Physico-chemical determinants of soluble intrabody expression in mammalian cell cytoplasm

Erik Kvaam1,2, Michael R. Sierks3, Charles B. Shoemaker4 and Anne Messer1,2,5

1New York State Department of Health, Wadsworth Center, David Axelrod Institute, 120 New Scotland Ave., PO Box 22002, Albany, NY 12201-2002, USA, 2Department of Biomedical Sciences, University at Albany, Albany, NY, USA, 3Department of Chemical Engineering, Arizona State University, Tempe, AZ, USA and 5Department of Biomedical Sciences, Tufts Cummings School of Veterinary Medicine, North Grafton, MA, USA

Received November 12, 2009; revised March 2, 2010; accepted March 12, 2010

Edited by Laurent Jespers

Soluble antibody fragments are desirable not only as potential therapeutic and diagnostic agents for extracellular targets but also as ‘intrabodies’ for functional genomics, proteomics and gene therapy inside cells. However, antibody fragments are notoriously aggregation-prone when expressed intracellularly, due in part to unfavorable redox potential and macromolecular crowding in cell cytoplasm. Only a small proportion of intrabodies are soluble in cytoplasm and little is known about the sequence determinants that confer such stability. By comparing the cytoplasmic expression of several related human single-chain variable fragments and camelid VHHs in mammalian cells, we report that intrabody solubility is highly influenced by CDR content and is improved by an overall negative charge at cytoplasmic pH and reduced hydrophilicity. We hypothesize that ionic repulsion and weak hydrophobic interactions compensate, to different extents, for impaired disulfide bond formation in cytoplasm, thereby decreasing the risk for intrabody aggregation. As proof of principle, we demonstrate that the soluble expression of an aggregation-prone positively charged intrabody is modestly enhanced via cis or trans acidification using highly charged peptide tags (3XFLAG tag, SV40 NLS). These findings suggest that simple sequence analysis and electrostatic manipulation may aid in predicting and engineering solubility-enhanced intrabodies from antibody libraries for intracellular use.

Keywords: intrabody/intracellular antibody/protein aggregation/scFv/VHH

Introduction

Recent advances in antibody engineering have fostered the development of recombinant antibody fragments that exhibit selective antigen specificities typical of conventional immunoglobulins but consist of single polypeptides. These recombinant molecules have been assembled into large non-immune libraries to rapidly screen for antibody fragments that bind antigen(s) with high affinity for drug-discovery purposes (reviewed in Hoogenboom, 2005). The smallest fragment capable of binding an antigen with favorable affinity is a single variable domain derived from antibody heavy- or light-chain (VH or VL). However, since VH and VL domains cooperatively assemble in immunoglobulin folds, single-domain antibody fragments tend to be unstable and aggregation-prone in physiological environments (Dudgeon et al., 2009). Two notable exceptions to this trend are highly stable VHH fragments derived from naturally occurring heavy-chain immunoglobulins that are devoid of light chains (Arbabi Ghahroudi et al., 1997), and engineered VH or VL domains in which protein stability has been artificially improved via molecular evolution or targeted amino-acid replacement (Davies and Riechmann, 1996; Colby et al., 2004a; Christ et al., 2007; Barthelemy et al., 2008). A preferred method for overcoming variable domain instability is to covalently fuse VH and VL domains together using a recombinant flexible linker typically composed of serine and glycine residues, thus forming a single-chain variable fragment (scFv) that recapitulates the antigen-binding site of a conventional immunoglobulin (Bird et al., 1988; Huston et al., 1988). However, as is the case for single-domain antibodies, the stability of scFvs can be poor, due in part to the fact that these recombinant domains are removed from their natural context within immunoglobulin folds, as well as the empirical observation that certain VH:VL domain combinations possess suboptimal folding and stability properties (Ewert et al., 2003).

Genes encoding VH and VL domains contain leader sequences that encode signal peptides, which direct antibody fragments to the lumen of the endoplasmic reticulum where an oxidative environment supports the formation of intradomain and interchain disulfide bonds that are critical for antibody assembly and stability prior to secretion from the cell. Removing the leader sequences of VH and VL genes through recombinant DNA techniques allows for the cytoplasmic expression of intracellular antibodies or ‘intrabodies’, which hold great promise as highly specific intracellular reagents for functional genomics, proteomics and gene therapy (reviewed in Lo et al., 2008; Messer et al., 2009). However, the interior of the cell poses significant challenges to intrabody folding, structure and function. First, the crowded nature of the cytoplasm promotes protein oligomerization and aggregation as a consequence of the numerous molecular interactions that occur within this compressed environment (reviewed in Ellis, 2001). Additionally, the redox environment of the cytoplasm inhibits intradomain disulfide formation, which normally contributes ~4–6 kcal/mol to the stability of antibody domains (Frisch et al., 1996; Worn and Pluckthun, 1998). Thus, although a few stable intrabodies have been identified through selective screening (Visintin et al., 2002; Tanaka et al., 2003; Auf der Maur et al., 2004) or have been artificially engineered to fold properly in intracellular environments (Proba et al., 1998; Philibert et al., 2007), most scFvs and single-domain antibody...
fragments are highly destabilized inside the cell, resulting in poor folding, degradation and low yield. These problems are further exacerbated by the degree at which intrabodies are ectopically expressed, as overexpression can aggravate the innate instability of these recombinant proteins and lead to rapid and precipitous formation of detergent-insoluble aggregates (Cattaneo and Biocca, 1999; Ohage and Steipe, 1999). Such intracellular aggregation not only reduces the yield of natively folded intrabody from cells but also can disrupt normal cellular physiology and trigger cell death (Sibler et al., 2003), the latter being especially relevant to conditions of proteotoxic stress such as protein conformational disorders (Gidalevitz et al., 2006).

