Original Article

An efficient method for refolding the extracellular portion of CD147 from the total bacterial lysate

Fei Song†, Xin Zhang†, Yifei Li, Qiang Ru, Xiaobai Ren, Bin Xia*, and Zhi-Nan Chen*

1Department of Cell Biology, Cell Engineering Research Centre, Cancer Biology of State Key Laboratory and State Key Discipline of Cell Biology, Fourth Military Medical University, Xi’an 710032, China
2Beijing Nuclear Magnetic Resonance Center, College of Chemistry and Molecular Engineering, and College of Life Science, Peking University, Beijing 100871, China
†These authors contributed equally to this work.
*Correspondence address. Tel: +86-29-84774545; Fax: 86-29-83293906; E-mail: zhinanchen@fmmu.edu.cn (Z.-N.C.)/Tel/Fax: +86-10-62758127; E-mail: binxia@pku.edu.cn (B.X.)

CD147, also known as basigin [1], M6 antigen [2], extracellular matrix metalloproteinase inducer (EMMPRIN) [3], and HAb18G [4,5], is a type I integral transmembrane glycoprotein ubiquitously expressed in different cell types and in human tumor cells at high levels [3,6]. CD147 belongs to the immunoglobulin superfamily (IgSF), and it is composed of N-terminal signal peptides, an extracellular portion with two Ig domains, a single α-helical transmembrane domain, and a short cytoplasmic domain with 39 residues [3].

CD147 was found to be highly expressed on the cell surface of most malignant tumors [7], and it could stimulate adjacent fibroblasts or tumor cells to produce matrix metalloproteinases (MMPs). It has been shown that CD147 also stimulates expression of vascular endothelial growth factor that leads to angiogenesis [8], and hyaluronan that causes anchorage-independent growth/multidrug resistance [9]. Thus, CD147 is an important molecule in tumor progression and is an attractive target for antitumor treatment [10].

Multiple functions except for MMP induction have been identified for CD147 in physiological and pathological conditions. CD147 was found to be involved in the activation and development of T cells [11–14], the formation and development of neural network, the escort of monocarboxylate transporter (MCT) to the plasma membrane [15], and served as a receptor for cyclophilin A (CyPA) [16,17]. Furthermore, plenty of studies have also shown that CD147 plays roles in fetal development, retinal function, spermatogenesis and fertilization, plaque formation within Alzheimer’s-stricken brains [18], HIV infection [19], and rheumatoid arthritis [20,21].

Although CD147 has been implicated in these physiological and pathological activities, the precise molecular mechanisms for its functions are still largely unclear. So far, only a few CD147-interacting molecules, such as cyclophilins, caveolin-1, MMPs, integrins, and MCT, have been characterized. The extracellular portion of CD147...
(residues 22–205, CD147EC) is responsible for many interactions of CD147 with its partners [10]. Meanwhile, it was reported that the self-association of CD147EC played important biological roles. CD147EC contains two Ig domains: each has one two-disulfide bond (C41-C87 and C126-C185). The first Ig domain is suggested to involve in the MMP induction via self-association [22–24], and is responsible for the association with integrins [25]. The second Ig domain is required for association with CyPA as a receptor [16]. Therefore, structural characterization of these interactions will provide insight into understanding of their mechanism in tumor progression, viral infection and rheumatoid arthritis pathogenesis, etc.; and may inspire the design of therapeutic drugs.

Our laboratory has recently solved the crystal structure of the CD147EC, and it is consisted of an N-terminal Ig C2 domain and a C-terminal Ig I domain connected by a 5-residue linker [26]. Currently, we are using nuclear magnetic resonance (NMR) spectroscopy to characterize the solution structure along with its interactions with other proteins. Although we were able to obtain soluble CD147EC expressed in Escherichia coli for the crystal structure determination, the soluble protein yield is quite low (~2 mg per liter culture). Most of the CD147EC protein expressed in E. coli was found to form inclusion body, which made it very inefficient and expensive for producing isotopically labeled NMR protein samples at a reasonable concentration. It was reported that high soluble protein yield can be achieved through refolding of the extracellular portion of CD147 via a 2-step dialysis protocol [27].

