Published in final edited form as:

Methods Enzymol. 2012 ; 503: 293–319. doi:10.1016/B978-0-12-396962-0.00012-4.

Engineering and Identifying Supercharged Proteins for Macromolecule Delivery into Mammalian Cells

David B. Thompson[‡], James J. Cronican[‡], and David R. Liu^{*}

Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138 USA

Abstract

Supercharged proteins are a class of engineered or naturally occurring proteins with unusually high net positive or negative theoretical charge. Both supernegatively and superpositively charged proteins exhibit a remarkable ability to withstand thermally or chemically induced aggregation. Superpositively charged proteins are also able to penetrate mammalian cells. Associating cargo with these proteins, such as plasmid DNA, siRNA, or other proteins, can enable the functional delivery of these macromolecules into mammalian cells both *in vitro* and *in vivo*. The potency of functional delivery in some cases can exceed that of other current methods for macromolecule delivery, including the use of cell-penetrating peptides such as Tat, and adenoviral delivery vectors. This chapter summarizes methods for engineering supercharged proteins, optimizing cell penetration, identifying naturally occurring supercharged proteins, and using these proteins for macromolecule delivery into mammalian cells.

Introduction

Most medicines are small molecules, chemical compounds containing less than approximately one hundred atoms. For many diseases, however, small molecule-based therapies have not been found. Recent efforts to discover bioactive molecules including human therapeutics have increasingly focused on macromolecules— for the purpose of this chapter, proteins or nucleic acids. Indeed, ~180 protein drugs are currently prescribed including insulin, erythropoietin, interferons, and a variety of antibodies.¹ Due to their significant folding energies, macromolecules are able to adopt large, stable three-dimensional conformations suitable for strong binding to targets even when they lack hydrophobic clefts commonly associated with small-molecule binding. Moreover, the strength of macromolecule-target binding can be sufficient to interfere with native protein-protein or protein-nucleic acid interfaces that have traditionally been difficult to address using small molecules.² The stability, size, and complexity of macromolecules can result in specificities that are not easily achievable using small molecules, as demonstrated by certain antibodies and RNAi agents.

Unfortunately macromolecules are typically not able to diffuse into cells and as a result, virtually all existing macromolecule therapeutics address extracellular targets. Perhaps the most challenging and widespread impediment to the broader use of proteins and nucleic acids as therapies is therefore the delivery of macromolecules into cells and into subcellular locations of interest.

^{*}To whom correspondence should be addressed: Professor David R. Liu, Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138 USA, drliu@fas.harvard.edu, (617) 496–1067 (office), (617) 496–5688 (fax).

[‡]These authors contributed equally to this work.

To address this challenge, various macromolecule delivery approaches have been developed including electroporation,³ ultrasound-mediated plasmid delivery,⁴ viral delivery,⁵ nebulization,⁶ and direct chemical modification.⁷ In addition, other strategies associate a macromolecule with a non-viral delivery vehicle such as lipidoids,⁸ liposomes,⁹ dendrimers,¹⁰ cationic polymers,¹¹ inorganic nanoparticles,¹² carbon nanotubes,¹³ cell-penetrating peptides,¹⁴ a small molecule,¹⁵ or receptor ligands¹⁶. In 2009, we reported the use of cationic supercharged proteins as general and potent vectors for macromolecule delivery into mammalian cells.

In this chapter, we summarize the basic properties of supercharged proteins, as well as methods for engineering them, identifying naturally occurring supercharged proteins within a proteome database, and applying them to deliver nucleic acids and proteins *in vitro* and *in vivo*.

Discovery and Basic Properties of Supercharged Proteins

We reported the creation and characterization of supercharged proteins in 2007.¹⁷ Engineered supercharged proteins are the product of extensive mutagenesis in which solvent-exposed residues throughout the protein's surface were substituted with either acidic or basic amino acids. In several cases tested, the resulting proteins retain much of their original activity. For example, we generated supercharged GFP proteins with a wide range of net theoretical charges (−30 to +48) that possess nearly identical excitation and emission spectra as starting GFP (Fig. 1a).¹⁷ Mutations that alter the excitation and emission maxima of GFP, resulting in blue, cyan, and yellow fluorescent proteins, are also amenable to supercharging (Fig. 1b).¹⁸ Likewise, supercharged streptavidin (a tetramer with +52 net theoretical charge) also retains the ability to tetramerize and bind biotin, albeit at a reduced affinity.¹⁷ Lastly, supercharged glutathione *S*-transferase (a dimer with −40 net theoretical charge) retains much of its catalytic activity.¹⁷ The preservation of function across a diverse set of proteins following extensive surface mutagenesis is a testament to the mutability of residues identified in the supercharging method (described below), and to the potential generality of protein supercharging for generating aggregation-resistant and cell-penetrating protein reagents.

Aggregation Resistance of Supercharged Proteins

Supercharged proteins are remarkably resistant to both thermal- and chemical-induced aggregation and can efficiently refold to regain much of their original function. We demonstrated this aggregation resistance with supercharged GFP, streptavidin, and glutathione *S*-transferase.¹⁷ Boiled supercharged proteins denature and lose their activity similar to their wild-type counterparts. However, their ability to avoid aggregation events including the association of exposed hydrophobic residues, even in the unfolded state, enables supercharged proteins to refold and regain much of their original activity after cooling (Fig. 2).¹⁷ Furthermore, extended exposure to 40% 2,2,2-trifluoroethanol (TFE), conditions that induce the denaturation and aggregation of typical proteins, does not result in any measurable aggregation of +36 GFP.¹⁷ These observations suggest that supercharged proteins can avoid common aggregation pathways upon thermal or chemical denaturation, and in the absence of aggregation can refold into functional proteins once conditions are restored that favor folding. Note that the net charge, rather than the number of charged residues, is necessary for these unusual properties; indeed, wild-type proteins and supercharged proteins can contain similar numbers of charged residues even though their total net charge differs dramatically.¹⁷

In support of this model, we have also created a conditionally supercharged protein in the form of a GFP containing 39 histidine residues, primarily at the positions previously

substituted with Lys/Arg in +36 GFP. This His₃₉ GFP also folds and fluoresces, and like +36 GFP recovers fluorescence when boiled and cooled at low pH (Fig. 3). As the pH increases past ~6, however, the histidine side chains become neutral, the protein decreases in charge, and the ability to refold after boiling is lost (Fig. 3)

The high charge of supercharged proteins enables them to reversibly complex with oppositely charged macromolecules, including nucleic acids. For example, upon mixing +36 GFP with nucleic acids, complexes form that sequester the nucleic acid out of solution and generate particles that can be isolated by centrifugation.¹⁷ These complexes can be dissociated by the addition of high concentration of salt, presumably by competing the electrostatic interactions necessary for complex formation.¹⁷

