JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Preparation of Unglycosylated Human Caseinomacropeptide by Engineering DAB *Escherichia coli*

FANGJIE LIU, LIANG LIAO, AND JINCHUN CHEN*

College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, People's Republic of China

Human caseinomacropeptide (hCMP) is 65 amino acids in length and was originally derived from the C terminus of human milk κ-casein. As it is highly abundant in both essential amino acids and branched amino acids, it could be developed as a practical food and even as medicinal nutrition for patients. This study was undertaken to prepare recombinant hCMP without glycosylation using recombinant plasmid and prokaryotic expression system. The gene encoding hCMP was chemically synthesized and directly inserted into the pET28a(+) vector and then expressed in *Escherichia coli* BL21(DE3). The maximum amount of soluble protein was obtained by incubation with 0.5 mM isopropyl-α-D-thiogalactopyranoside at 30 °C for 4 h and accounted for 40% of the total intracellular protein. Most of the expressed fusion proteins, located in the cell periplasm and cytoplasm, could be adsorbed by nickel affinity chromatography and eluted with buffer containing 300 mM imidazole. The fusion proteins were cleaved by enterokinase to remove the 6-His tag. Gel filtration chromatography with Sephadex G-10 was performed for desalting and purification. A final yield of 25 mg of the mature protein with high purity up to 99% was obtained from 1 L of *E. coli* culture. The purified protein was confirmed by MALDI-TOF-MS analysis. This study overcame the problem of glycosylation in hCMP and established a novel approach for the preparation of unglycosylated hCMP.

KEYWORDS: Unglycosylated hCMP; expression; purification; identify

INTRODUCTION

Human caseinomacropeptide (hCMP), originally derived from the C terminus of human milk κ-casein, consists of 65 amino acid residues and is generally glycosylated and phosphorylated (*1*). Its primary structure gives it high nutritional value for human health. hCMP contains five essential amino acids, which account for over half (34/65) of the whole macropeptide, and has nearly 25% (16/65) branched amino acid residues, although it is low in leucine. Additionally, phenylalanine is the only aromatic amino acid among the 65 residues. Obviously, hCMP as a natural polypeptide possesses a good *F* value, which is necessary for a liver patient's nutrition.

According to the literature, hCMP is known to have much more powerful biological functions than other CMPs. For example, hCMP can promote the growth of beneficial bacteria including *Bifidobacteria* (*2–4*), and its growth-stimulating activity is 4 times higher than that of bovine CMP (*5*). Moreover, hCMP can inhibit fibrinogen binding to its platelet receptors, which would otherwise stimulate platelet aggregation and give rise to fibrin, which is responsible for thrombus formation (*6*). Sollier et al. have reported that hCMP has a higher potential antithrombotic activity than bovine caseinomacropeptide (bCMP) (*7*). In addition, hCMP plays a role in the inhibition of the adhesion of oral *Actinomyces* to human cell membranes and the binding of cholera toxin to its receptors (*8*). Therefore, hCMP is thought to be a potential ingredient for medicinally functional foods and pharmaceuticals.

However, hCMP is not commercially available yet due to the lack of a natural source. An alternative approach to the preparation of hCMP is via DNA recombinant technology. Park et al. (*9*) reported that recombinant hCMP had been produced by the yeast *Saccharomyces cerevisiae* as a secretory product and modified by glycosylation after translation. Additionally, hCMP showed lower glycosylation compared to the natural bCMP. There is no doubt that hCMP had been successesfully expressed with the engineered yeast *S. cerevisiae* and that it could meet the requirements for patients as a functional food. If it could be digested with some proteinase to become small peptides, it would be easily taken up by intestinal cells. However, it cannot be further developed as an injection for severely ill patients because of the differences in sugar residues of rhCMP from yeast *S. cerevisiae* cells and human mammary grand cells. It is well-known that several sugar residues could be antigenic determinants: different sugar side chains on hCMP could induce an immune response if it was utilized as an injection. Therefore, prokaryotic *Escherichia coli* was employed as the host cell to produce unglycosylated hCMP, which has

* Corresponding author (telephone 86-10-64439673; fax 86-10-84858845; e-mail jingchunchen@hotmail.com).

