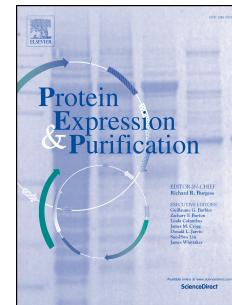


Accepted Manuscript

Automated High-Throughput Dense Matrix Protein Folding Screen Using a Liquid Handling Robot Combined with Microfluidic Capillary Electrophoresis

Philip An, Dwight Winters, Kenneth W. Walker



PII: S1046-5928(15)30107-8

DOI: [10.1016/j.pep.2015.11.015](https://doi.org/10.1016/j.pep.2015.11.015)

Reference: YPREP 4832

To appear in: *Protein Expression and Purification*

Received Date: 24 September 2015

Revised Date: 6 November 2015

Accepted Date: 17 November 2015

Please cite this article as: P. An, D. Winters, K.W Walker, Automated High-Throughput Dense Matrix Protein Folding Screen Using a Liquid Handling Robot Combined with Microfluidic Capillary Electrophoresis, *Protein Expression and Purification* (2016), doi: 10.1016/j.pep.2015.11.015.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 **Automated High-Throughput Dense Matrix Protein Folding Screen Using a Liquid**
2 **Handling Robot Combined with Microfluidic Capillary Electrophoresis**

4 Philip An¹, Dwight Winters¹ and Kenneth W Walker^{1,2}

5
6
7 ¹Therapeutic Discovery, Amgen, Inc., One Amgen Center,
8 Thousand Oaks, CA 91320-1789, United States

10 ²Author to whom correspondence should be addressed

13 *Contact information*

14 Kenneth Walker
15 One Amgen Center Dr.
16 Thousand Oaks, CA 91320
17 T: 805-447-0520
18 kennethw@amgen.com

20 Running title: “Automated High-Throughput Dense Matrix Protein Folding Screen”

Abstract

2 Modern molecular genetics technology has made it possible to swiftly sequence, clone and mass-
3 produce recombinant DNA for the purpose of expressing heterologous genes of interest;
4 however, recombinant protein production systems have struggled to keep pace. Mammalian
5 expression systems are typically favored for their ability to produce and secrete proteins in their
6 native state, but bacterial systems benefit from rapid cell line development and robust growth.
7 The primary drawback to prokaryotic expression systems are that recombinant proteins are
8 generally not secreted at high levels or correctly folded, and are often insoluble, necessitating
9 post-expression protein folding to obtain the active product. In order to harness the advantages
10 of prokaryotic expression, high-throughput methods for executing protein folding screens and the
11 subsequent analytics to identify lead conditions are required. Both of these tasks can be
12 accomplished using a Biomek 3000 liquid handling robot to prepare the folding screen and to
13 subsequently prepare the reactions for assessment using Caliper microfluidic capillary
14 electrophoresis. By augmenting a protein folding screen with automation, the primary
15 disadvantage of *E. coli* expression has been mitigated, namely the labor intensive identification
16 of the required protein folding conditions. Furthermore, a rigorous, quantitative method for
17 identifying optimal protein folding buffer aids in the rapid development of an optimal production
18 process.

Keywords

high-throughput, automation, protein folding, dense matrix, folding screen, liquid handling robot,
microfluidic capillary electrophoresis, inclusion bodies, *E. coli*

1 Abbreviations and Symbols**2**

3 Inclusion bodies (IB), guanidine hydrochloride (GuHCl), dithiothreitol (DTT), L-arginine (L-
4 arg), L-cysteine (L-cys), cystamine-2HCl (cystamine), attenuated total reflectance Fourier
5 transform infrared spectroscopy (ATR-FITR), microfluidic capillary electrophoresis (MCE),
6 molecular weight (MW), full width at half maximum (FWHM), Protein A (Pro A), final bulk
7 product (FBP), UV absorbance at 280 nm (A280), double-washed IB (dwIB).

1

Introduction

2 The cornerstone of preparative-scale protein production, whether for industrial or basic
3 research purposes, is the heterologous expression of recombinant genes of interest, and the most
4 common expression systems employed are bacterial and mammalian [1-3]. Mammalian systems
5 are generally the most capable host cells for producing properly folded secreted proteins;
6 however, they require more time to establish, necessitate longer culture times and greater
7 maintenance expense as compared to prokaryotic systems. Bacterial systems are the fastest and
8 most economical; however, they are not as efficient at protein secretion, and they possess only
9 rudimentary protein folding machinery. Prokaryotes generally struggle to fold complex
10 molecules with multiple subunits, and this is exacerbated when disulfide bonds are required [4].

11 Bacterial systems can rapidly produce recombinant proteins in high quantities, making
12 them an attractive solution given the increasingly value-centric philosophy of research
13 organizations. Bacterial host cells expressing recombinant gene products, particularly
14 extracellular proteins, will generally deposit the protein in insoluble, inactive aggregates called
15 inclusion bodies (IB) which can account for up to 26% of the host cell mass [5,6]. IBs are
16 formed from the intermolecular interactions of over-expressed proteins that accumulate in the
17 host cell [7,8]. Some benefits of proteins in this state are their high level of purity (up to 90%
18 target protein), protection from proteolysis and the convenience of being able to separate the IBs
19 from soluble host cell proteins by centrifugation and/or filtration [6,9,10]. While the benefits are
20 clear and easily realized, the challenge is to bring the proteins out of their aggregated, non-native
21 state and into a properly folded, active one.

22 The properly folded structures of proteins are dependent upon the amino acid sequence
23 and the conditions of the surrounding solution, thereby opening the possibility to *in vitro*

1 renaturation from a misfolded state [11]. The IBs must first be solubilized and fully reduced in
2 order to convert the misfolded mass of polypeptide chains to a fully unfolded state so that proper
3 folding can commence. A strong chaotrope such as 6 M guanidine hydrochloride (GuHCl) and a
4 reductant such as dithiothreitol (DTT) are generally employed to facilitate solubilization [12,13].
5 Following solubilization, the denaturants must be removed or reduced in a controlled fashion in
6 order to allow the polypeptide chains to fold into functional proteins. Methods of performing
7 renaturation include dilution, buffer exchange and immobilization of the protein on a solid phase
8 [9]. Buffer exchange by dialysis is an inefficient means of conducting a high-throughput process
9 because dialysis can be slow depending on the extent of exchange required, potentially consume
10 large amounts of dialysate and will require individual dialysis units and vessels for each sample.
11 Immobilization onto a solid matrix, or on column folding, is highly protein specific since the
12 matrix can interfere with the folding process and success will often depend on empirical
13 optimization experiments, which are not readily adapted to high-throughput processing [14]. In
14 order to preserve throughput and economy, dilution of the solubilized IBs into a folding buffer is
15 the most efficient way of reducing the concentration of the denaturants, which is why it is the
16 most frequently used method in high-throughput applications [14-21].

17 The breadth of possible folding conditions is infinite, thus a large number of experiments
18 with high-throughput capability employing a systematic and thorough exploration of diluents
19 conducive to protein folding is required. The power of the folding matrix, which is directly
20 proportional to the scope of variables tested, needs to be taken into consideration, since fewer
21 conditions diminish the probability of success while excessive conditions become uneconomical.
22 One method for selecting conditions is the fractional factorial approach, which seeks to explore a
23 small subset of folding conditions over a diverse panel of variables [20,22,23]. Numerous

1 commercially available protein folding products are built on the fractional factorial principle of
2 folding condition evaluation (iFold [EMD Millipore, Billerica, Massachusetts, USA], QuickFold
3 [AthenaES, Baltimore, Maryland, USA], Pierce Protein Refolding Kit [Thermo Fisher Scientific,
4 Rockford, Illinois, USA] and BioAssay Protein Refolding Kit [Biomol, Hamburg, Germany]).
5 However, the drawback to using a fractional factorial method for crafting a screening strategy is
6 that it creates a sparse matrix which generates diverse data with multiple variables, whereas a
7 dense matrix can generate trend data and better elucidate the interdependence of variables. In
8 the example presented, human Fc was chosen as the test protein due to its native structure
9 requiring intermolecular and intramolecular disulfide bonds; however, the matrixdescribed
10 herein has been successfully utilized to fold a variety of proteins including, but not limited to,
11 TGF β family members, receptor extracellular domains, cysteine knot proteins, polypeptides,
12 immunoglobulins and immunoglobulin fusions.

13 Physically assembling a 96-condition matrix and the ability to construct it efficiently is
14 germane to establishing the system as a practical tool for screening on demand. Assembly of the
15 matrix near the time the folding screen will be conducted is important, since many of the
16 components degrade rapidly when exposed to water. For this purpose, various programmable
17 mechanical instruments have been used such as liquid handling robots, and even an automated
18 chromatography system outfitted with multiple sample lines and a fraction collector as the matrix
19 destination [16,20,21]. In the example described herein, a dense matrix to explore the effects of
20 pH, urea, L-arginine (L-arg), glycerol and the redox couple of L-cysteine (L-cys) and cystamine-
21 2HCl (cystamine) on protein folding, and a Biomek (Beckman Coulter, Brea, California, USA)
22 XYZ robot with a custom working deck was employed to automate liquid handling to assemble
23 the matrix.

