Exploring three different expression systems for recombinant expression of globins: Escherichia coli, Pichia pastoris and Spodoptera frugiperda

An Brackea, David Hoogewijsb, Sylvia Dewildeab

aDepartment of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, Antwerp 2610, Belgium
bDepartment of Medicine/Physiology, University of Fribourg, Chemin du Musée 5, CH 1700 Fribourg, Switzerland

ARTICLE INFO

Keywords:
Globin
Protein expression
Androglobin
Neuroglobin

ABSTRACT

Globins are among the best investigated proteins in biological and medical sciences and represent a prime tool for the study of the evolution of genes and the structure-function relationship of proteins. Here, we explore the recombinant expression of globins in three different expression systems: Escherichia coli, Pichia pastoris and the baculovirus infected Spodoptera frugiperda. We expressed two different human globin types in these three expression systems: I) the well-characterized neuroglobin and II) the uncharacterized, circular permutated globin domain of the large chimeric globin androglobin. It is clear from the literature that E.coli is the most used expression system for expression and purification of recombinant globins. However, the major disadvantage of E. coli is the formation of insoluble aggregates. We experienced that, for more complex multi-domain globins, like the chimeric globin androglobin, it is recommended to switch to a higher eukaryotic expression system.

Introduction

Globins can be defined as small heme-containing proteins which can mostly reversibly bind oxygen. They occur in the three kingdoms of life: Archaea, Prokaryotes and Eukaryotes and exist as single domain or chimeric proteins [1,2]. The best known single domain globins are the vertebrate hemoglobin (Hb) and myoglobin (Mb), playing respectively a role in oxygen transport and storage. Examples of chimeric globins are the flavohemoglobins (FHbs) in Escherichia coli and yeast, consisting of an N-terminal globin domain coupled to a ferredoxin reductase-like domain [3], and the globin coupled sensors (GCS), reported in Bacteria and Archaea, comprising an N-terminal globin domain linked to variable gene regulatory domains [4]. FHbs and GCS do not occur in vertebrates. The recent and rapid accumulation of genomic information has resulted in a substantial increase in newly recognized globins. Bioinformatic genome surveys revealed that more than 50% of the bacterial, approximately 20% of the archaeal and almost 90% of the eukaryote genomes contain genes encoding globins [1,5-7]. Despite their great variety in function and amino acid sequence, most of them share a similar three-dimensional structure, referred to as the globin fold, characterized by a three-over-three alpha helical folding. Few globins, however, display a two-over-two alpha helical folding and are called ‘truncated globins’ [8].

Globins are among the most investigated proteins in biological and medical sciences and represent a major tool for the study of gene evolution and structure-function relationship of proteins. Sperm whale Mb was the first protein of which the three-dimensional structure was revealed by X-ray crystallography [9]. The structure of human Hb followed five years later [10]. In the early times of protein crystallization and structure determination, the investigated proteins were purified from their natural context, e.g. Mb out of sperm whale muscle tissue and Hb out of human erythrocytes. However, under normal physiological conditions, most proteins are present in low cellular concentrations and cannot be purified out of the tissue itself. Since the early eighties it has been possible to express proteins recombinantly in heterologous host cells like the bacterium Escherichia coli (E.coli) [11,12] or the yeast Saccharomyces cerevisiae (S.cerevisiae) [13]. These host cells can produce a large amount of recombinant protein, that enables the characterization of proteins with a low cellular abundance. In addition, this technique facilitates the study of protein-function relationships by site-directed mutagenesis. This revolution in protein biochemistry caused an exponential increase in knowledge on the structure and function of proteins in general.

Neuroglobin (Ngb) was first described in 2000 as a vertebrate globin, preferentially expressed in brain and other nerve tissue [14]. The estimated amount of Ngb in the vertebrate brain under non-pathological conditions is only in the micromolar range. It was possible to perform a biochemical and structural characterization of Ngb subsequent to the recombinant expression of the protein in E.coli [15].

