Protein Expression and Purification 100 (2014) 10–18

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Expression and purification of soluble recombinant full length HIV-1 $Pr55^{Gag}$ protein in *Escherichia coli*

William J. McKinstry [a,1], Marcel Hijnen [b,c,1], Hanumant S. Tanwar [a,d], Lindsay G. Sparrow [a], Sureshbabu Nagarajan [d,e], Son T. Pham [d,e], Johnson Mak [b,c,d,e,*]

[a] CSIRO Materials Science and Engineering, Parkville, Victoria, Australia
[b] Centre for Virology, Burnet Institute, Melbourne, Australia
[c] Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia
[d] School of Medicine, Deakin University, Geelong, Australia
[e] CSIRO, Australian Animal Health Laboratory, Geelong, Australia

ARTICLE INFO

Article history:
Received 30 January 2014
and in revised form 23 April 2014
Available online 6 May 2014

Keywords:
HIV-1
Gag
$Pr55^{Gag}$
Recombinant protein expression
Purification

ABSTRACT

The HIV-1 Gag precursor protein, $Pr55^{Gag}$, is a multi-domain polyprotein that drives HIV-1 assembly. The morphological features of HIV-1 suggested $Pr55^{Gag}$ assumes a variety of different conformations during virion assembly and maturation, yet structural determination of HIV-1 $Pr55^{Gag}$ has not been possible due to an inability to express and to isolate large amounts of full-length recombinant $Pr55^{Gag}$ for biophysical and biochemical analyses. This challenge is further complicated by HIV-1 Gag's natural propensity to multimerize for the formation of viral particle (with ~2500 Gag molecules per virion), and this has led $Pr55^{Gag}$ to aggregate and be expressed as inclusion bodies in a number of *in vitro* protein expression systems. This study reported the production of a recombinant form of HIV-1 $Pr55^{Gag}$ using a bacterial heterologous expression system. Recombinant HIV-1 $Pr55^{Gag}$ was expressed with a C-terminal His×6 tag, and purified using a combination of immobilized metal affinity chromatography and size exclusion chromatography. This procedure resulted in the production of milligram quantities of high purity HIV-1 $Pr55^{Gag}$ that has a mobility that resembles a trimer in solution using size exclusion chromatography analysis. The high quantity and purity of the full length HIV Gag will be suitable for structural and functional studies to further understand the process of viral assembly, maturation and the development of inhibitors to interfere with the process.

Introduction

The HIV-1 Gag precursor protein, $Pr55^{Gag}$[2], is the major structural protein that drives the formation of HIV-1 particles. Both $Pr55^{Gag}$ of HIV-1 and retroviral Gag are critical as these virus assembly machineries are responsible: to select viral genomic RNA for packaging; to direct the protein to the inner plasma membrane for oligomerization; and to hijack host ESCRT proteins to facilitate virion particle egress [1,2], through different domains within the polyprotein. Owing to its importance in HIV-1 function, understanding the biochemical and biophysical property of HIV-1 $Pr55^{Gag}$ is paramount to HIV-1 biology. However, despite HIV-1 being discovered nearly 3 decades ago [3,4], large quantities of recombinant full-length $Pr55^{Gag}$ proteins are still unavailable for biochemical and biophysical analyses. Perhaps the only exception is work from Carlson and Hurley [5], in which, recombinant $Pr55^{Gag}$ was generated by fusing the C-terminus of HIV-1 protein with a 42 kDa expression/purification tag, maltose-binding protein (MBP). Alternatively, most laboratories prepare HIV-1 Gag protein by deleting the last 52 amino acids from the C-terminus ($p6^{Late}$ domain) to stabilize the expression and the purification of HIV-1 Gag [6,7]. It is likely that the C-terminus fusion of MBP has helped to stabilize HIV-1 $Pr55^{Gag}$ during expression and purification. The C-terminus $p6^{Late}$ domain is important for recruitment of host cell factors ESCRT [8,9], and recombinant Gag proteins lacking $p6^{Late}$ domain are not ideal for $Pr55^{Gag}$ characterization.

