

Biotechnology based process for production of a disulfide-bridged peptide

Animesh Goswami^{1*}, Steven L. Goldberg¹, Ronald L. Hanson¹, Robert M. Johnston¹, Olav K. Lyngberg¹, Yeung Chan¹, Ehrlic Lo¹, Steven H. Chan¹, Nuria de Mas¹, Antonio Ramirez¹, Richard Doyle², Wei Ding³, Mian Gao⁴, Stanley Krystek⁴, Changhong Wan⁴, Yeoun jin Kim⁴, Deepa Calambur⁴, Mark Witmer⁴, James W. Bryson^{4*}

¹Chemical Development, R&D, Bristol-Myers Squibb, 1 Squibb Drive, New Brunswick, NJ 08903,

²Chemical Development Operations, R&D, Bristol-Myers Squibb, New Brunswick, NJ 08903,

³Analytical and Bioanalytical Development, R&D, Bristol-Myers Squibb, New Brunswick, NJ 08903,

⁴Discovery, R&D, Bristol-Myers Squibb, Route 206 and Province Line Road, Princeton, NJ 08543,

* Corresponding author: Animesh Goswami, Chemical Development, R&D, Bristol-Myers Squibb, 1 Squibb Drive, New Brunswick, NJ 08903, USA, E-mail: animesh.goswami@bms.com, Tel: 732-227-6225, FAX: 732-227-3994.

SUPPORTING INFORMATION

List of contents

1. Cloning and expression of individual chain constructs
2. Maltose binding fusion protein and generation of PKC and TAT
3. IF2 fusion proteins with methionine cleavage site and generation of PKC and TAT
4. IF2 fusion proteins with tryptophan cleavage site and generation of PKC and TAT
5. SET fusion proteins with methionine and enterokinase cleavage site
6. Approach to PKC and TAT from SUMO fusion proteins
7. Approach to PKC and TAT via Inteins
8. CD spectra of heterodimeric fusion protein
9. Fermentations in 5L scale for production of heterodimeric protein
10. Isolation of heterodimeric fusion proteins from cell paste
11. Tangential flow filtration for the preparation of clarified heterodimeric fusion protein solution
12. Heterodimeric fusion protein after separation on Ni-Sepharose column
13. ^1H -NMR spectra of disulfide-bridged peptide
14. ^{13}C -NMR spectra of disulfide-bridged peptide

1. Cloning and expression of individual chain constructs

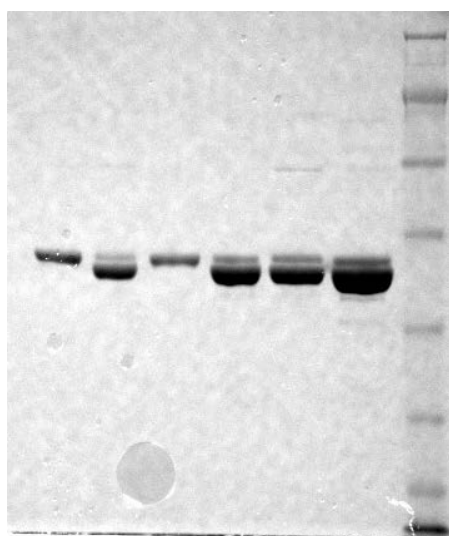
Synthetic genes prepared (for maltose binding protein fusion protein) contained four domains: (1) the maltose binding protein (MBP); (2) six consecutive histidine residues (His6 tag, or H6); (3) a bovine enterokinase (EK) cutting site (DDDDK), and (4) either the PKC or the TAT peptide. Synthetic genes encoding both MBP-His₆-EK-TAT (GenBank Accession numbers KU685471) and MBP-His₆-EK-PKC (GenBank Accession numbers KU685472) were ligated to expression plasmid pBMS2004, inserted into competent cells, and kanamycin resistant colonies with correct sequences for the maltose binding fusion protein (Fig 2a) were isolated. PKC and TAT maltose binding fusion protein plasmids were separately transformed into chemically-competent *Escherichia coli* BL21Gold cells and single colonies were isolated on LB kanamycin agar plates. A single colony from each plate was inoculated into 10 mL MT5-M2 Km medium in a 50 mL flask. MT5-M2 Km medium contains 2% pea hydrolyzate (Hy-Pea, Marcor, Carlstadt, NJ), 1.85% yeast extract (Tastone 154, Sensient, Indianapolis, IN), 0.6% Na₂HPO₄, 0.125% (NH₄)₂SO₄, 4.0% (w/v) glycerol, pH adjusted to 7.2 with 10 N NaOH prior to autoclaving. Sterile kanamycin sulfate was added to a final concentration of 50 µg/mL post-autoclaving. The flask was incubated at 30°C, 250 rpm for 22 hr. Each culture was diluted into 50 mL fresh MT5-M2 Km medium cultures at an initial OD₆₀₀ of 0.2. At an OD₆₀₀ of ca.1.0, the cultures were induced with isopropyl β-D thiogalactoside (IPTG) from a filter sterilized 1 M solution to a final concentration of 0.1 to 1.0 mM. Cells were pelleted by centrifugation at 5000g. To analyze the formation of fusion proteins, cells were suspended in phosphate buffer pH 7 containing 1 mM DTT and disrupted using the FastPrep unit (MP Biomedicals). The samples were centrifuged and the supernatant was analyzed by SDS-PAGE. Both desired fusion proteins were highly overexpressed and the heterologous proteins were found almost exclusively in the soluble protein fraction. The best expression conditions for BL21Gold(MBP-PKC) and BL21Gold(MBP-TAT) was overnight induction with 0.1 mM IPTG and 0.5 mM IPTG, respectively.

