Published in final edited form as:

Biotechnol Bioeng. 2019 March ; 116(3): 526–535. doi:10.1002/bit.26893.

Engineering an EGFR-Binding Gp2 Domain for Increased Hydrophilicity

Feifan Du, Max A. Kruziki, Elizabeth J. Zudock, Yi Zhang, Patrick S. Lown, and Benjamin J. Hackel

University of Minnesota – Twin Cities, Department of Chemical Engineering and Materials Science, 421 Washington Avenue SE, Minneapolis, MN 55455

Abstract

The Gp2 domain is a 45 amino-acid scaffold that has been evolved for specific, high-affinity binding towards multiple targets, and was proven useful in molecular imaging and biological antagonism. It was hypothesized that Gp2 may benefit from increased hydrophilicity for improved physiological distribution as well as for physicochemical robustness. We identified seven exposed hydrophobic sites for hydrophilic mutations and experimentally evaluated single mutants, which yielded six mutations that do not substantially hinder expression, binding affinity or specificity (to epidermal growth factor receptor), and thermal stability. Eight combinations of these mutations improved hydrophilicity relative to the parental Gp2 clone as assessed by reverse-phase HPLC ($p < 0.05$). Secondary structures and refolding ability of the selected single mutants and all multi-mutants were unchanged relative to the parental ligand. A variant with five hydrophobic-to-hydrophilic mutations was identified with enhanced solubility as well as reasonable binding affinity ($K_d = 53 - 63$ nM), recombinant yield (1.3 ± 0.8 mg/L), and thermal stability ($T_m = 53 \pm 3$ °C). An alternative variant with a cluster of three leucine-to-hydrophilic mutations was identified with increased solubility, nominally increased binding affinity ($K_d = 13 - 28$ nM) and reasonable thermal stability ($T_m = 54.0 \pm 0.6$ °C) but reduced yield (0.4 ± 0.3 mg/L). In addition, a 7 °C increase in the midpoint of thermal denaturation was observed in one of the single mutants (T21N). These mutants highlight the physicochemical tradeoffs associated with hydrophobic-to-hydrophilic mutation within a small protein, improve the solubility and hydrophilicity of an existent molecular imaging probe, and provide a more hydrophilic starting point for discovery of new Gp2 ligands towards additional targets.

Keywords

protein engineering; hydrophilicity; solubility; Gp2; synthetic ligand

Introduction

Molecular binding ligands are important tools in molecular therapy (Feldwisch and Tolmachev, 2012; Scott et al., 2012), diagnostics (Knowles and Wu, 2012; Stern et al., 2013), and biotechnology (Kiessling et al., 2006; Spencer et al., 1993). Antibodies are the

dominant molecular format and effective (Carter and Lazar, 2018), but disadvantages have spurred efforts to find alternatives (Banta et al., 2013; Škrlec et al., 2015) in particular applications. For example, large size not only limits tissue penetration (Thurber et al., 2008a; Thurber et al., 2008b) and slows clearance (Schmidt and Wittrup, 2009) of unbound agent thereby elevating background activity or signal, but also provides substantial additional surface for unwanted non-specific interaction (Chen et al., 2013). Furthermore, multidomain architecture complicates bispecific formatting (Holliger and Hudson, 2005). A multitude of alternative protein scaffolds have emerged (Banta et al., 2013; Škrlec et al., 2015) to address these opportunities. One such instance is Gp2, a 45 amino acid protein scaffold with two solvent-exposed loops that have been successfully engineered to provide high affinity binding while retaining thermal stability (Kruziki et al., 2015). Evolved Gp2 domains have proven successful as a molecular imaging agent (Case et al., 2018; Kruziki et al., 2016) and an insulin receptor antagonist (Chan et al., 2017).

Beyond size and domain architecture, other physical and chemical characteristics, such as hydrophilicity, solubility and stability, are significant for protein utility and developability (Jacobs et al., 2010; Jain et al., 2017). In particular, surface hydrophilicity of a protein can impact its solubility (Hayakawa and Nakai, 1985; Schein, 1990), which is a critical property for production (Marston et al., 1986; Sørensen and Mortensen, 2005), storage (Wang, 1999), and use. In addition, hydrophilic mutations can benefit stability (Pedone et al., 2001), although this is not always the case (Strub et al., 2004). Moreover, hydrophilic engineering has been used to tailor physiological distribution of a protein for molecular imaging (Hackel et al., 2012). Engineering a protein for increased surface hydrophilicity can improve functional utility, but because the majority of random mutations are detrimental (Tokuriki et al., 2007), care must be taken to identify effective mutants.

Rationalized site-directed mutagenesis has been a popular strategy to modify hydrophilicity and solubility. For example, hydrophilic substitutions of surface-exposed hydrophobic residues have been effective in increasing protein solubility in multiple cases (Das and Georgiadis, 2001; Jenkins et al., 1995; Lawson et al., 2009; Malissard and Berger, 2001; Mosavi and Peng, 2003; Roosild and Choe, 2005), with even single mutations providing benefit (Das and Georgiadis, 2001; Jenkins et al., 1995). More specifically, upon mutation of six exposed leucines to arginines, a protein that was only soluble from pH 3 to 4.25 at 100 μ M became soluble from pH 4 to 7 at higher concentrations (Mosavi and Peng, 2003). Progressive substitutions of lipid-facing hydrophobic residues with hydrophilics chosen based on biophysical tests made KchAfu104, a potassium-selective ion channel, more soluble and less likely to aggregate (Roosild and Choe, 2005). Hydrophilic mutation of residues according to solubility and charge differences between a human glycosyltransferase and bovine/murine versions resulted in 70% increase in expression, 10-fold increase in solubility, and nominally unchanged catalytic activity (Malissard and Berger, 2001). Select hydrophilic mutations to 11 β -hydroxysteroid dehydrogenase type 1 increase soluble recombinant yield and mildly impact enzymatic activity, with up to 5 ± 2 -fold increase or 12 ± 4 -fold decrease in k_{cat}/K_m depending on mutant (Lawson et al., 2009). Increased solubility was observed for three out of 29 mutants of an integrase protein with one or more solvent-exposed hydrophobic sites mutated to lysine and/or alanine (Jenkins et al., 1995). As evidenced by this modest success rate, not all hydrophobic-to-hydrophilic mutations