$Pr55^{Gag}$ relies on its poly-protein domains to facilitate particle formation and virion release. Briefly, several thousand copies of

* Corresponding author at: School of Medicine, Faculty of Health, Deakin University, Geelong, Victoria 3220, Australia. Tel.: +61 3 5227 5062; fax: +61 3 5227 5555.
E-mail address: johnson.mak@deakin.edu.au (J. Mak).

[1] These authors have contributed equally to this work.

[2] *Abbreviations used:* $Pr55^{Gag}$, HIV-1 Gag precursor; MA, matrix; CA, capsid; NC, nucleocapsid; p6, late domain; SP1, space peptide 1; SP2, spacer peptide 2; IPTG, isopropyl-β-D-thiogalactopyranoside; MALDI-TOF-MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; PBS, phosphate-buffered saline; SDS, sodium dodecyl sulfate; PAGE, poly acrylamide gel electrophoresis; TEM, transmission electron microscopy; VLP, virus-like particle.

http://dx.doi.org/10.1016/j.pep.2014.04.013
1046-5928/

the Gag poly-protein assemble at cholesterol- and sphingolipid-rich micro-domains on the plasma membrane of virus producing cells [10–14], where $Pr55^{Gag}$ organize themselves into a lattice-like structure and form spherical budding virus-like particles [15]. During or immediately after budding, $Pr55^{Gag}$ undergoes stepwise sequentially proteolytic processing by HIV protease, which resulted in a series of conformational rearrangement events to form mature infectious virus particle [16–19]. $Pr55^{Gag}$ consists of four major protein components (from N-terminus to C-terminus) known as matrix (MA, $p17^{MA}$), capsid (CA, $p24^{CA}$), nucleocapsid (NC, $p7^{NC}$), and the $p6^{Late\ Domain}$, while short linker peptides, SP1 and SP2 flank either side of the NC domain. Importantly, these different domains contribute to the oligomerization of the mature proteins and the precursor $Pr55^{Gag}$.

From the N-terminus of $Pr55^{Gag}$, MA is a highly basic protein domain (Gag_{1-132}) with a core molecular weight of ~17 kDa, and it exists in the context of $Pr55^{Gag}$ precursor, $p39^{MA-CA-SP1}$ intermediate and $p17^{MA}$ mature protein. The myristylation of Gly2 in MA targets the Gag protein to the plasma membrane [1,20], and mutation of this glycine myristylation site prevents virus budding and leads to accumulation of Gag within the host cell [21]. The structure of MA has been determined by both X-ray crystallography and NMR, and comprises 5 alpha helices and a three strand mixed beta sheet [22–24] which forms a trimer in the virion shell in the mature HIV-1 particle. It is presumed that the MA trimerizes in the context of $Pr55^{Gag}$ precursor protein during HIV-1 assembly [23].

The second major domain in $Pr55^{Gag}$ is CA (24 kDa), and it forms a fullerene shell surrounding the viral RNA genome and core associated proteins in mature viruses. CA is comprised of two subdomains, referred to as the NTD (N-terminal domain) and CTD (C-terminal domain) connected by a flexible linker. The NTD is comprised of seven α-helices and an N-terminal β-hairpin loop, with biochemical and genetic analyses demonstrating this domain is important for mature capsid assembly [25]. The CTD is roughly spherical, consisting of four α-helices and a 3_{10} helix, and forms dimers in solution [26]. The linker region joining both NTD and CTD is also important for capsid assembly and viral infectivity [27]. Within the mature viral capsid core, CA assembles into both hexamer and pentamer structures [28,29], and recent cryo-electron microscopy identified cone-shaped structures comprising 12 pentamers with either 216 or 186 hexamers [30], which is in agreement with previous estimation/model of the structure of the fullerene cone capsid core [31,32]. While it is almost certain that the arrangement of CA–CA interactions during HIV assembly are different from the CA–CA interactions in the mature virion core, yet our current understanding of CA–CA structure during assembly is extremely limited.

The third domain, nucleocapsid (NC) is a highly basic protein (~7 kDa) and contains two CCHC-type zinc finger motifs [33]. The zing finger motifs enable Gag to recruit genomic RNA molecules to newly forming virions, and to assist in Gag–Gag interactions during virus-like particle formation [34,35,19]. The ability of NC to interact with genomic RNA is critical for the oligomerization and eventual formation of virus like particles *in vitro*. However, the precise mechanism on how NC mediated this process remains unexamined due to the lack of full length $Pr55^{Gag}$ for biochemical and biophysical analysis.

