616 (1992).

25. Conrath, K.E. *et al.* Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the camelidae. *Antimicrob. Agents Chemother.* **45**, 2807–2812 (2001).

26. Alvarez-Rueda, N. *et al.* Generation of llama single-domain antibodies against methotrexate, a prototypical hapten. *Mol. Immunol.* **44**, 1680–1690 (2007).

27. Brohawn, S.G., Partridge, J.R., Whittle, J.R. & Schwartz, T.U. The nuclear pore complex has entered the atomic age. *Structure* **17**, 1156–1168 (2009).

28. Fernandez-Martinez, J. *et al.* Structure-function mapping of a heptameric module in the nuclear pore complex. *J. Cell Biol.* **196**, 419–434 (2012).

29. Kirchhofer, A. *et al.* Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* **17**, 133–138 (2010).

30. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).

31. Ries, J., Kaplan, C., Platonova, E., Eghlidi, H. & Ewers, H. A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat. Methods* **9**, 582–584 (2012).

32. Dolman, N.J., Kilgore, J.A. & Davidson, M.W. A review of reagents for fluorescence microscopy of cellular compartments and structures, part I: BacMam labeling and reagents for vesicular structures. *Curr. Protoc. Cytom.* **65**, 12.30 (2013).

33. DeGrasse, J.A. *et al.* Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Mol. Cell. Proteomics* **8**, 2119–2130 (2009).

34. Matz, M.V. *et al.* Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**, 969–973 (1999).

35. Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. & Remington, S.J. Novel chromophores and buried charges control color in mFruits. *Biochemistry* **45**, 9639–9647 (2006).

36. Xia, N.S. *et al.* Bioluminescence of *Aequorea macrodactyla*, a common jellyfish species in the East China Sea. *Mar. Biotechnol. (NY)* **4**, 155–162 (2002).

37. Goldflam, M., Tarrago, T., Gairi, M. & Giralt, E. NMR studies of protein-ligand interactions. *Methods Mol. Biol.* **831**, 233–259 (2012).

38. Georgescu, J. *et al.* Backbone $H^N$, N, $C^\alpha$ and $C^\beta$ assignment of the GFPuv mutant. *J. Biomol. NMR* **25**, 161–162 (2003).

39. Khan, F., Stott, K. & Jackson, S. $^1$H, $^{15}$N and $^{13}$C backbone assignment of the green fluorescent protein (GFP). *J. Biomol. NMR* **26**, 281–282 (2003).

40. Zuiderweg, E.R. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1–7 (2002).

41. Battistutta, R., Negro, A. & Zanotti, G. Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein. *Proteins* **41**, 429–437 (2000).

42. Neri, D., Momo, M., Prospero, T. & Winter, G. High-affinity antigen binding by chelating recombinant antibodies (CRAbs). *J. Mol. Biol.* **246**, 367–373 (1995).

43. Silverman, J. *et al.* Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nat. Biotechnol.* **23**, 1556–1561 (2005).

44. Vanlandschoot, P. *et al.* Nanobodies®: new ammunition to battle viruses. *Antiviral Res.* **92**, 389–407 (2011).

45. Huang, L., Muyldermans, S. & Saerens, D. Nanobodies®: proficient tools in diagnostics. *Expert Rev. Mol. Diagn.* **10**, 777–785 (2010).

46. Revets, H., De Baetselier, P. & Muyldermans, S. Nanobodies as novel agents for cancer therapy. *Expert Opin. Biol. Ther.* **5**, 111–124 (2005).

47. Vincke, C. *et al.* General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *J. Biol. Chem.* **284**, 3273–3284 (2009).

48. Els Conrath, K., Lauwereys, M., Wyns, L. & Muyldermans, S. Camel single-domain antibodies as modular building units in bispecific and bivalent antibody constructs. *J. Biol. Chem.* **276**, 7346–7350 (2001).