Here, we report a novel protocol for efficiently refolding CD147EC from total bacterial lysate and the procedure is very simple and economical. We were able to obtain over 25 mg of well-folded CD147EC from 1 l of E. coli culture in M9 minimal medium, and the refolded CD147EC was functionally active in stimulating MMP-9 expression in fibroblast cells.

Materials and Methods

Construction of expression plasmid

The DNA fragment of CD147EC encoding residues 22–205 was amplified via polymerase chain reaction (PCR) with primers 5′-GCCGAATTCATATGGCTGCTGGTAC CGTTTTCACCACCGTGAAGACCTG-3′ and 5′-CAATTACTCGAGTTAGTGCGTGCGCACGCGGAGCG-3′, in which Ndel and Xhol restriction enzyme sites have been added at the 5′ and 3′ ends, respectively. A stop codon was introduced after the codon of residue 205. PCR experiment was carried out over 30 cycles of 95°C for 1 min, 56°C for 1 min, and 72°C for 1 min using pfu polymerase. The Ndel and Xhol digested PCR product was ligated into prokaryote expression vector pET-32a (Novagen, Gibbstown, USA), and new DNA construct was verified by DNA sequencing.

DNA fragments encoding the N-terminal Ig domain (N-CD147EC, residues 22–103) and the second Ig domain (C-CD147EC, residues 99–205) were PCR-amplified and inserted into the vector pET32a between Ndel and Xhol cloning sites. The expressed protein includes a hexahistidine tag at the C-terminus.

Protein expression

The recombinant plasmids pET32a/CD147EC, pET32a/N-CD147EC, and pET32a/C-CD147EC were transformed into BL21(DE3) E. coli cells, respectively. A single colony of transformed cells was inoculated into 50 ml of Luria-Bertani (LB) medium with 100 μg/ml of ampicillin and the bacteria were grown overnight at 37°C. This starter culture was transferred into 11 of LB medium for continuing growth, supplemented with 100 μg/ml of ampicillin. When OD600 value of the culture reached 0.6, the protein synthesis was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 100 mg/l. After a certain time of incubation at 37°C, the bacteria were harvested by centrifugation (12,000 g). The cell pellet of CD147EC was resuspended in 20 ml of a lysis buffer with 50 mM Tris and 50 mM NaCl (pH 8.5), and the cell pellet of N-CD147EC and C-CD147EC were resuspended in 20 ml of a lysis buffer with 50 mM sodium phosphate and 300 mM NaCl (pH 8.0), and then stored at −20°C.

15N isotopic labeling

BL21(DE3) E. coli cells harboring the expression plasmids were first grown in 1 l of LB medium. The cells were centrifuged gently for 5 min at 3500 g when OD600 value of the culture reached 0.6. The cells were then resuspended in 500 ml of M9 minimal medium. The M9 minimal medium (pH 7.0) contained 17.1 g Na2HPO4 12H2O, 3 g KH2PO4, 4 g/l glucose, and 1 g/l15NH4Cl. Cells were incubated at 37°C for 1 h for recovery, and protein expression was then induced with the addition of 100 mg/l of IPTG. The cells continued growing at 37°C before harvest. For C-CD147EC, the pH of bacterial culture in M9 medium was adjusted to 8.3 by adding NaOH at 4 h after induction and was maintained at this level till harvest.

Protein refolding from total bacteria lysate

The frozen cells were first lysed by freezing and thawing, followed by extensive sonication on ice. Utrapure urea was added to the cell lysate to a final concentration of 8 M, and the urea was solubilized by stirring at room temperature until the mixture was clear. Oxygen was removed thoroughly from the solution by repeated vacuuming and
argon gas purging to maintain a deoxidizing condition. Then, 10 mM of DTT was added into solution to reduce the disulfide bond. Afterwards, the denatured cell lysate was diluted into eight-fold degassed refolding buffer (lysis buffer added 1 mM DTT) expeditiously and agitated for 5 min to allow the protein to refold. The forming of disulfide bonds was triggered by adding of two-fold undegassed lysis buffer subsequently. The refolding protein mixture was incubated for an additional 30 min at room temperature in air, and was then centrifuged at 18,000 g for 30 min at 4°C to remove insoluble components.

