Protein Expression and Purification 145 (2018) 25–31

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Strategizing for the purification of a multiple Big domain-containing protein in native conformation is worth it!

Asmita D. Pawar[a,*], Deepshikha Verma[b], Venu Sankeshi[a], Rajeev Raman[a], Yogendra Sharma[a,c]

[a] CSIR-Centre for Cellular and Molecular Biology (CCMB), Uppal Road, Hyderabad 500 007, India
[b] Department of Chemical Sciences, Tata Institute of Fundamental Research, 1, Homi Bhabha Road, Colaba, Mumbai, 400005, India
[c] Academy of Scientific and Innovative Research (AcSIR), New Delhi, India

ARTICLE INFO

Keywords:
Arig
Big_2 domain
Multidomain
In vitro refolding
Soluble expression
Sorbitol
Sucrose
Native conformation
2D [^{1}H-^{15}N]-HSQC

ABSTRACT

The reliability and accuracy of conformational or functional studies of any novel multidomain protein rely on the quality of protein. The bottleneck in structural studies with the complete Big_2 domain containing proteins like LigA, LigB or *Mp*IBP is usually their large molecular size owing to their multidomain (> 10–12 domains) architectures. Interestingly, a soil bacterium *Paenarthrobacter aurescens* TC1, harbours a gene that encodes a protein comprising of four predicted Big_2 domains. We report here the expression and purification of this novel, multiple Big_2 domains containing protein, Arig of *P. aurescens* TC1. During overexpression, recombinant Arig formed inclusion bodies and hence was purified by on-column refolding. The refolded Arig revealed a β-sheet conformation and a well-resolved near-UV CD spectra but did not exhibit a well-dispersed 2D [^{1}H-^{15}N]-HSQC NMR spectrum, as expected for a well-folded β-sheet native conformation. We, therefore, further optimized Arig overexpression in the soluble fraction by including osmolytes. CD spectroscopic and 2D [^{1}H-^{15}N]-HSQC analyses consolidate that Arig purified alternatively has a well-folded native conformation. While we describe different strategies for purification of Arig, we also present the spectral properties of this novel all-β-sheet protein.

1. Introduction

The quality of a protein purified following its overexpression in the bacterial system is primarily related to whether it is purified from the soluble fraction or from inclusion bodies. If a protein is purified from the soluble fraction, it usually exists in its native and active conformation [1–3]. When coming from inclusion bodies, its dissolution in urea or other denaturants, and subsequent in vitro refolding is needed to obtain a protein in as good as its native form [4–6]. This, however, may not always lead to its native conformation in our experience.

We present here a case of a multidomain protein belonging to Big_2 superfamily named Arig (**Ar**throbacter **i**mmuno**g**lobulin-like) [7] where we demonstrate a comparison of the purification strategies and spectral features of in vitro refolded Arig purified from inclusion bodies and the native conformation protein. Arig is annotated as hypothetical surface protein in the genome of the bacterium *P. aurescens* [7,8] and is encoded by the gene (GenBank accession ABM07665.1; locus AAur_3620) as per the recent NCBI database (www.ncbi.nih.gov) (Fig. 1A). The Big domain-containing proteins are generally hypothetical surface proteins prevalent in bacteria. Only a few of them have been structurally and functionally characterized like those in the adhesins viz. intimin [9,10], invasin [11,12] or ice adhesin RII domains of *Mp*IBP [13–15]. The four domains of Arig share a striking ~70–80% sequence homology to the adhesins LigA and LigB of *Leptospira interrogans* serovar Pomona that belong to a less characterized Big_2 group within the Big superfamily [16–18]. Interestingly, Big_2 domains of LigA and LigB have been implicated in host pathogenesis via their Ca^{2+}-dependent association with host extracellular matrix proteins like fibronectin, fibrinogen, elastin or tropoelastin [19–22]. Diagnostic ELISAs and vaccines for leptospirosis have been developed using LigA and LigB domains [23–27] signifying why these potent virulence factors need to be studied at the molecular level.

