CHARACTERIZATION OF THE FOLDING AND ASSEMBLY OF SINGLE-CHAIN ANTIBODIES

by

Nathaniel Macapagal

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Chemical Engineering

Spring 2011

CHARACTERIZATION OF THE FOLDING AND ASSEMBLY

OF SINGLE-CHAIN ANTIBODIES

by

Nathaniel Macapagal

Approved: _____

Anne S. Robinson, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Norman J. Wagner, Ph.D.
Chair of the Department of Chemical Engineering

Approved: _____

Michael J. Chajes, Ph.D.
Dean of the College of Engineering

Approved: _____

Charles G. Riordan, Ph.D.
Vice Provost for Graduate and Professional Education

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The focus of this thesis is to study the refolding of single-chain antibodies (scFvs), 4M5.3 and 4-4-20, expressed as inactive inclusion bodies (IBs) to their active monomer form. Previous work on the chemical refolding for 4M5.3 revealed two monomer forms, where one was less active (A1) than the other (A2) as determined by fluorescence quenching assays. Numerous biophysical assays were employed to determine structural differences, which may explain the difference in binding characteristics. The biophysical assays and examples in literature pointed to a cis/trans isomer difference at proline residue 100 for A1 (trans form) and A2 (cis form). Numerous mutational studies were conducted to stabilize the cis form of Pro100 in 4M5.3 but none of the mutants generated only one form of the molecule, suggesting that there maybe a different proline residue responsible for the difference in the structure (hydrodynamic radius and surface hydrohpbicity) of A1 and A2 or some other structural reason for the difference, such as differences in salt-bridging, hydrogen bonding, hydrophobic interactions or a combination of all these factors at the interface between the light and heavy domains of the scFv.

A different method for refolding 4M5.3 using hydrostatic pressure from inclusion bodies, which does not use any denaturants, also, showed A1 and A2 after purification by size-exclusion chromatography showing that the formation of the less active form is independent of the refolding method. The use of hydrostatic pressure to refold scFvs from IBs was extended to another scFv, 4-4-20, and to the addition of

their ligand, fluorescein, during refolding in order to study its effect on the refolding process. The inclusion of ligand increased yields of scFv by two-fold for 4-4-20 and by four-fold for 4M5.3, which is the higher affinity mutant for fluorescein. Two methods were investigated to obtain the free form of the scFvs by removing the bound ligand. A controlled-dilution and filtration method which uses a low concentration of denaturant to dislodge the ligand and then filtration with renaturation buffer to flush the denaturant away was found to yield the highest amount of free active monomer for 4-4-20 and 4M5.3. However, the scFvs produced by this method were found to be of lower quality, where they had a slower on-rate and a faster off-rate for their ligand, as compared to chemical refolding.

The final chapter of this thesis investigated ways of measuring the macromolecular folding rate of each scFv and, also, the folding rates of each domain and interface for each scFv. The determination of the rate of refolding for 4M5.3 and 4-4-20 by dilution and measuring the rate of fluorescence quenching was successful. This method would be useful for measuring the rate of refolding for very small amounts of protein or if the tryptophan (Trp) residues in the molecule were removed, since there was a way to track the rate of active protein appearance from unfolded protein. Trp to phenylalanine (Phe) mutations were made to 4M5.3 and 4-4-20 with the strategy of creating mutants that would have a fluorescent Trp residue in the domain or interface of interest as a tracker for the rate of folding for that domain. Trp mutants for each scFv were successfully created as verified by DNA sequencing and

their activity verified by fluorescence quenching assay. However, we were unsuccessful at measuring a rate of refolding for the domains and interfaces of 4M5.3 and 4-4-20 using a rapid dilution method and tracking Trp fluorescence. The rapid dilution method was unsuccessful at refolding adequate amounts of refolded protein to detect enough signal from Trp fluorescence.

Chapter 1

INTRODUCTION

Many important and novel recombinant protein therapeutics are either currently on the market or are in pre-clinical development (*1*). A large subsection of these protein therapeutics are monoclonal antibodies, such as, Synagis®, Herceptin®, Mylotarg® and Humira® (*2*). Antibodies consist of an Fc region, which acts as the signaling part of the antibody that binds receptors on other immune system cells to elicit a response, and Fv regions that bind specific antigens that are targeted by the antibody (Figure 1). An antibody can be enzymatically or recombinantly broken down into different architectures depending on the desired application. For example, an Fc protein molecule on its own may be used to trigger a general immune response without needing the Fv region to bind a specific molecule. A Fab fragment or the Fv fragment on its own can be used as a blocking molecule on a receptor molecule on the surface of a cell to stop a specific cellular response. The single-chain antibody (scFv) format is the focus of this thesis, which consists of the Fv fragment held together by a polypeptide linker. Interest in this format has grown due to its possible therapeutic and diagnostic uses which enable it to perform better than the whole anti-body molecule due to its smaller size, while retaining the binding properties of the larger molecule (*3*). Examples of smaller formats currently on the market are Lucentis® and Cimzia® (*2*) which are Fab fragments and not whole antibodies.

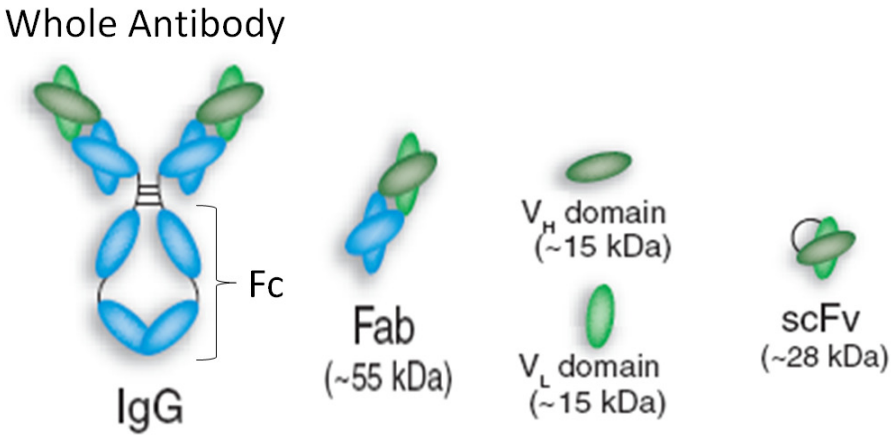


Figure 1. Schematic diagram showing the domains of a whole antibody and some of the different architectures used for diagnostic or therapeutic purposes based on the whole IgG (*3*).

Most antibodies used for diagnostic or therapeutic purposes are expressed from mammalian cell culture which has the disadvantage of not being as cost-effective as bacterial culture. Mammalian cell cultivation is characterized by low volumetric yields, long cultivation times, and expensive capital equipment and raw materials (*4*). On the other hand, bacterial cultivation processes are based on inexpensive equipment and raw materials in which high volumetric yields and fast growth can be achieved. However, cultivation of recombinant proteins in bacteria, such as *Escherichia coli*, frequently yields inactive protein in the form of inclusion bodies (*4*).

Bacterial inclusion bodies are insoluble higher-order aggregates of the native protein, typically, made up of misfolded monomer species that still retain some of their secondary structure characteristics. These dense (~1.3mg/ml), highly hydrated aggregates contain very little host protein, ribosomal components or DNA/RNA

fragments and are held together by non-covalent hydrophobic or ionic interactions or a combination of both. Inclusion bodies held covalently together by non-native disulfide bonds can also form from over expressed proteins that normally contain native disulfide bonds (*5, 6*).

There are several methods used to obtain active protein from inclusion bodies, which typically rely on chemical denaturation and refolding through dilution, dialysis or dilution and filtration (*7*). It is important to study this method of protein production because there are numerous proteins that are either on the market or in clinical development that express the drug product in bacterial hosts as inclusion bodies and rely on refolding methods to produce the active protein (*4*). Also, protein refolding of soluble aggregates which form during processing of protein therapeutics using different hosts is used extensively and the information obtained from this thesis regarding how proteins form their native structure could potentially be valuable for those applications (*8, 9*).

Two single-chain antibodies (scFv's), 4-4-20 and 4M5.3, were used in this thesis as model proteins to investigate the *in vitro* refolding of recombinant proteins from *E. coli* inclusion bodies. 4-4-20 is an anti-fluorescein-biotin scFv architecture that was generated from an anti-fluorescein-biotin antibody by Denzin et al (*10*). 4M5.3 is a high-affinity mutant of 4-4-20 that was generated by random mutagenesis coupled with yeast surface-display selection to elucidate the relationship between changes in structure and increased binding of the variable region to its antigen (*11-13*).

Both 4-4-20 and 4M5.3 quench the fluorescence of fluorescein-biotin upon binding by 94% for 4-4-20 and by 95% for 4M5.3 (*11*). A space-fill graphical representation of 4M5.3, derived from the X-ray crystallography structure coordinates (*12*) (PDB ID:1X9Q), is shown in figure 2 (A), which depicts the light chain (blue) and heavy chain (red) binding to fluorescein-biotin (green). Figure 2 (B) is a different depiction of 4M5.3 showing the beta-sheet (arrows), alpha-helical (cylinders) and random-coil content of the scFv.

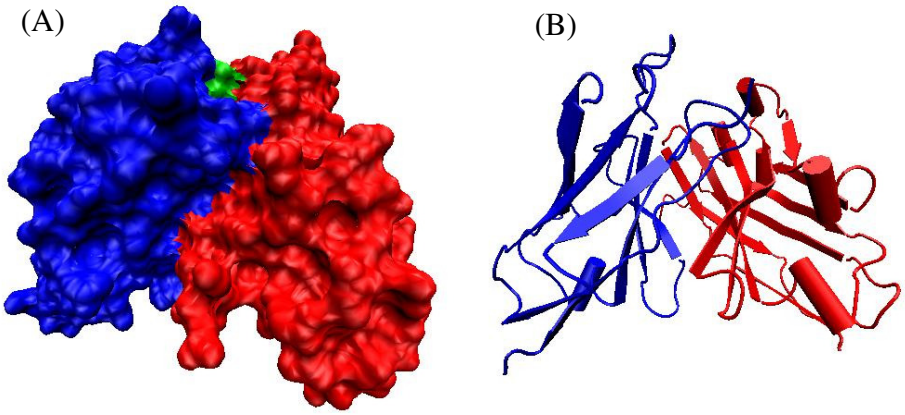


Figure 2. (A) Space-fill depiction of 4M5.3 rendered from PDB ID: 1X9Q using Swiss-PDB viewer. The light and heavy chains of the scFv are shown in blue and red, respectively. The ligand, fluorescein, is shown in green. (B) Depiction of 4M5.3 showing secondary structure motifs.

Fluorescein-biotin absorbs light at an excitation maxima of 494nm and emits light at 521nm (*14*). When bound to refolded scFv, the intensity of the emitted light diminishes. This characteristic was used to develop assays for determining folding kinetics and binding parameters at equilibrium (*11-13, 15*), which gives an idea on the

quality of the refolded scFv and how it compares to published values for 4-4-20 and 4M5.3 that were expressed and purified using different processes.

These two scFvs form inclusion bodies when expressed in *E. coli* under a strong lac Z promoter cultured at 37°C and required chemical solubilization and refolding to obtain the active monomeric scFv. Also, investigated was the use of hydrostatic pressure as an alternative to chemical denaturation for producing active scFv. Several biophysical assays were used to determine the activity and structure of the protein obtained by the processes studied. Based on the results and findings of the different assays, possible mechanisms of refolding for these two scFvs were postulated and tested using additional techniques.

Previous results for the chemical refolding of 4M5.3 revealed the presence of 2 active monomeric forms of 4M5.3, where one form had lower affinity for fluorescein-biotin than the other form (*16*). Chapter 2 of this thesis probed for a possible mechanism for the observations found regarding the two active monomeric forms and the quality of the scFvs produced by chemical refolding from inclusion bodies. The formation of the less active monomer could be unique to the process of chemical refolding from inclusion bodies.

Numerous studies have previously shown that high hydrostatic pressures can dissociate oligomers of protein trapped in an insoluble aggregate and pressures greater than 400MPa can fully unfold a protein molecule (*17-20*). The theory behind this lies

in the fact that increasing pressure shifts a system to its lowest volume. The lowest volume for macromolecules is, in general, the most solvated state, where waters force their way into the voids of a folded protein molecule pushing protein-protein interactions apart and replacing them with protein-water interactions. Therefore, the application of hydrostatic pressure drives the solvation of protein surfaces, thereby, denaturing them. The advantages of this method over the use of chemical denaturants are: it costs less per processing run because there is no need for expensive chemical denaturants which then need to be disposed of after refolding, higher yields per run where protein refolding can occur at much higher concentrations because the presence of pressure keeps folding intermediates from aggregating, less processing steps because there is no need for dilution or dialysis to remove the denaturant and refold the protein.

Chapter 3 of this thesis examined the use of hydrostatic pressure as an alternative method for refolding scFvs from inclusion bodies to see if the formation of the less active monomer was specific to the folding process. The kinetics of domain folding could be the reason for the multiple-active forms and yields that were observed in chapters 2 and 3.

Chapter 4 covers folding experiments conducted on 4M5.3 and 4-4-20 to elucidate their folding mechanisms which could have explained the phenomena we were seeing with the two production methods studied. A determination of macro-molecular scFv folding rate was developed by using fluorescence quenching and Trp

to Phe mutations were conducted on each scFv, in order, to track the folding rate of different domains for the molecule.

Chapter 2

MULTIPLE-ACTIVE CONFORMATIONS OF 4M5.3 AFTER CHEMICAL REFOLDING

Background

Refolding from inclusion bodies in *E.* coli offers a potentially lower cost method for protein production than eukaryotic expression. To study the process of refolding scFv proteins from inclusion bodies, this study has focuses on two well-characterized scFvs, 4M5.3 and 4-4-20, which bind to the ligand fluorescein.

4M5.3 (Figure 2) is a high-affinity mutant of 4-4-20 engineered by the Wittrup group at MIT by directed-evolution using random mutagenesis and yeast-surface display selection, which differs from 4-4-20 by twelve mutations in the heavy domain and two mutations in the light domain (*11-13*). These scFvs quench the fluorescence of fluorescein-biotin by greater than 90% upon binding. These molecules have a quick and easy spectrofluorimetric assay for the detection of proper folding and activity (*10*); however, they readily form inclusion body aggregates when expressed in *E. coli (15).*

While investigating *in vitro* refolding methods for 4-4-20 and 4M5.3 obtained from *E. coli* inclusion bodies, previous work in our laboratory showed multiple active forms of 4M5.3 (*15*). Here we present our characterization and folding studies aimed at identifying the distinguishing features of the active 4M5.3 isoforms, as well as how these forms result from the inclusion body folding process.

Materials and Methods

Expression of 4M5.3 and 4-4-20 scFv Inclusion Bodies

4M5.3 and 4-4-20 genes were inserted into the pET21-d plasmid vector (Novagen) and were expressed under an inducible *lac* promoter. Inclusion bodies of each scFv were made in 2L shake flasks where cell density was allowed to reach an OD between 0.6 and 0.8 prior to induction of protein expression using IPTG. The culture is then allowed to reach saturation between 4-8 hours before collection of the cell pellet via centrifugation for 30min at 3176 x g.

Isolation of Inclusion Bodies

Bacterial cells were lysed by three freeze-thaw cycles (1hr at -80ºC followed by 1hr at 37ºC and incubation with lysozyme at a 0.001mg/ml concentration. Endogenous DNA was digested by treatment with DNAse (0.1mM final conc) at 37ºC for 1hour. Inclusion bodies, which were part of the insoluble material, were pelleted by centrifugation at 3176xg for 30min. The soluble fraction was decanted and the insoluble fraction was resuspended with tris-buffered saline (TBS, 50mM Tris-HCl, 150mM NaCl, pH8.0). The resuspended pellet was centrifuged again at 3176xg for 30min and the soluble fraction was then decanted and discarded.

