Journal of Biomolecular NMR, **13:** 149–159, 1999.
KLUWER/ESCOM

Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation

Matthew J. Wood & Elizabeth A. Komives*
Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093-0601, U.S.A.

Received 16 September 1998; Accepted 19 October 1998

Key words: $(^{15}NH_4)_2SO_4$, ^{13}C-glucose, disulfide bond, glycosylation, protein expression

Abstract

Heterologous expression in *Pichia pastoris* has many of the advantages of eukaryotic expression, proper folding and disulfide bond formation, glycosylation, and secretion. Contrary to other eukaryotic systems, protein production from *P. pastoris* occurs in simple minimal defined media making this system attractive for production of labeled proteins for NMR analysis. *P. pastoris* is therefore the expression system of choice for NMR of proteins that cannot be refolded from inclusion bodies or that require post-translational modifications for proper folding or function. The yield of expressed proteins from *P. pastoris* depends critically on growth conditions, and attainment of high cell densities by fermentation has been shown to improve protein yields by 10–100-fold. Unfortunately, the cost of the isotopically enriched fermentation media components, particularly $^{15}NH_4OH$, is prohibitively high. We report fermentation methods that allow for both ^{15}N-labeling from $(^{15}NH_4)_2SO_4$ and ^{13}C-labeling from ^{13}C-glucose or ^{13}C-glycerol of proteins produced in *Pichia pastoris.* Expression of an 83 amino acid fragment of thrombomodulin with two N-linked glycosylation sites shows that fermentation is more cost effective than shake flask growth for isotopic enrichment.

Abbreviations: HSQC, heteronuclear single quantum coherence spectroscopy; TM, thrombomodulin; EGF, epidermal growth factor; TMEGF45, the fourth and fifth EGF-like domains of thrombomodulin.

Introduction

P. pastoris yeast can be grown on minimal nutrients and can be used to express large quantities of proteins that are not easily obtained from *E. coli* (Cregg et al., 1993). Proteins with disulfide bonds and proteins requiring glycosylation or other post-translational modifications are often difficult, if not impossible, to obtain from *E. coli* expression. These types of proteins can be produced in their correctly folded native states and with full activity from *P. pastoris*. For *P. pastoris*, isotopic enrichment will never be as cost effective as is possible in *E. coli*, because protein yield is correlated with high cell density, most of the isotopes end up in cell mass, and thus the yield of desired protein per gram of labeling reagents is much lower.

*To whom correspondence should be addressed.

Thrombomodulin (TM) is a 70 kDa endothelial cell surface glycoprotein consisting of an extracellular N-terminal domain, six epidermal growth factor (EGF)-like domains, an O-glycosylated domain, a membrane spanning domain, and an intracellular C-terminal domain. TM allosterically modulates thrombin resulting in a change in macromolecular specificity, inhibiting fibrinogen clotting while enhancing protein C activation. The TM fragment composed of the fourth and fifth EGF-like domains (TMEGF45) is the smallest fragment that retains both anticoagulant activities (Figure 1). We have expressed this fragment in *P. pastoris*. The protein is correctly folded, and is glycosylated with high mannose-type sugars at two N-linked glycosylation sites. The attached glycans are important for protein solubility, particularly at the high concentrations required for NMR analysis (D. Meininger, unpublished results). We have determined

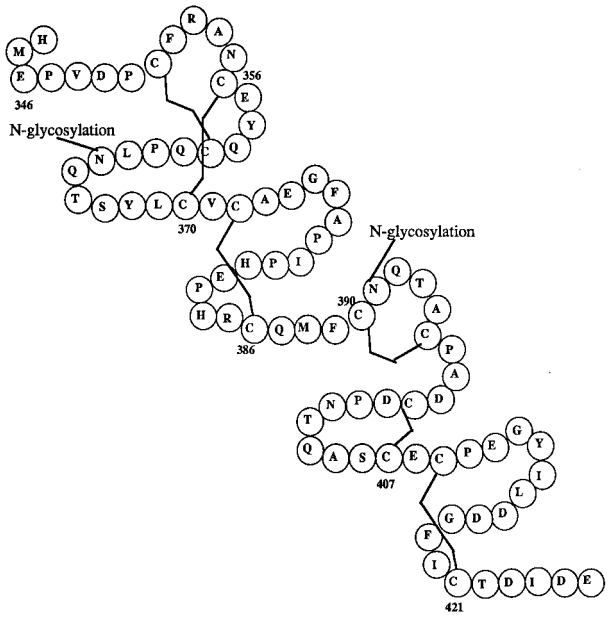

Figure 1. Schematic diagram of the protein TMEGF45 produced by *P. pastoris* fermentation. The protein has an extra two amino acids (H, M) at the N-terminus that are not part of the native sequence but result from subcloning. The disulfide bonding pattern is shown as lines between the cysteines and was determined chemically (White et al., 1996). The two N-linked glycosylation sites are indicated.

the structures of each of the individual EGF domains that comprise this fragment by synthesizing and refolding each polypeptide in vitro. Neither domain has much regular secondary structure as is typical for EGF domains (Meininger et al., 1995; Sampoli Benitez et al., 1997). Based on this observation, as well as the requirement of the sugars for solubility, we desired to carry out NMR experiments on ^{15}N-labeled protein.

As with many other proteins, TMEGF45 can be produced in *P. pastoris* grown either in shake flask cultures or by fermentation, however, *P. pastoris* produce much more protein at high cell densities. Protein yields improve 10–100 fold in high cell density fermentation as compared to shake flask cultures (Clare et al., 1991a,b; White et al., 1995). Laroche et al. (1994) reported a yield of 1.7 g/L from fermentation, and 70 mg/L for shake flasks, and were able to obtain 27 mg of ^{15}N and ^{13}C-labeled protein with greater than 98% enrichment from shake flask culture. For most proteins including TMEGF45, the yields from shake flask cultures, even after optimization, are only a few milligrams per liter, and isotopic enrichment to produce NMR quantities of protein requires many liters of culture medium.

Optimized fermentation protocols are available for production of large quantities of protein from *P. pastoris*, often higher yields are achieved than can be obtained in *E. coli* (Siegel and Brierley, 1989; Brierley et al., 1990; Chen et al., 1996). These protocols utilize a batch phase followed by a fed-batch phase in which the cell density is increased to upwards of 300 g/L wet cell weight. Higher yields of protein are obtained if induction is begun at these high cell densities (Cregg et al., 1987). These protocols seem ideally suited to isotopic labeling because the basal media components are all salts and defined nitrogen (NH_4OH) and carbon (glycerol followed by methanol) sources.