Androglobin (Adgb) was described in 2012 as the fifth mammalian
globin type (in addition to Hb, Mb, Ngb and cytoglobin (Cygb)) and it displays predominant expression in testis tissue in mice and human [16]. Adgb characteristically comprises four domains, an N-terminal ∼350 residues calpain-like cysteine protease domain, a region of ∼300 residues without known motifs/domains, followed by an ∼150 residue circularly permuted globin domain, and a second, uncharacterized ∼750 residues C-terminal domain. Interestingly, the globin domain, which normally consists of eight consecutive α-helices (named A-H), is circularly permuted and split into two parts within Adgb. The part containing helices A and B has been shifted in the C-terminal direction and is separated from the main globin sequence (helices C-H) by a calmodulin-binding IQ motif. Alignments of the split Adgb globin domain with mammalian Mb, Ngb, and Cygb sequences revealed that Adgb—despite its rearrangement—conforms to the criteria of the “globin fold” tertiary structure [16]. This was confirmed by molecular modelling of the human Adgb (ADGB) globin domain 3D structure, showing that the helix C-H segment alone is able to produce a bona fide globin fold [16]. In vitro structural and functional analysis of ADGB is necessary to better understand its molecular function during spermatogenesis, for which a recombinant form of ADGB is required.

In this paper, we explore three different expression systems for the expression of globins: a bacterial (E.coli), a yeast (Pichia pastoris (P.pastoris)) and an insect cell expression system (baculovirus infected Spodoptera frugiperda (Sf9)). These three expression systems are widely used and in previous studies they have been compared for their recombinant expression of e.g. G-protein-coupled receptors [17], endostatin [18] and rabbit liver carboxylesterases [19]. We investigate their advantages and disadvantages for the expression of globins in general. Furthermore, we used these three different expression systems to express the two different human globin types that were described above: I) the single domain globin neuroglobin (NGB), which has already been biochemically and structurally characterized and II) the uncharacterized chimeric globin CH-IQ-AB.

Material and methods

Construction of the recombinant expression vectors

The cDNA sequence of the globin domain of human ADGB (referred to as CH-IQ-AB) and human Ngb were cloned into the recombinant expression vectors pET23a (E.coli), pPicZa (P.pastoris) and pDEST10 (Spodoptera frugiperda 9 (Sf9)). The cDNA of full length ADGB was only cloned into the pDEST10 vector (Table 1). Full length ADGB was RT-PCR amplified, with Ncol and Xbal overhangs (Table 2) from human testis cDNA and cloned into the pENTR4 vector. The CH-IQ-AB domain was PCR-amplified for further cloning from the pENTR ADGB vector. The NGB domain was PCR-amplified from pET3a Ngb construct, which was already available in our laboratory [15]. For the cloning of CH-IQ-AB in the pPicZa vector, the cDNA of the fragment was synthesized by Invitrogen GeneArt, with optimized codons for expression in P.pastoris.

The amplifications of the fragments coding for the CH-IQ-AB and Ngb were performed by PCR. PCR amplification was performed as follows: initial denaturation at 94 °C for 5 min, followed by 35 cycles (94 °C 1 min; 55 °C 1 min; 72 °C 1 min) and a final elongation at 72 °C for 10 min. The forward and reverse primers sequences, including restriction sites/attB recombination sites, were in house designed and are listed in Table 2. Recombinant pET23a CH-IQ-AB, pET23a Ngb, pPicZa CH-IQ-AB and pPicZa Ngb were created using standard restriction enzyme cloning. Two mutations were made in the pET23a CH-IQ-AB construct resulting in the replacement of two cysteines (on position C3 and E2) using the QuikChange™ site-directed mutagenesis method (Stratagene). The recombinant pDEST10 CH-IQ-AB and pDEST10 Ngb vector were created using the Gateway cloning technique: first a pENTR CH-IQ-AB and pENTR Ngb vector were created by a BP recombination reaction between attB CH-IQ-AB/Ngb PCR product and pDONR221 vector and second the pDEST10 Ngb and pDEST10 CH-IQ-AB were created by a LR recombination reaction between pENTR CH-IQ-AB/Ngb and the pDEST10 vector. The pDEST10 ADGB (full length) was created using the Gateway cloning technique: LR recombination reaction between pENTR4 ADGB and the pDEST10 vector. The constructed recombinant expression vectors were verified via direct sequencing.