Expression and fermentations of IF2, SET, SUMO, and Intein fusion proteins with different cleavage sites were carried out in a similar manner. Various *E. coli* strains were used for expression, e.g. BL21Gold, BL21(DE3), MM294, W3110, LE392.

2. Maltose binding fusion protein and generation of PKC and TAT

Recombinant *E. coli* cells (1.98 g TAT-fusion protein cells or 1.79 g PKC-fusion protein cells) were disrupted by sonication in 10 ml buffer at 4°C. The extracts were clarified by centrifugation and the His-tagged fusion proteins were purified on columns containing 2 ml Ni-NTA agarose (Qiagen) packed in buffer and eluted with buffers containing increasing imidazole concentration. The fractions containing the purified fusion proteins were identified using SDS gels and dialyzed into the enterokinase cleavage buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 2 mM CaCl₂). Protein concentrations were determined with the Bio-Rad protein assay. To 1.7 ml samples containing 29 mg MBK-TAT or 43 mg MBP-PKC were added a 34 µl solution containing 0.068 µg enterokinase (New England Biolabs) (~0.0002% w/w of fusion protein) and the samples were incubated for 19h at 28°C. Samples were diluted 1:20 for analysis by SDS-PAGE.

SDS gel showing cleavage of the MBP fusion proteins by enterokinase (EK)

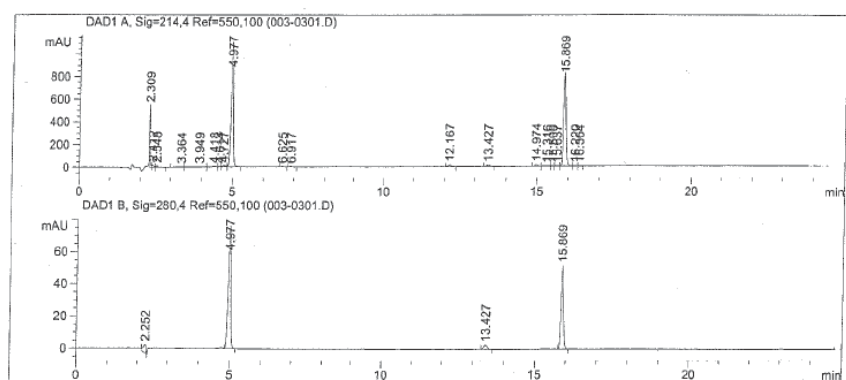


Lanes 1-4 2 μ l, lanes 5-6 5 μ l of EK-treated sample. Lane 1, MBP-TAT; lanes 2 and 5, MBP-TAT+EK; lane 3, MBP-PKC; lanes 4 and 6, MBP-PKC +EK, lane 7, protein molecular weight markers (fusion protein is between 51 and 39 kD markers).

TCEP.HCl (0.05 ml of 100 mM solution) was added to 0.5 ml samples of enterokinase treated fusion proteins. After a 1-h incubation at room temperature the solutions were centrifuged, and the supernatants were centrifuged through a 10000 MW cutoff membrane. Trifluoroacetic acid (TFA, 15 μ l) was added to each sample to acidify the Tris-buffered solution and keep the cysteines reduced. PKC and TAT peptides were quantitated by HPLC and analyzed by LC-MS.

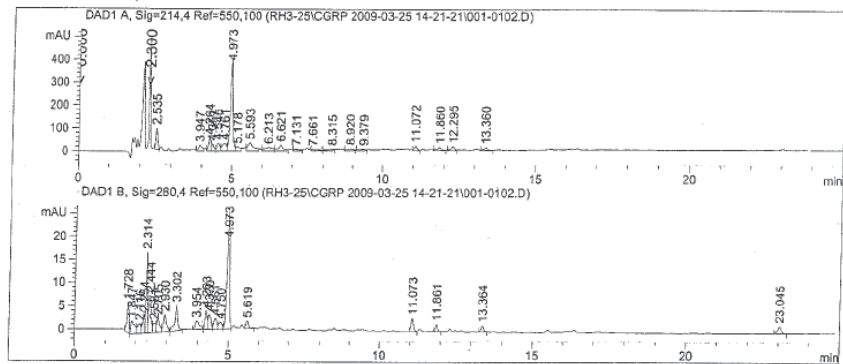
HPLC:

TAT 4.9 min, PKC 15.8 min



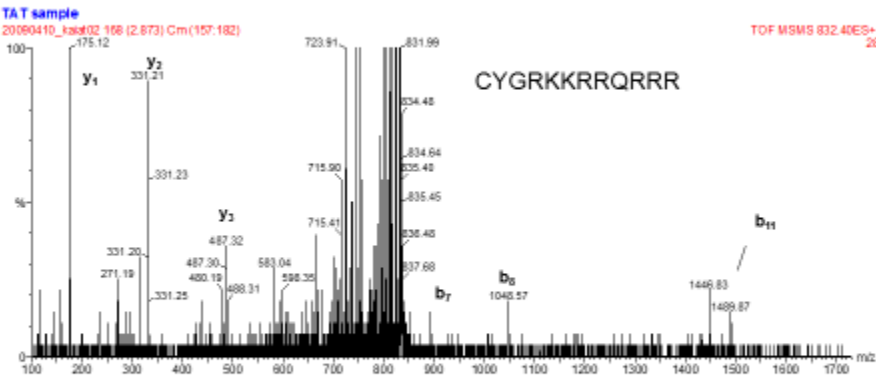
Sample from Enterokinase cleavage of MBP-EK-TAT

HPLC: TAT 4.9 min



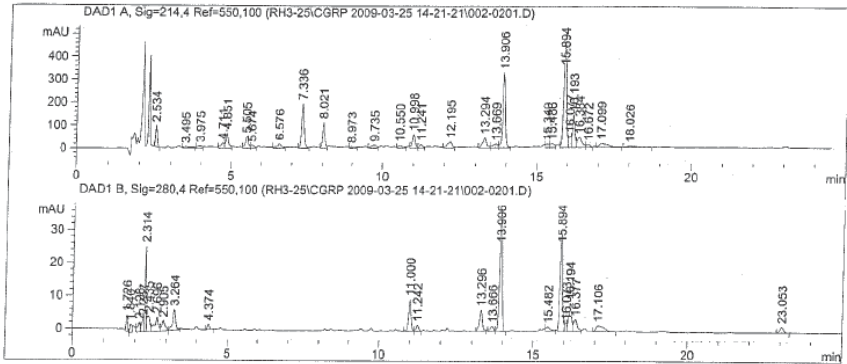
Mass Spec

MS/MS of Carrier_TAT peptide +2 precursor



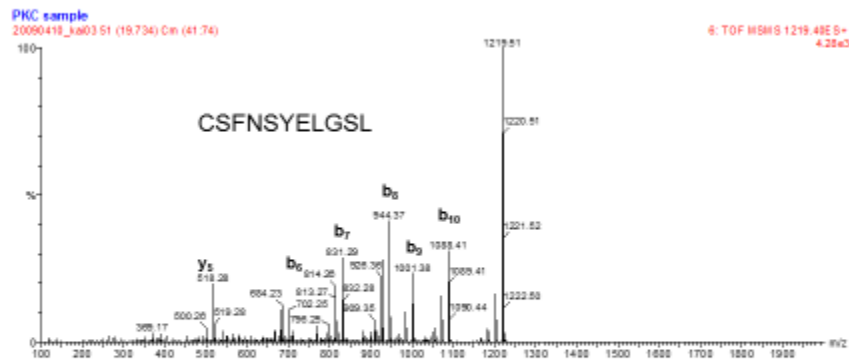
Sample from Enterokinase cleavage of MBP-EK-PKC

HPLC: PKC 15.8min



Mass spec

MS/MS of Cargo_PKC peptide +1 precursor



3. IF2 fusion proteins with methionine cleavage site and generation of PKC and TAT

Synthetic genes encoding IF2-His₆-Met-TAT/PKC (GenBank Accession numbers KU685473/KU685475) containing methionine as the cleavage site were created (Figure 2b). Both were expressed in *E. coli* and cells were produced by fermentation. Cell paste from *E. coli* strains BL21Gold(pBMS2004-IF2-His₆-Met-TAT (13.04 g) and BL21Gold(pBMS2004-IF2-His₆-Met-PKC (6.52 g) were processed separately as above using 10 ml Ni-NTA agarose. 7.97 mg IF2-His₆-Met-TAT fusion protein and 21.6 mg IF2-His₆-Met-PKC fusion protein were obtained per g cell paste. CNBr cleavage of solution containing 5 mg of fusion proteins were conducted by mixing with CNBr (100 equiv) at 25°C at pH 1.5 under closed condition followed by distillation under reduced pressure to remove excess CNBr and MeSCN in a ventilated hood taking appropriate precautions. HPLC showed yields (~13%) of 54 µg TAT and 38 µg PKC, respectively. LC-MS confirmed the respective masses.

