Protein Expression and Purification 104 (2014) 65–70

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Soluble full-length expression and characterization of snRNP protein U1-68/70 K

Carolin Richter [a,c], Thomas Simon [a], Iris Asen [a], Gerald Brenner-Weiss [b], Jürgen Hubbuch [c,*]

[a] DIARECT AG, Freiburg, Germany
[b] Karlsruhe Institute of Technology, Campus Nord, Institute of Functional Interfaces, Analytical Biochemistry, Eggenstein-Leopoldshafen, Germany
[c] Karlsruhe Institute of Technology, Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte-Ring 1, 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 19 February 2014
and in revised form 9 August 2014
Available online 6 September 2014

Keywords:
Cell-free expression
Autoimmunity
snRNP complex
Chaperones

ABSTRACT

The autoantigen U1-68/70 K is the dominant diagnostic marker in Mixed Connective Tissue Disease (MCTD) that until recently could not be expressed in its full-length form (Northemann et al., 1995, [16]). Using cell-free expression screening, we successfully produced the snRNP protein U1-68/70 K in a soluble full-length form in *Escherichia coli* cells. The protein length and identity was determined by Western Blot and MS/MS analysis. Additionally, its reactivity in the autoimmune diagnostic was confirmed. Establishment of a cell-free expression system for this protein was important for further elucidation of protein expression properties such as the cDNA construct, expression temperature and folding properties; these parameters can now be determined in a fast and resource-conserving manner.

Introduction

The U1-68/70 K protein, a component of the nuclear spliceosomal U1-snRNP particle [1], is a major autoantigen in autoimmune diseases such as Mixed Connective Tissue Disease (MCTD)[1] or Systemic Lupus Erythematodes (SLE; [2,3]). Biotechnological production of U1-68/70 K as a diagnostic autoantigen for detection of autoantibodies in patient sera has so far been difficult, most likely due to the sequence and structural peculiarities of the protein.

U1-68/70 K (total length 437 amino acids) has a long repetitive sequence between amino acids 231 and 393 with a very large proportion of basic (41% arginine-residues) and acidic (30%) amino acids [4,5]. Crystal structure analysis of the U1-snRNP particle [6] reveals that the first 60 amino acids of U1-68/70 K wrap around the common heptameric Sm snRNP-core in an unusual extended conformation devoid of regular secondary structure. A helical section (amino acids 61–89) and an RNA binding RRM-domain (amino acids 100–180) contact the U1-RNA component of the U1-snRNP. The crystal structure analysis [6] however does not include the C-terminal half of U1-68/70 K, which is thought to be unstructured and to provide binding sites for numerous constitutive and alternative splicing factors [7,8] and possibly for RNA in protamine-like fashion [4]; these activities may be regulated by serine phosphorylation [9].

Screening of a large panel of MCTD patient sera identified four major continuous domains within the human U1-68/70 K as autoantibody targets, referred to as regions A′, B′, C′ and D′ [10–13]. Recombinant *Escherichia coli*-based production of U1-68/70 K has been possible only for fragments with the antigenic epitopes, but not for the full-length U1-68/70 K protein [14,15]. Northemann et al. [16] detected an inhibitory element within the full-length sequence of U1-68/70 K (sequence X) and proposed that this element interferes with translation. Expression of a protein containing the inhibitory sequence X could inhibit trans-actively the synthesis of other *E. coli* proteins indicating that full-length expression of U1-68/70 K is impossible [16]. Experiments with various deletions of the inhibitory sequence have shown that the number of deletions correlates with the expression level of the truncated U1-68/70 K protein [TS, unpublished data]. Interestingly, the inhibitory element corresponds to part of the charged arginine-rich unstructured region of U1-68/70 K.

To shed some light into this, several expression strategies have been evaluated so far [16].

In prior work the baculovirus-based expression of the U1-68/70 K protein in Sf9 insect cells was evaluated. Despite the transient nature of this expression system, recombinant expression of U1-68/70 K is difficult to verify and is nowhere near the levels of overexpression seen for other intracellular proteins (including other

* Corresponding author. Tel.: +49 721 608 42557; fax: +49 721 608 46240.
E-mail addresses: Carolin.Richter@kit.edu (C. Richter), Thomas.Simon@diarect.com (T. Simon), Iris.Asen@diarect.com (I. Asen), Gerald.Brenner-Weiss@kit.edu (G. Brenner-Weiss), Juergen.Hubbuch@kit.edu (J. Hubbuch).

[1] *Abbreviations used:* MCTD, Mixed Connective Tissue Disease; SLE, Systemic Lupus Erythematodes; PBS, phosphate buffered saline; SDS, sodium dodecyl sulfate; MS/MS, matrix-assisted laser desorption/ionization-time-of-flight.

http://dx.doi.org/10.1016/j.pep.2014.08.009
1046-5928/

spliceosome components) in the baculovirus system. Explanations for this expression behavior are not available, but the observations underline the "difficult" nature of this protein.

In recent years, *in vitro* translation has become an important tool for rapid and cost reduced screening of different protein expression conditions. Combining *in vitro* expression with high throughput parameter optimization [17] followed by transfer to cell-based expression is a valid approach for folded and soluble difficult-to-express proteins [18–20]. Variables such as the comparison of eukaryotic or prokaryotic systems, requirements of auxiliary factors (e.g. the addition of chaperones, detergents or cofactors) or protein- and process-specific ones (e.g. temperature, time) can be estimated [21]. However, the full-length expression of the U1-68/70 K autoantigen has not been reported in literature so far.

