Post-print of: Unzueta, Ugutz, et al. "Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A- Electronic supplementary material", in Applied Microbiology and Biotechnology, vol. 99, issue 14 (july 2015), p. 5863-5874. The final publication is available at Springer via DOI 10.1007/s00253-014-6328-9

**Applied Microbiology and Biotechnology**

**Electronic Supplementary Material for:**

# Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A.

Ugutz Unzueta[a,b,c], Felicitas Vázquez[a,c],*, Giulia Accardi[a,b,§], Rosa Mendoza[a,c], Verónica Toledo-Rubio[a,c,ǂ], Maria Giuliani[d,‡], Filomena Sannino[d,e], Ermenegilda Parrilli[d], Ibane Abasolo[c,f], Simo Schwartz, Jr.[c,f], Maria L. Tutino[d], Antonio Villaverde[a,b,c], José L. Corchero[c,a,b] and Neus Ferrer-Miralles[a,b,c,#]

Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain[a], Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain[b], CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain[c], Dipartimento di Scienze Biochimiche, Facoltà di Medicina e Chirurgia, Università degli studi di Palermo, 90127 Palermo, Italy[d], Institute of Protein Biochemistry, CNR, Via Pietro Castellino 111, I-80131 Naples, Italy[e], VHIR Vall d'Hebron Inst Recerca, CIBBIM Nanomed, Barcelona 08035, Spain[f]

*Present Adress: GREENALTECH S.L., Barcelona 08028, Spain.

§Present Adress: Department of Pathobiology and Medical and Forensic Biotechnology, University of Palermo, 90134 Palermo, Italy.

ǂPresent Adress: BioIngenium S.L., Barcelona 08028, Spain.

‡ Present address: Novartis Vaccines and Diagnostics, via Fiorentina, 1 53100 Siena, Italy.

#Adress correspondence to:

Neus Ferrer-Miralles, neus.ferrer@uab.cat, Tel: +34 93 581 2148, FAX: +34 93 581 2011

## Materials and Methods

### Plasmids

In addition to the four expression vectors described in the main body of the article, four expression vectors were used in this work for expression in *E. coli* strains (Suppl. Figure S1a) described as follows: a) pET22b-GLA constructed by the insertion of a *Nde*I-*Eco*RI DNA fragment encoding the mature form of the *h*GLA with a His-tag and TEV protease cleavage site in pET22b; b) pDAss-GLA, constructed by the insertion of a *Bam*HI-*Hin*dIII DNA fragment encoding the mature form of the *h*GLA protein with a His-tag and TEV protease cleavage site in parental vector pDAss. pDAss is a pTrc99-derived expression vector containing the leader peptide of disulfide isomerase I protein (DsbA) for periplasmic secretion and was kindly provided by Juan-Miguel Puertas and Jean-Michel Betton from the Unité de Microbiologie Structurale of the Institut Pasteur, Paris; c) pGEX4T2-GLAmut1, derived from pGEX4T2-GLA, was constructed by site-directed mutagenesis to optimise codon usage of R392 and R402 of *h*GLA ORF using Quikchange Lightning site-directed mutagenesis kit (Stratagene 210518) (Figure S1b). R392 and R402 codons were modified in two consecutive mutagenesis steps. Wild type R392 and R402 AGG codon were mutated to CGC and CGT codons respectively; d) pGEX4T2-GLAmut2, modified *h*GLA gene from pGEX4T2-GLAmut1 was subjected to a third mutagenesis cycle to eliminate a putative *E. coli* transcriptional terminator (Nudler and Gottesman, 2002) since analysis of sequence propensity in the formation of secondary structure using Vector NTI suite 9 (informax, Invitrogen) revealed that it was energetically favorable (ΔG=-2.7Kcal/mol) (Suppl. Figure. S1b).