49. Jähnichen, S. *et al.* CXCR4 nanobodies (VHH-based single variable domains) potently inhibit chemotaxis and HIV-1 replication and mobilize stem cells. *Proc. Natl. Acad. Sci. USA* **107**, 20565–20570 (2010).

50. Roovers, R.C. *et al.* A biparatopic anti-EGFR nanobody efficiently inhibits solid tumour growth. *Int. J. Cancer* **129**, 2013–2024 (2011).

51. Ulrichts, H. *et al.* Antithrombotic drug candidate ALX-0081 shows superior preclinical efficacy and safety compared with currently marketed antiplatelet drugs. *Blood* **118**, 757–765 (2011).

52. Ormö, M. *et al.* Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).

53. Pletneva, N.V. *et al.* Yellow fluorescent protein phiYFPv (*Phialidium*): structure and structure-based mutagenesis. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1005–1012 (2013).

54. Wall, M.A., Socolich, M. & Ranganathan, R. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat. Struct. Biol.* **7**, 1133–1138 (2000).

55. Kelley, L.A. & Sternberg, M.J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).

## ONLINE METHODS

**Isolation of $V_H$H antibodies.** For detailed information on the entire nanobody identification procedure, please see the **Supplementary Protocol**. In short, a 5 year old female llama (*Lama glama*), Barbie, was immunized with recombinant GFP-His$_6$, and a 4 year old male llama, Marley, with recombinant mCherry-His$_6$ through subcutaneous injections of 5 mg of protein with CFA. Three additional injections of 5 mg protein with IFA were performed at three week intervals. Serum bleeds were obtained 10 days after the final injection. 2.5 ml of serum was diluted tenfold in 20 mM sodium phosphate, pH 7.0, and incubated with Protein G-agarose resin for 30 min. The flow-through was then incubated for 30 min with Protein A-agarose resin. Both resins were washed with 20 mM sodium phosphate, pH 7.0, and bound $V_H$H IgG was eluted with 100 mM acetic acid, pH 4.0, and 500 mM NaCl (Protein G resin), or 100 mM acetic acid, pH 3.5, and 150 mM NaCl (Protein A resin). These elutions were pooled and dialyzed into PBS. 3 mg of this $V_H$H fraction was then incubated with Sepharose-conjugated GFP. This resin was washed with 10 mM sodium phosphate, pH 7.4, and 500 mM NaCl, followed by 1-4.5 M MgCl$_2$ in 20 mM Tris, pH 7.5, and then equilibrated in PBS. The resin was then digested with 0.3 mg/ml papain in PBS plus 10 mM cysteine, for 4 h at 37 °C. The resin was then washed with (i) 10 mM sodium phosphate, pH 7.4, and 500 mM NaCl; (ii) PBS plus 0.1% Tween-20; (iii) PBS; and (iv) 0.1 M NH$_4$OAc, 0.1 mM MgCl$_2$, 0.02% Tween-20. Bound protein was then eluted for 20 min with 0.1 M NH$_4$OH and 0.5mM EDTA, pH 8.0. These elutions were dried down in a SpeedVac and resuspended in LDS plus 25 mM DTT. The samples were alkylated with iodoacetamide and run on a 4–12% Bis-Tris gel. The ~15 kDa band corresponding to the digested $V_H$H region was then cut out and prepared for MS. All animal procedures were performed according to protocols approved by the Capralogics, Inc., Institutional Animal Care and Use Committee.