improve solubility (Trevino et al., 2008). Deletion mutation was beneficial for the enzyme *HhaI*: replacing the C-terminal tetrapeptide that was suspected to contribute to aggregation with a single glycine residue led to a significantly higher solubility (Daujotyte et al., 2003). Introduction of charged residues has also been effective to increase solubility (Calloni et al., 2005; Thompson et al., 2012). Alternative to design, randomized mutagenesis followed by specific selections can achieve improved solubility. More specifically, one can mutate the hydrophobic sites on the protein surface first with error-prone PCR, then select the hydrophilic mutants by *in vivo* screening (Japrun et al., 2005; Maxwell et al., 1999) or high-throughput screening (Asano et al., 2011). Besides mutagenesis, other approaches to control solubility were also tested, such as attachment of short peptide tags (Islam et al., 2015; Kato et al., 2007) and shielding aggregation-promoting patches on the protein surface with carbohydrate moiety (Wu et al., 2010).

To enhance physicochemical robustness and utility of Gp2, we aimed to increase the hydrophilicity of its surface. A variant which was previously evolved to bind the epidermal growth factor receptor (EGFR), Gp2_{E2.2.3} (Kruziki et al., 2015), was used as the model system because it is a valuable lead molecule for molecular imaging of cancer. We used a rational approach to identify solvent-exposed sites that may benefit from hydrophilic mutation and selected candidate amino acids for mutation. We evaluated single mutants for retention of expression and binding affinity. Combinations of tolerable mutations were evaluated for recombinant yield, hydrophilicity, aggregation propensity, secondary structure, thermal stability and binding affinity. Single and combined mutations provided benefits in multiple functional properties, albeit with tradeoffs in other properties, thereby yielding a set of hydrophilic, soluble Gp2 options.

Results

Mutant Candidate Identification

The wild-type Gp2 has a hydrophobic interior and a predominantly hydrophilic exterior (Figure 1A). However, there are four hydrophobic residues with reasonable solvent exposure (F13, L22, L26, and L28) and four mildly hydrophilic sites with extensive solvent exposure (Y18, T21, W31, and T41; Figure 1B). We hypothesized that mutation of these sites could improve the physicochemical robustness of the Gp2 domain. These sites, with the exception of W31 which is already mutated to arginine within the EGFR-binding Gp2_{E2.2.3}, were mutated to more hydrophilic amino acids. Sites in the two solvent-exposed loops were excluded because of their presumed impact on target binding. Prolines at sites 16 and 45 were conserved as they are insufficiently solvent exposed (35% and 41% accessibility, respectively) to expect substantial benefit given their already medium hydrophilicity; moreover, their proline side chains provide unique conformational strength. Other sites were insufficiently exposed (<30% accessibility) or already strongly hydrophilic.

Mutational options were considered based on hydrophobicity, observation in natural Gp2 homologs, and theoretical destabilization computed in FoldX (Schymkowitz et al., 2005) (see *Materials and Methods* for details). We sought to increase hydrophilicity while not impacting the charge density and distribution; thus, mutations to charged D, E, N, and Q or partially cationic H were not considered. Hydrophobic leucine was mutated to mid-

hydrophilic serine or alanine. L26 was also mutated to asparagine or threonine because of the frequent observation of these mutants in natural Gp2 homologs (glutamine was not observed in homologs). Mid-hydrophilic threonine and tyrosine were mutated to highly hydrophilic asparagine because FoldX calculations predicted each to be well tolerated. Hydrophobic F13 was mutated to the more hydrophilic aromatic tyrosine; F13N was also tested. While additional amino acid options would have yielded more information regarding mutational tolerance and potentially identified beneficial mutations, this relatively focused group (Table 1) was chosen to enable experimental rigor in an efficient manner.

Single Mutant Characterization

A Gp2 domain engineered to bind EGFR was chosen as the parental clone, with which all potential mutants were implemented individually. Mutant genes were generated by site-directed mutagenesis and verified by sequencing. Mutants were expressed in *E. coli* with a C-terminal His₆ tag and purified via metal affinity chromatography. All mutants yielded recoverable quantities of Gp2 in soluble fractions of *E. coli*.

Target binding was evaluated for the mutants by titration on A431 human epidermoid carcinoma cells expressing EGFR, with binding detected by flow cytometry (Figure 2). With only one mutant tested at each site, Y18N, T21N and T41N exhibited comparable binding to the parental clone. At site L22, the serine mutant exhibited nominally stronger binding than parental at all concentrations, whereas the alanine mutant had reduced binding. At site 26, L26T and L26A were comparably strong binders, contrary to weak binding by L26N and L26S. Because threonine is more hydrophilic than alanine, L26T was selected for further study. Similarly, while L28S and L28A both exhibited reasonable binding to EGFR, L28S was chosen for further evaluation because serine is more hydrophilic than alanine. At site F13, mutation to asparagine was detrimental to the target binding. Although mutation to tyrosine at the same site was not as detrimental, mutant F13Y was still a weak binder compared to most other mutants tested. In addition, tyrosine is only mildly hydrophilic, thus site 13 was conserved as phenylalanine in further studies. Target specificity was assessed by incubation on EGFR-negative HEK293 cells at the highest titration concentration (243 nM), with binding detected by flow cytometry (Figure S1). None of the mutants showed substantial off-target binding.

Thermal stabilities of the six selected single mutants were tested by thermal denaturation monitored by circular dichroism spectroscopy. All mutants exhibited thermal stabilities of at least 65 °C (Figure 3). L22S and T41N hindered thermal stability by about 10 °C relative to the parental whereas T21N dramatically elevated thermal stability. Intriguingly, T21N was the one mutation predicted by FoldX to have a dramatically improved stability (Table 1).