**Protein purification**

For CD147EC purification, the supernatant was loaded onto a DE52 anion exchange column (Whatman, Piscataway, USA) pre-equilibrated with 50 mM Tris buffer with 50 mM NaCl, pH 8.5. The protein was eluted with elution buffer (50 mM Tris and 350 mM sodium chloride, pH 8.5) through a linear gradient profile of 15-fold column volume (300 ml). The fraction containing CD147EC was pooled and concentrated using Amicon Ultra-15 Centrifugal Filter Units from Millipore (Billerica, USA) and was further purified by gel filtration. For CD147EC, the filter membrane molecular weight cut off (MWCO) is 10 kDa, and that for the two Ig domains is 3 kDa.

For N-CD147EC or C-CD147EC purification, the refolding supernatant was loaded directly onto a Ni²⁺-NTA agarose column (Qiagen, Hilden, Germany), pre-equilibrated in 50 mM sodium phosphate buffer (pH 8.0) with 10 mM imidazole and 300 mM NaCl. The target protein was eluted with 200 mM imidazole, and was further purified by gel filtration.

For gel filtration, a preparative Superdex-75 size-exclusion column (GE Healthcare, Wisconsin, USA) equilibrated with 50 mM Tris buffer (pH 7.0) with 50 mM NaCl using an ÄKTA purifier100 system (GE Healthcare). Protein concentration was estimated using extinction coefficients at 280 nm of 26,720 M⁻¹ cm⁻¹ for CD147EC, 12,615 M⁻¹ cm⁻¹ for N-CD147EC, and 14,105 M⁻¹ cm⁻¹ for C-CD147EC. The extinction coefficient values (at 280 nm) were predicted based on protein amino acid sequences using online software ProtParam tool from Expasy (www.expasy.org).

**NMR spectroscopy**

All NMR samples contained ~1 mM protein in 50 mM Tris buffer (pH 7.0) in 90% H₂O/10% D₂O, with 50 mM NaCl, and Complete Cocktail Proteinase Inhibitor Tablets (Roche, Basel, Switzerland). NMR spectra were acquired at 298 K on a Bruker Avance 600 MHz spectrometer with a triple resonance CryoProbe. Data were processed with NMRpipe [28] and analyzed using NMRView [29].

**Results**

**Overexpression of recombinant CD147EC in E. coli**

The expression plasmid pET32a/CD147EC (contains the DNA fragment encoding residues 22–205 of CD147) was transformed into E. coli strain BL21(DE3) for protein production. The bacteria were grown in LB medium at 35°C. When the OD₆₀₀ value reached 0.6, IPTG was added at 100 mg/l to induce the protein expression. We compared protein expression levels for different incubation time after induction, and found that maximal expression was achieved...
at 8 h after induction. As shown on the SDS-PAGE gel [Fig. 1(A)], CD147EC is overexpressed in E. coli with a strong band at molecular weight of $\sim$20 kDa. However, over 90% of the expressed CD147EC protein is in the form of inclusion body [Fig. 1(D), lanes 1 and 2], and there is only a marginal amount of the expressed protein in soluble form ($< 2$ mg/l).

To produce isotopically labeled protein for NMR study, it is necessary to grow the bacteria in M9 minimal medium. We first grew the bacteria in LB medium until the OD$_{600}$ value reached 0.6, and the cells were collected and transferred into a moiety volume of M9 minimal medium for growth. After 1 h of incubation for recovery, protein expression was induced with 100 mg/l of IPTG. Longer incubation time after induction (16 h instead of 8 h) is required for bacteria grown in M9 minimal medium to achieve comparable protein yield as that in LB medium [Fig. 1(A), lanes 9 and 10].

As NMR study requires large quantity of soluble protein, efforts have been made to increase the yield of soluble form of CD147EC. Since lowering bacterial growth temperature is frequently found to be helpful with the yield of soluble protein [30,31], we first tried to grow the bacteria at lower temperatures (25 and 15°C). Unfortunately, no increase of soluble protein expression level was observed for CD147EC, and the total protein expression level was decreased at lower temperatures. We have also tried to express CD147EC with different fusion tags, such as GST and thioredoxin. However, none of these fusion tags led to higher yield of soluble CD147EC, and the recombinant fusion proteins were still mainly in inclusion bodies. As CD147EC has two disulfide bonds, the lack of soluble expression of CD147EC may account for the lack of oxidative environment in E. coli cytosol. Protein disulfide isomerase (PDI) is known to be able to catalyze the formation of protein disulfide bonds [32]. We thus tried to express the protein as a PDI fusion protein. Although most of PDI-CD147 fusion protein was expressed in soluble form, the total protein expression level was quite low, which resulted in a pure CD147EC protein yield of $<2$ mg/l after PDI was cleaved off the fusion protein.

