Protein Engineering, Design & Selection, 2016, vol. 29 no. 12, pp. 563–572
doi: 10.1093/protein/gzw054
Advance Access Publication Date: 13 October 2016
Original Article

Original Article

Engineering a targeted delivery platform using Centyrins

Shalom D. Goldberg, Rosa M.F. Cardoso, Tricia Lin, Tracy Spinka-Doms, Donna Klein, Steven A. Jacobs, Vadim Dudkin, Gary Gilliland, and Karyn T. O'Neil*

Janssen Research and Development, L.L.C., 1400 McKean Road, Spring House, PA 19477, USA

*To whom correspondence should be addressed. E-mail: KOneil@its.jnj.com
Edited by: Daniel Raleigh

Received and Revised 7 September 2016; Accepted 20 September 2016

Abstract

Targeted delivery of therapeutic payloads to specific tissues and cell types is an important component of modern pharmaceutical development. Antibodies or other scaffold proteins can provide the cellular address for delivering a covalently linked therapeutic via specific binding to cell-surface receptors. Optimization of the conjugation site on the targeting protein, linker chemistry and intracellular trafficking pathways can all influence the efficiency of delivery and potency of the drug candidate. In this study, we describe a comprehensive engineering experiment for an EGFR binding Centyrin, a highly stable fibronectin type III (FN3) domain, wherein all possible single-cysteine replacements were evaluated for expression, purification, conjugation efficiency, retention of target binding, biophysical properties and delivery of a cytotoxic small molecule payload. Overall, 26 of the 94 positions were identified as ideal for cysteine modification, conjugation and drug delivery. Conjugation-tolerant positions were mapped onto a crystal structure of the Centyrin, providing a structural context for interpretation of the mutagenesis experiment and providing a foundation for a Centyrin-targeted delivery platform.

Key words: centyrin, cysteine scanning, EGFR, FN3 domain, targeted delivery

Introduction

Targeted delivery is a key component of many emerging therapeutics. Technologies that enable delivery of a drug payload to a specific cell type can play an important role in the therapeutic index of a drug candidate and enable clinical development of drug candidates that would otherwise have significant toxicities. Some therapeutic platforms, including many nucleic acid modalities, are unable to cross cellular membranes and require a delivery mechanism to provide access to their intracellular targets. Antibodies are a classical targeting moiety and significant advances have been made over the last several years to develop antibody drug conjugates (ADCs) for delivery of cytotoxic payloads to tumors. Two ADCs have been approved by the FDA since 2011, and over 30 others are in clinical trials (Beck and Reichert, 2014). While ADCs are generally considered to be the most advanced targeted delivery platform, protein conjugates have alternative applications in biopharmaceuticals including therapeutics, diagnostics and tool reagents (Steiner and Neri, 2011).

'Alternative scaffold' proteins—non-antibody proteins that can be engineered to bind to a variety of therapeutic targets with high affinities—have great potential utility for targeted delivery due to their small size, lack of disulfide bonds and stability. These properties allow for the use of novel methods of purification; improved penetration into tissues; and easy formatting into multi-specific molecules (Binz and Pluckthun, 2005; Skerra, 2000). In addition, homogenous alternative scaffold protein conjugates can often be prepared in a straightforward and site-specific manner.

Covalent conjugation to proteins is often accomplished via lysine side chains using an activated ester or isothiocyanate, or via cysteine side chains with a maleimide, haloacetyl derivative or activated

disulfide (Brinkley, 1992). However, since most proteins have multiple lysine and cysteine residues, heterogeneous mixtures of product with different numbers of conjugated molecules at a variety of amino-acid positions often result. The heterogeneity of the first generation of ADCs led to manufacturing complications and, in some cases, deleterious effects on biological activity (Agarwal and Bertozzi, 2015). More recently, techniques for generating site-specific protein conjugates have been developed (Rabuka, 2010), including incorporation of unnatural amino acids, fusion of the protein of interest to a 'self-labeling tag' such as SNAP or DHFR or a tag that is recognized and modified specifically by another enzyme such as sortase A, lipoic acid ligase and formylglycine-generating enzyme. These methods are being applied to protein conjugates generally and to ADCs in particular (Mack *et al.*, 2014). Recent reports suggest that the conjugation site on the antibody can have significant effects on the activity of the conjugate and highlight the importance of careful consideration of conjugation site for producing optimized molecules (Shen *et al.*, 2012; Strop *et al.*, 2013).

Smaller, alternative scaffold proteins may be especially well-suited for targeted delivery applications. Many scaffolds lack cysteine residues, thereby allowing introduction of unique cysteines to enable site-specific chemical conjugation. The Centyrin, a consensus fibronectin type III (FN3) domain sequence based on human tenascin (Jacobs *et al.*, 2012), has been engineered to bind to multiple cytokines and cell-surface receptors with nanomolar and sub-nanomolar affinity (Diem *et al.*, 2014). They are easily expressed in *Escherichia coli* and can be linked together into bi- and multi-specific formats without loss of activity. Centyrins also possess excellent biophysical properties including stability to heat, pH, denaturant and organic solvents, reversible unfolding and monodispersity. Centyrins have no cysteine residues and are thus amenable to conjugate production by introduction of a unique cysteine residue at a specific site. This approach has been demonstrated previously for other FN3-based scaffolds (Mamluk *et al.*, 2010). We anticipated that the excellent biophysical properties of Centyrins would allow for mutation and conjugation at one or even multiple positions in the scaffold. Additionally, the small size (~100 residues) and ease of expressing Centyrins allows for all amino-acid positions in the protein to be assessed in order to identify the optimal conjugation sites. Here, we present a high-throughput method for comprehensive interrogation of the entire Centyrin domain to identify ideal sites for conjugation to a maleimide-linked small molecule with tolerant positions mapped onto a crystal structure of the parent Centyrin. We demonstrate that the Centyrin scaffold is highly amenable to conjugation, with 26 of the 94 positions tolerant to cysteine mutation and payload conjugation with little effect on biophysical properties or biological activity. These data provide the foundation for a targeted delivery platform based on Centyrin domains.

Materials and Methods

Gene synthesis

The amino-acid sequence of the parent Centyrin P54AR4-83v2 (83v2) with a C-terminal His_6 tag was back-translated into a nucleic acid sequence encoding the protein using preferred codons for *E. coli* expression, and a synthetic gene was produced (DNA 2.0). Genes of the cysteine variants and of the W26A variant were identical to the parent protein with the exception of the cysteine or alanine codon. All genes were cloned into a pJexpress401 vector (DNA 2.0) for expression driven by a T5 promoter sequence and transformed into *E. coli* strain DH10B or BL21 GOLD. Glycerol stocks of all members of the 83v2 'Cys-scan' library were arrayed sequentially into a 96-well plate.