Purification and Isolation of 4M5.3 Forms

4M5.3 produced as inclusion bodies were solubilized using 6M guanidine HCl with and without β-mercaptoethanol, as described previously (*15*). The denatured protein was refolded with or without disulfide shufflers (10mM glutathione (oxidized), 50mM glutathione (reduced)) in 6M guanidine HCl by dialysis overnight, followed by overnight dialysis against 3M Guanidine hydrochloride (Sigma), then overnight dialysis against 1.5M Urea (Fisher), and finally overnight dialysis against phosphate-buffered saline (PBS; 150mM NaCl, 10 mM sodium phosphate buffer, pH 8) or low-salt buffer (LSB; 10mM sodium phosphate buffer, pH 8). Two buffer systems were investigated to compare to published equilibrium dissociation constants for 4-4-20 and 4M5.3 that were determined in these buffers. A 350 ml prep grade Superdex 75 (from GE Lifesciences) or Sephacryl 200 (GE) size-exclusion chromatography (SEC) column equilibrated with PBS was used to isolate the folded 4M5.3 species. Sample concentrations were determined by absorbance readings at 280 nm using a molar extinction coefficient of 46,850. Samples were stored at 4 °C prior to analysis.

Fluorescein-Biotin Quenching Activity Assay

Equal concentrations of the two purified active forms (A1 and A2) were prepared in PBS or LSB. A Hitachi 4500 fluorimeter (excitation at 490 nm and emission at 512 nm) equipped with a temperature controller and stirrer, in order to maintain a constant 25°C and a well-mixed system (vortex formation without cavitation), was used to measure the fluorescence intensity of 2ml of PBS or LSB buffer containing 20nM fluorescein. This fluorescein-biotin solution was titrated with

small additions of folded protein (5 µl) and the decrease in emission intensity was monitored to determine protein activity. Active protein was titrated until no further change in intensity was observed. The fluorescein-biotin solution was also titrated with buffer containing no scFv to correct for dilution effects on fluorescein-biotin intensity.

Free Thiol Determination

The Thiol and Sulfide Quantitation Kit (Molecular Probes) was used to assay for the presence of free thiols in the A1 and A2 species of 4M5.3. Thiols were also quantitated using the traditional Ellman's reagent, although the enzymatic amplification step of the Quantitation Kit enabled us to detect as little as 0.2 µM free thiol (0.2 nanomole in a 1 mL reaction), a sensitivity that is about 100-fold greater than that achieved using Ellman's reagent. L-cysteine was used as a standard, with Ellman's reagent detection used for accurate standard determination. Cystamine was added to the reaction to improve detection of poorly accessible thiols or those with high pKa values.

K_D Determination by Fluorescence Quench Titration

2 mL of 20nM fluorescein-biotin in PBS was placed in a Hitachi Spectrophotometer with the sample excited by light at 485 nm and the emitted light detected at 512 nm. Known volumes of active scFv were added to the fluorescein-biotin and the fluorescence intensity was measured until equilibrium was reached

(approximately 10 min). Titrations were repeated several times and the values were then fit for single-site binding according to equations 2.1-2.3:

$$I = m([F] + (1 - Q_{max})[B]) \quad (2.1)$$

$$K_D = \frac{([P_T]-[B])}{[B]} \quad (2.2)$$

$$[F_T] = [F] - [B] \quad (2.3)$$

where *I* is the measured intensity at equilibrium, *m* is a constant determined from a standard curve of fluorescein-biotin concentration versus intensity, [F] is the free fluorescein-biotin concentration, Q_{max} is the fluorescence quenching constant (between 0.95 and 0.98), [B] is the bound fluorescein-scFv complex, K_D is the equilibrium dissociation constant, $[P_T]$ is the total protein concentration in the system at that equilibrium point, and $[F_T]$ is the total fluorescein-biotin concentration in the system at that equilibrium point.

The three equations were solved at each equilibrium point by using a least-squares fit method (K_D, Active (bound) scFv concentration [B], and Q_{max}) to determine K_D as described previously (*11, 15*).

Association Rate Constant ($k_{association}$) Determination by Stopped-Flow Fluorimetry

Active scFv samples were injected with 5 nM, 10 nM, and 20 nM fluorescein-biotin in PBS or LSB into a stopped-flow fluorimeter. Excitation was set at 485nm and a bandpass filter was used, such that all the emitted light greater than 495nm was detected. Values of the emitted light were compared to a standard curve of free fluorescein-biotin concentration versus emission intensity, in order to determine fluorescein-biotin concentration in the sample. Concentration vs. time data was fit to a one-site binding model (equations 2.5-2.8), where the assumption was made that the off-rate of fluorescein-biotin from the scFv was negligible. F, L and B represent free scFv, unbound ligand and scFv-ligand complex, respectively. The value for k_a was solved using a least-squares fit method in MATLAB.

Binding Model:

$$F + L \Leftrightarrow B \tag{2.5}$$

$$\frac{dB}{dt} = k_{association} FL - k_{dissociation} B \tag{2.6}$$

$$k_{dissociation} \approx 0 \tag{2.7}$$

$$\frac{dB}{dt} = k_{association} FL \tag{2.8}$$

$$K_D = \frac{k_{dissociation}}{k_{association}} \tag{2.9}$$

Site Directed Mutagenesis

The Quik-Change II® Kit (Stratagene) was used to create point mutations in the 4M5.3 gene. Complementary primers from Integrated DNA Technologies were used to create the P8A, P100A, P100G, and V99Y. Positive clones from the mutagenesis were purified using Wizard DNA Miniprep Kits (Promega) and mutated genes were sequenced (DNA Sequencing Facility, U. Penn) to verify that the correct product was obtained.

Results

Identification of Species After Refolding

Functional 4M5.3 was recovered from solubilized *E. coli* inclusion bodies by a variety of dialysis and dilution and filtration based refolding techniques (*15*). Regardless of the folding technique, 4M5.3 always partitioned into reproducible, discrete species separable by size exclusion chromatography (SEC). Four distinct peaks are detected by absorbance at 280nm during size-exclusion chromatography of 4M5.3 protein refolded from inclusion bodies via dialysis under reducing conditions (Figure 3).

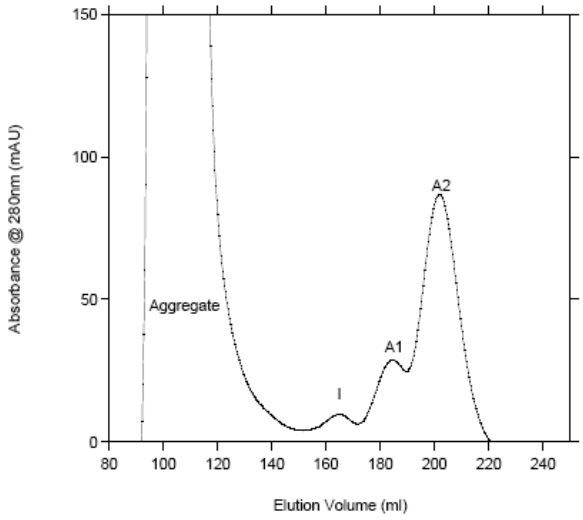


Figure 3. Size-exclusion chromatogram showing the 4M5.3 species produced after refolding by dialysis.

Inactive aggregate material, eluting in the void volume of a Superdex 75HR column, is the dominant species and typically accounts for approximately 65% of the total 4M5.3. Occasionally, inactive dimer-sized aggregate was observed for 4M5.3, although this usually represented a very small fraction of the total 4M5.3. Surprisingly, three peaks eluted from the column at expected monomeric residence times. The same refolded sample was taken and analyzed via Native PAGE followed by washing with fluorescein-biotin to detect active species (Figure 4) Active species detection was done by visualizing the bands that bound fluorescein by scanning the gel with light at 495nm and measuring the emitted light at 512nm using a Typhoon Scanner (GE).

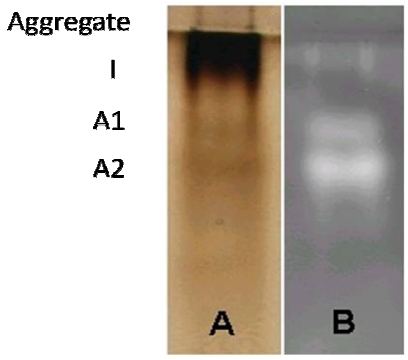


Figure 4. Silver stained (A) and fluorescein-washed (B) native PAGE images of 4M5.3 protein refolded via dialysis (*16*).

Based on gel mobility and the ability of the band to quench in-gel fluorescence, the peaks were identified as aggregate, an inactive monomer (I), an active monomer A1, and another active monomer A2. Small fractions were collected

during the SEC run and analyzed for activity to confirm the identity of the species. The identity of each of the peaks of scFv was also confirmed by Maldi-MS (*16*).

Different refolding conditions were investigated to see whether we could observe changes in the species distributions. First, non-reducing refolding was carried out with inclusion bodies of 4M5.3. Interestingly, we observe the disappearance of the inactive monomer form (I) when refolding is performed under non-reducing conditions (*16*). When high salt (PBS) and low salt (LSB) buffers were used, the same distribution of species was observed (data not shown), indicating that electrostatic effects did not influence the appearance of the different monomer species during non-reducing refolding.

Mass spectrometry was performed on SEC-purified samples of A1 and A2 to to rule out differences in post-translational modifications. The molecular weights of the A1 (28016 Da) and A2 (28023 Da) samples were approximately that expected for unmodified, monomeric 4M5.3 based on sequence (28188 Da), and is within the error limits of molecular weight determination using this equipment. N-terminal sequencing of the first 12 amino acids verified that the proteins had the same N-terminal sequence (*16*).

Disulfide bond differences between A1 and A2 were investigated by SDS PAGE and free thiol determination. Purified samples of aggregate, A1, and A2 were analyzed by SDS PAGE with and without the addition of reducing agent. SDS gel

mobilities of the A1 and A2 species were unaffected by the addition of reducing agent (*16*).

A colorimetric assay using 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB), was used to detect free cysteines in purified samples of aggregate, A1, and A2 as well as unseparated 4M5.3 batches. DTNB reacts with free thiol groups to produce the yellow-colored thionitrobenzoate. No color change, as monitored by absorbance at 405 nm, was observed for any of the purified samples or for a 4M5.3 batch before purification into its components, indicating that no free cysteines were present. A more sensitive assay using the thiol and sulfide quantitation kit (Molecular Probes) also determined that no free cysteines were present in either A1 or A2 (data not shown). Therefore, the four cysteines of 4M5.3 appear to form the correct intradomain bonds. Cross-domain disulfides are highly unlikely in either active form, as the structure of the protein would have to be substantially perturbed to allow cysteines in the heavy and light domain to come in close proximity for bonding.

Activity Characteristics

Two of the three monomeric forms of 4M5.3 were found to bind and quench fluorescein-biotin (Figure 3). However, we were interested in any apparent differences in the on ($k_{association}$) and off ($k_{dissociation}$) rates for fluorescein-biotin binding between the two species. $k_{association}$ was measured by stop-flow fluorimetry using refolded and purified A1 and A2 to quench the fluorescence of Fl-Bio in PBS and LSB.

Qualitatively, Figure 5 shows that A1 binds slower than A2 in both PBS and LSB and that the rate of association of A1 and A2 is faster in LSB compared to PBS. $k_{association}$ was determined from this raw data as described in the materials and methods section of this chapter.

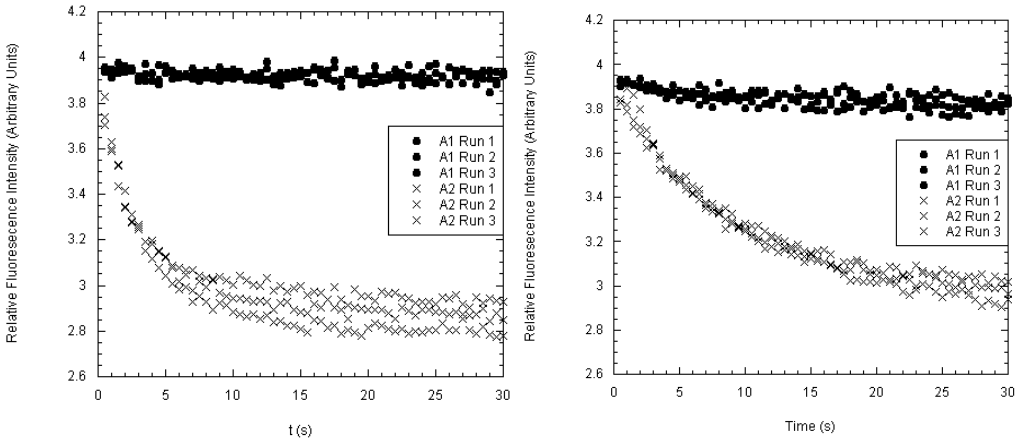


Figure 5. Comparison of rate of association ($k_{association}$) for A1 and A2 in LSB and PBS. Equimolar amounts (measured by A_{280}) of A1 and A2 were added to Fl-Bio in LSB. The relative fluorescence intensity was measured using stop-flow fluorimetry.

The equilibrium dissociation constant K_D was experimentally determined by titrating known amounts of A1 and A2 into Fl-Bio in LSB and PBS and measuring the fluorescence at equilibrium and numerically solving for K_D using a least-squares method with a three-parameter search model (K_D, Active scFv concentration, and Qmax). Figure 6 shows the results of the three-parameter fit for numerically solving for the K_D of A1 and A2 in LSB and PBS. Based on the K_D value, $k_{dissociation}$ can be calculated as K_D multiplied by $k_{association}$.

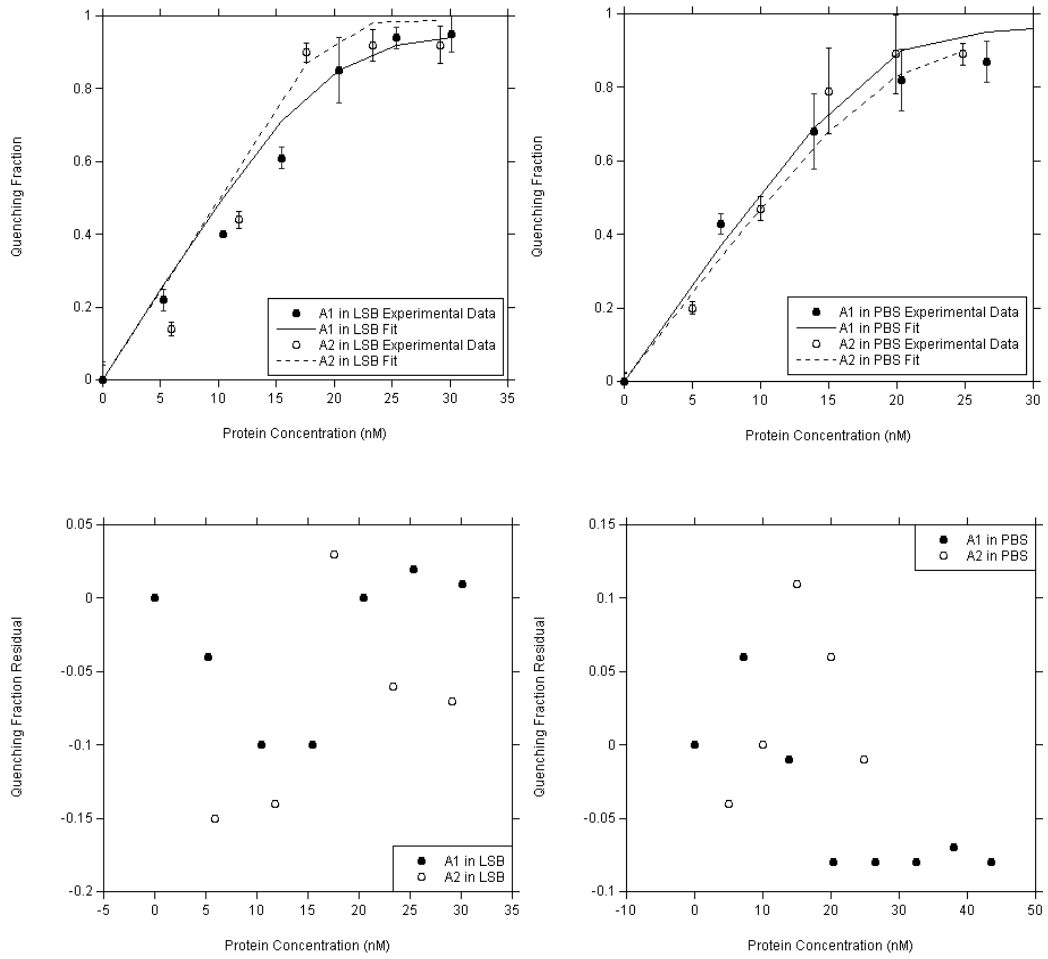


Figure 6. Graphs showing the fluorescence quenching at equilibrium for different concentrations of A1 and A2 refolded by dialysis into LSB and PBS added to Fl-Bio in LSB and PBS. The open and closed circles represent the raw data with error bars from the noise of the signal that were determined from one refolding experiment, while the solid and dashed lines represent the numerical fit used to determine KD. Also, shown below each of the fits are the residuals for each equilibrium point.

