Effect of Tandem Rare Codon Substitution and Vector-Host Combinations on the Expression of the EBV gp110 C-Terminal Domain in Escherichia coli

Sung Jean Park,* Suk Kyeong Lee,† and Bong Jin Lee* ,1

*Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, San 56-1, Shillim-Dong, Kwanak-Gu, Seoul 151-742, Korea; and †Research Institute of Immunobiology, Catholic Research Institutes of Medical Science, The Catholic University of Korea, 505 Banpo-dong, Seocho-Gu, Seoul 137-701, Korea

Received October 12, 2001, and in revised form November 20, 2001

Gp110 of Epstein-Barr virus (EBV) is a glycoprotein that functions exclusively during the assembly of EBV nucleocapsid and the release of infectious EBV. Its C-terminal tail domain (gp110 CTD) is essential for gp110's function and may provide signals that are responsible for the assembly and release of EBV. In the present study, to get large amounts of gp110 CTD for structural analysis, the effects of vector system, codon usage, and host strain on expression levels of gp110 CTD in Escherichia coli have been investigated. The coding region of gp110 CTD (11 kDa) was subcloned into the expression vectors pSE 280, pET-15b, pET-29a, pMAL-c2x, and pGEX-4T-1. Except the pMAL-c2x construct, all the others failed to express detectable amounts of recombinant gp110 CTD. Substituting a tandem rare AGA (Arg) codon with a synonymous CGC (Arg) codon facilitated expression of the recombinant protein, while a protease-deficient host E. coli strain helped in the accumulation of a soluble form of gp110 CTD fusion. The secondary structures of the obtained recombinant gp110 CTD purified from soluble extracts and inclusion bodies were compared using circular dichroism analysis. In aqueous solutions, both samples equally adopt a mixed \( \alpha \)-helix and \( \beta \)-sheet conformation as well as a partly unordered structure. Notably, in the membrane-mimicking environments the helical propensity of gp110 CTD increased up to the previously predicted level based on its sequence, suggesting that gp110 CTD may fold into a more stable conformation through interactions with the cell membrane.

Key Words: EBV; gp110; tandem AGA; circular dichroism; micelle.

Epstein-Barr virus (EBV)\(^2\) is classified among the \( \gamma \)-herpesviruses and establishes latent infection in human B lymphocytes (1). Glycoprotein 110 (gp110), one of the nine glycoproteins encoded by EBV (2–8), has been known to be essential for virus assembly in vivo (9, 10). Gp110 is a homologue of gB, which is the most highly conserved glycoprotein among herpesviruses, with the herpes simplex virus (HSV) gB being the prototype (11). While gp110 has considerable amino acid and predicted secondary structure homology to gB (12), gp110 localizes predominantly to the inner and outer membranes of the nucleus in infected cells, unlike gB, which is mainly found in the plasma membrane (13–16), indicating that the biological role of gp110 is fundamentally different from that of gB (10, 17). Consistent with such observations, antibodies directed against gp110 fail to neutralize EBV infectivity (15).

Gp110 has a primary sequence of 857 amino acids with nine potential sites for N-linked glycosylation, which are potential membrane-spanning domains that consist of 70 amino acids, and an unusual C-terminal domain characterized by a long stretch of basic amino acids (12,13). The C-terminal tail domain of gp110

\(^2\)Abbreviations used: EBV, Epstein-Barr virus; gp110, glycoprotein 110; HSV, herpes simplex virus; CTD, C-terminal tail domain; ER, endoplasmic reticulum; CD, circular dichroism; DPC, dodecyl phosphocholine; SDS, sodium dodecyl sulfate.

To whom correspondence and reprint requests should be addressed. B. J. Lee, Fax: 82-2-872-3632; E-mail: lbj@nmr.snu.ac.kr; S. K. Lee, Fax: 82-2-532-0575; E-mail: sukklee@cmc.cuk.ac.kr.