4. IF2 fusion protein with tryptophan cleavage site and generation of PKC and TAT

Synthetic genes encoding IF2-His₆-Trp-TAT/PKC (GenBank Accession numbers KU685474/KU685476) containing tryptophan as the cleavage site were created (Figure 2b). Both were expressed in *E. coli*, and cells were produced by fermentation. Cell pellets (1.3 g each) of *E. coli* BL21Gold(pBMS2004-IF2-His₆-W-TAT and BL21Gold(pBMS2004-IF2-His₆-W-PKC containing His-tagged TAT and PKC IF2 protein fusions were suspended in 6 ml of binding buffer and disrupted by sonication at 4°C. The sonicates were clarified by centrifugation and applied to columns containing 2 ml Ni Sepharose 6 Fast Flow (GE Healthcare) packed in binding buffer. The columns were washed with 10 ml binding buffer, then eluted with 1 ml elution buffer per fraction. Binding buffer (pH 7.4) contained 20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole and elution buffer (pH 7.4) contained 20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole. Fractions were analyzed on NuPAGE Bis-Tris 12% 1 mm 10 lane gels using MES running buffer and reducing conditions. Protein concentration in the

eluted fractions was measured with the BioRad protein assay using BSA as standard. The eluted fractions were dialyzed into 80% acetic acid/20% water or precipitated by addition of 2 volumes of absolute ethanol for cleavage studies. The overexpressed TAT fusion protein was completely bound by the affinity column. 44 mg protein was obtained in the eluted fractions. The overexpressed PKC fusion protein was not entirely captured by the affinity column. 48 mg protein was obtained in the eluted fractions.

For cleavage, 1mL fusion protein solution in 80% acetic acid (about 10mg/mL) was mixed with a solution of BNPS-skatole (2mg) in acetic acid (0.5mL) and phenol (10mg). After shaking for 18 hr at 25°C, 100mM TCEP.HCl solution (0.1mL) was added and the mixture was shaken for 1 hr. The resulting mixture was analyzed by SDS-PAGE, HPLC and LC-MS.

5. SET fusion proteins with methionine and enterokinase cleavage site and generation of PKC and TAT

SET fusion proteins with methionine cleavage sites (GenBank Accession numbers TAT - KU685477, PKC-KU685478) (Figure 2c) were expressed in *E. coli*. Fermentation provided 6.98 mg TAT and 1.91 mg PKC fusion proteins per g of cells. Cleavage of fusion proteins by CNBr provided PKC. However, no TAT was detected. Instead, a truncated TAT peptide with loss of two terminal arginine was observed.

SET fusion proteins with enterokinase cleavage site (GenBank Accession numbers TAT - KU685479, PKC-KU685480) (Fig 2d) were prepared in similar fashion providing about 3.53 mg TAT fusion protein and 1.39 PKC fusion protein per g of cells. Gel analysis after enterokinase treatment indicated complete cleavage of the PKC fusion protein and most of the TAT fusion protein was cleaved. Addition of 5 mM dithiothreitol or TCEP during the cleavage of the TAT fusion protein allowed

complete cleavage. LC-MS analysis of the samples found the intact SET partner 1-56 but PKC or TAT was not found. LC-MS analysis indicated enterokinase digestion also removed the two C-terminal amino acids from TAT.

6. Approach to PKC and TAT from SUMO fusion proteins

SUMO or small-ubiquitin modifier protein of *Saccharomyces cerevisiae* [Saitoh, H., Pu, R. T., and Dasso, M. (1997) SUMO-1: Wrestling with a New Ubiquitin-Related Modifier. *Trends in Biochemical Science* 22, 374-376] was shown to be a very good fusion partner because of its generally high level of soluble expression in *E. coli*. In addition, SUMO is a relatively small protein of about 11.5 kD and is specifically cleaved at GG-X (X representing the fusion partner) by the Ulp1 protease [Mossessova, E. and Lima, C. D. (2000) Ulp1-SUMO Crystal Structure and Genetic Analysis Reveal Conserved Interactions and a Regulatory Element Essential for Cell Growth in Yeast. *Molecular Cell* 5, 865-876] which is commercially available or can be produced by recombinant *E. coli* [Lee, CD., Sun, H.-C., Hu, S.-M., Chiu, C.-F. Chiu, Homhuan, A., Liang, S.-M., Leng, C.-H., Wang, T.-F. (2008) An improved SUMO fusion protein system for effective production of native proteins. *Protein Science* 17, 1241-1248]. The native gene encoding SUMO was used as an N-terminal fusion partner with the PKC and TAT peptides. However, no expression of the fusion protein could be detected. A new leader sequence that included a spacer region between the His₆ tag and start of the SUMO protein was used (Fig 2e), leading to good levels of soluble fusion protein. A commercial preparation of Ulp1 protease did specifically cleave the SUMOInv-PKC and SUMOInv-TAT proteins, albeit with different efficiencies. SDS gels showed cleavage of the fusion proteins by the commercial and our in-house produced SUMO proteases. In-house SUMO proteases with 1 or 2 His tags were prepared and both cleaved the SUMO-TAT and SUMO-PKC fusion proteins. However analysis by LC-MS showed only very low levels of PKC together with peptide fragments of PKC and no TAT in the cleaved samples. A peptide with a similar retention time to TAT was found that was 12 Da larger. The TAT sequence is highly basic due to