Expression in Escherichia coli

pPET23a CH-IQ-AB and pPET23a Ngb were transformed into competent E.coli BL21(DE3)pLysS cells. Subsequently, cells were grown in a shaking incubator (160 rpm) at 25 °C in TB medium (1.2% bacitryptone, 2.4% yeast extract, 0.4% glycerol, 72 mM potassium phosphate buffer, pH 7.5) containing 200 µg/ml ampicillin, 30 µg/ml chloramphenicol, and 1 mM δ-aminolevulinic acid. The expression of the C-terminal His-tagged fusion proteins was induced at A550nm = 0.8 by the addition of isopropyl-1-thio-D-galactopyranoside (IPTG) to a final concentration of 0.4 mM, and expression was continued overnight. The cells were harvested and resuspended in ∼1/20 of the initial volume (resuspension buffer = 50 mM Tris-HCl pH 8.0, 300 mM NaCl). The cells were then exposed to three freeze-thaw steps and were sonicated until completely lysed. The extract was clarified by centrifugation (12000g, 10min, 4 °C) and both supernatant and pellet fraction were screened for expression of recombinant protein (CH-IQ-AB, Ngb) using SDS-PAGE.

Expression in Pichia pastoris

pPicZa CH-IQ-AB and pPicZa Ngb were linearized with restriction endonuclease SacI and subsequently 5 µg plasmid was transformed into P.pastoris X33 electrocompetent cells by electroporation: cells were pulsed using the manufacturer’s instructions for Saccharomyces cerevisiae (Bio-rad electroporation system). Transformed cells were plated out on YPSD agar plates (1% yeast extract, 2% peptone, 2% dextrose, 1M sorbitol, 2% agar) containing 100 µg/ml Zeocin and were incubated 4 days at 30 °C. The Mut Phenotype was determined by streaking the colonies out on both MD (1,34% YNB, 4 × 10−5% biotin, 2% dextrose) and MM (1,34% YNB, 4 × 10−5% biotin, 0,5% methanol) agar plates containing 100 µg/ml Zeocin. 8 Mut− colonies, of which the presence of the insert was confirmed by PCR using the recommended sequencing primers, were inoculated in 25 ml BMGY medium (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH6.0, 1,34% Yeast Nitrogen Base, 4 × 10−5% biotin, 1% glycerol) and grown overnight at 30 °C in a shaking incubator (200 rpm) until they reached an OD600 of ∼0.5. Subsequently, the cells were harvested by centrifugation (2000g, 10min, RT) and the cell pellet was resuspended to an OD600 of 1 in BMMY medium (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH6.0, 1,34% Yeast Nitrogen Base, 4 × 10−5% biotin, 0,5% methanol) containing 1 mM δ-aminolevulinic acid and further grown at 30 °C in a shaking incubator (200 rpm). Methanol will induce the expression of the C-terminal His-tagged proteins, so every 24 h 100 µmol methanol was added to final concentration of 0.5% to maintain induction. 1 ml of culture expression was collected every 24 h to analyze expression levels with Western blot. For this, the cells were centrifuged (4000g, 20min, RT); the supernatant was screened for secreted expression of recombinant protein; the cell pellet was resuspended in SDS-loading buffer and screened for intracellular expression. After 96 h all

Table 1 Overview of the different expression vectors used in this paper.