470
EBV gp110 EXPRESSION

EBV gp110 (gp110 CTD) located in the nucleoplasm may provide signals that are responsible for the assembly of EBV nucleocapsids and their egression through the nuclear membrane (9). Residues 816 to 841 in the tail domain containing four consecutive arginines especially are essential for retention in the nuclear/endoplasmic reticulum (ER) membranes. A recent mutagenesis experiment supported that not only the positive charge of consecutive arginines but also their proper structural configuration is an important factor for gp110's localization to the nuclear/ER membranes (18). Besides the role of gp110 CTD in lytic replication, the QKRAA sequence in the C-terinal tail domain, which is also found in several human pathogens, is related to enhanced humoral and cellular immune responses in early rheumatoid arthritis patients (19–21). This implies the possibility that gp110 CTD has another function associated with human immune reactions.

While the biological function of gp110 has been investigated by several groups of investigators, the structure and the mechanism of its function have not yet been cleared. Pellet et al. (12) reported significant sequence and predicted secondary structure homology between the herpes simplex virus 1 glycoprotein B and EBV gp110. However, there has been no direct experimental evidence to support the predicted secondary structure. To get enough protein for structure analysis, it is necessary to develop an optimized expression system. Recombinant gp110 CTD has previously been expressed using the pMAL system in Escherichia coli by Papworth et al. (22) and was used to obtain an antiserum from an immunized rabbit, which facilitated the analysis of EBV gp110. With this antiserum, expression of gp110 was successfully detected in several EBV-positive B cell lines by immunofluorescence and immunoprecipitation (22).

In this study, the effects of a vector system including the pMAL system, codon usage, and host strain on expression levels of the recombinant gp110 CTD were investigated. The secondary structure of the obtained recombinant gp110 CTD was examined by circular dichroism (CD) and the CD data were compared with the previously proposed model structure (12).

MATERIALS AND METHODS

Materials

Restriction endonucleases were purchased from New England Biolabs Inc. (Beverly, MA), Boehringer Mannheim GmbH (Mannheim, Germany), and Promega (Madison, WI). Cloned Pfu DNA polymerase was obtained from Stratagene (La Jolla, CA). The T7 sequencing version 2.0 DNA sequencing kit, radioactive [α-35S]dATP, plasmid pGEX-4T-1, and glutathione-Sepharose 4B resin were supplied by Amersham Pharmacia Biotech (Uppsala, Sweden). Oligonucleotides used for DNA sequencing and PCR amplification reaction were obtained from Genotech Co. (Dae-Jun, Korea). Dodecyl phosphocholine (DPC) was purchased from Avanti Polar Lipids Inc. (Alabaster, AL). Sodium dodecyl sulfate (SDS) was obtained from Sigma (St. Louis, MO). All materials were of reagent or biotechnology grade.

Construction of Plasmids

To select a system that can express the recombinant gp110 CTD at a high level, various plasmids carrying the sequence coding for gp110 CTD (including amino acids 758 to 857 of EBV gp110) were constructed. The plasmid pSVgp110 encoding the full sequence of gp110 (23) was used as a template to prepare gp110 CTD constructs. The plasmids were generated by inserting the PCR fragment containing the gp110 CTD region into the appropriately digested vectors. The gp110 CTD constructs prepared using pSE 280, pET-15b, pET-29a, pMAL-c2x, and pGEX-4T-1 were designated pj 1, pj 2, pj 3, pj 4, and pj 5, respectively. The primer pairs used in PCR amplification are shown in Table 1. The constructed plasmids were confirmed by restriction analysis and DNA sequencing.

Mutagenesis of Tandem AGA Codons

To facilitate expression of gp110 CTD, the consecutive AGA codons were replaced with the major synonymous CGC codons by site-directed mutagenesis. PCR amplification was performed with Pfu DNA polymerase using two overlapping primers, 5'-GTTTTCCAGGCCTACG and 5'-GA TCGTGATAGCGGCGGCGGCGTAGGCCTGGAAAA (sense) and 5'-GA TCGTGATAGCGGCGGCGGCGTAGGCCTGGAAAA (antisense), containing the synonymous codons (underlined). The constructed silent mutant plasmids were confirmed by restriction analysis and DNA sequencing. The plasmids pj 1, pj 2, and pj 5 were mutated and designated pj 1-m, pj 2-m, and pj 5-m, respectively (Table 2).