As a consequence of the interdependence between antibody structure and function, there exists an excellent correlation between intrabody stability and efficacy inside cells (Zhu et al., 1999; Worn et al., 2000; Rajpal and Turi, 2001). Thus, maximizing intrabody stability and improving solubility are critical goals for downstream applications in basic and clinical research. Efforts to engineer intrabody stability or solubility through molecular evolution (Fisher and DeLisa, 2009), random or targeted amino-acid replacement (Martineau et al., 1998; Aires da Silva et al., 2004) or as protein fusions (Strube and Chen, 2004; Shaki-Loewenstein et al., 2005) have met with success but have often neglected to elucidate underlying molecular reasons for why one intrabody may be more stable in cell cytoplasm than another. Here, we analyze the soluble expression of several related human scFv and camelid VHH intrabodies in mammalian cell cytoplasm to identify simple physico-chemical determinants that correlate with enhanced intracellular solubility. We report that soluble expression is highly influenced by CDR content as well as by the overall charge and hydrophaticity of the intrabody sequence. These physico-chemical parameters are readily modified by cloning short, highly charged peptide tags onto intrabodies that effect intracellular aggregation, thus providing a possible mechanism for how charged protein fusion partners reported in the literature may improve intrabody solubility in cell cytoplasm. We conclude that simple sequence analysis and non-invasive electrostatic manipulation may aid in predicting and engineering solubility-enhanced intrabodies from antibody libraries for intracellular use.

Materials and methods

Expression plasmids

Fluorescently labeled intrabodies were assembled as described previously (Kvam et al., 2009) using a universal EGFP-tagging vector that was constructed by PCR amplification of humanized enhanced green fluorescent protein (EGFP), lacking a start codon, with a forward primer that introduced a (Gly4Ser)4 linker. cDNA for scFv or VHH intrabodies (lacking 5’ leader sequences or 3’ hinge regions) were PCR amplified, without stop codons, using reverse primers that introduced an HA epitope tag (YPYDVPDYA) and cloned upstream and in-frame to (Gly4Ser)4-EGFP in pcDNA3.1(−) (Invitrogen), resulting in the following en face orientation: Kozak sequence-start-intrabody-HA-(Gly4Ser)4-EGFP-stop. Amino-acid changes in recoded C4 intrabody (rcC4) were introduced by re-synthesizing cDNA with the desired changes (GeneArt). The HA tag on scFv-D5 was replaced via re-amplification of intrabody cDNA using a reverse primer that introduced a 3XFLAG epitope tag ([DYKDDDK]3), followed by re-ligation upstream of (Gly4Ser)4-EGFP in pcDNA3.1(−) at identical restriction sites. Similarly, canonical (TPPKKKRKV) or inverted (VKRKKKPKT) SV40 nuclear localization sequences were cloned onto the 3’ end of scFv-D5 and rcC4 via re-amplification of HA-tagged intrabody cDNA using corresponding reverse primers that introduced these sequences, and then re-ligated upstream of (Gly4Ser)4-EGFP in pcDNA3.1(−) at identical restriction sites. Expression vectors for nuclear localization signal (NLS)-mRFP, a live-cell fluorescent nuclear marker, httex1-25Q-mRFP and httex1-72Q-EGFP were described previously (Kvam et al., 2009). All expression plasmids were confirmed by DNA sequencing.

Cell culture and transfection

Undifferentiated ST14A cells derived from embryonic day 14 rat striatal primordia were propagated at 33°C essentially as described (Ehrlich et al., 2001). Endotoxin-free plasmid DNA was isolated with an EndoFree Plasmid Maxi kit (Qiagen) and transiently transfected into ST14A cells using jetPEI DNA transfection reagent (Polyplus-transfection Inc.) at an N/P ratio of 8. Total amounts of transfected plasmid were reduced to 60% of the manufacturer’s recommended amount, and culture medium was replaced ~3 h post-transfection to minimize cellular toxicity. For all transfections, intrabody plasmids were applied at an equal ratio (1:1) with empty vector, and cells were analyzed 48 h post-transfection.

Live-cell imaging and aggregation counts

ST14A cells expressing fluorescently labeled reporters were imaged directly in 6- or 12-well culture dishes using an Olympus IX70 inverted microscope equipped with an Olympus IX-FLA Inverted Reflected Light Fluorescence Observation attachment and an RGB Mirror Cube filter wheel (Olympus). To maximize the detection of the fluorescent signal, the culture medium was replaced with sterile PBS immediately prior to imaging, and cells were observed, without fixation, using a ×40 lens. Images were captured 48 h post-transfection with a SPOT RT Color CCD camera and SPOT Advanced software (Diagnostic Instruments). Digital images were overlayed and cropped using Adobe Photoshop. Intrabody aggregation propensity was scored in live ST14A cells ~48 h post-transfection by counting the number of cells containing small or large fluorescent aggregates relative to cells exhibiting completely diffuse signal among 200–400 cells/well in at least three independent transfections. These counts were averaged and expressed as percentages by dividing the number of cells containing intrabody aggregates by the total number of cells scored (aggregate plus diffuse cells). Intracellular huntingtin aggregates were scored in a similar manner as described previously (Kvam et al., 2009) using intrabodies encoded by the pAAV-MCS expression vector (Stratagene).

Cell fractionation and western blotting

Approximately 48 h post-transfection, cell medium was harvested and combined with detached cells from 6-well culture dishes. Cells were pelleted and washed twice with PBS, and whole-cell lysates were collected at 4°C in 50 μl/well of RIPA lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% NP40, 0.25% sodium deoxycholate, 0.1% SDS)
supplemented with 1× Complete Protease Inhibitor Cocktail (Roche). Lysates were cleared by microcentrifugation (16,000g, 10 min at 4°C) to pellet insoluble material, and soluble protein content was quantified using Bio-Rad Dc Protein Assay reagents (Bio-Rad Laboratories). Soluble lysates were diluted to 2 μg/μl in 2× denaturing sample buffer (125 mM Tris pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.02% bromophenol blue) and boiled. Approximately 20 μg of soluble cell lysate (10 μl) was resolved by SDS-PAGE and transferred onto PVDF membranes (PerkinElmer). Intrabodies were probed with monoclonal α-HA (1:1000, Covance) or polyclonal α-GFP (1:2000, AbCam) Abs, whereas monoclonal α-actin (1:1500, Sigma) Ab was used to probe endogenous actin as a loading control. Blots were immunolabeled with HRP-conjugated goat-anti-mouse or goat-anti-rabbit Abs (1:2000, Santa Cruz) control. Blots were immunolabeled with HRP-conjugated goat-anti-mouse or goat-anti-rabbit Abs (1:2000, Santa Cruz) where appropriate and were detected by ECL (PerkinElmer). Western blot signals were quantified using ImageJ software (http://rsb.info.nih.gov/ij/), and statistical analysis was performed by one-way ANOVA with StatView software.