<table>
<thead>
<tr>
<th>Vector</th>
<th>Tag</th>
<th>Promoter</th>
<th>Expression host</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET23a</td>
<td>6xHis, C-terminal</td>
<td>T7</td>
<td>Escherichia coli BL21(DE3)pLysS</td>
</tr>
<tr>
<td>pPicZa</td>
<td>6xHis, C-terminal</td>
<td>AOX</td>
<td>Pichia pastoris X33</td>
</tr>
<tr>
<td>pDEST10</td>
<td>6xHis, N-terminal</td>
<td>Polyhedrin</td>
<td>Spodoptera frugiperda Sf9</td>
</tr>
</tbody>
</table>

http://doc.rero.ch
Expression in Spodoptera frugiperda 9 insect cells

pDEST10 CH-IQ-AB, pDEST10 NGB and pDEST10 ADG were transformed into MAX Efficiency™ DH10Bac™ competent E.coli (ThermoScientific™) to generate recombinant baculovirus DNA. The presence of the insert in the baculovirus DNA was verified by PCR using the recommended sequencing primers. 1 μg of pure recombinant baculovirus DNA was subsequently transfected into Sf9 insect cells (ThermoScientific™) using Cellfectin™ (Invitrogen) or anti-ADGB antibody (polyclonal anti-ADGB goat antibodies). To verify a correct heme-incorporation of the purified recombinant globin, UV/visible spectra were taken (200 nm–700 nm). Spectral measurements were made with a spectrophotometer (Shimadzu UV-2101PC). The ferrous deoxy sample was obtained by adding an excess of sodium dithionite.

SDS-page and Western blot analysis

Samples were first denatured and reduced by the addition of an equal amount of loading buffer (4% SDS, 10% v/v β-mercaptoethanol, 125 mM Tris-HCl pH 6.8, 20% glycerol) and by incubating them at 95 °C for 5min. Subsequently, equal mass samples were separated by SDS-PAGE (15% acrylamide gel for NGB and CH-IQ-AB and 7.5% acrylamide for ADG). For SDS-PAGE analysis, proteins were stained with Coomassie Brilliant Blue G-250. For Western blot analysis, proteins were transferred to polyvinylidene difluoride (PVDF) membranes according to the manufacturer’s instructions (Immobilon™-P PVDF membrane, pore size 0.45 μm, Millipore). The membranes were probed with mouse-derived anti-His Tag antibody (6x-His Tag Monoclonal Antibody (4E3D10H2/E3)Invitrogen) or anti-ADGB antibody (polyclonal anti-ADG antibody HPA036540 SIGMA) (1:5000 in 5% Milk TBS buffer); overnight at 4 °C, followed by an incubation at room temperature for 1 h with an HRP-conjugated antibody (1:5000 in TBS) (Polyclonal Goat anti-mouse/rabbit Immunoglobulins – HRP, DAKO). Subsequently, Luminata Forte Western HRP substrate (Millipore) was added to the membrane and the targeted proteins were visualized using a SynGene imager.
Results and discussion

Recombinant expression in *Escherichia coli*

The by far most popular host for recombinant protein expression is *E. coli*: 86% of the protein structures entered in the protein databank (PDB) [20] were expressed recombinantly in *E. coli*. This can be explained by the fact that, for a large scale protein production, the *E. coli* system is the easiest, quickest and cheapest method. There are many commercial and non-commercial expression vectors available with different N- and C-terminal tags and many different *E. coli* strains, which are optimized for special applications. The first globins recombinant expressed in *E. coli* were human Mb [21] and Hb [12], followed rapidly by many other globin types of many different organisms. In most cases, globins are small globular proteins, with no specific requirements in folding or post-translational processing and thus *E. coli* is often the perfect host for their large scale recombinant expression.

