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.

[ǂ]Present 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

Abstract

Obtaining high levels of pure proteins remains the main bottleneck of many scientific and biotechnological studies. Among all the available recombinant expression systems *Escherichia coli* facilitates gene expression by its relative simplicity, inexpensive and fast cultivation, well-known genetics and the large number of tools available for its biotechnological application. However, recombinant expression in *E. coli* is not always a straightforward procedure and major obstacles are encountered when producing many eukaryotic proteins and especially membrane proteins, linked to missing post-translational modifications, proteolysis and aggregation. In this context, many conventional and unconventional eukaryotic hosts are under exploration and development, but in some cases linked to complex culture media or processes. In this context, alternative bacterial systems able to overcome some of the limitations posed by *E. coli* keeping the simplicity of prokaryotic manipulation are currently emerging as convenient hosts for protein production. We have comparatively produced a "difficult-to-express" human protein, the lysosomal enzyme alpha-galactosidase A (*h*GLA) in *E. coli* and in the psychrophilic bacterium *Pseudoalteromonas haloplanktis* TAC125 cells (*P. halopanktis* TAC125). While in *E. coli* the production of active *h*GLA was unreachable due to proteolytic instability and/or protein misfolding, the expression of *h*GLA gene in *P.halopanktis* TAC125 allows obtaining active enzyme. These results are discussed in the context of emerging bacterial systems for protein production

38 that represent appealing alternatives to the regular use of *E. coli* and also of more complex eukaryotic
39 systems.
40

41 **Keywords:**
42 recombinant protein, expression systems, *Escherichia coli*, *Pseudoalteromonas haloplanktis*
43 TAC125, human alpha-galactosidase A, Fabry's disease.
44

45 **Introduction**

46 The deep genetic and physiological characterization, short generation time, ease of handling,
47 established fermentation know-how and finally the capacity to accumulate foreign proteins to a high
48 percentage of the total cellular protein content have made *E. coli* the most widely used prokaryotic organism
49 for recombinant protein production. However, there are disadvantages in using *E. coli* as an expression host.
50 For instance, *E. coli* is not capable of producing eukaryotic post-translational modifications, such as
51 glycosylation, phosphorylation or disulphide bridge formation in the reducing cytoplasm, which can be
52 critical for the production of folded, active proteins. Even so, several strategies can be implemented to obtain
53 specific post-translational modifications as the use of *E. coli* strains which have been engineered to maintain
54 oxidizing conditions in the cellular milieu or to send the recombinant protein to the periplasmic space to allow
55 cysteine bridging (Inaba, 2009; Nozach et al., 2013; Prinz et al., 1997). In addition, a N-linked protein
56 glycosylation system has been recently identified in the human enteropathogen *Campylobacter jejuni* that can
57 be transferred to *E. coli* (Wacker et al., 2006) opening up the possibility of engineering recombinant products
58 needing glycosylation for research and industrial applications.

59 On the other hand, some proteins, especially large and membrane proteins, simply fail to be
60 produced in *E. coli* and they occur as proteolysed species or deposited in inclusion bodies (Vallejo and Rinas,
61 2004). Others, are subjected to premature termination of translation in the presence of repetitive DNA
62 sequences or rare codons (Daly and Hearn, 2005; Sallach et al., 2009) or have a low rate of internal
63 translation initiaton (Ferreira et al., 2013; Nakamoto, 2009). A large number of studies describe the
64 conversion of proteins accumulated in inclusion bodies into soluble forms. Mainly, these methods can be
65 categorized into three alternative approaches. In the first one, factors influencing the amount of recombinant
66 protein present in the insoluble fraction can be modified through careful optimization and control of the
67 production conditions, leading to the expression of the recombinant protein in its soluble version. In this
68 context, conventional approaches include gene expression at low-temperature, use of promoters with different
69 transcriptional strengths, modifications of growth media and the use of folding modulators (Kolaj et al.,
70 2009). Alternatively, the protein can be extracted from inclusion bodies either under native or denaturing
71 conditions (Martinez-Alonso et al., 2009). Finally, the target protein can be engineered to achieve soluble
72 expression through fusion of solubility-enhancing tags (Sahdev et al., 2008; Torres et al., 2012).

73 In the recent years, a novel prokaryotic expression system has been developed based on the use of
74 the psychrophilic bacterium *P. halopanktis* TAC125 by driving the expression of genes of interest by both
75 basal or inducible promoters (Duilio et al., 2004a; Duilio et al., 2004b). In addition, as was established for *E.*

coli expression system, mutation of genes coding for proteases greatly reduces proteolysis of the recombinant protein (Parrilli et al., 2008). The system has demonstrated to be especially useful in improving protein solubility in relation to the widely used *E. coli* expression system and gives higher protein yield for secreted proteins (Cusano et al., 2006; Giuliani et al., 2014; Vigentini et al., 2006). Proteomics analysis point out the triger factor chaperone as the main factor involved in protein folding during recombinant protein expression under cold temperatures (Piette et al., 2010).

The human α-galactosidase A (EC 3.2.1.22; α-Gal A or GLA) is the lysosomal exoglycosidase responsible for the hydrolysis of terminal α-galactosyl moieties from various glycoconjugates. In the late 1980's, the full-length cDNA and the entire genomic sequence encoding mammalian GLA were isolated and characterized (Calhoun et al., 1985; Gotlib et al., 1996). The 1.4 kb full-length cDNA encodes a peptide of 429 residues, which includes a 31-residue amino-terminal signal peptide. This enzyme precursor is a glycopeptide (of approx. 55 kDa) that is processed by cleavage of the signal peptide and by oligosaccharide modifications in the Golgi and lysosomes to form the mature, active, homodimeric enzyme (approx. 100 kDa). After cleavage of the signal peptide, the glycopeptide undergoes modification of its N-linked oligosaccharide moieties in the Golgi apparatus, and then it is transported to the lysosome via the mannose-6-phosphate receptor (M6PR)-mediated pathway. Mutations in the *h*GLA gene resulting in deficient or absent enzymatic activity are the basis of Fabry's disease, a disorder characterized by progressive glycosphingolipid deposition in vascular lysosomes leading to early demise from renal, cardiac, or cerebral vascular disease (Garman and Garboczi, 2002).