Crystallization and structure determination

The His_6-tagged 83v2 W26A Centyrin was expressed in *E. coli* BL21 GOLD cells (Agilent). A one-liter culture in Terrific Broth (TB; Teknova) was grown at 37°C with shaking to an OD_{600} of 1.2–1.4, and then induced with isopropyl thiogalactoside (IPTG) at 1 mM and the temperature reduced to 30°C. Cells were harvested after ~20 h of growth and then lysed with BugBuster HT protein extraction reagent (Novagen) supplemented with 0.2 mg/ml recombinant lysozyme (Sigma). Centyrin was purified from clarified lysate by nickel affinity chromatography with Superflow resin (Qiagen) followed by size-exclusion chromatography in phosphate-buffered saline (PBS) on a TSKgel G3000SW 21.5 × 600 mm column (Tosoh Bioscience). The buffer was exchanged to 50 mM NaCl, 20 mM HEPES pH 7.4 by dialysis and the protein was concentrated to 15.7 mg/ml by ultrafiltration (Amicon Ultra-4; 3 kDa cutoff) for crystallization trials.

After several rounds of microseeding, crystals suitable for X-ray diffraction were obtained at 20°C from 25% PEG 3 kDa, 1 M sodium acetate, 0.1 M Tris pH 8.5 using the sitting drop vapor-diffusion method and equal drop volumes of protein and well solutions. For X-ray data collection, crystals were soaked for a few seconds in a cryo-protectant solution containing mother liquor supplemented with 25% glycerol, and then flash frozen in liquid nitrogen. X-ray diffraction data were collected using a Pilatus 6 M detector on beamline 17-ID at the Advanced Photon Source (Argonne National Laboratory) from one cryogenically frozen crystal, and the diffraction data were processed with the program HKL2000 (Otwinowski and Minor, 1997). The structure was determined by molecular replacement with Phaser (McCoy *et al.*, 2007) using the coordinates of another one of our anti-EGFR Centyrins (data not shown) as search model. The structure was refined with PHENIX (Adams *et al.*, 2010) and model adjustments were carried out using COOT (Emsley and Cowtan, 2004). All other crystallographic calculations were performed with the CCP4 suite of programs (Krissinel *et al.*, 2004). All molecular graphics were generated with the PyMOL Molecular Graphics System, version 1.8 (Schrödinger, LLC).

Accession numbers

The atomic coordinates and structure factors for the W26A variant of Centyrin 83v2 have been deposited in the RCSB Protein Data Bank (accession code 5L2H).

High-throughput expression and purification

One milliliter of LB broth in a 96-well DeepWell block supplemented with kanamycin (50 μg/ml) was inoculated with bacteria from frozen glycerol stocks and grown overnight at 37°C with shaking. The next day, 50 μl of the overnight culture was added to 5 ml of TB in 24-well blocks supplemented with 50 μg/ml kanamycin and the culture was grown at 37°C with shaking to an OD_{600} of 1.2–1.4. IPTG was then added to a final concentration of 1 mM and the temperature reduced to 30°C. Cells were harvested after ~20 h of growth by centrifugation at 2250× *g* for 15 min. Cell pellets were resuspended in 600 μl of BugBuster HT (Novagen) supplemented with rLysozyme (Novagen) to 1 kU/ml and β-mercaptoethanol to