Expression of Recombinant gp110 CTD

Each gp110 CTD construct was transformed into either E. coli BL21 or DH5α. The transformants were cultured in TBG-M9 broth supplemented with ampicillin (50 mg/ml). After incubation at 37°C for 7 h, protein expression was induced by adding IPTG to a final concentration of 1 mM. After 4 h postinduction at 37°C, cells were harvested by centrifugation at 8000 rpm at 4°C for 15 min (Beckman J 2-MC). Gp110 CTD was obtained from either soluble extract or inclusion bodies depending on the E. coli strain used. Gp110 CTD from soluble extract is designated “soluble gp110 CTD” or “soluble protein” while gp110 CTD from inclusion bodies is designated “refolded gp110 CTD” or “refolded protein” throughout the text.
TABLE 1
Primers Used for Plasmid Construction

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>PCR primers used for construction</th>
<th>Endonuclease sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJ1</td>
<td>5'-TTATCCATGGCTCAAGATGCCACGCCC-3'</td>
<td>NcoI</td>
</tr>
<tr>
<td></td>
<td>5'-TATCTGGGCTCAAATCTCTGTCG-3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>pJ2</td>
<td>5'-TGGATTCGGCAGATGTCGAGCAGCCG-3'</td>
<td>BamHI</td>
</tr>
<tr>
<td>pJ3</td>
<td>5'-TTATCCATGGCAGATGTCGAGCAGCCG-3'</td>
<td>NcoI</td>
</tr>
<tr>
<td></td>
<td>5'-TATCTGCAGCAGATGTCGAGCAGCCG-3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>pJ4</td>
<td>5'-TATGGATCCAGATGTCGAGCAGCCG-3'</td>
<td>BamHI</td>
</tr>
<tr>
<td>pJ5</td>
<td>5'-TTATGAGCTTGAGATGTCGAGCAGCCG-3'</td>
<td>EcoRI</td>
</tr>
</tbody>
</table>

Note. Restriction sites are underlined.

Purification of gp110 CTD

Purification of soluble gp110 CTD. The bacterial cell pellet was suspended in 100 ml of the lysis buffer (50 mM Tris–Cl (pH 8.0), 1 mM dithiothreitol (DTT), 10% glycerol, and 0.1 mM PMSF) containing 0.2 mg/ml of lysozyme. Bacterial lysis was performed by sonication (4 × 30-s pulses) until the lysate became clear. After lysis, the supernatant was applied to a glutathione-Sepharose 4B column (2.5 × 20 cm) equilibrated with phosphate-buffered saline (pH 7.4). Bound GST-gp110 CTD was eluted with 20 ml of the elution buffer (30 mM glutathione in 50 mM Tris–Cl, pH 8.0) at a flow rate of 1 ml/min. The fused GST was removed from eluted protein by treating with thrombin at 20°C for 16 h. After cleavage, the reaction mixture was loaded onto a Q-Sepharose column, followed by Blue-Sepharose 6B chromatography (Amersham Pharmacia Biotech). Protein samples were collected at the different purifying steps and analyzed by SDS–PAGE and immunoblotting.

Purification of refolded gp110 CTD. The culture, harvest, lysis, and purification of the refolded protein were identical to those for the soluble protein, except for the procedures described below. After centrifugation of the bacterial lysate, precipitates were solubilized in 100 ml of 50 mM Tris–Cl buffer (pH 8.0) supplemented with 6 M guanidine–HCl, 10 mM DTT, 10 mM EDTA, and 150 mM NaCl. One hundred milliliters of guanidine–HCl-solubilized protein was dialyzed overnight at 4°C with slow stirring against 2 L (× 2) of the refolding buffer containing 0.5 M arginine, 50 mM Tris buffer (pH 8.0), 150 mM NaCl, 10 mM DTT, and 1 mM PMSF (24–26). L-Arginine was removed by additional dialysis against 2 L of 50 mM Tris buffer (pH 8.0) containing 150 mM NaCl, 10 mM DTT, and 1 mM PMSF. Finally, the dialysate was clarified by centrifugation (12,000 rpm, 4°C, 30 min).