However, the high degree of accumulation of soluble protein during recombinant expression is not always accepted by the metabolic system of the host. Sometimes, the recombinant expressed protein is toxic to the host cell and causes a cellular stress response, which results in poor and irregular protein expression. An example of such a situation is the expression of microbial FHbs. These chimeric globins possess the ability to generate superoxide and peroxide in the presence of reductant and NAD(P)H and the accumulation of these reactive oxygen species are toxic to the host cell. These problems can be overcome by expressing the FHbs under the control of an inducible promoter. Lewis et al. used the pBAD expression system, which allows tightly controlled, titratable expression of a protein through the regulation of arabinose concentration [22].

Another often encountered problem is the accumulation of target proteins into insoluble aggregates known as inclusion bodies. These proteins are misfolded and biologically inactive. Aggregation prone proteins require the existence of a number of molecular chaperones that interact reversibly with nascent polypeptide chains to prevent aggregation during the folding process. The formation of inclusion bodies could therefore result either from accumulation of high concentrations of folding intermediates or from inefficient processing by molecular chaperones [23]. Refolding from inclusion bodies is in many cases considered undesirable as it results most of the time in a poor recovery of bio-active protein [24]. Most used strategies for shifting the recombinant expression from inclusion bodies to soluble protein are: changing the expression temperature, media and transcription rates, trying various *E. coli* strains, including molecular chaperones, including tRNA complementation plasmids, expressing fragments, using fusion technology, stabilizing mRNA or screening for soluble variants. For further reading on these strategies we refer to the review paper of Sorensen et al. [23].

The formation of inclusion bodies during recombinant globin expression is a frequently encountered problem. The first recombinant

---

**Fig. 1.** A) SDS-PAGE (15%): Recombinant expression in *E. coli* of CH-IQ-AB (∼25 kDa) and NGB (∼18 kDa). I) *E. coli* total cell lysate II) Insoluble fraction of cell lysate III) Soluble fraction of cell lysate IV) Fraction after native purification with Nickel column. B) Optical absorption spectra of NGB expressed in *E. coli* (pET23a) and C) optical absorption spectra of refolded CH-IQ-AB, expressed in *E. coli* (pET23a).
expressed globins (Mb and Hb) were expressed as inclusion bodies and refolded in vitro into their functional form [12,21]. Nevertheless, by changing expression conditions most globins can be expressed and purified as soluble proteins [25].

Expression of NGB and CH-IQ-AB in E.coli

For the recombinant expression of human NGB and the globin domain of human ADGB (referred to as CH-IQ-AB) in E.coli we cloned them into the pET23a expression vector, which uses the strong inducible T7 promoter. The recombinant pET23a constructs were subsequently transformed in BL21(DE3)pLysS E.coli, where the recombinant expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). pET23a expresses the proteins with a C-terminal His-tag, which binds to several types of immobilized metal ions, like nickel. His-tagged proteins can thus easily be purified using nickel affinity chromatography. Both the soluble and insoluble fractions of the cell lysate were screened using SDS-PAGE for the recombinant over-expression of NGB and CH-IQ-AB (Fig. 1A).

NGB was expressed in a soluble form and was purified with nickel affinity chromatography. After purification of NGB, optical absorption spectra were recorded, which were characteristic for NGB [15]. NGB is a hexacoordinated globin which means that under deoxy conditions the protein forms two coordinative bonds with the iron atom of the heme group. The optical absorption spectrum of a hexacoordinated heme group has large amplitudes for the alpha band (562 nm) and the Soret band (428 nm). This observation indicates that the heme group was correctly incorporated in the polypeptide chain of NGB (Fig. 1B), consistent with a previous report [15].