The subunits of the glycosylated mature enzyme contain four putative N-glycosylation consensus sites at positions Asn139, Asn192, Asn215 and Asn408 and five disulfide bonds between residues Cys52-Cys94; Cys56-Cys-63, Cys142-Cys172, Cys202–Cys223, and Cys378–Cys382 (Garman and Garboczi, 2004; Saito et al., 2013). The human enzyme has been purified from a variety of sources and its physical and kinetic properties have been also characterized (Chen et al., 2000a; Chen et al., 2000b; Corchero et al., 2011; Corchero et al., 2012; Yasuda et al., 2004). However, the *h*GLA obtained from prokaryotic expression systems has neither been fully characterized nor purified (Hantzopoulos and Calhoun, 1987). Therefore, the objective of this work is the production and purification of active *h*GLA from prokaryotic cell factories.

Material and Methods

Bacterial strains and plasmids

The *E. coli* strains used in this work were DH5α for plasmid cloning and maintenance; BL21(DE3), *Rosetta* 2(DE3) and Rosetta-gami B(DE3) (Novagen) for recombinant protein expression and S17-1(λ*pir*) was used as donor in interspecific conjugation experiments (Tascon et al., 1993). *P. haloplanktis* TAC125 is a Gram-negative bacterium isolated from Antarctic seawater and is deposited and available at the Institut Pasteur Collection (CIP 108707) (Medigue et al., 2005). Antarctic bacteria transformation was achieved by intergeneric conjugation as previously reported (Duilio et al., 2004b).

112 The four *E. coli* expression vectors used in this work (Fig. 1a) are as follows: a) pReceiver-B01-GLA
113 (product EX-Q0172-B01, OmicsLink ORF Expression Clone), encoding a full-length version of *h*GLA under
114 the control of the T7 promoter; in this vector, 6xHis tag is fused to the N-terminus of the *h*GLA ORF for
115 further detection and purification purposes. b) pGEX4T2-GLA, encoding the mature form of the *h*GLA with a
116 GST (Glutathione S-transferase) fused at its N-terminus; a His-tag and a Tobacco etch Virus protease (TEVp)
117 cleavage site were also added in the designed PCR oligonucleotides of the *h*GLA gene for purification
118 purposes. c) pGEX4T2-Opt-GLA, synthetic, codon optimised *h*GLA gene (GeneArt, Invitrogen; KF500099)
119 cloned into pGEX4T2 as previously described. d) pGEX4T2-GLA-GFP, derived from pGEX4T2-GLA in
120 which the GFP coding sequence was inserted into the *Eco*RI cloning site. Some other four expression vectors
121 used in *E. coli* strains are described in Suppl. Figure S1a.

122 Finally, the psychrophilic gene expression vector pPM13psDs-GLA (e) was constructed as follows:
123 the pPM13psDs shuttle vector, containing the strong constitutive promoter PM13 (Duilio et al., 2004a)
124 correctly located with respect to a psychrophilic Shine-Dalgarno sequence and the 93bp encoding the signal
125 peptide of *P. haloplanktis* TAC125 DsbA, was double digested by *Sal*I-*Eco*RI. The synthetic codon optimised
126 *h*GLA gene (KF500099) was double digested by *Sal*I-*Eco*RI and cloned into pPM13psDs corresponding sites
127 (Fig. 1a).

128

129 **Recombinant human GLA production and cellular fractionation**

130 The production process in recombinant *E. coli* cells was performed by the Protein Production
131 Platform (http://ibb.uab.es/ibb) of the Institute of Biotechnology and Biomedicine (IBB-UAB) and the
132 Biomedical Networking Center (CIBER-BBN). In short, shake flask cultures were set at 37°C and 250 rpm in
133 LB rich medium, plus 100 µg/mL of ampicillin for plasmid maintenance in all expression *E. coli* vectors.
134 Expression of recombinant *h*GLA gene was induced when the absorbance at 550 nm reached values around
135 0.5, adding isopropyl β-D-1-thiogalactopyranoside (IPTG) at concentrations ranging from 0.1 mM to 1 mM.
136 Incubation proceeded then at different temperatures 16ºC, 20ºC, 25ºC and 37ºC depending on the strain as
137 indicated. After induction of gene expression, 1.5 mL culture samples were withdrawn at different times and
138 soluble and insoluble fractions were separated. In short, cells were harvested by centrifugation at 5,000 g (at
139 4ºC) for 15 min and resuspended in 200 µl of PBS buffer (7.5 mM Na_2HPO_4, 110 mM NaCl, 2.5 mM
140 NaH_2PO_4, pH 7.4) and further sonicated. Soluble and insoluble fractions were separated by centrifugation at
141 15,000 g for 15 min at 4ºC. Insoluble fractions were resuspended in PBS (same volumes than those from their
142 respective soluble fractions). Both fractions were stored at –80ºC until further analysis.

143 *P. haloplanktis* TAC125 (pPM13psDs-GLA) recombinant cells were cultured in aerobic conditions
144 at 4°C and 15°C in TYP broth (16 g/l yeast extract, 16 g/l bacto tryptone, 10 g/l marine salt mix) at pH 7.5,
145 supplemented with 100 µg/ml ampicillin, and the recombinant *h*GLA production and its cellular localization
146 were evaluated in cell samples withdrawn at different growth phases. Cell pellets (corresponding to
147 OD_{600nm}=10) were re-suspended in 1 ml of 50 mM Tris–HCl pH 8.0, 50 mM EDTA and disrupted by
148 ultrasonic treatment consisting of six cycles of 30 s on/ 1 min off, on ice. The mixture was centrifuged for 15