Cell Penetration of Supercharged Proteins

Many known non-viral macromolecule delivery vehicles are cationic. These vehicles typically bind to negatively charged components on the cell membrane and are endocytosed into cells, either by stimulating endocytosis or by simple membrane recycling.¹⁹ In some cases, a fraction of these endocytosed molecules are able to escape endosomes and access the intracellular environment.¹⁹ The HIV Tat cell-penetrating peptide is cationic and is one example of a delivery vehicle that is thought to use this macromolecule delivery mechanism.¹⁹ Engineered proteins with cationic regions have also been found to penetrate cells, including a pentamutant GFP containing a patch of five arginine residues.²⁰

We hypothesized and subsequently demonstrated that superpositively charged proteins such as +36 GFP can penetrate mammalian cells with potencies much greater than that of cationic peptides or modestly cationic engineered proteins.^{21,22} Live-cell fluorescence microscopy images of mammalian cells reveal that the entire cell membrane becomes associated with +36 GFP within seconds of exposure to low nanomolar concentrations of the protein.²¹ Within minutes, +36 GFP can be observed within the body of the cell as bright puncti, presumably contained within endosomes. Following uptake, +36 GFP-containing puncti can be found distributed throughout the extranuclear space.²¹

We have probed the requirements for +36 GFP uptake to better understand the mechanism by which it achieves such potent internalization. During internalization assays, the treated cells can be washed several times with heparin, a highly sulfated glycosaminoglycan, to remove surface-bound +36 GFP. This washing procedure effectively removes surface-bound +36 GFP as measured by both flow cytometry and confocal microscopy, while the intracellular puncti remain.^{21,23,24} Examination of cells washed in this manner therefore enables measurement of total internalized protein as well as measurements of uptake and trafficking kinetics. The high negative charge of heparin also enables its use as a competitor with the cell surface for +36 GFP binding. For example, at 4 °C endocytosis is known to be inhibited in mammalian cells.²⁵ At 4 °C, +36 GFP will bind to the outside of the cell membrane but will not be internalized. Washing cells exposed to +36 GFP with heparin at 4 °C will remove all +36 GFP fluorescence from the cell.²¹ These results indicate that +36 GFP does not passively traverse the lipid bilayers of the cell's plasma membrane but instead requires endocytosis to drive internalization.²¹

The surface of mammalian cells is decorated with sulfated proteoglycans that contribute greatly to the anionic nature of the cell surface. Treatment with sodium chlorate, an inhibitor of proteoglycan sulfation, inhibits internalization of +36 GFP, as well as cell-surface association of +36 GFP.²¹ Furthermore, cells that have been genetically modified to produce only non-sulfated proteoglycans are also incapable of internalizing +36 GFP.²¹ Although these experiments were performed with +36 GFP, the results likely apply to all superpositively charged proteins in light of our observation that +36 GFP acts as a

competitive inhibitor to the internalization of six unrelated naturally supercharged human proteins.²⁶ Together, the above observations suggest an essential role for electrostatic interactions, derived in large part from sulfated proteoglycans,²⁷ and subsequent endocytosis in supercharged protein internalization.

While it is clear that positive charge promotes internalization of supercharged proteins, the relationship between charge and cell-penetration is not so simple.²⁸ We have generated dozens of charged GFP variants from shuffling sub-sequences of starting GFP, +15 GFP, +25 GFP, +36 GFP, and +48 GFP.¹⁸ Using a single protein scaffold to display residues contributing to a variety of net charges allows examination of the role of charge magnitude and distribution without the complication of gross structural variation. We found scGFP cellular uptake to be highly charge-dependent and strongly sigmoidal, exhibiting both low-potency and high-potency forms with the transition occurring near +21 charge units (Fig. 4).²² Similarly, mammalian cells incubated with His₃₉ GFP in media of pH values from 4 to 8 exhibit a pH-dependent internalization whereas the uptake of +36 GFP and non-supercharged starting GFP (stGFP) are unaffected by changes in pH over this range (Fig. 5).

This sigmoidal charge-cell penetration relationship has not been observed for oligoarginine peptide reagents, where uptake efficiency generally *decreases* past ~+15 charge magnitude.²² It therefore appears that supercharged protein surfaces can attain higher net charges than cationic peptide reagents without eroding cell penetration capabilities. Even when comparing the cell-penetration potency of supercharged proteins with that of synthetic oligo-Lys/Arg peptides of similar or identical theoretical net charge, we observed that the cationic peptides were consistently outperformed by supercharged GFPs.²² The mechanism underlying this difference remains the subject of active investigation in our laboratory. It is tempting to speculate that the structure of the surface of a folded, globular protein may engage in cooperative binding or cross-linking of anionic cell-surface receptors more effectively than similarly charged unstructured peptides. Indeed, we have observed certain forms of endocytic stimulation from supercharged proteins that are not observed upon treatment with unstructured cationic peptides of similar theoretical net charge.²²

Macromolecule Delivery by Supercharged Proteins

Supercharged proteins are general and potent vehicles for delivery of macromolecules into mammalian cells. As summarized above, simple mixing of siRNA or plasmid DNA with +36 GFP results in the formation of electrostatic complexes. Incubation of these complexes with mammalian cells leads to delivery of the associated nucleic acid even into cell lines known to be resistant to lipid-mediated transfection.²¹

Supercharged proteins have also been shown to deliver a variety of functional protein molecules directly into cells by translational fusion to supercharged proteins.²³ We initially demonstrated delivery of mCherry, ubiquitin, and Cre recombinase into multiple cell types using +36 GFP.²³ These three proteins provide separate complementary measures of delivery in cultured mammalian cells. Delivery of mCherry provides a quantitative measure of simple protein uptake into target cells. Ubiquitin delivery enables a measure of endosomal escape due to the cytosol-specific cleavage of ubiquitin by deubiquitinating enzymes.²³ Cre recombinase functions a general measure of extra-endosomal delivery of functional proteins, as a positive delivery phenotype is only observed upon nuclear localization and enzymatic activity of the Cre protein in appropriate reporter cell lines. We observed effective protein delivery into mammalian cells in all three cases using supercharged proteins, including some examples *in vivo*.²³

Most recently we have demonstrated that these delivery properties are not unique to engineered cationic supercharged proteins but are instead present in a diverse class of

naturally occurring supercharged human proteins.²⁶ We analyzed the human proteome for proteins with unusually high net positive charge and found a large, diverse class of proteins (possibly > 2% of the human proteome) that potently delivers protein in functional form into mammalian cells both *in vitro* and in retinal, pancreatic, and white adipose tissues *in vivo* (Fig. 6).²⁶ These findings reveal a diverse set of macromolecule delivery agents for *in vivo* applications, and also raise the possibility that some human proteins may penetrate cells as part of their native biological functions.