Table 1 shows the summary of binding properties for A1 and A2 in PBS and LSB based on a representative refolding experiment out of several and the reported values for 4M5.3 produced in yeast, in the literature (*11, 12*). The data shows that A2 has better binding characteristics than A1 in both high and low-salt buffer systems,

since A2 binds fluorescein-biotin faster and tighter than A1 as shown by the experimental measurements of $k_{association}$ and K_D. There are substantial differences between the binding properties of A2 in LSB compared to previous results reported for 4M5.3. However, the $k_{association}$ and K_D values in PBS match up well with previously results. The previously reported binding properties for 4M5.3 were obtained using protein which was solubly expressed in yeast and affinity purified, which is quite different from the production method used in this study, which could explain for the discrepancy. Also, chemical alterations of the protein during the refolding process, such as, carbamylation, deamidation or backbone nicks by proteases, could alter the conformation and the activity of the scFv (*21-24*). These changes in structure require more involved assays, such as, more elaborate mass spectrometry experiments that could detect these modifications to the amino acids in the protein sequence.

Table 1. Table of binding properties with Fl-Bio for A1 and A2 in LSB and PBS. The data for A1 and A2 shown above are average values from three individual refolding batches of 4M5.3.

LSB	A1		A2		Reported Values for 4M5.3 (*11*)	log Difference of A2 from Reported Value
$k_{association}$ (M^{-1} s^{-1})	3×10^{7}	$\pm 4 \times 10^{5}$	5×10^{6}	$\pm 7 \times 10^{5}$	29×10^{6}	0
K_D by Direct Titration (M)	3×10^{-8}	$\pm 4 \times 10^{-8}$	2×10^{-14}	$\pm 4 \times 10^{-19}$	4.8×10^{-14}	0
Calculated $k_{dissociation}$ (s^{-1})	2×10^{-2}	$\pm 3 \times 10^{-2}$	1×10^{-7}	$\pm 2 \times 10^{-8}$	1.6×10^{-6}	1
PBS	A1		A2		Reported Values for 4M5.3 (*11*)	log Difference of A2 from Reported Value
$k_{association}$ (M^{-1} s^{-1})	2×10^{6}	$\pm 1 \times 10^{6}$	3×10^{6}	$\pm 1 \times 10^{6}$	6.0×10^{6}	0
K_D by Direct Titration (M)	3×10^{-9}	$\pm 4 \times 10^{-8}$	2×10^{-14}	$\pm 1 \times 10^{-18}$	2.7×10^{-13}	1
Calculated $k_{dissociation}$ (s^{-1})	9×10^{-3}	$\pm 1 \times 10^{-2}$	7×10^{-8}	$\pm 3 \times 10^{-8}$	1.4×10^{-6}	1

Mutational Studies

The previous characterization studies conducted on the isoforms of 4M5.3 suggested that the difference between the A1 and A2 conformations was likely due to a difference in the isomerization state of the backbone (*16*). One hypothesis is that *cis (A2)-trans (A1)* proline isomerization at position 100 (Kabat position 95) of the light chain is limiting. For other single-chain antibodies studied, this residue must be in the *cis* form to achieve activity (*25-27*). The *trans* isomer has demonstrated a slow

folding phase, representing isomerization to the *cis* state that can occur under native conditions.

Several mutations to 4M5.3 were created to test the hypothesis that proline isomerization at position 100 causes the formation of two active forms during refolding from inclusion bodies. The suspected proline residue at position 100 was mutated to alanine and glycine, in order to test this hypothesis. Alanine and glycine were selected because the trans conformation for these amino acids is much more energetically favorable than the cis conformation, as compared to proline. It was expected that only the A1 (trans) form of 4M5.3 would be generated after chemical refolding of these two mutants. P8A was created where the proline residue at position 8 was replaced with alanine to serve as a control for the P100A and P100G variants. Position 8 in the light chain is remote from the binding site and has not been shown to be involved in proline isomerization-limited folding of other single-chain antibodies. An additional mutant was created at position 99, V99Y, because it has been previously shown that cis-Pro is stabilized by n-terminal aromatic residues in small peptides (*28*). Therefore, the V99Y protein should predominantly form the A2 form of 4M5.3, if cis-trans isomerization were limiting.

Surprisingly, the mutations made to 4M5.3 did not change the number of species obtained, as observed by SEC analysis following chemical refolding of the protein variants from inclusion bodies. Figure 7 shows bar graphs depicting the relative abundance of aggregate, I, A1 and A2 after refolding by dialysis for each

mutant studied. This suggests that proline isomerization may be only one of multiple factors differentiating A1 and A2.

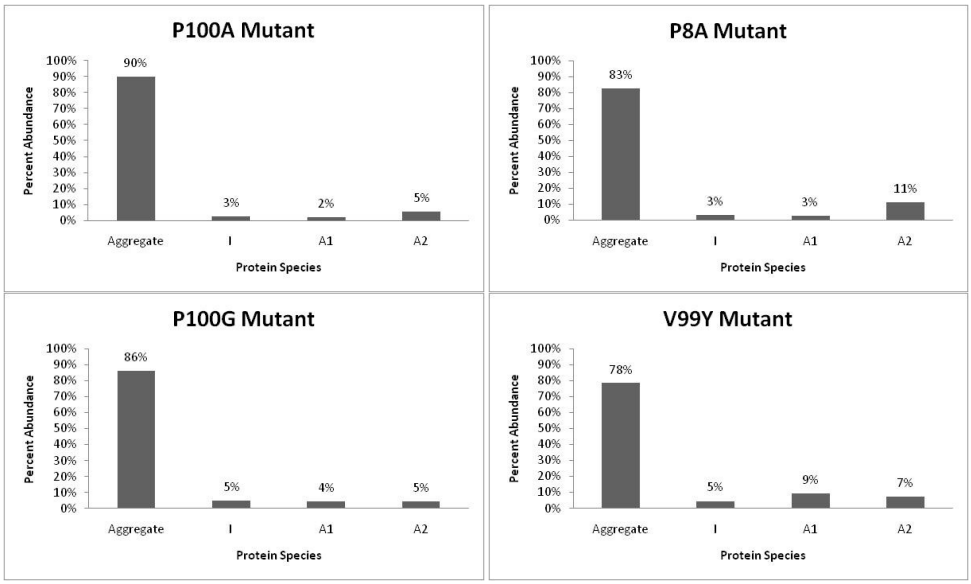


Figure 7. Bar graphs showing percent abundance of aggregate, I, A1 and A2 after refolding by dialysis for each mutant studied.

Discussion

After chemically refolding scFv 4M5.3 from inclusion bodies, four distinct species were detected during size-exclusion chromatography: Aggregate, inactive monomer (I), 2 active monomers (A1 and A2), when reducing conditions are present during the refolding process. When non-reducing conditions are used, the inactive monomer form does not occur, which provides strong evidence that incorrect inter-domain disulfide bond formation (one in V_L and one in V_H) was responsible for the observed inactive form (I). This agrees with the SEC data (Figure 3) showing that I has a larger hydrodynamic radius than A1 or A2 because it can be inferred from the structure of 4M5.3 (Figure 2) that inter-domain disulfides will create molecule with a larger hydrodynamic radius. Free thiol determination of A1 and A2 revealed no free cysteines in either molecule and there were no differences mobility during reduced and non-reduced SDS-PAGE. Different salt concentrations of the final refolding buffer did not change the appearance of two active monomer forms.

The conformational difference between A1 and A2 affected the ability of the molecule to bind fluorescein, where A2 was able to bind fluorescein-biotin faster and release it more slowly. The reported binding characteristics of 4M5.3 purified from yeast by Middlefort et al (*12, 13*) are different from the values obtained in this study. These differences can be attributed to a number of factors such as the host cell used in the recombinant expression, route of expression, and assay methodology. The binding

characteristics determined by Middlefort et al suggest that the quality of the protein produced by our method is lower due to its slower $k_{association}$ and higher K_D.

Interestingly, A1 and A2 both adopt the same hydrodynamic radius upon binding to fluorescein, which suggests that the conformational difference is energetically reversible (*16*). The activation energy for this transition is equivalent to trans-to-cis proline conversion in small peptides, which led us to believe that the cause for the conformational difference in A1 and A2 was a prolyl-isomerization difference at one of the five proline residues present in 4M5.3. Proline 100 was of particular interest, since previous studies have been conducted that have shown proline at this canonical site needed to be in the *cis* conformation for the similar scFvs to show activity (*29*).

To test this hypothesis, several mutants of 4M5.3 were generated. None of the protein variants generated produced only one active conformation, which suggests that either the prolyl-isomerization event was occurring at one of the other remaining 6 proline residues, or that differences in salt bridging or hydrophobic interactions between the light and heavy domains are causing the presence of both A1 and A2. Further mutational studies could be conducted to further investigate this hypothesis.

Understanding scFv folding is critical to improving our ability to refold single-chain antibodies for therapeutic uses, and finding two active monomers with such different ligand release rates is significant. The ability to understand and control the

formation of multiple active forms is essential to maximizing scFv affinity and ensuring scFv homogeneity. Also, it is clear that other antibody constructs, such as Fab, may demonstrate this ability to adopt multiple active conformations. It has been reported that an antibody raised against an aromatic ligand, thioredoxin, possesses several isoforms that exist in equilibrium (*30*). One difference between our studies and that of Tawfik and colleagues is that the two active conformations of 4M5.3 do not appear to be in equilibrium with one another under native conditions, but rather exist as stable yet trapped states.

Chapter 3

IN VITRO REFOLDING OF SINGLE-CHAIN ANTIBODY INCLUSION BODIES USING HYDROSTATIC PRESSURE

There are several methods used to obtain active protein from inclusion bodies, which typically rely on chemical denaturation and refolding through dilution (*7*). The use of hydrostatic pressure to obtain active protein from soluble inactive aggregates and from insoluble inclusion bodies is being studied here for the recovery of active recombinant single-chain antibodies (scFv). Anti-fluorescein single-chain antibodies called 4-4-20 and 4M5.3 were used in this study.

Previously, the use of hydrostatic pressure on scFvs from soluble aggregates resulted in misfolding and re-aggregation, once the pressure was released. Previous attempts at obtaining active refolded protein using hydrostatic pressure produced a modest 15% monomer yield from soluble aggregates in the presence of low concentrations of chemical denaturant and disulfide shufflers (*16*). To improve upon the yield previously obtained, this study refolded the proteins using hydrostatic pressure in the presence of its ligand, fluorescein. Theoretically, we can assume that the bound state for 4M5.3 has evolved to be the most thermodynamically favorable state, then it should be preferred over the aggregate form or some misfolded state. Therefore, during refolding the ligand may act as a template for the unfolded protein and prevent dissociation-pathway misfolding or re-aggregation.

Materials and Methods

Expression of 4M5.3 and 4-4-20 scFv Inclusion Bodies

4M5.3 and 4-4-20 genes were inserted into the pET21-d plasmid vector (Novagen) and were expressed under an inducible *lac* promoter. Inclusion bodies of each scFv were made in 2L shake flasks where cell density was allowed to reach an OD between 0.6 and 0.8 prior to induction of protein expression using IPTG. The culture is then allowed to reach saturation between 4-8 hours before collection of the cell pellet via centrifugation for 30min at 3176 x g.

Isolation of Inclusion Bodies

Bacterial cells were lysed by three freeze-thaw cycles (1hr at -80ºC followed by 1hr at 37ºC and incubation with lysozyme (0.001mg/ml). Endogenous DNA was (digested by treatment with DNAse (0.1mM) at 37ºC for 1hour. Inclusion bodies, which were part of the insoluble material, were pelleted by centrifugation at 3176xg for 30min. The soluble fraction was decanted and the insoluble fraction was resuspended with tris-buffered saline (TBS, 50mM Tris-HCl, 150mM NaCl, pH8.0). The resuspended pellet was centrifuged again at 3176xg for 30min and the soluble fraction was then decanted and discarded.

Pressure-Treatment of Inclusion Bodies

The pelleted inclusion bodies were resuspended with TBS. The total molar amount of protein was determined by pelleting 1ml of the resuspended inclusion bodies and then dissolving in 1ml 6M Gdn HCl. The concentration of protein in the solution was then determined by absorbance at 280nm. Fluorescein was added to a 10-fold molar excess of the total protein contained in the inclusion bodies. The samples were then placed in heat-sealed, air-free, sterile polyethylene transfer pipets. An R1-10-60 pressure vessel (High Pressure Inc.) was used to house the sealed pipets during pressure treatment. Pressure was applied to 40kpsi (2.8kbar) using a 37-5.75-60 Extra-Capacity Laboratory Pressure Generator (High Pressure Inc.) and held for 4, 24, and 48 hours.

Purification of Pressure-Treated Inclusion Bodies

After pressure treatment, the sample was centrifuged for 10min at 10,000xg to remove all the insoluble material, 0.22µm filtered and then loaded onto a Superdex 75 26/60 Prep Grade Size Exclusion Column (GE) equilibrated with TBS. The sample was fractionated and monitored by absorbance at 280nm. The peak corresponding to bound monomer was collected.

Post-Pressurization Yield Assay

To determine the amount of scFv liberated from the inclusion bodies by hydrostatic pressure, a sample of the scFv bound to fluorescein fraction was denatured by the addition of 6M Guanidine Hydrochloride 1:1 v/v to release bound fluorescein

and placed in a Hitachi Spectrometer. The fluorescence emission at 512nm (excitation at 485nm) was then monitored until equilibrium. The final fluorescence value was compared to a standard curve (fluorescein in 6M Gdn HCl concentration versus fluorescence intensity) to obtain an amount of fluorescein liberated from the scFv. Since ligand binding is monovalent, the amount of fluorescein measured equals the amount of active scFv produced by the pressurization step, if release is 100% efficient.

Fluorescein Stripping of Bound scFv

Two different methods were explored to obtain the apo protein. The Size-Exclusion Chromatography (SEC) release method involved treating the bound scFv fraction with 1.5M Guanidine Hydrochloride. Then following incubation at room temperature for 5 minutes, the sample was directly injected onto a Superdex 75 26/60 Prep Grade Size Exclusion Column (GE) previously equilibrated with TBS. The protein present in the sample was monitored by absorbance at 280nm and the peak corresponding to free scFv was collected.