10.1021/jf072585n CCC: $40.75
Published on Web 01/12/2008

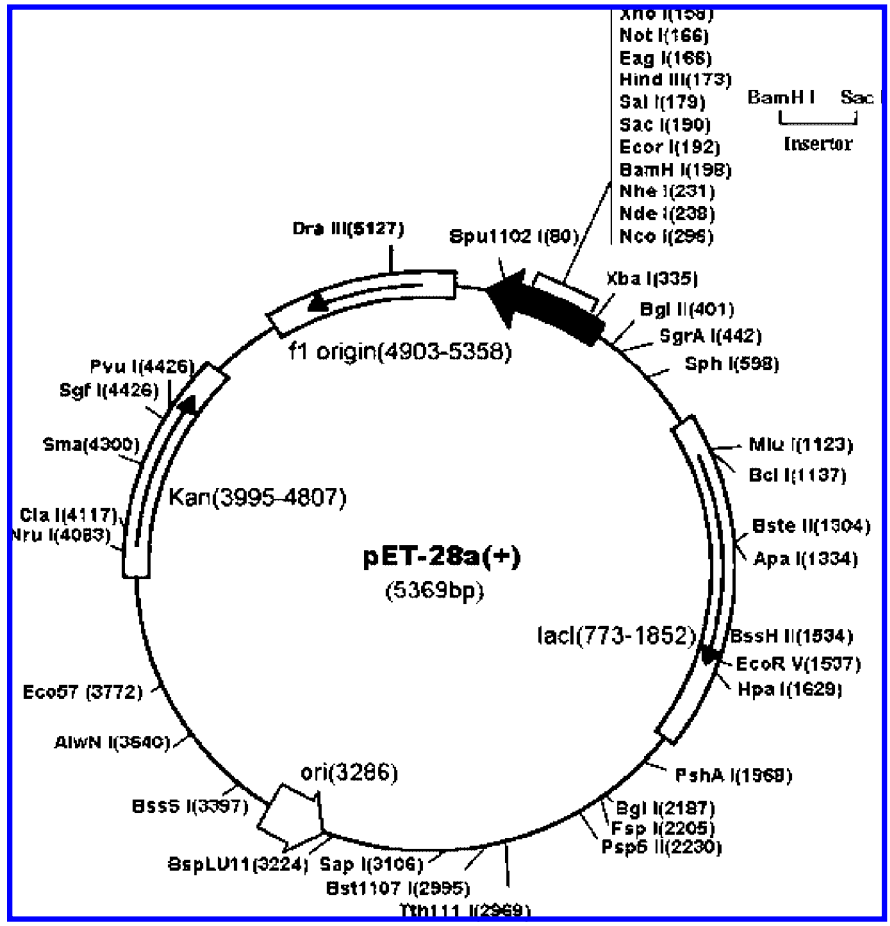

Figure 1. Schematic diagram of the expression vector pET28a(+)-hCMP construct.

the advantage of avoiding possible immune responses caused by alternative sugars on rhCMP. Herein, we report that 25 mg of the unglycosylated hCMP with high purity was obtained via engineering *E. coli* in a liter culture.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Cultivation Conditions. Both *E. coli* DH5α and *E. coli* BL21(DE3) (Novagen) were used as the hosts for cloning and protein expression, respectively. All *E. coli* strains were grown in Luria–Bertani (LB) medium and 2× YT medium at 37 °C, and the media were supplemented with kanamycin (30 μg mL^{-1}) or ampicillin (80 μg mL^{-1}). Plasmid pMD-18T (Novagen) was used as a cloning vector, and pET-28a(+) (Novagen) was used as an expression vector.

Construction of Recombinant hCMP Expression System. The hCMP gene was chemically synthesized on the basis of the amino acid sequence of hCMP (*10*) by using the preferred codons of *E. coli*. Upstream of the synthesized gene was a *Bam*HI site for ligation followed by an enterokinase site for the purpose of peptide maturation. Downstream of the gene a *Sac*I site was located for the direction of ligation during the construction of the expression vector. The gene was amplified by the Polymerase Chain Reaction (PCR) method using the forward primer 5′-GCCGGATCCGATGACGATGACAAAATTGC-GATTCCGCCGAAAAAGATTCA-3′ (the *Bam*HI site is underlined) and the reverse primer 5′-GGCGAGCTCTTACGCCGGGGTGCTG-GTCACCGCCA CGGTCGGGGTTTCCG- 3′ (the *Sac*I site is underlined).