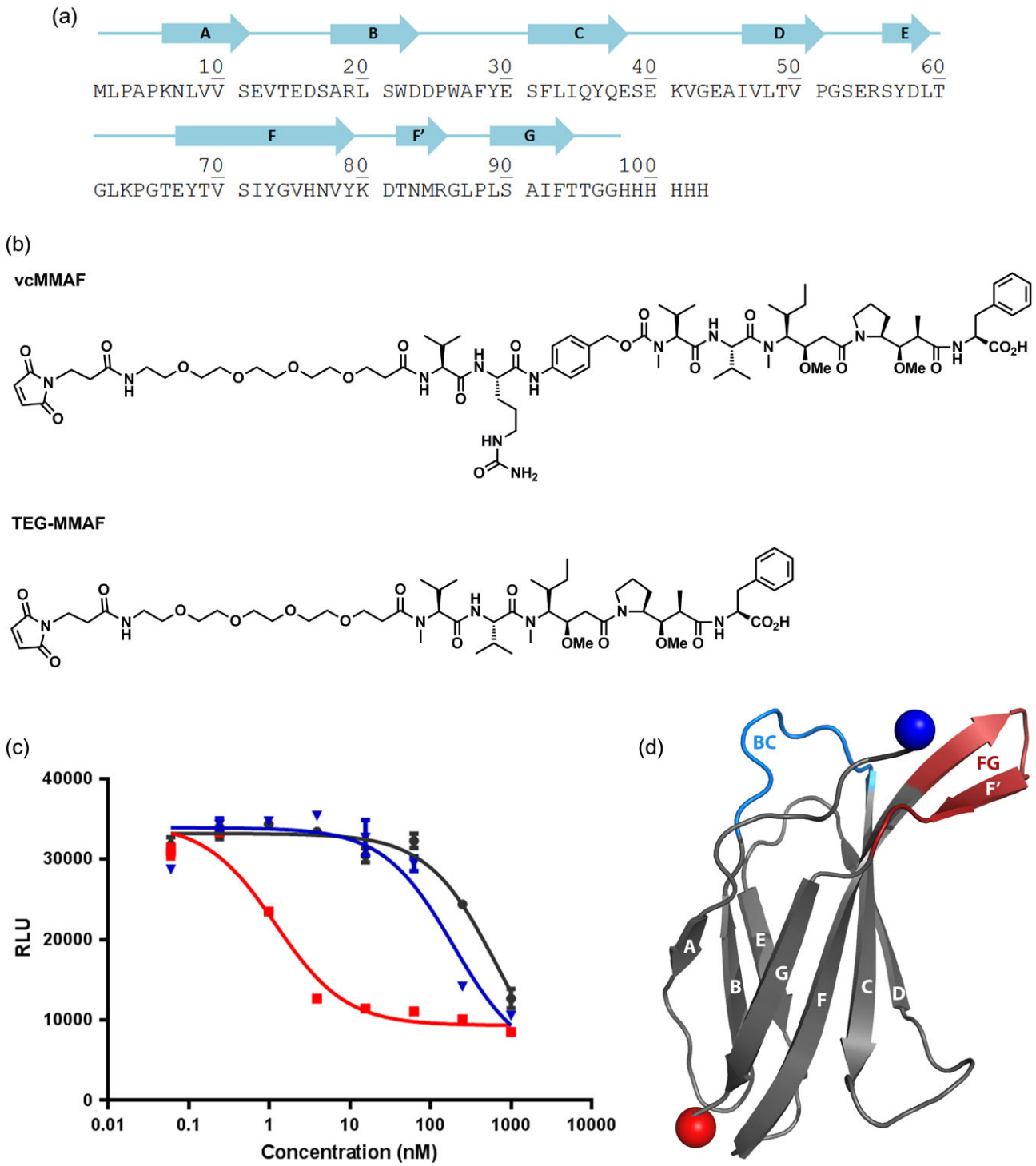

Fig. 1 (**a**) Sequence of 83v2 parent Centyrin. The secondary structure of the W26A variant of 83v2 is shown above the sequence with arrows and lines representing β-strands and loops, respectively. A unique F′ strand is observed in the 'FG loop' region of W26A. (**b**) Structures of maleimido-tetraethylene glycol-valine-citrulline-para-amino benzyloxycarbonyl-MMAF (vcMMAF) and maleimido-tetraethylene glycol-MMAF (TEG-MMAF). (**c**) Cytotoxicity of Centyrin-drug conjugate 83v2-vcMMAF. The Centyrin 83v2, modified with a C-terminal cysteine, was conjugated to MMAF via an enzyme-cleavable valine-citrulline linker using thiol-maleimide chemistry. Cytotoxicity of 83v2-vcMMAF (red) is shown in comparison with free MMAF (blue) and with a drug conjugate produced with a non-targeted Centyrin (TenconWT-vcMMAF; black) in the H1573 cell line, a lung cancer line that has high EGFR expression and a KRAS mutation. The IC_{50} of 83v2 is 2.6 nM; free MMAF is ~260 nM and Tencon1-vcMMAF is >500 nM. (**d**) Crystal structure of the W26A variant of 83v2. The loops that were randomized in the library used for discovery of 83v2 are shown in blue (BC loop) and red (FG loop). The β-strands are labeled A to G, and N- and C-termini are shown as blue and red spheres, respectively.

10 mM, and gently mixed at room temperature for 60 min. Lysate was clarified by centrifugation at 2250× *g* for 15 min. Centyrins in clarified lysate were bound to Ni-NTA resin in 96-well format using His-trap HP plates (GE Healthcare) by adding lysate to the wells and centrifugation at 100× *g* for 4 min. The resin was washed three times with buffer A (50 mM sodium phosphate pH 7.0, 0.5 M NaCl, 10 mM imidazole) supplemented with 10 mM β-mercaptoethanol, and then Centyrin was eluted with 2× 150 µl of buffer B (50 mM sodium phosphate pH 7.0, 0.5 M NaCl, 250 mM imidazole) supplemented with 10 mM β-mercaptoethanol and exchanged into PBS with 96-well PD MultiTrap plates (GE Healthcare).

Conjugation

Conjugation was integrated into the purification process as follows: bacteria was grown, induced and lysed as described above. Centyrins in clarified lysate were bound to Ni-NTA resin in 96-well format using His-trap HP plates by adding lysate to the wells and centrifugation at 100× *g* for 5 min. The resin was washed three times with buffer A, and then the maleimide-containing molecule—*N*-ethyl maleimide (NEM); the cytotoxic tubulin inhibitor momomethyl auristatin F (MMAF, Fig. 1b) with enzyme-cleavable valine-citrulline linker (vcMMAF); or MMAF with a non-cleavable tetraethylene glycol (TEG) linker (TEG-MMAF)—was added as 500 µl of a 1.5-mM solution. Following a 1-h RT incubation on a rotisserie shaker, excess maleimide was removed by centrifugation and three washes with buffer A. Conjugated Centyrin was eluted with 2× 150 µl of buffer B and exchanged into PBS with 96-well PD MultiTrap plates (GE Healthcare).

Mass spectrometry

The proteins were analyzed using an Agilent Model G6224 MS-TOF mass spectrometer system. The instrument was operated in positive electro-spray ionization mode and scanned from 600 to 3000 *m/z*. Instrument settings included: 4000 V as the capillary voltage; 250 V for the fragmentor; 65 V for the skimmer; 325°C as the gas temperature; drying gas flow of 12 l/min; nebulizer pressure of 45 psig. The mobile phase was 0.1% trifluoroacetic acid (TFA) in water (solution A) and 0.1% TFA in acetonitrile (solution B). The column was an Agilent Poroshell 120 EC-C18 2.7 µm, 50 × 4.6 mm running with a linear gradient of 20–80% of solution B at a flow rate of 0.6 ml/min and with an injection volume of 10 µl.

Analytical size-exclusion chromatography

Size-exclusion chromatography was used to determine the aggregation state of the Centyrin-NEM conjugates. Aliquots (10 µl) of each purified Centyrin were injected onto a Superdex 75 5/150 column (GE Healthcare) at a flow rate of 0.3 ml/min in a mobile phase of PBS pH 7. Elution from the column was monitored by absorbance at 280 nm.

EGFR binding assay

Relative binding affinity of the NEM-conjugated cysteine variants to EGFR was estimated by enzyme-linked immunosorbent assay (ELISA). Maxisorp plates (Nunc) were coated with 100 µl of 5 µg/ml EGFR-Fc (R&D Systems) overnight, then blocked using 200 µl Starting Block T20 (Thermo Fisher) for 1 h. Plates were washed three times with Tris-Buffered Saline, pH 7.4 with 0.05% Tween-20 (TBST). Centyrins were diluted to 500, 100 or 10 nM, and 100 µl was added to each well. After 1 h of incubation with gentle shaking, plates were washed three times with TBST. Anti-His IgG (Qiagen) was diluted 1:1000 in Starting Block, and 100 µl added per well. Plate was incubated with the antibody for 1 h—with gentle shaking and then washed three times with TBST. Peroxidase-affiniPure Goat anti-mouse IgG conjugated to HRP (Jackson Immuno Research) was diluted 1:5000 in Starting Block and 100 µl added per well. Plate was incubated with the antibody for 1 h with gentle shaking and then washed three times with TBST. Plate was treated with POD substrate (Roche) at 50 µl/well, and luminescence measured on an M5 Spectramax plate reader.

Thermal stability

The thermal stability of Centyrin-NEM conjugates was assessed by differential scanning calorimetry (DSC). Melting temperature was determined by heating 400 µl of the Centyrin solution from 25 to 100°C at a scan rate of 1°C/min in a VP-DSC instrument (MicroCal). Data were fit to a two-state unfolding model in order to calculate the melting temperature. After cooling, a second identical scan was completed on the sample in order to assess the degree of refolding.

Cytotoxicity assay

83v2 cysteine variants were conjugated to vcMMAF or TEG-MMAF as described above. Cell killing was assessed by measuring viability of the EGFR-overexpressing human tumor cell line NCI-H1573 (ATCC) following exposure to the Centyrin-MMAF conjugates. Cells were plated in black-well, clear bottomed, tissue culture-treated plates (Falcon) at 7000 cells/well in 100 µl/well of phenol red-free RPMI media (Gibco) with 5% fetal bovine serum (Gibco). Cells were allowed to attach overnight at 37°C in a humidified 5% CO_2 atmosphere. Medium was aspirated from 96-well plate and cells were treated with 50 µl of fresh media and 50 µl of 2× inhibitor in fresh media. Cell viability was determined by an end-point assay with Cell TiterGlo (Promega) at 70 h. IC_{50} values were determined by fitting data to the equation for a sigmoidal dose response with variable slope using GraphPad Prism software.