The second method to remove fluorescein from the bound complex was a controlled dilution and filtration (CDF) release method which involved placing the bound scFv fraction in 1.5M Guanidine Hydrochloride and dilution and filtration Amicon Pressure Concentrator fitted with a YM-10kDa membrane. The released fluorescein was diluted and washed to the filtrate with 3 dilution and filtration volumes of fresh 1.5M Guanidine Hydrochloride followed by 3 volumes of TBS. The

resulting protein sample was then analyzed by SEC to ensure that only one peak that corresponds to the free monomer was present.

K_D Determination by Fluorescence Quench Titration

2 mL of 20 nM fluorescein in PBS was placed in a Hitachi Spectrophotometer with the sample excited by light at 485 nm and the emitted light detected at 512 nm. Known volumes of active scFv were added to the fluorescein and the fluorescence intensity was measured until equilibrium was reached (approximately 10 min). Titrations were repeated several times and the values were then fit for single-site binding according to equations 3.1-3.3.

$$I = m([F] + (1 - Q_{max})[B]) \quad (3.1)$$

$$K_D = \frac{([P_T]-[B])}{[B]} \quad (3.2)$$

$$[F_T] = [F] - [B] \quad (3.3)$$

where *I* is the measured intensity at equilibrium, *m* is a constant determined from a standard curve of fluorescein concentration versus intensity, [F] is the free fluorescein concentration, Q_{max} is the fluorescence quenching constant (between 0.95 and 0.98), [B] is the bound fluorescein-scFv complex, K_D is the equilibrium dissociation constant, $[P_T]$ is the total protein concentration in the system at that equilibrium point, and $[F_T]$ is the total fluorescein concentration in the system at that equilibrium point.

The three equations were solved at each equilibrium point by using a least-squares fit method in MATLAB (K_D, Active (bound) scFv concentration [B], and Q_{max}) to solve for K_D as described previously (*11, 15*).

Association Rate Constant ($k_{association}$) Determination by Stopped-Flow Fluorimetry

Active scFv samples were injected with 5 nM, 10 nM, and 20 nM fluorescein in PBS or LSB into a stopped-flow fluorimeter. Excitation was set at 485nm and a bandpass filter was used, such that all the emitted light greater than 495nm was detected. Values of the emitted light were compared to a standard curve of free fluorescein concentration versus emission intensity, in order to determine fluorescein concentration in the sample. Concentration vs. time data was fit to a first-order one-site binding model (equations 3.5-3.9) where the assumption was made that the off-rate of fluorescein from the scFv was negligible. F, L and B represent free scFv, unbound ligand and scFv-ligand complex, respectively. The value for k_a was solved using a least-squares fit method in MATLAB.

Binding Model:

$$F + L \Leftrightarrow B \tag{3.5}$$

$$\frac{dB}{dt} = k_{association} FL - k_{dissociation} B \tag{3.6}$$

$$k_{dissociation} \approx 0 \tag{3.7}$$

$$\frac{dB}{dt} = k_{association} FL \tag{3.8}$$

$$K_D = \frac{k_{dissociation}}{k_{association}} \tag{3.9}$$

Dissociation Rate Constant ($k_{dissociation}$) Determination by Non-Fluorescent Competitor Assay

1ml of 20nM fluorescein was quenched to a minimum fluorescence intensity using a known molar amount of active scFv. A five-molar excess of a non-fluorescent competitor (5-Aminofluorescein, Sigma) with respect to the scFv was added to the quenched solution. The increase in fluorescence intensity (I) versus time was measured and fit to a first-order rate equation using MATLAB with respect to the bound scFv-fluorescein complex concentration (*B*)) (equation 3.10), in order to calculate the rate of dissociation, $k_{dissociation}$.

$$\frac{dI}{dt} = k_{dissociation}[B] \tag{3.10}$$

Results

Effect of Fluorescein on Pressure Refolding

In order to test the hypothesis that refolding proteins using hydrostatic pressure in the presence of ligand yields active scFv from inclusion bodies, IBs were over-expressed and recovered from a 1L culture of E. coli. The insoluble IBs were then resuspended in 10ml of TBS. A 1ml sample of the suspension was taken to get an approximate amount of the total protein present in the 1L culture by spinning down the IBs in the 1ml sample, solubilizing the IBs in 6M Gdn HCl in TBS, then measuring the absorbance by 280nm. An approximate amount of total protein in the IBs was then calculated from this absorbance value. The remaining 9ml of IBs resuspended in TBS was spiked with a 50-fold molar excess of fluorescein and then treated with hydrostatic pressure.

After pressure treatment at 40kpsi for 14days, with a 50-fold molar excess of fluorescein, a maximum yield of 2.79mg/L of culture of active scFv monomer was obtained from inclusion bodies resuspended in TBS as determined by the post-pressurization yield assay. The SEC chromatograms in figure 3.1 shows that the presence of fluorescein significantly increased the monomer protein population compared to pressure treatment without fluorescein for both 4-4-20 (Figure 8 (A)) and 4M5.3 (Figure 8 (B)). The higher affinity mutant, 4M5.3, showed a bigger increase (~150% increase) in monomer yield with the addition of fluorescein than the lower

affinity scFv, 4-4-20 (~50%). However, even with fluorescein present at these pressurization parameters, there is still a significant insoluble aggregate population that suggests inadequate aggregate denaturation or reaggregation following pressure release.

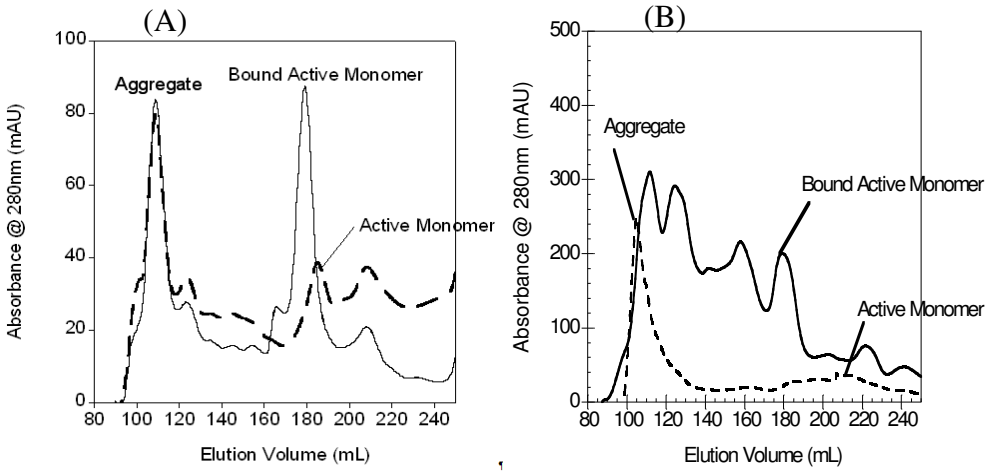


Figure 8. SEC chromatograms comparing pressure refolded 4-4-20 (A) and 4M5.3 (B) with (solid line) and without (dashed line) fluorescein. The identity of each peak for the pressure refolding without fluorescein experiment was determined by assaying each peak for fluorescence quenching activity, while peak identity for refolding with fluorescein was determined by fluorescence detection after the addition of denaturant to the protein sample from each peak in the chromatogram.

Ligand Removal after Pressure Treatment

Two methods were investigated to obtain the apo protein free of ligand. In one approach, the bound monomer fraction was treated for 5min at room temperature with 1.5M Guanidine HCl and then directly injected onto an SEC column equilibrated with TBS for on-column separation of the free monomer and fluorescein or by controlled dilution and filtration to remove the released fluorescein prior to exchanging the

guanidine hydrochloride buffer with TBS. Figures 9 (A) and (B) are SEC chromatograms for 4-4-20 and 4M5.3 using this rapid SEC release method.

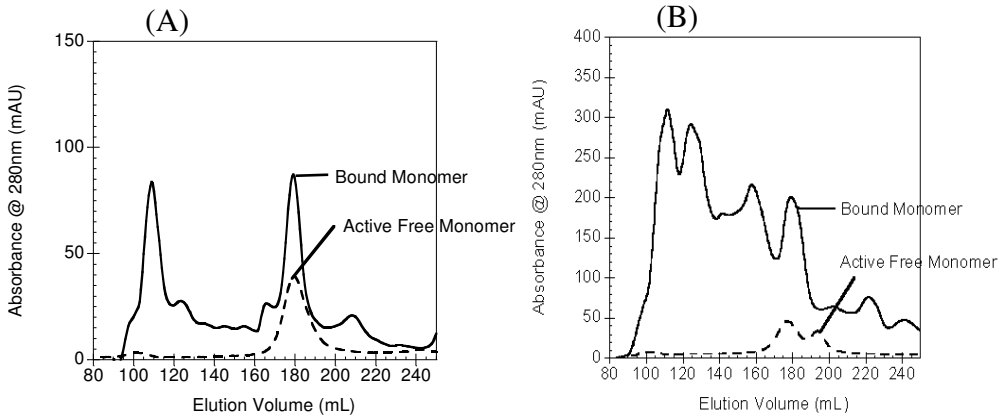


Figure 9. SEC chromatograms comparing fluorescein-stripped 4-4-20 (A) and 4M5.3 (B). Samples representing post-pressure treatment (solid line) and post-fluorescein stripping (dashed line) are shown. The identity of each peak was determined by assaying each peak for fluorescence quenching activity for the post-stripping fractions, while post-pressurization peaks were determined by fluorescence detection after the addition of denaturant to the protein sample from each peak in the chromatogram.

Table 2 shows the yields of free monomer from the different stripping methods three experimental runs that had different pressurization times. The SEC release method had an average recovery of 20% of free monomer, while the CDF method had an improved yield of around 65%.

Table 2. Comparison of active bound monomer yield for 4-4-20 and 4M5.3 held at 40kpsi with fluorescein for 5 and 14 days. Post-pressurization yield was determined by the addition of 6M Guanidine HCl to the active eluate fraction (or peak) and measuring the intensity of the fluorescence released at equilibrium and then determining the concentration of the released fluorescein using a standard curve. Since 4-4-20 and 4M5.3 are monovalent, the concentration of released fluorescein was equal to the concentration of the bound active monomer scFv recovered from pressure treatment. The yield of the apo scFv was determined by calculating the concentration of the active fraction by measuring the absorbance at 280nm.

Run	Hold Time at 40kpsi (days)	Post-Pressurization Yield (mg/L of culture)		Fluorescein-Stripping Method	Post-fluorescein Stripping Yield (mg/L of culture)	
		4-4-20	**4M53**		4-4-20	4M53
1	5	0.24	0.64	SEC release	0.04	0.05
2	14	0.27	1.17	SEC release	-	0.29
3	14	0.21	2.79	CDF release	0.14	-

Activity Analysis of Bound and Stripped Forms of the scFvs

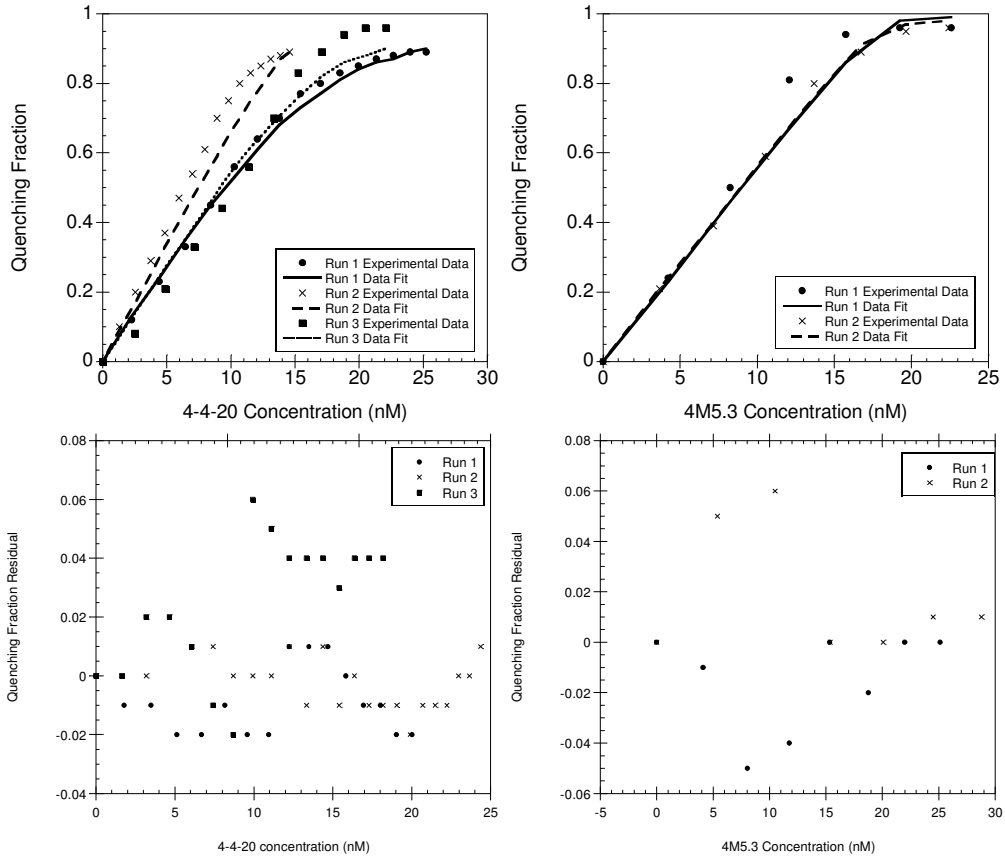


Figure 10. Numerical fit of quenching fraction versus protein concentration at each equilibrium point to determine K_D for 4-4-20 and 4M5.3. The numerical fit involved a three-parameter search on a set of non-linear simultaneous equations using MATLAB. Three independent experiments were conducted on three separate batches of 4-4-20 generated using the methods described and two independent experiments were done on two separate batches for 4M5.3. Each data point represents the fraction of fluorescein fluorescence quenched by a known amount of anti-fluorescein scFv in the sample at equilibrium. The plots at the beneath each fit show the residuals for each equilibrium point.

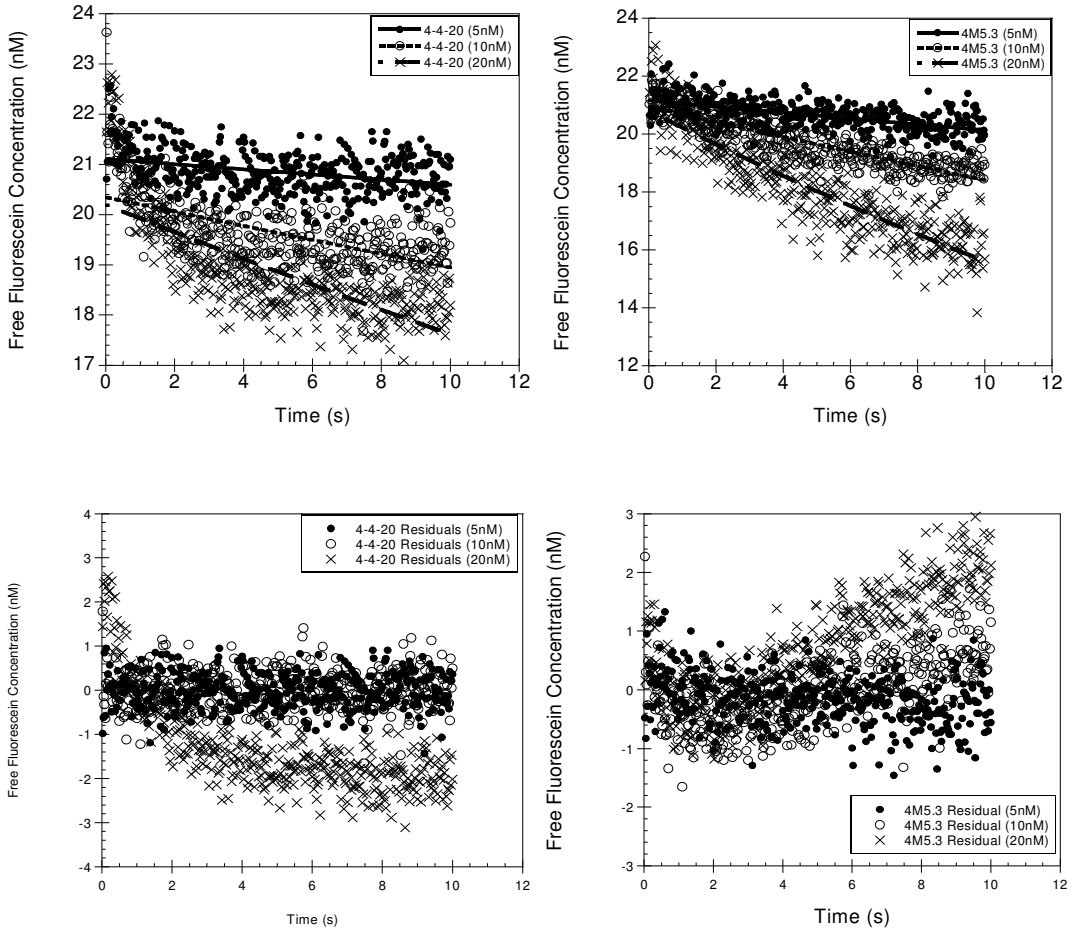


Figure 11. Graphs showing rate of association ($k_{association}$) for 4-4-20 and 4M5.3. Three separate assays were conducted for 4-4-20 and 4M5.3 at 5, 10 and 20nM concentrations of scFv. The rate of fluorescence decrease is equal to the rate of association and was numerically calculated by fitting a first-order rate equation with the data shown. The bottom graphs show the residual values from the data fit.