Multi-mutant Characterization

After identification of point mutations that improve protein hydrophilicity without detriment to binding affinity and thermal stability, and still enable recombinant expression, combinations of these mutations were assessed (Table 2). In particular, we tested a mutant with all six mutations, as well as all six variations of five mutations. In addition, we

constructed a mutant (mutant I) with the three mutations that change hydrophilicity the most, the leucine mutations; we also added Y18N to this triple mutant (mutant H).

Each clone, containing a C-terminal His₆ tag, was expressed at 1 L scale in BL21 (DE3) *E. coli* under the same culture conditions and purified by metal affinity chromatography. Concentration and purity were measured by absorbance at 280 nm (Pace et al., 1995), SDS-PAGE, and MALDI mass spectrometry. Yields were acceptably modest (0.3 ± 0.2 to 1.3 ± 0.8 mg/L) though nominally reduced for all of the multi-mutants compared to the yield of the parental clone (Figure 4), with one exception: clone A had negligible yield and, thus, was not considered for further analysis.

Hydrophilicities of all expressed mutants were evaluated by reverse-phase high-performance liquid chromatography (HPLC) on a C18 column using a 10–90% acetonitrile gradient over 15 minutes. Retention time of all the mutants (11.9 ± 0.1 to 12.4 ± 0.3 min) was similarly reduced ($p < 0.05$ for all multi-mutants) relative to parental (13.4 ± 0.1 min; Figure 5). Thus, any of the combinations of beneficial point mutations led to improved hydrophilicity as expected. It should be noted that hydrophobic chromatography retention time can be affected by protein aggregation or oligomerization. The parental clone has been previously validated as monomeric via size-exclusion chromatography (Kruziki et al., 2016), and mutants in this study were shown to exhibit minimal aggregation over a range of conditions (see next paragraph).

The relative solubility of Gp2 clones was assessed via a modified version of affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS). Gold particles (20 nm in diameter) were conjugated with anti-His₆ antibodies and His₆-tagged Gp2 variants. Gp2 aggregation resulted in a shift in the wavelength of maximum absorbance of the gold particles. This general approach has been validated as a high throughput and robust tool to evaluate the aggregation propensity of antibodies (Estep et al., 2015; Liu et al., 2014; Sule et al., 2011; Sule et al., 2013). The parental Gp2 exhibited substantial aggregation at mildly acidic pH under reduced salt (60 mM) and slight aggregation under higher salt (300 mM) conditions (Figure 6A). Conversely, clone I exhibited substantially reduced aggregation at these conditions. Analysis of all mutants revealed improved solubility relative to the parental variant (Figure 6B).

Purified protein samples were re-suspended in pH 5 sodium acetate solution for circular dichroism and thermal denaturation analysis. Secondary structures of all mutants were similar to that of the parental ligand, and all mutants can refold — from partially to fully — after thermal denaturation (Figure S2). Mutation combinations substantially hindered thermal stability with the T_m of penta-mutants ranging from 46 ± 4 °C to 53 ± 3 °C (Figure 7). Notably, the nominally least stable of the penta-mutants was clone C that reverted the beneficial T21N mutation back to wild-type. Clone I, with only the three leucine mutations, was similarly destabilized ($T_m = 54 \pm 1$ °C) thereby implicating the impact of these dramatic hydrophobic-to-hydrophilic mutations. This is reasonable given their partial interaction with the hydrophobic core (Figure 1).

Binding affinities of all mutants were evaluated by titration on EGFR⁺ A431 cells with binding detected by flow cytometry (Figure 8A). Three out of eight mutants (B, F, and H) exhibited substantially reduced binding strength compared to that of the parental clone while four mutants (C, D, E, and G) were mildly reduced (Figure 8B). Clone I, with only the three leucine mutations, exhibited nominally stronger binding ($K_d = 19$ nM with 95% confidence interval of 13 – 28 nM). Target specificity was assessed by incubation of a high concentration (243 nM) of mutants on EGFR-negative HEK293 cells, with binding detected by flow cytometry. All variants exhibited ~1% binding relative to parental binding on EGFR-positive A431 cells (Figure S1).

Discussion

The Gp2 domain offers a compelling option for synthetic ligand engineering in applications benefiting from small size and modularity. Motivated by potential benefits to developability and physiological distribution, we engineered the Gp2 scaffold for increased hydrophilicity. We mutated seven solvent-exposed hydrophobic sites of an EGFR-binding Gp2 to hydrophilic residues and assessed the functionalities of those mutations as well as combinations thereof.

While mutation of hydrophobic residues to hydrophilic will trivially increase the hydrophilicity of the unfolded polypeptide, the impact on folded protein is less obvious because mutation may modify folding to expose other hydrophobic residues to the solvent. Moreover, the impact of mutation on other properties including aggregation, affinity, stability, and recombinant yield are less obvious *a priori*. Combinations of such hydrophobic-to-hydrophilic mutations greatly improved the overall hydrophilicity of the molecule, as assessed by retention time of reverse-phase HPLC. Although secondary structure and ability to refold after thermal denaturation were relatively unchanged after the mutations, other characteristics of the molecule, such as aggregation propensity, binding affinity, recombinant yield, and thermal stability, were altered. Solubility of all multi-mutants were increased as assessed via a modified version of AC-SINS. Regarding affinity, the mutants ranged from nominally increased (mutant I) to decreased to 299 nM (95% confidence interval of 261 – 344 nM) (mutant H). The three leucine mutations (L22S, L26T, L28S; mutant I; Figure S3), collectively, are tolerated in terms of affinity. The addition of the Y18N mutation hinders affinity substantially, which is an interesting result given its distance from the evolved loops. Within the context of this hindered Y18N mutant, T21N and T41N each aid affinity. Recombinant yields in the soluble fraction of *E. coli* were all decreased for the multi-mutants compared to the starting clone, but appreciable amount of protein was still obtained for all multi-mutants except mutant A, which contained all six mutations. While the single mutants exhibited a range of stabilization to destabilization, the multi-mutants were uniformly destabilized. Future experiments will be needed to assess the enthalpic and entropic contributions to these impacts on thermal stability.