The sole nitrogen source in optimized fermentation protocols is NH_4OH, which serves two purposes. First, as the *P. pastoris* grow, they produce acid, and addition of NH_4OH functions as a base to adjust the pH of the growth medium. Second, NH_4OH is a nitrogen source that does not result in increased ionic strength of the medium, because the counter ion is $^{-}$OH. It is impractical to utilize $^{15}NH_4OH$, because several hundred milliliters of NH_4OH are required for a one liter fermentation, and the cost of these quantities of $^{15}NH_4OH$ is prohibitively high. For ^{15}N-labeling, we sought to replace the $^{15}NH_4OH$ with the much more economical nitrogen source, $(^{15}NH_4)_2SO_4$. The difficulty we encountered in attempts to substitute $(NH_4)_2SO_4$ for NH_4OH was build-up of an unacceptably high concentration of K_2SO_4 during the fermentation. We present here a solution to this problem that enables economical ^{15}N-labeling in *P. pastoris* fermentations.

Two carbon sources are typically utilized in protein production by fermentation of *P. pastoris*; glycerol and methanol. Although *P. pastoris* has been advertised as able to grow on methanol as the sole carbon source, they grow relatively slowly on methanol (doubling time 5.5–6.0 h). Therefore, to increase productivity of fermentation systems, which requires growing to high cell densities, alternate carbon sources (typically glycerol) are used to build cell density prior to induction of protein production on methanol (Wegner and Harder, 1987). Furthermore, high levels of protein expression in *P. pastoris* are correlated with beginning induction at a high cell density, and induction at lower cell densities results in concomitantly lower protein yields (Cregg et al., 1987). This is disappointing because methanol would be an inexpensive source of ^{13}C-label if it could be used as the sole carbon source in the entire fermentation process. ^{13}C-glycerol is one of the more expensive sources of ^{13}C-label. An alternative

carbon source that might be more economical than glycerol is glucose; however, the strong inducible promoter that is typically utilized in *P. pastoris* expression systems, the *AOX1* promoter, is severely repressed by glucose. In fact, suddenly switching the carbon source from glucose to methanol results in cell death because the cells cannot convert the methanol without the presence of alcohol oxidase which is produced from the *AOX1* gene, and the methanol can become toxic (Tschopp et al., 1987). We also report here a simple way to overcome this problem so that the cell mass can be built up on glucose.

Materials and methods

Reagents

All reagents were the highest grade possible from Fisher Scientific (Pittsburgh, PA) unless otherwise stated. Yeast nitrogen base (YNB), bactopeptone, yeast extract and casamino acids were from DIFCO distributed by VWR Scientific (Los Angeles, CA). Cupric sulfate·$5H_2O$, sodium iodide, manganese sulfate·H_2O, sodium molybdate·$2H_2O$, boric acid, cobalt chloride, zinc chloride, ferrous sulfate·$7H_2O$, and sulfuric acid were from Aldrich Chemicals (Milwaukee, WI). ^{14}C-glucose was obtained from New England Nuclear. The amount of total protein was determined by BCA assay (Pierce Chemicals, Rockford, IL). The BCA assay results for TMEGF45 were standardized by comparison to known quantities of TMEGF45 determined by quantitative amino acid analysis (White et al., 1995). The amount of TMEGF45 was determined by a protein C activation assay described previously (Lougheed et al., 1995).

Pichia pastoris strain

The *P. pastoris* strain that produces TMEGF45 was described previously (White et al., 1995), and contains a multicopy insert of a pPIC9K plasmid with a synthetic gene for TMEGF45 in the *pep4*$^-$ strain, SMD1168 that has a *mut*^{S} phenotype. This strain was observed to have similar growth characteristics to a strain that had a *mut*$^+$ phenotype except that less methanol is required during induction, and the cell mass does not increase during induction.

Growth media

In preparation for fermentation, cultures of *P. pastoris* were grown in BMM which contained 10% yeast nitrogen base (YNB), 1% (w/v) casamino acids, 2 ml/L biotin solution, and 2% (v/v) glycerol). YNB was made up of 100 g/L of either $(NH_4)_2SO_4$ or $(^{15}NH_4)_2SO_4$ [98%+ isotope enrichment (Cambridge Isotope Laboratories, MA)] and 34 g/L yeast nitrogen base [without amino acids and $(NH_4)_2SO_4$ (DIFCO Laboratories, Andover, MI)]. In preparation for fermentations that used glucose in place of glycerol, BMD media contained 2% (w/v) glucose in place of glycerol, but was otherwise identical to BMM.

Fermentation of *P. pastoris* with $(NH_4)_2SO_4$ was carried out using basal salts medium containing 16.6 g/L sodium phosphate monobasic, 2.3 g/L calcium sulfate·$2H_2O$, 11.7 g/L magnesium sulfate·$7H_2O$, 45 ml/L glycerol, 4 ml/L of a biotin solution (0.25 g/L) and 4 ml/L of PTM salts solution (2.0 g/L cupric sulfate·$5H_2O$, 0.08 g/L sodium iodide, 3.0 g/L manganese sulfate·H_2O, 0.2 g/L sodium molybdate·$2H_2O$, 0.02 g/L boric acid, 0.5 g/L cobalt chloride, 7.0 g/L zinc chloride, 22.2 g/L ferrous sulfate·$7H_2O$, and 5 ml/L sulfuric acid). The pH was controlled with a sterilized base solution of 4.4 M KOH and 7.4 M NaOH. During the fed batch phase, a glycerol solution containing 50% (v/v) glycerol, 12 ml/L biotin solution and 12 ml/L PTM salts solution was delivered to the vessel at a constant feed rate. Induction of protein production was achieved by feeding a methanol solution (50% (v/v) containing 6 ml/L PTM salts and 6 ml/L biotin solution). Fermentations in which glucose was substituted in place of glycerol used a 37% (w/v) solution of glucose in H_2O that was sterilized by filtration through a 0.22 μm filter.

Shake flask culture protocol

Shake flask cultures (1 L starting volume) were inoculated from a 10 ml culture of *P. pastoris* that was grown from a frozen stock overnight in BMM with shaking at 30 °C and grown in either BMM or BMD for two days at 30 °C with shaking (300 rpm). Cells were harvested and resuspended in 500 mL of the same growth medium but containing 2% methanol instead of glycerol or glucose. For the growths in glucose, cells were first resuspended in medium containing 0.1% glycerol for 2 h prior to addition of 2% methanol. Induction periods were typically 48 h.