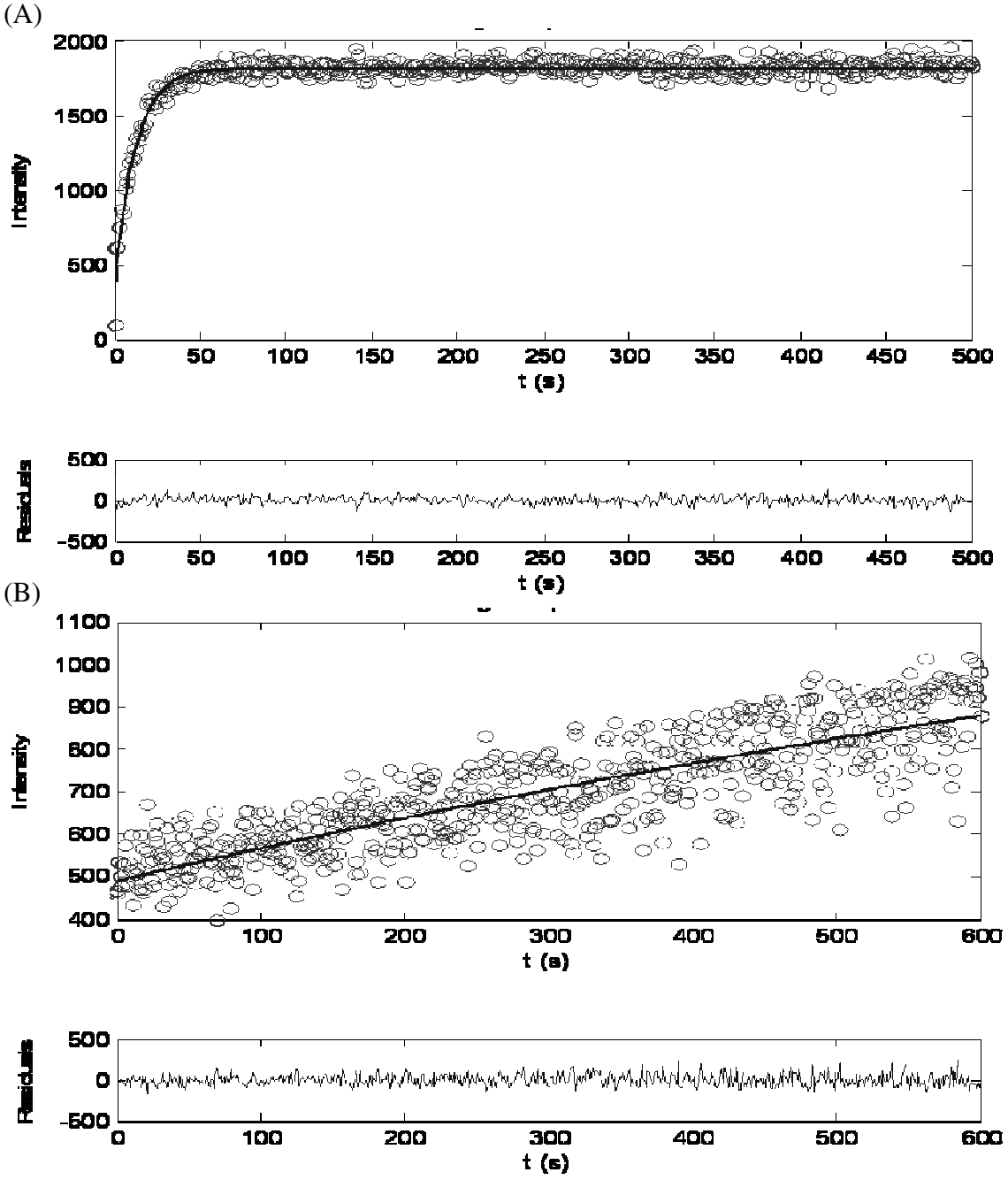


Figure 12. Graphs showing rate of dissociation ($k_{dissociation}$) for 4-4-20 (A) and 4M5.3 (B) via a competition assay with a non-fluorescent competitor. The rate of increase in fluorescence is equal to the rate of dissociation of fluorescein from the scFv. The data shown was fit to a first-order rate equation in order to numerically determine the value of $k_{dissociation}$ for 4-4-20 and 4M5.3.

Table 3. Summary of binding characteristics for pressure-treated active 4-4-20 and 4M5.3. The reported values for 4-4-20 and 4M5.3 are included for comparison and they compare favorably with the values generated in this study, even though the expression system used to generate the proteins and the assay methodology used was different.

Binding Property	4-4-20	Reported Values for 4-4-20 (*11*)	4M5.3	Values for 4M5.3 A2 in PBS from Ch.2
K_D (M)	1.4×10^{-9}	3.1×10^{-10}	1.6×10^{-10}	2×10^{-14}
$k_{association}$ ($M^{-1}s^{-1}$)	1.0×10^{6}	8×10^{7}	5.1×10^{5}	3×10^{6}
Calculated $k_{dissociation}$ (s^{-1})	1.5×10^{-3}	2.5×10^{-2}	8.0×10^{-5}	7×10^{-8}
Experimental $k_{dissociation}$ (s^{-1})	8.1×10^{-2}		1.1×10^{-3}	

K_D, $k_{association}$, and $k_{dissociation}$ post-fluorescein stripping were determined for 4-4-20 and 4M5.3 refolded by hydrostatic pressure from inclusion bodies (Table 3). The experimentally determined K_D for 4-4-20 and 4M5.3 by fluorescence quench titration, as described in section 2.2, were 1.4×10^{-9}M and 1.6×10^{-10}M, respectively. These values compare relatively well for the reported values of 3.1×10^{-10}M for 4-4-20 (*31*) and 2.7×10^{-13}M for 4M5.3 (*11*) despite a difference in protein expression process (bacterial expression in inclusion bodies versus soluble expression in yeast), purification procedure (chemical refolding followed by size-exclusion chromatography versus affinity chromatography) and assay methodology. The K_D for 4M5.3 produced by chemical refolding was 2×10^{-14} (Table 1). The value determined

by hydrostatic pressure refolding followed by ligand removal was 1.6×10^{-10} which is about four orders of magnitude less than the value determined by chemical refolding.

Independently determined $k_{association}$ and $k_{dissociation}$ values were determined by stopped-flow fluorimetry and by a non-fluorescent competitor assay, respectively, as described in the methods section of this chapter. Experimentally determined $k_{association}$ values for 4-4-20 and 4M5.3 were 5.18×10^5 $M^{-1}s^{-1}$ and 1.32×10^4 $M^{-1}s^{-1}$, respectively, while the reported values were 5×10^6 $M^{-1}s^{-1}$ and 29×10^6 $M^{-1}s^{-1}$. The experimentally determined $k_{dissociation}$ values for bound 4-4-20 and 4M5.3 were $8.1 \times 10^{-2}s^{-1}$ and $1.1 \times 10^{-3}s^{-1}$, respectively, which compares well for the reported value for 4-4-20 ($2.5 \times 10^{-2}s^{-1}$) but not for the reported value for 4M5.3 (1.6×10^{-6} s^{-1}).

Discussion

The addition of fluorescein to the refolding buffer greatly increases the amount of monomer obtained from inclusion bodies for both scFvs compared to hydrostatic pressure alone. We observed a 50% increase for 4-4-20 and a 150% increase in monomer species for 4M5.3. It is interesting that the higher affinity mutant is affected more by the presence of ligand. This gives evidence that the bound form of the molecule is the more thermodynamically favorable state than the aggregate form or some other misfolded state because the higher affinity mutant, which has a more thermodynamically stable bound form than 4-4-20, is much more affected by the presence of the ligand in the refolding buffer. Another piece of evidence for this is

4M5.3 is more affected by the increase in hold time at the elevated pressures, where we see an approximately 3-fold increase in yield by increasing the hold time from 5 to 14 days at 40kpsi, whereas, there is little or no difference in yield for 4-4-20 when held longer at 40kpsi. Our experiments were limited by the maximum pressure of 40kpsi on our equipment. Perhaps by conducting the experiments at higher hydrostatic pressures, our yields for the monomer form would increase and increasing pressure may also decrease our processing time.

Of the two methods explored to remove the bound ligand, the CDF method yielded the most free monomer. This result is likely due to the fast $k_{association}$ of the ligand to the scFv where in the SEC method, there is not enough time for the gel matrix to separate the released ligand from the free monomer upon injection, before the concentration of guanidine hydrochloride falls to a level that allows the ligand to rebind to the protein. In the CDF method, the liberated ligand is washed away before the guanidine concentration is brought down to non-denaturing levels. The yields of less than 1mg/L of culture obtained by the methods explored in this study is much less than the reported yields in the literature that report from 10-100mg of active protein per liter of bacterial culture (5, 6, 7, 9).

Determination of the binding characteristics for 4M5.3 and 4-4-20 gives an indication of how well the protein is refolded using the process described in this study. The dissociation equilibrium constants obtained using the methods described here are greater than the published values for the scFvs 4-4-20 and 4M5.3 suggesting that the

scFvs produced by this method are of lower quality, in terms of its ability to bind to their ligand. Both scFvs produced bind fluorescein slower and release the ligand more readily than previously determined by other methods. However, the methods explored here may lend to refolding difficult to obtain proteins that may require the presence of its ligand in order to refold to the monomer state.

The methods explored here show that pressure treatment with ligand followed by stripping with a low level of denaturant can produce active forms of the monomer from highly aggregated inclusion bodies, albeit, with lower yields than traditional methods using high concentrations of denaturant to solubilize IBs followed by dialysis, dilution or dilution and filtration.

Chapter 4

CHARACTERIZATION OF THE PROTEIN FOLDING PATHWAY FOR 4M5.3

In-vitro protein refolding methods are notorious for producing low yields of protein folded to the native state with most of the protein going to an aggregated or misfolded state, which is evident in the results presented in this thesis and in other examples in the literature (5, 6, 7, 9). The rate at which in-vitro protein folding occurs may be related to the yield of productive in-vitro folding to the native state, where a protein that folds relatively fast compared to diffusive rates may avoid aggregation compared to a slowly folding protein, which may need to be folded in more dilute environments to achieve equivalent yields. The rate at which a protein can fold in-vitro can be controlled by the rate of dilution or the rate of the chaotrope and/or reducing agent removal by dilution and filtration when using chemical refolding methods, or the rate of pressure release when using hydrostatic pressure. In this study, the in-vitro refolding rates by dilution from a chaotrope for anti-fluorescein scFvs 4M5.3 and 4-4-20 were indirectly determined by recovery of activity (fluorescein quenching rates) and also by measuring Trp fluorescence quenching upon folding. These data were used to compare to the published refolding rate for 4-4-20 and possibly use the determined refolding rate for improving yields during hydrostatic pressure refolding by matching the rate of pressure release to the scFv rate of folding.

Introduction

The folding and binding of an scFv to its ligand can be represented by equation 4.1 where U represents the concentration of unfolded scFv, F is the concentration of folded protein, L is the concentration free ligand and B is the concentration of the scFv ligand complex. k_F, k_U, k_a and k_d are the rate constants for the rate of folding, unfolding, ligand association and ligand dissociation, respectively. Indirect measurement of the scFv folding rate by fluorescein quenching was conducted by assuming the following:

1. The rate of ligand association ($8 \times 10^7 M^{-1}s^{-1}$ for 4-4-20 (2) and $5.1 \times 10^5 M^{-1}s^{-1}$ for 4M5.3 (3)) is much greater than the rate of protein folding, which is a reasonable assumption based on the folding rates of other proteins in the literature. (*32*)
2. The rate of folding is much greater than the rate of unfolding for a stable protein, which is true for these scFvs in a non-denaturing environment. (*32*)
3. The rate of ligand association is much greater than the rate of dissociation as determined previously by experiment in chapter 2.

The equation that represents the folding and binding of an scFv to its ligand can therefore be from simplified 4.1 to 4.2:

$$U \quad \Leftrightarrow \quad F + L \quad \Leftrightarrow \quad B \tag{4.1}$$

$$U \xrightarrow{k_F} F \tag{4.2}$$

The rate of folding constant can then be determined by using equations 4.3 to 4.6:

$$\frac{d(F(t))}{dt} = k_F U(t) \tag{4.3}$$

$$U(t) = U_0 - F(t) \tag{4.4}$$

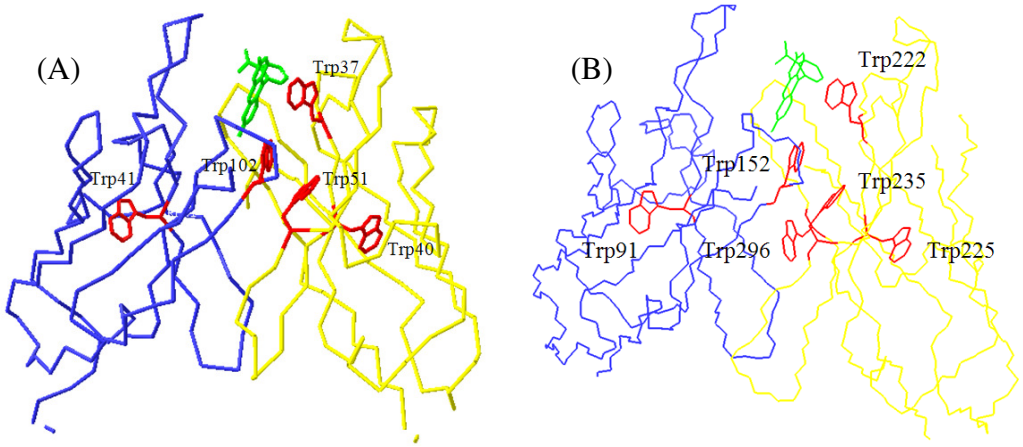
$$F(t) = \frac{1}{m}(I_0 - I(t)) \tag{4.5}$$

$$m = 1 \times 10^{11} M^{-1} \tag{4.6}$$

F is the folded protein concentration, U is the unfolded concentration and I is the fluorescence intensity. The decrease of fluorescence intensity ($I_0 - I(t)$) versus time was used to obtain the rate of folding. m was determined by the gradient of a standard curve of fluorescence intensity versus fluorescein concentration

In addition to measuring the folding rate of the entire scFv, measuring the rate of domain folding and interface formation was also attempted by creating tryptophan to phenylalanine mutants. Tryptophan has an absorbance maxima at 280nm and a fluorescence maxima at 340nm which is quenched when the tryptophan environment becomes solvated (*33*). Therefore, we can measure the rate of folding by measuring the rate of increase in tryptophan fluorescence as the domain folds or as the interface

between the 2 domains form. Analysis by this method requires that tryptophan residues are present in the domain being studied, while mutating the other tryptophans in the molecule to phenylalanine. Figures 13 (A) and (B) show the tryptophan locations in both 4M5.3 and 4-4-20 scFvs.