Upon completion of the folding reactions, analytical methods are necessary to identify the conditions that produced properly folded material. Activity assays, immunoassays and limited proteolysis assays address the folded state from a functional point of view, which is the highest level of determining whether a protein is in its native conformation; however, these methods require specific protein reagents and optimization, increasing the challenge of establishing a robust primary screen for novel proteins. Multispectroscopic monitoring (far and near UV circular dichroism, fluorescence and UV spectrometry) is a powerful set of spectroscopic tools that provide information about the secondary structure of a protein; however, proper performance requires the samples be exchanged into organic solvent, and in conjunction with the subsequent data deconvolution, result in low throughput analysis [14]. Recently, developments in attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FITR) have been used to assess protein folding over time; however, an automated high-throughput version of this technique is currently not available [24]. SDS-PAGE is a simpler means of assessing the fold quality of renatured proteins yet suffers from inconsistencies stemming from gel loading, staining/destaining and densitometer settings. Furthermore, the amount of labor and resources required to run gels on a large matrix of conditions is substantial [16, 25]. From a throughput standpoint, an automated SDS-PAGE mimetic with greater run-to-run consistency would be the most accessible and efficient technique for analyzing the folding matrix. The Caliper LabChip GXII (PerkinElmer, Waltham, Massachusetts, USA) is an instrument that utilizes microfluidic capillary electrophoresis (MCE) to perform high-throughput electrophoretic separations from samples loaded into a microplate, a form factor which accommodates the aforementioned automation robotics.

1 Quantitative data generated by MCE assessment of the folding conditions provide a
2 method to rank the conditions based on performance criteria set forth by the analyst, culminating
3 in the selection of a group of top-performing conditions. The top performing conditions are then
4 scaled-up to mimic production-like conditions in order to corroborate the screen findings and
5 identify a single ideal folding condition. The methods presented here demonstrate how hundreds
6 of tasks can be performed efficiently by one analyst in a timely manner using rigorously defined
7 methods, which may enable bringing molecules from a functionless state through a discovery-
8 level screen, ultimately leading to a production-level method for producing proteins of interest.

9

10 **Results**

11 The *E. coli* IB-based protein production process described herein is outlined in Figure 1.
12 Briefly, IBs containing human Fc were processed via an automated small-scale folding buffer
13 screen, and the folding reactions were subsequently analyzed on an MCE system. Data based on
14 the electrophoresis attributes of the products were employed to select a subset of lead folding
15 conditions. The lead folding conditions were then repeated in an intermediate-scale screen to
16 test their scalability and purification behavior, and the data collected from this screen culminated
17 in the selection of a single best condition for large-scale production. Performing these
18 experiments in a systematic fashion enabled collection of the necessary data to detect trends that
19 aid the analyst in selecting an optimized lead folding condition.

20 A Biomek (Beckman Coulter, Brea, California, USA) XYZ robot with a custom working
21 deck was employed to automate liquid handling by assembling the matrix in microplates on
22 Peltier-cooled plate shakers attached to the working deck (Fig. 2) from a set of stock solutions
23 containing the key refolding buffer components. The shakers mix the matrix components after

1 liquid transfer is completed, ensuring complete sample mixing and temperature consistency
2 during the folding reaction.

3 The small-scale folding screen consisted of 96 unique formulations, which comprise a
4 dense matrix that explores the effects of L-arg, L-cys, cystamine, glycerol, urea and pH in a
5 logical progression (Table 1). A Biomek 3000 liquid-handling robot assembled all 96
6 conditions, distributed across four separate 24-well plates located on chilled shakers from a set of
7 10 concentrated stock solutions located in various vessels on the custom working deck (Fig. 2).
8 Stock solution locations and liquid transfer volumes were specified in a Microsoft Excel
9 spreadsheet, and a built-in macro automatically exported the instructions to Biomek “worklists”
10 that direct the robot’s actions (supplemental data). After completion of the folding reactions, the
11 working deck and instruction set of the Biomek were reconfigured to automatically load the
12 reactions from the four 24-well plates into a single 96-well plate along with MCE running buffer
13 (supplemental data). Precipitation was observed with certain folding conditions located in the
14 24-well plates. A centrifugation step, per MCE manufacturer recommendation, was performed
15 to pellet precipitates that may have been transferred from the folding solution or formed during
16 plate preparation. The sample plate was then loaded into the MCE instrument for analysis and
17 electrophoretic data collection, which requires about 1.5 hours to assess all 96 conditions.

18 Although the theoretical molecular weight (MW) of native human Fc is 51 kDa, a
19 purified standard assessed using MCE migrates at 64 kDa compared to the protein standard
20 ladder. Engaging a peak search with the human Fc standard MW as a guideline (64 kDa \pm 5%)
21 and a minimum peak height of 10 fluorescence units, the analysis software identified product
22 peaks in 95 of the reactions with condition H12 (400 mM L-arg, 2 mM L-cys, 2 mM cystamine,
23 20% glycerol, 4 M urea, pH 10.5) being the only condition that did not produce an identifiable

1 eletropherogram peak. The first tier selection criterion was MCE peak height in fluorescence
2 units, and a cutoff limit for peak height was set at ≥ 842 fluorescence units in order to segregate
3 the top 30 performing folding conditions (Fig. 3). Within the top 30, higher pH was preferred in
4 each folding group up to 4 M urea, after which lower pH was preferred. The groups with 160
5 mM L-arg (A – D) were represented by 21 conditions above the cutoff limit, and they were more
6 successful than the groups with 400 mM L-arg (groups E – H), which were represented by only
7 nine conditions above the cutoff. Groups with a 1:4 redox couple ratio (A, B, E and F) had a
8 higher rate of success than groups with a 2:2 redox couple (C, D, G and H), represented by 19
9 conditions and 11 conditions above the cutoff, respectively. Groups without glycerol (A, C, E
10 and G) performed better in urea concentrations of 1 and 2 M, with nine conditions each above
11 the limit, than in 0 and 4 M urea, which had one and none, respectively. Ranking the groups by
12 representation in the top 30 proceeds as follows: seven from group A (160 mM L-arg, 1:4 redox
13 ratio and 0% glycerol) > six from B (160 mM L-arg, 1:4 redox ratio and 20% glycerol) > five
14 from C (160 mM L-arg, 2:2 redox ratio and 0% glycerol) > four from E (400 mM L-arg, 1:4
15 redox ratio and 0% glycerol) > three each from D (160 mM L-arg, 2:2 redox ratio and 20%
16 glycerol) and G (400 mM L-arg, 2:2 redox ratio and 0% glycerol) > two from F (400 mM L-arg,
17 1:4 redox ratio and 20% glycerol) (Table 1).

18 The electropherograms of the top 30 conditions, selected by peak height, were subjected
19 to further assessment, which employed peak sharpness (the ratio of peak height to full width at
20 half maximum (FWHM)) as the distinguishing factor, where a sharper peak equates to a greater
21 score (Fig. 4). Groups with 160 mM L-arg (A – D) generally produced higher scores for
22 sharpness than groups with 400 mM (E – G). In the presence of 0% glycerol, conditions with 1:4
23 redox ratios (A5 – 9 and E5 – 8) performed better than their 2:2 counterparts (C5 – 9 and G5 –

1 8). Groups with 20% glycerol (B, D and F) possessed a higher maximum sharpness score when
2 compared to their corresponding 0% glycerol groups (A, C and E). The significance of each
3 component was measured by their representation in the top 30 as follows, with the number of
4 samples for a given condition in parentheses: 160 mM L-arg (21) > 400 mM (9); 1:4 redox ratio
5 (19) > 2:2 (11); 0% glycerol (19) > 20% (11); 2 M urea (12) ≈ 1 M (10) > 4 M (7) > 0 M (1); pH
6 10.5 (13) ≈ 9.5 (12) > 8.5 (5). The top 10 performing conditions by sharpness (≥ 5555
7 fluorescence:FWHM) advanced to the final tier of selection.

8 Product purity by percent, measured by MCE, was introduced as a parameter for the final
9 selection of conditions for intermediate-scale production (Fig. 5). Measurements for MCE peak
10 height, peak sharpness, and purity were converted to relative performance scores with a mean of
11 72 for normalized comparison across all three attributes. The two highest scoring folding
12 conditions for purity were selected, followed by the two highest for peak sharpness and lastly the
13 two highest for peak height. If a condition was already selected by a preceding attribute, the next
14 best scoring condition for the attribute at hand was selected. All six conditions that were
15 advanced to intermediate-scale production share the same L-arg concentration (160 mM) and
16 those that contained 20% glycerol required higher urea concentrations (2 M and 4 M).

17 In order to identify trends, the mean of the measurements for each selection category of
18 peak height, sharpness, and purity were calculated (Table 2). A matching bias toward 160 mM
19 L-arg, 1:4 redox ratio, 0% glycerol, 2 M urea, and pH 10.5 exists between peak height and
20 sharpness. The four urea concentrations produced a bell-shaped curve for preference in all three
21 selection categories starting on the low end at 0 M, increasing through 1 M to the most preferred
22 concentration of 2 M and then back down again at 4 M. The three pH levels showed a
23 preference for increasing alkalinity in all three selection categories.

1 After determining the top six folding conditions by the two highest scores in each of the
2 selection criteria (peak height, sharpness and purity), this subset was scaled-up to investigate the
3 impact of conducting the folding experiments under conditions similar to a full-scale preparative
4 production. The intermediate-scale folding screen was conducted in 100 mL reaction volumes,
5 and after 60 hours of incubation the reactions were analyzed by SDS-PAGE (Fig. 6). All
6 conditions produced similar amounts of material with bands of comparable compactness;
7 however, conditions A9 and D12 possessed the lowest amount of high molecular weight species.
8 Purification of the intermediate-scale reactions utilized Protein A (Pro A) affinity
9 chromatography, followed by SP HP cation exchange chromatography. A NaCl gradient was
10 used to elute proteins from the SP HP column and the resulting fractions were analyzed by MCE.
11 Each SP HP fraction was analyzed for peak sharpness, the same secondary metric employed to
12 analyze the small-scale folding screen, to identify the most desirable fractions (Fig. 7). The three
13 highest scoring fractions for each sample were pooled as the final bulk product (FBP).