Results

The Centyrin 83v2 (Fig. 1) binds with high affinity (K_d ~0.1 nM) to human epidermal growth factor receptor (EGFR), a receptor tyrosine kinase that is overexpressed in many cancers (Supplementary Fig. S1a). 83v2 was identified by *in vitro* panning of Centyrin libraries that were constructed by randomizing the BC and FG loops of the Centyrin scaffold (Anderson *et al.*, 2014). 83v2 binds to $EGFR^+$ cells with high affinity and when conjugated to the cytotoxic tubulin inhibitor MMAF via a Val-Cit linker (83v2-vcMMAF), enables delivery of MMAF into cells in a receptor-dependent manner (Fig. 1c). Though 83v2 itself proved difficult to crystallize, a point mutant (W26A) was crystallized and the structure was solved (Fig. 1d). These factors make 83v2 an attractive candidate for the cysteine-scanning analysis.

Crystal structure of 83v2 W26A

The crystal structure of the W26A variant of Centyrin 83v2 was determined to 1.8 Å resolution (Supplementary Table SI, Fig. 1d). The W26A variant binds to purified EGFR with high affinity though with an ~10-fold reduction relative to the parent 83v2 protein (K_d ~1.5 nM, Supplementary Fig. S1b). The structure represents the full-

length Centyrin (residues 1–95) and it contains four structurally similar copies of the Centyrin monomer per asymmetric unit (root mean square deviation <1.0 Å for Cα superposition of monomers). The electron density map is unambiguous and allows clear positioning of the Centyrin amino-acid residues. The C-terminal His_6 tag is disordered, but the G96-G97 linker is visible and included in the structure.

The W26A Centyrin has the expected Ig-like fold of the fibronectin type III domain (Jacobs, *et al.*, 2012), a β-sandwich structure comprised of two anti-parallel β-sheets containing three and four β-stands (Fig. 1d). One sheet forms a concave face composed of strands C, D, G, F and the other forms a convex face composed of strands A, B, E. The N and C-termini are located at opposite ends of the molecule.

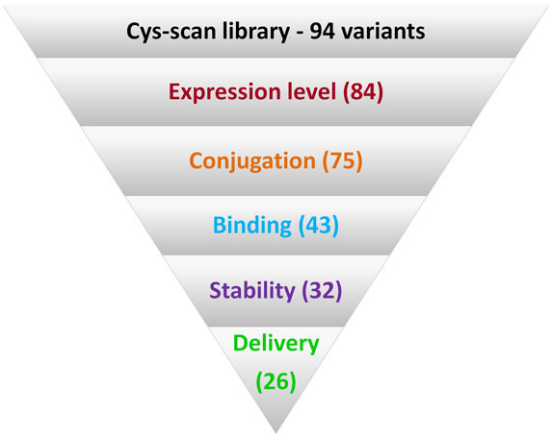

Fig. 2 Scheme for characterization of cys-scan library. A library of 94 single-cysteine variants was evaluated sequentially for expression in *E. coli*, conjugation to maleimides, binding to target receptor in a biochemical ELISA, biophysical properties (thermal stability and aggregation state) and functional delivery of payload in a cell-killing assay. Variants that did not meet the criteria at each step were excluded from further consideration. The number of variants that met the criteria at each step is shown in parentheses.

The BC, DE and FG loops, analogous to the complementarity-determining regions (CDRs) of antibodies, connect the β-strands on the N-terminal side of the protein. The FG loop of W26A is five amino acids longer than the typical FG loop of the Centyrin scaffold. In W26A, it protrudes from the Centyrin with a finger-like shape made up of two anti-parallel β-strands, one of which is a continuation of the core F strand while the other is a unique F' strand, which has not been seen in other Centyrin structures (Jacobs, *et al.*, 2012, Luo *et al.*, 2014, Teplyakov *et al.*, 2014).

Library design

A comprehensive library of 94 single-cysteine mutants of 83v2 was produced by gene synthesis. This covers every position of the 83v2 protein excluding the N-terminal methionine. These variants were compared with an 83v2 variant in which the cysteine is included between the scaffold and the His tag (G96C) and another in which the cysteine is placed at the C-terminus of the protein immediately after the His tag (denoted 83v2-CTC). The genes are identical at all positions except for the codon that is mutated to encode the cysteine. Genes were cloned into the pJ401 expression vector and transformed into *E. coli* for expression. The 96 variants were evaluated for purification yield, conjugation efficiency, binding to target, biophysical properties and delivery of a cytotoxic payload to $EGFR^+$ cells. All properties except expression level (measured with free protein) and cytotoxicity (measured with MMAF conjugate) were evaluated with NEM as a surrogate for the cytotoxic payload since it was necessary to minimize exposure to the potent cytotoxic. Variants with properties that were negatively altered relative to those of the parent protein (83v2-CTC) at each stage were excluded from further consideration. Filters were applied sequentially to identify a set of cysteine variants that met all of the desired criteria and the substitutions could be considered appropriate sites for conjugation (Fig. 2).

Protein expression

83v2 cysteine variants were expressed in 5 ml Terrific Broth cultures in 24-well plates and purified in 96-well format using His MultiTrap plates containing Ni-NTA resin. Expression of soluble