Immunoblotting

Protein samples were resolved by SDS–15% polyacrylamide gel electrophoresis prior to electrophoretic transfer to a nitrocellulose membrane. Western transfer was carried out in cold 0.025 M Tris–0.19 M glycine (pH 8.8), 20% methanol, and 0.1% SDS in a Bio-Rad mini-gel system at 300 mA for 1 h. Nitrocellulose membrane was then blocked overnight at 4°C with 5% skim milk.

TABLE 2
Constructed Plasmids and Their Expression

<table>
<thead>
<tr>
<th>Classification of constructs</th>
<th>Designation</th>
<th>Gene product</th>
<th>Vector</th>
<th>Promoter</th>
<th>Host strain</th>
<th>Expression result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>pJ1</td>
<td>gp110 CTD</td>
<td>pSE280</td>
<td>trc</td>
<td>DH5α, BL21</td>
<td>Undetectable</td>
</tr>
<tr>
<td></td>
<td>pJ2</td>
<td>His tag-gp110 CTD</td>
<td>pET-15b</td>
<td>T7</td>
<td>BL21(DE3)α</td>
<td>Undetectable</td>
</tr>
<tr>
<td></td>
<td>pJ3</td>
<td>S tag-gp110 CTD</td>
<td>pET-29a</td>
<td>T7</td>
<td>BL21(DE3)α</td>
<td>Undetectable</td>
</tr>
<tr>
<td></td>
<td>pJ4</td>
<td>MBP-gp110 CTD</td>
<td>pMAL-c2x</td>
<td>tac</td>
<td>TB1</td>
<td>Good but degraded</td>
</tr>
<tr>
<td></td>
<td>pJ5</td>
<td>GST-gp110 CTD</td>
<td>pGEX-4T-1</td>
<td>tac</td>
<td>DH5α, BL21</td>
<td>Undetectable</td>
</tr>
<tr>
<td>Silent mutant</td>
<td>pJ1-m</td>
<td>gp110 CTD</td>
<td>pSE280</td>
<td>trc</td>
<td>DH5α, BL21</td>
<td>Undetectable</td>
</tr>
<tr>
<td></td>
<td>pJ2-m</td>
<td>His tag-gp110 CTD</td>
<td>pET-15b</td>
<td>T7</td>
<td>BL21(DE3)α</td>
<td>Undetectable</td>
</tr>
<tr>
<td></td>
<td>pJ5-m</td>
<td>GST-gp110 CTD</td>
<td>pGEX-4T-1</td>
<td>tac</td>
<td>DH5α</td>
<td>Very good (as inclusion bodies)</td>
</tr>
</tbody>
</table>

α T7 vectors were transformed into BL21 (DE3) and its analogues such as BL21(DE3) pLysS and BL21(DE3)-RIL.
milk in TBS (20 mM Tris–HCl, 150 mM NaCl, pH 7.6). After washing with TTBS (Tris-buffered saline with 0.1% Tween 20, pH 8.0), the membrane was incubated for 60 min with polyclonal antibodies for gp110 CTD at room temperature. The membrane was then washed and incubated for 60 min with horseradish peroxidase-conjugated secondary antibody. After further washing, immunoreactive proteins were visualized using an enhanced chemiluminescence detection system according to the protocol of the manufacturer (Amersham Life Sciences, Amersham, Arlington Heights, IL).