14 The protein concentration for each pool was measured at each step by UV absorbance at
15 280 nm (A_{280}) (Table 3) and the corresponding step yields were calculated. The three highest
16 yielding conditions after purification were, in order of greatest to least, A8, A9 and A6. While
17 all six conditions shared the same L-arg concentration (160 mM), the three aforementioned
18 conditions also shared the same glycerol content (0%), distinguishing them from B9, B10 and
19 D12.

20 Quality assessment of the FBPs included purity by main peak percentage on HPLC-SEC,
21 MCE and SDS-PAGE. All HPLC-SEC main peak values were $\geq 94.6\%$ and MCE main peaks
22 were $\geq 99.6\%$ (Table 3). The samples, loaded with equivalent amounts of protein (2 μ g), ran at

1 the same molecular weight as the purified standard on SDS-PAGE and appeared to be of similar
2 quality and purity with the exception of B10 which shows a faint doublet (Fig. 8).

3

4 **Discussion**

5 Gene products are identified and sequenced with increasing swiftness as molecular
6 genetics technology advances, thus creating a bottleneck at the protein production phase.
7 Additionally, as protein engineering practices increase, the quantity of sequence variants in a
8 screening panel are also expanding, and the necessity to analyze the physical properties and
9 activity profiles of each variant demands high-volume protein production. Expressing the
10 proteins in *E. coli* attends to the issue of rapid output, but complex proteins with disulfide bonds
11 or multiple subunits are typically shuttled to the IBs during expression, thus requiring *in vitro*
12 folding to attain a bioactive, native conformation. Since optimal folding conditions often differ
13 between proteins, a logically assembled matrix comprised of reagents known to strongly
14 influence protein folding (chaotrope, kosmotrope, pH and redox couple) can be employed to
15 increase the probability of identifying good folding conditions for novel proteins (Table 1). The
16 96-condition folding matrix presented here is optimized for secreted disulfide-bond containing
17 proteins by exploring the effects of L-arg, urea, glycerol, L-cys, cystamine, and pH in a
18 systematic progression, and it provides the added convenience of 24 and 96-well microplate
19 formatting. Manually assembling a refold screen for a matrix of this size is time consuming and
20 labor intensive, and the commonly utilized method for assessing folding success by SDS-PAGE
21 introduces substantial subjectivity. The method described herein exploits robotics and other
22 forms of automation to simplify and standardize folding matrix experiments and their subsequent

1 analyses, and it utilizes a well-defined, non-subjective method to identify optimal folding
2 conditions in the most rapid and efficient manner possible.

3 In order to automate the small-scale 96-condition matrix assembly and subsequent
4 folding reaction and analysis, a Biomek 3000 is utilized to manage liquid transfer, and Peltier-
5 cooled shakers are employed to provide sample mixing and temperature control (Fig. 2). The
6 benefits of the Biomek 3000 are its flexible working deck with customizable configurations,
7 simplistic integration of 3rd party on-deck devices and relatively low cost. The attached
8 Thermoshake (INHECO, Martinsried, Germany) shakers (thermoshaker) are a robust solution for
9 this application due to their ability to always stop in the same position, wide temperature control
10 range (4 – 70 °C), versatile orbital speed (200 – 2,000 rpm, 2 mm shake amplitude, ≤ 1 kg load),
11 a footprint approximately the size of a microplate, reliable magnetic drive without moving parts,
12 single controller with manual and remote control that can handle up to six shakers, and USB-
13 enabled computer interface. In conjunction with hardware automation, a custom Microsoft Excel
14 spreadsheet was designed to simplify data organization and Biomek programming by allowing
15 the analyst to enter stock buffers and liquid transfer volumes which are then converted to
16 Biomek “worklists” through a macro. The Biomek software uses these worklists as instructions
17 for pipetting locations and volumes. Using this system, the actions involved in executing a
18 folding matrix with 96 unique buffers can be achieved by preparing 10 stock buffer solutions,
19 which equates to a 9.6-fold increase in efficiency just in liquid handling.

20 The next step in automation involves analyzing the small-scale refold matrix using an
21 objective, data-centric means of triaging the conditions. SDS-PAGE is a prevalent means of
22 protein quality assessment, but for a large quantity of samples, the entire process can be
23 laborious and the results cannot be accurately quantified. Preparing 96 samples and the requisite

1 gels and buffers for SDS-PAGE analysis is time consuming, and strict sample management
2 logistics must be employed to prevent sample handling and identification errors. In addition, this
3 process introduces subjectivity and can lack consistency across gels due to the non-linearity of
4 Coomassie staining, the varying amounts of staining and destaining between gels and the image
5 exposure settings selected by different analysts, thus making it difficult to generate reliable and
6 quantifiable values. By evaluating the samples on an MCE system, one retains the benefits of
7 SDS-PAGE quality assessment (identification of contaminating species, strong aggregates and
8 heterogeneous product banding) and takes advantage of an entirely digital detection system
9 which can process all 96 samples in a single 1.5-hour run, the results of which can be compared
10 side-by-side on a single display. Furthermore, the Biomek working deck can be easily modified
11 and programmed to automatically load all of the folding conditions into a 96-well plate for MCE
12 analysis, thus reducing the possibility of sample handling errors. Automated MCE plate loading
13 also substantially enhances the efficiency of investigating folding duration as an additional
14 dimension of comparative analysis by loading MCE plates at specific points over a time course.

15 MCE separates proteins treated with lithium dodecyl sulfate (LDS) by charge and
16 molecular weight through a microfluidic chip containing a denaturing gel matrix, and by using
17 fluorescence detection, it produces electropherograms with quantifiable peaks representing
18 various species present in the sample. Either a protein standard or an estimation of product MW
19 can be used as a means to instruct the software to automatically identify product peaks, and peak
20 identification tolerance can be adjusted to detect small peaks and separate incompletely resolved,
21 overlapping peaks. After identifying the product peaks, their traits are assessed for height,
22 sharpness, and purity to estimate folding success. Peak height is directly related to product yield,
23 which is why peak height is used as the first method of selection (Fig. 3). Difference in peak

height, and thus yield, may be attributed to the removal of highly-aggregated/precipitated forms during centrifugation of the MCE sample plate. The next selection criterion is peak sharpness (Fig. 4), where sharper peaks indicate greater sample homogeneity, due to co-migration through the gel matrix, than broader peaks, likely stemming from the differential migration of a mixture of folded forms, which travel through the gel matrix at altered rates due to variances in disulfide bond configuration and/or charge envelope. Finally the third criterion for folding condition selection is product purity (Fig. 5). Although product purity is not a direct measure of folding success, the ultimate goal of finding the most efficient way to produce purified protein makes initial purity a valuable asset, since such proteins are likely to require fewer purification steps, thus resulting in higher yields. Using these three selection criteria in a tiered manner provides the analyst with a quick and objective tool to distill the 96 condition matrix down to a small set of promising and diverse conditions (Table 1).

The final evaluation step requires reproducing the folding reactions at an intermediate-scale to assess protein behavior during scale-up. Instead of microplates on a thermoshaker, the intermediate-scale reactions are assembled in beakers and constantly mixed with a stir bar, which better mimics the aeration conditions encountered in a large-scale reactor. The larger volume also provides the analyst with ample starting material to conduct a purification process that will generate meaningful data regarding the protein's behavior and performance upon purification, since misfolded proteins usually display poor chromatographic behavior. After the purifying the intermediate-scale reactions, yield and purity comparisons are utilized to select the most desirable condition for full-scale protein production. In the human Fc test case described herein, all six intermediate-scale FBPs possess virtually identical purity, therefore the yield and purification efficiency is the determining factor (Table 3). The highest yielding condition, A8

1 (160 mM L-arg, 1:4 redox ratio, 0% glycerol, 2 M urea, pH 9.5) at 22.0 mg, also resulted in the
2 most efficiently purified protein, with an overall 65% recovery from Pro A followed by cation
3 exchange chromatography. The next most efficient is D12 at 62%, and the other conditions are
4 45% or lower. It is worth noting that the second lowest Pro A yielding condition is ultimately
5 the best overall folding condition indicating that this condition likely produced the largest
6 quantity of high-quality product with the least amount of contaminants.

7 The *in vitro* protein folding system presented here is centered on a robust folding
8 condition matrix coupled to a high level of automation and a well-defined quantitative analytical
9 method to enable rapid identification of optimal protein folding conditions. This high-
10 throughput and efficient protein folding process enables reduced timelines for both publication
11 and path to market for a wide variety of proteins.

12

13 Materials and Methods

14 Cell Culture

15 The constant region (Fc) from human IgG1 was cloned into a prokaryotic expression
16 vector and the resulting construct was transformed into *E. coli*. A 50 µl portion of a frozen
17 glycerol stock was used to inoculate 50 mL of Terrific Broth (Teknova, Hollister, California,
18 USA) plus Kanamycin. This culture was incubated at 37 °C in a shaking incubator overnight.
19 20 mL of the overnight culture was used to inoculate 1 liter of Terrific Broth plus kanamycin in a
20 2.5 L Thomson Ultra-Yield shake flask. The culture was incubated at 37 °C until the OD₆₀₀
21 reached 1.0. Homoserine lactone was added to the culture at a final concentration of 30 ng/mL.
22 The culture continued to be incubated at 37 °C for an additional 6 hours. The cells were
23 harvested by centrifugation and stored at -80 °C.