**Sequence analysis**

Intrabody amino-acid sequences were aligned with ClustalX software (www.clustal.org) and numbered according to the Kabat annotation using the Abnum tool (Abhinandan and Martin, 2008). Germline domain families and CDR residues were identified using the IMGT/V-QUEST tool (Brochet et al., 2008) in the international ImMunoGeneTics (IMGT) database (www.imgt.org). Isoelectric point and net charge at cytoplasmic pH (7.4) were inferred from the amino-acid sequence of each intrabody (including peptide tags) using Protein Calculator v3.3, developed by Chris Putnam at the Scripps Research Institute (www.scripps.edu/~cdputnam/protcalc). Grand average of hydropathicity (GRAVY) and aliphatic index (AI) were calculated from intrabody amino-acid sequences (including peptide tags) using the ProtParam tool at the ExPASy proteomics server of the Swiss Institute of Bioinformatics (Gasteiger et al., 2003).

**Results**

**Identification of solubility-associated intrabody sequence determinants from a semi-synthetic scFv human scaffold**

We previously characterized a conformation-specific human intrabody (scFv-6E) that is largely soluble in cell cytoplasm (Kvam et al., 2009). This scFv intrabody was originally selected from a CDR-modified scFv repertoire (Tomlinson I and J) based on a single human framework composed of the germline heavy-chain genes IGHV3-23/IGHJ2 and IGHJ4/IGH4b and the κ-light-chain genes IGKV1-39/O12 and IGKJ1, which are recombiantly fused together via a linker that encodes a flexible (Gly4Ser)3-pentadecapeptide (Holt et al., 2000; Barkhordarian et al., 2006). Because this semi-synthetic framework is derived from the human VH3 (IGHV3) and Vk1 (IGKV1) germline domain families, which form the basis for scFv intrabody consensus sequences identified by intracellular antibody capture methods (Tse et al., 2002; Visintin et al., 2002), human scFvs selected from the Tomlinson library are theoretically adapted for intracellular expression. As an added advantage, amino-acid diversity within the Tomlinson library is limited to only 18 hypervariable positions that make contact with antigens (Holt et al., 2000), thus rendering sequence comparison relatively straightforward. To rationally identify sequence determinants that influence soluble expression, we compared the cytoplasmic expression of several published and novel scFvs that were previously selected from the Tomlinson library (Table I). Amino-acid sequence alignments confirmed that all seven scFvs analyzed were identical in size and sequence, except for 18 hypervariable residues within CDR loops 2 and 3 in each variable domain, as anticipated (Supplementary Fig. S1). One scFv (scFv-J15) contained an additional mutation within the framework of the variable light chain (L196C→T) likely induced by the polymerase error during library construction (Supplementary Fig. S1). The coding sequence of each scFv (minus a secretion signal) was fused to EGFP (Fig. 1A) as described previously (Kvam et al., 2009) to create fluorescently labeled intrabodies that can be directly analyzed and imaged in live cells via fluorescence microscopy. Equal amounts of plasmid encoding GFP-tagged scFvs under control of the human cytomegalovirus (CMV) immediate-early promoter/enhancer were transiently transfected into a mammalian cell line (ST14A) that, in our experience, possesses favorable characteristics for intrabody screening purposes (see the ‘Discussion’ section). Drawing on the fact that insoluble intrabodies tend to form intracellular aggregates and fractionate as detergent-insoluble material (Cattaneo and Biocca, 1999), we scored the number of fluorescent aggregates visible after 48 h of scFv-GFP expression to assess the cytoplasmic solubility of each scFv. As shown in Fig. 1B, a wide spectrum of intrabody aggregation was observed, despite the fact that these intrabodies shared 92% amino-acid sequence identity. Cell fractionation revealed that the steady-state levels of intrabody in detergent-soluble extract generally correlated inversely with observed intracellular aggregation propensities (Fig. 1B, inset i). However, by pairwise analysis, we noted that some intrabodies (e.g. scFv-D1, A8 and J15) exhibited comparatively weaker fluorescence signals and lower levels on western blot, indicating a shorter half-life in the steady state for reasons

*Table I. Physico-chemical properties of affinity-selected Tomlinson scFv intrabody sequences*  

<table>
<thead>
<tr>
<th>Intrabody</th>
<th>Antigen</th>
<th>Isoelectric point</th>
<th>Net charge at pH 7.4</th>
<th>GRAVY</th>
<th>AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>6E</td>
<td>Amyloid fibrils</td>
<td>6.62</td>
<td>-0.5</td>
<td>-0.298</td>
<td>61.89</td>
</tr>
<tr>
<td>C10</td>
<td>TL-T1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.62</td>
<td>-0.5</td>
<td>-0.290</td>
<td>61.50</td>
</tr>
<tr>
<td>D1</td>
<td>TL-T1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.62</td>
<td>-0.5</td>
<td>-0.285</td>
<td>61.89</td>
</tr>
<tr>
<td>G7</td>
<td>httex1-39Q&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6.62</td>
<td>-0.5</td>
<td>-0.311</td>
<td>61.10</td>
</tr>
<tr>
<td>J15</td>
<td>α-Synuclein oligomers&lt;sup&gt;c&lt;/sup&gt;</td>
<td>7.19</td>
<td>-0.3</td>
<td>-0.335</td>
<td>61.50</td>
</tr>
<tr>
<td>A8</td>
<td>TL-T1&lt;sup&gt;d&lt;/sup&gt;</td>
<td>8.10</td>
<td>+1.5</td>
<td>-0.274</td>
<td>61.50</td>
</tr>
<tr>
<td>D5</td>
<td>Amyloid oligomers&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.62</td>
<td>+3.7</td>
<td>-0.370</td>
<td>60.71</td>
</tr>
</tbody>
</table>

<sup>a</sup>Isoelectric point, net charge, GRAVY and AI were estimated from the amino-acid sequence of each intrabody fused to an HA epitope tag (AAAYPYDVPDYA), as outlined in the ‘Materials and methods’ section.

<sup>b</sup>Edited from a CDR3-randomized sublibrary of scFv-D5 (M. Sierks).

<sup>c</sup>Selected against aggregated httex1-39Q using the method outlined in Barkhordarian et al. (2006).