Circular Dichroism Analysis

To characterize the secondary structure of the refolded and soluble gp110 CTD, CD analysis was performed using a 2-mm path length cuvette in a Jasco J-715 spectropolarimeter at 20°C. CD measurements were carried out in a wavelength range between 190 and 240 nm. The refolded and soluble gp110 CTD purified as described above were dialyzed against 5 mM potassium phosphate buffer (pH 7.4) and concentrated using a Centricon 30 unit (Amicon Inc., Beverly, MA). The concentration of protein used in CD measurements was 50 μM. The resultant spectra were corrected for the buffer signal. To examine the structural change in membrane-mimicking environment, the spectra for 50 μM gp110 CTD were also collected in 5 mM potassium phosphate buffer (pH 7.4) with 5 mM DPC or 10 mM SDS. A baseline of pure micelle solution was recorded and subtracted from each spectrum.

RESULTS

Construction of Plasmids and Expression of gp110 CTD

In order to compare the effect of vectors on the amount of recombinant gp110 CTD produced in E. coli, the coding region for gp110 CTD was subcloned into pSE280, pETs, pMAL-c2x, and pGEX-4T-1 (Fig. 1). All of these constructs except pl 4 failed to express any significant amount of recombinant protein (Table 2, Fig. 2A). Using a pMAL system that was previously used by Papworth et al. (22), gp110 CTD was successfully expressed, but the yield of the protein was relatively low due to degradation of the protein after cell lysis (Fig. 2B).

Codon Substitution of Tandem AGA to Facilitate Expression of gp110 CTD

To increase the translation efficiency of the gp110 CTD mRNA, rare codons of gp110 CTD constructs were replaced with synonymous major codons by site-directed mutagenesis. The plasmids pl 1, pl 2, and pl 5 were mutagenized so that codons AGA80 and AGA81 were replaced by the synonymous E. coli codons CGC80 and CGC81, resulting in pl 1-m, pl 2-m, and pl 5-m, respectively. When the DH5α strain was transformed with pl 1-m or pl 5-m and BL21(DE3) was transformed with pl 2-m, only pl 5-m expressed gp110 CTD at a high enough level (Fig. 2C). However, expression of a fusion protein with a molecular mass of about 37 kDa was confirmed mainly in the cell pellet, indicating that GST-gp110 CTD was expressed mainly as insoluble inclusion bodies (Fig. 3A, lane d). To obtain a soluble form of GST-gp110 CTD, we changed various expression conditions such as culture temperature (10–37°C), IPTG concentrations (0.1–1 mM), and induction time (1–6 h). While these modulations didn’t significantly improve the yield of the fusion protein obtained in the soluble extract (data not shown), switching the host strain from DH5α to BL21, which is deficient in the OmpT and Lon proteases, resulted in remarkably enhanced expression of the soluble GST-gp110 CTD (Fig. 3B, lane c).

Purification of Recombinant gp110 CTD

Both refolded and soluble forms of gp110 CTD were purified by stepwise chromatographic methods described...
FIG. 2. Protein expression from the plasmids constructed in this study. (A) Vectors were transformed into appropriate host strains (Table 2). Whole cell lysates before (lanes 1) and after (lanes 2) induction were separated by SDS–PAGE. Only MBP-gp110 CTD was expressed but the yield of purified gp110 CTD was very low. (B) SDS–PAGE of the recombinant protein MBP-gp110 CTD; lanes: a, induced whole-cell lysate; b, lysate supernatant; c, lysate pellet; d, eluate from amylose resin. The MBP-gp110 CTD and its degraded products after cell lysis are marked by arrows (on the right side is an enlarged picture of lane d). (C) Tandem rare codons AGA AGA for arginines in pJ1, pJ2, and pJ5 were replaced with CGC CGC to produce pJ1-m, pJ2-m, and pJ5-m, respectively. These silent mutant constructs were transformed into E. coli and the expression of gp110 CTD was checked after induction.