149 min at 10,000 *g* at 4°C; the resulting supernatant and pellet were collected as soluble and insoluble fractions,
150 respectively. Total *P. haloplanktis* TAC125 (pPM13psDs-GLA) cell extract was obtained by resuspending 0.5
151 OD_{600nm} cell pellet of into 400 µl of SDS loading dye, followed by 30 min treatment at 90°C.
152 The periplasmic cell extract was obtained by resuspending bacterial pellets in 1/20 of culture volume
153 of borate buffer (200 mM $Na_2B_4O_7$, 130 mM NaCl, 5 mM EDTA, pH 8) and incubating the mixture for 18
154 hours at 4°C. The suspension was then centrifuged at 8,000 rpm for 15min at 4°C. The supernatant was stored
155 as periplasmic extract, while the pellet was resuspended in 1/20 of original culture volume of SDS loading
156 dye, followed by 30 min treatment at 90°C and stored as cytoplasmic extract.
157
158 **Recombinant protein detection**
159 Total cell extracts, soluble, insoluble and periplasmic cell fractions were analysed by SDS-PAGE
160 and Coomassie staining. For Western-blot studies, upon SDS-PAGE proteins were blotted onto nitrocellulose
161 membranes, *h*GLA immunoreactive bands were developed using either a rabbit polyclonal anti-GLA serum
162 from Santa Cruz Biotechnology, Inc. (α-gal A H-104: sc-25823) raised against an epitope corresponding to
163 amino acids 316-429 at the C-terminus of the mature form of *h*GLA (Fig. 1b), a rabbit polyclonal anti-GLA
164 serum from Sigma Prestige Antibodies (product HPA000237) raised against amino acids 302-412 of the
165 mature form of *h*GLA, or a monoclonal antibody anti-His tag from GE Healthcare (product 27-471001), and
166 the corresponding secondary antibodies. GST was detected with mouse monoclonal anti-GST antibody (Santa
167 Cruz Biotechnology, Inc.; SC-138) and GFP with rabbit polyclonal anti-GFP antibodies (Santa Cruz
168 Biotechnology, Inc.; SC-8334).
169 For comparison purposes, gels were loaded with sample volumes adjusted according to the OD of the
170 culture. Samples for quantitative comparison were run in the same gel and processed as a set. Bands were
171 analysed with the Quantity One analysis software (BioRad).
172
173 **Protein purification**
174
175 GST-containing *h*GLA recombinant proteins were purified from Rosetta-gami B cell cultures
176 induced at 0.1 mM IPTG for 16 h at 20ºC at 250m rpm. Cell pellets were resuspended in PBS buffer (7.5 mM
177 Na_2HPO_4, 110 mM NaCl , 2.5 mM NaH_2PO_4, pH 7.4) in the presence of EDTA-free protease inhibitors
178 (Roche Applied Science) and cells were lysed using a French Press (Thermo FA-078A) at 1,100 psi. The
179 soluble fraction was separated by centrifugation at 15,000 g for 45 min at 4°C, filtered through 0.22-µm filters
180 and loaded onto 1 ml GSTrap HP column (GE Healthcare, 17-5281-05). Bound GST-GLA protein was eluted
181 with 5 column volumes of elution buffer (50 mM Tris-HCl and 10 mM reduced glutathione, pH 8.0). Positive
182 fractions were collected and dialysed against 0.01M acetic buffer (pH 4.5) and protein concentration was
183 estimated using the Bradford method.
184 The pellets of *E. coli* or *P. haloplanktis* TAC125 cells containing His-tagged *h*GLA proteins were
185 resuspended in buffer A (20 mM Tris-HCl pH 7.5, 500 mM NaCl, and10 mM imidazole) and cells disrupted
186 by sonication in the presence of EDTA-free protease inhibitors (Complete, 11873580001 from Roche). The

187 soluble cell fraction was separated by centrifugation at 15,000 g for 15 min at 4°C. After filtration through
188 0.22 µm filters, recombinant proteins were purified by affinity chromatography in Ni^{2+} columns (HiTrap
189 Chelating HP columns, 17-0408-01 from GE Healthcare) in an ÄKTA™ Purifier (GE healthcare) fast protein
190 liquid chromatography system. Positive fractions in elution buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl
191 and 500 mM Imidazole) were collected and dialyzed against 0.01M acetic buffer (pH 4.5). Protein
192 concentration was estimated using the Bradford method.

193 Purified proteins were characterised by N-terminal sequencing by Edman's automated degradation
194 using in an Applied Biosystems Procise 492 protein sequencer and molecular weight was experimentally
195 determined by mass spectrometry (UltraFlex MALDI-TOF mass spectrometer, Bruker Daltonics, Bremen,
196 Germany). Both analysis were performed at the Proteomics and Bioinformatics Unit of the Scientific
197 Technical Service, SepBioEs, of the Autonomous University of Barcelona.

198

199 **Time course of TEV cleavage reaction of GST-GLA-GFP**

200 Purified GST-GLA-GFP was dialysed against reaction buffer (50 mM Tris-HCl, pH 8.0, 0.5 mM
201 EDTA). 10 Units of TEVp (Invitrogen, 12575-015) were added to 20 µg of GST-GLA-GFP and incubated at
202 30ºC. 30 µl aliquots were removed at specified incubation times. Reaction was stopped by adding 30 µl stop
203 solution (125 mM Tris-HCl, pH 6.8; 4% SDS; 1.4 M β-mercaptoethanol; 20% (v/v) glycerol; 0.01%
204 bromophenol blue) and samples were stored at -20ºC until gel electrophoresis analysis. Protein samples when
205 performing *h*GLA activity assays were used without buffer addition after protease incubation.

206

207 **Enzyme assay of α-galactosidase A activity**.

208 The enzymatic *h*GLA activity was assayed with 4-methylumbelliferyl α-D-galactoside (4MU-α-Gal,
209 ref. M7633, Sigma Chemical, St. Louis, MO) as substrate, at a concentration of 2.46 mM in 0.01M acetic
210 buffer (pH 4.5). A typical 4MU-α-Gal assay was performed with a reaction mixture containing 100 µL of
211 substrate and 25 µL of enzyme solution. Enzymatic reactions took place in agitation, at 37ºC for 1 hour.
212 Reaction was stopped adding 1.25 mL of 0.2 M glycine-NaOH buffer (pH 10.4). The released 4-
213 methylumbelliferone (4-MU) was determined by fluorescence measurement at 365 and 450 nm as excitation
214 and emission wavelengths, respectively. Samples ranging from 5 to 500 ng/ml of 4-MU (ref. M1381, Sigma,
215 St. Louis, MO) in 0.2 M glycine–NaOH buffer (pH 10.4), were used as standard curve. One unit of enzyme
216 activity was defined as the amount of enzyme releasing 1 nmol of 4-MU per mg of enzyme and per hour.
217 *h*GLA obtained from human fibroblasts or HEK-293T cells were used as positive controls (Corchero et al.,
218 2011).

219

220 **Results**

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

222 Plasmid pReceiver-B01-GLA, which contains the complete ORF of *h*GLA gene, was transformed in *E. coli*
223 BL21(DE3) and Rosetta(DE3) *E. coli* strains (Fig. 1a). The Rosetta strain is able to compensate low