<sup>d</sup>Amino-acid sequence determinants that influence soluble expression, we compared the cytoplasmic expression of several published and novel scFvs that were previously selected from the Tomlinson library (Table I). Amino-acid sequence alignments confirmed that all seven scFvs analyzed were identical in size and sequence, except for 18 hypervariable residues within CDR loops 2 and 3 in each variable domain, as anticipated (Supplementary Fig. S1). One scFv (scFv-J15) contained an additional mutation within the framework of the variable light chain (L196C→T) likely induced by the polymerase error during library construction (Supplementary Fig. S1). The coding sequence of each scFv (minus a secretion signal) was fused to EGFP (Fig. 1A) as described previously (Kvam et al., 2009) to create fluorescently labeled intrabodies that can be directly analyzed and imaged in live cells via fluorescence microscopy. Equal amounts of plasmid encoding GFP-tagged scFvs under control of the human cytomegalovirus (CMV) immediate-early promoter/enhancer were transiently transfected into a mammalian cell line (ST14A) that, in our experience, possesses favorable characteristics for intrabody screening purposes (see the ‘Discussion’ section). Drawing on the fact that insoluble intrabodies tend to form intracellular aggregates and fractionate as detergent-insoluble material (Cattaneo and Biocca, 1999), we scored the number of fluorescent aggregates visible after 48 h of scFv-GFP expression to assess the cytoplasmic solubility of each scFv. As shown in Fig. 1B, a wide spectrum of intrabody aggregation was observed, despite the fact that these intrabodies shared 92% amino-acid sequence identity. Cell fractionation revealed that the steady-state levels of intrabody in detergent-soluble extract generally correlated inversely with observed intracellular aggregation propensities (Fig. 1B, inset i). However, by pairwise analysis, we noted that some intrabodies (e.g. scFvs-D1, A8 and J15) exhibited comparatively weaker fluorescence signals and lower levels on western blot, indicating a shorter half-life in the steady state for reasons...
Intracellular aggregation and detergent-soluble expression of related human scFv intrabodies generally correlate with predicted intrabody net charge at cytoplasmic pH (7.4) and GRAVY. (A) Diagram of GFP-labeled intrabody cassette for transient expression of fluorescent intrabodies in cultured mammalian cells. Non-coding elements, including the human cytomegalovirus immediate-early promoter/enhancer (P<sub>CMV</sub>), Kozak translation initiation sequence and bovine growth hormone polyadenylation signal (BGH pA), are shaded in gray. Coding elements are in white. Unless otherwise indicated, intrabodies were fused to an influenza hemagglutinin (HA) epitope tag for biochemical detection and linked to human codon-optimized EGFP for fluorescence detection via a flexible (Gly<sub>4</sub>Ser)<sub>4</sub> icosapeptide. (HA) The average percentage of transfected ST14A rat striatal progenitor cells exhibiting intracellular scFv-GFP aggregates 48 h post-transfection. Cells were scored for fluorescent intrabody aggregates among three independent replicates, as described in the ‘Materials and methods’ section. Detergent-soluble scFv-GFP levels from cleared cell lysates were analyzed by western blot (inset i; non, non-transfected cells). Estimates of scFv intrabody net charge and GRAVY were inferred from amino-acid sequence data.

Validation of solubility-associated physico-chemical determinants among cytoplasmic scFv intrabodies in the literature

To extend our analysis, we next conducted a literature search for intrabody studies that specifically measured cytoplasmic scFv solubility using cell fractionation and/or microscopy techniques and included corresponding intrabody sequence data. Net protein charge and average hydropathicity values were calculated for each published intrabody sequence and compared with the qualitative degree of soluble expression reported by the authors. As shown in Table II, we again observed a good correlation between reported soluble expression and estimated intrabody charge and/or hydrophilicity (GRAVY) among these published scFvs. Overall, intrabodies reported to be highly soluble in cell cytoplasm have predicted net negative charges at cytoplasmic pH and weaker hydrophilicity (i.e. less negative GRAVY scores), whereas those reported to be insoluble, aggregate-prone or otherwise detected at low levels in detergent-soluble cell extract have net positive charges and/or stronger hydrophilicity (i.e. more negative GRAVY scores). These findings could explain why, for example, the steady-state levels of two nearly identical human scFv intrabodies (G4E4 and G4G11, 97% sequence identity) selected against the protein-tyrosine kinase Syk were found to differ significantly in a study performed by Dauvillier et al. (2002), in light of our calculations showing that scFv-G4E4 is considerably more acidic (negatively charged) at cytoplasmic pH than scFv-G4G11 (Table II). Importantly, all scFvs in Table II represent diverse germline family pairings that are distinct from Tomlinson scFvs (V<sub>k</sub>3;V<sub>u</sub>1), suggesting that intrabody net charge and hydropathicity generally influence the solubility of many scFv intrabody frameworks. However, we found that net protein charge and GRAVY estimates generally failed to correlate with the reported solubility of a synthetic intrabody framework that has been hyperstabilized by molecular evolution (Philibert et al., 2007) (see the ‘Discussion’ section).
Determinants of intrabody solubility in cytoplasm

It is generally accepted that little correlation exists between intrabody solubility and activity inside cells, as aggregation-prone intrabodies can functionally recruit intracellular antigens (Cardinale et al., 1998; Yi et al., 2001). However, in certain cases where intrabodies are desired for cytoprotective applications, such as blocking the misfolding and aggregation of toxic proteins inside cells, aggregation-prone intrabodies are less favorable therapeutic agents, in part because misfolding-prone proteins, in general, can enhance the aggregation and toxicity of disease-related proteins such as expanded polyglutamine (Gidalevitz et al., 2006). A human intrabody (scFv-C4, Table II) that inhibits the aggregation of mutant huntingtin protein in cell and insect models of Huntington’s disease (Lecerf et al., 2001; Wolfgang et al., 2005; Kvam et al., 2009) offers such an example. Using knowledge-based antibody engineering methods for improving thermostability and folding efficiency in vitro (Ewert et al., 2004), a total of 37 amino-acid changes were incorporated into the framework of scFv-C4 without altering the composition of antigen-binding CDR regions (Supplementary Fig. S2) to generate a recoded version of scFv-C4 (rcC4) with enhanced in vitro stability according to the best predictive and structural modeling methods available at the time (Ewert et al., 2004). However, upon expressing rcC4 as an intrabody in ST14A striatal progenitor cells and assaying for the inhibition of intracellular huntingtin aggregation as a biological read-out of intrabody activity, rcC4 unexpectedly demonstrated lower efficacy in blocking the aggregation of a mutant huntingtin reporter protein (httex1-72Q) compared with our original scFv-C4 intrabody (Fig. 2A).