Fermentation protocol for ^{15}N-labeling with glycerol as the carbon source

The production of large amounts of ^{15}N-labeled TMEGF45 by fermentation was carried out in a BioFlow 3000 fermentor (New Brunswick Scientific, Edison, NJ) equipped with a 1.25 L bioreactor. A

10 ml culture of *P. pastoris* was grown from a frozen stock overnight in BMM with shaking at 30 °C. The culture was used to inoculate a 100 mL culture of BMM in a 500 mL Erlenmeyer flask and was grown overnight shaking at 30 °C. The fermentation media, 0.7 L of the basal salts, was sterilized in the fermentor vessel by autoclaving, and after slow cooling was allowed to oxygenate and the pH was slowly adjusted by addition of the base solution. Fermentation began by addition of 50 ml of *P. pastoris* culture, as well as 2.6 mL each of PTM salts solution and biotin solution. A solution of 11.5 g of $(^{15}NH_4)_2SO_4$ dissolved in 50 mL of H_2O which had been sterilized by filtration through a 0.22 μm filter was also added. The pH was maintained at 5.3 by addition of the KOH and NaOH base mixture with a maximal flow rate of 0.15 mL/min. During the initial growth period, the *P. pastoris* cells use the glycerol that is present in the basal salts medium, and this batch growth phase typically lasted less than 24 h. Towards the end of the batch growth phase, the cells started to utilize O_2 at a high rate requiring pure O_2 supplementation to maintain the dissolved O_2 levels at 30%. When the glycerol in solution was exhausted, the cells immediately ceased using O_2, which was observed as a rapid increase in dissolved O_2. At this point, the wet cell weight (wcw) was approximately 140 g/L and the fed batch phase was begun by addition of another 11.5 g of $(^{15}NH_4)_2SO_4$ dissolved in 50 mL of H_2O, and by initiation of the glycerol feed. A total of 160 mL of 50% glycerol (100 g) was fed, at a rate of 0.11 mL/min over 20 h. The cells reached a final wcw of 290 g/L. At this point, to reduce the high salt concentration, the media was exchanged by harvesting the cells with centrifugation at 1050 × g for 20 min in 1 L bottles in a Sorvall RC3B centrifuge. The pelleted cells were resuspended in 0.7 L of freshly autoclaved basal salts media (pH 5.3) devoid of glycerol. Fermentation was resumed with the addition of 23 g of $(^{15}NH_4)_2SO_4$ dissolved in 100 mL of H_2O, 2.6 ml PTM salts and 2.6 ml biotin. Methanol induction was initiated with a solution of 50% methanol. The flow rate of methanol was increased from 0.015 mL/min to 0.15 mL/min over a period of 20 h. At this point the methanol feed rate was held constant a 0.15 mL/min until the end of the induction phase, which terminated after 60 h. The cell were separated from the culture supernatant by centrifugation at 4200 × g for 1 h in a Sorvall RC3B centrifuge. A solution of 0.5 M EDTA was added to the culture supernatant to a final concentration of 10 mM. The culture supernatant was stored at −70 °C until protein purification.

Glucose fermentations

Fermentations were carried out to ascertain the feasibility of ^{13}C-labeling via feeding of glucose instead of glycerol to determine if glucose would be a more economical carbon source. For the glucose fermentation, a 37% (w/v) solution of glucose containing 12 mL/L PTM salts and 12 mL/L biotin solution was used in place of the glycerol solution during the batch and fed batch phases of the fermentation. For these experiments, 0.7 L of basal salts media were autoclaved in the bioreactor and 100 ml of the 37% (w/v) glucose solution (37 g) were pumped into the vessel before inoculation with the 50 mL culture. The wcw was 100 g/L at the end of the batch phase when the rapid increase in dissolved O_2 indicated exhaustion of the glucose in the growth medium. A solution of 37% glucose in H_2O was fed at rate of 0.15 mL/min until the wcw was 290 g/L. During the fed batch phases, an additional 340 ml of 37% glucose (126 g of glucose) was added. The cells were harvested, resuspended in 0.7 L of basal salts, and returned to the bioreactor. Before beginning the induction, 23 g of $(NH_4)_2SO_4$ dissolved in 100 ml of H_2O was added. In order to derepress the *AOX1* promoter, a solution of 50% glycerol in H_2O was fed at a rate of 0.15 mL/min for 20 min (1.9 g glycerol was added during this derepression period). The cells were then induced in the same manner as described for the glycerol fermentations.

Purification of TMEGF45

The TM fragments were secreted into the culture medium, which was separated from the cells and the protein was purified initially on QAE Sephadex followed by HiLoad 26/10 Q Sepharose chromatography as described previously (White et al., 1995). The partially purified TMEGF45 was acidified to pH 3.5 with HCl and immediately injected on a Waters C18 reverse-phase column (19 × 300 mm) equilibrated with 0.1% trifluoroacetic acid (TFA). TMEGF45 was eluted by a gradient of 100% TFA for 10 min, 0–20% acetonitrile over 10 min, 20–40% acetonitrile over 40 min at a flow rate of 10 mL/min. Yields were approximately 50% in all cases.

TMEGF45 characterization

The purified TMEGF45 was characterized by N-terminal sequencing which confirmed the cleavage of

the α-factor leader sequence after the Lys-Arg cleavage site and the addition of His-Met to the N-terminus resulting from the *Nde I* restriction site (White et al., 1995). For mass spectrometry analysis, both ^{15}N-labeled and unlabeled TMEGF45 were treated with endoglycosidase H (Sigma Chemicals, St. Louis, MO). The reaction was carried out using 0.2 mg of protein in 400 μL of 100 mM sodium acetate buffer, pH 5.5, reacted for 12 h at 37 °C with 10 mU of endoglycosidase H. The products were purified by chromatography on a Vydac C18 reverse-phase HPLC column (4.6 × 250 mm) with a gradient of 0–50% acetonitrile over 50 min. The mass spectra were collected on a Kratos Kompact MALDI 4 (resolution 100 ppm). NMR samples were prepared by resuspending lyophilized TMEGF45 in 0.5 mL of 50 mM sodium phosphate buffer containing 10% D_2O to a final concentration of 1 mM and the pH was adjusted to 6.5 with NaOH. The ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled TMEGF45 was acquired at 310 K on a Bruker DRX600 spectrometer. Data were acquired with a spectral width of 1946 Hz and 256 points in the ^{15}N (F1) dimension and a spectral width of 3005 Hz and 512 points in the ^{1}H (F2) dimension. The data were apodized with a sine-bell squared function in both dimensions and processed using the Felix 97 software (Molecular Simulations, Inc., San Diego, CA).

^{14}C-Glucose fermentation

Fermentation using ^{14}C-glucose as the sole carbon source during batch and fed batch phases of the fermentation, was carried out to ascertain what percent of the overall carbon incorporated into TMEGF45 resulted from glucose versus methanol. The fermentation was carried out in an identical manner as described for the ^{12}C glucose fermentations, the only difference being that the glucose solution used was 440 ml of 37% glucose with 1 mCi of ^{14}C-glucose added. The methanol added during the induction phase was not radioactive.