224 concentrations of certain tRNAs of *E. coli* and to improve recombinant protein translation. It should be
225 noticed that in the *h*GLA ORF at least 4 critical Arginine codons are present in the DNA sequence (Fig. 1b).
226 Strains BL21(DE3) and Rosetta(DE3), with and without the expression plasmid, were submitted to gene
227 expression induction conditions and analysed by SDS-PAGE followed by Western blot. Nitrocellulose
228 membranes were developed either with a polyclonal anti-GLA antibody (Sigma-Aldrich Co.) raised against
229 an epitope corresponding to amino acids 302-412 of the *h*GLA protein or with the monoclonal anti-His tag
230 (Fig. 2a, b). Results showed that the polyclonal antibody from Sigma and the anti-his antibody detect a main
231 protein band corresponding to the degradation product of about 40-42 kDa (expected MW 49.6 kDa), both
232 antibodies detecting this band as the main immunoreactive product, while samples from not transformed cells
233 did not show any immunoreactive band. Also, both antibodies detected bands that might be associated to
234 protein aggregates or oligomers, while proteolysis fragments were detected with the anti His-tag monoclonal
235 antibody but not with the polyclonal antibody, indicating a specific C-terminal cleavage. Finally, equivalent
236 protein bands were detected upon tRNA supplementation. According to these results, only a product of
237 approximately 42 kDa and probably other aggregated and degraded protein products, recognized by the
238 monoclonal anti-His tag and the anti-GLA antibodies, could be obtained in these strains transformed with the
239 pReceiver-B01-GLA expression vector.
240 In addition, cell fractionation studies performed in *E. coli* BL21(DE3)/pReceiver-B01-GLA cultures showed
241 that the *h*GLA produced at different induction temperatures was exclusively detected in the insoluble cell
242 fraction (Suppl. Figure S2).
243
244 **Recombinant production of mature form of *h*GLA in *E. coli***
245 *h*GLA ORF present in pReceiver-B01-GLA codes for the 429 amino acid *h*GLA precursor including
246 31 amino acid residues corresponding to the signal peptide, which is naturally processed in the ER to a 398
247 amino acids mature form (Lemansky et al., 1987) (Fig. 1b). Since eukaryotic signal peptides are not processed
248 in prokaryotic hosts, nucleotides coding exclusively for the mature form of the *h*GLA ORF (32-429; Fig. 1b)
249 were amplified by PCR and transferred to *E. coli* expression vectors to improve protein folding in the
250 heterologous expression system (Suppl. Figure S1a, expression vectors a) and b)). In addition, to promote the
251 formation of disulfide bonds in the *h*GLA product, the experiments were performed in either Rosetta-gami *E-*
252 *coli* strain or the protein was sent to the oxidizing environment present in the periplasm. Expression of the
253 mature form of *h*GLA in *E. coli* under these conditions produced aggregated forms of the recombinant protein
254 and protein bands compatible with degradation products (Suppl. Figure S3 and Suppl. Figure S4).
255 It has been reported that purification of full-length human proteins can be achieved by fusion to soluble
256 proteins as GST in *E. coli* (Braun et al., 2002), this DNA fragment was also inserted into pGEX4T-2, which
257 adds a GST fusion protein at the N-terminal end of the desired protein product (Figure 1a).
258 Protein expression experiments were carried out with transformed Rosetta-gami B(DE3) cells with
259 pGEX4T2-GLA with the aim of using the fused GST for solubilisation and purification purposes. Several
260 protein induction conditions were assayed (Suppl. Figure S5). Specific protein bands were detected in all
261 tested conditions in both soluble and insoluble fractions. A predominant band was observed in the insoluble

262 fraction with an apparent weight below 75 kDa that would correspond to the expected GST tagged
263 recombinant *h*GLA, apart from other protein bands compatible with aggregates and proteolysed forms.
264
265 **Purification and characterisation of the produced human GLA**
266 Soluble GST-fused protein was purified by affinity chromatography using a GSTrap HP 1 mL column (GE
267 Healthcare). Elution peak contains highly pure protein sample obtained in one single chromatographic step,
268 indicating that the chosen purification strategy was appropriate (Fig. 3). However, molecular weight of the
269 purified protein was again smaller than expected. N-terminal sequencing of the protein band demonstrated
270 that corresponded to the initial amino acids of the recombinant protein, although the first methionine was only
271 partially removed. In *E. coli*, methionine aminopeptidase removes initial methionine with variable efficiency
272 depending on the penultimate amino acid and it has been demonstrated that the presence of an acidic residue
273 as L present in *h*GLA reduces its activity up to 20% (Liao et al., 2004; Xiao et al., 2010a). Moreover, mass
274 spectrometry of the affinity purified sample revealed that the exact molecular weight of the product was 68.9
275 kDa (theoretical weight 73.6 kDa). This size reduction in 4.7 kDa is compatible with the elimination of 29
276 amino acids at the C-terminus of the protein since as it has been demonstrated, the N-terminal end remains
277 intact. Therefore, in this proteolized *h*GLA protein, the last amino acid corresponded then to S401 after which
278 an Arginine coded by a rare codon has to be added at position 402. In addition, at position 392 another
279 Arginine rare codon is present (Fig. 1b).
280 On the basis of the above-mentioned results, and to avoid the presence of rare codons close to the proteolytic
281 point of the polypeptide chain, site-directed mutagenesis of the mentioned Arginine residues in the *h*GLA
282 (R402 and R392 rare codons) were performed using the QuikChange Lightning site-directed mutagenesis kit
283 (Stratagene) with oligonucleotides designed by PrimerX (www.Bioinformatics.org). The resulting mutant
284 pGEX4T2-GLA (pGEX4T2-GLAmut1; Figure S1a supplementary information) was sequenced and used in
285 expression experiments. Western blot analyses of the resulting protein samples showed that a GST-*h*GLA
286 protein could be obtained but its molecular weight corresponded to the 69 kDa truncated protein previously
287 obtained (data not shown).
288 Further analysis of the RNA sequence of the *h*GLA revealed the presence of a putative *E. coli* transcriptional
289 terminator motif surrounding the nucleotide position at which translation seems to finish (Nudler and
290 Gottesman, 2002) (Suppl. Figure S1b). Oligonucleotides were designed to eliminate hairpin loop secondary
291 structures in that region and were used to mutate the pGEX4T2-GLA DNA sequence using the QuikChange
292 Lightning site-directed mutagenesis kit (Stratagene) generating pGEX4T2-GLAmut2 (Suppl. Figure S1a).
293 Protein expression experiments showed that under these conditions, the resulting recombinant protein was
294 still truncated, showing an apparent molecular weight of 69 kDa (data not shown). Protein aggregation was
295 observed in protein samples stored at 4ºC for 2 weeks and confirmed in Superdex 200 10/300 GL (GE
296 Healthcare) gel filtration chromatography since a protein peak was obtained in the void volume of the
297 column. Activity assays of purified GLA protein fused to GST were performed using purified human GLA
298 from a mammalian expression system as positive control (Corchero et al., 2012) as well as cellular extracts of

299 *E. coli* cells transformed with pReceiver-B01-GLA and pET22b-GLA but enzymatic activity was not detected
300 in any sample (Table 1).