To investigate the basis for this phenomenon using current methods, we calculated the net charge and hydrophaticy of rcC4 using amino-acid sequence data (Supplementary Fig. S2). Unlike scFv-C4, which is soluble in cell cytoplasm (Table II; net charge −0.5, GRAVY score −0.282), rcC4 is strikingly basic (net charge +1.5) and more hydrophilic (GRAVY score −0.303). These findings suggested that rcC4 may in fact be aggregation-prone in the cytoplasmic environment, despite extensive engineering to improve its in vitro stability. Indeed, live-cell imaging revealed that GFP-labeled rcC4 intrabody is significantly aggregation-prone in cell cytoplasm (Fig. 2B) and was consequently detected at reduced detergent-soluble levels in the steady state compared with scFv-C4 intrabody (Fig. 2B, inset). Importantly, we observed that non-aggregated rcC4 intrabody sequestered a native huntingtin reporter protein (httex1-25Q) as efficiently as our original scFv-C4 intrabody using a classical antibody-antigen re-targeting assay for

![Fig. 2. Correlation of protein net charge and hydropathy with intrabody efficacy in situ](image-url)

### Table II. Physico-chemical properties of published scFv intrabodies reported as showing favorable or poor soluble expression in cytoplasm

<table>
<thead>
<tr>
<th>scFv</th>
<th>Antigen</th>
<th>VH family</th>
<th>VL family</th>
<th>Tag</th>
<th>pI</th>
<th>Net charge at pH 7.4</th>
<th>GRAVY score</th>
<th>Soluble expression in cytoplasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA10</td>
<td>Giantin</td>
<td>VH1</td>
<td>VA1</td>
<td>his6-myc</td>
<td>5.54</td>
<td>−8.6</td>
<td>−0.281</td>
<td>Good</td>
</tr>
<tr>
<td>3C1-A2</td>
<td>mCASP3</td>
<td>VH3</td>
<td>VA1</td>
<td>myc</td>
<td>5.69</td>
<td>−2.4</td>
<td>−0.283</td>
<td>Good</td>
</tr>
<tr>
<td>G4E4</td>
<td>mSyk</td>
<td>VH4</td>
<td>VA1</td>
<td>myc</td>
<td>6.53</td>
<td>−1.3</td>
<td>−0.235</td>
<td>Good</td>
</tr>
<tr>
<td>C4</td>
<td>Huntingtin</td>
<td>VH3</td>
<td>VA2</td>
<td>HA</td>
<td>6.62</td>
<td>−0.5</td>
<td>−0.282</td>
<td>Good</td>
</tr>
<tr>
<td>3C1-B1</td>
<td>mCASP3</td>
<td>VH1</td>
<td>VA3</td>
<td>myc</td>
<td>7.28</td>
<td>−0.2</td>
<td>−0.352</td>
<td>Poor</td>
</tr>
<tr>
<td>F7</td>
<td>PLC-γ1</td>
<td>mMVH1</td>
<td>mVk4</td>
<td>Xpress</td>
<td>7.80</td>
<td>+0.6</td>
<td>−0.549</td>
<td>Poor</td>
</tr>
<tr>
<td>G4G11</td>
<td>mSyk</td>
<td>VH4</td>
<td>VA1</td>
<td>myc</td>
<td>7.75</td>
<td>+0.7</td>
<td>−0.265</td>
<td>Poor</td>
</tr>
<tr>
<td>NT73</td>
<td>RNA pol B’</td>
<td>mVH5</td>
<td>mVk1</td>
<td>E tag-his6</td>
<td>8.43</td>
<td>+3.5</td>
<td>−0.368</td>
<td>Poor</td>
</tr>
</tbody>
</table>

*Human or mouse (m) variable domain germline families and intrabody isoelectric point (pI), net charge and GRAVY were estimated using published intrabody amino-acid sequences including the listed tags, as outlined in the ‘Materials and methods’ section.

1Nizak et al. (2003); GenBank accession number AAQ56715.
2Rajpal and Turi (2001).
3Dauvillier et al. (2002); GenBank accession number AAN02491.
4Kvam et al. (2009); GenBank accession number ACA53373.
5Yi et al. (2001); GenBank accession number AAK56283.
6Dauvillier et al. (2002); GenBank accession number AAN02492.
7Soluble expression score is relative to scFv-G4E4.
8Lammerski et al. (2006); GenBank accession number AAZ85393.
detecting intracellular protein–protein interactions (Sibler et al., 2003), thus indicating that rcC4 intrabody possesses qualitatively similar binding affinity. However, in cells containing visible aggregates of rcC4 intrabody, the subcellular re-targeting of huntingtin was less complete in this assay (Supplementary Fig. S3). Together with our functional analysis of rcC4 in the presence of aggregation-prone mutant huntingtin (Fig. 2A), these findings illustrate that intrabody solubility and efficacy can be linked for certain intracellular applications and suggest that protein charge and hydrophilicity should be considered when engineering antibody fragments in silico for specific intracellular use.