Figures 13 (A) and (B). Tryptophan locations in 4M5.3 (A) and 4-4-20 (B) are highlighted in red in the figures above. Molecular renditions for 4M5.3 and 4-4-20 were generated using x-ray crystallography coordinates from PDB ID 1X9Q and 1FLR, respectively, using Swiss-PDB Viewer (www.expasy.org/spdbv). Blue and yellow signifies the light and heavy chains, respectively. Shown in green is the fluorescein ligand in the binding pocket.

Table 4 summarizes the mutants for 4M5.3 and 4-4-20 that were created in order to measure the folding kinetics of each domain in the scFv. Trp37 in 4M5.3 and Trp322 in 4-4-20 was not mutated because it was hypothesized that fluorescence quenching activity may be affected if the Trp that appears to be interacting with fluorescein because of its proximity was mutated.

Table 4. Summary of mutants created for 4M5.3 and 4-4-20 to measure the unfolding/folding rates of the light chain, heavy chain in each of the scFv's.

Mutant	Tryptophans Mutated to Phenylalanine	Description
4M5.3 TrpV_H	41, 51, 102	Trp40 left in the core of the heavy chain used to track the folding/unfolding rate of the heavy chain.
4M5.3 TrpV_L	40, 51, 102	Trp41 left in the core of the light chain used to track the folding/unfolding rate of the light chain.
4M5.3 Trp_I	40, 41, 51	Trp102 left at the interface used to track the folding/unfolding rate of the interface.
4-4-20 TrpV_H	91, 235, 296	Trp225 left in the core of the heavy chain used to track the folding/unfolding rate of the heavy chain.
4-4-20 TrpV_L	225, 235, 296	Trp91 left in the core of the light chain used to track the folding/unfolding rate of the light chain.
4-4-20 Trp_I	91, 225, 296	Trp235 left at the interface used to track the folding/unfolding rate of the interface.

Materials and Methods

Creation of 4M5.3 and 4-4-20 Tryptophan Mutants

4M5.3 and 4-4-20 genes were inserted into the pET21-d plasmid vector (Novagen) and selected tryptophan residues were mutated to phenylalanine using the Quick-Change II Mutagenesis Kit (Stratagene) using oligonucleotide primers (Operon) listed in. Mutagenesis was verified by DNA sequencing (University of Pennsylvania) via Edman degradation.

Table 5. Table of primers used to generate 4M5.3 and 4-4-20 mutants used to measure the unfolding/folding rates of different domains in each scFv.

Mutant	Tryptophans Mutated to Phenylalanine	Primers Used
4M5.3 $TrpV_H$	41, 51, 102	W41F primer: TATTTACGTTTCTACCTGCAGAAGCC W51F primer: GTACACATGTTCCGTTCACGTTCGGTGGAGGC W102F primer: CCAGAGAAAGGACTGGAGTTCGTAGCACAATTTAG
4M5.3 $TrpV_L$	40, 51, 102	W40F primer: GGATGAACTTCGTCCGCCAGTCTCC W51F primer: GTACACATGTTCCGTTCACGTTCGGTGGAGGC W102F primer: CCAGAGAAAGGACTGGAGTTCGTAGCACAATTTAG
4M5.3 Trp_I	40, 41, 51	W40F primer: GGATGAACTTCGTCCGCCAGTCTCC W41F primer: TATTTACGTTTCTACCTGCAGAAGCC W51F primer: GTACACATGTTCCGTTCACGTTCGGTGGAGGC
4-4-20 $TrpV_H$	91, 235, 296	W91F primer : CAGTAATGGAAACACCTATTTACGTTTCTACCTGCAGAAGCC W235F primer: CTCCAGAGAAAGGACTGGAGTTCGTAGCACAAATTAGAAACAAA W296F primer: GGTTCTTACTATGGTATGGACTACTTCGGTCAAGGAACCTC
4-4-20 $TrpV_L$	225, 235, 296	W225F primer: TTAGTGACTACTGGATGAACTTCGTCCGCCAGTCTCC W235F primer: CTCCAGAGAAAGGACTGGAGTTCGTAGCACAAATTAGAAACAAA W296F primer: GGTTCTTACTATGGTATGGACTACTTCGGTCAAGGAACCTC
4-4-20 Trp_I	91, 225, 296	W91F primer : CAGTAATGGAAACACCTATTTACGTTTCTACCTGCAGAAGCC W225F primer: TTAGTGACTACTGGATGAACTTCGTCCGCCAGTCTCC W296F primer: GGTTCTTACTATGGTATGGACTACTTCGGTCAAGGAACCTC

Expression of 4M5.3 and 4-4-20 scFv Inclusion Bodies

4M5.3 and 4-4-20 genes were inserted into the pET21-d plasmid vector (Novagen) and were expressed under an inducible *lac* promoter. Inclusion bodies of

each scFv were made in 2L shake flasks where cell density was allowed to reach an OD between 0.6 and 0.8 prior to induction of protein expression using IPTG. The culture is then allowed to reach saturation between 4-8 hours before collection of the cell pellet via centrifugation for 30min at 3176 x g.

Isolation of Inclusion Bodies and Inclusion Body Solubilization

Bacterial cells were lysed by three freeze-thaw cycles (1hr at -80ºC followed by 1hr at 37ºC and incubation with lysozyme at a 0.001mg/ml concentration. Endogenous DNA was digested by treatment with DNAse (0.1mM final conc) at 37ºC for 1hour. Inclusion bodies, which were part of the insoluble material, were pelleted by centrifugation at 3176xg for 30min. The soluble fraction was decanted and the insoluble fraction was resuspended with tris-buffered saline (TBS, 50mM Tris-HCl, 150mM NaCl, pH8.0). The resuspended pellet was centrifuged again at 3176xg for 30min and the soluble fraction was then decanted and discarded. Inclusion bodies were solubilized using 6M guanidine HCl, followed by overnight dialysis against 3M Guanidine hydrochloride, and finally overnight dialysis against phosphate-buffered saline (PBS; 150mM NaCl, 10 mM sodium phosphate buffer, pH 8), as described previously (*15*).

scFv Purification

A 350 ml prep grade Superdex 75 size-exclusion chromatography (SEC) column equilibrated with PBS was used to isolate the folded scFv species. Sample

concentrations were determined by absorbance readings at 280 nm using a molar extinction coefficient of 46,850. Samples were stored at 4 °C prior to analysis. The activity of each fraction was determined by fluorescence quench titration.

Fluorescein Quenching Activity Assay

A Hitachi 4500 fluorimeter (excitation at 490 nm and emission at 512 nm) equipped with a temperature controller and stirrer was used to measure the fluorescent intensity of 2ml of PBS buffer containing 20nM fluorescein. This fluorescein solution was titrated with small additions of folded scFv of known concentration (5-10 □l) and the decrease in emission intensity was monitored to determine protein activity. Active protein was titrated until no further change in intensity was observed. The fluorescein solution was also titrated with buffer containing no scFv to correct for dilution effects on fluorescein intensity. Titrations were repeated several times and the values were then fit to a least-squares method for single site binding according to equations 4.6 to 4.8:

$$I = m([F] + (1 - Q_{max})[B]) \quad (4.6)$$

$$K_D = \frac{([P_T]-[B])}{[B]} \quad (4.7)$$

$$[F_T] = [F] - [B] \quad (4.8)$$

where *I* is the measured intensity at equilibrium, *m* is a constant determined from a standard curve of fluorescein concentration versus intensity, [F] is the free

fluorescein concentration, Q_{max} is the fluorescence quenching constant (between 0.95 and 0.98), [B] is the bound fluorescein-scFv complex, K_D is the equilibrium dissociation constant, $[P_T]$ is the total protein concentration in the system at that equilibrium point, and $[F_T]$ is the total fluorescein concentration in the system at that equilibrium point.

The three equations were solved at each equilibrium point by using a three-parameter search model (K_D, Active (bound) scFv concentration [B], and Q_{max}) to solve for K_D as described previously (*11, 15*).

Macromolecular Rate of Folding Assays

A Hitachi 4500 fluorimeter equipped with a temperature controller and stirrer was used to measure folding rates by measuring fluorescence intensity of fluorescein (excitation at 490 nm and emission at 512 nm) with time. The fluorescence of 2ml of PBS buffer containing 20nM fluorescein was measured after adding 200nM of unfolded scFv in 6M Guanidine Hydrochloride, TBS, pH8.0 in a 10ul volume. The decrease in fluorescent intensity is measured with time. Rate data was fit to a first-order rate equation to solve for the rate of folding constant, k_F (Equation 4.9) using MATLAB.

$$\frac{d(F(t))}{dt} = k_F U(t) \tag{4.9}$$

The rate of scFv folding was also measured by intrinsic tryptophan fluorescence. The scFv was first unfolded by incubation for 10min in 6M Guanidine Hydrochloride, TBS (pH8.0) and then refolded by dilution into TBS at pH8.0. During refolding, the increase in intrinsic tryptophan fluorescence (excitation at 280nm and emission at 340) with time was measured using a hitachi 4500 fluorimeter. Rate data was fit to a first-order rate equation to solve for the rate of scFv folding (Eq 4.3).

scFv Domain Rate of Folding Assay

To measure the folding rates of the different domains of 4M5.3 and 4-4-20, the purified tryptophan knockout mutants were first denatured in 6M Guanidine Hydrochloride, TBS, pH8.0. The denatured protein was then refolded by dilution by adding 100µl of the denatured protein (~0.05mg/ml) to 2ml of TBS. During refolding by dilution, the increase in intrinsic tryptophan fluorescence (excitation at 280nm and emission at 340) with time was measured using a hitachi 4500 fluorimeter. Rate data was fit to a first-order rate equation to solve for the rate of folding constant for each of the Trp mutants.

Results

Macromolecular Rate of scFv Folding

Initial experiments that attempted to measure the rate of productive scFv folding from a denatured state by dilution using fluorescein fluorescence quenching are shown in figure 14.

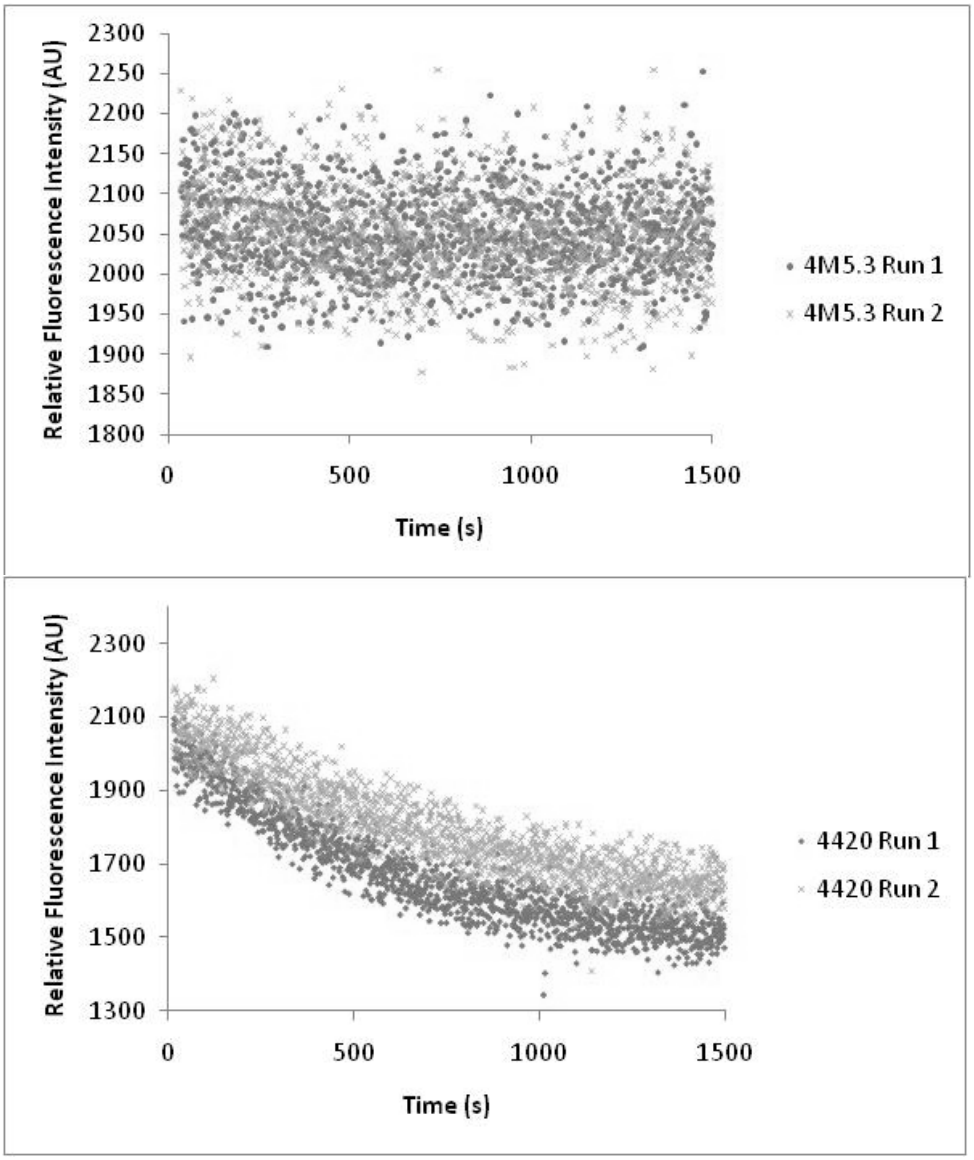


Figure 14. The graphs above show fluorescence quenching of fluorescein by refolding denatured scFv by dilution. The rate at which the fluorescence intensity decreases is also the rate at which active scFv is folded from a denatured state. Two independent refolding runs are shown for 4M5.3 and 4-4-20.

Figure 14 shows that 4M5.3 fluorescence intensity decrease plateaus at around 500s where as the plateau for 4-4-20 is at around 1500s. There is also a difference in the total decrease in fluorescence intensity where the total decrease in intensity for

4M5.3 is around 100AU and for 4-4-20 the decrease in intensity is around 500AU, even though equimolar amounts of denatured scFv were added to the reaction sample. The rate of folding constants ($k_{folding}$) for 4M5.3 and 4-4-20 were determined by fitting the data shown in figure 14 to a single-exponential rate equation. These values are shown in Table 6. A rate of 0.00074 s^{-1} was previously experimentally determined for 4-4-20 using tryptophan fluorescence (*31*) compares fairly well with the average value of 0.0000184 determined for 4-4-20 using the method described herein. There is currently no refolding rate published for 4M5.3. 4M5.3 had a faster calculated rate of refolding compared to 4-4-20.

Table 6. Summary of rate of folding constants for 4M5.3 and 4-4-20 as determined by fluorescence quenching of fluorescein rate data. These values were calculated by fitting the rate of fluorescence decrease data to a single-exponential rate equation.