**RT-PCR and DNA sequencing.** Bone marrow aspirates were obtained from immunized llamas concurrent with serum bleeds. Bone marrow plasma cells were isolated on a Ficoll gradient using Ficoll-Paque (GE Healthcare). RNA was isolated from approximately $1 \times 10^7$ to $6 \times 10^7$ cells using Trizol LS reagent (Life Technologies), according to the manufacturer's instructions. cDNA was reverse-transcribed using Ambion RETROscript (Life Technologies). A nested PCR was then performed with IgG specific primers. In the first step, CALL001 (5′-GTCCTGGCTGCTCTTCTACAAGG-3′) and CALL002 (5′-GGTACGTGCTGTTGA ACTGTTCC-3′) primers were used to amplify the IgG variable domain into the CH2 domain[25]. The approximately 600- to 750-bp band from $V_H$H variants lacking a CH1 domain was purified on an agarose gel. Next, for 454 sequencing, $V_H$H regions were specifically reamplified using framework 1- and 4-specific primers with 5′ 454 adaptor sequences: 454-VHH-forward (5′-<u>CGTATCGCCTCCCTCGCGCCATCAG</u> ATGGCT[C/G]A[G/T]GTGCAGCTGGTGGAGTCTGG-3′) and 454-VHH-reverse (5′-<u>CTATGCGCCTTGCCAGCCCGCT CAG</u>GGAGACGGTGACCTGGGT-3′) (adaptor sequences are underlined)[25]. The approximately 400 bp product of this reaction was gel purified, then sequenced on a 454 GS FLX system after emPCR amplification, on one Pico Titer Plate. For Illumina MiSeq sequencing, the second PCR was instead performed with

random 12-mers replacing adaptor sequences, to aid in cluster identification: MiSeq-VHH-forward (5′-NNNNNNNNNNNN ATGGCT[C/G]A[G/T]GTGCAGCTGGTGGAGTCTGG-3′) and MiSeq-VHH-reverse (5′-NNNNNNNNNNNN GGAGACGGTG ACCTGGGT-3′). The product of this PCR was gel purified, ligated to MiSeq adaptors before library preparation using Illumina kits, and run on a MiSeq sequencer with $2 \times 300$-bp paired-end reads.

**Database preparation.** The protein sequence databases used for identification were prepared by translating sequencing reads in all 6 reading frames, and for each read the longest open reading frame (ORF) was selected. The selected ORF was digested with trypsin *in silico* and a list of unique tryptic peptides of 7 amino acids or longer was constructed and saved in a FASTA file. It is important to construct a FASTA file only containing unique peptides because even though most search engines can handle some sequence redundancy, they are not well equipped to handle the extreme redundancy that is provided by next-generation sequencing of the single-chain antibody locus, and search engines either become very slow or crash if presented with such extreme redundancy.

**Mass spectrometry.** Gel sections containing $V_H$H domains were excised, destained, and dehydrated. The dehydrated gel slices were then subjected to in-gel digestion with proteomic-grade trypsin (80 μL; 25 ng trypsin, 25 mM ammonium bicarbonate) (Promega) at 37 °C overnight. The gel was extracted once with extraction solution (140 μL; 67% acetonitrile, 1.7% formic acid). The resulting proteolytic digest was cleaned with a STAGE tip[56] and loaded onto a home-packed reverse phase C18 column (75 μm I.D., 15 μm tip) (New Objective) with a pressurized bomb. The loaded peptides were subsequently separated with a linear gradient (0% to 42% acetonitrile, 0.5% acetic acid, 120 min, 150 nL/min after flow splitting) generated by an Agilent 1260 HPLC and directly sprayed into an LTQ-Velos-Orbitrap mass spectrometer (Thermo Scientific) for analysis. In the mass spectrometer, a survey scan was carried out in the orbitrap (resolution = 30,000, AGC target = 1E6) followed by tandem MS in the ion trap (AGC target = 5E3) of the top twenty most intense peaks. Tandem MS was carried out with collision-induced dissociation (isolation width = 2 Th, CE = 35%, activation time = 5 ms). Internal calibration was used for improved mass accuracy (lock mass $m/z$ = 371.1012). In order to scan more peptides, both predictive AGC and dynamic exclusion were enabled (repeat counts: 2, repeat duration: 12 s, exclusion duration: 60 s). Single and unassigned charge species were excluded from tandem MS scans. The raw files were converted into mzXML format with ReAdW (version 4.3.1).