The PCR product was cloned into pMD-18T followed by nucleotide sequencing. The hCMP gene was inserted directly into pET28a(+) by both *Bam*HI and *Sac*I sites to generate the pET-28a(+)-hCMP construct, which was transformed into *E. coli* DH5α strain for sequencing again. Finally, the plasmid pET-28a(+)-hCMP was transformed to *E. coli* BL21(DE3) for the expression of hCMP.

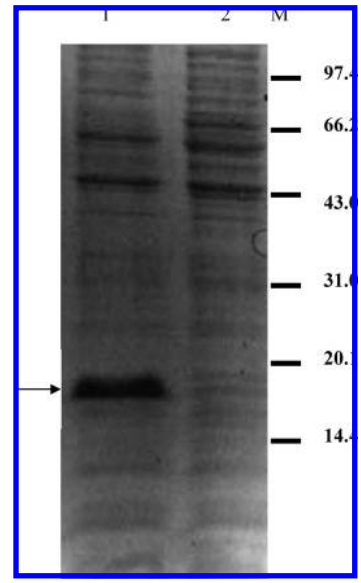

Figure 2. Tricine−SDS-PAGE analysis of expression of recombinant protein induced by IPTG at a final concentration of 0.5 mM. M, protein molecular mass markers (14.4, 20.1, 31.0, 43.0, 66.2, and 97.4 kDa); lane 1, *E. coli* BL21 (pET28a-hCMP), desired fusion protein shown by arrow; lane 2, *E. coli* BL21 (pET-28a) as control.

Expression of Recombinant hCMP. The *E. coli* BL21(DE3) containing the pET-28a-hCMP recombinant plasmid was grown overnight at 37 °C in 15 mL of LB liquid medium with 30 μg mL^{-1} kanamycin (Kan). The culture was diluted 1:100 with 2× YT medium in the presence of 30 μg mL^{-1} Kan and grown at 37 °C with vigorous shaking. When the optical density at 600 nm reached 0.6–0.8, inducing

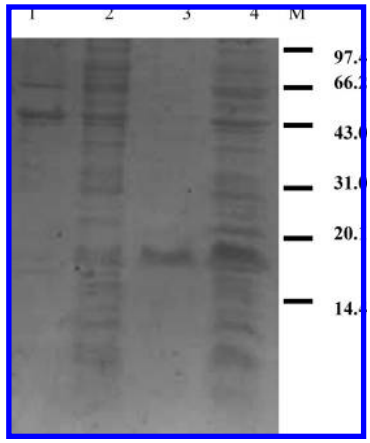

Figure 3. Tricine−SDS-PAGE analysis of the recombinant protein in different fractions of *E. coli* BL21(DE3). M, protein molecular mass markers (14.4, 20.1, 31.0, 43.0, 66.2, and 97.4 kDa); lane 1, inclusion body from *E. coli* BL21 /pET28a-hCMP; lane 2, supernatant from *E. coli* BL21 /pET28a-hCMP; lane 3, protein from periplasmic space of *E. coli* BL21 /pET28a-hCMP; lane 4, crude cell extract from *E. coli* BL21 /pET28a-hCMP.

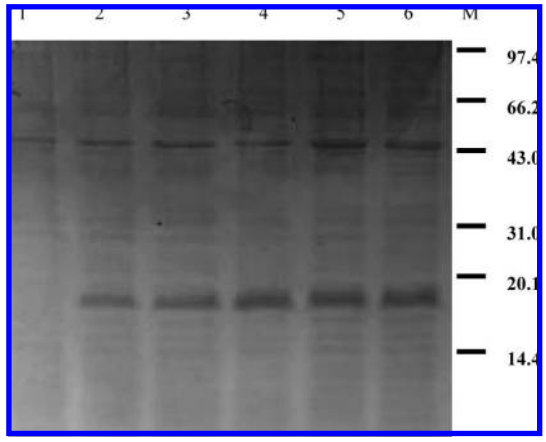

Figure 4. Tricine−SDS-PAGE analysis of the hCMP at different times after induction by 0.5 mM IPTG. M, protein molecular mass markers (14.4, 20.1, 31.0, 43.0, 66.2, and 97.4 kDa); lanes 1−6 represent time points after induction by 0.5 mM IPTG.