### Periplasm fractionation in *E. coli* cell cultures

Periplasm fractionation in pDAss-GLA transformed *E. coli* cells was obtained by osmotic stress. Treated cells were centrifuged and periplasmic cell fraction was recovered from the supernatant leaving concentrated spheroplasts in the pellet (Miot and Betton, 2007).

## Results

### Recombinant expression of human *GLA* ORF in *E. coli*

Suppl. Figure S2 shows the detection of the *h*GLA produced by *E. coli* BL21(DE3) strain transformed with pReceiver-B01-GLA. As it can be seen, the protein was produced under all the assayed conditions although in the insoluble cell fraction only. However, when compared with the molecular weight markers, it was observed that the molecular weight of the obtained product seemed slightly lower than expected (theoretical molecular mass of the recombinant *h*GLA≈ 49.6 kDa). As seen in Suppl. Figure 1a, this recombinant *h*GLA contains an N-terminal His-tag that can be detected in the insoluble cell fraction of induced cultures indicating that the recombinant *h*GLA is probably C-terminally proteolysed. In addition, an anti-GLA antibody raised against the last 100 amino acids of GLA (Santa Cruz Biotechnology, data not shown) could also detect *h*GLA in the insoluble cell fraction indicating that less than 100 amino acids have been proteolysed at the C-terminus. Other positive protein bands were also detected which could correspond to either aggregated GLA products or proteolysed forms of the His-tagged *h*GLA.

**Recombinant production of mature form of *h*GLA in *E. coli***

The complete ORF of *h*GLA includes its secretion signal, it has been observed that the eukaryotic signal peptides are not processed in prokaryotic expression systems, therefore, the secretion signal included in *h*GLA gene was eliminated and the resulting gene was fused to a N-terminal His-tag and Tobacco Etch Virus (TEV) protease cleavage site and used in expression experiments either to be expressed in the cytoplasm of transformed cells or sent to the periplasm by adding specific secretion signals (pET22b-GLA and pDAss-GLA respectively in Suppl. Figure S1a).

Rosetta-gami B(DE3) cells were transformed with pET22b-GLA vector and used to produce a N-terminally his-tagged mature form of *h*GLA in shake cultures induced at either 20ºC or 37ºC at different IPTG concentrations (Suppl. Figure S3). The Rosetta-gami B(DE3) strain is able to compensate low concentrations of certain tRNAs of *E. coli* and improve recombinant protein translation in addition to providing an oxidizing cytoplasm in which protein disulphide bridges can be formed. However, the theoretical 47.8 kDa protein band corresponding to the his-tagged *h*GLA was not detected. In fact, only a specific 25 kDa protein band was detected when using anti-GLA antibodies in cultures induced at 20ºC in both soluble and insoluble fractions regardless of the IPTG concentration, a protein band which can be barely detected when induction of gene expression was performed at 37ºC. When samples were developed with the monoclonal anti His-tag antibody, no bands were detected suggesting an N-terminal proteolysis (data not shown). The mature form of *h*GLA in the Rosetta-gami B(DE3) strain seemed to be completely translated since it could be detected by antibodies raised against the C-terminal end of the protein but it seems that is degraded to a proteolytically reluctant smaller polypeptide of 25 kDa.

Protein expression experiments were carried out with transformed Rosetta 2(DE3) cells with pDass-GLA to test the presence of *h*GLA in the periplasmic fraction of the cells (Suppl. Figure 4). Specific bands were detected in all the assayed conditions although aggregated forms were also observed in positive lanes. A band of approximately 40 kDa appeared in all lanes which corresponded to a proteolysed product as observed in the preceding experiments. Correspondingly, a protein band displaying slightly lower molecular weight was detected exclusively in periplasmic samples as the result of the periplasmic secretion signal peptide processing to get into the periplasmic space.