**MS-based identification of $V_H$H sequences.** The MS search was performed on the custom database of tryptic peptides using the X! Tandem search engine. Then the identified peptides, filtered by expectation value, were mapped to the sequences translated from DNA sequencing reads (longest ORF only, as described above). The CDR regions were located within the sequence on the basis of approximate position in the sequence and the presence of specific leading and trailing amino acids. For example, to locate the CDR3 region, the algorithm searched for the left anchor YXC (X representing any amino acid) between position 93 and 103 of

the sequence, and the right anchor WG between position $n - 14$ and $n - 4$ of the sequence, where $n$ is the length of the sequence. Once the peptides were mapped to the sequences and their CDR regions, a metric was calculated to rank each sequence as a potential candidate on the basis of the bioinformatics evidence available. The factors included in the metric were: MS coverage and length of individual CDR regions, with CDR3 carrying highest weight; overall coverage including framework region; and a count of the reads producing the sequence. Finally, sequences with similar CDR3 regions were grouped together, allowing for the identification of the highest-confidence sequence corresponding to a particular CDR3. A sequence was assigned to a group where its hamming distance to an existing member was 1, i.e., there was one amino acid difference in the sequence, and different groups that have one shared sequence were further combined. By choosing sequence hits from different groups for production, we maximized the overall sequence diversity of the candidate pool. The candidate list was displayed for manual inspection as an interactive HTML page with CDR regions annotated and peptide mapping information and the ranking metrics shown for each sequence. All algorithms described above were implemented in Perl. An example of a candidate list view is shown in **Supplementary Figure 3**.

**Web-based application for nanobody sequence identification: 'Llama Magic'.** The pipeline that was used for identification of the nanobody sequences has been automated and can be accessed through a web-based interface at http://www.llamamagic.org/. Source code is available at https://github.com/FenyoLab/llama-magic. Llama Magic allows upload of FASTA files containing reads from high-throughput DNA sequencing. Once uploaded, the reads will be automatically translated and digested to create an MS-searchable database of tryptic peptides, as described above. Next, the MS (MGF) files can be uploaded for a selected tryptic peptide sequence database, and the parent and fragment error can be chosen for the X! Tandem search. Once the MGF files are uploaded, the X! Tandem search will be executed and the matching peptides saved. Then (i) annotation of CDR regions, (ii) mapping of the identified peptides and (iii) ranking and grouping of candidates are performed automatically, producing an interactive display of the candidate list showing detailed information regarding each sequence and its corresponding rank. Llama Magic is implemented in Perl, HTML and JavaScript. Manual inspection was performed to make sure (i) long CDR3 peptides, which embrace both variable regions and framework regions, have fragmentation pattern within the variable regions; (ii) CDR3 peptides are unique enough (uniqueness score <100).

**Cloning.** Nanobody sequences were codon-optimized for expression in *Escherichia coli* and cloned into pCR2.1 after gene synthesis (Eurofins MWG Operon), incorporating BamHI and XhoI restriction sites at 5′ and 3′ ends, respectively. A pelB leader sequence was cloned into pET21b at NdeI and BamHI restriction sites using complementary primers: 5′-TATGAAATACTTATTG CCTACGGCAGCCGCTGGATTGTTATTACTCGCGGCCCAG CCGGCCATGGCTG-3′ and 5′-GATCCAGCCATGGCCGGCTG GGCCGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAA TAAGTATTTCA-3′. Nanobody sequences were then subcloned into pET21b-pelB using BamHI and XhoI restriction sites, with

primers also encoding a PreScission Protease (GE Healthcare) cleavage site just before the C-terminal His$_6$ tag.