expression was performed by adding isopropyl-α-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. After induction for 6 h at 30 °C, cells were harvested by centrifugation at 10000*g* for 10 min at 4 °C. Total proteins from induced culture and a control were analyzed by Tricine−SDS-PAGE. The protein expressed in *E. coli* BL21(DE3) was further identified as His-tagged hCMP by performing Western blot analysis using an anti-His tag antibody (*11*). Protein concentration was estimated by using the Bradford assay with bovine serum albumin as standard. The purity of protein was determined by densitometry of Coomassie blue R-250 stained SDS-PAGE gel.

Purification of Recombinant hCMP. The harvested cells were resuspended in buffer A [25 mM phosphate buffer (PB), 500 mM NaCl, 10 mM imidazole, pH 7.5] and disrupted by sonication. The cell lysate was centrifuged at 10000*g* for 30 min at 4 °C, and the supernatant was filtered through a 0.45 μm hydrophilic polypropylene membrane to prevent clogging of the chromatography medium. The supernatant was applied to a column containing 10 mL of Ni^{2+}-chelating Sepharose FF equilibrated with buffer A, and the column was washed with the same buffer to remove unbound proteins. The His-tagged hCMP was eluted with buffer B (25 mM PB, 500 mM NaCl, 300 mM imidazole, pH 7.5), and all of the eluted fractions were analyzed by Tricine−SDS-PAGE. Fractions containing the hCMP were pooled and dialyzed against 5 mM PBS (pH 7.5) overnight at 4 °C to desalt.

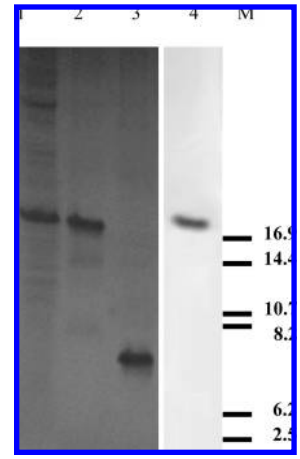

Figure 5. rhCMP purification analyzed on Tricine−SDS-PAGE and identified by Western blotting: lane 1, crude cell extract from *E. coli* BL21 /pET28a-hCMP; lane 2, elution using 300 mM imidazole in wash buffer; lane 3, protein solution after Sephadex G-10 chromatography; lane 4, Western blotting of purified hCMP with anti-His_6 tag monoclonal antibody; M, protein molecular mass markers (2.5, 6.2, 8.2, 10.7, 14.4, and 16.9 kDa).

Table 1. Purification of rhCMP from BL21(DE3)/pET28a(+)-hCMP[a]

purification step	total protein (mg)	object protein (mg)	purity[b] (%)	yield[c] (%)
cell lysate	305	122	40	100
IMAC	115	103.5	90	84.8
cleavage	94	41.4[d]	44	54.8[e]
Sephadex G-10	25[d]	24.8[d]	99	32.8[e]

[a] Star from 1 L of bacteria culture or about 4 g (wet weight) cell pellet. [b] Determined by densitometry based on SDS-PAGE stained by Coomassie blue R-250. [c] Purification yield was calculated on the basis of the amount of object protein. [d] Protein amount of rhCMP. [e] Purification yield of rhCMP is calculated on the basis of the following formula: yield = protein amount of rhCMP/(122 × 6722/10855); 10855 and 6722 are the molecular weights of His_6- rhCMP and rhCMP, respectively.

The desalted fusion protein was digested with enterokinase overnight at 25 °C to remove the His-tag from the fusion protein. The reaction mixture was purified again with a 10 mL Ni^{2+}-chelating Sepharose FF column equilibrated with buffer A. The penetration was collected and desalted with a Sephadex G-10 column using buffer C (5 mM PB, pH 8.0) at 10 mL/min. The purified proteins were lyophilized and stored at −70 °C.