Samples were also tested against a monoclonal anti-His antibody revealing the presence of macromolecular complexes and a specific band of 42 kDa only present in the insoluble cellular fraction, indicating that the N-terminal end in both soluble and periplasm samples had been removed. Therefore, the mature form of the *h*GLA could be translated and sent to the periplasm although it is degraded at least N-terminally.

**References**

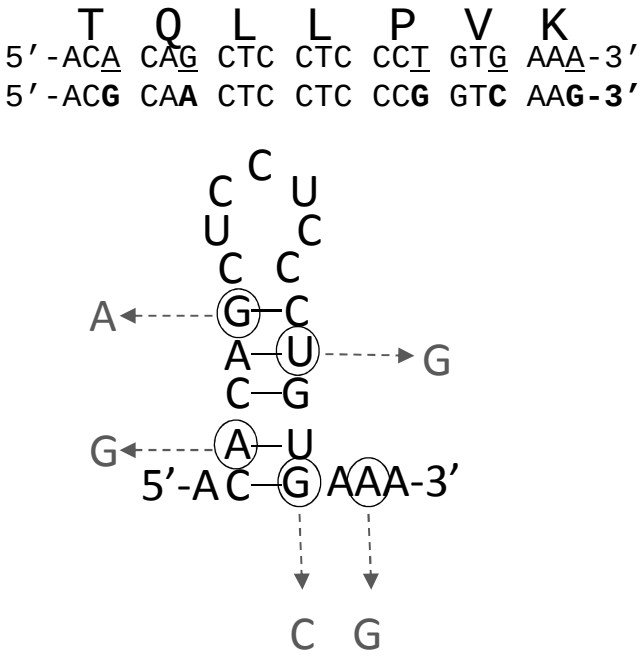
Miot M, Betton JM (2007) Optimization of the inefficient translation initiation region of the cpxP gene from *Escherichia coli*. Protein Sci 16: 2445-2453.

Nudler E, Gottesman ME (2002) Transcription termination and anti-termination in *E. coli*. Genes Cells 7: 755-768.

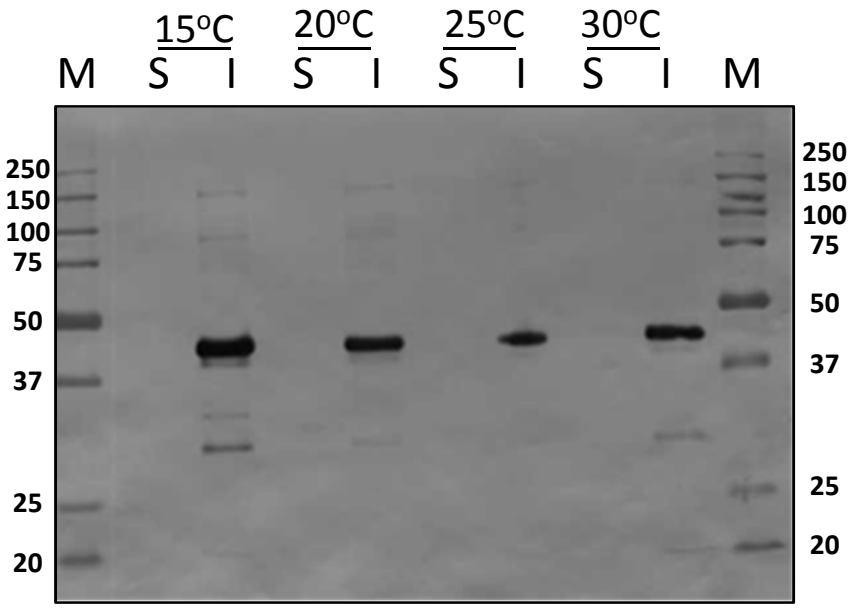
**A**

| Plasmid name | Recombinant GLA | Cell compartment | Size | Molecular weight |
|---|---|---|---|---|
| a) pET22b-GLA | GLA | cytoplasm | 420 aa | 47.8 kDa |
| b) pDAss-GLA | GLA | periplasm | 415 aa | 47.3 kDa |
| c) pGEX4T2-GLAmut1 | GST GLA | cytoplasm | 640 aa | 73.6 kDa |
| d) pGEX4T2-GLAmut2 | GST GLA | cytoplasm | 640 aa | 73.6 kDa |