Refolding and purification of CD147EC

We have developed a relatively simple and efficient method for refolding CD147EC using total cell lysate. This method is inspired from that used in the reconstitution of iron–sulfur cluster of ferredoxins [33]. First, bacterial cells were lysed by freeze–thawing followed by sonication, and then urea was dissolved in the cell lysate to a final concentration of 8 M. The mixture was degassed by vacuum and purged with argon gas until no more air bubble coming out under vacuum. Then, DTT was added to a final concentration of 10 mM to reduce all disulfide bonds, and the mixture was degassed again. Finally, the mixture was first diluted with eight-fold
degassed refolding buffer (50 mM Tris, 50 mM NaCl, and 1 mM DTT, pH 8.5), followed by two-fold undegassed refolding buffer, to allow the denatured protein to fold and the disulfide formation.

The refolded protein was first purified with DE52 anion-exchange column using a sodium chloride gradient for elution. CD147EC was in fraction 2 [Fig. 1(B)] of the elution profile as indicated by immunoblotting using HAb18 antibody, an anti-CD147 monoclonal antibody [34] [Fig. 1(D)], lanes 9 and 10. The protein was further purified by fast protein liquid chromatography using Superdex-75 SEC column, and a single peak corresponding to monomeric CD147EC was eluted from the column [Fig. 1(C)]. The purity of the purified protein was estimated to be over 95% by SDS-PAGE analysis. The refolded CD147EC run faster on the SDS-PAGE gel without DTT than that with DTT [Fig. 1(D), lanes 7 and 8], which is an indication for intra-molecular disulfide bond formation. The final yield of the purified protein was 25–30 mg/l culture (Table 1).

Table 1 Summary of the purified protein yield for CD147EC, N-CD147EC, and C-CD147EC refolded from inclusion body or total cell lysate

<table>
<thead>
<tr>
<th>Protein sample</th>
<th>Refolding material</th>
<th>Protein amount (mg)</th>
<th>Final amount of purified protein (mg)</th>
<th>Final yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD147EC</td>
<td>Inclusion body</td>
<td>30</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>CD147EC</td>
<td>Total cell lysate</td>
<td>30</td>
<td>26</td>
<td>86</td>
</tr>
<tr>
<td>N-CD147EC</td>
<td>Total cell lysate</td>
<td>70</td>
<td>51</td>
<td>73</td>
</tr>
<tr>
<td>C-CD147EC</td>
<td>Total cell lysate</td>
<td>28</td>
<td>21</td>
<td>75</td>
</tr>
</tbody>
</table>

*Results from an experiment with 1 l of bacterial culture.

Characterization of refolded CD147EC

To examine whether the refolded CD147EC adopts the right structure, sensitivity enhanced 2D 1H-15N heteronuclear single-quantum coherence (HSQC) spectrum of uniformly 15N-labeled sample is collected. The 2D 1H-15N HSQC experiment selectively detects NH pairs and correlates each 15N to the directly attached 1H. This spectrum is considered as an NMR ‘fingerprint’ of protein structure, because chemical shifts of the NH group reflect the unique environment encountered by each residue in a folded structure. As shown in Fig. 2(A), the 2D 1H-15N HSQC spectrum of the refolded CD147EC shows wide chemical shift dispersion characteristic of a folded protein, and the spectrum of the refolded CD147EC is indistinguishable from that of the solubly expressed protein in E. coli. In addition, the spectrum is almost identical to that of the previously reported CD147EC refolded with the 2-step dialysis protocol [27].