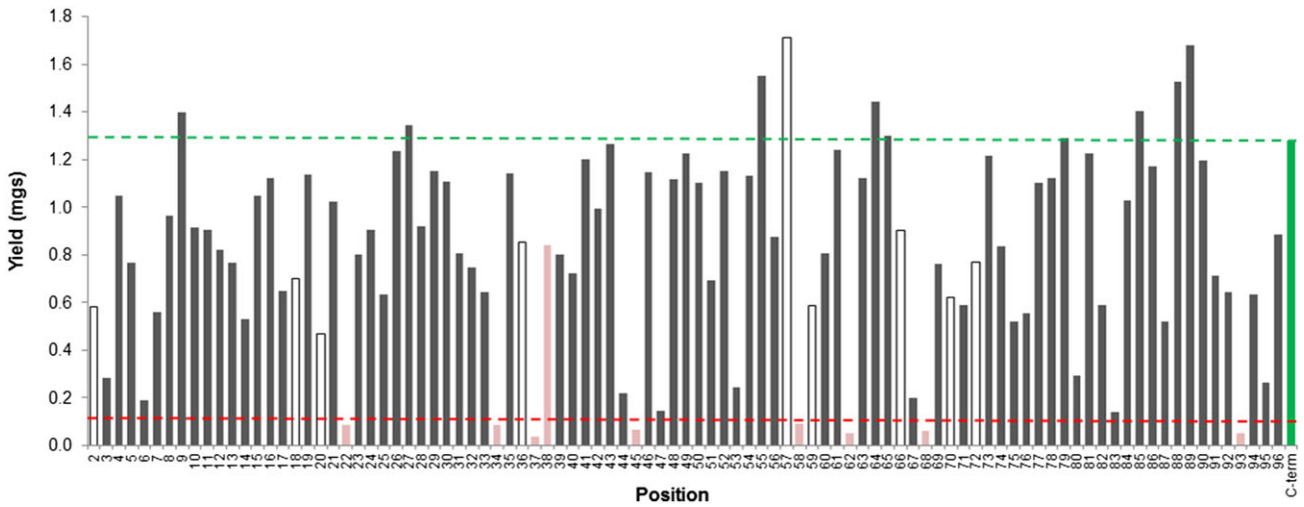

Fig. 3 Expression and conjugation of cys-scan variants. Protein yield after expression in 5 ml cultures of *E. coli* and high-throughput Ni-NTA purification is shown for each variant. Centyrin 83v2 with C-terminal cysteine is shown in green; all variants with yields <0.1 mg (red line) are shown in light red. Variants with poor conjugation efficiency (<80% conjugated by mass spectrometry) are shown in white with outline.

Centyrin was determined for each variant by UV-Vis spectrophotometry of the elutions from the Ni-NTA plate (Fig. 3). The parent '83-cys' Centyrin produced 1.28 mg of protein from the 5-ml culture (256 mg/l). Ten of the variants had equal or higher yields relative to 83v2-CTC with an additional 26 having at least 80%. The eight variants with expression levels <0.1 mg/5 ml—W22C, I34C, Q37C, A45C, D58C, L62C, Y68C and F93C—were excluded from further consideration. E67C and T95C expressed at low levels though slightly above the cutoff, and were also excluded from further consideration.

Conjugation efficiency

Conjugation of the variants to NEM was incorporated into the Ni-NTA plate purification. Cells were lysed in the presence of β-mercaptoethanol to ensure that the cysteines remain in the reduced, free thiol form. After binding to the resin, Centyrins were washed extensively to remove reducing agent and then exposed to NEM, followed by additional washing and then elution. The conjugation efficiency of each variant was assessed by mass spectrometry (data not shown). Most variants conjugated efficiently to NEM (>80% conjugated by mass spectrometry) under the tested high-throughput conditions (Fig. 3). Eight variants (A18C, L20C, Y36C, Y57C, L59C, T66C, V70C and I72C) showed poor conjugation efficiency and were excluded from further characterization. In addition, L2C had no detectable conjugate or Centyrin by mass spectrometry, possibly indicating poor protein or conjugate stability, and was excluded from further characterization.

EGFR binding

Binding of NEM-conjugated Centyrins to recombinant EGFR was measured by ELISA. A full dose–response curve was generated for the parent 83v2 protein (Fig. 4a), and this curve was used to select the concentration (10 nM) for characterizing the variants. The criterion for considering a variant to have retained target binding was initially set as >60% of the raw ELISA signal of the parent at 10 nM.

Of the 77 variants tested for EGFR binding, 57 had at least 60% of the ELISA signal of the parent 83-cys protein when evaluated at

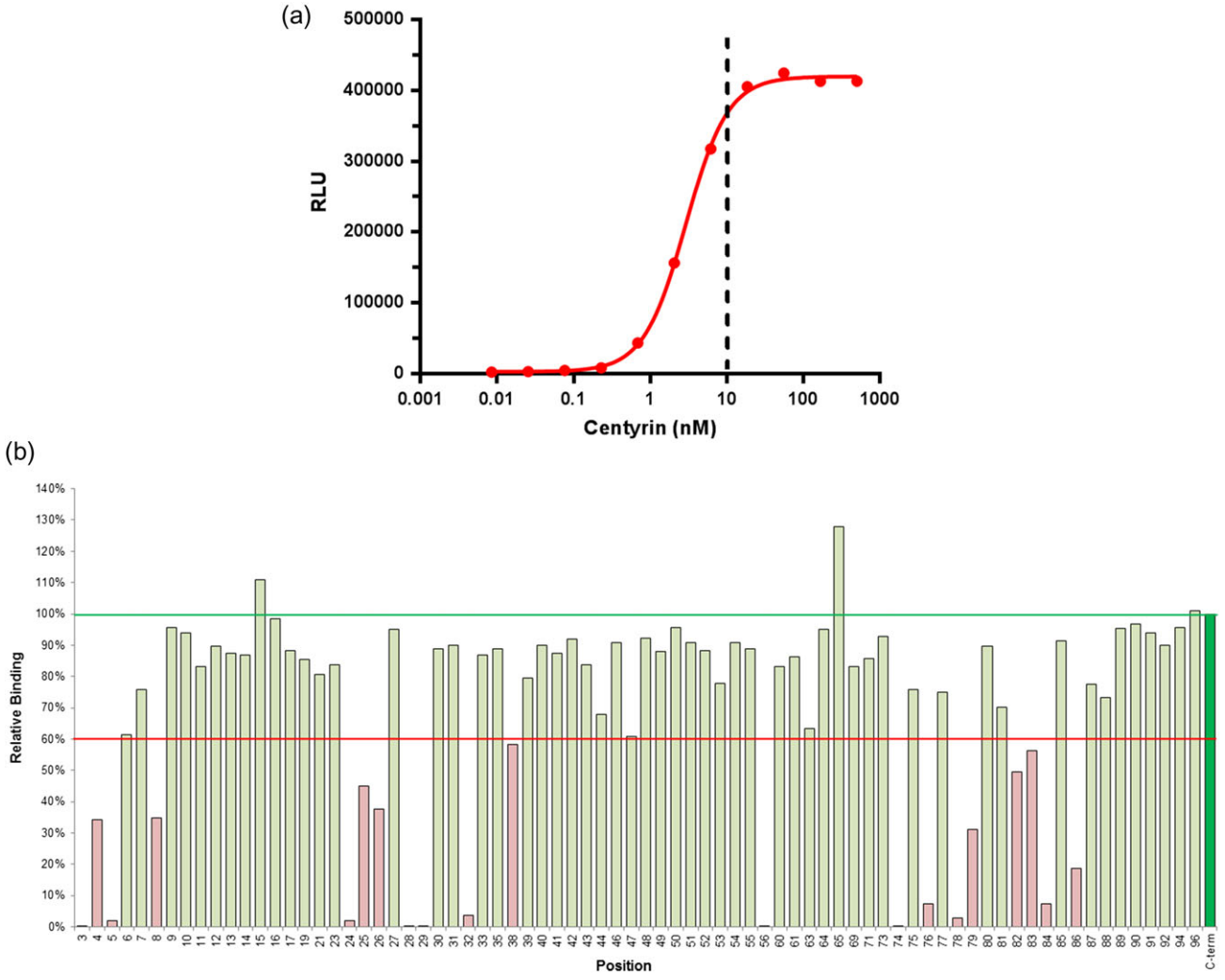

Fig. 4 EGFR binding. (**a**) Binding of NEM-conjugated Centyrin variants to the target receptor was determined by biochemical ELISA with immobilized receptor (extracellular domain). A full dose–response curve is shown for binding of parent 83v2 Centyrin to receptor. Binding at 10 nM is near-maximal, as indicated by the dotted line. (**b**) Variants with >50% of the signal of the parent Centyrin when assayed at 10 nM (red line) were considered to have acceptable target binding to move forward. Variants with <60% are shown in light red.