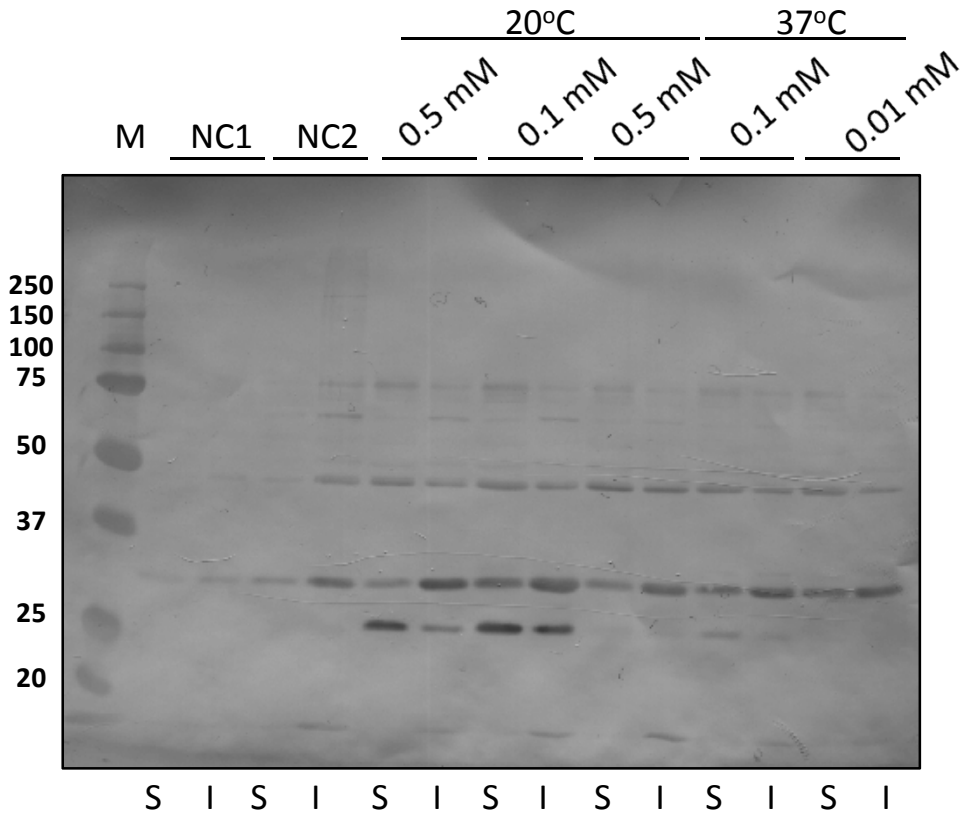
**B**



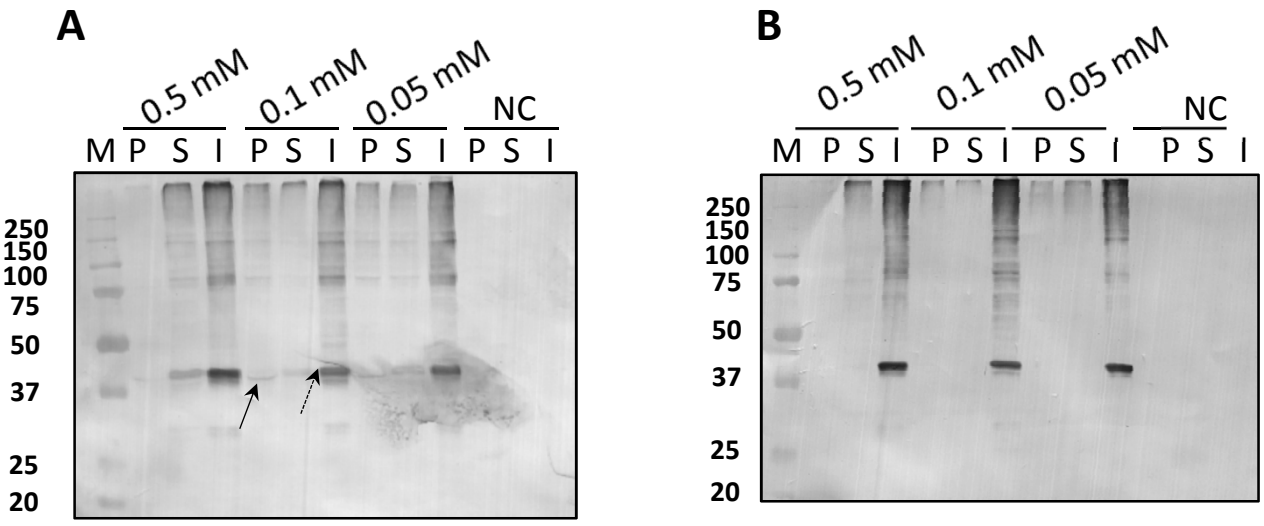

Suppl. Figure S1 (a) Expression vectors used in recombinant *h*GLA production assays. Expression vectors used in *E. coli* strains: a) to d). GST: Glutathione S-transferase, GLA: *hGLA* gene. *Tobacco Etch Virus* protease cleavage site (TEVp) location is marked by an arrow tip. His-tag is marked by a stripped box. GLA mut1 gene corrects two rare codons for arginine of the human GLA coding sequence to make it compatible with the *E. coli* expression system. GLAmut2 gene corresponds to GLAmut1 coding sequence with silent mutations to avoid putative transcriptional terminator (b) Hairpin configuration of putative *E. coli* transcriptional terminator in RNA derived sequence of underlined amino acids in Fig. 1b. Base specific modifications are depicted in bold. DNA coding sequence is shown before and after site-directed mutagenesis to obtain pGEX4T2-mut2.