With acknowledgement of such tradeoffs amongst physicochemical properties, there are multiple options to explore depending on the specific application. For positron emission tomography imaging of EGFR, clone I is the most compelling mutant because it has substantially increased solubility and hydrophilicity (1.2 min. improvement over the parental

clone in retention time, $p < 0.005$), the highest thermal stability ($T_m = 54.0 \pm 0.6$ °C) and the highest EGFR-binding affinity ($K_d = 19$ nM with 95% confidence interval of 13 – 28 nM) among multi-mutants and parental clone. Yet clone I exhibits low soluble recombinant yield (0.4 ± 0.3 mg/L). Conversely, clone E (Figure S2) is intriguing for future application because it has the highest yield (1.3 ± 0.8 mg/L), the second-highest binding affinity ($K_d = 58$ nM with 95% confidence interval of 53 – 63 nM), improved solubility, and comparable stability among the multi-mutants while also being more hydrophilic than the starting clone (1.2 min. improvement in retention time, $p < 0.05$). Therefore, clone E would be the best choice if yield is paramount and/or affinity requirement is flexible (*e.g.* in multivalent applications). In an alternative use, if a hydrophilic mutant was sought as the basis of a new combinatorial library for Gp2 discovery and evolution, the single mutation T21N would be recommended because of its heightened stability ($T_m = 83$ °C). To summarize, the optimal hydrophilic Gp2 mutant will depend on which properties are deemed critical and, thus, should be identified based on context; clone I is optimal if binding affinity is critical; clone E provides an option with better recombinant yield; T21N is the best conservative option to be considered in future *de novo* ligand discovery libraries.

The approach utilized in this study, hydrophobic-to-hydrophilic mutagenesis, was proven effective in increasing hydrophilicity and/or solubility in the past (Das and Georgiadis, 2001; Jenkins et al., 1995; Lawson et al., 2009; Malissard and Berger, 2001; Mosavi and Peng, 2003; Roosild and Choe, 2005), albeit not in all the cases (Trevino et al., 2008). In this study, solubility and hydrophilicity were increased for every multi-mutant. In hydrophilic mutation of integrase, increased solubility was only observed in a small fraction of mutants (Jenkins et al., 1995). The expression level of all the multi-mutants were reduced, contrary to the 70% increase in the expression level of a human glycotransferase (Malissard and Berger, 2001).

Additionally, a more thorough sampling of the mutational space might reveal improved variants, perhaps including one that is globally optimal across many Gp2 ligands (*e.g.* against different targets), and also broadly optimized for multiple physicochemical properties. Such sampling could potentially be implemented using yeast display. Higher throughput selection techniques could also examine more mutations and their combinations, but such assays need development.

Materials and Methods

Computational analysis

PyMol was used to visualize the wild-type Gp2 solution structure. Solvent accessible surface area (SASA) was computed as the median of the four Gp2 structures (PDB code: 2WNM, 2LMC, 4LK0, 4LLG) using a 1.4 Å probe through the Accessible Surface Area tool from Center for Information Biology at Ochanomizu University (*i.e.* ASA server). Prediction of mutational destabilization was made using FoldX (Schymkowitz et al., 2005) in terms of change in folding energy. Natural frequency was concluded as observed in 34 sequences of Gp2 homologs, with an average of 68% identity, which were identified through a BLAST search of the wild-type Gp2 sequence.

Mutation cloning

The gene for EGFR-binding Gp2 (Gp2_{E2.2.3} (Kruziki et al., 2015)) and each of the mutated versions was shuttled into pET-22b vector plasmid with a C-terminal His₆ tag (Novagen, EMD Millipore, Billerica, MA). Genes for both single-mutants and multi-mutants (see *Supplementary Data* Table S1 for sequences of parental and multi-mutants; Figure S4 for DNA sequence of the parental clone) were prepared by sequential annealing and extension of overlapping oligonucleotides (purchased from IDT). Plasmids of both inserts (genes for Gp2 domain and its variations) were cleaned up and verified through agarose gel electrophoresis before ligation. All plasmids were confirmed by DNA sequencing before used in bacterial expression.

Protein production and purification

BL21 (DE3) *Escherichia coli* were transformed with pET-Gp2-His₆ plasmid and grown overnight in lysogeny broth (LB) medium with 50 mg/L kanamycin. Five milliliters of overnight culture were added to 1 L of LB medium with 50 mg/L kanamycin, grown at 37 °C to an optical density of 0.6 – 1 units, and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 16 ± 4 hours at 30 °C. Cells were pelleted, resuspended in 10 – 20 mL of lysis buffer (50 mM sodium phosphate (pH 8.0), 0.5 M NaCl, 5% glycerol, 5 mM CHAPS, and 25 mM imidazole), and underwent four freeze-thaw cycles. The insoluble fraction was removed by centrifugation at 12,000 g for 10 min. and then by filtration using 0.2 μm syringe-filters (Whatman, GE Healthcare). The remaining soluble fraction was then purified by metal affinity chromatography using HisPur cobalt resin (Thermo Scientific). Resulting Gp2 in elution buffer (PBS with 150 mM imidazole) was further purified and analyzed by reverse-phase high-performance liquid chromatography (RP-HPLC) on a C18 column (Waters XBridge Peptide BEH C18, 300 Å, 5 μm) with a 10–90% gradient of acetonitrile in water with 0.1% trifluoroacetic acid over 15 minutes.