10 nM (Fig. 4b). Of the 20 proteins that did not, 7 had very weak or no binding to EGFR, defined as <2% of the ELISA signal relative to parent—P3C, P5C, D24C, F28C, Y29C, S56C and G74C, and the others had a range of signals (A4C, L8C, P25C, W26C, F32C, E38C, H76C, V78C, Y79C, T82C, N83C, M84C and G86C). Four of these variants have cysteines replacing residues in the BC binding loop, and seven have cysteines in the FG binding loop. Since so many of the variants met the criteria for cell binding, we decided to increase the stringency and eliminate some additional variants. K6C and V47C were right near the pre-determined cutoff (61%) and were also excluded. Both the variants E44C and S53C were between 60 and 80% of the parent signal, and were also discarded since it was decided to apply more stringent criteria to positions in loops.

Of the variants that were found to retain EGFR binding, two have cysteines in the BC loop and five have cysteines in the FG loop. These seven variants were excluded from further evaluation because they are located in the 'variable' segments of the Centyrin and cysteine introduction at these positions was not anticipated to be broadly applicable to multiple Centyrin binders. The P88C variant immediately adjacent to the FG loop was also discarded due to its proximity to the binding region. In addition, the G95C variant was excluded from further evaluation since the position is located in the segment linking the Centyrin to the His tag. Thus, 43 variants were advanced to biophysical characterization.

Biophysical characterization

Cysteine variants that expressed well, conjugated efficiently, and retained EGFR binding were next evaluated for biophysical properties. Aggregation state of the NEM-conjugation proteins was estimated by size-exclusion chromatography in PBS buffer. Thermal stability and reversibility of unfolding was determined by DSC. The aggregation state of each protein was determined by comparing the area of the 280-nm absorbance peak corresponding to monomeric protein to the peaks corresponding to higher order species. The parental protein, 83v2 with C-terminal cysteine, was 97% monomeric. All variants were at least 75% monomeric except for V13C (66%), which also did not meet the thermal stability cutoff.

Thermal stability was measured by DSC. Melting temperatures and degree of refolding were determined for each NEM-conjugated variant. The T_m of the 83v2 parent and of 83v2cys was 71°C and close to 100% of the protein refolded upon cooling. Eleven of the variants tested had a T_m <65°C (V13C, T14C, R19C, E30C and G52C), unfolded irreversibly (E40C and G43C) or both (V10C, V50C, P51C and P64C) and were excluded from further consideration. There were also four variants that had T_ms increased by ≥2°C relative to the parent: D16C, Q35C, E54C and S71C (Table I).

Cytotoxicity

The 32 variants that met all of the preceding criteria were tested for activity in a cell-based cytotoxicity assay. The variants were conjugated to the cytotoxic tubulin inhibitor MMAF via an enzyme-cleavable Val-Cit linker or a non-cleavable TEG linker (Fig. 1b) using the high-throughput conjugation method. The 83v2-CTC parent and the non-EGFR-binding Centyrin TenconWT with a C-terminal cysteine were also included as controls. Cell killing was measured in the EGFR-overexpressing, Kras mutant human tumor cell line H1573.

Table I reports IC_{50} values obtained from this analysis. The average IC_{50} of two replicates of the 83v2cys-vcMMAF conjugate was 0.7 nM. Four of the thirty-two vcMMAF conjugates appeared to be less potent than the parent in this assay (IC_{50} at least 2× as high):

Table I. Thermal stability and cytotoxicity of variants[a]

Variant	T_m (°C)	Reversible unfolding	vcMMAF conjugate IC_{50}	TEG-MMAF conjugate IC_{50}
N7C	71	Y	*0.2*	2.2
V9C	69	Y	0.4	1.5
V10C	46	N	–	–
S11C	68	Y	0.8	0.7
E12C	71	Y	1.0	0.6
V13C	58	Y	–	–
T14C	63	Y	–	–
E15C	70	Y	0.3	1.0
D16C	*73*	Y	0.3	0.8
S17C	68	Y	0.6	0.8
R19C	62	Y	–	–
S21C	70	Y	0.3	2.8
E30C	63	Y	–	–
S31C	71	Y	0.8	1.5
L33C	71	Y	2.9	1.2
Q35C	*75*	Y	0.9	0.6
S39C	65	Y	0.6	1.1
E40C	67	N	–	–
K41C	70	Y	0.6	0.5
V42C	71	Y	1.0	0.8
G43C	65	N	–	–
I46C	69	Y	1.0	0.8
L48C	67	Y	0.9	0.8
T49C	72	Y	0.5	1.2
V50C	54	Y	–	–
P51C	63	N	–	–
G52C	61	Y	–	–
E54C	*76*	Y	*0.2*	0.7
R55C	65	Y	0.4	1.1
T60C	67	Y	0.4	1.7
G61C	66	Y	0.8	1.4
K63C	65	Y	1.1	0.7
P64C	60	N	–	–
G65C	70	Y	0.5	1.2
T69C	72	Y	4.1	3.5
S71C	*73*	Y	0.9	1.7
Y73C	70	Y	1.5	5.5
V75C	68	Y	2.6	4.1
L89C	70	Y	0.7	0.7
S90C	72	Y	0.9	1.0
A91C	67	Y	0.8	1.0
I92C	70	Y	0.6	0.8
T94C	69	Y	0.3	0.8
83v2-CTC	71	Y	0.7	0.6
83v2	71	Y	NA	NA

[a]Melting temperatures and reversibility were measured with NEM conjugates. Cytotoxicity was determined with MMAF conjugates via either a cleavable Val-Cit linker or a non-cleavable TEG linker. Gray shading indicates a variant that missed the cutoff. Italics type indicates a variant that is improved relative to the parent 83v2-CTC.