The supernatant from this fermentation was purified and a sample was characterized by mass spectrometry to determine the average molecular weight of glycosylated TMEGF45, which was consistent with results from other fermentations. Dilutions of the initial glucose solution and of purified TMEGF45 were counted with a Packard 1600TR Liquid Scintillation Analyzer. Carbon content of both glucose samples and TMEGF45 samples were calculated so that they could be directly compared with the counts per minute (cpm) of each sample. Glucose and TMEGF45 sam-

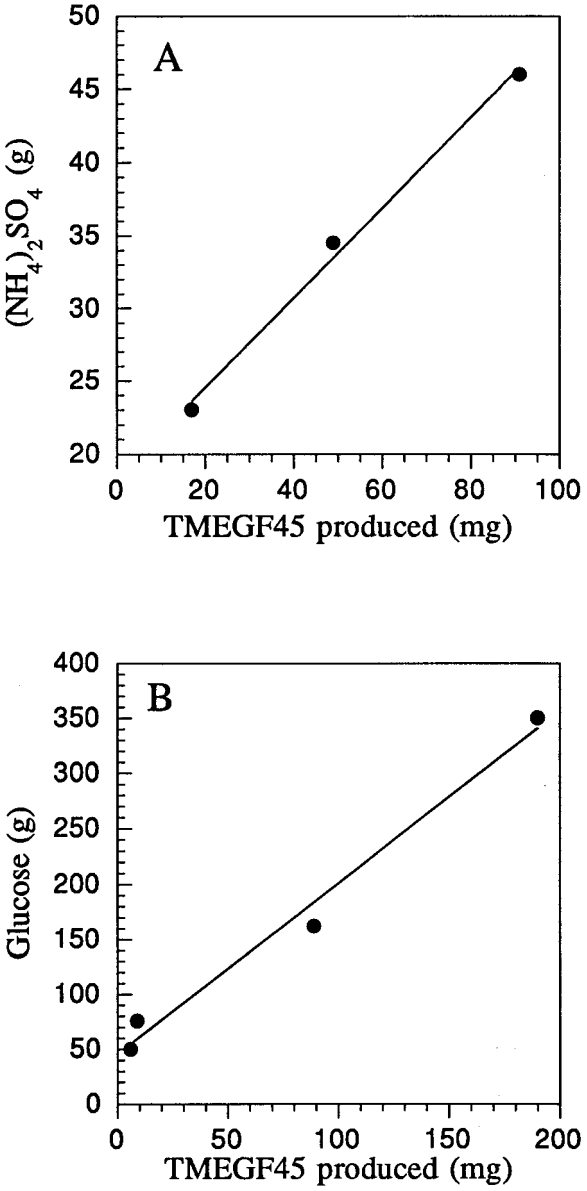

Figure 2. (A) Relationship between the yield of TMEGF45 and the amount of nitrogen fed as $(NH_4)_2SO_4$. The fermentation was as described in Materials and methods wherein 1/2 of the $(NH_4)_2SO_4$ was added during the growth phase and 1/2 during the induction phase and a total of 46 g was added. For the fermentations wherein less than 46 g of $(NH_4)_2SO_4$ was added, each addition of $(NH_4)_2SO_4$ was decreased proportionately. (B) Relationship between the yield of TMEGF45 and the amount of carbon fed as glucose. Induction was begun once all the glucose was fed even though the wet cell weight was lower for the lower amounts of glucose at the time of induction.

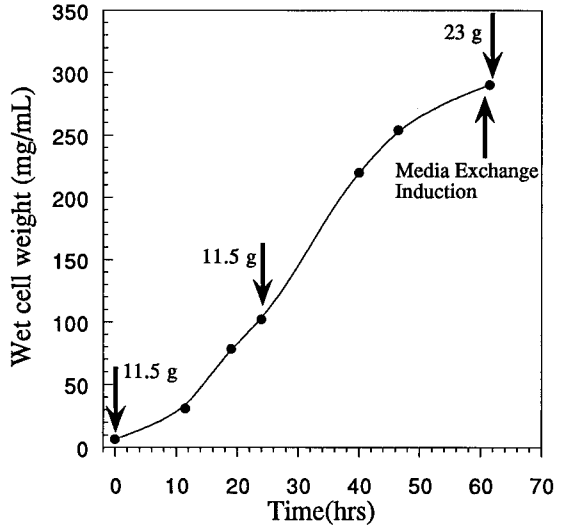

Figure 3. Profile of a typical *P. pastoris* fermentation using $(^{15}NH_4)_2SO_4$ as the sole nitrogen source. Time 0 corresponds to inoculation of the fermentor vessel. Arrows indicate additions of $(^{15}NH_4)_2SO_4$ at initial inoculation, at the beginning of the fed batch phase. Cells were harvested by centrifugation after the wet cell weight reached 300 g/L and resuspended in fresh basal salts devoid of glycerol and 23 g of $(^{15}NH_4)_2SO_4$ was added. Induction with 50% methanol was initiated immediately after the media exchange.

ples containing the same amount of carbon would be expected to have the same cpm if 100% of the carbon incorporated into TMEGF45 came via the glucose.

Results

Production of TMEGF45 by fermentation using $(^{15}NH_4)_2SO_4$

Our first goal was to work out fermentation conditions utilizing $(NH_4)_2SO_4$ because ^{15}N-labeling by way of $(^{15}NH_4)_2SO_4$ is much more economical than labeling by way of $^{15}NH_4OH$. The difficulty encountered in utilizing $(NH_4)_2SO_4$ as the sole nitrogen source is the build-up of K_2SO_4 salt in the growth medium, and initial attempts at fermentation failed as a result of the high ionic strength of the growth medium. The ionic strength of the media after enough $(NH_4)_2SO_4$ had been added was equivalent to a 1.2 M solution of NaCl, and we observed that the *P. pastoris* cells stopped growing under these conditions. We first tried to utilize less $(NH_4)_2SO_4$ hoping to achieve a fermentation where the ionic strength at the end was not prohibitively high. Figure 2A shows that indeed, the ammonium source is limiting and increases in the ammonium fed resulted in increases in protein yield up to a molar equivalent to the recommended amount of NH_4OH. Fortunately, the high ionic strength resulting from addition of the required amount of $(NH_4)_2SO_4$ could be alleviated by incorporation of a media exchange step at the midpoint of the fermentation (Figure 3). Several different schemes including continuous feeding and batch-wise addition were explored for the addition of the $(NH_4)_2SO_4$. Due to the increased ionic strength from addition of $(NH_4)_2SO_4$, all protocols involved addition of 1/2 of the $(NH_4)_2SO_4$ in the fed batch phase and 1/2 in the induction phase. Protocols involving continuous feeding gave similar results to those using batch-wise addition but resulted in unnecessary waste of $(NH_4)_2SO_4$. Optimal protein production was obtained by addition of two portions of $(NH_4)_2SO_4$ during the growth phase in which the cell mass was built up, the media was then exchanged, and the second half of the $(NH_4)_2SO_4$ was added at the beginning of the induction phase (Figure 3). This procedure resulted in equivalent yields of TMEGF45 using $(^{15}NH_4)_2SO_4$ as the sole nitrogen source instead of NH_4OH (Table 1).