**Influence of net charge and hydrophobicity on the soluble expression of camelid VHH intrabodies**

We next tested whether protein net charge and hydrophobicity also influence the soluble expression of camelid VHH intrabodies in mammalian cell cytoplasm. Camelid VHHs are experimentally attractive single-domain antibody fragments because they have evolved to fold properly in the absence of light chains and are therefore considered to be innately more stable than corresponding human single-domain VHs (Davies and Riechmann, 1996; Muyldermans, 2001). Indeed, intracellular experiments show that camelid VHHs are readily adaptable as soluble cytoplasmic intrabodies (Rothbauer et al., 2006). We obtained several VHHs that were selected by phage display from an immunized alpaca VH library (Maass et al., 2007) against two different serotypes of *Botulinum* neurotoxin light-chain protease domains (BoNT LCs; C.B. Shoemaker, manuscript in preparation). We fused three BoNT LC serotype A-binding VHHs and two BoNT LC serotype B-binding VHHs (Table III) to GFP (Fig. 1A) in order to create VHH-GFP ‘chromobodies’ (Rothbauer et al., 2006) for visualizing intracellular solubility. Sequence alignment revealed that each VHH shared highest homology with the human germline heavy-chain gene products of IGHV3-23/DP47 and IGHJ4/IH4b (Supplementary Fig. S4), which also comprise the VH framework of Tomlinson human scFv, consistent with previous observations that camelid VHHs are close relatives of the human V3 (IGHV3) subfamily (Ewert et al., 2002). In contrast to our Tomlinson scFv data, we noted that the aggregation propensity of each GFP-labeled VHH intrabody (Fig. 3) correlated better with the observed steady-state levels of VHH-GFP protein in detergent-soluble extract, as revealed by western blot (Fig. 3, inset i). Among four VHHs estimated to possess weak net negative charges (less than or equal to –1.3), we observed a direct correlation between greater aggregation propensity and stronger hydrophilicity (i.e. more negative GRAVY scores) (Fig. 3). However, one VHH (H7) bearing a strikingly greater negative charge (–5.0) exhibited the best solubility within this group despite having a high GRAVY score (Fig. 3). This result suggested that a strong net negative charge, alone, may be sufficient to confer greater intrabody solubility in cell cytoplasm. To validate this observation, we analyzed the sequence of a single-domain intrabody (V12.3-his6-HA, GenBank accession AAV87178) that exhibits excellent cytoplasmic solubility and blocks the intracellular aggregation of mutant huntingtin protein in neuronal cells (Colby et al., 2004a,b; Southwell et al., 2008, 2009). Our analysis revealed that this soluble single-domain intrabody carries a high GRAVY score (–0.375) but also a very strong negative charge (–6.5) at cytoplasmic pH. Thus, we conclude that net charge and hydrophobicity contribute unequally to intrabody aggregation propensity, since hydrophobicity correlates with solubility only when intrabody net charge is weakly acidic. Such a trend is also evident in our previous analysis of scFv intrabodies (Fig. 1B, Table II).

---

**Table III. Physico-chemical properties of affinity-selected VHH intrabodies**

<table>
<thead>
<tr>
<th>Intrabody</th>
<th>Antigen</th>
<th>Isoelectric point</th>
<th>Net charge at pH 7.4</th>
<th>GRAVY score</th>
</tr>
</thead>
<tbody>
<tr>
<td>B8</td>
<td>BoNT/A light chain</td>
<td>7.09</td>
<td>–0.3</td>
<td>–0.214</td>
</tr>
<tr>
<td>B10</td>
<td>BoNT/B light chain</td>
<td>6.62</td>
<td>–0.5</td>
<td>–0.231</td>
</tr>
<tr>
<td>C3</td>
<td>BoNT/A light chain</td>
<td>5.97</td>
<td>–1.3</td>
<td>–0.279</td>
</tr>
<tr>
<td>D4</td>
<td>BoNT/A light chain</td>
<td>6.62</td>
<td>–0.5</td>
<td>–0.364</td>
</tr>
<tr>
<td>H7</td>
<td>BoNT/A light chain</td>
<td>5.15</td>
<td>–5.0</td>
<td>–0.395</td>
</tr>
</tbody>
</table>

*VHHs were selected by phage display from an immunized alpaca VH library against two different serotypes (A or B) of BoNT light-chain protease domains. Isoelectric point, net charge and GRAVY were estimated from the amino-acid sequence of each VHH fused to an HA epitope tag (AAAYPDVDPDYA) as outlined in the ‘Materials and methods’ section.*

---

**cis or trans acidification improves soluble expression of an aggregation-prone, positively charged human scFv intrabody**

As a proof of principle for the observed relationship between net charge and intrabody solubility, we next tested whether electrostatic manipulation of an aggregation-prone, positively charged intrabody (scFv-D5) selected from the human scFv Tomlinson library (Fig. 1B) can reduce aggregation and improve soluble intrabody expression in mammalian cell cytoplasm. Toward this goal, we replaced the existing HA epitope tag on scFv-D5 (Fig. 1A) with a 3XFLAG epitope tag [(DYKDDDDK)₃] which is rich in negatively charged aspartic acid residues. Physico-chemical sequence analysis confirmed that this subtle modification nearly inverted the
predicted total charge of scFv-D5, from +3.7 to −3.3, at cytoplasmic pH (Supplementary Table S1). As before, we expressed scFv-D5-3XFLAG as a fluorescent intrabody and assayed for detergent-soluble expression and aggregation propensity by western blotting and live-cell analysis. As shown in Fig. 4A, 3XFLAG-tagged scFv-D5 intrabody exhibited an almost 4-fold reduction in aggregation propensity and a slight improvement in the steady-state soluble level (inset i), when compared with HA-tagged scFv-D5 intrabody. Thus, acidification by a negatively charged 3XFLAG epitope tag is effective in reducing the aggregation propensity of an intrinsically unstable intrabody in cis. Interestingly, the 3XFLAG tag also increased the predicted hydrophilicity (GRAVY score) of scFv-D5, although without increasing aggregation, analogous to the case of camelid VHH-H7 above, which also carries a large net negative charge (Fig. 3). These results validate that net charge is an important factor for controlling the aggregation propensity of an intrabody.

As independent confirmation of the influence of charge on intrabody folding and solubility, we exploited the unique features of NLS peptides, which are typically rich in positively charged lysine and arginine residues. These peptide motifs mediate protein trafficking to the nucleus by binding highly acidic nuclear transport receptors called importins or karyopherins, which exert a chaperone-like function in blocking the ionic aggregation of positively charged NLS domains within the cytoplasm (Jakel et al., 2002). We tested whether an SV40 NLS (TPPPKKRKVP) improved the soluble expression of scFv-D5 in trans by recruiting acidic nuclear transport chaperones, when compared with an inverted NLS (VKRKKPPPT), which is equally as positively charged but defective in recruiting importins for nuclear entry (Lobl et al., 1990; Gorlich et al., 1994). Live-cell imaging confirmed that fluorescent scFv-D5 intrabody tagged with a canonical SV40 NLS localized exclusively to cell nuclei, as indicated by perfect co-localization with the nuclear marker NLS-mRFP, whereas scFv-D5 fused to an inverted and non-functional sequence (rNLS) showed a more diffuse distribution in cells (Fig. 4B). Physico-chemical sequence analysis revealed that both tags increased the estimated charge of scFv-D5 from +3.7 to +8.7 at cytoplasmic pH, suggesting that these intrabodies are equally unstable. However, we observed that the steady-state levels of NLS-tagged scFv-D5 intrabody were significantly higher in the detergent-soluble extract compared with either rNLS-tagged or unmodified intrabody (Fig. 4C). These findings are consistent with the idea that NLS-mediated interactions between scFv-D5 and acidic nuclear transport proteins transiently acidify and enhance the soluble expression of this normally unstable intrabody in trans.