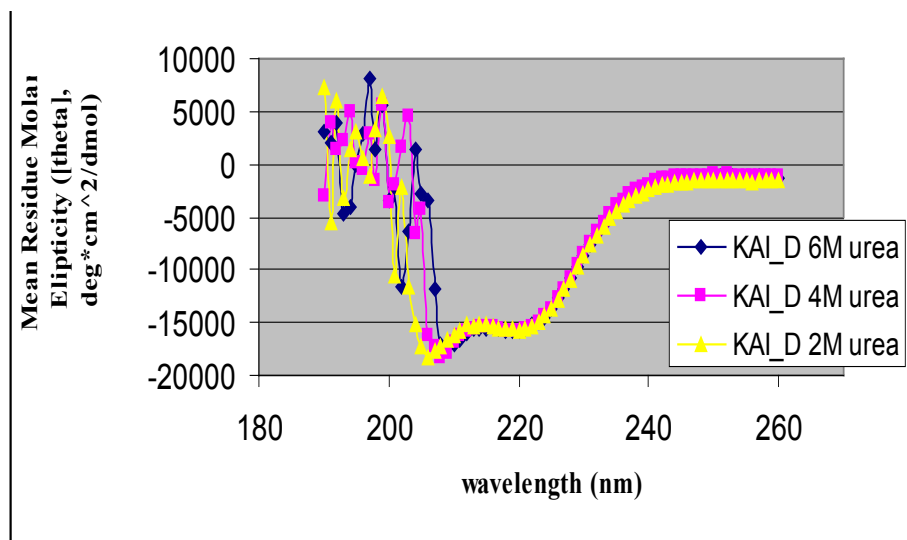
the presence of several Arg residues and this may affect the binding and/or catalytic efficiency of the protease.

7. Approach to PKC and TAT via Inteins

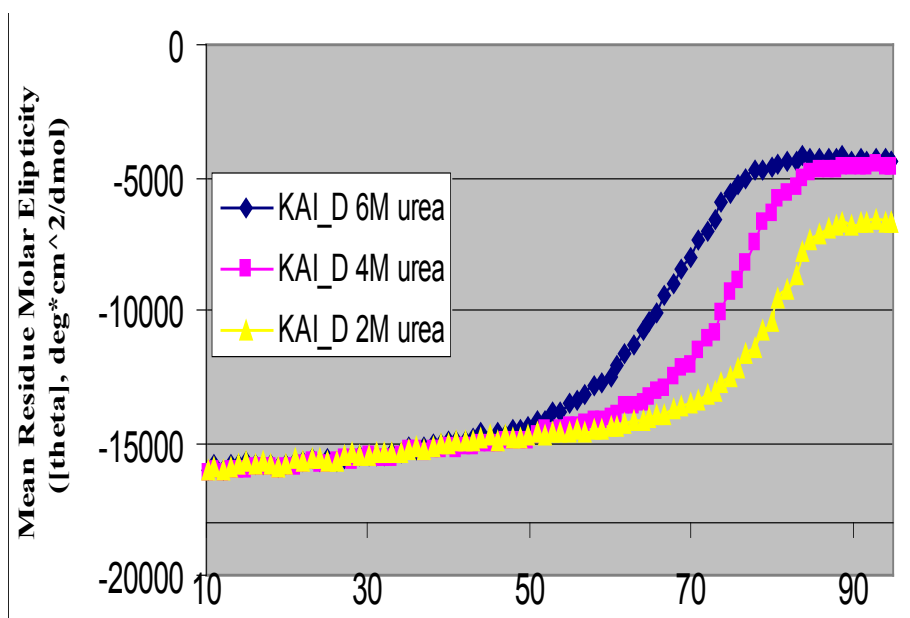
The self-splicing “Inteins” were also tried as fusion protein partner [Evans, T. C., Benner, J., Xu, M.-Q. (1999) The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J. Biol. Chem.* 274, 18359-18363; Mathys, S., Evans, T.C., Chute, I.C., Wu, H., Chong, S., Benner, J., Liu, X.-Q., Xu, M.-Q. (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* 231, 1-13; Southworth, M.W., Amaya, K., Evans, J., T.C., Xu, M.-Q., Perler, F.B. (1999) Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *BioTechniques* 27, 110-120]. The intein constructs contained a chitin binding domain (CBD) for affinity purification on chitin beads fused by inteins to either the amino terminal (DnaB) or carboxy terminal (gyrA) ends of the PKC and TAT peptides (Fig 2f). For fusion proteins CBD-DnaB-PKC and CBD-DnaB-TAT, SDS gel analysis of the fractions before and after elution with cleavage conditions showed no cleavage of the overexpressed fusion proteins and TAT and PKC peptides were not detected by HPLC. For fusion proteins TAT-gyrA-CBD and PKC-gyrA-CBD, the TAT fusion protein was purified by the chitin column but the PKC fusion protein did not bind to the column. SDS gel analysis of the fractions before and after cleavage conditions showed no cleavage of the overexpressed TAT fusion protein and TAT and PKC peptides were not detected by HPLC.