301 A global codon optimisation of *h*GLA gene was attempted next. For this purpose, a synthetic GLA gene was
302 designed and constructed (GeneArt; Invitrogen) which was cloned within the pGEX4T2 vector to produce a
303 pGEX4T2-Opt-GLA expression vector with an optimised codon usage (Fig. 1a, expression vector c)). In that
304 construct, the His-tag purification module was fused at the C-terminus to differentiate from the previous GST-
305 *h*GLA constructs in which the his-tag module was located at the N-terminus of *h*GLA. The resulting
306 expression vector was used to transform the Rosetta-gami B(DE3) strain. Protein expression experiments of
307 the resulting strain indicated that once again, a truncated protein was produced and hence, production of the
308 intact *h*GLA does not depend on optimised codon usage (Fig. 4a). In addition, the detection of the GST-GLA
309 recombinant protein with anti-GST antibodies demonstrated the presence of GST in the final protein product
310 (Fig.4b). However, the same samples were not detected by anti-his antibodies indicating that truncation of
311 GLA protein occurred at the C-terminus (Fig. 4c). On the other hand, GST-GLA obtained from the expression
312 of the non-optimised gene contained in pGEX4T2-GLA was detected by both anti-GST and anti-His
313 antibodies since in this construction His-tag peptide was located N-terminally between GST and GLA coding
314 sequences (Fig.1a).

315 Since an incomplete product was obtained regardless of the strategies attempted so far, a protease cleavage
316 analysis was carried out using the ExPaSy PeptideCutter program. However, no specific protease cleavage
317 site was found at the C-terminal end of the truncated GLA product obtained. Nevertheless, to confirm these
318 findings, the next attempted strategy was the fusion of another protein (green fluorescent protein, GFP, (Hsieh
319 et al., 2010)) at the C-terminus of the GST-GLA with the aim to protect GLA from a possible proteolysis at
320 this end (Murby et al., 1991). pGEX4T2-GLA-GFP (Fig. 1a, expression vector d)) was obtained and
321 transformed into Rosetta-gami B(DE3). Protein expression assays using anti-GLA (Fig. 5a) and specifically
322 anti-GFP western-blot experiments (Fig. 5b) demonstrated that the whole fused protein GST-His-GLA-GFP
323 could be produced by this strain in contrast to previous experiments. Purification and subsequent TEV
324 proteolysis of the fused protein (Fig. 5c) did not succeeded in producing an active *h*GLA product despite the
325 strategies attempted (GST and His-tag affinity chromatography), and only background signal was detected
326 when determining the enzymatic activity of the obtained product corresponding to the signal background of
327 the negative control (data not shown).

328

329 **Recombinant production of mature form of *h*GLA in *P. haloplanktis* TAC125**

330 Recombinant *P. haloplanktis* (pPM13psDs-GLA) cells were aerobically grown in complex rich medium at
331 two different temperatures (4 and 15°C) and samples were withdrawn at different times of cultivation. The
332 *h*GLA production was evaluated by SDS-PAGE followed by western blotting analysis (using specific anti
333 GLA polyclonal antibodies) of total soluble protein extracts and it revealed highest protein yield after 24 h of
334 cultivation at 15°C (Fig. 6a and 6b). No recombinant GLA protein was detected in insoluble protein extracts

335 (data not shown). A cell fractioning followed by western blotting analysis demonstrated that recombinant
336 GLA produced at higher temperature is also fully localized in the bacterial periplasm (Fig. 6c), while when
337 produced at 4°C, a fraction of the protein was still associated to the cytoplasm, likely deriving from a less
338 efficient translocation across bacterial inner membrane. Furthermore, at the lower temperature, the
339 recombinant *h*GLA translocated in the periplasm was also subjected to host encoded proteolytic cleavage, as
340 demonstrated by the appearance of a lower molecular weight specific band on western blotting analysis
341 shown in Fig. 6c.
342 Soluble protein extracts from *P. haloplanktis*/pPM13psDs-GLA grown at 15°C for 24 h were used for His-
343 trap affinity purification (His-trap 1 mL column, GE-Healthcare) according to the protocol reported in the
344 methods section. A linear gradient was used for elution (20-500 mM imidazole in 15 min at 0.5 mL/min
345 flow), thus obtaining a partial purified protein (about 70 % purity) which has been used for determination of
346 enzymatic activity (Fig. 6d). Although the purified enzyme has the tendency to precipitate, it turned out to be
347 active and a value of 77.4±0.3 μmol $h^{-1}mg^{-1}$ prot of activity was recorded (Table 1).
348

349 **Discussion**

350

351 Recombinant protein production in heterologous expression systems represents a major bottleneck in the
352 development of biopharmaceutical, industrial and research applications (Liras, 2008). Even though high
353 throughput techniques for cloning and protein production have been already described (Barnard et al., 2010;
354 Nettleship et al., 2010; Savitsky et al., 2010; Xiao et al., 2010b), purification of challenging proteins is still a
355 matter of trial an error approaches. As expression system, *E. coli* offers many advantages over eukaryotic
356 systems referring to rapidity, simplicity and expenditure (Demain and Vaishnav, 2009). However it has some
357 disadvantages which in some cases are difficult to override as the lack of many posttranslational
358 modifications and high tendency to form aggregated species (Panda, 2003). In the present study, a human full-
359 length protein, GLA, which has been purified from eukaryotic expression systems (Chen et al., 2000a; Chen
360 et al., 2000b; Corchero et al., 2012; Yasuda et al., 2004) and none prokaryotic source (Hantzopoulos and
361 Calhoun, 1987), has been selected to try available expression methods in *Escherichia coli* and the Antarctica
362 psychrophilic bacterium *P. haloplanktis* TAC125. In 1987, Hantzopolus and coauthors described the
363 production of an active *h*GLA in *E. coli* (Hantzopoulos and Calhoun, 1987). However, such results have not
364 been reproduced or continued since then, and moreover, no purification protocol has been published to date.
365 The interest in developing a production and purification system of human GLA from a prokaryotic source
366 relies on the high cost of the actual enzymatic replacement therapy in Fabry's disease treatment which
367 consists in recombinant protein obtained from mammalian cells having a great economical impact in patient
368 expenditure. We have shown that *h*GLA can be detected in *E. coli* as either aggregated and proteolytically
369 cleaved forms in any experimental condition tested. Aggregation of proteins seems to be driven by the
370 presence of specific amino acid sequences (Espargaro et al., 2012), saturation of the folding machinery (Kolaj
371 et al., 2009) or to the lack of the required posttranslational modifications as glycosylation (Ioannou et al.,
372 1998). However, full-length recombinant *h*GLA has been produced when fused to GST and GFP, although
373 purified protein is not correctly folded lacking biological activity. The absence of enzymatic activity might be
374 related to either unfolded protein species or deletion of the C-terminal end of the protein which seems to be
375 critical (Miyamura et al., 1996). In addition, GLA has 4 glycosylation sites added after signal peptide removal
376 in the Golgi apparatus. This sugar moieties (especially glycosylation site located at residue Asn215) play a
377 crucial role in both folding and solubility and consequently in enzymatic activity (Ioannou et al., 1998).
378 However, when producing *h*GLA in the psychrophilic expression system based in the use of the antarctic
379 bacteria *P. haloplanktis* TAC125, the functional full-length enzyme can be produced and purified. It has been
380 described in comparative studies that recombinant protein production at low temperature has a positive effect
381 in protein yield (Dragosits et al., 2011; Duilio et al., 2004a) as well as in protein conformation (Vigentini et
382 al., 2006). The ability of the *P. halopanktis* TAC125 to produce active *h*GLA over the proteolyitic and
383 aggregation tendency of the protein in *E. coli* migth be related to the presence of high peptidyl prolyl cis-trans
384 isomerases genes found in *P. haloplanktis* TAC125 genome, and more specifically to the upregulation of the
385 main molecular chaperone, trigger factor (Giuliani et al., 2014; Piette et al., 2010). The folding activity in the
386 psychrophilic expression system can counteract the high tendency of the *h*GLA to adopt unsuitable