METHODS

Theory Underlying Protein Supercharging

Our initial motivation for generating supercharged proteins was the desire to test the hypothesis that highly charged proteins were less likely to aggregate than less charged proteins. Supercharged proteins enabled the limits of this hypothesis to be tested. In theory, protein surfaces rather than buried side chains are the most likely regions of a protein to tolerate these substitutions. Protein folding is driven primarily by the loss of solvation of hydrophobic residues.²⁹ Water molecules, which possess limited freedom when interacting with hydrophobic amino acid side chains, gain entropy as the hydrophobic residues collapse and fold into the interior of the protein. Amino acid side chains that do not become buried upon protein folding remain solvent-exposed and are in a very similar environment in the folded or unfolded state. Due to the similar environments in the folded and unfolded states, solvent-exposed amino acids are thought to contribute less energy towards stabilizing the folded state on the average compared with buried residues.³⁰ In idealized cases in which the side chain of an amino acid is completely solvent exposed and makes no interactions with the rest of the protein, the identity of the side chain should be irrelevant to the thermodynamics of protein folding. This simple model was supported by our experiments in which up to 36, 32, and 30 mutations were installed in GFP, in the streptavidin tetramer, and in the GST dimer, respectively, without abolishing the fluorescence, biotin-binding, or catalytic function of these three proteins. Because three to five random mutations in a 50 kDa protein typically eliminates protein function,³⁰ it is highly unlikely that such a large number of mutations (up to 15% of the total number of residues) could be made to these proteins without destroying their fold and function were they not restricted to the most solvent-exposed amino acids.

Engineering Supercharged Proteins

Engineering a supercharged protein first requires that the structure of the parent protein has been determined and is available for analysis. PDB files containing three-dimensional proteins structures are widely available on public databases including the Research Collaboratory for Structural Bioinformatics (<http://www.rcsb.org>). To supercharge GFP, for example, the structure of an optimized GFP was downloaded as the file 2B3P.pdb. In the case of GFP, the solvent-exposed residues were selected by manually inspecting the crystal structure using a PDB imaging software such as PYMOL. Engineering of supercharged streptavidin and GST proteins was performed computationally by ranking residues by their average number of neighboring atoms (within 10 Å) per side-chain atom (AvNAPSA). Charged or highly polar solvent-exposed residues (Asp, Glu, Arg, Lys, Asn, and Gln) were mutated either to Glu (unless the starting residue was Asn, in which case to Asp) for negative supercharging; or to Lys for positive supercharging. The AvNAPSA Perl script is provided upon request.

As an example, supercharging streptavidin proceeds as follows:

- 1) Download the pdb file of interest. In this case, 1stp.pdb can be downloaded from <http://www.pdb.org/pdb/explore/explore.do?structureId=1STP>.

- 2) Install Perl. (Available at <http://www.perl.org/get.html>.)
- 3) Save the AvNAPSA file to the same folder as the 1stp.pdb file. For example, save both files to C:\Temp.
- 4) In the command prompt, enter: C:\Temp>perl avnapsa 1STP.pdb
- 5) The AvNAPSA program will output a list of values associated with each amino acid in the protein.
- 6) Copy the AvNAPSA amino acid list to a spreadsheet program and rank the list by AvNAPSA values, starting from the residues with the lowest AvNAPSAs (the highest degree of solvent exposure).
- 7) To positively supercharge the protein, change the polar non-positive amino acids (D, E, N, or Q) to lysine (K), starting from the top of the list. To negatively supercharge the protein, change the polar non-negative amino acids (R, K, or Q) to glutamate (E), except for asparagine (N) which should be changed to aspartate (D). The mutations should be made in order of increasing AvNAPSA values. The total number of mutations to supercharge a protein is dependent upon the size of the protein being analyzed and the desired degree of supercharging.
- 8) Optional: if a family of protein sequences related to the protein of interest is available, perform a sequence homology alignment and avoid mutation of evolutionarily conserved residues, regardless of AvNAPSA.
- 8) Rank the sequence-altered amino acid list by residue number to reform the correct amino acid sequence.

If a PDB file is not available for a desired protein, a related protein PDB file may be identified by searching the PDB database for sequence homology to the desired protein. BLASTP is a powerful application for this purpose (<http://blast.ncbi.nlm.nih.gov/>). A family of related structures can be used as a template to generate a PDB file for the desired protein using a modeling program such as MODELLER (<http://salilab.org/modeller/>). The generated PDB file can be used as the input for the supercharging algorithm or to model the finished supercharged protein.

The supercharging method is compatible with other means of identifying mutable residues. For example, Biopython (biopython.org) has built-in modules for calculating amino acid exposure by the number of neighboring amino acid alpha carbons within a specified radius of each amino acid alpha carbon or by the number of alpha carbons in the half-sphere defined by the alpha carbon to beta carbon vector. Other programs such as DSSP (<http://swift.cmbi.ru.nl/gv/dssp/>) are also able to assign accessible surface area values to each amino acid to provide another estimation of relative solvent exposure. Mutable amino acids can also be identified without any necessary dependence on solvent exposure by using protein design software such as ROSETTADesign. Each implementation for the purposes of supercharging will likely have different advantages and disadvantages. Incomplete knowledge of a protein's folding and functional requirements can result in non-functional supercharged variants. It is therefore useful to combine the above method with techniques for generating and screening libraries of supercharged proteins for variants that function in the desired context.

Screening for Functional Supercharged Proteins

Incomplete knowledge of a protein's folding and functional requirements can result in non-functional supercharged variants for a variety of reasons presented below. It is therefore

useful to combine the above method with techniques for generating and screening libraries of supercharged proteins for variants that function in the desired context.

While solvent-exposed amino acids should be more easily mutated without impairing protein structure or function, the relationship between solvent exposure and mutability is imperfect. For example, the folding energy of most proteins is 5–15 kcal/mol while the strength of a hydrogen bond is ~0–9 kcal/mol depending on its environment.²⁹ If a solvent-exposed side chain contributes one or more intramolecular hydrogen bonds or other electrostatic interactions then it may not be mutable. Indeed, there is evidence that some naturally thermostable proteins are stabilized by the presence of multiple surface salt bridges.³¹ Similarly, since the native protein structure is in equilibrium with an ensemble of non-native folded states, a mutated side chain may stabilize a non-native structure in which it is no longer as solvent-exposed or in which it makes new interactions, reducing the relative stability of the native protein. Moreover, side chains regardless of their degree of solvent exposure may play important roles in the expression and folding of some proteins.²⁹ Finally, some residues in X-ray or even NMR structures of proteins may adopt non-native conformations that alter their apparent AvNAPSA values.