1 *Inclusion body preparation*

2 The frozen cell paste was homogenized in five volumes of 50 mM Tris, 5 mM EDTA, pH
3 8.0 using a stationary homogenizer, and the resultant suspension was passed through an M-110S
4 Microfluidizer (Microfluidics Corporation, Irvine, California, USA) twice at 15,000 PSI. The
5 lysate was then centrifuged at 25,000 RCF to pellet the insoluble fraction, while the soluble
6 fraction was decanted to waste. The pellet was resuspended, with half the starting lysis volume,
7 in water using an Omni TH (Omni International, Kennesaw, Georgia, USA) handheld
8 homogenizer and centrifuged again at the same RCF, and the soluble portion was subsequently
9 decanted to waste; this process was repeated once again resulting in double-washed IB (dwIB).
10 The dwIB were stored at -80 °C.

11 The dwIB were first resuspended in one volume (by weight) of water using a handheld
12 homogenizer, after which 8 volumes of 8 M GuHCl (Teknova) were added and homogenization
13 continued until no visible solids were observed. Solid DTT (Sigma-Aldrich, St. Louis, Missouri,
14 USA) was added to the solubilized dwIB to a final concentration of 10 mM and agitated until no
15 visible DTT was observed. The reduced, solubilized dwIBs were incubated at 37 °C in a water
16 bath for one hour.

17 *Small-scale folding matrix design*

18 The small-scale folding matrix employs 96 unique combinations of six components.
19 Each of the six components were explored in multiple dimensions: two concentrations of L-arg-
20 HCl (160 mM and 400 mM) (Sigma-Aldrich), St. Louis, Missouri, USA), two redox couple
21 ratios of L-cys to cystamine (1 mM:4 mM and 2 mM:2 mM) (Sigma-Aldrich), two
22 concentrations of glycerol (0% and 20%) (Sigma-Aldrich), four concentrations of urea (0 M, 1

1 M, 2 M and 4 M) (J. T. Baker, Center Valley, Philadelphia, USA) and three different pHs
2 buffered by 50 mM Tris (Teknova) (8.5, 9.5 and 10.5) (Table 1).

3 *Liquid-handling robot setup*

4 A Biomek 3000 robot with a custom working deck assembled the folding conditions for
5 the small-scale screen (Fig. 1). Three micropipetting tools, P20, P200L, P1000L, were stored in
6 a holder at position A2 on the working deck, with corresponding tip sizes located in positions A3
7 – A5. The L-arg and Tris were formulated together as concentrated stock solutions, 160 mM L-
8 arg, 50 mM Tris and 400 mM L-arg, 50 mM Tris, respectively, at each of three pHs, 8.5, 9.5,
9 10.5 comprising six stock solutions total. Each of these stock solutions occupied a separate well
10 in a 24-well plate (Whatman, Little Chalfont, Buckinghamshire, United Kingdom) at position
11 B2. An 80% glycerol and a freshly prepared 9 M urea stock were stored in separate 40 mL
12 reservoirs (Beckman Coulter) at position B3. The water was kept in an inverted pipette tip box
13 lid at position B4. The freshly prepared redox reagents, L-cys and cystamine, were formulated at
14 100 mM each and placed along with the reduced, solubilized dwIB in separate wells in a 24-well
15 plate (Whatman) at position B5. The liquid transfer volumes and locations were specified in an
16 Excel spreadsheet and then exported as Biomek “worklists” through a macro programmed to
17 perform that task. Protein folding took place in four 24-well cell culture plates (Corning Life
18 Sciences, Tewksbury, Massachusetts, USA) at positions A1, A6, B1, and B6 on thermoshakers
19 set to maintain a temperature of 4 °C with constant agitation.

20 Four Inheco Thermoshakes were connected to an Inheco Multi TEC Control unit, which
21 was then connected to the control computer. Due to the height of the thermoshakers, placing
22 them on the Biomek working deck would not provide enough clearance for 1 mL pipette tips. In
23 order to fit the four thermoshakers and keep them within the reach of the robot, the left and right

1 deck extensions were removed and the thermoshakers were attached using universal “L”
2 brackets. Anti-vibration rubber pads (1/8 inch thickness) were glued to the bottom of each
3 bracket to reduce vibration. The thermoshaker software driver was written in Visual Basic 2008
4 (Microsoft, Redmond, Washington, USA) based on manufacturer examples and dynamic-link
5 library (Microsoft) components, and its operation was made accessible from the Biomek
6 software using a built in command line module which is entered as follows: (executable driver
7 filename) (shaker address) (shaker function) (on or off); for example “ts.exe 1 S 1” where
8 “ts.exe” is the driver filename, the first “1” specifies shaker number one (use “0” for all four), the
9 “S” stands for shaking (or “T” for temperature), and the last “1” signifies thermoshaker
10 activation (“0” for off).

11 *MCE analysis*

12 Following a 60-hour folding reaction, another Biomek automated operation was used to
13 prepare the samples for the subsequent analytical step. The vessels in positions B2 – B5 were
14 removed, an empty 96-well PCR plate (BioRad, Hercules, California, USA) was placed in B2
15 and a 24-well plate (Whatman) containing LabChip GXII protein analysis denaturing LDS
16 sample buffer with 23.8 mM iodoacetamide (Sigma-Aldrich) was placed in B3. Due to the less
17 complex nature of the liquid handling in this step, the Excel spreadsheet and macro were not
18 employed and Biomek programming was performed directly in the control software. The
19 Biomek transferred 21 µL of LabChip sample buffer with IAA to each well in the PCR plate, and
20 then 6 µL of each folding reaction was transferred to individual wells in the PCR plate. The
21 PCR plate was then covered with an adhesive film and placed in a heating block at 85 °C for 5
22 minutes; the subsequent steps were carried out per manufacturer instructions. Upon processing
23 completion, the Peak Find utility under Analysis Settings was adjusted for maximum peak

1 identification tolerance (Slope Threshold Sample = 0.01) and maximum shoulder/split-peak
2 identification tolerance (Inflection Threshold Sample = 999.0), thus allowing the discovery of
3 small peaks and liberally dissecting potentially overlapping peaks.

4 *Intermediate-scale folding reactions and purification*

5 The intermediate folding reactions were conducted at a 100-mL scale in 250-mL beakers.
6 The stock buffers used to formulate the folding solutions were 1 M Tris base (Teknova), 1.6 M
7 L-arg-HCl (Sigma-Aldrich), 50% glycerol (Teknova), 0.5 M L-cys (Sigma-Aldrich) and 0.5 M
8 cystamine (Sigma-Aldrich). Solid urea (Sigma-Aldrich) was used to ensure the freshness of the
9 component. The components were combined in individual beakers, mixed to homogeneity,
10 titrated to the appropriate pH and then brought up to the final volume with water. Four mL of
11 reduced, solubilized dwIB were added to the reactions with gentle stirring, and then the mixture
12 was incubated at 4 °C for 60 hours with gentle stirring. Following the incubation, a 6 µL aliquot
13 of each reaction was run on a non-reduced 4 – 12% NuPAGE gel using MES running buffer
14 (Life Technologies, Carlsbad, California, USA), per manufacturer instructions.

15 Purification was conducted on an AKTA Purifier (GE Healthcare Life Sciences, Little
16 Chalfont, Buckinghamshire, UK) liquid chromatography system. The refold solution was loaded
17 directly onto a 5 mL MabSelect HiTrap (GE Healthcare Life Sciences) column, washed with five
18 column volumes (CV) of Dulbecco's phosphate buffered saline (Life Technologies) and eluted
19 step-wise with five CV of 100 mM acetic acid, pH 3.2. The entire elution peak measured by
20 A₂₈₀ was pooled, titrated to pH 5.0 using 1 N hydrochloric acid and diluted with five volumes of
21 water. The MabSelect purified material was then loaded onto a 5 mL SP Sepharose HP HiTrap
22 (SPHP) (GE Healthcare Life Sciences), washed with five CV of 20 mM sodium acetate (Sigma-

1 Aldrich), pH 5.0, 50 mM NaCl and eluted with a 20 CV NaCl gradient from 50 to 450 mM. The
2 SPHP fractions were analyzed MCE and pooled by product purity to achieve a FBP.

3 *FBP quality assessment*

4 Product purity of the FBP was measured by MCE and reconfirmed by running 2 µg of
5 each sample on a non-reducing 4 – 12% NuPAGE gel using MES running buffer. High-
6 molecular weight quantity was measured by HPLC size exclusion chromatography using a Sepax
7 Zenix-C SEC-300 (Newark, Delaware, USA) connected to an Agilent 1260 Infinity LC System
8 (Santa Clara, California, USA) running a 50 mM sodium phosphate, 250 mM NaCl, pH 6.9
9 mobile phase at 1 mL/min.

10

11 **Supplementary Material**

12 Supplemental figures 1-5 describe additional configurations of the custom Biomek deck, scripts,
13 and SDS-PAGE analysis of the scaled-up folding reactions.

14

15 **Acknowledgements**

16 We would like to acknowledge Alex Mladenovic and Randy Hecht for contributing to the
17 construction and programming of the Biomek, Tom Boone for protein folding condition
18 guidance and Jeff Lewis for expressing the recombinant protein described in this manuscript.

19

20 **References**

21 [1] C.J. Huang, H. Lin, X.M. Yang, Industrial production of recombinant therapeutics in
22 *Escherichia coli* and its recent advancements, J. Ind. Microbiol. Biotechnol. 39 (2012)
23 383–399. doi:10.1007/s10295-011-1082-9.