387 conformations. On the other hand, in the *E. coli* cellular environment, the nascent *h*GLA polypeptides adopt
388 unstable conformational species that are detected by cellular proteases, mainly at the C-terminus of the
389 protein, leading to the formation of inactive aggregation-prone protein species. Even though the full-length
390 protein can be produced in *E. coli*, the limited chaperone activity in this expression system renders inactive
391 protein. Therefore, microbial host range might be expanded to psychrophilic expression systems when
392 difficult to express proteins such as full-length mammalian proteins are to be produced.
393

394 **Acknowledgments**

395 This work was supported by ERANET-IB08-007 project from the European Union and its linked national
396 project EUI2008-03610 to AV. We also appreciate the support from EME2007-08 to NFM from Universitat
397 Autonoma de Barcelona, from Antartide 2010 to MLT and EP, from MIUR Azioni Integrate Italia-Spagna
398 2010 Prot. IT10LECLM9 to MLT, from MINECO (IT2009-0021) to AV and LT, from AGAUR (2009SGR-
399 108 to AV). AV is also supported by The Biomedical Research Networking Center in Bioengineering,
400 Biomaterials and Nanomedicine (CIBER-BBN, Spain), an initiative funded by the VI National R&D&i Plan
401 2008–2011, Iniciativa Ingenio 2010, Consolider Program, CIBER Actions and financed by the Instituto de
402 Salud Carlos III with assistance from the European Regional Development Fund. PS has received predoctoral
403 fellowship from ISCIII, and AV has been distinguished with an ICREA ACADEMIA award (Catalonia,
404 Spain).

405

406 **Conflict of interests**

407

408 The authors declare no conflict of interests.

409

410

411 **References**

412

413 Barnard GC, Kull AR, Sharkey NS, Shaikh SS, Rittenhour AM, Burnina I, Jiang Y, Li F, Lynaugh H,
414 Mitchell T, Nett JH, Nylen A, Potgieter TI, Prinz B, Rios SE, Zha D, Sethuraman N, Stadheim TA,
415 Bobrowicz P (2010) High-throughput screening and selection of yeast cell lines expressing monoclonal
416 antibodies. J Ind Microbiol Biot 37: 961-971.

417 Braun P, Hu YH, Shen BH, Halleck A, Koundinya M, Harlow E, LaBaer J (2002) Proteome-scale purification
418 of human proteins from bacteria. Proc Natl Acad Sci U S A 99: 2654-2659.

419 Calhoun DH, Bishop DF, Bernstein HS, Quinn M, Hantzopoulos P, Desnick RJ (1985) Fabry disease:
420 isolation of a cDNA clone encoding human alpha-galactosidase A. Proc Natl Acad Sci U S A 82: 7364-7368.

421 Chen YS, Jin M, Egborge T, Coppola G, Andre J, Calhoun DH (2000a) Expression and characterization of
422 glycosylated and catalytically active recombinant human alpha-galactosidase a produced in *Pichia pastoris*.
423 Protein Expression and Purification 20: 472-484.

424 Chen YS, Jin M, Goodrich L, Smith G, Coppola G, Calhoun DH (2000b) Purification and characterization of
425 human alpha-galactosidase A expressed in insect cells using a baculovirus vector. Protein Expression and
426 Purification 20: 228-236.

427 Corchero JL, Mendoza R, Lorenzo J, Rodriguez-Sureda V, Dominguez C, Vazquez E, Ferrer-Miralles N,
428 Villaverde A (2011) Integrated approach to produce a recombinant, His-tagged human alpha-galactosidase A
429 in mammalian cells. Biotechnol Prog 27: 1206-1217.

430 Corchero JL, Mendoza R, Ferrer-Miralles N, Montras A, Martinez LM, Villaverde A (2012) Enzymatic
431 characterization of highly stable human alpha-galactosidase A displayed on magnetic particles. Biochemical
432 Engineering Journal 67: 20-27.

433 Cusano AM, Parrilli E, Marino G, Tutino ML (2006) A novel genetic system for recombinant protein
434 secretion in the Antarctic *Pseudoalteromonas haloplanktis* TAC125. Microb Cell Fact 5: 40.

435 Daly R, Hearn MT (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool
436 in protein engineering and production. J Mol Recognit 18: 119-138.

437 Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms.
438 Biotechnology Advances 27: 297-306.

439 Dragosits M, Frascotti G, Bernard-Granger L, Vazquez F, Giuliani M, Baumann K, Rodriguez-Carmona E,
440 Tokkanen J, Parrilli E, Wiebe MG, Kunert R, Maurer M, Gasser B, Sauer M, Branduardi P, Pakula T,
441 Saloheimo M, Penttila M, Ferrer P, Tutino ML, Villaverde A, Porro D, Mattanovich D (2011) Influence of
442 Growth Temperature on the Production of Antibody Fab Fragments in Different Microbes: A Host
443 Comparative Analysis. Biotechnology Progress 27: 38-46.

444 Duilio A, Madonna S, Tutino ML, Pirozzi M, Sannia G, Marino G (2004a) Promoters from a cold-adapted
445 bacterium: definition of a consensus motif and molecular characterization of UP regulative elements.
446 Extremophiles 8: 125-132.