Protein Digestion and MALDI-TOF-MS Analysis. Coomassie-stained gel bands were cut with a scalpel blade into 1 mm pieces, destained with 50% acetonitrile (ACN) containing 50 mM $(NH4)_2CO_3$, and washed in Milli-Q water until the gels became clear. The gel spots were then dehydrated with 100% ACN for 5 min and dried by Speed-Vac for 20–30 min. After being dried, the gels were rehydrated with about 15 μL of cold trypsin solution (12.5 ng/μL) and then incubated at 37 °C for 16–24 h. The gel fragments were extracted twice using 15–25 μL of 50% ACN/0.1% trifluoroacetic acid (TFA) for 30–60 min, and then the two extracts were combined and completely dried by Speed-Vac.

Protein digests (0.5 μL) were spotted on a MALDI target plate, followed by 0.5 μL of CHCA matrix (8 mg/mL in 50% ACN, 0.1% TFA) and then dried. Analysis was performed in reflector mode on a 4700 Proteomics Analyzer TOF/TOF mass spectrometer. For protein identification, PeptideMass (http://www.expasy.org/tools/peptide-mass.html) was used to cleave the hCMP amino acid sequence with trypsin and compute the masses of the generated peptides. All measured peptide masses were then compared with theoretical peptide masses generated by PeptideMass to determine the presence of peptides originating from hCMP.

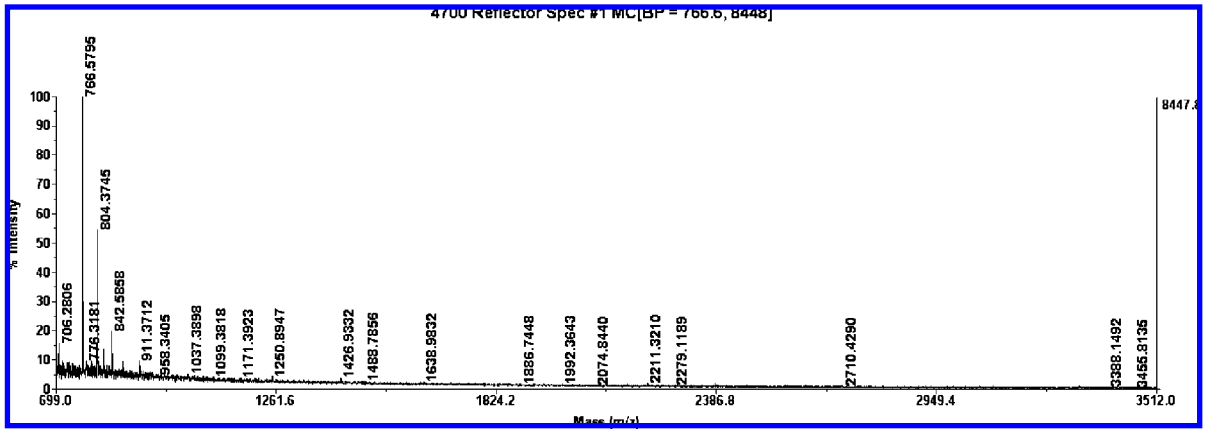

Figure 6. MADLI-TOF-MS spectra of rhCMP subjected to SDS-PAGE and subsequent trypsin digestion.

RESULTS

Construction of Recombinant hCMP Expression System. The chemically synthesized hCMP gene was amplified by PCR, producing a single amplified 231 bp DNA fragment. The resultant DNA product was ligated into pMD18-T, which is an AT cloning vector for DNA sequencing (data not shown). The insert with correct sequence was cut from pMD18-hCMP and inserted into the pET28a vector. The final pET28a-hCMP construct was able to express the recombinant hCMP protein as a fusion protein with a hexahistidine tag (**Figure 1**).

Expression of Recombinant hCMP. Cell cultures were grown in 1 L of 2× YT medium with 30 μg mL^{-1} Kan at 30 °C for 4 h. To confirm the expression of the fusion protein, total proteins from induced cells and a control culture were analyzed by Tricine−SDS-PAGE (**Figure 2**). A clear band from the induction culture showed on the gel. In contrast, this band did not show on the gel from the control culture. The target protein in different fractions was investigated for further optimization of the induction conditions and purification. Total proteins of the whole cell and the proteins located in the cell periplasm, cytoplasm, and inclusion body were extracted individually and analyzed by Tricine−SDS-PAGE (**Figure 3**). Most of the target proteins were located in the cell periplasmic space and cytoplasm, whereas the target peptide located in the inclusion body was at a low relative level.