Given the challenges associated with the correct prediction of many mutable residues in a protein, it is prudent to generate a collection of candidate supercharged variants of a protein and screen or select those with desired properties. For example, using the AvNAPSA method, one can generate a hypothetical protein that contains many more supercharging mutations than the final protein is likely to withstand, then fragment and shuffle the gene encoding the “overcharged” protein with the gene encoding the wild type protein, and screen or select to isolate variants that possess the desired function. For example, we designed by manual inspection (prior to development of the AvNAPSA method) a GFP with a net theoretical charge of +36 containing 29 mutations. When the gene corresponding to this protein was overexpressed, +36 GFP expressed well and was fluorescent. However, when we designed using the same principles a GFP with a net theoretical charge of –39, the resulting protein did not express in *E. coli*. The following is a summary of the method we used to generate negatively supercharged GFP variants:

1. The starting GFP and –39 GFP genes were ordered as two sets of overlapping 40-base pair oligonucleotides and resuspended in water to 100 μ M.³²
2. The oligonucleotides for starting GFP and –39 GFP were mixed in a 1:20 molar ratio and 6 μ L of the mixture was added to 28 μ L water, 4 μ L 10X T4 Ligase Buffer and 2 μ L T4 polynucleotide kinase (10 U/ μ L, New England Biolabs). The kinase reaction was incubated for 20 min at 37 °C, 2 min at 94 °C and then cooled at 0.1 °C/s to 70 °C.
3. While the oligonucleotides were being phosphorylated, a mixture of 10 μ L 10X T4 Ligase Buffer and 50 μ L water was warmed to 70 °C and added to the kinase reaction once it reached 70 °C. The reaction was incubated for another 30 min at 70 °C, then cooled at 0.1 °C/sec to 16 °C. A 5 μ L aliquot was removed as the sample before ligation.
4. To the reaction was added 2 μ L of T4 DNA ligase (5 U/ μ L, Invitrogen). The reaction was incubated for 1 hour at 16 °C. Another 5 μ L aliquot was removed as the sample after ligation.
5. The ligation product (1 μ L) was used as the template for a PCR reaction using a terminal 5' forward oligo and the 3' reverse oligo. The integrity of the library was confirmed by running out the “before ligation”, “after ligation” and PCR samples on a 1.2% agarose gel containing ethidium bromide. Ligation should result in

formation of a broad smear upward that is present in the sample after ligation but not before ligation. An ideal PCR reaction will predominantly contain a single band at the expected gene length.

6. The resulting shuffled library of GFP genes was then digested with restriction enzymes, purified by agarose gel electrophoresis, ligated into a protein expression plasmid, and transformed into a cloning strain of *E. coli*.
7. The resulting colonies from transformation were scraped from the agar plate, grown in liquid culture and harvested for plasmid purification.
8. The plasmid library was transformed into a protein expression host such as BL21(DE3) cells and plated on an IPTG-containing agar plate.
9. The next morning, the plate was illuminated with ultraviolet light to reveal colonies that contain functional fluorescent GFP clones. These colonies were picked for sequencing of the plasmid and larger scale protein expression and purification.

In this manner, -30 GFP, which contained 15 of the 20 planned mutations, and -25 GFP, which contained 12 mutations, were identified. As the costs of gene synthesis decrease, libraries can also be generated by *in vitro* recombination of full-length genes using methods such as DNA shuffling,³³ StEP,³⁴ and NEBNext (New England Biolabs). While most proteins are not fluorescent, a wide variety of screens that use chromogenic³⁵ or fluorogenic assays,³⁶ or that couple the function of a protein to transcription of a reporter gene,³⁷ can also be used to identify functional protein variants

Identifying Naturally Supercharged Proteins

One concern of using engineered proteins such as +36 GFP for *in vivo* applications is that the protein may provoke an immune system response. To identify naturally occurring supercharged human proteins that may be less immunogenic than engineered supercharged proteins, we sorted the human proteome to identify the most positively charged proteins (Fig. 7). We found that these proteins are also able to penetrate mammalian cells and functionally deliver associated macromolecules *in vitro* and *in vivo*. The naturally supercharged human protein (NSHP) Python script that we use to identify naturally supercharged human proteins from the PDB and Swissprot databases will be provided upon request. The following is a protocol for identifying supercharged proteins from collection of protein sequences:

1. Supercharged proteins can be identified by first collecting all of the primary sequences for the proteins you are interested in. A curated collection of all known natural proteins can be downloaded from the Uniprot website (<http://www.uniprot.org/downloads>) as uniprot_sprot.fasta.gz. The uniprot_sprot.fasta file can be extracted and each protein entry will contain the name of the protein, the source organism, whether the protein is experimentally confirmed or predicted, and the protein primary sequence in FASTA format. Another important source of proteomic data is from the Research Collaboratory for Structural Bioinformatics website (www.rcsb.org) which contains only proteins with X-ray crystallography or NMR structural information. The advanced search function on the RCSB website allows for filtering for proteins that are from the source organism, such as *Homo sapiens*, and expression host, such as *Escherichia coli*. The resulting list will provide proteins that will likely express in sufficient yield for macromolecule delivery experiments and will also likely have accompanying biochemical or biological information.
2. A scripting program such as Python or Perl is used to generate a calculated molecular weight and net theoretical charge from the primary sequences of the

protein entries. It is helpful to create a graph of the proteome with net theoretical charge on the x-axis and molecular weight on the y-axis. Supercharged proteins can be identified from this dot plot by manual inspection, or the proteins can be ranked by net theoretical charge or by the ratio of net theoretical charge to molecular weight. Our experience working with naturally supercharged proteins and with various charged GFPs suggests that a positive charge to molecular weight (in kDa) ratio greater than 0.75 represents a good starting point to identify proteins that can penetrate cells.

3. Genes corresponding to the supercharged protein can be ordered from various companies. Genes for natural proteins, for example, can be ordered from companies with cDNA libraries such as Open Biosystems (www.openbiosystems.com). Genes for proteins identified through the PDB database can be requested from the lab that produced the structural information. Genes can also be ordered from gene synthesis companies with codon optimization for recombinant expression.

This approach is distinct from and complementary to approaches that identify cell penetrating peptides such as QSAR, which is based on algorithms developed using a empirical training set of known cell-penetrating and non-penetrating examples.³⁸

Protein Expression

In our experience, supercharged proteins generally express well in standard *E. coli* protein expression strains such as BL21(DE3). Protein-specific variations in protocol may be necessary, and parameters including induction time and temperature can significantly impact the yield of the expressed supercharged protein on a case-by-case basis.

Small-scale expression cultures should be tested to determine the optimal expression conditions and the solubility of the expressed protein before performing large-scale expression experiments. The following procedure can be used as a starting point for protein-specific protocol optimization:

Materials

LB broth, 2xYT broth and agar plates containing appropriate antibiotic.

pET expression plasmid.

Competent BL21(DE3) cells.

Resuspension Buffer: PBS with 2 M NaCl, 20 mM imidazole pH 7.5, with one tablet of EDTA-free Complete Protease Inhibitor (Roche) per 50 mL buffer.

Elution Buffer: PBS with 2 M NaCl, 500 mM imidazole pH 7.5.

Ni-NTA agarose resin (Qiagen).

1. Transform the supercharged protein pET expression plasmid into BL21(DE3) and plate on 2xYT media with appropriate antibiotic. Incubate overnight at 30 °C to minimize the risk of colony overgrowth, autoinduction, and toxicity from leaky protein expression.
2. The following day, pick a medium sized, isolated colony and inoculate a 5 mL overnight seed culture of 2xYT media. Incubate overnight at 30 °C.
3. The following day, pellet the seed culture by centrifugation at $4000 \times g$ for 5 min and resuspend the pellet in LB media.