CH-IQ-AB was expressed in E.coli as insoluble aggregates, also known as inclusion bodies and it was not possible to purify CH-IQ-AB as a soluble protein even after alternating the conditions (Fig. 1A). Different strategies were used to shift the expression from inclusion bodies to stable soluble protein. By changing time of induction and lowering the temperature during protein expression we were able to avoid inclusion body formation, however CH-IQ-AB remained unstable and precipitated over time. We tried to increase the solubility of the globin domain by the addition of reducing agents or detergents and by changing the pH or salt concentrations of the resuspension buffer. Because CH-IQ-AB possesses two cysteine residues, which can possibly cause intermolecular disulfide bridge formation and hence precipitation, we mutated the cysteines into serines. Despite these efforts, CH-IQ-AB remained unstable. As such, CH-IQ-AB was expressed as inclusion bodies, purified under denaturing conditions (8M urea) and the recombinant protein was refolded in vitro by dialysis and adding free heme, as previously done for human CYGB [26]. In contrast to the predictions of the computer three-dimensional modeling [16] the refolded globin domain was not stable and precipitated over time. Despite the precipitation, we were able to measure an UV-visible spectrum of the refolded globin domain. The optical absorption spectra of the refolded ferrous deoxygenated CH-IQ-AB domain displayed the typical spectra of a hexacoordinated globin (high α band (557 nm) and Soret band (421 nm), low beta band (527 nm)), like NGB (Fig. 1C). However, the purification of the refolded CH-IQ-AB was not efficient and only low yields could be obtained. The refolded globin domain remained unstable, precluding biological characterizations, including ligand binding kinetic determination and crystallization.

Folding studies of apomyoglobin (apoMb) [27] and plant hemoglobin [28] showed that the folding of a globin happens based on the contact order of not only local-residue contacts (as is seen with very small proteins), but also on long-range amino acid interactions. A stepwise-folding process of apoMb has been suggested in the following order; helices G, A, H, B, E, C and D. During this folding pathway, there are two folding intermediates, if one of these intermediates is misfolded, aggregation occurs. We can imagine that the folding of the circular permutated globin of ADGB, where helices A and B are placed after helices C-H, is more complicated than the folding of a typical globin with a normal consecution of the helices, like NGB.

Interestingly, the effect of circular permutation in Mb was investigated by Ribeiro et al. [29]. A circular permutated Mb was engineered, where helices A and B were shifted towards the C-terminus (similar to the globin domain of ADGB). Subsequently, this protein was expressed in E.coli as inclusion bodies and it was afterwards refolded in vitro. The refolded circular permutated myoglobin displayed absorption spectra similar to the WT Mb. However, the circular permutation caused a significant decrease in the stability of the protein and a large fraction formed stable aggregates, similar to our experience with the circular permutated globin domain of ADGB.

Most plausible, the globin domain of ADGB needs long-range interactions from the amino acid sequence of the full-length ADGB to form a stable globin fold. Another possibility is that the human globin domain is simply not able to fold correctly in the prokaryotic E.coli system. Therefore, we further investigated whether the globin domain can be natively folded in a eukaryotic expression system.

Recombinant expression in Pichia pastoris

E.coli as expression system might be the obvious first choice, but it is not always the best choice, as we have seen for CH-IQ-AB. E.coli is a prokaryotic system and does not contain the subcellular machinery of eukaryotes needed for post-translational modifications, including proteolytic processing, folding, disulfide bond formation, glycosylation and phosphorylation. Many eukaryotic proteins need to be modified following translation in order to become active and/or adapt the proper structure. Yeast is an eukaryotic organism and, similar to E.coli, can be grown to very high densities, which makes them especially useful for large scale protein production. The two most common yeast strains used for recombinant protein expression are S.cerevisiae and the methotlycotic yeast P.pastoris.

Only few cases are reported where globins are recombinantly expressed in a yeast system [30–32]. An example of a globin that has been successfully expressed in P.pastoris is the Crocodylus siamensis hemoglobin. This crocodile Hb displays potent antibacterial and anti-inflammatory activity, as well as strong antioxidant properties [33–35]. Furthermore, it comprises six cysteine residues in the α-globin chain and three cysteine residues in the β-globin chain [35]. Each of these cysteine residues can form disulfide bonds both of the intermolecular and intramolecular formation. The expression in E.coli resulted in low expression yields and precipitation after purification [32]. To overcome these problems the crocodile Hb was successfully expressed in P.pastoris.