	$k_{folding}$ for Experimental Run (Ms^{-1})		Average $k_{folding}$ (Ms^{-1})
	Run 1	**Run 2**	
4M5.3	1.57×10^{-7}	2.82×10^{-7}	2.19×10^{-7}
4-4-20	2.12×10^{-5}	1.56×10^{-5}	1.84×10^{-5}

In-vitro Refolding of Trp Mutants

Following successful mutagenesis of the 4M5.3 and 4-4-20 genes to generate tryptophan mutants that can be used to track the folding of a particular domain, each gene was transformed into *E. coli* and expressed as insoluble inclusion bodies. The mutagenesis was confirmed by DNA sequencing (data not shown). Size-exclusion purification of the active scFvs were conducted after solubilization and refolding of the inclusion bodies. The chromatograms from those runs are shown in figures 3 and 4.

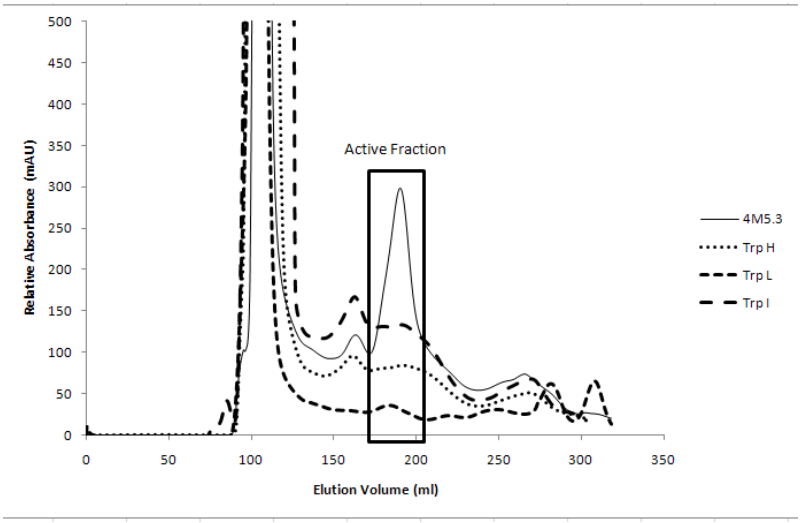


Figure 15. SEC chromatograms for 4M5.3 Trp mutants which were expressed as inclusion bodies in E. coli, chemically solubilized, renatured by dialysis and purified by size-exclusion chromatography.

The majority of the protein for the tryptophan mutants of 4M5.3 and 4-4-20 after inclusion body solubilization and chemical refolding are in the form of aggregate (~80-90%) which appears as a large peak at around the ~110ml elution volume. The

active protein fractions from size-exclusion chromatography that elute at ~190ml for 4M5.3 and at ~150ml for 4-4-20 were confirmed by fluorescence quench titration assays.

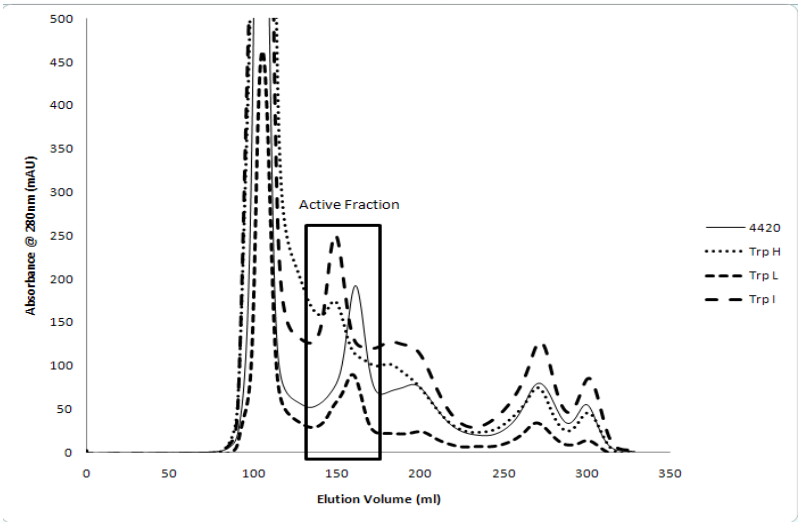


Figure 16. SEC chromatograms for 4-4-20 Trp mutants which were expressed as inclusion bodies in E. coli, chemically solubilized, renatured and purified by size-exclusion chromatography.

Tracking Trp Mutant Rate of Folding by Trp Fluorescence

After isolation of the active fractions for each of the mutants, attempts to measure the refolding rate of the VL, VH and interface domains were made by tracking the intrinsic fluorescence of the remaining tryptophans in the mutants with time. Figures 5 and 6 show the rate of change in fluorescence intensity versus time for 4M5.3 and 4-4-20 and their respective TrpH, TrpL and TrpI mutants. All, except 4M5.3, showed a decrease in fluorescence upon refolding by dilution which suggests that the remaining tryptophans are not burying into a hydrophobic environment. The

expected result is an increase in fluorescence intensity with time as the tryptophans get buried into the hydrophobic core of the protein. A folding rate cannot be calculated based on the decreasing fluorescence rate data shown in figures 17 and 18.

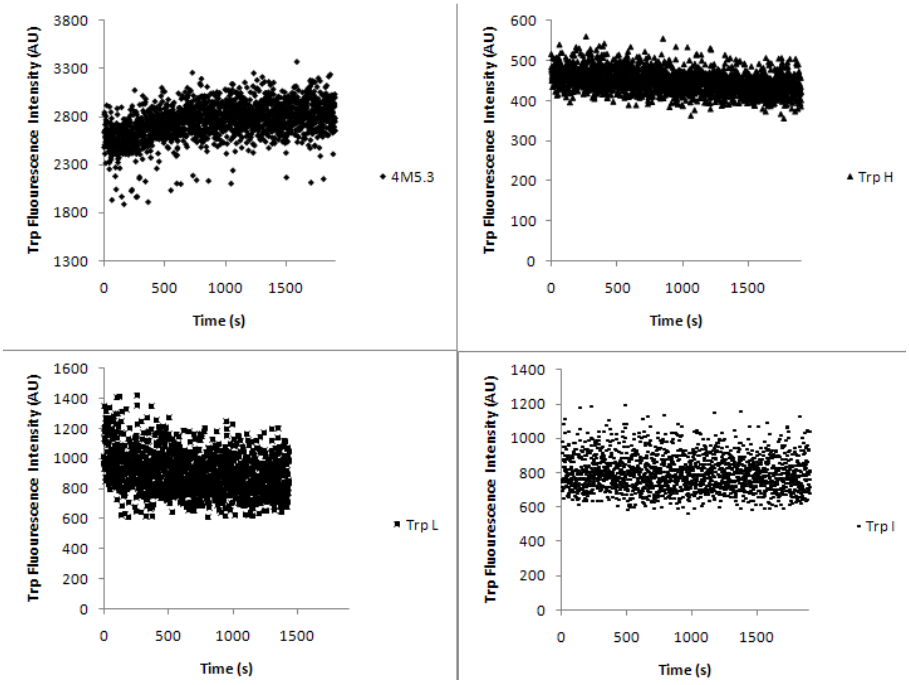


Figure 17. Graphs of Trp fluorescence versus time for 4M5.3 and its Trp mutants.

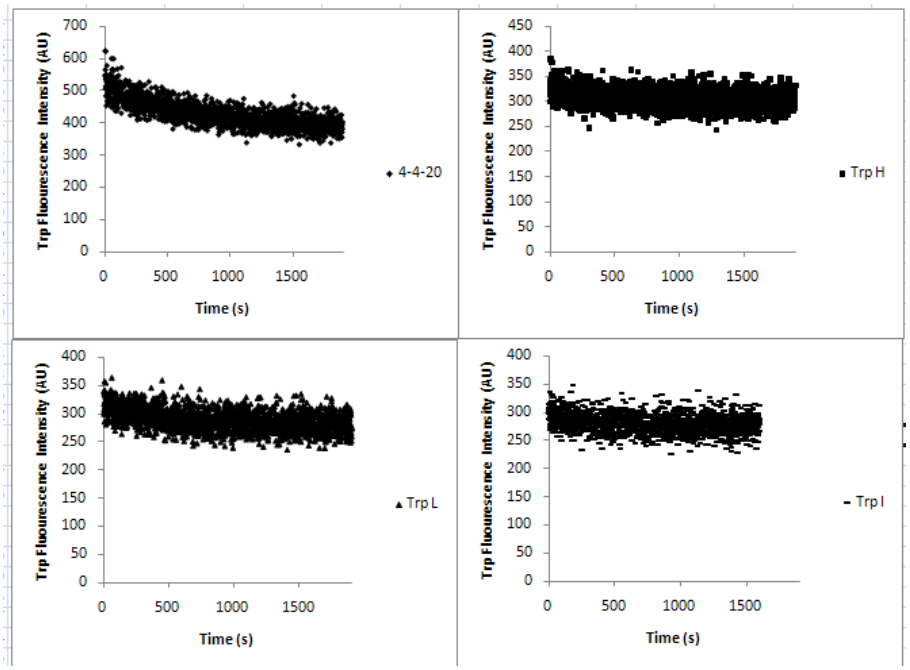


Figure 18. Graphs of Trp fluorescence versus time for 4-4-20 and its Trp mutants.

A possible explanation for the decreasing tryptophan fluorescence during refolding by dilution is aggregation which can be tracked by measuring the fluorescence intensity of the sample with time at an excitation wavelength of 350nm and an emission wavelength of 350nm. An increasing fluorescence intensity is indicative of a protein aggregation event occurring during the refolding. These measurements for the tryptophan mutants of 4M5.3 and 4-4-20 are shown in figures 19 and 20. 4M5.3 Trp H and Trp I show an increase in fluorescence intensity at 350nm, which is indicative of an aggregation event, but Trp L does not. For the tryptophan mutants of 4-4-20 only Trp I shows an increase in fluorescence intensity at 350nm. For the mutants that do not show an increase in fluorescence intensity at

350nm and shows no increase in Trp fluorescence duing refolding, some other unproductive misfolding event is occurring for these mutants other than aggregation.

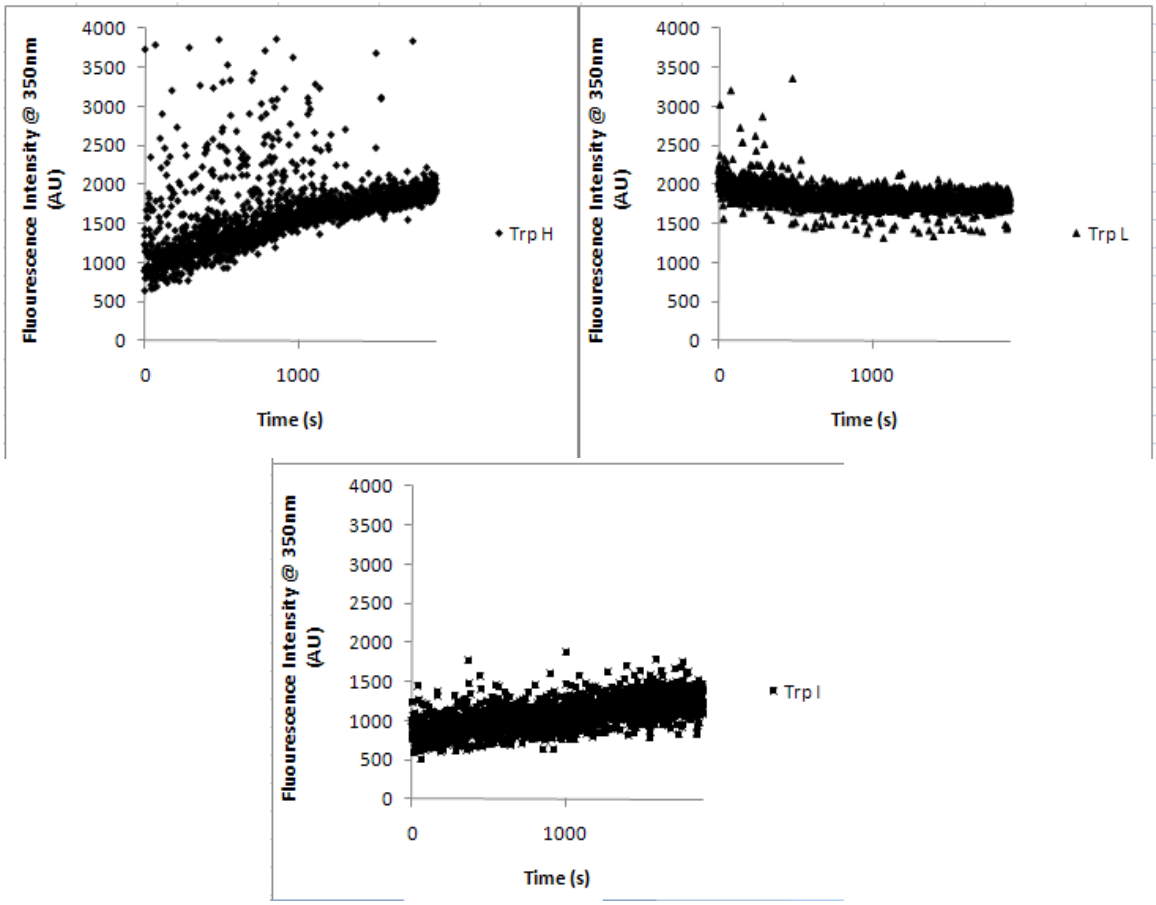


Figure 19. Graphs showing emission at 350nm versus time for 4M5.3 Trp mutants. The sample was excited with light also at 350nm. An increase in fluorescence intensity is indicative of aggregation.

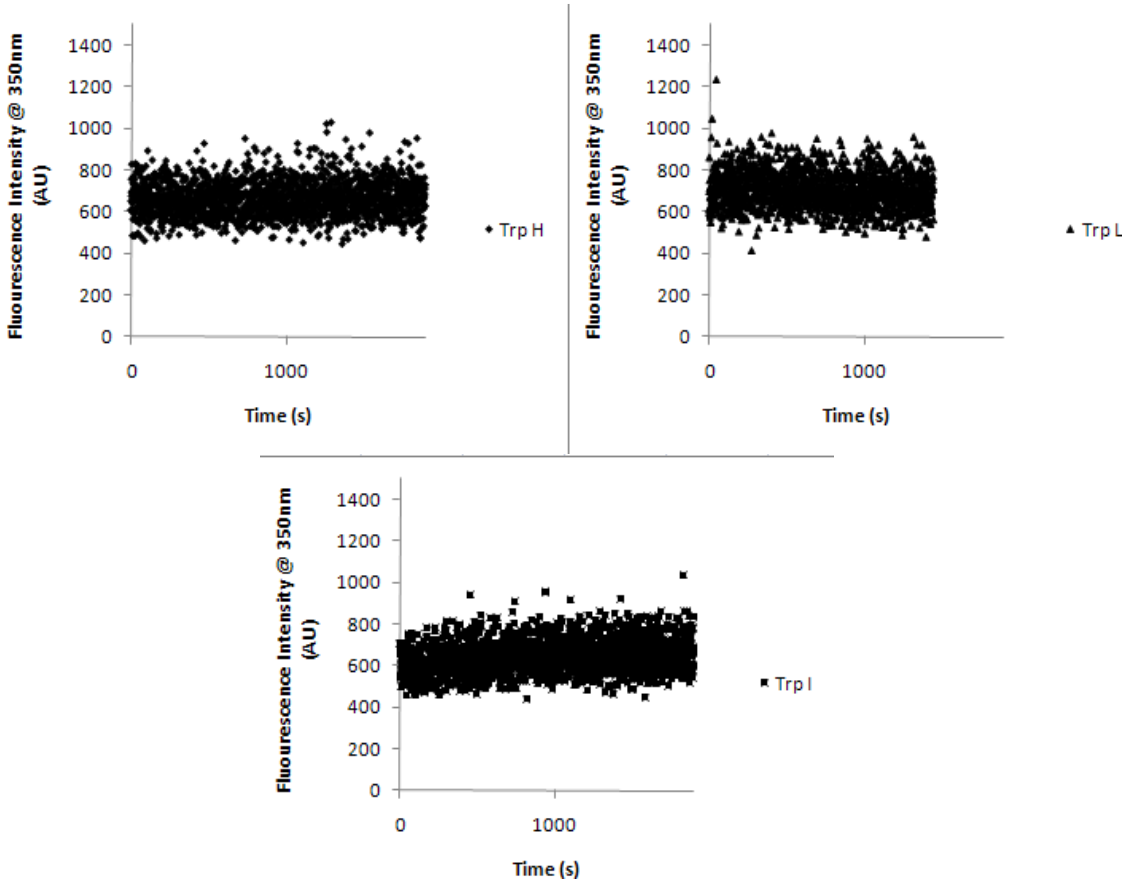


Figure 20. Graphs showing emission at 350nm versus time for 4-4-20 Trp mutants. The sample was excited with light also at 350nm. An increase in fluorescence intensity is indicative of aggregation.