**Purification of nanobodies.** pelB-fused nanobodies were expressed under a T7 promoter in Arctic Express (DE3) cells (Agilent), induced with IPTG at a final concentration of 0.1 mM. Cells were induced for 18–20 h at 12 °C, then pelleted by a 10-min spin at 5,000$g$. The periplasmic fraction was then isolated by osmotic shock[17]. This fraction was bound to His-Select nickel affinity resin (Sigma), washed with His wash buffer (20 mM sodium phosphate, pH 8.0, 1 M NaCl, 20 mM imidazole), and eluted with His elution buffer (20 mM sodium phosphate, pH 8.0, 0.5 M NaCl, 0.3 M imidazole). The elution was then dialyzed into PBS.

**Fluorescent protein–binding assays.** 2 µg of fluorescent protein were added to 50 µl of 2 mg/mL *E. coli* lysate diluted in binding buffer (20 mM HEPES, pH 7.4, 350 mM NaCl, 0.01% Tween-20, 0.1 M PMSF, 3 µg/mL pepstatin A). This was incubated with 25 µL of nanobody-Dynabead slurry. After a 30-min incubation at 4 °C, beads were washed with binding buffer and bound protein was eluted with 15 µL LDS. Elutions were run on a 4–12% Bis-Tris gel.

**$K_d$ determinations.** SPR measurements were obtained on a Proteon XPR36 Protein Interaction Array System (Bio-Rad). Recombinant GFP or mCherry was immobilized on a ProteOn GLC sensor chip: the chip surface was first activated with 50 mM sulfo-NHS and 50 mM EDC, run at a flow rate of 30 µL/min for 300 s. The ligand was then diluted to 5 µg/mL in 10 mM sodium acetate, pH 5.0, and injected at 25 µL/min for 180 s. Finally, the surface was deactivated by running 1 M ethanolamine-HCl (pH 8.5) at 30 µL/min for 300 s. This led to immobilization of approximately 600–800 response units (RU) of ligand.

$K_d$ values of recombinant nanobodies were determined by injecting 4 or 5 concentrations of each protein, in triplicate, with a running buffer of 20 mM HEPES, pH 8.0/150 mM NaCl/0.01% Tween. Proteins were injected at 50 µL/min for 120 s, or 100 µL/min for 90 s, followed by a dissociation time of 600 s. Between injections, residual bound protein was eliminated by regeneration with 4.5 M MgCl$_2$ in 10 mM Tris, pH 7.5, run at 100 µL/min for 36 s. Binding sensorgrams from these injections were processed and analyzed using the ProteOn Manager software. Binding curves were fit to the data with a Langmuir model, using grouped on rate, off rate and $R_{max}$ values.

**Immunofluorescence microscopy.** HeLa cells were cultured on coverslips in DMEM medium with 10% FBS and penicillin/ streptomycin at 37 °C with 8% CO$_2$ in a humidified environment. Cells tested negative for mycoplasma. Cells were transfected with CellLight tubulin-GFP or mitochondria-GFP BacMam 2.0 reagents (Life Technologies) using 4 µL of reagent per 5,000 cells, and processed after 18–20 h. Cells were fixed in ice-cold methanol for 10 min (for tubulin-GFP) or in 2% paraformaldehyde for 10 min (for mitochondria-GFP). Cells were permeabilized with 0.5% Triton X-100 for 10 min. and blocked for 1 h with 10% goat serum/1% BSA in PBS. They were then incubated for 1 h at room temperature with recombinant nanobody conjugated to Alexa Fluor 568 succinimidyl ester (Life Technologies), diluted to 100 ng/mL in 1% BSA in PBS. Cells were washed four times

with PBS/0.01% BSA, with 300 nM 4,6-diamidino-2-phenylindole (DAPI) included in the final wash, then mounted with ProLong Diamond (Life Technologies).

Wild-type and Sec13-GFP–tagged *T. brucei* strains were cultured to a cell density of $1 \times 10^7$ as previously described[33]. Cells from each strain were mixed 1:1 and fixed for 10 min with cold 4% formaldehyde. Approximately $1 \times 10^6$ cells were spotted onto coverslips, allowed to settle for 30 min, permeabilized with 0.1% Triton for 5 min, and blocked with 10% goat serum/1% BSA in PBS for 30 min. Cells were then stained, washed, and mounted identically to HeLa cells.