4. Inoculate a 1 L expression culture of LB media with the resuspended overnight seed culture. Grow the culture at 37 °C to an OD₆₀₀ of ~0.6 and induce expression with 1 mM IPTG.
5. Following induction, incubate the expression culture for 4 hours at 30 °C.
6. Harvest the culture by centrifugation at 6000 × g for 10 min. Pellets can be stored at –80 °C or processed immediately.

Protein Purification

The following protocol is for standard His-tagged proteins; other purification protocols are also suitable for supercharged proteins. The protein samples should be kept on ice at all times.

1. Thaw frozen pellets and resuspend in 40 mL resuspension buffer.
2. Split the the resuspended pellets into two 20 mL fractions and lyse each by sonication in a Sonic Dismembrator 550 (Fisher Scientific) for 3 min with a cycle of 1 s on and 1 s off at power level of 4.
3. Pellet cell debris by centrifugation at 4000 × g for 10 min. For soluble protein preparations, transfer the supernatant to a new 50 mL conical tube. For insoluble protein preparations, resuspend and homogenize the pellet in resuspension buffer containing 1% Tween 20 to remove the lipid contents of the pellet. Pellet the still insoluble inclusion bodies by centrifugation at 4000 × g for 10 min. Resuspend and homogenize in resuspension buffer containing 8 M urea, and incubate with agitation for 30 min at room temperature.
4. Add 1 mL of settled Ni-NTA resin to the lysate and incubate at 4 °C for 30–45 min with gentle agitation on a rocker or rotary.
5. Pellet the Ni-NTA resin by centrifugation at 2500 × g for 1 minute. Discard the supernatant.
6. Transfer the Ni-NTA resin to a column.
7. Wash the resin first with 20 mL PBS + 2 M NaCl, then 15 mL of PBS + 2 M NaCl + 20 mM imidazole and finally elute with 4 mL PBS + 2 M NaCl + 500 mM imidazole.
8. Dialyze the protein against 1 L of PBS or an appropriate buffer at 4 °C for one hour, replace with 2 L of fresh buffer, and dialyze overnight. Concentrate as necessary by ultrafiltration.

Upon desalting or dialysis into buffers with lower salt concentration, a small amount of precipitate may form, giving the solution a cloudy appearance. This is likely due to complexation of the supercharged protein with co-purifying charged cell components such as phospholipids and nucleic acids. Pellet these precipitates to recover the remaining soluble fraction. Most supercharged proteins and their fusions in our experience are soluble in PBS. Certain fusion proteins, including fusions to Cre recombinase, may require additional salt (up to 500 mM) or other additives to promote protein solubility and stability.

The protein preparation may either be used as is, or subjected it to further purification steps including ion exchange chromatography and endotoxin removal (described below). Quantitate proteins either spectrophotometrically or via bicinchoninic acid protein assay. Protein expression and size should be confirmed by PAGE and/or by MALDI-TOF mass spectrometry. Following quantitation of protein, multiple aliquots of should be prepared at

volumes sufficient for single sets of assays to minimize freeze-thaw cycles. Aliquots should be snap frozen in liquid nitrogen and stored at -80°C .

Cation Exchange And Endotoxin Removal

Co-purified anionic contaminants may still be present following Ni-NTA-based purification. The presence of anionic contaminants may alter the interaction of supercharged proteins with desired anionic molecules in downstream applications. For example, high levels of endotoxin (>0.5 EU/mL) can alter phenotypes in cellular assays. As such, we recommend further purification steps for most applications, though the Ni-NTA purified material may be sufficient for some experiments. The following is a representative FPLC protocol for purifying superpositive proteins.

- 1) The cation exchange column is first washed with deionized water and equilibrated with five column volumes of PBS. We use an Akta FPLC and a 1 mL HiTrap Capto SP XL column.
- 2) Dialyze or desalt protein into PBS prior to injection onto FPLC cation exchange column. Many of the supercharged protein-Cre fusion proteins are not soluble in PBS and were dialyzed against PBS + 0.5 M NaCl. To cation exchange these proteins, they were diluted 1:4 into PBS immediately prior to injection onto the column.
- 2) After injection, five column volumes are passed through the column to remove contaminant proteins and degradation products.
- 3) We recommend an elution protocol utilizing a linear gradient from PBS to PBS + 1 M NaCl over 10 column volumes.
- 4) Following completion of the elution cycle, pool fractions, concentrate the sample to ~ 1.5 mg/mL and dialyze into storage buffers as needed. To avoid introducing molecules that could influence cell penetration or delivery, we dialyze proteins against PBS for storage except for Cre proteins which are dialyzed against PBS + 0.5 M NaCl.

We have found that Ni-NTA purification, cation exchange and anion exchange often do not fully remove endotoxin from protein samples. If the protein is to be used in endotoxin-sensitive applications, we use an endotoxin removal column and often repeat this process until endotoxin levels are acceptably low (e.g., <0.5 EU/mL).

1. Endotoxin levels are measured using an endpoint chromogenic limulus amoebocyte lysate endotoxin assay kit (Lonza). Standard endotoxin is diluted to create a standard curve from 0 to 1 EU/mL according to the manufacturer's protocol. Protein is known to inhibit the endotoxin reaction and we have found that +36 GFP concentrations greater than $1.5\ \mu\text{M}$ are inhibitory.
2. The protein sample is then diluted in endotoxin free water at non-inhibitory concentrations such as 0, 0.25, 0.5, and 1 % of the endotoxin sample.
3. To remove endotoxin, we use a polymyxin B column (Pierce catalog number 20344) according to the manufacturer's instructions. We equilibrate the column with five volumes of freshly prepared PBS.
4. Protein is eluted with five additional volumes of PBS, concentrated and retested for endotoxin levels. We have found some protein preparations to initially contain >30 EU/mL endotoxin. The polymyxin B columns can remove as much as $>90\%$ of endotoxin but may need to be repeated to lower endotoxin concentrations to acceptable levels.

Construction of Fusion Proteins

Unfused supercharged proteins are typically purified by affinity chromatography with either N- or C-terminal His₆-tagging. For fusion of the supercharged protein to other protein cargoes, the orientation of the fusion partners and affinity tags can be optimized to maximize the yield of the full-length fusion. We have tested and optimized the constructs in multiple cases, and have arrived at a general-purpose starting architecture shown in Fig. 8.

We recommend that the supercharged protein domain be located at the N-terminus for initial attempts. We have found that this orientation generally results in higher expression levels and provides advantages in purification described below. A long flexible glycine-serine linker, (GGS)₉, ensures sufficient spacing between folded domains and stable linkage of the fusion partners. The fusion partner of interest is located at the C-terminus, followed by a C-terminal His₆ tag. This orientation allows for highly efficient isolation of full-length fusion protein without additional size-exclusion or gel-filtration chromatography steps. The fusion is first pulled down by the C-terminal tag using Ni-NTA and then further purified via cation exchange to isolate proteins possessing an intact N-terminal supercharged protein domain. This construction and purification strategy ensures that both ends of the fusion are present in the final collected material.