Protein purification and characterization

^{15}N-labeled TMEGF45 was characterized first by N-terminal sequencing, and then by mass spectrometry and two-dimensional heteronuclear NMR spectroscopy. The protein had the correct N-terminal sequence of that shown in Figure 1 out to 10 amino acids. TMEGF45 has two N-linked glycosylation sites, and expression in *P. pastoris* results in attachment of two N-acetylglucosamines (GlcNAc) and 20–42 mannose sugars at both of these sites. The major isomers were those with 25–30 mannose sugars or 13–15 per N-linked glycan. Endoglycosidase H removes the oligosaccharides from both sites resulting in a protein of molecular weight 9690 g/mol. This corresponds to the unglycosylated molecular weight (9282 g/mol) plus two GlcNAc sugars (408 g/mol). The fully ^{15}N-labeled protein that was treated with endoglycolsidase H had a molecular weight of 9796 g/mol which is within the instrument error of the unglycosylated molecular weight of 9388 g/mol plus two GlcNAc sugars (410 g/mol) (Figure 4). By mass spectrometry, then, the protein was at least 98% ^{15}N-enriched.

Despite the presence of two large oligosaccharides, ^{15}N-labeled TMEGF45 yields well-resolved HSQC spectra under optimized conditions of pH and temperature (Figure 5). The protein appears to aggregate at low pH, and the quality of the HSQC spectrum improved markedly as the pH was increased from 4.5 to

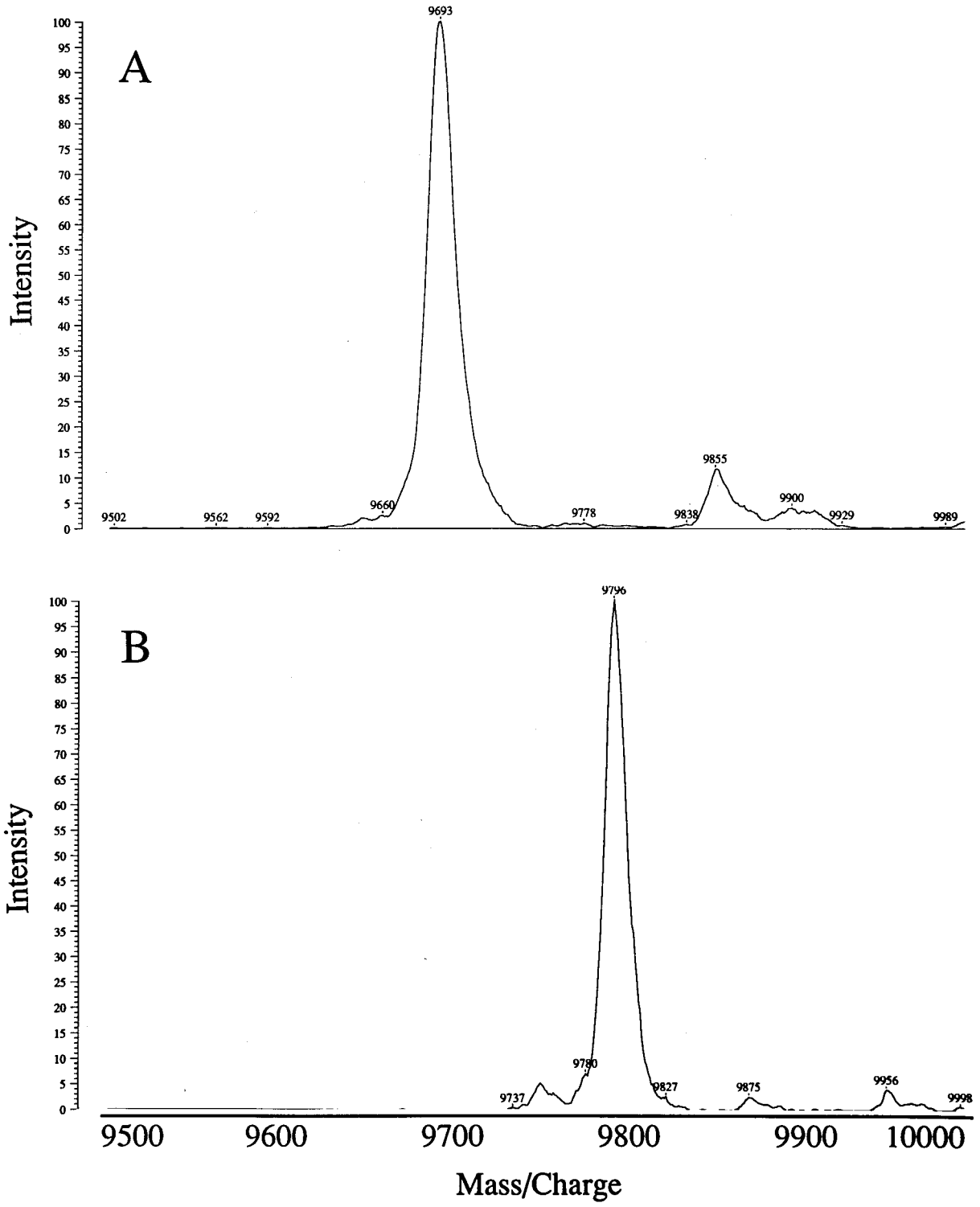

Figure 4. MALDI mass spectra of unlabeled (A) and ^{15}N-labeled (B) TMEGF45 treated with Endo H. The mass of the unlabeled sample (A) was 9693 Da, which compares very favorably to 9690 Da, the mass calculated from the sequence accounting for the addition of two N-acetylglucosamine monosaccharides. The mass of the ^{15}N-labeled sample (B) was 9796 Da, which is within instrument error of 9798 Da, the mass calculated from the ^{15}N-labeled protein including two ^{15}N-acetylglucosamine monosaccharides.