447 Duilio A, Tutino ML, Marino G (2004b) Recombinant protein production in Antarctic Gram-negative
448 bacteria. Methods in molecular biology (Clifton, N J) 267: 225-237.

449 Espargaro A, Villar-Pique A, Sabate R, Ventura S (2012) Yeast prions form infectious amyloid inclusion
450 bodies in bacteria. Microbial Cell Factories 11: 89.

451 Ferreira JP, Overton KW, Wang CL (2013) Tuning gene expression with synthetic upstream open reading
452 frames. Proc Natl Acad Sci U S A 110: 11284-11289.

453 Garman SC, Garboczi DN (2002) Structural basis of Fabry disease. Mol Genet Metab 77: 3-11.

454 Garman SC, Garboczi DN (2004) The molecular defect leading to Fabry disease: structure of human alpha-
455 galactosidase. J Mol Biol 337: 319-335.

456 Giuliani M, Parrilli E, Sannino F, Apuzzo GA, Marino G, Tutino ML (2014) Recombinant production of a
457 single-chain antibody fragment in *Pseudoalteromonas haloplanktis* TAC125. Appl Microbiol Biotechnol 98:
458 4887-4895.

459 Gotlib RW, Bishop DF, Wang AM, Zeidner KM, Ioannou YA, Adler DA, Disteche CM, Desnick RJ (1996)
460 The entire genomic sequence and cDNA expression of mouse alpha-galactosidase A. Biochem Mol Med 57:
461 139-148.

462 Hantzopoulos PA, Calhoun DH (1987) Expression of the Human Alpha-Galactosidase-A in *Escherichia coli*-
463 K-12. Gene 57: 159-169.

464 Hsieh JM, Besserer GM, Madej M, Bui HQ, Kwon S, Abramson J (2010) Bridging the gap: A GFP-based
465 strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini.
466 Protein Sci 19: 868-880.

467 Inaba K (2009) Disulfide Bond Formation System in *Escherichia coli*. J Biochem 146: 591-597.

468 Ioannou YA, Zeidner KM, Grace ME, Desnick RJ (1998) Human alpha-galactosidase A: glycosylation site 3
469 is essential for enzyme solubility. Biochem J 332: 789-797.

470 Kolaj O, Spada S, Robin S, Wall JG (2009) Use of folding modulators to improve heterologous protein
471 production in *Escherichia coli*. Microbial Cell Factories 8: 9.

472 Lemansky P, Bishop DF, Desnick RJ, Hasilik A, Vonfigura K (1987) Synthesis and Processing of Alpha-
473 Galactosidase A in Human Fibroblasts. Evidence for Different Mutations in Fabry Disease. J Biol Chem 262:
474 2062-2065.

475 Liao YD, Jeng JC, Wang CF, Wang SC, Chang ST (2004) Removal of N-terminal methionine from
476 recombinant proteins by engineered *E. coli* methionine aminopeptidase. Protein Sci 13: 1802-1810.

477 Liras A (2008) Recombinant proteins in therapeutics: haemophilia treatment as an example. International
478 archives of medicine 1: 4.

479 Martinez-Alonso M, Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A (2009) Learning about protein
480 solubility from bacterial inclusion bodies. Microb Cell Fact 8: 4.

481 Medigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin PN, Cheung F, Cruveiller S, D'Amico S, Duilio A,
482 Fang G, Feller G, Ho C, Mangenot S, Marino G, Nilsson J, Parrilli E, Rocha EPC, Rouy Z, Sekowska A,
483 Tutino ML, Vallenet D, von Heijne G, Danchin A (2005) Coping with cold: The genome of the versatile
484 marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. Genome Res 15: 1325-1335.

485 Miyamura N, Araki E, Matsuda K, Yoshimura R, Furukawa N, Tsuruzoe K, Shirotani T, Kishikawa H,
486 Yamaguchi K, Shichiri M (1996) A carboxy-terminal truncation of human alpha-galactosidase A in a
487 heterozygous female with Fabry disease and modification of the enzymatic activity by the carboxy-terminal
488 domain - Increased, reduced, or absent enzyme activity depending on number of amino acid residues deleted.
489 Journal of Clinical Investigation 98: 1809-1817.

490 Murby M, Cedergren L, Nilsson J, Nygren PA, Hammarberg B, Nilsson B, Enfors SO, Uhlen M (1991)
491 Stabilization of recombinant proteins from proteolytic degradation in *Escherichia coli* using a dual affinity
492 fusion strategy. Biotechnol Appl Biochem 14: 336-346.

493 Nakamoto T (2009) Evolution and the universality of the mechanism of initiation of protein synthesis. Gene
494 432: 1-6.

495 Nettleship JE, Assenberg R, Diprose JM, Rahman-Huq N, Owens RJ (2010) Recent advances in the
496 production of proteins in insect and mammalian cells for structural biology. Journal of Structural Biology
497 172: 55-65.

498 Nozach H, Fruchart-Gaillard C, Fenaille F, Beau F, Ramos OHP, Douzi B, Saez NJ, Moutiez M, Servent D,
499 Gondry M, Thai R, Cuniasse P, Vincentelli R, Dive V (2013) High throughput screening identifies disulfide
500 isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in *E.
501 coli*. Microb Cell Fact 12: 37.

502 Nudler E, Gottesman ME (2002) Transcription termination and anti-termination in *E. coli*. Genes Cells 7:
503 755-768.

504 Panda AK (2003) Bioprocessing of therapeutic proteins from the inclusion bodies of *Escherichia coli*.
505 Advances in biochemical engineering/biotechnology 85: 43-93.

506 Parrilli E, De VD, Cirulli C, Tutino ML (2008) Development of an improved *Pseudoalteromonas*
507 *haloplanktis* TAC125 strain for recombinant protein secretion at low temperature. Microb Cell Fact 7: 2.

508 Piette F, D'Amico S, Struvay C, Mazzucchelli G, Renaut J, Tutino ML, Danchin A, Leprince P, Feller G
509 (2010) Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic
510 bacterium *Pseudoalteromonas haloplanktis* TAC125. Mol Microbiol 76: 120-132.

511 Prinz WA, Aslund F, Holmgren A, Beckwith J (1997) The role of the thioredoxin and glutaredoxin pathways
512 in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. J Biol Chem 272: 15661-15667.

513 Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression
514 systems: a review of the existing biotechnology strategies. Molecular and Cellular Biochemistry 307: 249-
515 264.

516 Saito S, Ohno K, Sakuraba H (2013) Comparative study of structural changes caused by different
517 substitutions at the same residue on alpha-galactosidase A. PLoS One 8: e84267.