L33C, T69C, Y73C and V75C (gray shading). Additionally, two conjugates gave IC_{50}s that were over three times as potent as the parent: N7C and E54C (italics). The average IC_{50} of four replicates of the 83v2cys-TEG-MMAF conjugate was 0.6 nM. The T69C, Y73C and V75C variants had IC_{50}s significantly higher than the parent with these conjugates as well (5–9×), while the L33C variant was ~2× that of the parent. N7C and S21C were 3- to 4-fold higher than the parent. Unlike the vcMMAF conjugates, none of the

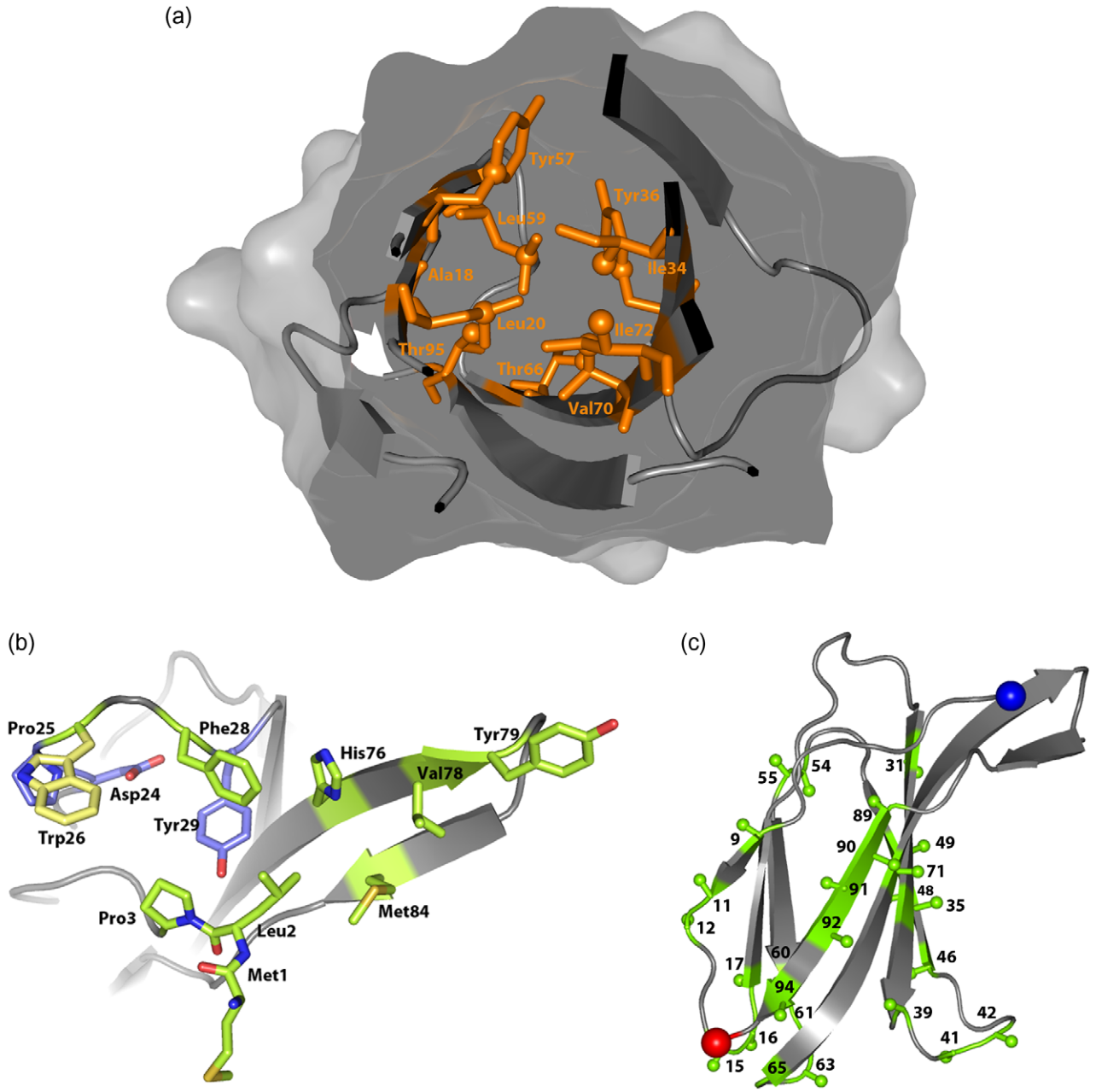

Fig. 5 Location of conjugation sites on the structure of the W26A variant of 83v2. (**a**) Poor conjugation sites. The 83v2 protein is shown as a cartoon rendering with the outer surface representation. Positions at which cysteine variants conjugate poorly to NEM are shown as orange sticks, with a sphere indicating the atom position γ (approximately where the reactive sulfur would be in the cysteine variant). (**b**) Proposed EGFR binding site. Structure is shown from a 'top' view. The positions in the BC and FG loops and at the N-terminus proposed to comprise the EGFR binding site are shown as sticks in light green. The Trp26 side chain that is not present in the X-ray structure was inserted using PyMol without moving any other atoms, and is colored yellow. Secondary BC loop positions that are likely to affect EGFR binding only through their contact with the binding positions are shown in blue. (**c**) Optimal cysteine locations. The 26 positions determined to be potential conjugation sites are shown in green. If applicable, the C_α-C_β bond of each residue is shown as a stick and the β-carbon is shown as a sphere to indicate the direction of the side chain. N- and C-termini are shown as blue and red spheres, respectively.

TEG-MMAF variants had IC_{50}s that were much lower than the parent 83v2cys-PEG_4MMAF conjugate, which demonstrates the impact the linker type can have on cytotoxicity.

Discussion

Cysteine engineering of therapeutic proteins such as antibodies is becoming increasingly important with the advent of targeted delivery technologies; however, prediction of optimal sites for cysteine incorporation remains a challenge. Computational tools have been developed to complement empirical approaches (Voynov *et al.*, 2010), although optimization of conjugation sites for large proteins can often be labor intensive. Centyrins have the advantage of small size, making a comprehensive cysteine scanning experiment feasible and relatively fast. A complete cysteine scan was conducted of the EGFR-binding Centyrin 83v2, including those positions in the

variable binding loops and those predicted to be buried in the protein core. For the most part, the cysteine variants behaved as expected. The cysteines in buried positions generally failed to conjugate efficiently to even the small maleimide NEM (Fig. 5a). Many of the variants with cysteines in the binding loops had reduced binding to the EGFR target protein, though some retained EGFR binding. In the BC loop, cysteine variants at the Asp23 position at the edge of the loop as well as at Ala27 in the middle of the loop retained EGFR binding while all other variants were significantly impaired, suggesting that these residues are either directly involved in contact with EGFR or that they affect Centyrin folding. In the FG loop, variants with cysteines at Asn77, Lys80, Asp81, Arg85 and Leu87 retained most of their EGFR binding, while T82C and N83C retained some binding as well. Taken together, this suggests a map of the EGFR binding site that includes the BC loop, one face of the beta sheet within the FG loop, and the N-terminus of the protein (Fig. 5b). The close proximity of the N-terminus to the variable loops may explain why many of the A strand variants show poor EGFR binding—they may comprise a portion of the site that contacts EGFR, or they play a secondary role affecting the positioning of the BC and/or FG loops. Besides the variable loops and the A strand, a significant majority of positions in the scaffold are tolerant to mutation.