under Materials and Methods. The yields of protein purified from inclusion bodies and soluble extracts were approximately 5 and 9 mg per liter of E. coli culture, respectively (Table 3). SDS–PAGE showed that the purified gp110 CTD from inclusion bodies and soluble extracts appeared homogeneous and identical under reducing conditions, with an apparent molecular mass of approximately 11 kDa (Figs. 3A, lane h, and 3B, lane g). The inclusion bodies from pJ5-m/DH5α were dissolved with 6 M guanidine–HCl and refolded by dialysis in Tris–Cl buffer containing L-arginine (Fig. 3A, lane e). The refolding was facilitated by an increase in the arginine concentration and a decrease in the protein concentration (Fig. 4). The limiting concentration for optimal recovery of GST-gp110 CTD was about 10 μg/ml in refolding buffer without L-arginine (Fig. 4A), while elevated up to 300 μg/ml when 0.5 M L-arginine was added to the refolding buffer (Fig. 4A). After the refolding process, 53% of the total GST-gp110 CTD present in the guanidine–HCl-solubilized sample was recovered (Fig. 4). Slow or rapid dilution is known to be an alternative to dialysis but was not effective in the refolding of GST-gp110 CTD and optimized dialysis gave a more improved yield than the dilution method in our study (data not shown). With other refolding methods using detergents, glycine, and glycerol, the GST-gp110 CTD aggregated during dialysis, and the quantity of recovered soluble protein was very low (data not shown).

Immunoblotting

To confirm the identity of the purified peptide, immunoblot analysis was carried out for the samples taken...
FIG. 3. SDS–PAGE of the recombinant GST-gp110 CTD. (A) Purification of gp110 CTD from inclusion bodies; lanes: a, uninduced whole-cell lysate; b, induced whole-cell lysate; c, lysate supernatant; d, lysate pellet; e, supernatant after dialysis using L-arginine; f, eluate from glutathione-Sepharose 4B chromatography; g, products partially cleaved by thrombin at 4°C for 1 h; h, purified refolded gp110 CTD. (B) Purification of gp110 CTD from soluble extracts; lanes: a, uninduced whole-cell lysate; b, induced whole-cell lysate; c, lysate supernatant; d, lysate pellet; e, eluate from glutathione-Sepharose 4B chromatography; f, products partially cleaved by thrombin at 4°C for 1 h; g, purified soluble gp110 CTD.

Circular Dichroism Analysis

Secondary structure analysis of the gp110 CTDs purified by different protocols. The far-UV CD spectra were used to analyze the secondary structure of the two forms of gp110 CTD. Figure 6 shows the CD spectra of the soluble and refolded gp110 CTD. Both spectra showed similar profiles, suggesting that gp110 CTD purified by the two different protocols have the same conformational characteristics. The CD curves reveal that the structure of gp110 CTD is not so highly ordered, even though the presence of two negative extremes at 222 and 205 nm and the positive intensity below 200 nm point to a partially helical conformation adopted by the protein (27). The intensity of negative minimum at 222 nm was relatively small compared to that of the whole α-helical proteins, which indicates that gp110

TABLE 3

<table>
<thead>
<tr>
<th>Purification step</th>
<th>Total protein (mg)</th>
<th>Purity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Soluble gp110 CTD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell lysate</td>
<td>285 (60)a</td>
<td>21b</td>
</tr>
<tr>
<td>GST affinity column</td>
<td>84 (58)a</td>
<td>69b</td>
</tr>
<tr>
<td>Thrombin digest</td>
<td>77 (16)c</td>
<td>21d</td>
</tr>
<tr>
<td>Final product</td>
<td>9</td>
<td>99</td>
</tr>
<tr>
<td><strong>Refolded gp110 CTD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusion bodies</td>
<td>115 (54)a</td>
<td>47b</td>
</tr>
<tr>
<td>Refolding by dialysis</td>
<td>61 (36)a</td>
<td>59p</td>
</tr>
<tr>
<td>GST affinity column</td>
<td>42 (35)a</td>
<td>83b</td>
</tr>
<tr>
<td>Thrombin digest</td>
<td>40 (10)d</td>
<td>25d</td>
</tr>
<tr>
<td>Final product</td>
<td>5</td>
<td>99</td>
</tr>
</tbody>
</table>