To get high-level expression of recombinant hCMP, the recombinant *E. coli* BL21 was induced with 0.5 mM IPTG for 0, 2, 3, 4, 5, and 6 h (**Figure 4**). The results showed that when the incubation period was up to 6 h, the yield of the target protein increased as the induction time was prolonged, but when the incubation time was >6 h, the yield of the product was not significantly boosted. The maximum amount of soluble protein was obtained when the induction was carried out with 0.5 mM IPTG at 30 °C for 4 h. Under these conditions the amount of expressed fusion protein could reach 40% of the total cell protein of *E. coli*.

Purification of Recombinant hCMP. Ni^{2+}-chelating chromatography was employed to adsorb the fusion peptides with 6× His from the culture supernatant. A substantial amount of protein from the supernatant was bound to the resin. To remove nonspecifically bound proteins, washing buffer including 100 mM imidazole was used according to our previously optimized results. Most of the recombinant protein was eluted by the buffer containing 300 mM imidazole. The fractions were desalted and concentrated. The recombinant hCMP was further confirmed by Western blotting using anti-His6 tag monoclonal antibody (**Figure 5**, lane 4). The purity of the fusion protein as determined by densitometry could reach approximately 90% (**Figure 5**).

For the maturation of the fusion proteins, it was necessary to remove the 6× His tag from the fusion proteins. The cleavage reaction was carried out, and 94 mg of desalted fusion protein was incubated with 2200 units of enterokinase at 20 °C for 12 h. About 95% fusion protein was cleaved to release rhCMP from the His tag.

Once the His tag was cleaved off, the 6× His tag was removed by nickel affinity chromatography and rhCMP was recovered and subsequently desalted by Sephadex G-10 chromatography with almost 60% recovery (**Figure 5**). After the whole purification procedure, a final yield of 25 mg of purified rhCMP with high purity (up to 99%) was obtained from 1 L of culture (**Table 1**).

Protein Digestion and MALDI-TOF-MS Analysis. To characterize the rhCMP, the purified protein was resolved on Tricine−SDS-PAGE stained with Coomassie G-250, and the appropriate band at 6.7 kDa was cut out and subjected to trypsin digestion for subsequent MALDI-TOF analysis. A representative mass spectroscopy output for trypsin-produced peptides from the 6.7 kDa band was obtained (**Figure 6**). The main spectrum showed a typical MALDI-TOF spectrum with the main peak stemming from peptide IAIPPKK (766.6 Da) originating from rhCMP. This confirmed that the recombinant product was the correct protein, hCMP.

DISCUSSION

Adopting DNA recombinant technology to prepare hCMP has great advantages, including high protein yield and low cost. Moreover, hCMP can be selectively produced in prokaryotic or eukaryotic systems. Recently, Kim has cloned the hCMP gene in *S. cerevisiae* for expression as a secretory product, and high-level expression of the recombinant hCMP was observed (*9*). However, the recombinant hCMP produced by *S. cerevisiae* has a different sugar residue composition from that of human mammary gland cells as a result of their differences in O-glycosylation (*12–14*). Consequently, the recombinant hCMP is unsuitable for injection into patients with severe liver disorders, as an immune response could be caused by the different sugar chains on rhCMP from *S. cerevisae*.

To conquer this problem, we have cloned the hCMP gene in *E. coli* to prepare recombinant hCMP without glycosylation. The hCMP gene was inserted into the pET-28a vector under the control of the T7 *lac* promoter, which promotes high-level expression of recombinant protein in *E. coli*. A final yield of 25 mg of unglycosylated protein with purity up to 99% was obtained from 1 L of *E. coli* culture.

We have investigated the cellular localization of rhCMP and, interestingly, a part of the target protein was directed into the

periplasmic space, which is located between the cell wall and the plasma membrane of *E. coli*. It is known that the target protein can be directed to the periplasmic space by adding cleavable signal sequences to the target protein. These can be classified as either signal recognition particle (SRP) dependent or non-SRP-dependent (*15*). However, no signal sequence was found in the recombinant hCMP; thus, it is presumed that the hCMP is directed into the periplasmic space owing to its mature domain, which is recognized by SecB, the major chaperone of export. Nevertheless, the mechanism of this phenomenon remains unclear, and further investigation is necessary.