**Discussion**

Recombinant antibody stability is highly unpredictable in physiological environments, a fact which has complicated the use of antibody fragments in research and clinical applications. Such unpredictability is especially pertinent for intrabodies that are expressed and retained inside cells under conditions that foster aggregation. Although methods have been developed to select for soluble intrabodies within cell cytoplasm (Visintin et al., 2002; Contreras-Martinez and DeLisa, 2007), most antibody fragments, especially those adapted from monoclonal antibodies or selected from immune libraries under oxidizing conditions (i.e. phage display), are blindly tested as intrabodies by laboriously expressing these fragments inside cells. Only then is it found that many fragments are unstable, aggregation-prone or
potentially less effective as intrabodies in the reducing cell environment. Here, we show through sequence analyses and expression comparisons that net charge and, to a lesser extent, GRAVY are important determinants of intrabody aggregation propensity in cell cytoplasm and may be used to estimate the feasibility of expressing antibody fragments as soluble cytoplasmic intrabodies. By comparing the aggregation propensity and steady-state levels of several related human scFvs and camelid V\textsubscript{HH}s in mammalian cells, we demonstrate that elevated intrabody acidity (i.e. net negative charge at cytoplasmic pH) and weak hydrophilicity (i.e. less negative GRAVY scores) correlate with greater cytoplasmic solubility, whereas increased basicity (net positive charge) and/or hydrophilicity associate with a greater risk of intracellular aggregation. These observations were validated by comparing the estimates of net charge and hydrophaticity (GRAVY) to the published solubility of several cytoplasmic intrabodies in the literature. Moreover, our findings are supported by \textit{in vitro} results from other groups who show that aggregation-resistant human V\textsubscript{H}s and camelid V\textsubscript{HH}s possess acidic isoelectric points, and thus carry net negative charges at neutral pH (Jespers \textit{et al.}, 2004; Arbabi-Ghahroudi \textit{et al.}, 2009a; Dudgeon \textit{et al.}, 2009).

Net protein charge is a well-documented determinant of protein aggregation inside cells, given that a significant proportion of the cell is occupied by polyanionic (negatively charged) molecules such as RNA. As a consequence, proteins with a net positive charge can easily aggregate in the polyanionic environment of the cytoplasm via multivalent ionic interactions (Jakel \textit{et al.}, 2002). Perhaps for this reason, a majority of cytoplasmic proteins possess net negative charges at cytoplasmic pH, consistent with the idea that the cellular proteome has evolved so as to limit non-specific ionic aggregation (Chan \textit{et al.}, 2006). Although we find that net charge is an important determinant of intrabody solubility, total hydrophilicity (GRAVY) is a secondary correlate of aggregation propensity for weakly acidic intrabodies. Thus, weakly acidic intrabodies possessing reduced hydrophilicity (less negative GRAVY scores) generally exhibited lower aggregation propensities than similarly charged intrabodies having stronger hydrophilicity (more negative GRAVY scores) in this study, suggesting that weak hydrophobic interactions play important roles with regard to intrabody stability in the absence of a strong negative charge. Therefore, we speculate that intrabody folding within the reducing environment of the cytoplasm occurs through a ‘two-hit’ process whereby electrostatic repulsion and hydrophobicity compensate, to different extents, for impaired disulfide bond formation, thereby enabling some intrabodies to fold more natively. Classical \textit{in vitro} approaches may offer additional insights into such intrabody folding states, although these biochemical assays may not exactly replicate the intracellular condition.

Peptide tags are experimentally necessary for discriminating recombinant intrabodies from the cellular cytoplasmic milieu. In this study, we demonstrate that short, highly charged peptide tags can profoundly affect intrabody solubility either in \textit{cis} (3XFLAG tag) or in \textit{trans} (SV40 NLS), likely by altering net intrabody charge. These findings provide a possible explanation for the results of several groups who report improved cytoplasmic scFv solubility upon fusion to highly acidic proteins such as the transcription elongation factor NusA (Zheng \textit{et al.}, 2003) or maltose-binding protein (Bach \textit{et al.}, 2001; Shaki-Loewenstein \textit{et al.}, 2005), both of which are thought to inhibit protein aggregation by electrostatic repulsion (Su \textit{et al.}, 2007; Zou \textit{et al.}, 2008). Likewise, aggregation-prone Ig-like domains, derived from cell surface receptor proteins in the human immunoglobulin superfamily, are fully solubilized when fused to small acidic peptide tags that increase electrostatic repulsion and minimize aggregation (Zhang \textit{et al.}, 2004). Additional methods for antibody ‘isoelectric engineering’, such as adding short stretches of negatively charged glutamic acid residues (Tan \textit{et al.}, 1998), mutating exposed basic surface residues to acidic or neutral residues (Hugo \textit{et al.}, 2002) or replacing hydrophobic residues at the variable/constant domain interface with acidic residues (Nieba \textit{et al.}, 1997), also have proven effective for improving scFv folding and half-life. These prior findings highlight why net charge was the most consistent determinant of intrabody aggregation propensity in the present study. We note that the some of the most common epitope tags used in biological applications (i.e. HA, Myc and FLAG) are highly acidic at cytoplasmic pH, although choosing which tag is appropriate for intrabody acidification is largely dependent on the starting physico-chemical properties of the antibody in question (e.g. see Supplementary Table S1). However, given the numerous ways to design and express recombinant antibody domains, it is possible that certain aggregation-prone intrabodies might prove to be impervious to isoelectric engineering due to yet undefined sequence or structural parameters.