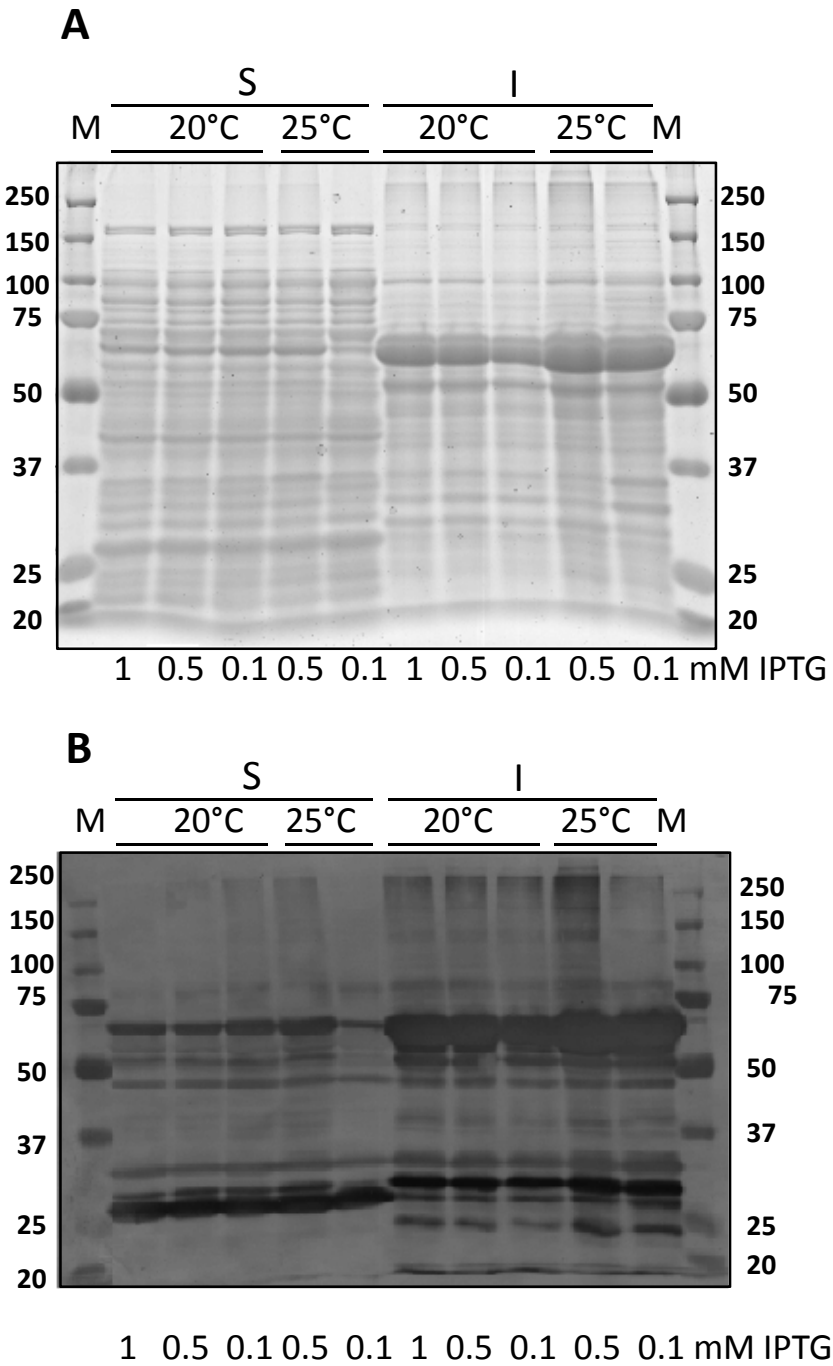

**Suppl. Figure S2** Western blot developed with monoclonal antibody against His-tag of soluble and insoluble fractions of induced BL21(DE3)/pReceiver-B01-GLA cultures. Gene expression induction was set at 15°C, 20°C, 25°C or 30°C for 16 h. Volumes corresponding to equal OD values were loaded in all the lanes. S and I correspond to the soluble and insoluble cell fractions of induced cultures. Molecular weight marker standards are indicated in kDa at both blot sides (Dual color, BioRad).




**Suppl. Figure S3** Western blot analysis of soluble (S) and insoluble (I) cellular fractions of Rosetta-gami B(DE3) cell cultures transformed with pET22b-GLA under different IPTG concentrations and induction temperatures, developed with polyclonal anti-GLA antibody (Santa Cruz Biotechnology). Non induced cell cultures are noted as NC lanes. Molecular weight marker standards are indicated in kDa (Dual color, BioRad).




**Suppl. Figure S4** Western blot analysis of soluble (S), insoluble (I) and periplasmic cellular fraction (P) of Rosetta 2(DE3) cell cultures transformed with pDass-GLA developed with (a) polyclonal anti-GLA antibody (Sigma). (b) monoclonal anti-his tag (GE Healthcare). The unprocessed form of the *h*GLA is marked by an arrow with discontinued line and the secreted form of *h*GLA is marked by an arrow. Molecular weight marker standards are indicated in kDa (Dual color, BioRad).




**Suppl. Figure S5** ( a) Coomassie blue stained SDS-PAGE of soluble and insoluble cellular fractions of Rosetta-gami B(DE3) cell cultures transformed with pGEX4T2-GLA at different temperatures and IPTG concentrations. (b) Western blot analysis of the same samples using a polyclonal anti-GLA antibody (Santa Cruz Biotecnology). Molecular weight marker standards are indicated in kDa (Dual color, BioRad).