Protein verification and quantification

Protein yield was first quantified by light absorbance at 280 nm (Pace et al., 1995) of Gp2 in elution buffer. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was also used (with NuPAGE 4–12% Bis-Tris Gel) to quantify amount of protein, especially those in the earlier elutions since the earlier elutions tend to be less pure. A voltage of 200 V was supplied for 30 min. for SDS-PAGE, and 1 g/L bovine serum albumin in PBS (1 g/L PBSA) was used as a reference to analyze the resulting gel. Protein mass was verified by matrix-assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS) with protein samples in acetonitrile solution (after RP-HPLC). In cases of more than one type of protein existing in one sample, area under the curve from RP-HPLC and/or MALDI-TOF-MS results was used to finalize the yield of one single clone.

The self-aggregation propensity of Gp2 was evaluated using AC-SINS via a modified protocol of the procedures published previously by Tessier *et al.* (Estep et al., 2015; Liu et al., 2014; Sule et al., 2011; Sule et al., 2013). The mixture of His₆-tagged antibody (320 μg/mL; Proteintech) and rabbit IgG (80 μg/mL; Rockland) in 78 mM acetate buffer (pH 4.3) was incubated with 20 nm diameter gold particles (4.6×10¹¹ particles/mL; Ted Pella) at a volume ratio of 1:9 at room temperature for two hours. Poly (ethylene glycol) methyl ether

thiol MW 2000 (final concentration: 0.1 μM ; Nanocs) was added one hour after incubation started. The coated nanoparticles were filtered (0.2 μm PVDF) and eluted by 1/10th volume of PBS (BioWhittaker®). Five μL of coated nanoparticle solution in PBS was incubated with 45 μL of His₆-tagged Gp2 in desired buffers (20 mM acetate buffer for pH 3.3 – 5.6; 20 mM phosphate buffer for pH 5.9 – 7.9; 5 – 50 $\mu\text{g/mL}$ Gp2) for 2.5 hr in a 384 well plate (Corning) at room temperature. The negative controls were prepared by the same procedures except no His₆-tagged Gp2 was added. The absorbance spectra from 510 nm to 580 nm were measured using a microplate reader (SYNERGY). The maximum absorbance wavelength of each sample was analyzed by MATLAB, and the maximum absorbance wavelength shift was calculated by setting 529 nm as the reference wavelength based on multiple control measurements.

Affinity measurement

A431 human epidermoid carcinoma cells (provided by Professor Daniel Vallera, Therapeutic radiology department of UMN) were cultured in Dulbecco's Modified Eagle's Medium (supplemented with 10% fetal bovine serum and 1% Pen Strep) and incubated at 37 °C in humidified air with 5% CO₂. Cells to be used in flow cytometry were washed with PBS, detached using Trypsin/EDTA (Corning) for 8–10 minutes, placed in a cold block from 4 °C during experimentation to prevent EGFR internalization and stored in 4 °C fridge for at most 3 days when not in use. Cells were counted (by Countess II FL) before labeling to ensure an excess of EGFR to Gp2 molecules (at least 15-fold). Cells were labeled with varying concentrations (2, 9, 27, 81, 243 nM for single mutants and 5, 11, 26, 58, 130, 294, 664, 1500 nM for multi-mutants) of purified Gp2 for at least 40 min at 4 °C, pelleted and washed with 1 g/L PBSA, then labeled with fluorescein-conjugated rabbit anti-His₆ antibody (Abcam ab1206) for 10 min at 4 °C followed by another wash with PBSA. The median fluorescence signal was quantified by flow cytometers BD FACSCanto A V07300428, BD LSR II H1160 or BD LSR II H4710. The outliers of fluorescence signals at each concentration of each clone were identified by the Grubbs test and discarded (with 95% confidence), then the rest of the signals were averaged. The equilibrium dissociation constant, K_d , was calculated by minimizing the sum of squared residuals assuming a 1:1 binding interaction:

$$S = S_{background} + \Delta S \frac{[Gp2]}{[Gp2] + K_d}$$

The 95% confidence intervals of the dissociation constants were determined by calculating $\chi^2/\chi_{\min}^2$ for a range of K_d values and identifying the critical values (at which the ratio equals $1 + nF/\nu$ where n = the number of fitted parameters, F is the F statistical distribution value at $\alpha = 0.05$, and ν = degrees of freedom).

Structural and thermal stability characterization

After RP-HPLC, protein samples were lyophilized and resuspended in 100 mM sodium acetate (pH 5) for characterization via circular dichroism spectroscopy (JASCO J-815 Spectropolarimeter). For each clone, secondary structure was assessed by measuring the molar ellipticity from 260 to 200 nm at room temperature. For uniform data reliability, data

was retained up to high tension (HT) voltage of 700 V if the HT exceeded 700 V before wavelength reached 200 nm. Thermal denaturation was conducted by heating the sample from 25 to 95 °C monitored at 218 nm. Melting temperature was calculated by minimizing the sum of squared residues assuming a two-state unfolding equation. Immediately after thermal denaturation, molar ellipticity was measured from 200 to 260 nm at 95 °C to confirm the denaturation, as well as at 25 °C after cooling to check for refolding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding:

This work was funded by the National Institutes of Health (R01 EB023339 and R21 EB021511) and the University of Minnesota (Interdisciplinary Doctoral Fellowship to M.K.).