Note. One liter of high-density culture contained about 4 g of wet cell weight.

a The amount of total protein was determined by protein assay kit (Pierce) with BSA as a standard; numbers in parentheses are estimated amounts of GST fusion by SDS–PAGE.

b Purity is defined as the percentage of GST fusion in the purified protein preparation.

c Numbers in parentheses are estimated amounts of cleaved gp110 CTD by SDS–PAGE.

d Purity is defined as the percentage of cleaved gp110 CTD in the purified protein preparation.
of gp110 CTD in micelle solutions exhibited evidence of significant \( \alpha \)-helical structure, as indicated by the strong positive band at about 190 nm and negative bands at 208 and 222 nm (Fig. 7). As shown in Table 4, the induced helicity of gp110 CTD in micelle solutions originated from decreased random coiling, while \( \beta \)-structural elements were not much changed. These data apparently indicate that the largely unordered regions of gp110 CTD have intrinsic \( \alpha \)-helical propensities in its amino acid sequences. Gp110 CTD in DPC micelle solution appeared to adopt a slightly less helical conformation than in SDS micelle solution (Fig. 7 and Table 4). This distinct conformation may result from the difference in the charged head group of the micelle, with which gp110 CTD is supposed to interact.

DISCUSSION

Many recombinant viral proteins are poorly produced in \textit{E. coli} (35, 36). The adoption of a generally applicable expression method is precluded in these cases and various strategies to optimize protein production in \textit{E. coli} should be considered, for example, substitution of the vector, modulation of translational efficiency, using fusion partners, changing fermentation conditions, and so forth. In this study, the factors limiting the expression of gp110 CTD in \textit{E. coli} can be summarized as follows: First, the tandem AGA present in the gp110 CTD coding region negatively affects translational efficiency. Nine percent of 100 amino acids of gp110 CTD are encoded
by rare codons and 5% are encoded by the rarely used arginine codons AGG and AGA. It is noteworthy that the diminished expression of recombinant protein in E. coli depends not only on the prevalence of these codons but also on their relative locations in the gene (37–40). Second, the susceptibility of gp110 CTD to proteases (especially Lon and OmpT proteases) may cause degradation of the protein, considering that gp110 CTD could be obtained in soluble extracts by simply switching the host E. coli strain to protease-deficient BL21. Third, toxicity of the protein may also influence expression. Gp110 CTD was not expressed from pl 1-m and pl 2-m up to detectable levels in E. coli BL21 analogues that are Lon and OmpT protease-deficient. Thus, it seemed that the tandem AGA and the susceptibility to proteases were not primary causes for the failed expression of gp110 CTD in pl 1-m/BL21 and pl 2-m/BL21 (DE3). It is noteworthy that the optical density of growing E. coli did not increase after these two clones were induced by IPTG (data not shown), in agreement with the results of Popworth et al., which showed that several truncated gp110 were not expressed in E. coli (22). These data imply that the gene product may be toxic to the host cell. It is very probable that GST fusion could somehow help the host cell to escape from the toxic effects of gp110 CTD. Without replacing the tandem AGA, gp110 CTD was expressed well in the pMAL system. This could be explained as a result of decreased toxicity by MBP fusion and/or enhanced translation by stabilization of the mRNA transcribed from the MBP-gp110 CTD fusion construct (41–44). Gp110 CTD could be obtained with over 99% purity either from inclusion bodies (pl 5-m/DH5α) or from soluble extracts (pl 5-m/BL21). As the proper conditions of refolding vary from protein to protein, the choice of a refolding system is very important for a good result. In the case of gp110 CTD, the refolding of the protein could be sufficiently achieved using L-arginine.