The C-terminal domain, $p6^{Late}$ contains N- and C-terminal helical domains connected by a flexible hinge region [36], containing two "late domain" motifs (PTAP and $LYPX_nL$) that bind and recruit host cell ESCRT proteins TSG101 and Alix that are essential for virus budding, and the LXSLFG motif that is required for the incorporation of the viral accessory protein Vpr into virus particles [37–39]. Despite the identification of these important motifs and amino acids within HIV $p6^{Late}$ domain, in the absence of soluble $Pr55^{Gag}$, it is currently unknown how any of these interactions would affect the folding $Pr55^{Gag}$ and the dynamics of Gag oligomerization during HIV assembly.

Previously, researchers had expressed individual MA, CA, and NC domains and intermediate forms of Gag, including mutations that result in the formation of stable monomers in *Escherichia coli* for structural and functional studies [15,6,40,41,7]. This is the first reported use of a bacterial heterologous protein expression system to produce soluble recombinant wild-type, full length, HIV-1 $Pr55^{Gag}$ protein, without the need for solubility tags that can be used for a large array of biochemical and biophysical assays to dissect the mechanisms of HIV assembly.

Materials and methods

Bacterial strains and media

All transformation steps were carried out using the *E. coli* strains DH5α and BL21(DE3)pLysS using standard heat shock of $RbCl_2$ competent cells. For the production of plasmid DNA in DH5α cells and the production of recombinant protein in BL21(DE3)pLysS, cells were cultured in LB media [1% (w/v) peptone, 0.5% (w/v) yeast extract, and 0.5% NaCl]. The media was supplemented with the antibiotics kanamycin (50 μg/mL; DH5α) and both kanamycin and chloramphenicol (35 μg/mL; BL21(DE3)pLysS).

Plasmid construction

The HIV molecular clone pNL4.3 [42] served as template DNA for amplifying the full length wild-type HIV-1 $Pr55^{Gag}$ gene. The target gene was amplified using oligonucleotide primers designed according to the full length gene sequence of HIV-1 $Pr55^{Gag}$ and the restriction sites for *NcoI* and *XhoI* were included at the ends of the primers for ligation dependent cloning into a modified pET28a protein expression vector. The forward primers were 5′-AACCATGGGTGCGAGAGCGTCGGTA-3′ and the reverse primer was 5′-AAACTCGAGTTGTGACGAGGGGTCGCT-3′, with the underline regions representing the recognitions sites for the restriction endonucleases *NcoI* and *XhoI*, respectively. This vector is a derivative of the pET28a in which the N-terminal His×6 tag and protease cleavage site were removed and replaced with a C-terminal His×6 tag to facilitate protein purification using IMAC. The total PCR reaction volume was 100 μL including 10 μL of the template DNA, and 2 μL of Vent DNA polymerase (New England Biolabs). PCR conditions were as follows: initial denaturation 95 °C for 5 min, then 35 cycles of denaturation at 95 °C for 45 s, primer annealing at 58 °C for 45 s, followed by primer extension at 72 °C for 90 s. A final extension step proceeded at 72 °C for 7 min. The size of PCR product was confirmed by electrophoresis on a 1.2% agarose gel, treated with SYBRSafe DNA stain and visualized by UV illumination. The remaining amplified DNA was column purified (Qiagen). The purified HIV-1 $Pr55^{Gag}$ coding DNA sequence was excised with *NcoI* and *XhoI* at 37 °C for 60 min, column purified and then ligated using T4 DNA ligase into the linearized modified pET28a vector, that have previously been cleaved with *NcoI* and *XhoI*. The $Pr55^{Gag}$–pET28a expression plasmid was verified by DNA sequencing.

Analytical protein expression studies

Small scale recombinant protein expression studies were performed by inoculating a single colony into 5 mL LB media containing antibiotics overnight at 37 °C, with shaking at 160 rpm. The overnight culture was used to inoculate 5 mL fresh LB media containing antibiotics at a starting $OD_{600\ nm}$ 0.1 and the culture was grown at 37 °C, 160 rpm, until an $OD_{600\ nm}$ of 0.8 was reached.