Nucleic Acid Delivery by Supercharged Proteins *In Vitro*

The non-viral delivery of siRNA and plasmid DNA into mammalian cells are valuable both for research and therapeutic applications⁹. Purified +36 GFP protein (or other superpositively charged protein) is mixed with siRNAs in the appropriate serum-free media and allowed to complex prior addition to cells. Inclusion of serum at this stage inhibits formation of the supercharged protein-siRNA complexes and reduces the effectiveness of the treatment. The following protocol has been found to be effective for a variety of cell lines.²¹ However, pilot experiments varying the dose of protein and siRNA should be performed to optimize the procedure for specific cell lines.

1. One day before treatment, plate 1×10^5 cells per well in a 48-well plate.
2. On the day of treatment, dilute purified +36GFP protein in serum-free media to a final concentration 200 nM. Add siRNA to a final concentration of 50 nM. Vortex to mix and incubate at room temperature for 10 min.
3. During incubation, aspirate media from cells and wash once with PBS.
4. Following incubation of +36 GFP and siRNA, add the protein-siRNA complexes to cells.
5. Incubate cells with complexes at 37 °C for 4 hours.
6. Following incubation, aspirate the media and wash three times with 20 U/mL heparin PBS. Incubate cells with serum-containing media for a further 48 hours or longer depending upon the assay for knockdown.
7. Analyze cells by immunoblot, qPCR, phenotypic assay, or other appropriate method.

We have further found +36 GFP to be an effective plasmid delivery reagent in a range of cells. As plasmid DNA is a larger cargo than siRNA, proportionately more +36 GFP protein is required to effectively complex plasmids. For effective plasmid delivery we have developed a variant of +36 GFP bearing a C-terminal HA2 peptide tag, a known endosome-disrupting peptide derived from the influenza virus hemagglutinin protein. The following protocol has been effective in a variety of cells, but as above it is advised that plasmid DNA and supercharged protein doses be optimized for specific cell lines and delivery applications.

1. One day before treatment, plate 1×10^5 per well in a 48-well plate.
2. On the day of treatment, dilute purified +36 GFP protein in serum-free media to a final concentration 2 μ M. Add 1 μ g of plasmid DNA. Vortex to mix and incubate at room temperature for 10 min.
3. During incubation, aspirate media from cells and wash once with PBS.
4. Following incubation of +36 GFP and plasmid DNA, gently add the protein-DNA complexes to cells.
5. Incubate cells with complexes at 37 °C for 4 hours.
6. Following incubation, aspirate the media and wash with PBS. Incubate cells in serum containing media and incubate for a further 24 to 48 hours.
7. Analyze plasmid delivery (for example, by plasmid-driven gene expression) as appropriate.

Protein Delivery by Supercharged Proteins *In Vitro*

Protein delivery is a valuable strategy for experimental biological research and therapeutics. The direct delivery of proteins without exogenous nucleic acids minimizes the risk of oncogenic mutation or undesired genome alteration. Furthermore, protein delivery is particularly well suited for applications in which prolonged activity of the delivered molecule is not necessary or potentially harmful. Finally, the activation of innate antiviral immune responses to purified proteins is potentially much lower compared to nucleic acid-based reagents, whose presence extracellularly and within the endocytic pathway is known to induce strong antiviral responses.³⁹ Here we describe a representative protocol for the delivery of proteins fused to supercharged proteins that has been effective across several cell lines:

1. One day before treatment, plate 1×10^5 per well in a 48-well plate.
2. On the day of treatment, dilute purified +36 GFP fusion protein into serum-free media.
 - a. If delivering fluorescent reagents such as +36 GFP-mCherry to monitor uptake, significant levels of uptake are observed dose-dependently in treatments ranging from 25 nM to 2 μ M. Internalization can be visualized within 15 min of treatment, rapidly increasing over the course of two hours before reaching a plateau after 4 hours of incubation. These parameters should be varied as needed for the assay of interest.
 - b. If using ubiquitin-+36GFP as a means to monitor cytosolic localization of internalized proteins, cells should be treated with no more than 100 nM protein in serum-free media for 1 hour. The use of higher doses of protein in this assay is not recommended, as removal of extracellular proteins by washing can be increasingly incomplete or variable, confounding downstream analysis. Treatment with a non-cleavable ubiquitin-+36GFP, such as the G76V ubiquitin mutant protein, is recommended as a control for non-specific cleavage by extracellular proteases prior to internalization and within endosomes post-internalization.
 - c. If delivering +36 GFP-Cre protein (or other functional protein cargo of interest), we recommend treatment with ~10 nM to 2 μ M. Incubate cells with proteins for at least 4 hours prior to removal and further incubation in serum-containing media.

3. If desired, treatment with endosome-disrupting reagents such as chloroquine can be included during the incubation with proteins. Cell-line specific optimization of these agents should be performed, as they are generally toxic. For chloroquine, we recommend a starting dose of ~100 μ M chloroquine during the protein treatment, with an optional continued treatment for up to 12 hours following removal of the protein-containing media. Optimization of dose and duration of endosome disruption treatment should be performed to maximize protein response while minimizing cytotoxicity.
4. Following incubation with protein, wash cells at least three times with PBS containing 20 U/mL heparin to remove cell surface-bound protein. This washing is required for measurement of internalized protein signal, as any remaining extracellular protein can confound interpretation of results.
5. Analyze cells as appropriate.
 - a. For fluorescent protein fusion uptake assays: immediately trypsinize cells in wells to detach from plate. This treatment also works to further remove any remaining uninternalized protein signal, as surface proteins are extensively cleaved during trypsin treatment. Following quenching of trypsin with serum-containing media, we recommend analysis by flow cytometry to determine the extent of protein internalization between treatments.
 - b. For ubiquitin fusion analysis: immediately following incubation and washing, lyse cells on ice directly in wells with cold LDS sample buffer containing 1 mM PMSF. Scrape cells to ensure consistent and complete removal. Analyze by western blot to observe the fraction of ubiquitin-+36GFP fusion protein that has specifically cleaved by cytosolic deubiquitinating enzymes. The cleaved products will run ~8 kDa smaller on a 12% SDS-PAGE gel and can be readily quantified by standard densitometry methods.
 - c. For Cre protein fusion delivery analysis: incubate cells for 24 to 48 hours post-treatment to allow reporter signal to maximize. Cells should be analyzed for reporter protein expression as appropriate. We have previously used a fluorescent Cre reporter, floxed tdTomato BSR cells, which carry a genomically integrated floxed STOP cassette followed by the tdTomato fluorescent protein. Following successful Cre recombination, the STOP is deleted to allow high expression of tdTomato from the upstream CAG enhancer and CMV promoter. To analyze recombination driven by +36 GFP-delivered Cre protein, trypsinize cells and analyze by flow cytometry. Recombinant cells will display strong fluorescence at 581 nm corresponding to expression of the tdTomato protein. Treatment with 1 μ M to 2 μ M +36 GFP-Cre as described above will yield 20–30% recombination in this cell line, with levels as high as 70% being achieved upon co-treatment with 100 μ M chloroquine at doses as low as 200 nM +36 GFP-Cre.