- 1 [2] G.L. Rosano, E.A. Ceccarelli, Recombinant protein expression in microbial systems.,
2 Front. Microbiol. 5 (2014) 341. doi:10.3389/fmicb.2014.00341.
- 3 [3] J. Zhang, Mammalian Cell Culture for Biopharmaceutical Production, Man. Ind. Microbiol.
4 Biotechnol. (2010) 157–178.
- 5 [4] W. Swietnicki, Folding aggregated proteins into functionally active forms, Curr. Opin.
6 Biotechnol. 17 (2006) 367–372. doi:10.1016/j.copbio.2006.05.011.
- 7 [5] F. a. O. Marston, P. a. Lowe, M.T. Doel, J.M. Schoemaker, S. White, S. Angal, Purification
8 of Calf Prochymosin (Prorennin) Synthesized in *Escherichia coli*, Bio/Technology. 2
9 (1984) 800–804. doi:10.1038/nbt0984-800.
- 10 [6] F. a. O. Marston, The purification of eukaryotic polypeptides synthesized in *Escherichia*
11 *coli*., Biochem. J. 240 (1986) 1–12.
- 12 [7] A. Mukhopadhyay, Construction and deconstruction of bacterial inclusion bodies, Adv
13 Biochem Eng Biotechnol. 56 (1997) 61–109.
- 14 [8] J.M. Schoemaker, a H. Brasnett, F. a Marston, Examination of calf prochymosin
15 accumulation in *Escherichia coli*: disulphide linkages are a structural component of
16 prochymosin-containing inclusion bodies., EMBO J. 4 (1985) 775–780.
- 17 [9] L.F. Vallejo, U. Rinas, Strategies for the recovery of active proteins through refolding of
18 bacterial inclusion body proteins., Microb. Cell Fact. 3 (2004) 11. doi:10.1186/1475-2859-
19 3-11.
- 20 [10] B. Fischer, I. Sumner, P. Goodenough, Isolation, renaturation, and formation of disulfide
21 bonds of eukaryotic proteins expressed in *Escherichia coli* as inclusion bodies., Biotechnol.
22 Bioeng. 41 (1993) 3–13. doi:10.1002/bit.260410103.

- 1 [11] C.B. Anfinsen, Principles that govern the folding of protein chains., *Science*. 181 (1973)
2 223–230. doi:10.1126/science.181.4096.223.
- 3 [12] R. Rudolph, H. Lilie, In vitro folding of inclusion body proteins, *FASEB J.* 10 (1996) 49–
4 56. doi:10.1021/bi2012965.
- 5 [13] E. Clark, Refolding of recombinant proteins, *Curr. Opin. Biotechnol.* 9 (1998) 157–63.
6 doi:10.1016/S0958-1669(98)80109-2.
- 7 [14] A.P.J. Middelberg, Preparative protein refolding., *Trends Biotechnol.* 20 (2002) 437–43.
8 doi:10.1016/S0167-7799(02)02047-4.
- 9 [15] B. Coutard, E.G.J. Danchin, R. Oubelaid, B. Canard, C. Bignon, Single pH buffer refolding
10 screen for protein from inclusion bodies, *Protein Expr. Purif.* 82 (2012) 352–359.
11 doi:10.1016/j.pep.2012.01.014.
- 12 [16] V. Dechavanne, N. Barrillat, F. Borlat, A. Hermant, L. Magnenat, M. Paquet, et al., A high-
13 throughput protein refolding screen in 96-well format combined with design of experiments
14 to optimize the refolding conditions, *Protein Expr. Purif.* 75 (2011) 192–203.
15 doi:10.1016/j.pep.2010.09.008.
- 16 [17] M.W. Qoronfleh, L.K. Hesterberg, M.B. Seefeldt, Confronting high-throughput protein
17 refolding using high pressure and solution screens, *Protein Expr. Purif.* 55 (2007) 209–224.
18 doi:10.1016/j.pep.2007.05.014.
- 19 [18] A.M. Buswell, M. Ebtinger, A. a. Vertés, A.P.J. Middelberg, Effect of operating variables
20 on the yield of recombinant trypsinogen for a pulse-fed dilution-refolding reactor,
21 *Biotechnol. Bioeng.* 77 (2002) 435–444. doi:10.1002/bit.10148.

- 1 [19] L. Lin, J. Seehra, M.L. Stahl, High-throughput identification of refolding conditions for
2 LXRbeta without a functional assay., Protein Expr. Purif. 47 (2006) 355–66.
3 doi:10.1016/j.pep.2005.11.010.
- 4 [20] Y. Bourne, R. Vincentelli, F. Frassinetti, C. Valencia, C. Bignon, C. Cambillau, et al.,
5 High-throughput automated refolding screening of inclusion bodies, Protein Sci. 13 (2004)
6 2782–2792. doi:10.1110/ps.04806004.
- 7 [21] C. Scheich, F.H. Niesen, R. Seckler, K. Büßow, An automated in vitro protein folding
8 screen applied to a human dynactin subunit., Protein Sci. 13 (2004) 370–80.
9 doi:10.1110/ps.03304604.
- 10 [22] D.A. Tobbell, B.J. Middleton, S. Raines, M.R.C. Needham, I.W.F. Taylor, J.Y. Beveridge,
11 et al., Identification of in vitro folding conditions for procathepsin S and cathepsin S using
12 fractional factorial screens., Protein Expr. Purif. 24 (2002) 242–54.
13 doi:10.1006/prep.2001.1573.
- 14 [23] N. Armstrong, A. de Lencastre, E. Gouaux, A new protein folding screen: application to
15 the ligand binding domains of a glutamate and kainate receptor and to lysozyme and
16 carbonic anhydrase., Protein Sci. 8 (1999) 1475–83. doi:10.1110/ps.8.7.1475.
- 17 [24] C. Walther, S. Mayer, A. Jungbauer, A. Dürauer, Getting ready for PAT: Scale up and
18 inline monitoring of protein refolding of Npro fusion proteins, Process Biochem. 49 (2014)
19 1113–1121. doi:10.1016/j.procbio.2014.03.022.
- 20 [25] S. Lee, J.F. Carpenter, B.S. Chang, T.W. Randolph, Y. Kim, Effects of solutes on
21 solubilization and refolding of proteins from inclusion bodies with high hydrostatic
22 pressure, Protein Sci. 15 (2006) 304–313. doi:10.1110/ps.051813506.For.

1 [26] M. Manning, W. Colón, Structural basis of protein kinetic stability: Resistance to sodium
2 dodecyl sulfate suggests a central role for rigidity and a bias toward β -sheet structure,
3 Biochemistry. 43 (2004) 11248–11254. doi:10.1021/bi0491898.

4

Table & Figure Legends

1
2
3
4 **Table 1. 96-condition folding matrix.** The folding matrix is depicted as a 96-well plate with
5 each well representing a unique folding condition containing one of each of the following
6 components: 160A, 400A = 160 mM and 400 mM L-arginine, respectively; 1:4, 2:2 = 1 mM to 4
7 mM and 2 mM to 2 mM L-cysteine to cystamine-2HCl, respectively; 0%, 20% = 0% and 20%
8 glycerol, respectively; 0U, 1U, 2U, 4U = 0 M, 1 M, 2 M, 4 M urea, respectively; 8.5, 9.5, 10.5 =
9 pH 8.5, 9.5, 10.5, respectively. The distribution of top rated folding conditions are highlighted:
10 yellow = MCE peak height; blue = MCE peak height and sharpness; green = MCE peak height,
11 sharpness and purity

12
13
14

15 **Table 2.** Folding component trends by average measured MCE score. The data generated in the
16 areas of height (fluorescence), sharpness (height/FWHM in fluorescence/sec), and purity (%)
17 were averaged for each folding buffer component and their respective standard error of the
18 means calculated. The highlighted numbers represented the best performers in each category. L-
19 arg = L-arginine; Redox = L-cysteine:cystamine-2HCl

1 **Table 3. Purification yield and purity data.** Yields were calculated from UV absorbance at
2 280 nm. SEC main peak (MP) percentage was obtained by HPLC analysis and software
3 integration on the SP HP pool. MCE MP percentage was obtained by software integration on the
4 SP HP pool. 160A = 160 mM L-arginine, 1:4, 2:2 = 1 mM to 4 mM and 2 mM to 2 mM L-
5 cysteine to cystamine-2HCl, respectively; 0%, 20% = 0% and 20% glycerol, respectively; 1U,
6 2U, 4U = 1 M, 2 M, 4 M urea, respectively; 8.5, 9.5, 10.5 = pH 8.5, 9.5, 10.5, respectively.
7
8

9 **Supplemental Table 1. Excel spreadsheet header for generating Biomek worklists.** The
10 Biomek uses instructions called “worklists” to direct its movement and liquid handling; a
11 Microsoft Excel spreadsheet macro is used to generate the worklists. Each component required
12 for assembling a complete fold condition is listed as a different Component. A Tip Type is
13 specified as well as the Location of the component in the working deck. The Solution recipe and
14 Total Vol. fields are for reference only and are not required information for the Biomek. Each
15 row below the header represents a different condition and the amount to be transferred (in μ L) is
16 entered for each component column.

1 **Figure 1. Biomek instrument with high-throughput protein folding setup.** Positions A1,
2 A6, B1 and B6 are the microplates on thermoshakers where the folding reactions are conducted;
3 C is the thermoshaker controller. The pipetting tools are located at A2, and the pipette tips are
4 located at A3 (P1000), A4 (P250) and A5 (P20). The chemical reagents are located at B2 (24-
5 deepwell microplate for L-arg/Tris buffers), B3 (modular reservoir for urea and glycerol), B4
6 (inverted tip box lid for water) and B5 (24-deepwell microplate for cystamine-2HCl, L-cysteine,
7 dwIBs).