Table 1. Comparison of TMEGF45 yields obtained from various *P. pastoris* culture conditions

	TMEGF45 (mg)	$(NH_4)_2SO_4$[a] (g)	$^{14}NH_4OH$ (g)	Glycerol (g)	Glucose (g)	Methanol (g)
Shake flask[b]	5	15	–	20	–	8
Fermentation 1[c]	91	–	225	143	–	200
Fermentation 2	90	46	–	143	–	200
Fermentation 3	89	46	–	–	162	200

[a]The $(NH_4)_2SO_4$ used in these fermentations was $(^{15}NH_4)_2SO_4$ for fermentation 2 and $(^{14}NH_4)_2SO_4$ for fermentation 3. We observed no difference in yields for fermentations using $(^{14}NH_4)_2SO_4$ versus $(^{15}NH_4)_2SO_4$.
[b]The cells were initially grown in 1 L of culture medium. The culture was then centrifuged and the cells were resuspended in 500 mL of induction medium.
[c]Fermentations were carried out in a 1.25 L vessel with a beginning volume of 800 mL and an ending volume of 600 mL.

6.5. The quality of the HSQC spectrum also increased as the temperature was increased from 25 °C to 37 °C, presumably due to faster tumbling of the highly glycosylated protein, for which the total molecular weight is approximately 15 kDa.

Alternate carbon sources for ^{13}C-labeling

Several issues were important to resolve before optimal carbon labeling could be achieved. First, because *P. pastoris* grow poorly on methanol, either glycerol or glucose must be used as the carbon source before induction. Because market prices vary greatly for ^{13}C-labeled glycerol and glucose, we wanted to ascertain if it was possible to obtain high yields of protein from *P. pastoris* grown on glucose instead of glycerol. Our first attempts to grow the *P. pastoris* on glucose failed because the cells did not continue to grow after induction as indicated by lack of O_2 usage. This was in spite of the fact that the glucose was exhausted from the culture medium and the media was exchanged with fresh at the beginning of the induction phase. We assumed that this was because the *AOX1* promoter was completely repressed by growth on glucose, as had been reported previously (Tschopp et al., 1987). To overcome this problem, after the media exchange we fed a 50% glycerol solution at a feed rate of 0.15 ml/min for 20 min (total glycerol fed = 1.9 g, circa 1% of the total carbon fed). This short period was sufficient to completely derepress both the promoter required for production of alcohol oxidase for cell viability as well as the *AOX1* promoter required for TMEGF45 mRNA production. Table 1 shows comparisons for identical fermentations carried out using glycerol or glucose as the sole carbon source before induction. To obtain circa 90 mg of TMEGF45, either 143 g of glycerol or 162 g of glucose were required. Based on current prices, the cost is approximately the same, and either carbon source can be utilized. Since both of these carbon sources are expensive, we explored the possibility of limiting the growth on these carbon sources and inducing at a lower wcw. Figure 6 shows that protein yield increases logarithmically with wcw, and therefore the most important factor in obtaining high yields of protein is to attain a high wcw. Since wcw is measured in g/L, this means that fermentation of larger, more dilute volumes is not cost effective. As with the nitrogen source, it appears that carbon is also limiting protein production and increasing carbon results in increasing protein yield (Figure 2B).

It was also of interest to explore how much carbon in the final protein comes from the cell mass that is accumulated during the fed batch phase and how much is from the methanol. Anecdotal reports suggested that for *P. pastoris*, unlike *E. coli*, much of the carbon in the final protein comes from cell mass accumulated during the growth phase, however no quantitative studies had been carried out. Using ^{14}C-labeled glucose as the sole carbon source, we have ascertained that 70% of the carbon in the TMEGF45 produced by fermentation comes from cell mass accumulated prior to induction. Only 30% of the carbon comes from the methanol based on specific activity of final purified protein compared to starting ^{14}C-glucose.

Discussion

The advent of isotopic enrichment of proteins, and the pulse sequences that utilize these heteronuclei, have revolutionized the speed and accuracy of NMR solution structure analysis, and increased the size of