ABBREVIATIONS USED

hCMP, human caseinomacropeptide; bCMP, bovine caseinomacropeptide; IPTG, isopropyl-α-D-thiogalactopyranoside; Kan, kanamycin; PCR, Polymerase Chain Reaction; PB, phosphate buffer; ACN, acetonitrile; TFA, trifluoroacetic acid.

ACKNOWLEDGMENT

We sincerely thank Drs. Tim A. Burke and Sa Qila, Department of Molecular Cardiology, Cleveland Clinic Lerner Research Institute, for their critical reading and valuable suggestions.

LITERATURE CITED

(1) El-Salam, M. H. A.; El-Shibiny, S.; Buchheim, W. Characteristics and potential uses of the casein macropeptide. *Int. Dairy J.* **1995**, *6*, 327–341.

(2) Azuma, N.; Yamauchi, K.; Mitsuoka, T. Bifidus growth-promoting activity of glycomacropeptide derived from human κ-casein. *Agric. Biol. Chem.* **1984**, *48*, 2159–2162.

(3) Janer, C.; Pelaez, C.; Requena, T. Caseinomacropeptide and whey protein concentrate enhance *Bifidobacterium lactis* growth in milk. *Food Chem.* **2004**, *86*, 263–267.

(4) Metwally, M. M.; El-Shibiny, S.; Dieb, S. M. El.; El-Etriby, H. M. M.; Assem, F. A. Large scale preparation and growth promoting effects on *Bifidobacterium* of glycomacropeptide from sweet whey. *Egypt. J. Dairy Sci.* **2001**, *29*, 29–36.

(5) Modler, H. W. Bifidogenic factors − sources, metabolism and applications. *Int. Dairy J.* **1994**, *4*, 383–407.

(6) Rutherfurd, K. J.; Gill, H. S. Peptides affecting coagulation. *Br. J. Nutr.* **2000**, *84*, 99–102.

(7) Sollier, C. B. D.; Drouet, L.; Pignand, G.; Chevallier, C.; Caen, J.; Fiat, A. M.; Izquierdo, C.; Jollěs, P. Effect of κ-casein split peptides on platelet aggregation and on thrombus formation in the guinea-pig. *Thromb. Res.* **1996**, *81*, 427–437.

(8) Brody, E. P. Biological activities of bovine glycomacropeptide. *Br. J. Nutr.* **2000**, *84*, 39–46.

(9) Kim, Y. J.; Park, S.; Oh, Y. K.; Kang, W.; Kim, H. S.; Lee, E. Y. Purification and characterization of human caseinomacropeptide produced by a recombinant *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **2005**, *41*, 441–446.

(10) Mercier, J.-C.; Chobert, J.-M. Comparative of the amino acid sequences of the caseinomacropeptides from seven species. *FEBS Lett.* **1976**, *72*, 208–214.

(11) Li, M.; Huang, D. On-column refolding purification and characterization of recombinant human interferon-k1 produced in *Escherichia coli*. *Protein Expr. Purif.* **2007**, *53*, 119–123.

(12) Strahl-Bolsinger, S.; Tanner, W. Protein O-glycosylation in *Saccharomyces cerevisiae*: purification and characterization of the dolichyl-phosphate-D-mannose-protein *O*-D-mannosyltransferase. *Eur. J. Biochem.* **1991**, *196*, 185–190.

(13) Ernst, J. F.; Mermod, J. J.; Richman, L. H. Site-specific O-glycosylation of human granulocyte/macrophage colony-stimulating factor secreted by yeast and animal cells. *Eur. J. Biochem.* **1992**, *203*, 663–667.

(14) Steen, P. V.; Redd, P. M.; Dwek, R. A.; Opdenakker, G. Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 151–208.

(15) Huber, D.; Boyd, D.; Xia, Y.; Olma, M. H.; Gerstein, M.; Beckwith, J. Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. *J. Bacteriol.* **2005**, *187*, 2983–2991.

Received for review August 30, 2007. Revised manuscript received December 11, 2007. Accepted December 13, 2007.

JF072585N