8

9

10 **Figure 2. Process flow diagram.** The protein of interest follows a path indicated by the boxes
11 beginning with IBs isolated from E. coli cell paste. Following IB solubilization, a primary small-
12 scale screen is initiated to test a matrix of folding conditions with analytics performed by MCE.
13 The top performers from the small-scale screen are processed on an intermediate-scale and
14 purified to test and observe the protein's behavior in a process simulating preparative-scale
15 handling. Assessment of the intermediate-scale results culminate in the selection of one lead
16 condition for full-scale production.

17

18

19 **Figure 3. First tier small-scale protein folding condition selection by MCE peak height.** The
20 96-condition folding screen was analyzed by MCE. Analysis settings were adjusted to maximize
21 peak recognition and automatically identify the product peak by MW. Peak heights were
22 compared, and the top 30 conditions were identified for advancement to second tier selection,
23 represented by those above the dotted line.

1 **Figure 4. Second tier small-scale protein folding condition selection by MCE peak**
2 **sharpness.** The top 30 peaks by height were examined for peak sharpness, determined by a ratio
3 of MCE peak height/FWHM in fluorescence/sec. Greater sharpness scores equate to a narrower
4 width while also taking into account the overall peak size by height. The top 10 scoring
5 conditions were identified for advancement to final selection, represented by those above the
6 dotted line.

7

8

9 **Figure 5. Final selection of small-scale folding conditions by relative MCE performance in**
10 **purity, peak sharpness, and peak height for intermediate-scale screening.** The top 10
11 conditions were compared for purity, peak sharpness, and peak height, with each attribute
12 represented by a relative numeric score. The scores were adjusted to an equivalent mean ($\bar{x} =$
13 71.1) for comparability. The two highest scores for purity (A6: 78.9%, B9: 75.5%) were
14 selected, then the 2 highest scores for peak sharpness (D12: 7,068 fluorescence/sec, B10: 6,478
15 fluorescence/sec), and finally the 2 highest scores for peak height (A8: 1643 fluorescence units,
16 A9: 1603 fluorescence units), each indicated by “*”. If a condition was already selected by a
17 preceding attribute, the next best scoring condition was selected.

18

19

20

1 **Figure 6. Non-reducing SDS-PAGE of intermediate-scale fold conditions.** A NuPAGE 4-
2 12% Bis-Tris gel, with MES running buffer was used for the analysis. Lane 1: huFc standard (2
3 μg), 2: Life Technologies Mark12 MW standards, 3 – 8: Refold conditions A6, A8, A9, B9, B10,
4 D12 (6 μL each). Each aliquot was prepared in 4x NuPAGE sample buffer with iodoacetamide
5 (IAA) added to a final concentration of 50 mM in order to alkylate free sulfhydryls.
6
7

8 **Figure 7. Intermediate-scale SP HP fraction selection by MCE peak sharpness.** The
9 fractions resulting from SP Sepharose HP purification were analyzed for peak sharpness. The
10 three fractions possessing the sharpest peaks were selected for pooling (indicated by “*”).
11
12

13 **Figure 8. Non-reducing SDS-PAGE of final bulk products.** A NuPAGE 4-12% Bis-Tris gel,
14 with MES running buffer was used for the analysis. Lane 1: huFc standard (2 μg), 2: Life
15 Technologies Mark12 MW standards, 3 – 8: A6, A8, A9, B9, B10, D12 (2 μg each). Each
16 aliquot was prepared in 4x NuPAGE sample buffer with IAA added to a final concentration of 50
17 mM in order to alkylate free sulfhydryls.

1 **Supplemental Figure 1.** Biomek automated small-scale fold working deck setup. Biomek
2 software graphical representation of Biomek deck components required for the small-scale fold
3 method.

4

5

6 **Supplemental Figure 2.** Biomek automated LabChip microplate load working deck setup.
7 Biomek software graphical representation of Biomek deck components required for loading a
8 microplate with the small-scale folds for subsequent LabChip analysis.

Table 1.

	0U 8.5	0U 9.5	0U 10.5	1U 8.5	1U 9.5	1U 10.5	2U 8.5	2U 9.5	2U 10.5	4U 8.5	4U 9.5	4U 10.5
160A 1:4, 0%	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
160A 1:4, 20%	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
160A 2:2, 0%	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
160A 2:2, 20%	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
400A 1:4, 0%	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
400A 1:4, 20%	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
400A 2:2, 0%	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
400A 2:2, 20%	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

Table 2.

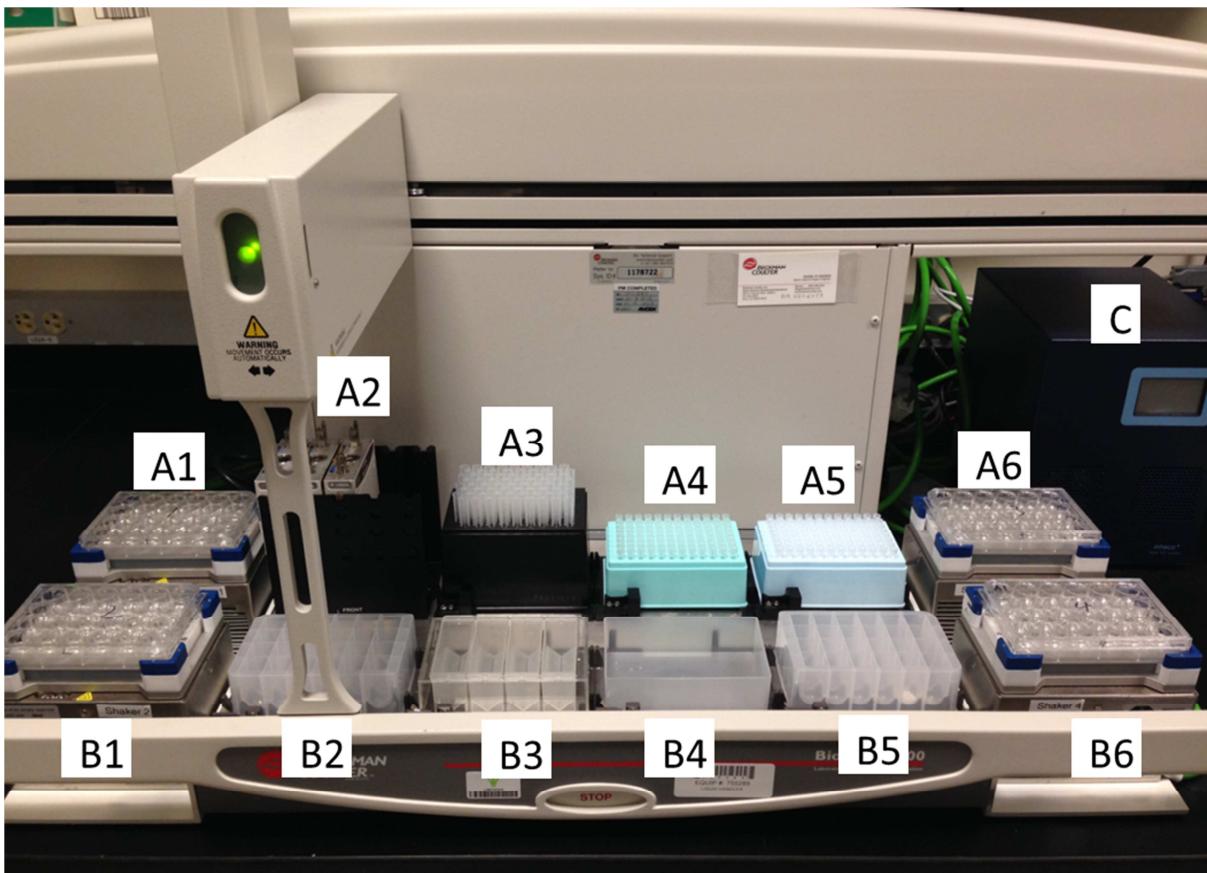
Component	Quantity	Height	Sharpness	% Purity
L-arg (mM)	160	779 ± 84	2879 ± 338	58 ± 3
	400	423 ± 55	1514 ± 222	60 ± 3
Redox (mM:mM)	1:4	707 ± 82	2593 ± 324	58 ± 3
	2:2	495 ± 66	1620 ± 229	60 ± 3
Glycerol (%)	0	677 ± 74	2419 ± 297	61 ± 3
	20	525 ± 76	1974 ± 305	57 ± 3
Urea (M)	0	240 ± 58	859 ± 224	48 ± 4
	1	717 ± 87	2683 ± 339	67 ± 3
	2	815 ± 115	3070 ± 450	69 ± 3
	4	633 ± 118	2174 ± 500	52 ± 5
pH	8.5	467 ± 80	1724 ± 328	51 ± 3
	9.5	587 ± 97	2168 ± 388	60 ± 4
	10.5	748 ± 94	2697 ± 373	65 ± 3