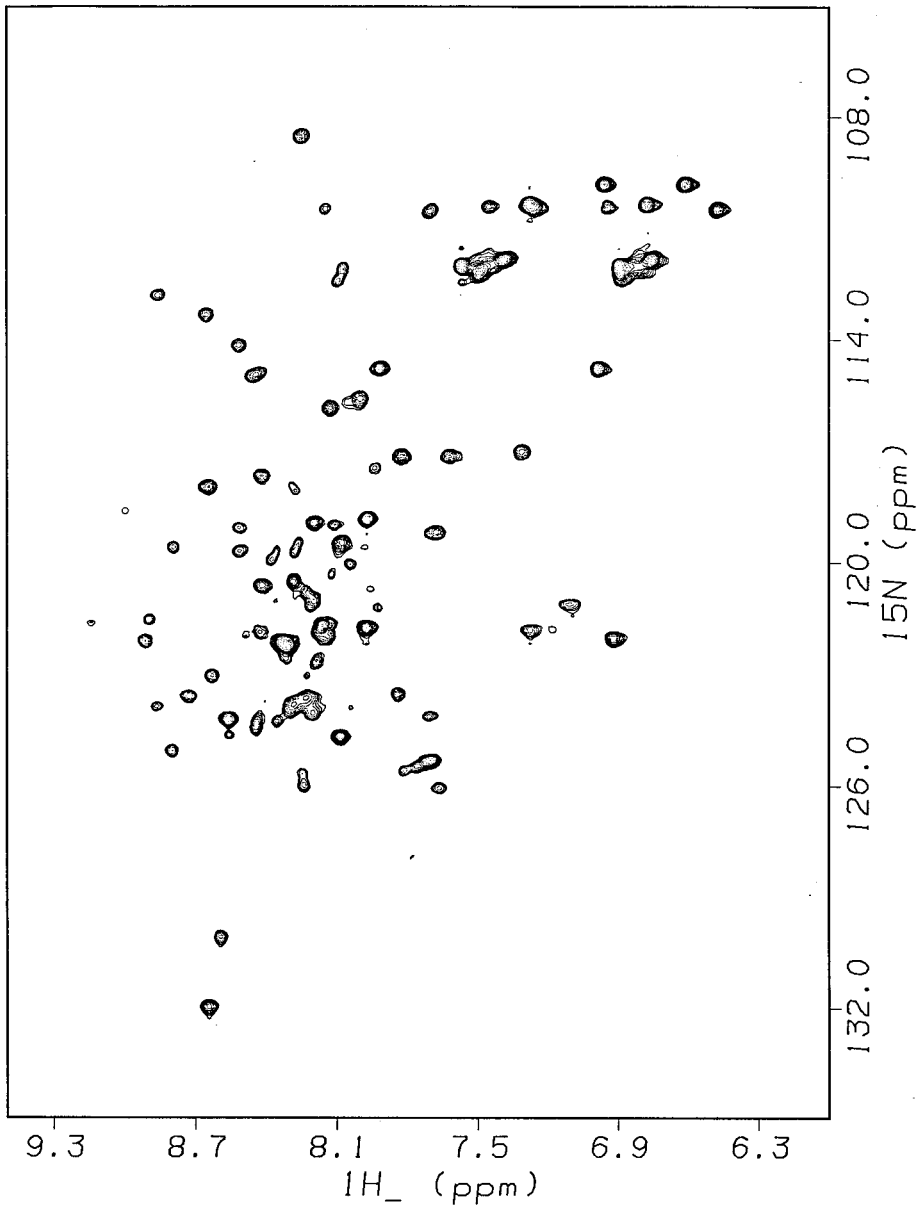

Figure 5. The ^{1}H-^{15}N HSQC of glycosylated ^{15}N-labeled TMEGF45 was collected at 310 K on a Bruker DRX600. The TMEGF45 concentration was 1 mM in 50 mM sodium phosphate, 10% D_2O at pH 6.5. Data were acquired with a spectral width of 1946 Hz and 256 points in the ^{15}N (F1) dimension and a spectral width of 3005 Hz and 512 points in the ^{1}H (F2) dimension and apodized with a sine-bell squared function in both dimensions.

macromolecular structures that can be studied (Fesik et al., 1990; Clore and Gronenborn, 1997). Protein structures and complexes exceeding 30 kDa can now be solved by NMR (Clore and Gronenborn, 1997). Today, a critical requirement for high resolution NMR protein structure determination is that the protein be overexpressed with uniform ^{15}N, ^{13}C and in some cases ^{2}H isotopic enrichment. While techniques to carry out such isotopic enrichment in *E. coli* have been available for some years (Venters et al., 1991), labeling of proteins that cannot be expressed in *E. coli* due to requirements of chaperones for folding, or post-translational modifications, is much more difficult. We now report a high-yielding protocol for isotopic

enrichment in *P. pastoris* by fermentation which we have used to produce tens of milligrams of purified TMEGF45, a small anticoagulant glycoprotein that contains six disulfide bonds and two N-linked glycans. This protein could not be expressed in *E. coli*, but could be obtained in high yields from fermentation in *P. pastoris* (White et al., 1995).

Chinese hamster ovary cells (CHO cells) have been used to produce uniformly labeled glycosylated proteins. Human chorionic gonadotropin was uniformly labeled in CHO cells using media composed of algal extracts containing ^{13}C and ^{15}N-labeled amino acids (Lustbader et al., 1996). CHO cells grown on a combination of isotopically enriched amino acids enabled the structural characterization of CD2, a glycosylated protein of much interest in signal transduction (Wyss et al., 1997). Characterization of the interaction between the N-linked glycans of CD2 and the protein gave important information regarding the differences between complex mammalian glycans and the simpler high mannose type glycans attached to the CHO cell-expressed CD2. Interestingly, Wyss and Wagner showed that only the two GlcNac sugars of the core $Man_3GlcNAc_2$ pentasaccharide interacted with the CD2 protein (Wyss et al., 1995). This core is conserved across all N-linked glycan types (Wyss and Wagner, 1996). *P. pastoris* is only capable of attaching high mannose type glycans, and the fact that the core is all that interacts with the protein suggests that biologically relevant information may be obtained from studies of glycosylated proteins produced in *P. pastoris* even if the glycans are not identical to those found in the native mammalian protein. We chose to produce TMEGF45 in *P. pastoris* because higher protein yields at a lower cost than from CHO cells was expected. We and others have observed much higher yields of protein produced in *P. pastoris* grown in fermentation culture than in shake flasks (Cregg et al., 1987; Clare et al., 1991a, b; White et al., 1995).

Cost comparisons show that if a protein can be expressed in *E. coli*, this is the most economical expression host for isotopic enrichment of proteins, with reported yields of 75 mg/L of Human Carbonic Anhydrase II from growth medium containing 1 g of $^{15}NH_4Cl$ and 2 g of ^{13}C-glucose (Venters et al., 1991). If expression is not possible in *E. coli*, shake flask cultures of *P. pastoris* may be the next best option. Two reports have utilized *P. pastoris* yeast expression in shake flask cultures with varying yields (Laroche et al., 1994; Wiles et al., 1997). Laroche et al. report the highest yields ever (27 mg) for *P. pastoris* shake flask

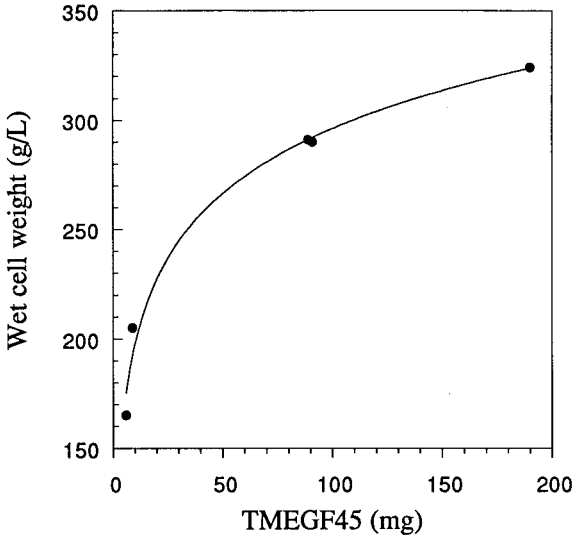

Figure 6. Relationship between the amount of TMEGF45 produced and wet cell weight. Much more TMEGF45 was produced at wet cell weights near 300 g/L, and the best fit of the data was obtained with a logarithmic curve.

production of ^{15}N- and ^{13}C-labeled tick anticoagulant protein using 5 g/L of $(^{15}NH_4)_2SO_4$, 10 g/L of ^{13}C-glucose and 80 g/L of ^{13}C-methanol (Laroche et al., 1994) (note that 2 L of growth medium was required due to media exchange before induction). A more typical yield of 2–3 mg/L of ^{15}N-labeled Vaccinia virus complement control protein from a minimal 2.5 g of $(^{15}NH_4)_2SO_4$ has also been reported (Wiles et al., 1997). If yields are too low in shake flasks, a better option may be fermentation, particularly if protein purification from low level amounts is difficult. The advantage of fermentation is that high concentrations of expressed, secreted proteins are obtained in the culture supernatant facilitating protein purification. Comparison of the results obtained from fermentation show that (Table 1) for ^{15}N-labeling, fermentation resulted in 2 mg protein produced per gram of $(^{15}NH_4)_2SO_4$ and this compares favorably with the 0.3 mg – 1 mg protein produced per gram of $(^{15}NH_4)_2SO_4$ in shake flasks. The data in Table 1 also show a favorable comparison for ^{13}C-labeling by fermentation versus shake flasks.