8. CD spectra of a representative heterodimeric fusion protein

a. Wavelength scan in 6M, 4M and 2M urea



b. T_m in 6M, 4M and 2M urea



9. Fermentations in 5L scale for production of heterodimeric protein

One vial containing one mL of the culture was inoculated into 200 mL MT5-M2 Km medium in a 1 L flask. Flasks were shaken at 30°C, 250 rpm, for one day. An appropriate volume of the grown culture to give a predicted starting OD₆₀₀ of about 0.7 was added to 2 L of MT5-M2 Km medium in a 5 L fermentor. After appropriate growth, a filter sterilized-solution of IPTG was added. The cells were harvested by centrifugation (5000g for 10 min).

10. Isolation of heterodimeric fusion proteins from cell paste

All buffers contained 20 mM sodium phosphate, 0.5 M NaCl and 2 M urea and were at pH 7.4 for initial studies but pH 8 for final batches. Binding buffer contained 20 mM imidazole, wash buffer contained 50 mM imidazole and elution buffer contained 0.5 M imidazole. Cells were suspended in binding buffer at 15% w/v then disrupted by sonication for small volumes or microfluidization at 13000 psig for larger volumes. Crude extracts were clarified by centrifugation (small scale) or cross-flow filtration (0.1 µm filter cassette). Clarified extracts were applied at 2 ml/min to a 16 mm diameter column containing 10 ml Ni Sepharose 6 Fast Flow packed in binding buffer. All operations and storage of protein solutions were at 4°C except the filtration and the column purification were carried out at ambient temperature. The column was washed with 80 ml binding buffer containing 20 mM imidazole, 40 ml wash buffer containing 50 mM imidazole, and eluted with 80 ml buffer containing 0.5 M imidazole. Wash fractions were 50 ml and eluted fractions were 10 ml. Column effluent was monitored continuously at 280 nm and fractions were assayed by SDS PAGE, BioRad protein assay, Agilent 2100 Bioanalyzer Labchip 80 under reducing and non-reducing conditions and by HPLC. The identity of the HPLC peaks was established by LC-MS.

A sample of fusion protein was treated with TCEP to reduce the disulfide bond then dialyzed

into solutions at pH 7 or 8 with various additives and lacking the reducing agent to determine optimum conditions for formation of the disulfide bond. One day of aging after the affinity column gave a maximum yield of the disulfide linked heterodimer, whereas the extract held for 1 day was only about half converted to the disulfide-linked heterodimer. Formation of the disulfide linked dimer occurred more quickly after the protein had been purified on the Ni sepharose column rather than in the extract, probably because of the removal of reducing agents found in the cell extract by the affinity purification. In initial experiments ethanol or trichloroacetic acid were used to precipitate the fusion protein from the column eluate prior to CNBr cleavage, but better results were obtained by using the eluate directly for the CNBr cleavage.

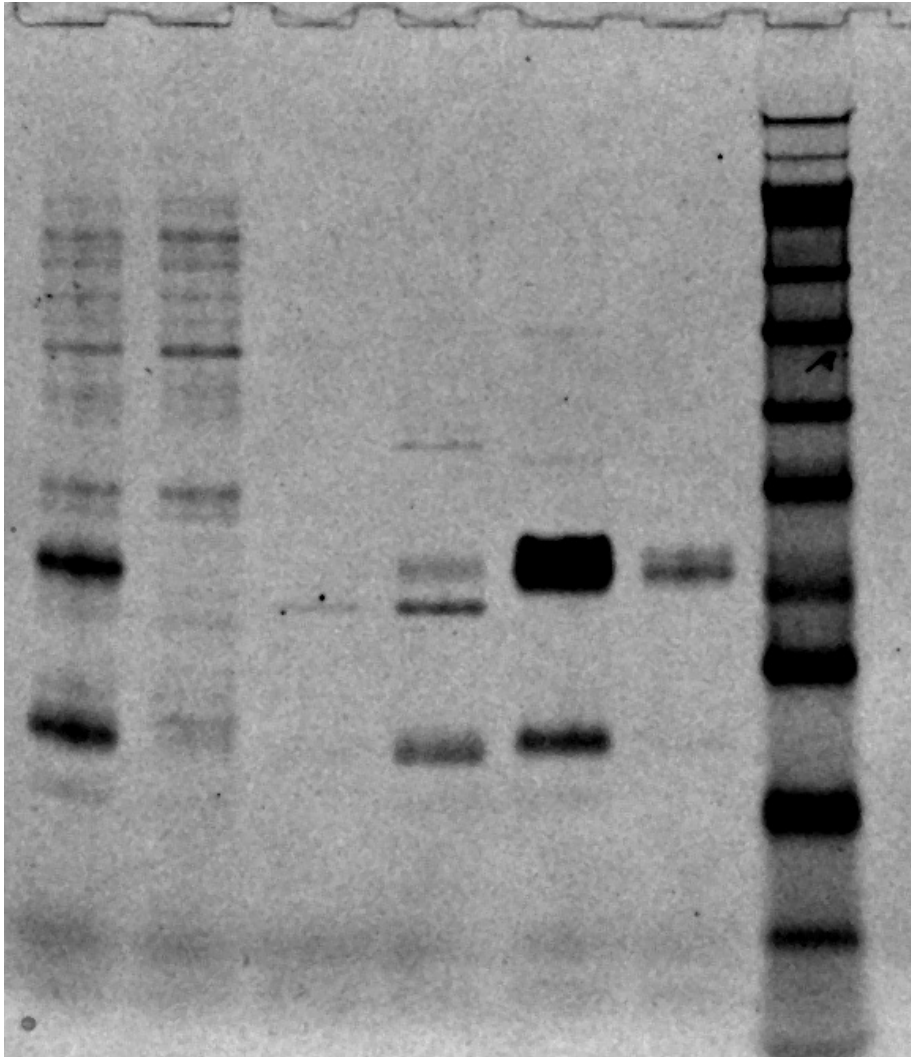
11. Tangential flow filtration for the preparation of clarified heterodimeric fusion protein solution