The solution and crystal structures of a few individual Big domains have demonstrated their β-sheet sandwich structures [14,28,29]. However, the structure of an entire multidomain assembly is yet to be demonstrated. The complete Big_2 multidomain proteins are too large proteins to be recombinantly expressed for spectroscopic or functional characterization. The cloned Arig is relatively small in size with only four predicted Big_2 domains (Fig. 1A and B). We, therefore, sought to express Arig in well-folded conformation to investigate its structural features reliably. During initial attempts of overexpression, the protein aggregated into inclusion bodies and hence was purified by on-column

Abbreviations: Big, Bacterial immunoglobulin-like; HSQC, Heteronuclear single quantum coherence; HEPES, hydroxyethyl piperazine ethane sulfonic acid; CD, circular dichroism

* Corresponding author.
E-mail address: pawarasmita@ccmb.res.in (A.D. Pawar).

https://doi.org/10.1016/j.pep.2017.12.008
Received 21 August 2017; Received in revised form 15 December 2017; Accepted 21 December 2017
Available online 27 December 2017
1046-5928/

(A)

```
          10         20         30         40         50         60
MRDALRQRLL AEASLHARTH PADLGLALAE GMSTVRRRII RAWVVAAVVS VALMAAFVAA
          70         80         90        100        110        120
VFAAPDDEGG PGPVTLQSIT VTASSQLLRP DTKLDLTARG TYSNDETKEL VDGLTWTSDK
         130        140        150        160        170        180
PEVAAVSPIG EVTAATAGTV VITAMHSDVK GMLNLTVTDG PGLLTGLRIK PDQTTVSLGG
         190        200        210        220        230        240
TTQLIAEGTF SDNSIGNRNE PAIWESRNPG IATVDGLGMV TGMNPGVATI SATEGTIQAT
         250        260        270        280        290        300
AAITVTDPGT PNPTVVPNGC VVDPATLRVK HTQTQPLTAF ITATDGTRVP PTTVKWSSTD
         310        320        330        340        350        360
SKIANVNTSG SVTGTGPGST RITALCVDAN GKEWQATASV TVEPVARSII IGPAGPHRLT
         370        380        390        400        410        420
PQQSIQLAAT VTFSDGSTST KAAVSWNASS SPPLVVIVSR SGVVTGVRNG RANVTASLDG
         430        440 441
ATSNPITVDV AITFPDEGEI G
```

(B)

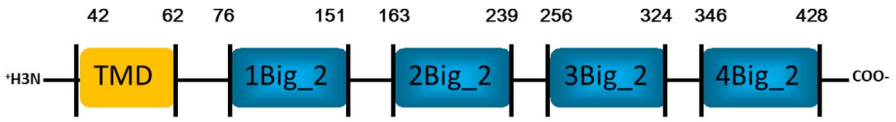

Fig. 1. (A) The annotated amino acid sequence corresponding the translated gene, locus AAur_3620 of *Paenarthrobacter aurescens* TC1. The sequence of Arig, which was cloned into pET21b is underlined, and the predicted four Big_2 domains are highlighted in blue. The N-terminal 73 residues containing the 20-amino acid residue transmembrane domain segment were excluded. **(B)** Schematic representation of the organization of predicted Big_2 domains of Arig, a 38.5 kDa protein along with the excluded N-terminal transmembrane domain. Numbers denote the first and the last residues of each predicted domain. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

refolding. When we did not get well-dispersed cross peaks on recording $^{1}H,^{15}N$-HSQC spectra for in vitro refolded Arig, we optimized further to overexpress the protein in the soluble fraction. We then compared the spectral properties of Arig purified by refolding from inclusion bodies with that of the natively purified protein. This study lays emphasis on overexpressing and purifying a multiple Big domain-containing bacterial protein from the soluble fraction.