Discussion

The refolding rate of $0.0000184s^{-1}$ for 4-4-20 determined by fluorescence quench titration compares fairly well with the published value of $0.00074s^{-1}$ which suggests that the value determined for productive folding of 4M5.3 of $2.19 \times 10^{-7}\ s^{-1}$ is probably accurate to within an order of magnitude. The method described here gives an alternative method to intrinsic tryptophan fluorescence for measuring the rate of scFv folding. The decrease in signal for 4M5.3 was not as great as the decrease in

signal for 4-4-20 even though the initial protein concentrations prior to refolding were equal. This is probably due to unproductive folding pathways such as aggregation or misfolding that 4M5.3 is more prone to than 4-4-20. It is also possible that the time-scales investigated were not sufficient if protein folding is slow.

Generation of scFv Trp mutants of 4M5.3 and 4-4-20 from inclusion bodies was successful. However, tracking the intrinsic fluorescence with time to determine a rate of folding for each scFv domain was not achieved because an increase in fluorescence was not seen for any of the mutants upon refolding by dilution. One possible explanation for the decrease in intrinsic tryptophan fluorescence of all the Trp mutants and 4-4-20 may be aggregation upon dilution. An increase in fluorescence intensity at an emission and excitation wavelength of 350nm is indicative of an aggregation event. Both Trp H and Trp I for 4M5.3 showed an increase in fluorescence at 350nm which suggests that they reaggregate upon refolding by dilution. Trp L for 4M5.3 showed a decrease in fluorescence at 350nm. Perhaps, some other unproductive misfolding event, such as precipitation out of solution, is occurring other than aggregation. Only 4-4-20 Trp I showed an increase in fluorescence intensity at 350nm which suggests that 4-4-20 Trp I may aggregate upon refolding by dilution. 4-4-20 Trp L and Trp H may misfold in some other pathway other than aggregation because there is no increase in fluorescence intensity at 350nm and no increase in intrinsic tryptophan fluorescence.

Experiments to determine the optimal conditions for Trp mutant refolding by dilution will need to be conducted so that the rate of increase in intrinsic tryptophan fluorescence can be measured and, hence, a rate of folding for that domain. Factors that affect protein folding such as protein concentration, pH, temperature, ionic strength and additional folding additives will need to be evaluated in its effect to effectively allow 4M5.3 and 4-4-20 Trp mutants to refold by dilution. Once adequate refolding conditions for 4M5.3 and 4-4-20 Trp mutants have been determined, the refolding rate of each domain can be determined by measuring the increase of intrinsic fluorescence, as the tryptophans bury themselves into the hydrophobic cores of the folded scFv.

Chapter 5

SUMMARY OF CONCLUSIONS AND SUGGESTIONS ON FUTURE WORK

Upon chemical refolding of 4M5.3 from inclusion bodies and purification via SEC, four distinct species of 4M5.3 were identified as aggregate, an inactive monomer (I), and two active monomers (A1 and A2), in order of larger to smaller hydrodynamic radius. When reducing agent and disulfide shufflers were not used during refolding, the inactive species disappears suggesting that (I) may be a foldamer with incorrect intra-domain disulfides because correct inter-domain disulfides are more likely formed as the protein is translated from the ribosome due to protein sequence and it can be inferred from the structure of 4M5.3 (figure 2) that a species with intra-domain disulfides will have a larger hydrodynamic radius, which is supported by the SEC data. A1 was found to have slower and looser binding kinetics to fluorescein-biotin than A2 via binding characterization assays. A2 also showed slower and looser binding kinetics compared to published values but are within an order of magnitude, which could be explained by differences in scFv expression, purification and assay methodology.

Possible structural differences between A1 and A2 were investigated using different biophysical assays, in order to explain the difference in hydrodynamic radius and activity. Reduced SDS-PAGE and mass spectrometry revealed that A1 and A2 have the same molecular weight, even though A1 shows a larger hydrodynamic radius than A2 as seen by SEC. N-terminal sequencing by Edman degradation showed that

both A1 and A2 have the same N-terminal sequence. Free disulfides were detected in neither A1 nor A2 as reported by a sensitive free-disulfide quantification assay. There were differences detected in surface hydrophobicity between A1 and A2 as reported by a fluorometric assay using ANS suggesting a difference in protein conformation, which is shown on SEC.

Previous work has shown that the thermodynamics of conversion from A1 to A2 is similar to cis/trans conversion of Pro for small peptides, suggesting that the difference between A1 and A2 are cis/trans isomers at a certain Pro residue. Pro100 on the light chain of 4M5.3 was targeted for mutation studies based on this data and reported data that suggests this residue is important for activity in other scFvs. Mutational studies that would favor the cis form of the molecule were carried out. None of the mutations to 4M5.3 succeeded in only forming the cis form (A2) of the molecule. Suggesting that there may be another cis/trans Pro residue responsible for the difference in A1 and A2 or some other cause for the structural difference, such as, differences in salt-bridges, hydrogen bonding and hydrophobic interactions.

The most straightforward approach to elucidate the structural differences between A1 and A2 would be to purify enough A1 and A2 that have been radio-labeled (by growing them in *E. coli* with radio-labeled media) by SEC, after chemical refolding to a suitable concentration for nuclear magnetic resonance (NMR) work. Generating enough radio-labeled A1 and A2 would allow one to obtain an average of atomic coordinates for each atom in the protein molecule in solution and solve its

structure by NMR. X-ray crystallography is another option for solving the structures of A1 and A2 but the crystallization process might affect the conformation of A1 and A2 and both molecules might not retain the original structural differences that were there before crystallization.

Chapter 3 of this thesis covered the use of hydrostatic pressure as an alternative for solubilizing and refolding scFvs from IBs. The conditions investigated were able to produce active scFvs that bind its ligand. The addition of ligand during the hydrostatic pressure refolding process greatly increases the yield. This supports the hypothesis that the addition of ligand acts as a template for the scFv during refolding. It was also observed that the scFv (4M5.3) with the higher affinity for its ligand generated more scFv as compared to the the scFv with a lower affinity (4-4-20) when refolded with its ligand and when incubated at longer times with ligand.

Two methods for obtaining the free protein pressure refolding were investigated. The CDF method of treating the bound scFv complex with low level of denaturant, filtering away the ligand, then exchanging the buffer to remove the denaturant, was the better process and was able to produce approximately 1mg of active protein per liter of bacterial culture. Assays that measured the binding kinetics of the scFvs produced by this method suggested that the scFvs were of lower quality where ligand binding values were slower and looser compared to published values and to values obtained by chemical refolding for 4M5.3 in Chapter 2.

Further work in improving the yields for this 2 step-process (protein solubilization and refolding by hydrostatic pressure, ligand-stripping by CDF) could involve a design of experiments (DOE) approach to identify the important factors, such as, pressure, time of incubation, rate of pressure release, buffer conditions, pH and temperature for the hydrostatic pressure step using active scFv yield as the output to measure the importance of each factor and the interactions between factors. Possible important factors for the ligand-stripping could be denaturant concentration, rate of addition and removal of denaturant, pH and buffer conditions using active scFv yield and binding characteristics as the outputs to measure the importance of each factor and the interactions between factors. Once the important factors for each unit operation are identified one could use a surface response methodology to optimize each factor to maximize the desired outputs, which are yield and binding characteristics.

Methods for measuring the macro-molecular and scFv domain rate of folding were investigated in chapter 5. The determination of the rate of refolding for 4M5.3 and 4-4-20 by dilution and measuring the rate of fluorescence quenching was successful. This method would be useful for measuring the rate of refolding for very small amounts of protein or if the Trp residues in the molecule were removed, since there was a way to track the rate of active protein appearance from unfolded protein. Trp to Phe mutations were made to 4M5.3 and 4-4-20 with the strategy of creating mutants that would have a fluorescent Trp residue in the domain or interface of

interest as a tracker for the rate of folding for that domain. Trp mutants for each scFv were successfully created as verified by DNA sequencing and their activity verified by fluorescence quenching assay. However, we were unsuccessful at measuring a rate of refolding for the domains and interfaces of 4M5.3 and 4-4-20 using a rapid dilution method and tracking Trp fluorescence. The rapid dilution method was unsuccessful at making adequate amounts of refolded protein to detect enough signal from Trp fluorescence.

A DOE approach to finding the critical factors in the refolding by rapid dilution method could be carried out to identify the factors and the interactions of factors that affect the total amount of scFv refolded. Critical factors in the refolding by rapid dilution method could be protein concentration, pH, temperature, buffer concentration. Once the critical factors are identified, they could be optimized using a response surface model to find the best conditions for refolding each of the Trp mutants, so that a rate of refolding could be determined for each domain and interface for 4M5.3 and 4-4-20. This rate data could be useful for optimizing the rate denaturant removal or hydrostatic pressure release to optimize the yield for these refolding methods.

REFERENCES

1. Leader, B., Baca, Q. J., and Golan, D. E. (2008) Protein therapeutics: a summary and pharmacological classification, *Nat Rev Drug Discov 7*, 21-39.

2. Shukla, A. A., and Thömmes, J. Recent advances in large-scale production of monoclonal antibodies and related proteins, *Trends in Biotechnology 28*, 253-261.

3. Holliger, P., and Hudson, P. J. (2005) Engineered antibody fragments and the rise of single domains, *Nat Biotech 23*, 1126-1136.

4. Klaus, G., and Andreas, P. (2006) Manufacturing of recombinant therapeutic proteins in microbial systems, *Biotechnology Journal 1*, 164-186.

5. Singh, S. M., and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins, *Journal of Bioscience and Bioengineering 99*, 303-310.

6. Burgess, R. R., Richard, R. B., and Murray, P. D. (2009) Chapter 17 Refolding Solubilized Inclusion Body Proteins, in *Methods in Enzymology*, pp 259-282, Academic Press.

7. Clark, E. D. B. (2001) Protein refolding for industrial processes, *Current Opinion in Biotechnology 12*, 202-207.

8. Manning, M. C., Patel, K., and Borchardt, R. T. (1989) Stability of Protein Pharmaceuticals, *Pharmaceutical Research 6*, 903-918.

9. Jungbauer, A., and Kaar, W. (2007) Current status of technical protein refolding, *Journal of Biotechnology 128*, 587-596.

10. Denzin, L. K., and Voss, E. W., Jr. (1992) Construction, characterization, and mutagenesis of an anti-fluorescein single chain antibody idiotype family, *J. Biol. Chem. 267*, 8925-8931.

11. Boder, E. T., Midelfort, K. S., and Wittrup, K. D. (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity, *Proceedings of the National Academy of Sciences of the United States of America 97*, 10701-10705.

12. Midelfort, K. S., Hernandez, H. H., Lippow, S. M., Tidor, B., Drennan, C. L., and Wittrup, K. D. (2004) Substantial Energetic Improvement with Minimal

Structural Perturbation in a High Affinity Mutant Antibody, *Journal of Molecular Biology 343*, 685-701.

13. Midelfort KS, W. K. (2006) Context-dependent mutations predominate in an engineered high-affinity single chain antibody fragment, *Protein Sci. 15*, 10.

14. Sun, W.-C., Gee, K. R., Klaubert, D. H., and Haugland, R. P. (1997) Synthesis of Fluorinated Fluoresceins, *The Journal of Organic Chemistry 62*, 6469-6475.

15. Sinacola, J. R., and Robinson, A. S. (2002) Rapid refolding and polishing of single-chain antibodies from Escherichia coli inclusion bodies, *Protein Expression and Purification 26*, 301-308.

16. Sinacola, J. R. (2003) Anti-Fluorescein Single-Chain Antibodies: Characterization of Multiple Active Conformations and Maximization of Functional Yield, in *Chemical Engineering*, University of Delaware, Newark.

17. Michael Gross, R. J. (1994) Proteins under pressure, *European Journal of Biochemistry 221*, 617-630.

18. Ryan L. Crisman, T. W. R. (2009) Refolding of proteins from inclusion bodies is favored by a diminished hydrophobic effect at elevated pressures, *Biotechnology and Bioengineering 102*, 483-492.

19. Schoner, B. E., Bramlett, K. S., Guo, H., and Burris, T. P. (2005) Reconstitution of functional nuclear receptor proteins using high pressure refolding, *Molecular Genetics and Metabolism 85*, 318-322.

20. Chura-Chambi, R. M., Genova, L. A., Affonso, R., and Morganti, L. (2008) Refolding of endostatin from inclusion bodies using high hydrostatic pressure, *Analytical Biochemistry 379*, 32-39.

21. Susanna, S. J. L., and Anton, P. J. M. (2007) A simplified bioprocess for human alpha-fetoprotein production from inclusion bodies, *Biotechnology and Bioengineering 97*, 99-117.

22. Cabrita, L. D., Bottomley, S. P., and El-Gewely, M. R. (2004) Protein expression and refolding - A practical guide to getting the most out of inclusion bodies, in *Biotechnology Annual Review*, pp 31-50, Elsevier.

23. Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies, *Protein Expression and Purification 28*, 1-8.

24. Ventura, S., and Villaverde, A. (2006) Protein quality in bacterial inclusion bodies, *Trends in Biotechnology 24*, 179-185.

25. Jäger, M., Gehrig, P., and Plückthun, A. (2001) The scFv fragment of the antibody hu4D5-8: Evidence for early premature domain interaction in refolding, *Journal of Molecular Biology 305*, 1111-1129.

26. Freund, C., Gehrig, P., Baici, A., Holak, T. A., and Plückthun, A. (1997) Parallel pathways in the folding of a short-term denatured scFv fragment of an antibody, *Folding & Design 3*, 39-49.

27. Jäger, M., and Plückthun, A. (1999) Folding and assembly of an antibody Fv fragment, a heterodimer stabilized by antigen, *Journal of Molecular Biology 285*, 2005-2019.

28. Taylor, C. M., Hardre, R., Edwards, P. J. B., and Park, J. H. (2003) Factors Affecting Conformation in Proline-Containing Peptides, *Organic Letters 5*, 4413-4416.

29. Hoyer, W., Ramm, K., and Plückthun, A. (2002) A kinetic trap is an intrinsic feature in the folding pathway of single-chain Fv fragments, *Biophysical Chemistry 96*, 273-284.

30. James, L. C., Roversi, P., and Tawfik, D. S. (2003) Antibody Multispecificity Mediated by Conformational Diversity, *Science 299*, 1362-1367.

31. Mallender, W. D., Carrero, J., and Voss, E. W., Jr. (1996) Comparative Properties of the Single Chain Antibody and Fv Derivatives of mAb 4-4-20, *J. Biol. Chem. 271*, 5338-5346.

32. Weissman, J. S. (1995) All roads lead to Rome? The multiple pathways of protein folding, *Chemistry & biology 2*, 255-260.

33. Eftink, M. R., and Ghiron, C. A. (1981) Fluorescence quenching studies with proteins, *Analytical Biochemistry 114*, 199-227.