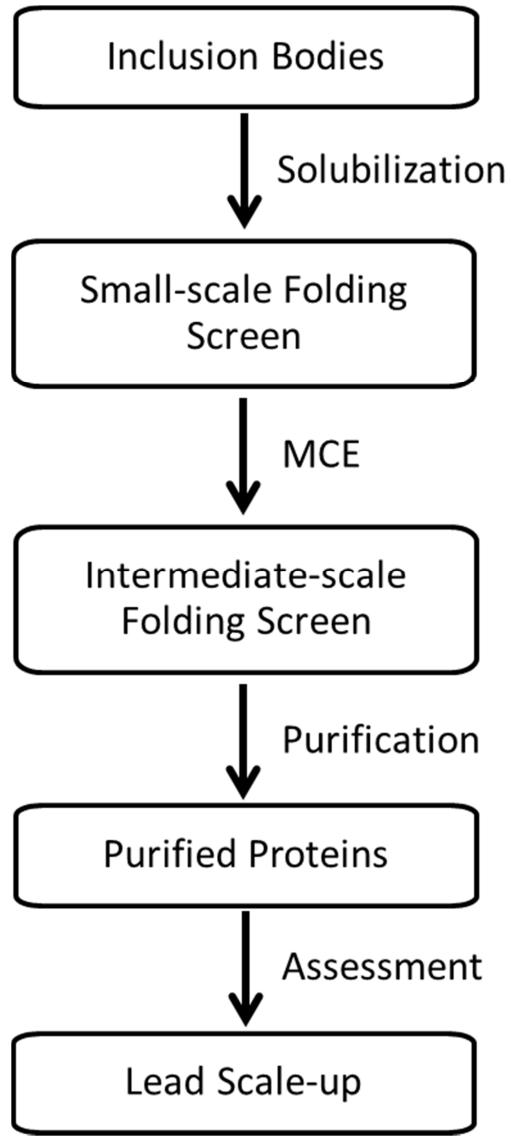
Table 3.

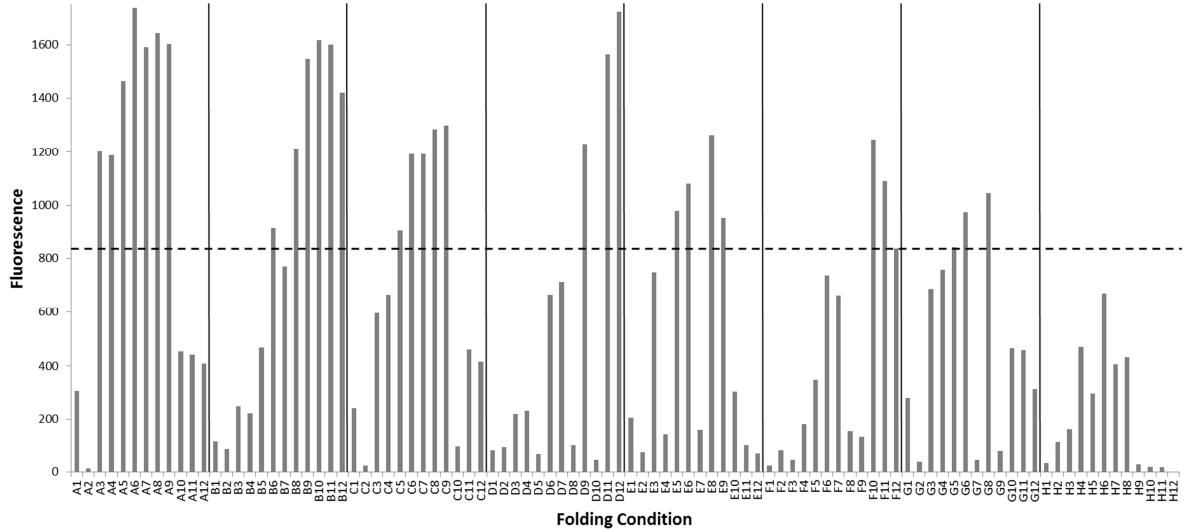
Folding Condition	Pro A Yield (mg)	SP HP Yield (mg)	SEC (%MP)	MCE (%MP)
A6: 160A, 1:4, 0%, 1U, 10.5	48.5	19.8	95.0	99.9
A8: 160A, 1:4, 0%, 2U, 9.5	33.9	22.0	95.1	100.0
A9: 160A, 1:4, 0%, 2U, 10.5	48.6	21.2	95.2	99.9
B9: 160A, 1:4, 20%, 2U, 10.5	37.8	16.9	95.3	99.6
B10: 160A, 1:4, 20%, 4U, 8.5	35.2	11.9	94.6	99.7
D12: 160A, 2:2, 20%, 4U, 10.5	28.7	17.7	95.7	99.8

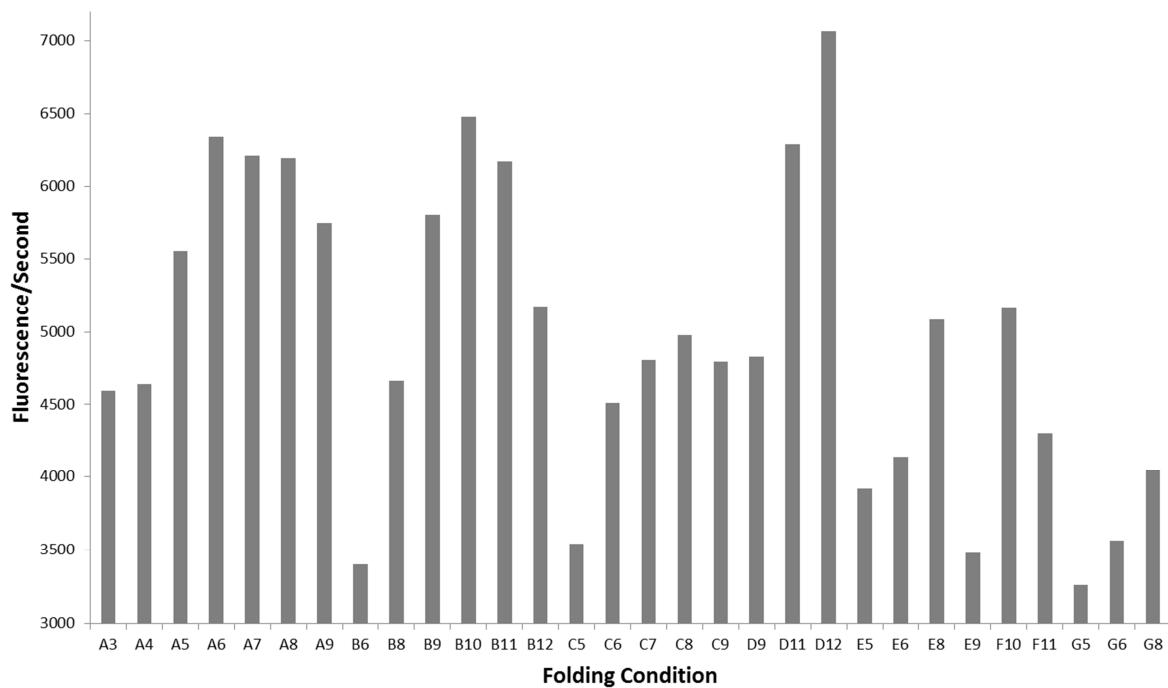
Supplemental Table 1.

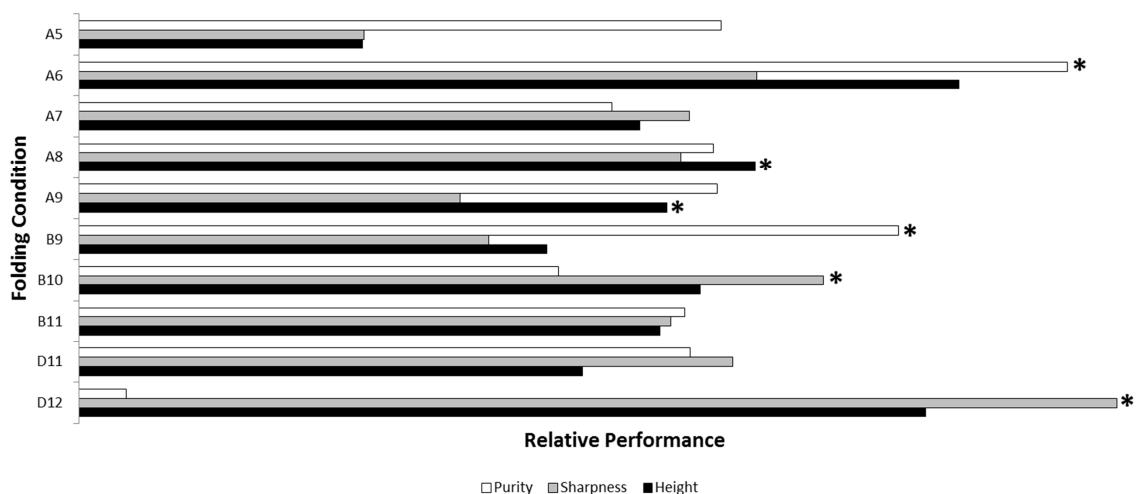
Component:	1	2	3	4	5	6	7	8	9	10	11	12	13
Tip Type:	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	200	200	200
Location:	B2-1	B2-2	B2-3	B2-4	B2-5	B2-6	B3-1	B3-3	B4	B4	B5-1	B5-2	B5-8
Solution:	1.6M L-arg-HCl	1.6M L-arg-HCl	1.6M L-arg-HCl	1.6M L-arg-HCl	1.6M L-arg-HCl	1.6M L-arg-HCl	9M urea	80% glycerol	water	water	100mM cystamine	100mM cysteine	dwIBs
Plate well	0.5M Tris, pH 8.5	0.5M Tris, pH 9.5	0.5M Tris, pH 10.5	0.2M Tris, pH 8.5	0.2M Tris, pH 9.5	0.2M Tris, pH 10.5							Solute 1:9
Total Vol.	3200	3200	3200	8000	8000	8000	37320	24000	73938	23142	2880	5760	7680