518 Sallach RE, Conticello VP, Chaikof EL (2009) Expression of a recombinant elastin-like protein in *Pichia*
519 *pastoris*. Biotechnol Prog 25: 1810-1818.

520 Savitsky P, Bray J, Cooper CD, Marsden BD, Mahajan P, Burgess-Brown NA, Gileadi O (2010) High-
521 throughput production of human proteins for crystallization: The SGC experience. Journal of Structural
522 Biology 172: 3-13.

523 Tascon RI, Rodriguez-Ferri EF, Gutierrez-Martin CB, Rodriguez-Barbosa I, Berche P, Vazquez-Boland JA
524 (1993) Transposon Mutagenesis in *Actinobacillus pleuropneumoniae* with A Tn10 Derivative. J Bacteriol
525 175: 5717-5722.

526 Torres LL, Ferreras ER, Cantero A, Hidalgo A, Berenguer J (2012) Strategies for the recovery of active
527 proteins through refolding of bacterial inclusion body proteins. Microb Cell Fact 11: 105.

528 Vallejo LF, Rinas U (2004) Strategies for the recovery of active proteins through refolding of bacterial
529 inclusion body proteins. Microb Cell Fact 3: 11.

530 Vigentini I, Merico A, Tutino ML, Compagno C, Marino G (2006) Optimization of recombinant human nerve
531 growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. J Biotechnol 127:
532 141-150.

533 Wacker M, Feldman MF, Callewaert N, Kowarik M, Clarke BR, Pohl NL, Hernandez M, Vines ED, Valvano
534 MA, Whitfield C, Aebi M (2006) Substrate specificity of bacterial oligosaccharyltransferase suggests a
535 common transfer mechanism for the bacterial and eukaryotic systems. Proc Natl Acad Sci U S A 103: 7088-
536 7093.

537 Xiao Q, Zhang F, Nacev BA, Liu JO, Pei D (2010a) Protein N-terminal processing: substrate specificity of
538 *Escherichia coli* and human methionine aminopeptidases. Biochemistry-US 49: 5588-5599.

539 Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccosanti C, Conover K, Everett JK, Hamilton K,
540 Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma LC, Maglaqui M, Mao L, Mitra S, Patel
541 D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna G, V, Tong SN, Wang D, Wang H, Zhao L, Montelione
542 GT, Acton TB (2010b) The high-throughput protein sample production platform of the Northeast Structural
543 Genomics Consortium. Journal of Structural Biology 172: 21-33.

544 Yasuda K, Chang HH, Wu HL, Ishii S, Fan JQ (2004) Efficient and rapid purification of recombinant human
545 alpha-galactosidase A by affinity column chromatography. Protein Expression and Purification 37: 499-506.
546
547
548

549 **Figures**

550

551 **Fig. 1** (a) Expression vectors used in recombinant *h*GLA production assays. Expression vectors used in *E. coli*
552 strains: a) to d). Expression vector used in *Pseudoalteromonas haloplanktis* is shown in e). GST: Glutathione
553 S-transferase, GLA: *h*GLA gene, GLAopt: *h*GLA gene codon optimised for *E. coli* in c) and gene codon
554 optimised for *Pseudoalteromonas haloplanktis* in e). *Tobacco Etch Virus* protease cleavage site (TEVp)
555 location is marked by an arrow tip. His-tag is marked by a stripped box. Signal peptide of *h*GLA in a) is
556 marked by a stippled box. (b) Amino acid sequence of *h*GLA. The first 31 amino acids correspond to the
557 peptide signal that is marked in italics. The residues coded by rare codons are shown in bold. Putative
558 translation termination sequence is underlined.

559

560 **Fig. 2** Crude extracts of induced *E. coli* strains BL21(DE3) and Rosetta(DE3), not transformed (-) or
561 transformed with pReceiver-B01-GLA (+). Samples were collected 3 hours post-induction at 30°C, and
562 equivalent amounts of protein sample were loaded in each lane. (a) Western-blot developed with polyclonal
563 anti-GLA from Sigma. (b) Western-blot developed with monoclonal anti-His tag. Molecular weight marker
564 standards are indicated in kDa (Dual color, BioRad).

565

566 **Fig. 3** (a) Chromatogram of recombinant GST-GLA protein on GSTrap FF 1 ml column obtained from
567 soluble cellular fraction of Rosetta-gami B(DE3) induced cell culture. (b) Coomassie blue stained SDS-PAGE
568 of soluble, insoluble cell fractions and protein samples of affinity chromatography. Flow through (FT), wash
569 (W), fractions 5-10 of elution peak. Molecular weight marker standards are indicated in kDa (Dual color,
570 BioRad).

571

572 **Fig. 4** Detection of GST-GLA protein in Rosetta-gami B(DE3)/pGEX4T2-Opt-GLA induced cell cultures by
573 Western blot analysis using (a) polyclonal anti-GLA antibody (Sigma), (b) polyclonal anti-GST, (c)
574 monoclonal anti-His. NC: non-induced cell cultures, C: purified GST-GLA protein, T: total cellular fraction,
575 S: soluble fraction, I: insoluble fraction. Molecular weight marker standards are indicated in kDa (Dual color,
576 BioRad).

577

578 **Fig. 5** Detection of GST-GLA-GFP protein in Rosetta-gami B(DE3)/pGEX4T2-GLA-GFP induced cell
579 cultures by Western blot analysis. (a) polyclonal anti-GLA antibody (Sigma). (b) polyclonal anti-GFP
580 antibody. F: pooled positive protein fractions obtained in GST affinity chromatography, W: wash fraction,
581 FT: Flow through, S: soluble cell fraction, I: insoluble cell fraction. (c) Release of *h*GLA protein from
582 purified GST-GLA-GFP recombinant protein by Tobacco Etch Protease cleavage. Cli: Initial protein sample
583 after protein purification. Cl0: protein sample at time 0. Cl1-Cl4: protein sampl at 1-4 hours. Molecular
584 weight marker standards are indicated in kDa (Dual color, BioRad).

585

586 **Fig. 6** (a) Coomassie blue stained SDS-PAGE of soluble cell fractions of *P. haloplanktis* transformed with
587 pPM13psDs-GLA at different temperatures. (b) Western blot analysis of the same samples using a polyclonal
588 anti-GLA antibody (Santa Cruz Biotecnology). (c) Detection of recombinant *h*GLA in induced *P.*
589 *haloplanktis* cell cultures. T: total cell extract, P: periplasm, C: cytoplasm, NC: negative control. (d)
590 Coomassie blue stained SDS-PAGE of *h*GLA purification by affinity chromatography. T: total cell extract, 1:
591 unbound protein, 2 and 3: washes, 4 eluted protein.