Our experiments confirm anecdotal information that much of the carbon in the protein produced in *P. pastoris* results from cell mass that was accumulated during the growth phase, and not much is from the methanolic carbon. Still, 30% of the carbon does come from the methanol, so in order to obtain highly enriched samples for NMR analysis, it is necessary

to incorporate both ^{13}C glycerol or glucose and ^{13}C-methanol. The strain used in these studies was of the *mut*^{S} phenotype. This means that the expression cassette was inserted into the genome of the *P. pastoris* at the *AOX1* locus, and the *P. pastoris* grow more slowly on methanol. This has the advantage that during induction, less methanol is required, and the cell density does not continue to increase making the protein production per cell more efficient. Growth of a *mut*$^{+}$ strain would require at least twice the methanol, and from our observations, would not yield significantly more protein. Thus, we would recommend fermentation of a *mut*^{S} phenotype strain, if protein yields are comparable.

Although the cost of ^{13}C-labeling by fermentation is cheaper than any other option at present, it is still moderately high. The fermentation protocols presented here are economical for labeling only with $(^{15}NH_4)_2SO_4$, costing less than US $2000 for 90 mg of protein. The carbon-labeling exploration experiments were therefore carried out in such a way that the nitrogen was not the limiting reagent. Exploration into inducing at lower wcw by limiting the carbon source in the growth phase showed that this is not a useful approach because substantially less protein is obtained at only moderately lower wcw (Figure 6). In the 1 L fermentation vessel, 90 mg of TMEGF45 was produced from 143 g glycerol or 162 g glucose. This amount of protein is sufficient for at least 10 NMR samples, and therefore on a per sample basis, the cost is not prohibitive. A sample of 9 mg of ^{13}C-labeled TMEGF45 (using both ^{13}C-glycerol and ^{13}C-methanol) would cost US $4000 at current market prices. The limitation lies in the minimal volume that can be achieved in the fermentation vessel, which currently is 700 ml, and optimal production in this volume results in 10 NMR samples-worth of protein! Hopefully, smaller volume fermentation vessels will be available in the future.

Acknowledgements

We thank David Meininger for early experiments and helpful advice, and Martin Gleeson for helpful discussions. We thank Siv Garrod in the Taylor lab for N-terminal sequencing. We also thank Ken Harris in the Suizdak lab at The Scripps Research Institute for the mass spectra. This work was supported by NIH grant HL47463 and the American Heart Association. M.J. Wood acknowledges support from the NIH training grant in Heme and Blood Proteins.

References

Brierley, R.A., Bussineau, C., Kosson, R., Melton, A. and Siegel, R.S. (1990) *Ann. New York Acad. Sci.*, **589**, 350–362.

Chen, Y.L., Cino, J., Hart, G., Freedman, D., White, C.E. and Komives, E.A. (1996) *Process Biochem.*, **32**, 107–111.

Clare, J.J., Rayment, F.B., Ballantine, S.P., Sreekrishna, K. and Romanos, M. (1991a) *Bio/Technology*, **9**, 455–460.

Clare, J.J., Romanos, M., Rayment, F.B., Rowedder, J.E., Smith, M.A., Payne, M.M., Sreekrishna, K. and Henwood, C.A. (1991b) *Gene*, **105**, 205–211.

Clore, G.M. and Gronenborn, A.M. (1997) *Nat. Struct. Biol.*, **4**, 849–853.

Cregg, J.M., Tschopp, J.F., Stillman, C., Siegel, R., Akong, M., Craig, W.S., Buckholz, R. G., Madden, K.R., Kellaris, P.A., Davis, G.R., Smiley, B.L., Cruze, J., Torregrossa, R., Velicelebi, G. and Thill, G.P. (1987) *Bio/Technology*, **5**, 479–485.

Cregg, J.M., Vedvick, T.S. and Raschke, W.C. (1993) *Bio/Technology*, **11**, 905–910.

Fesik, S.W. and Zuiderweg, E.R.P. (1990) *Q. Rev. Biophys.*, **23**, 97–131.

Laroche, Y., Storme, V., De Meutter, J., Messens, J. and Lauwereys, M. (1994) *Bio/Technology*, **12**, 1119–1124.

Lougheed, J.L., Bowman, C.A., Meininger, D.P. and Komives, E.A. (1995) *Protein Sci.*, **4**, 773–780.

Lustbader, J.W., Birken, S., Pollak, S., Pound, A., Chait, B.T., Mirza, U.A., Ramnarain, S., Canfield, R.E. and Brown, J.M. (1996) *J. Biomol. NMR*, **7**, 295–304.

Meininger, D.P., Hunter, M.J. and Komives, E.A. (1995) *Protein Sci.*, **4**, 1683–1695.

Sampoli Benitez, B., Hunter, M.J., Meininger, D.P. and Komives, E.A. (1997) *J. Mol. Biol.*, **273**, 913–926.

Siegel, R.S. and Brierley, R.A. (1989) *Biotechnol. Bioeng.*, **34**, 403–404.

Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A. and Gingras, T.R. (1987) *Nucleic Acids Res.*, **9**, 3859–3876.

Venters, R.A., Calderone, T.L., Spicer, L.D. and Fierke, C.A. (1991) *Biochemistry*, **30**, 4491–4494.

Wegner, G.H. and Harder, W. (1987) *Antonie Van Leeuwenhoek*, **53**, 29–36.

White, C.E., Hunter, M.J., Meininger, D.P., White, L.R. and Komives, E.A. (1995) *Protein Eng.*, **8**, 1177–1187.

White, C.E., Hunter, M.J., Meininger, D.P., Garrod, S. and Komives, E.A. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 10177–10182.

Wiles, A.P., Shaw, G., Bright, J., Perczel, A., Campbell, I.D. and Barlow, P.N. (1997) *J. Mol. Biol.*, **272**, 253–265.

Wyss, D.F., Dayie, K.T. and Wagner, G. (1997) *Protein Sci.*, **6**, 534–542.

Wyss, D.F., Choi, J.S., Li, J., Knoppers, M.H., Willis, K.J., Arulanandam, R.N., Smolyyar, A., Reinherz, E.L. and Wagner, G. (1995) *Science*, **269**, 1273–1278.

Wyss, D.F. and Wanger, G. (1996) *Curr. Opin. Biotechnol.*, **7**, 409–416.