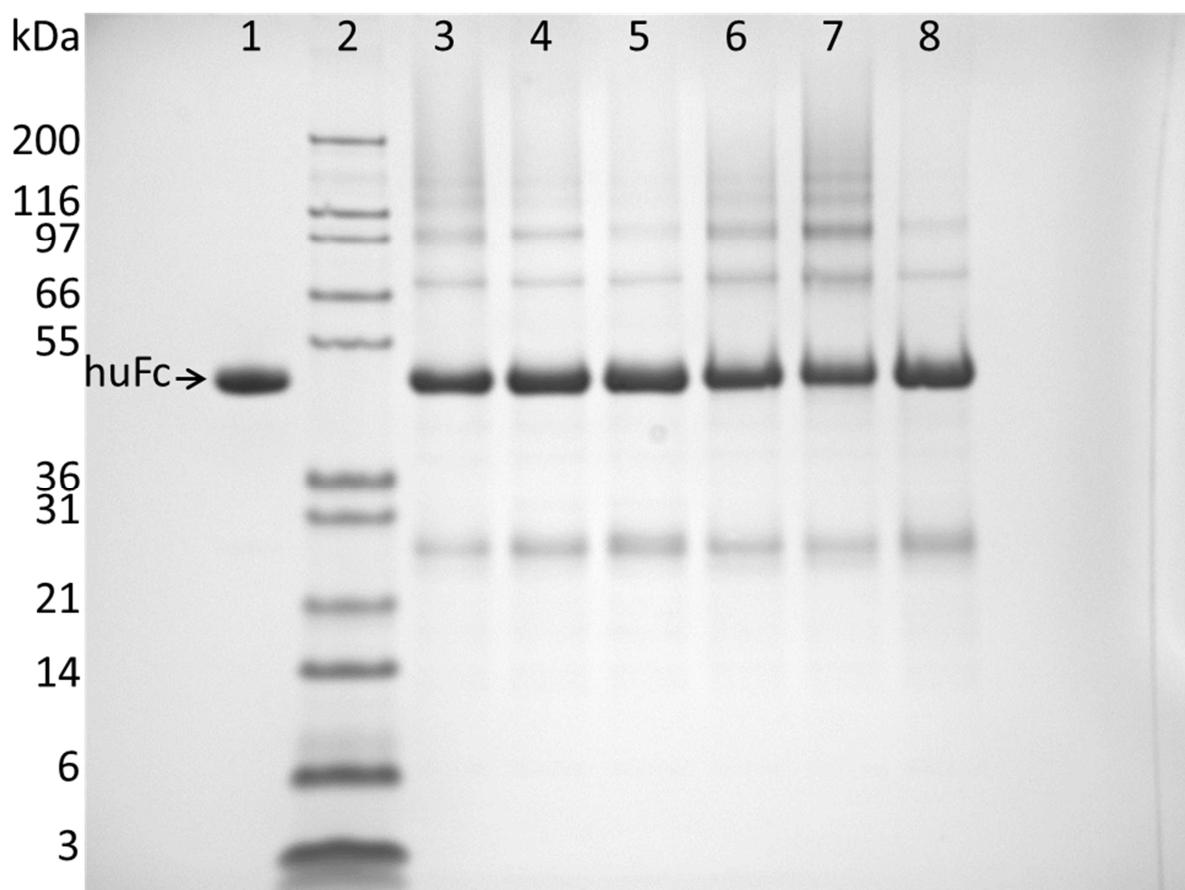


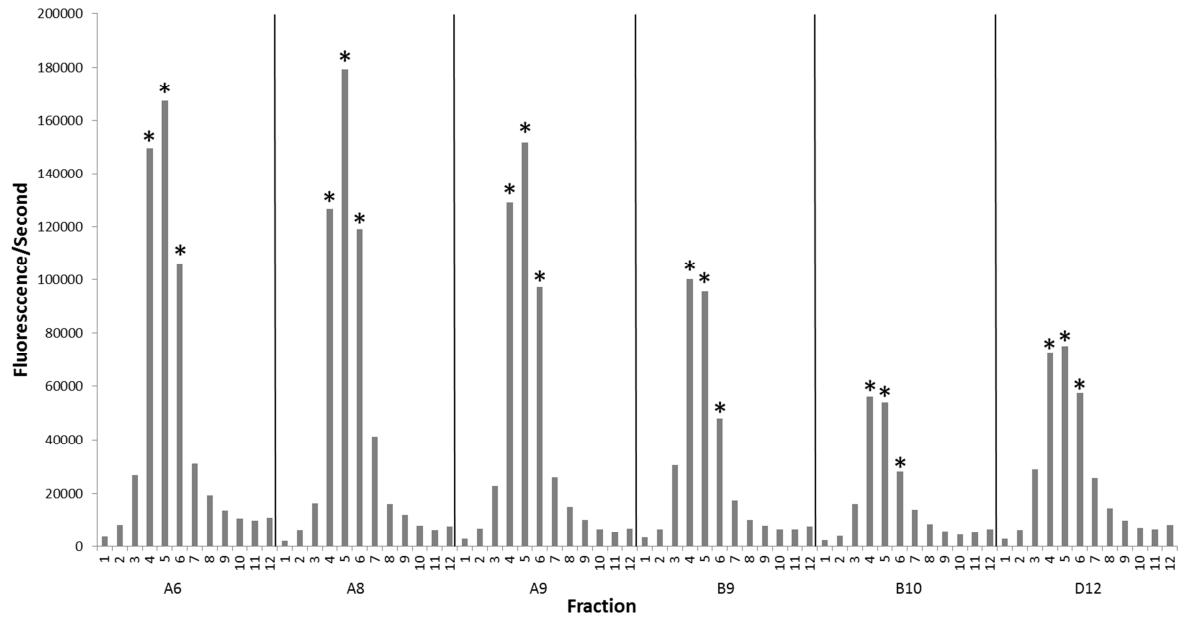


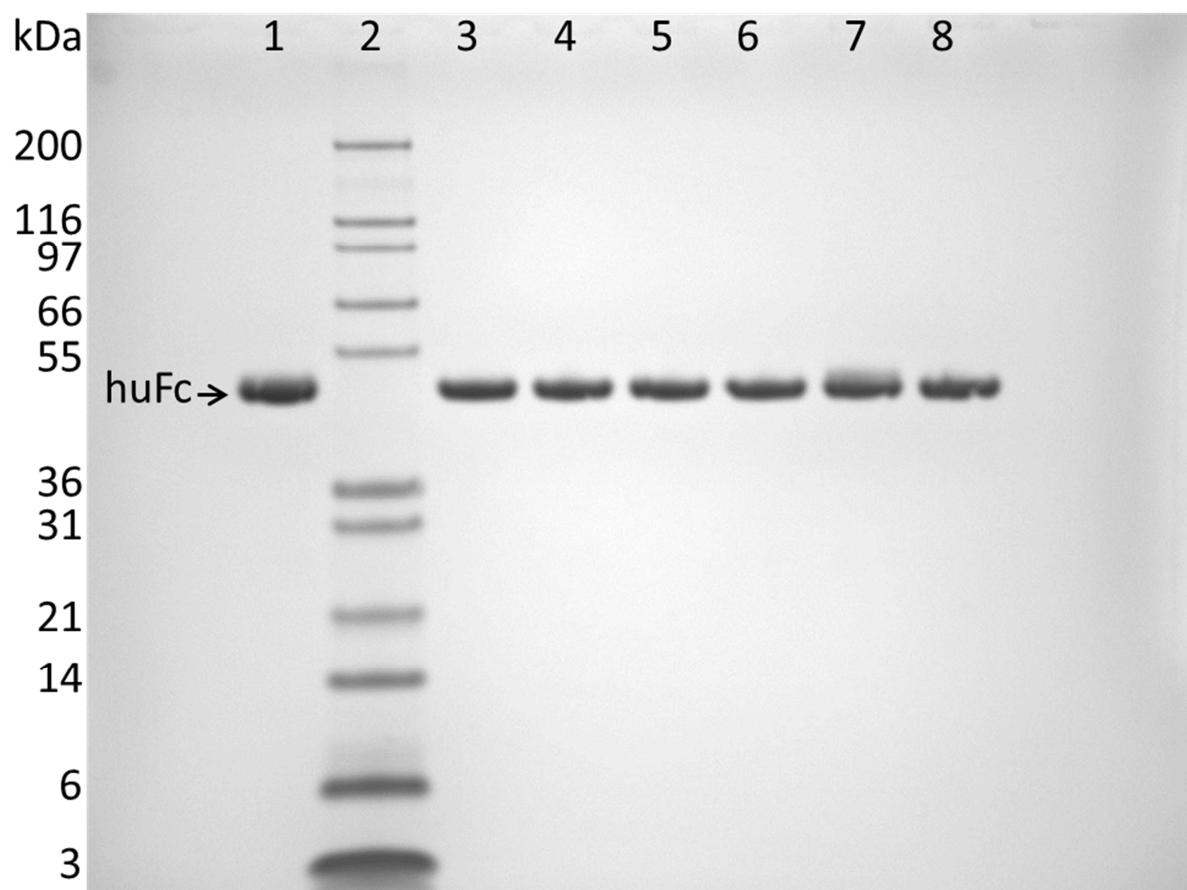












- An automated disulfide bond containing protein folding screen
- Employs a curated matrix that has been very successful for a wide variety of proteins
- Automates matrix generation as well as folding progress analytics
- Employs a quantitative defined method for lead condition identification
- Greatly reduces the time and labor required to fold novel proteins