Since no functional assay of gp110 CTD exists, the structural integrity of gp110 CTD was probed by comparing CD spectra of the soluble gp110 CTD and the refolded gp110 CTD. The experimental CD data were also compared with the published model structure (12). The CD spectrum of the refolded gp110 CTD was identical to that of the soluble gp110 CTD, indicating that the renaturation of gp110 CTD from inclusion bodies was complete (45). Overall the secondary structure of gp110 CTD in an aqueous buffer can be described as an α + β structure including a large portion of unordered structure. However, the proposed model structure of gp110 CTD consists of a short β-sheet and four α-helical segments (12), which is inconsistent with our experimental results. Numerous proteins involved in important cellular regulatory functions lack an intrinsic globular structure under physiological conditions, and their unstructured domains fold into a stable structure upon binding to the targets (reviewed in 46). The abundance of random coil in gp110 CTD may be converted into more ordered structure when gp110 CTD interacts with other cellular components inside the cell. On the basis of gp110’s unique localization to the nuclear/ER membrane, as well as the presence of the localizing signal in the C-terminal domain, we postulated that ER/nuclear membrane could interact with the C-terminal domain, affecting its conformation. Notably, in SDS and DPC micelle solutions, the CD curves of gp110 CTD showed drastically increased helicity, suggesting that the C-terminal domain indeed interacts with the micelles. The increased helical propensity in micelle solutions is consistent with the proposal of Pellett et al. (12).

The predicted structure of gp110 CTD by a combination of hydrophilic, helical wheel, and empirically based secondary structure analyses (12) showed an interesting feature. All four helices share the feature of an amphipathic α-helix by helical wheel analysis and

<table>
<thead>
<tr>
<th>Secondary Structural Elements of gp110 CTD in Various Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents (%) Gp110 CTD in buffer In 5 mM DPC In 10 mM SDS</td>
</tr>
<tr>
<td>α-Helix</td>
</tr>
<tr>
<td>β-Sheet</td>
</tr>
<tr>
<td>β-Turn</td>
</tr>
<tr>
<td>Random coil</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Note. The fractions of α-helix, β-sheet, β-turn, and random coil were calculated using the data from the CD spectra shown in Fig. 7 by JASCO built-in software, SSE (Secondary Structure Estimation).
the longest $\alpha$-helix (gp110$_{788-816}$) among them is expected to interact with other proteins that are yet unrevealed (12). Amphipathic helices are most common among structures of membrane proteins. Often, the helices exposed to the exterior of the membrane are stabilized by the interaction of the polar residues with water and the head group of the membrane via electrostatic interaction and hydrogen bonding (30, 33, 34). Based on the results of CD analysis, it is expected that the structural properties of gp110 CTD are directly influenced in a membrane environment. The unusual enrichment of basic residues such as Arg, which is important for gp110’s unique localization, may be required for stabilizing the global fold of gp110 CTD by interaction with the polar lipid head groups protruding from the surface of the ER/nuclear membrane.

In conclusion, an efficient system for the expression of gp110 CTD in E. coli was developed. Changing the fusion partner and substituting tandem AGA codons with synonymous codons were greatly effective in the expression of gp110 CTD, in addition to using a protease-deficient host strain. This result might help to solve problems encountered in expressing other cloned genes. The current study suggests that the C-terminal domain is not so highly ordered on its own and forms a more ordered structure upon interaction with target molecules. As suggested earlier (12, 18), the target molecules are supposed to be proteins primarily, but here we show the possibility of direct interaction between gp110 CTD and membrane components. Further defined studies are required to delineate whether gp110 CTD adopts a stable fold through interaction with the ER/nuclear cell membrane in vivo and such conformational change is a prerequisite for the interaction of gp110 CTD with other proteins to complete its biological role.

ACKNOWLEDGMENTS

This work was supported by Grant 981-0508-045-2 from The Korea Science and Engineering Foundation. It was also supported in part by the 2001 BK21 Project for Medicine, Dentistry, and Pharmacy. We gratefully thank Dr. H. W. Kang, Dr. S. H. Park, and H. E. Kim for their valuable comments.

REFERENCES