An *S. cerevisiae* W303 strain with Htb2 genomically tagged at the C terminus with mCherry was grown to mid-log phase and allowed to settle on concanavalin A–coated coverslips. Yeast were fixed in 4% paraformaldehyde/2% sucrose/PBS and blocked and permeabilized for 30 min in 0.25% Triton/2% milk/PBS[31]. Cells were stained overnight at 4 °C with nanobody diluted to 3.3 μg/mL in 0.25% Triton/1% BSA/PBS. They were then washed five times with 0.01% BSA in PBS, the final two washes for 5 min. Cells were mounted in 70% glycerol/PBS.

All images were obtained on a DeltaVision Image Restoration microscope (Applied Precision/Olympus), with an Olympus 100×/1.40–numerical aperture (NA) objective, or 60×/1.42 objective in the case of HeLa cells. Raw images were processed by a deconvolution algorithm using softWorX software (Applied Precision/GE Healthcare).

**Affinity isolations of tagged protein complexes.** Recombinant nanobodies were conjugated to epoxy-activated magnetic Dynabeads (Life Technologies), with minor modifications to published IgG coupling conditions[57]. 10 μg recombinant protein were used per 1 mg of Dynabeads, with conjugations carried out in 0.1 M sodium phosphate, pH 8.0, and 1 M ammonium sulfate, with an 18- to 20-h incubation at 30 °C. Affinity isolations of yeast Nup84-GFP were carried out as previously described, using binding buffer consisting of 20 mM HEPES, pH 7.4, 500 mM NaCl, 2 mM MgCl$_2$, 0.1% CHAPS, 0.1M PMSF, and 3 μg/mL pepstatin A[57]. For each experiment, 50 μL of bead slurry were used with 0.5 g of yeast cells. Similar conditions were used for HTB2-mCherry isolations (from yeast with HTB2 genomically tagged at the C terminus with mCherry[58]), except lysate was sonicated four times for 10 s before centrifugation, and the binding buffer consisted of 20 mM HEPES, pH 8.0, 300 mM NaCl, 110 mM KOAc, 0.1% Tween-20, 0.1% Triton X-100, 0.1 M PMSF, and 3 μg/mL pepstatin A. Isolations of RBM7-GFP from HeLa cells were performed as previously described[4]. 10 μl of bead slurry were used with 100 mg of cells, in a binding buffer of 20 mM HEPES, pH 7.4, 300 mM NaCl, 0.5% Triton X-100, with Complete Protease Inhibitor, EDTA free (Roche).

For Nup84-GFP isolations, yield and signal-to-noise ratio were quantified by calculating band densities from SDS-PAGE gels using ImageJ software. Signal or yield was defined as the combined densities of Nup84-GFP, Nup120, and Nup85/Nup145c bands. Noise was estimated from the density of the region of the lane between Nup85/Nup145c and Seh1, which does not correlate to known Nup84-complex components. Experiments were performed in duplicate.

To determine affinity isolation yields, samples of resuspended lysate were taken before and after Dynabead binding. These were run on a 4–12% Novex Bis-Tris gel in MES running buffer (Life Technologies) and probed by western blotting using mouse anti-GFP antibody (Roche, cat. no. 11814460001) diluted 1:1,000 in TBST/2% dry milk, and an anti-mouse, HRP-conjugated secondary antibody (GE Healthcare, cat. no. NA931V) diluted 1:3,000 in TBST/2% dry milk. Signals were quantified using ImageJ software.

**Fluorescence spectra.** Samples of recombinant GFP at 0.5 μM in PBS were mixed with either buffer or 10 μM of a LaG protein. Fluorescence spectra were obtained on a Synergy Neo (BioTek) microplate reader. Excitation spectra from 300 nm to 530 nm were taken at an emission wavelength of 560 nm, and emission spectra were measured from 450 nm to 600 nm at an excitation wavelength of 425 nm.