Protein Delivery by Supercharged Proteins *In Vivo*

Supercharged proteins are also able to deliver protein *in vivo*. The applications for protein delivery are constrained by the physical properties of the cargo protein and the supercharged protein fusion. Injection of Cre fused to a supercharged protein, for example, leads to functional delivery only near the injection site as most of the protein fusion precipitates, likely as a result of both the lower salt concentrations in plasma, and coprecipitation with

serum proteins and other plasma constituents. This localization of functional protein delivery can be advantageous for certain applications such as protein delivery specifically to cells of the retina, pancreas or white adipose tissue.²⁶ Before performing experiments with animals, consult your local Experimental Review Board or Institutional Animal Care and Use Committee to receive approval for your experimental protocol.

The following is a standard protocol for injection of protein into the subretinal space of mice⁴⁰:

1. Anesthetize adult mice and keep on a 37 °C pad during and after surgery.
2. Expose the eyeball pulling down the skin around the eye. Make a small incision in the sclera near the lens using a 30-gauge needle.
3. Insert an injection needle such as a Hamilton syringe with a 32- or 33-gauge blunt-ended needle into the eyeball through the incision until you feel resistance.
4. Inject 1 µL of protein sample.
5. For visualization of protein delivery with a β-galactosidase reporter, harvest retinae three days post-injection, fix with 0.5% glutaraldehyde for 30 min, stain with X-Gal overnight, and image.
6. For sectioning, embed retinae in 50% OCT, 50% of 30% sucrose in PBS and store at −80 °C.
7. Cut retinae into 30 µm sections and image for X-Gal on a brightfield microscope.

The following is a standard protocol for pancreatic injections:

1. Spike in adenovirus GFP (1.5×10^8 pfu) or an injection dye to each 100 µL of protein samples that are not fluorescent to identify the sections of the pancreas that have been injected.
2. Anesthetize adult mice and keep on a 37 °C pad during and after surgery.
3. Shear hair from the abdomen and sterilize three times with betadine and ethanol.
4. Make an incision in the abdominal wall, identify and expose the pancreas. Using a 27-gauge needle, directly inject 100 µL of the protein sample into 2–3 foci of the dorsal lobe with a 3/10cc insulin syringe.
5. Replace the pancreas into the abdominal cavity and close the incision with a suture and surgical staple. Provide mice with 48 hours of analgesia to facilitate recovery.
6. After five days, remove the pancreases and isolate the GFP⁺ (or other delivery reporter-positive) regions of the pancreata with a fluorescent dissecting microscope.
7. For visualization of protein delivery, for example with a β-galactosidase reporter, fix the pancreas tissues in 20 mLs of 4% paraformaldehyde by rocking at 4 °C for 1.5 hrs, and wash three times with 20 mLs of PBS by rocking at 4 °C for 5 min. Equilibrate the pancreas in 25 mLs of 30% sucrose PBS by rocking at 4 °C overnight. Place the pancreas into a mold and cover with OCT. Incubate at room temperature for 30 min, freeze on dry ice and store at −80 °C. Cut the frozen molds into 12 µm sections. Immunostain the sections with anti-beta galactosidase antibody, and image with a fluorescence microscope.
8. For quantification of protein delivery with a β-galactosidase reporter, suspend the pancreas in 200 µL of β-galactosidase lysis buffer (2.5 mM EDTA, 0.25% NP-40, 250 mM Tris pH.4) plus one Complete Protease Inhibitor tablet (Roche) per 10 mL.

Homogenize pancreas tissues by electric pestle (2×15 s) in a 1.5 mL tube and incubate on a rotating drum at 4 °C for 30 min. Centrifuge lysates at 13,000 *G* for 5 min and use 20 μ L of supernatant for the β -galactosidase assay (Stratagene). Incubate the β -galactosidase assay reactions at 37 °C for ~30 min and assay for absorbance at 575 nm. The β -galactosidase assay can be normalized to a standard curve using commercially available β -galactosidase enzyme (Abcam). The pancreas supernatants (5 μ L) can also be added to a bicinchoninic acid protein assay (Pierce) to quantify total protein concentration against a BSA standard.

The following is a standard protocol for subcutaneous injection near white adipose tissue:

- 1) Adult mice (anesthesia optional) are injected subcutaneously above the pelvic bone on either side of the mouse abdomen with 100 μ L of protein sample. A 3/10cc insulin syringe is inserted underneath the mouse skin and prior to injection the tip is angled to confirm that the syringe is underneath the skin but has not penetrated the peritoneum.
- 3) For visualization of protein delivery, for example with a luciferase reporter, after three days, inject the mice with 400 μ L of 7.5 μ g/mL luciferin. Sacrifice the mice after five min mice, extract white adipose tissue, and image for luminescence using the IVIS molecular imaging system (Caliper).
- 4) For quantification, suspend the white adipose tissue in 2 mL of 0.1% Triton in PBS plus one Complete protease inhibitor tablet (Roche) per 10 mL buffer. Homogenize the white adipose tissues using an Omni tissue homogenizer (2×30 s) in a 15 mL tube then incubated on a rotating drum at 4 °C for 30 min. Lysates were centrifuged at 13,000 *G* for 5 min and 40 μ L of supernatant was added to 200 μ L of the luciferase reaction buffer (Stratagene). The white adipose tissue supernatant was also added (5 μ L) to a bicinchoninic acid protein assay (Pierce) to quantify total protein concentration against a BSA standard.

Supercharged proteins have been demonstrated to deliver functional protein to the retina, pancreas, and white adipose tissue in mouse *in vivo*. These proteins are likely to be useful for other localized applications such as intramuscular injections or injection to other organs.

CONCLUSION

Supercharged proteins are a class of engineered and naturally existing proteins with highly positive or negative net theoretical charge (typically > 1 net charge unit per kD of molecular weight). Both negatively and positively supercharged proteins display remarkable chemical and biological properties, including resistance to chemically or thermally induced aggregation. Superpositively charged proteins can bind and potently penetrate mammalian cells, and can effectively deliver both nucleic acid and protein cargoes into cells in functional form. Protein delivery by supercharged proteins can also function *in vivo* in multiple tissue types of therapeutic interest, including the retina, the pancreas, and white adipose tissue. The potency of functional delivery effected by supercharged protein reagents can, in some tissues, exceed the current best methods for macromolecule delivery including the Tat peptide and adenoviral vectors. As such, supercharged proteins represent a powerful tool to deliver macromolecules into cells and tissues of interest, in both basic research and prospective therapeutic applications.