2. Methods

2.1. Genomic DNA and cloning

P. aurescens TC1 (earlier *Arthrobacter aurescens* TC1) [8,30] was a kind gift from Dr. Michael Sadowsky, University of Minnesota, USA. Genomic DNA of *P. aurescens* TC1 was extracted by classical chloroform–phenol method [31] from the bacterium cultured in R2A (HiMedia, India) broth. The gene locus (AAur_3620) that codes for the hypothetical surface protein (GenBank accession no. ABM07665.1) without the N-terminal 73-residue coding segment was amplified by PCR using the forward primer with *Nde*I restriction site 5′GCCCCATATGGTGACTCTGCAGTCGATC3′ and reverse primer with *Xho*I restriction site 5′AAAACTCGAGGCCGATCTCGCCCTCGTCGG3'. The 1104 bp amplicon was gel extracted and cloned into *Nde*I and *Xho*I sites in a pET21b bacterial expression vector. Competent *Escherichia coli* DH5α was transformed with the recombinant plasmid and plated onto Luria-Bertani agar with 100 μg/mL ampicillin (LB_{amp}). The recombinant plasmid was prepared using plasmid miniprep kit (Machery-Nagel, Germany) and the cloned gene named '*arig*' was confirmed by sequencing using the plasmid T7 promoter and T7 terminator specific primers.

2.2. Arig overexpression and purification

a) From inclusion bodies

The recombinant pET21b plasmid with *arig* gene was transformed into *E. coli* BL21 (DE3) cells. Arig was overexpressed as a C-terminal 6X His-tagged protein within transformed *E. coli* BL21 (DE3) when grown in LB_{amp} broth. Expression cultures were incubated at 18 °C for 14 h and 37 °C for 4–6 h after varying 0.6–1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) inductions. Arig was solubilized from inclusion bodies after ultrasonic cell disruption in buffer containing 50 mM Tris, pH 8, 100 mM KCl (buffer A) and 4 M urea as denaturant as per the protocol of vendor (Qiagen, Hilden, Germany). Phenylmethanesulfonyl fluoride was added for preventing proteolytic degradation of overexpressed protein and DNase addition ensured DNA-free protein. The sonicated cell lysate was centrifuged at 18,000 rpm for 50 min for cell-free lysate to be loaded on Ni-NTA affinity column. The Ni-NTA bound Arig was then washed with buffer A (without urea) for spontaneous on-column refolding. The on-column refolded protein was eluted in buffer A with 200 mM imidazole. All buffers were freshly prepared, filtered and chilled prior to use. The protein in the eluates was confirmed by SDS-PAGE. Similarly, ^{15}N-labeled Arig was also overexpressed in inclusion bodies by *E. coli* BL21 (DE3), in M9 minimal medium supplemented with 1% ^{15}N-NH_4Cl. ^{15}N-labeled Arig was also purified by the above on-column refolding strategy.

b) Altered strategy for soluble overexpression and purification

Unlabeled and ^{15}N-labeled Arig was overexpressed in the soluble fraction adopting a simple known modification of the use of osmolytes at the bacterial culture stage [32,33]. *E. coli* BL21(DE3) transformed with *arig* were grown in Luria-Bertani broth and M9 minimal medium supplemented with ^{15}N-NH_4Cl at 37 °C till an O.D of 0.6 is reached. Thereafter, the culture was supplemented with 0.8% sucrose and 0.8% sorbitol of the total culture volume [34]. After an acclimatization time of 10–15 min, the cells were induced with 0.8 mM IPTG for overexpressing the recombinant 6X His-tagged Arig. Cells were resuspended in buffer containing 50 mM KH_2PO_4, pH 7.5, 300 mM KCl (buffer B), lysed by ultrasonication at 4 °C and centrifuged at 15,000 rpm for 50 min. The supernatant with Arig was loaded on a Ni-NTA column for affinity binding and washed with a buffer B containing 10 mM imidazole. Ni-NTA resin bound Arig was eluted in buffer B with 200 mM imidazole. The purity of eluted Arig was judged by SDS-PAGE.