References

- Asano Y, Dadashpour M, Yamazaki M, Doi N, Komeda H. 2011 Functional expression of a plant hydroxynitrile lyase in *Escherichia coli* by directed evolution: Creation and characterization of highly in vivo soluble mutants. *Protein Eng. Des. Sel* 24:607–616. [PubMed: 21729945]
- Banta S, Dooley K, Shur O. 2013 Replacing antibodies: engineering new binding proteins. *Annu. Rev. Biomed. Eng* 15:93–113. [PubMed: 23642248]
- Calloni G, Zoffoli S, Stefani M, Dobson CM, Chiti F. 2005 Investigating the effects of mutations on protein aggregation in the cell. *J. Biol. Chem* 280:10607–13. [PubMed: 15611128]
- Carter PJ, Lazar GA. 2018 Next generation antibody drugs: Pursuit of the “high-hanging fruit.” *Nat. Rev. Drug Discov* 17:197–223. [PubMed: 29192287]
- Case BA, Kruziki MA, Johnson SM, Hackel BJ. 2018 Engineered Charge Redistribution of Gp2 Proteins through Guided Diversity for Improved PET Imaging of Epidermal Growth Factor Receptor. *Research-article. Bioconjug. Chem* 29:1646–1658. [PubMed: 29579383]
- Chan JY, Hackel BJ, Yee D. 2017 Targeting Insulin Receptor in Breast Cancer Using Small Engineered Protein Scaffolds. *Mol. Cancer Ther* 16:1324–1334. [PubMed: 28468775]
- Chen J, Sawyer N, Regan L. 2013 Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* 22:510–5. [PubMed: 23389845]
- Das D, Georgiadis MM. 2001 A directed approach to improving the solubility of Moloney murine leukemia virus reverse transcriptase. *Protein Sci* 10:1936–41. [PubMed: 11567084]
- Daujotyte D, Vilkaitis G, Manelyte L, Skalicky J, Szyperski T, Klimasauskas S. 2003 Solubility engineering of the HhaI methyltransferase. *Protein Eng* 16:295–301. [PubMed: 12736373]
- Estep P, Caffry I, Yu Y, Sun T, Cao Y, Lynaugh H, Jain T, Vásquez M, Tessier PM, Xu Y. 2015 An alternative assay to hydrophobic interaction chromatography for high-throughput characterization of monoclonal antibodies. *MAbs* 7:553–561. [PubMed: 25790175]
- Feldwisch J, Tolmachev V. 2012 Engineering of affibody molecules for therapy and diagnostics. *Methods Mol. Biol* 899:103–126. [PubMed: 22735949]
- Hackel BJ, Sathirachinda A, Gambhir SS. 2012 Designed hydrophilic and charge mutations of the fibronectin domain: towards tailored protein biodistribution. *Protein Eng. Des. Sel* 25:639–647. [PubMed: 22691700]
- Hayakawa S, Nakai S. 1985 Relationships of Hydrophobicity and Net Charge to the Solubility of Milk and Soy Proteins. *J. Food Sci* 50:486–491.
- Holliger P, Hudson PJ. 2005 Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol* 23:1126–1136. [PubMed: 16151406]

- Islam MM, Nakamura S, Noguchi K, Yohda M, Kidokoro S, Kuroda Y. 2015 Analysis and Control of Protein Crystallization Using Short Peptide Tags That Change Solubility without Affecting Structure, Thermal Stability, and Function. *Cryst. Growth Des* 15:2703–2711.
- Jacobs SA, Wu SJ, Feng Y, Bethea D, O'Neil KT. 2010 Cross-interaction chromatography: A rapid method to identify highly soluble monoclonal antibody candidates. *Pharm. Res* 27:65–71. [PubMed: 19911257]
- Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, Sharkey B, Bobrowicz B, Caffry I, Yu Y, Cao Y, Lynaugh H, Brown M, Baruah H, Gray LT, Krauland EM, Xu Y, Vásquez M, Wittrup KD. 2017 Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci* 114:201616408.
- Japrun D, Chusacutanachai S, Yuvaniyama J, Wilairat P, Yuthavong Y. 2005 A simple dual selection for functionally active mutants of *Plasmodium falciparum* dihydrofolate reductase with improved solubility. *Protein Eng. Des. Sel* 18:457–464. [PubMed: 16120637]
- Jenkins TM, Hickman AB, Dyda F, Ghirlando R, Davies DR, Craigie R. 1995 Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci U S A* 92:6057–6061. [PubMed: 7597080]
- Kato A, Maki K, Ebina T, Kuwajima K, Soda K, Kuroda Y. 2007 Mutational analysis of protein solubility enhancement using short peptide tags. *Biopolymers* 85:12–18. [PubMed: 16948129]
- Kiehlmann LL, Gestwicki JE, Strong LE. 2006 Synthetic multivalent ligands as probes of signal transduction. *Angew. Chemie - Int. Ed* 45: 2348–2368.
- Knowles SM, Wu AM. 2012 Advances in immuno-positron emission tomography: antibodies for molecular imaging in oncology. *J. Clin. Oncol* 30:3884–3892. [PubMed: 22987087]
- Kruziki MA, Bhatnagar S, Woldring DR, Duong VT, Hackel BJ. 2015 A 45-Amino-Acid Scaffold Mined from the PDB for High-Affinity Ligand Engineering. *Chem. Biol* 22:946–956. [PubMed: 26165154]
- Kruziki MA, Case BA, Chan JY, Zudock EJ, Woldring DR, Yee D, Hackel BJ. 2016 Cu-Labeled Gp2 Domain for PET Imaging of Epidermal Growth Factor Receptor. *Mol. Pharm* 13:3747–3755. [PubMed: 27696863]
- Kyte J, Doolittle RF. 1982 A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol* 157:105–132. [PubMed: 7108955]
- Lawson AJ, Walker EA, White SA, Dafforn TR, Stewart PM, Ride JP. 2009 Mutations of key hydrophobic surface residues of 11 β -hydroxysteroid dehydrogenase type 1 increase solubility and monodispersity in a bacterial expression system. *Protein Sci* 18:1552–1563. [PubMed: 19507261]
- Liu Y, Caffry I, Wu J, Geng SB, Jain T, Sun T, Reid F, Cao Y, Estep P, Yu Y, Vaquez M, Tessier PM, Xu Y. 2014 High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. *MAbs* 6:483–492. [PubMed: 24492294]
- Malissard M, Berger EG. 2001 Improving solubility of catalytic domain of human beta-1,4-galactosyltransferase 1 through rationally designed amino acid replacements. *Eur. J. Biochem.* 268:4352–8. [PubMed: 11488931]
- Marston FA, Marston F, Georgiou G, Valax P, Martinez-Alonso M, Garcia-Fruitos E, Ferrer-Miralles N, Rinas U, Villaverde A, Martinez-Alonso M, Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A, Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz R, Aris A, Vera A, Gonzalez-Montalban N, Aris A, Villaverde A, Garcia-Fruitos E, Aris A, Villaverde A, Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A, Garcia-Fruitos E, Rodriguez-Carmona E, Díez-Gil C, Ferraz R, Vazquez E, Corchero J, Garcia-Fruitos E, Seras-Franzoso J, Vazquez E, Villaverde A, Díez-Gil C, Krabbenborg S, Garcia-Fruitos E, Vazquez E, Rodriguez-Carmona E, Ratera I, Martinez-Alonso M, Garcia-Fruitos E, Villaverde A, Martinez-Alonso M, Vera A, Villaverde A, Groot N de, Ventura S, Groot N de, Ventura S, Peternel S, Gaberc-Porekar V, Komel R, Peternel S, Grdadolnik J, Gaberc-Porekar V, Komel R, Jevsevar S, Gaberc-Porekar V, Fonda I, Podobnik B, Grdadolnik J, Menart V, Tokatlidis K, Dhurjati P, Millet J, Beguin P, Aubert J, Worrall D, Goss N, Nahalka J, Gemeiner P, Bucko M, Wang P, Nahalka J, Nidetzky B, Nahalka J, Vikartovska A, Hrabarova E, Nahalka J, Patoprsty V, Nahalka J, Gonzalez-Montalban N, Natalello A, Garcia-Fruitos E, Villaverde A, Doglia S, Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A, Nahalka J, Dib I, Nidetzky B, Roessl U,