The recovery of the desired fusion protein from the crude fermentation centrate by tangential flow filtration was performed using a single use TFF cassette (2.5 m², 0.1 µm) on a Novasep/ TanenX Tangential Flow Filtration System (TFF System). In this microfiltration process, the filtrate stream contains the fusion protein while the retentate stream contains the unwanted cell debris. This pilot system processed approximately 16 L of fermentation centrate each time. After the installation of a new TFF cassette, approximately 50 L of binding buffer solution was used to equilibrate and to completely wet the polymeric membrane prior to start of the microfiltration operation. The entire microfiltration process was maintained under a nearly constant filtrate flux rate of 120 mL/ min by restricting the filtrate flow-rate using a peristaltic pump. The 16 L of centrate was concentrated in the TFF system to a minimal volume of about 1.5 to 2 L. Then 9 L of binding buffer was used to dilute the concentrate and was re-concentrated to a minimal working volume. This dilution/ concentration procedure was repeated a total of three times for optimal recovery of the fusion protein in the filtrate streams. For better quantitation of the concentrated retentate stream at the end of the microfiltration, 4.5 L of binding buffer was recirculated in the membrane for 5 to 10 minutes to displace any fusion protein which may have

been trapped on the membrane pores.

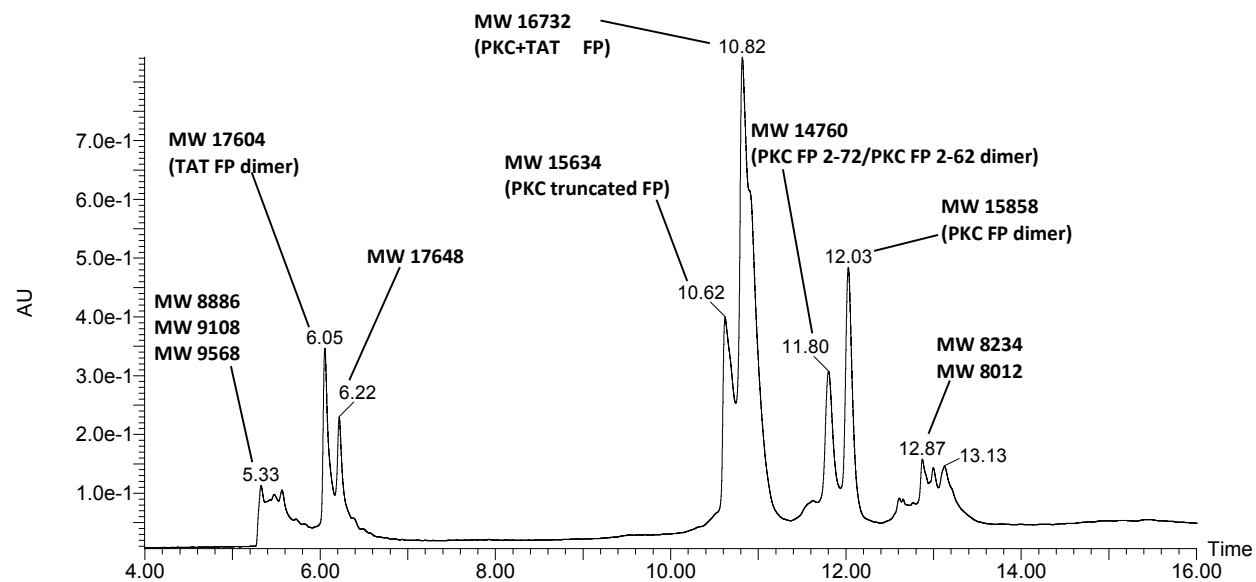
12. Heterodimeric fusion protein construct B after separation on Ni-Sepharose column