It has been reported that the treatment of stromal cells with the soluble CD147 protein can stimulate the expression of several MMPs [35]. To test whether the refolded CD147EC is capable of stimulating MMP expression in fibroblasts, fibroblast cells were treated with CD147EC for 24 h and the expression of MMP-9 was determined using quantitative real-time PCR. The result showed that the refolded CD147EC did significantly stimulate MMP-9 expression in a dose-dependent manner (Fig. 3). As a control, the unfolded CD147 EC could not stimulate expression of MMP-9.

Based on these results, we conclude that the refolded CD147EC using our rapid dilution method should adopt the native 3D structure.
Expression and refolding of individual Ig domains of CD147EC

The crystal structure revealed that the two Ig domains of CD147EC are relatively independent [26]. To simplify subsequent NMR studies, expression plasmids for the two Ig domains of CD147EC with C-terminal hexahistidine tag were constructed. The coding sequence for residues 22–103 (N-CD147EC) and residues 99–205 (C-CD147EC) of CD147 were cloned into the pET32a expression vector, and the resulting plasmids were transformed into BL21(DE3), respectively.

N-CD147EC was highly expressed in LB medium at 8 h after induction with 100 mg/l of IPTG at 35°C, and SDS-PAGE analysis showed that a major protein band of the expected 9.8 kDa size for N-CD147EC accounted for about 60% of the total cell protein (Fig. 4). It took longer induction time (15 h) to obtain the maximal production for C-CD147EC. As shown by SDS-PAGE gel (Fig. 4), C-CD147EC was overexpressed in E. coli with a strong band at molecular weight of ~12.8 kDa that accounted for about 23% of the total cell proteins. Both Ig domains expressed in E. coli were mainly in the insoluble fraction as CD147EC did. Using the same refolding method for CD147EC (as shown above), we were able to refold these two proteins successfully, each with one disulfide bond. The refolded protein was purified by Ni²⁺-NTA affinity chromatography, followed by Superdex-75 SEC column (Fig. 5). The purification summary is presented in Table 1.

Correct foldings for the two Ig domains were confirmed by NMR spectroscopy [Fig. 2(B,C)].

Using the same procedure for labeling CD147EC, we successfully labeled N-CD147EC in M9 minimal medium with ¹⁵NH₄Cl as sole nitrogen source, whereas the expression level of C-CD147EC in M9 minimal medium was very poor [Fig. 4(B), lane 3]. We thus compared the bacteria growth condition in LB and M9 media, and found that the pH of the culture changed significantly when bacteria harboring the expression plasmid of C-CD147EC were grown in LB medium. During the period before induction, the pH of culture decreased from 7.20 to 6.77 as bacteria growing. When the OD₆₀₀ value of the bacterial culture reached 0.6, 100 mg/l of IPTG was added to induce the protein expression. After 4 h, the pH of the culture was increased to 7.18, and protein expression could be detected by SDS-PAGE. Afterward, the pH of the culture was

---

**Figure 3** Refolded CD147EC stimulates MMP-9 expression in fibroblast cells

Fibroblast cells were treated with the indicated concentrations of refolded and unfolded CD147EC for 24 h and the fold change in the expression of MMP-9 determined by quantitative real-time PCR. Results from three independent experiments are shown. Error bars represent standard error. Analysis of variance was performed using analysis of variance and the statistical significance of the fold change was determined using Tukey’s procedure. Significant gene expression changes over controls are indicated by symbols: *P < 0.05 compared with unfolded group.

**Figure 4** Expression of N-CD147EC and C-CD147EC

(A) SDS-PAGE analysis of the expression of N-CD147EC in LB medium culture. Lanes 1–2, total lysates of cells grown in LB medium before induction and 8 h after induction, respectively. (B) SDS-PAGE analysis of the expression of C-CD147EC in LB and M9 medium culture. Lanes 1–3, total lysates of cells grown in LB medium 8 and 15 h after induction and grown in M9 medium 15 h after induction, respectively. Lane 4, total lysates of cells grown in M9 medium 12 h after induction, the pH of M9 medium was adjusted to 8.2 by adding NaOH at 4 h after induction and was maintained at this level afterwards.