- Nahalka J, Nidetzky B, Nahalka J, Mislovicova D, Kavcova H, Sengupta D, Heilshorn S, Tran N, Webster T, Villaverde A, Vallejo L, Rinas U, Arai Y, Yamaguchi H, Fukuoka S, Wanarska M, Hildebrandt P, Kur J, Huang L, Wang J, Zhong Z, Peng L, Chen H, Xu Z, Rodriguez-Carmona E, Cano-Garrido O, Seras-Franzoso J, Villaverde A, Garcia-Fruitos E, Peternel S, Komel R, Villa R, Lotti M, Gatti-Lafrancini P. 1986 The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. *Biochem. J* 240:1–12. [PubMed: 3548705]
- Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR. 1999 A simple in vivo assay for increased protein solubility. *Protein Sci* 8:1908–1911. [PubMed: 10493593]
- Mosavi LK, Peng Z -y. 2003 Structure-based substitutions for increased solubility of a designed protein. *Protein Eng. Des. Sel* 16:739–745.
- Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. 1995 How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4:2411–2423. [PubMed: 8563639]
- Pedone E, Saviano M, Rossi M, Bartolucci S. 2001 A single point mutation (Glu85Arg) increases the stability of the thioredoxin from *Escherichia coli*. *Protein Eng* 14:255–60. [PubMed: 11391017]
- Roosild TP, Choe S. 2005 Redesigning an integral membrane K⁺ channel into a soluble protein. *Protein Eng. Des. Sel* 18:79–84. [PubMed: 15788421]
- Schein CH. 1990 Solubility as a function of protein structure and solvent components. *Bio/Technology* 8:308–317. [PubMed: 1369261]
- Schmidt MM, Wittrup KD. 2009 A modeling analysis of the effects of molecular size and binding affinity on tumor targeting. *Mol. Cancer Ther* 8:2861–2871. [PubMed: 19825804]
- Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005 The FoldX web server: an online force field. *Nucleic Acids Res* 33:W382–8. [PubMed: 15980494]
- Scott AM, Wolchok JD, Old LJ. 2012 Antibody therapy of cancer. *Nat. Rev. Cancer* 12:278–287. [PubMed: 22437872]
- Škrlec K, Štrukelj B, Berlec A. 2015 Non-immunoglobulin scaffolds: A focus on their targets. *Trends Biotechnol* 33:408–418. [PubMed: 25931178]
- Sørensen HP, Mortensen KK. 2005 Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact* 4:1. [PubMed: 15629064]
- Spencer DM, Wandless TJ, Schreiber SL, Crabtree GR. 1993 Controlling signal transduction with synthetic ligands. *Science* 262:1019–1024. [PubMed: 7694365]
- Stern LA, Case BA, Hackel BJ. 2013 Alternative non-antibody protein scaffolds for molecular imaging of cancer. *Curr. Opin. Chem. Eng* 2:425–432.
- Strub C, Alies C, Lougarre A, Ladurantie C, Czaplicki J, Fournier D. 2004 Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochem* 5:1–6. [PubMed: 14715092]
- Sule SV, Dickinson CD, Lu J, Chow CK, Tessier PM. 2013 Rapid analysis of antibody self-association in complex mixtures using immunogold conjugates. *Mol. Pharm* 10:1322–1331. [PubMed: 23383873]
- Sule SV, Sukumar M, Weiss IV WF, Marcelino-Cruz AM, Sample T, Tessier PM. 2011 High-throughput analysis of concentration-dependent antibody self-association. *Biophys. J* 101:1749–1757. [PubMed: 21961601]
- Thompson DB, Cronican JJ, Liu DR. 2012 Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. *Methods Enzymol* 503:293–319. [PubMed: 22230574]
- Thurber GM, Schmidt MM, Wittrup KD. 2008a Antibody tumor penetration: transport opposed by systemic and antigen-mediated clearance. *Adv. Drug Deliv. Rev* 60:1421–1434. [PubMed: 18541331]
- Thurber GM, Schmidt MM, Wittrup KD. 2008b Factors determining antibody distribution in tumors. *Trends Pharmacol. Sci* 29:57–61. [PubMed: 18179828]
- Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. 2007 The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J. Mol. Biol* 369:1318–1332. [PubMed: 17482644]
- Trevino SR, Scholtz JM, Pace CN. 2008 Measuring and increasing protein solubility. *J. Pharm. Sci* 97:4155–4166. [PubMed: 18240286]

- Wang W 1999 Instability, stabilization, and formulation of liquid protein pharmaceuticals. *Int. J. Pharm* 185:129–188. [PubMed: 10460913]
- Wu SJ, Luo J, O'Neil KT, Kang J, Lacy ER, Canziani G, Baker A, Huang M, Tang QM, Raju TS, Jacobs S a., Teplyakov A, Gilliland GL, Feng Y. 2010 Structure-based engineering of a monoclonal antibody for improved solubility. *Protein Eng. Des. Sel* 23:643–651. [PubMed: 20543007]

Figure 1. Gp2 Hydrophilicity.