SDS gel analysis of Ni Sepharose column fractions



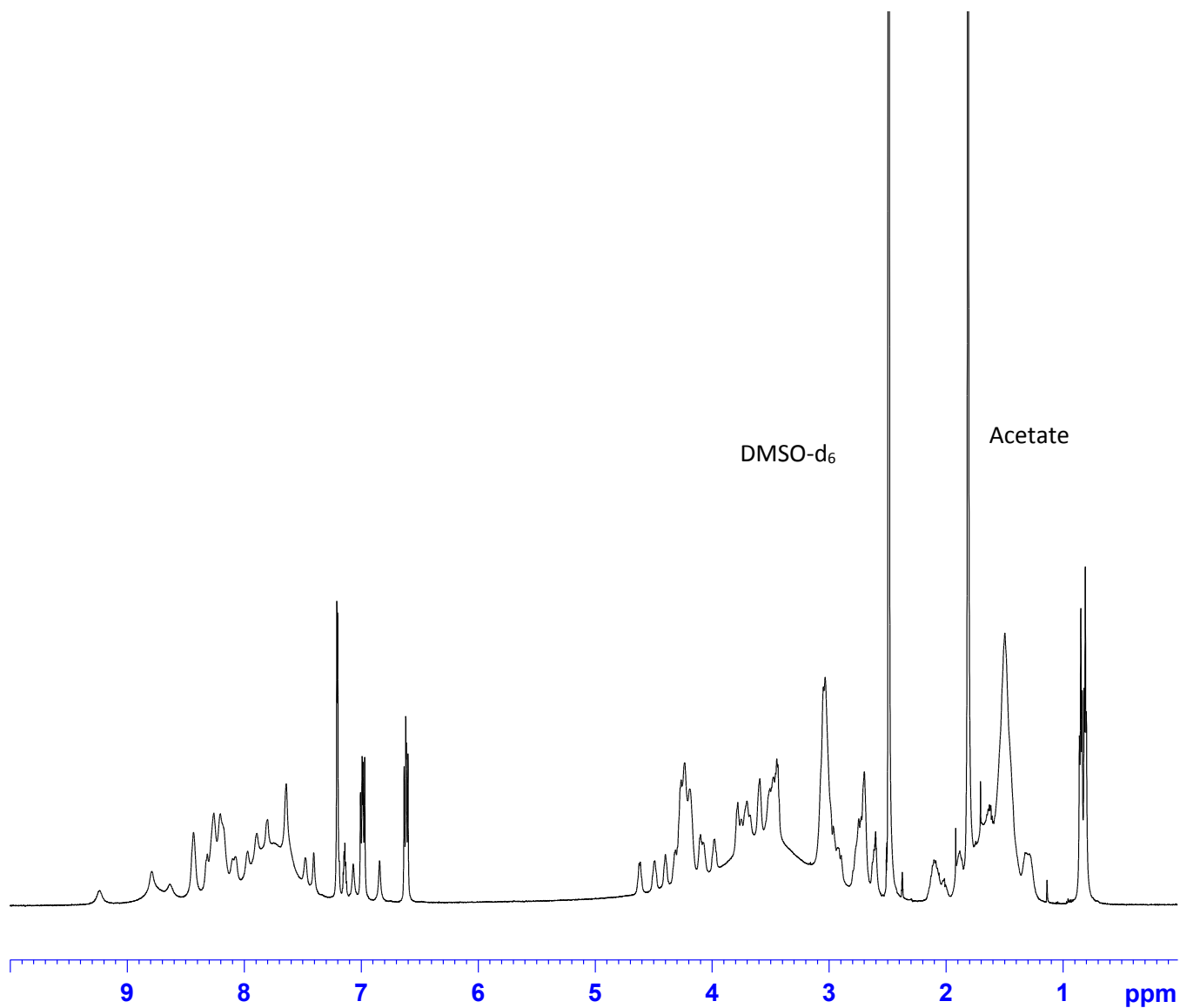
Lane 1, filtrate; lane 2, column flow through; lane 3, wash (20 mM imidazole); lane 4, 0.05 M imidazole wash; lane 5, peak fraction a2 diluted 1:10 with column buffer; lane 6, peak fraction a2 diluted 1:50; lane 7, protein molecular weight markers.

LC-MS analysis of peak fraction a2 from Ni Sepharose column



13. ^1H -NMR spectra of disulfide-bridged peptide

^1H NMR spectrum (600.67 MHz, DMSO- d_6), Acetate salt



14. ^{13}C -NMR spectra of disulfide-bridged peptide

^{13}C NMR spectrum (151.05 MHz, DMSO-d_6), Acetate salt