By comparing human scFvs selected from the CDR-modified Tomlinson repertoire (Holt \textit{et al.}, 2000), we additionally demonstrate a striking influence of CDR content on intrabody solubility and aggregation propensity. This observation is substantiated by studies of \textit{in vitro} thermo-stability among CDR-modified human V\textsubscript{H} repertoires and camelid V\textsubscript{HH}s (Bond \textit{et al.}, 2003; Jespers \textit{et al.}, 2004; Arbabi-Ghahroudi \textit{et al.}, 2009b; Dudgeon \textit{et al.}, 2009), as well as by the findings of CDR loop grafting experiments, prompting some to conclude that nature has evolved a multitude of antibody frameworks as a consequence of selective pressures posed by hyervariable CDR residues on overall antibody structure and function (Honegger \textit{et al.}, 2009). An important implication of this study is that codons should be selected carefully when creating CDR-only repertoires so as to limit the frequency of positively charged amino-acid residues that potentially increase the risk for ionic aggregation at physiologic pH. Our aggregation-prone scFv-D5 intrabody, which was originally selected from the Tomlinson J scFv library (Emadi \textit{et al.}, 2007), offers one such example. CDR diversity was introduced into the Tomlinson J library by means of NNK codon degeneracy (where \( N = G, A, T \) or \( C \), and \( K = G \) or \( T \), according to IUPAC nucleotide nomenclature), which codes for all 20 standard amino acids (and an amber stop codon) but is statistically and empirically redundant in arginine residues (Barbas \textit{et al.}, 1992). Consequently, CDR3 of the light chain of scFv-D5 shows a high incidence of arginines (Supplementary Fig. S1), which increases the net positive charge of this intrabody. In contrast, our less aggregation-prone scFv-6E and scFv-C10 intrabodies were selected from the Tomlinson I library (Barkhordarian \textit{et al.}, 2006; Giomarelli \textit{et al.}, 2007), which was diversified by means of DVT codon degeneracy (where \( D = A, G \) or \( T \), and


V = A, C or G nucleotides) that codes for only eight possible amino acids (serine, threonine, tyrosine, glycine, alanine, cysteine, asparagine and aspartic acid), all of which closely mimic natural amino-acid bias within the CDRs of human antibodies (Tomlinson and Winter, 2005) and fail to increase the net protein charge at neutral pH. Thus, the usage of degenerate codons designed to maximize amino-acid diversity within an intrabody framework, such as NNK, may inadvertently jeopardize cytoplasmic solubility.

In our experience, rat ST14A striatal progenitor cells are ideal for studying intrabody expression. Aside from their high transfectability and ease in culturing, ST14A cells maintain a relatively constant morphology during routine passaging, unlike some other cell types that tend to decline in size and health over time. Thus, confounding influences of cell volume and macromolecular crowding on intrabody solubility are theoretically controlled for in ST14A cells. Interestingly, we have observed that some intrabodies (particularly semi-synthetic VH3:VL Tomlinson scFvs) are increasingly aggregation-prone when expressed in morphologically smaller mammalian cells (e.g. HEK293) or when ectopically overexpressed from stronger promoters (E.K., unpublished results). These phenomena may reflect the empirical observation that certain VH3:VL domain combinations show suboptimal stability and folding (Ewert et al., 2003), a fact which is likely exacerbated by the overall concentration of intrabody in the cell.

In summary, we report that intrabody solubility in mammalian cell cytoplasm correlates with estimated protein charge and hydropathy for a representative group of human scFv and camelid VH3 intrabodies. However, given the multiple ways in which to engineer antibody fragments, we stress that these physico-chemical measurements may not accurately predict intrabody solubility for every case. In fact, we note that sequence-based estimates of net charge and hydropathy fail to correlate with the published solubility of cytoplasmic intrabodies derived from a hyperstabilized framework which has undergone several rounds of selection by molecular evolution (Philibert et al., 2007). It is likely that mutations introduced into synthetic frameworks by serendipity impart a greater degree of context-dependent thermodynamic stability than can be anticipated solely on the basis of charge or hydropathy. Related to this fact, our findings may not necessarily apply to intrabodies that have acquired unusual sequence alterations that deviate significantly from consensus frameworks (such as large duplications, deletions or insertions that can arise aberrantly during library construction, chain shuffling or site-directed mutagenesis). Furthermore, although soluble expression is influenced by intrabody aggregation propensity, other parameters such as protease susceptibility and toxicity can significantly influence steady-state intrabody levels (Zhu et al., 1999). This fact likely explains why steady-state levels of certain Tomlinson scFv intrabodies in our study (e.g. scFv-D1) failed to correlate well with either observed aggregation propensity or the estimates of net charge and hydropathy. Importantly, all intrabodies in the present study were expressed in the absence of cognate antigen(s), a condition also known to variably affect intrabody half-life (Sibler et al., 2005). Thus, although net negative charge and, to a lesser extent, reduced hydrophilicity are positive indicators for reduced intrabody aggregation in cell cytoplasm, these sequence-based parameters do not guarantee that cellular intrabody levels will accumulate to a high degree in the steady state.

**Supplementary data**

Supplementary data are available at PEDS online.

**Acknowledgements**

We thank Drs Annemarie Honegger (Universität Zürich, Switzerland) for expert advice on the in silico antibody engineering of scFv-C4, Elena Cattaneo (University of Milano, Italy) for the rat ST14A cell line, Daniel McVicar (National Cancer Institute, Frederick, MD) for providing anti-TLT1 scFvs C10, D1 and A8 and Ray Truant (McMaster University, Ontario, Canada) for plasmid encoding mRFP1 (Q66T). We also thank Kathleen Dumas and Bin Yuan for help in vector construction, and David Butler, Tyisha Hathorn, Kevin Manley and Abigail Snyder-Keller for helpful discussions on this study.

**Funding**

This work was supported by the National Institutes of Health (5R21 NS061257-02 to A.M. and M.R.S.; 5R01 NS053912-04 to A.M. and N01 AI30050 and U54 AI057159 to C.B.S.), the Michael J. Fox Foundation to M.R.S., and the Arizona Biomedical Research Foundation to M.R.S. Summer student work on this project was supported by the NSF-REU Program (DBI0243690) to the Wadsworth Center.

**References**