(A) Wild-type Gp2 solution structure with amino acids labeled based on hydrophilicity (PDB ID: 2WNM). (B) *Left*, Hydrophobicity, as reported by Kyte and Doolittle (Kyte and Doolittle, 1982), is plotted against SASA, obtained by ASA server (see *Materials and Methods* for details). The twelve sites in the two loop regions, which are typically diversified in Gp2 engineering are shown as open triangles. *Right*, Sites for potential beneficial mutations are labeled as green. Site W31 is excluded from mutational analysis because it is the hydrophilic arginine within the context of the EGFR-binding Gp2_{E2.2.3}.

Figure 2. Relative binding of clones with single mutation.

The parental (P) Gp2 domain and each mutant were incubated with EGFR⁺ A431 cells at concentrations of 3 to 243 nM; binding was detected by flow cytometry, then normalized relative to the signal obtained with 243 nM of the parental clone. Mutants are ordered by the decreasing hydrophilicity of the final amino acid within each site. Mutants selected for the combined mutations are marked with asterisks. Data is represented as mean $\pm$ standard deviation of 2 – 7 replicates (except L22A (3 and 9 nM) and L28A, 27 nM, which were single trials).

Figure 3. Thermal stability.

Purified Gp2 (13 – 63 μ M in 100 mM sodium acetate, pH 5) was heated from 25 °C to 95 °C at a ramping rate of 1 °C/min and monitored by circular dichroism spectroscopy at a wavelength of 218 nm. The midpoint of thermal denaturation (T_m) was computed assuming two-state unfolding. *Left*, data for Y18N. *Right*, tabulated data for the parent and selected single mutants.

Figure 4. Final yield.

All mutants were produced in 1 L of *E. coli* culture and purified via metal affinity chromatography. Yields were evaluated based on absorbance at 280 nm, SDS-PAGE, and mass spectrometry. Bars represent mean, and circles represent specific samples (n = 2–3).

Figure 5. Retention time.

All mutants were purified and analyzed by RP-HPLC using a 10–90% acetonitrile gradient over 15 min. Value and error bars represent mean and standard deviation ($n = 2-5$).

Figure 6. Aggregation propensity.

All mutants were purified and analyzed via AC-SINS for aggregation. Gold nanoparticles were conjugated with anti-His₆ antibodies and purified Gp2 domains. The shift in the wavelength of maximum absorbance was recorded. (A) Parental clone and mutant I were evaluated at 60 mM salt (solid) and 300 mM salt (dashed). Average $\pm$ standard deviation of triplicate samples shown. * indicates $p < 0.05$ between P and I; ** indicates $p < 0.005$ (B) Variants evaluated at 60 mM salt. Average of triplicates shown. Standard deviation of P, E, and I are shown for examples. Other deviations suppressed for visual clarity (average relative standard deviation = 18%).

Figure 7. Midpoints of thermal denaturation.

Purified Gp2 domains were evaluated via circular dichroism as in Figure 4. On average, T_m of the mutants are 25 °C lower than T_m of the parental clone. Value and error bars represent mean and standard deviation ($n = 2-3$).

Figure 8. Binding characterization.

The parental (P) Gp2 domain and each mutant were incubated at varying concentrations with EGFR⁺ A431 cells. Binding was quantified by flow cytometry. Dissociation constants (K_d) were calculated assuming a single-site binding model using data from multiple ($n = 3 - 6$) trials. (a) Representative binding titrations for mutant E ($K_d = 58$ nM with 95% confidence interval of 53 – 63 nM, $n = 5$ trials) and mutant I ($K_d = 19$ nM with 95%

confidence interval of 13 – 28 nM, n = 3 trials). (b) Dissociation constants. Uncertainties on the dissociation constants were calculated based on χ^2 confidence interval.

Table 1**Mutant information.**

Sites in the wild-type Gp2 were selected based on hydrophilicity and SASA (as introduced in Figure 1B), as well as the theoretical stability predicted using FoldX (with stated uncertainty of 0.5 kcal/mol) and natural frequency as observed in 34 sequences of Gp2 homologs.

WT	Site	SASA	Mutant	G _{fold} (kcal/mol)	Natural Frequency
F	13	47%	N	1.7	0%
			Y	0.2	6%
Y	18	57%	N	0.9	0%
T	21	48%	N	-0.7	3%
L	22	39%	S	2.3	0%
			A	1.9	0%
L	26	33%	N	1.7	15%
			S	2.0	0%
			T	1.8	24%
			A	1.1	0%
L	28	37%	S	1.9	0%
			A	1.0	3%
T	41	55%	N	-0.1	0%

Table 2
Combinatorial mutation summary.

The six beneficial point mutations identified by results introduced previously, namely Y18N, T21N, L22S, L26T, L28S, T41N, were combined to construct the following multi-mutants.

Name	AA at site 18, 21, 22, 26, 28, 41	Number of mutants
Parental(P)	YTLLLT	-
A	NNSTSN	6
B	YNSTSN	5
C	NTSTSN	5
D	NNLTSN	5
E	NNSLSN	5
F	NNSTLN	5
G	NNSTST	5
H	NTSTST	4
I	YTSTST	3