In this study, we evaluated the potential of an *E. coli* cell-free translation system for producing soluble full-length human U1-68/70 K protein. Furthermore it was elucidated if data from cell-free expression provide useful informations regarding optimizing cell-based expression strategies. Several parameters, which were determined via the *E. coli in vitro* expression system, could be successfully transferred to a cell-based approach. The resulting expression strategy allows now for the first time expression of a soluble, full-length and immunologically active U1-68/70 K autoantigen in *E. coli* cells. This result enables further characterization of the human U1-68/70 K full-length protein in either structural or functional studies.

Materials and methods

DNA template generation: cloning and codon optimization

Work was carried out with the alternatively spliced shorter U1-68 K isoform of human U1-68/70 K (UniProt ID: P08621 – isoform 2). Full-length as well as the truncated (lacking the 66 AA inhibitory sequence X – Fig. 1) cDNA constructs were cloned into different vector systems:

- pET24d (Merck Millipore, Germany) for *E. coli* cell-free and cell-based expression.
- pCDF-Duet-1 (Merck Millipore, Germany) for co-expression with pET24d-based chaperone constructs in *E. coli* cells.

All vectors included C-terminal hexahistidine tags, whereas expression vectors of the original human cDNA sequence additionally included N-terminal hexahistidine tags (see Supplementary material). All constructs were verified by resequencing (Solvias AG, Switzerland). In addition to constructs with the original human cDNA sequence, full-length and truncated U1-68/70 K cDNAs were codon-optimized for *E. coli* expression. Web-based bioinformatic tools were: codon usage (http://www.entelechon.com); GC content (http://www.bioinformatics.org); RNA secondary structure (www.genebee.msu.su). Gene synthesis was done by Entelechon GmbH (Germany).

Coding regions for the chaperones dnaK, dnaJ and grpE were PCR-amplified from *E. coli* BL21 and cloned into pET24d. For chaperone co-expression, ribosome binding site/chaperone cassettes were combined in a single pET24d construct, with expression of a polycistronic mRNA driven from a single T7 promoter.

Preparation of bacterial cell-free extract

For cell-free expression the RNaseE-mutant *E. coli* BL21 Star™ (DE3) strain (Life Technologies GmbH, Germany) was transformed with pRARE2 (isolated from *E. coli* Rosetta2, Merck Millipore, Germany). Bacterial cell-free extract was prepared according to Kim et al. [22] using the simplified procedure (S12) including slight modifications.

In vitro expression method

E. coli cell-free expression was carried out either in 50 µl (analytic) or 500 µl (preparative) reaction volumes. The cell-free reaction mixture (Kim et al. [23] with slight modifications) was adapted to the S12 extract. To the cell-free reactions 6% of DnaK supplement (5 Prime, Germany) was added. Expression temperature was set to a value between 15 and 37 °C and reactions were incubated overnight (~12–14 h) in a thermo mixer at 300 rpm. Negative controls excluded plasmid DNA and were performed for all cell-free expression methods; control background was analyzed in parallel to the product-containing reactions.

In vivo expression method

BL21 Star™ (DE3) bacteria transformed with the respective cDNA expression constructs were grown as overnight pre-cultures in MDG media at 37 °C. Expression cultures were grown with PepYMD-505 (Studier's ZYM-505 medium with added aspartic acid and NZ-Amine replaced by peptone; [24]). At a bacterial density of OD_{600nm} = 0.6, cultures were induced by addition of 1 mM IPTG. For protein expression, bacterial cultures were incubated overnight at 25 °C at 300 rpm in an incubation shaker (Infors, Switzerland).

Protein analysis

Overnight *E. coli* cell-free reactions were centrifuged (16,000×*g*; 5 min), the pellet was washed once with phosphate buffered saline (PBS) (1.5 M NaCl, 37 mM NaH_2PO_4 and 163 mM Na_2HPO_4) and solubilized in 1% sodium dodecyl sulfate (SDS). *E. coli* cell lysates from *E. coli* cell-based expression cultures were prepared by lysozyme treatment, addition of Triton X-100 to 1% final concentration and freeze-thaw steps. Lysates were centrifuged (16,000×*g*; 5 min) for fractionation of soluble vs. insoluble components. Identical volumes of reducing SDS sample buffer were added to cell-free and cell-based samples. All samples were denatured for 5 min at

```
                                                    Inhibitory sequence X [16]
                                250        260        270        280        290
U1-68/70 K            KERERRRSRS RDRRRRSRSR DKEERRRSRE RSKDKDRDRK RRSSRSRERA
U1-68/70 K_truncated  KERERRRSRS RDRRRRSRSR ---------- ---------- ----------
                     ********** **********
                                300        310        320        330        340
U1-68/70 K            RRERERKEEL RGGGGDMAEP SEAGDAPPDD GPPGELGPDG PDGPEEKGRD
U1-68/70 K_truncated  ---------- ---------- ---------- ----- GPDG PDGPEEKGRD
                                                          **** **********
```

Fig. 1. Amino acid sequence comparison of the full-length and truncated U1-68/70 K at the inhibitory sequence X region.