**Figure 5** SDS-PAGE analysis of the purification of N-CD147EC and C-CD147EC

M, protein molecular weight marker; lane 1, refolding solution from total cell lysates for N-CD147EC; lane 2, flow through of Ni-NTA-agarose column for N-CD147EC; lane 3, purified N-CD147EC from Ni-NTA-agarose column; lane 4, refolding solution from total cell lysates for C-CD147EC; lane 5, flow through of Ni-NTA-agarose column for C-CD147EC; lane 6, purified C-CD147EC from Ni-NTA-agarose column.
increased continuously. After 15 h of induction, the pH of the LB medium reached 8.15, and the maximal C-CD147EC expression level was achieved for C-CD147EC (Fig. 6). However, when the bacteria were grown in M9 medium, the pH of the culture decreased continuously to 6.60 at 17 h after induction, with very little expression of C-CD147EC. We also compared cell density for bacteria grown in the LB and M9 media, and OD600 values of the bacterial culture were comparable between the LB and M9 media during the time. Therefore, we suspected that high pH may facilitate the expression of C-CD147EC. Accordingly, we adjusted the pH value of the culture to 8.2 by adding NaOH into M9 medium at 3.5 and 8 h after induction, and the pH was stabilized at 8.2 thereafter till harvest (15 h after induction). The result was very encouraging, and we were able to significantly increase the protein expression level using this strategy [Fig. 4(B), lane 4]. We also monitored the pH change during the bacterial growth for expressing N-CD147EC and CD147EC in LB and M9 medium. For these two proteins, the pH of bacterial culture was continuously increased until it reached eight in either LB or M9 medium, while protein expression levels were comparable for both media.

Discussion

Inclusion body, as it is well known, has both advantages and disadvantages [36,37]. The inclusion body is usually resistant to proteolytic degradation; often contains almost exclusively the overexpressed protein [38]. In this study, we presented a novel refolding protocol for obtaining soluble CD147EC with high yield, in which denatured protein in 8 M urea is first diluted with 8-fold degassed buffer in a reduced condition to allow protein refolding, and the formation of disulfide bonds is catalyzed by subsequently adding 2-fold undegassed buffer, which introduce molecular oxygen into the system. Compared with the previously described 2-step dialysis protocol which takes over 60 hours for the refolding procedure [27], our new method needs only a few hours and no additive (e.g. arginine) is required for preventing precipitation. Therefore, our new protocol is simple, fast and cost-efficient, and can be easily scaled up to prepare large amount of protein.

It is noteworthy that the refolding protocol presented here is characterized by the use of total bacterial lysate instead of pure inclusion bodies. It was frequently addressed that the inclusion body contaminants may interfere with the target protein refolding process and result in severe aggregation [36]. In our study, we have also tried to refold CD147EC through rapid dilution using purified inclusion body. However, it is found that the protein tends to aggregate and form white precipitation after dilution with refolding buffer, which results in an extremely low yield of soluble protein. On the contrary, refolding with total cell lysate did not show much precipitation [Fig. 1(D), lane 2]. It is possible that bacteria proteins in the refolding mixture may play some assistant roles in the CD147EC refolding. Therefore, the protein yield for refolding using total cell lysate is better than using pure inclusion bodies (Table 1). Refolding of ferredoxin with the total cell lysate was previously reported, and high refolding yield was also achieved [33].

Moreover, we believe that this method may be also applicable for refolding other proteins with disulfide bonds. At least, it provides another novel approach for refolding disulfide bond containing proteins from inclusion body, which is still largely dependent on trial-and-error experiments [39].

An interesting finding in this study is that the bacterial growth may cause different changes in the pH of culture between LB and M9 media, which results in different protein expression levels for C-CD147EC. In the LB medium, the pH goes higher to >8 as bacteria growing, and the expression level is high. However, in the M9 medium, the bacteria growth decreases the pH of the culture to as low as 6.6, which in turn cause poor protein production. This problem can be overcome by adjusting the pH of the M9 medium during the bacterial growth. This finding may be helpful for those who have trouble in producing protein in the M9 medium but not in the LB medium.

Acknowledgement

All NMR experiments were carried out at Beijing NMR Center, Peking University.

Funding

This work was supported by the grants from the National Natural Science Foundation of China (30900234 F. Song), and the Major State Basic Research Development Program
of China (973 Program; 2009CB521700 ZN Chen and B. Xia).

References