**Phylogenetic analysis.** Phylogenetic trees and alignments were generated from LaG amino acid sequences using the Phylogeny.fr web service[59,60].

**Mapping of nanobody binding epitopes on GFP by NMR.** Three variants of GFP were used in the preparation of NMR samples: GFP-His$_6$ (EGFP), the variant used for immunization; GFPuv, the variant for which backbone $^{15}$N-$^1$H chemical shift assignments were available from BMRB 5666 (ref. 39) and a crystal structure was available from PBD 1B9C (ref. 41); and GFPuv_A206K (GFPuv_M), a monomeric version of GFPuv[61]. **Supplementary Table 1** summarizes the amino acid sequences of the three GFP variants.

All NMR samples contained between 500 and 20 μM $^{15}$N-GFP either alone or in the presence of a 1–1.2 molar excess of LaG, 10 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl and 90% H$_2$O/10% D$_2$O. All NMR spectra (2D $^1$H-$^{15}$N HSQC) were measured at 310 K on a Bruker Avance DPX-600 MHz spectrometer equipped with a TCI cryoprobe.

Backbone $^1$H-$^{15}$N assignments of GFPuv were obtained from a comparison between a $^1$H-$^{15}$N HSQC spectrum of GFPuv alone and a simulated $^1$H-$^{15}$N HSQC based on BMRB 5666 (ref. 39) (**Supplementary Fig. 14a**). Owing to a very high similarity between the two, $^1$H-$^{15}$N backbone assignment of GFPuv was obtained for 97% of $^1$H-$^{15}$N backbone resonances for which assignment was available in BMRB 5666. The accuracy of the GFPuv assignment was verified by mapping the binding site of a previously identified nanobody, GFP-Trap[29], on GFPuv. The crystal structure of the GFP/GFP-Trap complex is available from PDB 3K1K (ref. 29). The X-ray crystallography–derived binding site (obtained by analysis of PDB 3K1K by the Protein Interfaces, Surfaces and Assemblies (PISA) service at the European Bioinformatics Institute[62]; http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) overlaps with the one determined by the chemical shift perturbation method, thereby confirming our assignment of GFPuv residues (**Supplementary Fig. 16**).

Backbone $^1$H-$^{15}$N assignments of GFPuv_M were obtained from a comparison between a $^1$H-$^{15}$N HSQC spectrum of GFPuv and that of GFPuv_M (**Supplementary Fig. 14a**). Assignment was verified by mapping the dimerization site of GFPuv and comparing it to the crystal structure of PDB 1B9C (ref. 41) (analyzed for interacting residues using PISA[62]).

All chemical shift differences were calculated using equation (1)

$$CSD = \sqrt{\frac{\left(\frac{\Delta\delta N}{5}\right)^2 + \Delta\delta H^2}{2}} \quad (1)$$

where CSD is the total chemical shift difference and $\Delta\delta N$ and $\Delta\delta H$ are the chemical shift differences in the free and bound states between the amide nitrogens and protons, respectively. The CSD cutoff for binding-site residues was 0.05 p.p.m. for GFP-Trap binding site and for GFPuv dimerization site and 0.03 p.p.m. for all LaG binding sites.

All LaG binding-site residues are listed in **Supplementary Table 2**, and their respective $^1$H-$^{15}$N HSQC spectra are shown in **Supplementary Figure 15** overlaid with the $^1$H-$^{15}$N HSQC spectrum of the free GFPuv_M.

56. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
57. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).
58. Rout, M.P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
59. Dereeper, A., Audic, S., Claverie, J.M. & Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* **10**, 8 (2010).
60. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
61. Zacharias, D.A., Violin, J.D., Newton, A.C. & Tsien, R.Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
62. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).