The cysteine scanning experiment also provided some unexpected insights into the Centyrin scaffold. Most of the mutations in the CD loop resulted in proteins that expressed poorly (Gln37, Glu38 and Ala45) or did not exhibit reversible unfolding (Glu40, Gly43). We hypothesize that the CD loop plays an important role in protein folding or stability that has not been appreciated previously. Similarly, many of the variants with mutations in the DE loop had reduced thermal stability (Val50, Pro51 and Gly52), while mutation of a different DE loop position (Glu54) results in increased thermal stability. Thus, the DE loop likely plays an important role in folding or stability of the Centyrin domain.

Of the 32 variants tested in the cytotoxicity assay, only 6 had even a small reduction in potency (up to 10-fold) when conjugated to the MMAF payload via either the non-cleavable ethylene glycol linker or the enzyme-cleavable Val-Cit linker. Three of these were less potent with either linker (T69C, Y73C and V75C); Tyr73 and Val75 are near the FG loop involved in binding that may explain the effect that mutation has on activity. The other three variants, especially N7C and S21C, had differences that appeared to be linker-based. The release mechanism of drug is expected to vary considerably between the two linkers as the Val-Cit linker is cleaved upon internalization by the enzyme cathepsin B in the endosome and does not require degradation of the Centyrin, while the PEG-linked drug is released only by degradation of the protein to its component amino acids. We propose that the N7C and S21C proteins are more resistant to protease cleavage than the parent protein, which reduces activity in conjugates prepared with the PEG linker. This example highlights the value of using a non-biased, comprehensive approach to identify optimal conjugation sites.

The 26 positions that were determined to be optimal conjugation sites are well distributed across the protein (Fig. 5c). Both loop and β-sheet positions are well represented in the final group. It is interesting to note that the β-sheet on the concave face of the molecule (made by C, D, F and G β-strands) has a number of potential conjugation sites while the β-sheet on the convex face (made by A, B and E β-strands) has fewer and they are mainly at the ends of the strands. The β-sheet on the convex face has also been randomized in a different Centyrin library (Diem *et al.*, 2014), indicating that many of these positions are tolerant to substitution, and our data are consistent with this conclusion. Most of the optimal positions have side chains that are solvent exposed in the crystal structure; however, there are a handful of positions on the edges of the sheets with side chains that point toward the protein core that were also successful conjugation sites (Ile46, Leu48, Leu89 and Ala91). We also found that seven of the nine serine residues in 83v2 were identified as potential conjugation sites, an unsurprising result since serine-to-cysteine is a conservative mutation and serines are often located on the protein surface. Meanwhile, none of the variants at aromatic positions (Phe/Tyr/Trp) or at prolines were determined to be good conjugation sites.

Conclusion

In this study, we have developed a comprehensive picture of the relationship between structure and function of the Centyrin 83v2 and its ability to be chemically conjugated and achieve efficient targeted delivery. We anticipate that the locations identified in this study will be largely translatable to other Centyrins that bind to different targets since positions near to the binding site were excluded via our filters. Identification of these tolerant positions will be of great use in advancing a Centyrin-based targeted delivery platform, including those with multiple small molecules per Centyrin. Additional studies to evaluate *in vivo* profiles of these variants, such as pharmacokinetics and assessment of conjugate stability, may help to further elucidate optimal sites for conjugation in the Centyrin scaffold. Further experiments to assess the effect of incorporating multiple cysteines in one Centyrin domain will also be valuable. In addition, the scanning mutagenesis experiments have provided further insight into the Centyrin 83v2 and the parent scaffold by identifying positions that have a significant effect on binding and stability. Finally, the approach described provides a general model that could be useful in identifying conjugation sites in targeting proteins of various types beyond Centyrins.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

The authors thank Ed Swift for producing the 83v2 W26A protein for crystallization and Joe Bourghol for the ProteOn analysis.

References

Adams,P.D., Afonine,P.V., Bunkoczi,G. *et al.* (2010) *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

Agarwal,P. and Bertozzi,C.R. (2015) *Bioconjug. Chem.*, **26**, 176–192.

Anderson,M., Attar,R., Diem,M.D. *et al.* (2014) US20140155325

Beck,A. and Reichert,J.M. (2014) *mAbs*, **6**, 15–17.

Binz,H.K. and Pluckthun,A. (2005) *Curr. Opin. Biotechnol.*, **16**, 459–469.

Brinkley,M. (1992) *Bioconjug. Chem.*, **3**, 2–13.

Diem,M.D., Hyun,L., Yi,F., Hippensteel,R., Kuhar,E., Lowenstein,C., Swift, E.J., O'Neil,K.T. and Jacobs,S.A. (2014) *Protein Eng. Des.Sel.*, **27**, 419–429.

Emsley,P. and Cowtan,K. (2004) *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

Jacobs,S.A., Diem,M.D., Luo,J., Teplyakov,A., Obmolova,G., Malia,T., Gilliland,G.L. and O'Neil,K.T. (2012) *Protein Eng. Des. Sel.*, **25**, 107–117.

Krissinel,E.B., Winn,M.D., Ballard,C.C., Ashton,A.W., Patel,P., Potterton,E.A., McNicholas,S.J., Cowtan,K.D. and Emsley,P. (2004) *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2250–2255.

Luo,J., Teplyakov,A., Obmolova,G., Malia,T.J., Chan,W., Jacobs,S.A., O'Neil,K.T. and Gilliland,G.L. (2014) *Proteins*, **82**, 1527–1533.

Mack,F., Ritchie,M. and Sapra,P. (2014) *Semin. Oncol.*, **41**, 637–652.

Mamluk,R., Carvajal,I.M., Morse,B.A. *et al.* (2010) *mAbs*, **2**, 199–208.

McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) *J. Appl. Crystallogr.*, **40**, 658–674.

Otwinowski,Z. and Minor,W. (1997) *Methods Enzymol.*, **276**, 307–326.

Rabuka,D. (2010) *Curr. Opin. Chem. Biol.*, **14**, 790–796.

Shen,B.Q., Xu,K., Liu,L. *et al.* (2012) *Nat. Biotechnol.*, **30**, 184–189.

Skerra,A. (2000) *J. Mol. Recognit.*, **13**, 167–187.

Steiner,M. and Neri,D. (2011) *Clin. Cancer Res.*, **17**, 6406–6416.

Strop,P., Liu,S.H., Dorywalska,M. *et al.* (2013) *Chem. Biol.*, **20**, 161–167.

Teplyakov,A., Obmolova,G., Malia,T.J. *et al.* (2014) *Proteins*, **82**, 1359–1369.

Voynov,V., Chennamsetty,N., Kayser,V., Wallny,H.J., Helk,B. and Trout,B.L. (2010) *Bioconjug. Chem.*, **21**, 385–392.