References Cited

1. Overington JP, Al-Lazikani B, Hopkins AL. Nat Rev Drug Discov. 2006; 5:993–996. [PubMed: 17139284]

2. Arkin MR, Wells JA. *Nat Rev Drug Discov.* 2004; 3:301–317. [PubMed: 15060526]
3. Jantsch J, et al. *J Immunol Methods.* 2008; 337:71–77. [PubMed: 18514219]
4. Shen ZP, Brayman AA, Chen L, Miao CH. *Gene Ther.* 2008; 15:1147–1155. [PubMed: 18385766]
5. Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. *Nature.* 2008; 455:627–632. [PubMed: 18754011]
6. Durcan N, Murphy C, Cryan SA. *Mol Pharm.* 2008; 5:559–566. [PubMed: 18491918]
7. Akdim F, et al. *Am J Cardiol.* 2010; 105:1413–1419. [PubMed: 20451687]
8. Akinc A, et al. *Nat Biotech.* 2008; 26:561–569.
9. Akinc A, et al. *Mol Ther.* 2010; 18:1357–1364. [PubMed: 20461061]
10. Patil ML, et al. *Biomacromol.* 2009; 10:258–266.
11. Watanabe K, et al. *Mol Biosyst.* 2009; 5:1306–1310. [PubMed: 19823746]
12. Lee SK, Han MS, Asokan S, Tung CH. *Small.* 2011; 7:364–370. [PubMed: 21294265]
13. Kam NW, Liu Z, Dai H. *J Am Chem Soc.* 2005; 127:12492–12493. [PubMed: 16144388]
14. Jo D, et al. *Nat Biotech.* 2001; 19:929–933.
15. Soutschek J, et al. *Nature (London).* 2004; 432:173–178. [PubMed: 15538359]
16. Nishina K, et al. *Mol Ther.* 2008; 16:734–740. [PubMed: 18362929]
17. Lawrence MS, Phillips KJ, Liu DR. *Journal of the American Chemical Society.* 2007; 129:10110–10112. [PubMed: 17665911]
18. Thompson DB, Liu DR.
19. Wadia JS, Stan RV, Dowdy SF. *Nature Medicine.* 2004; 10:310–315.
20. Fuchs SM, Raines RT. *ACS Chem Biol.* 2007; 2:167–170. [PubMed: 17319644]
21. McNaughton BR, Cronican JJ, Thompson DB, Liu DR. *Proceedings of the National Academy of Sciences of the United States of America.* 2009; 106:6111–6116. [PubMed: 19307578]
22. Thompson DB, Liu DR. Submitted.
23. Cronican JJ, et al. *ACS Chemical Biology.* 2010; 5:747–752. [PubMed: 20545362]
24. Veldhoen S, Laufer SD, Trampe A, Restle T. *Nucleic Acids Res.* 2006; 34:6561–6573. [PubMed: 17135188]
25. Harding C, Heuser J, Stahl P. *J Cell Biol.* 1983; 97:329–339. [PubMed: 6309857]
26. Cronican JJ, et al. 2011 Submitted.
27. Gump JM, June RK, Dowdy SF. *J Biol Chem.* 2010; 285:1500–1500. [PubMed: 19858185]
28. Turcotte RF, Lavis LD, Raines RT. *FEBS J.* 2009; 276:3846–3857. [PubMed: 19523116]
29. Dill KA, Ozkan SB, Shell MS, Weikl TR. *Annu Rev Biophys.* 2008; 37:289–316. [PubMed: 18573083]
30. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. *J Mol Biol.* 2007; 369:1318–1332. [PubMed: 17482644]
31. Karshikoff A, Ladenstein R. *Biochem Sci.* 2001; 26:550–556.
32. Stemmer WP, Cramer A, Ha KD, Brennan TM, Heyneker HL. *Gene.* 1995; 164:49–53. [PubMed: 7590320]
33. Zhao H, Arnold FH. *Nucleic Acids Res.* 1997; 25:1307–1308. [PubMed: 9092645]
34. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. *Nat Biotech.* 1998; 16:258–261.
35. Huimin Z, Frances HA. *A.* 1997; 94:7997–8000.
36. Yanisch-Perron C, Vieira J, Messing J. *Gene.* 1985; 33:103–119. [PubMed: 2985470]
37. Karimova G, Pidoux J, Ullmann A, Ladant D. *Proc Natl Acad Sci.* 1998; 95:5752–5756. [PubMed: 9576956]
38. Jones S, Holm T, Mäger I, Langel U, Howl J. *CHEM BIOL.* 2010; 17:735–744. [PubMed: 20659686]
39. Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C. *Science.* 2004; 303:1529–1531. [PubMed: 14976261]
40. Matsuda T, Cepko CL. *Methods Mol Biol.* 2008; 423:259–278. [PubMed: 18370205]

Figure 1. Electrostatic surface potentials of -30 GFP, stGFP, +36 GFP, and +48 GFP colored from -25 kT/e (red) to +25 kT/e (blue).

Figure 2. UV-illuminated samples of purified GFP variants (“native”), those samples heated 1 min at 100 °C (“boiled”), and those samples subsequently cooled for 2 h at 25 °C (“cooled”).

Figure 3.

UV-illuminated samples of His₃₉ GFP (“native”) at different pH values and after those samples were heated 1 min at 100 °C (“boiled”), and subsequently cooled for 2 h at 25 °C (“cooled”).

Figure 4.

Properties of supercharged GFP variants. (A) The charge-dependence of supercharged GFP uptake in cultured HeLa cells treated with 200 nM protein for 4 hours at 37 °C. (B) The excitation and emission spectra of blue, cyan, green and yellow fluorescent variants of +36 GFP. The yellow variant is notable, as it has a large stokes shift with a 400 nm absorption, and a 520 nm emission maxima.

Figure 5. His₃₉ GFP penetrates mammalian cells in a pH-dependent manner consistent with the protonation state of its histidine side chains.

Figure 6. Natural supercharged human proteins (NSHPs) deliver active proteins *in vitro* and *in vivo*. (A) Percent recombined cells among floxed tdTomato BSR cells incubated with NSHP-Cre fusions as measured by flow cytometry. (B) Adult floxed LacZ mice were injected subretinally with Cre fusion proteins. Recombination results in LacZ activity, which was visualized with X-gal stain (blue) 3 days after injection. (D) Adult floxed LacZ mice injected in the pancreas with Cre fusion proteins exhibit recombination in the exocrine tissues as indicated by LacZ immunostaining (red) 5 days after injection. (E) Adult floxed luciferase mice injected subcutaneously with Cre fusion proteins exhibit recombination in the white adipose tissue as visualized by luminescence 3 days after injection. White adipose tissue was extracted and placed to the right of each mouse.

Figure 7. Plot of human proteins expressed from *E. coli* within the Protein Data Bank. The blue dots represent proteins with positive charge:molecular weight ratios exceeding +0.75/kD.

Figure 8. Optimized supercharged protein fusion architecture for expression, purification and protein delivery.