As the CH-IQ-AB globin domain of ADGB displays similar problems of precipitation after purification and comprises also three cysteine residues (two in the α-helices and one in the IQ motif). We recombinantly expressed CH-IQ-AB in the P.pastoris system.

Expression of NGB and CH-IQ-AB in P.pastoris

For the expression of NGB and CH-IQ-AB we chose for methotlycotic yeast P.pastoris because it is widely used, cost-effective and suitable for large-scale protein production. cDNA of NGB and CH-IQ-AB were cloned into the pPIC3a vector, which places a native S.cerevisiae α-secretion signal before the recombinant protein. The advantage of a secreted expression is that P.pastoris secretes very low levels of native proteins, facilitating further purification steps. The pPIC3a vector uses the AOX promoter for a high-level, methanol-induced expression. After induction of recombinant expression, every 24 h a sample was taken for analysis of expression of CH-IQ-AB and NGB using western blot. NGB is expressed and secreted by P.pastoris (Fig. 2A). 96 h post-induction, NGB was purified out of the expression medium and optical absorption spectra were taken (Fig. 2B). These spectra evidence proper folding and incorporation of heme in NGB in line with previous descriptions [31]. Despite our efforts to optimize the cDNA sequence of CH-IQ-AB to the codon usage of P.pastoris, we could not detect recombinant expression.

Expression of NGB and CH-IQ-AB in P.pastoris

For the expression of NGB and CH-IQ-AB we chose for methotlycotic yeast P.pastoris because it is widely used, cost-effective and suitable for large-scale protein production. cDNA of NGB and CH-IQ-AB were cloned into the pPIC3a vector, which places a native S.cerevisiae α-secretion signal before the recombinant protein. The advantage of a secreted expression is that P.pastoris secretes very low levels of native proteins, facilitating further purification steps. The pPIC3a vector uses the AOX promoter for a high-level, methanol-induced expression. After induction of recombinant expression, every 24 h a sample was taken for analysis of expression of CH-IQ-AB and NGB using western blot. NGB is expressed and secreted by P.pastoris (Fig. 2A). 96 h post-induction, NGB was purified out of the expression medium and optical absorption spectra were taken (Fig. 2B). These spectra evidence proper folding and incorporation of heme in NGB in line with previous descriptions [31]. Despite our efforts to optimize the cDNA sequence of CH-IQ-AB to the codon usage of P.pastoris, we could not detect recombinant expression.
of CH-IQ-AB in *P. pastoris* (Fig. 2A), neither in the secreted fraction nor in the intracellular fraction.

This may be explained by the possibility that CH-IQ-AB cannot be expressed as a stable soluble form because it requires the context of the full length ADGB protein for proper folding as was previously suggested. In *E. coli*, misfolded proteins are stored as inclusion bodies. In eukaryotic systems, protein folding occurs in a different way; proteins are folded and processed in the endoplasmic reticulum (ER) and/or the Golgi apparatus and sometimes secreted into the extracellular environment through vesicular transport (as is in the case of recombinant expression of CH-IQ-AB in *P. pastoris*). In eukaryotes aberrant folding properties of the target protein and/or high level production can lead to the accumulation of unfolded or even aggregated protein in the ER, which can initiate the unfolded protein response (UPR) and ER-associated degradation [36–40]. This can thus possibly explain the fact that we cannot detect expression of CH-IQ-AB in *P. pastoris*.

**Recombinant expression in Sf9 insect cells**

Finally, we opted for a third expression system, one that is much closer to mammalian cells; the baculovirus-insect cell system. This expression system is a binary system, consisting of a recombinant baculovirus as the vector and lepidopteran insect cells as the host [41]. In eukaryotes the insect cells can carry out more complex post-translational modifications and have a better folding machinery for mammalian proteins. Furthermore, baculoviral vectors can accommodate large DNA inserts, which allows for the production of multi-subunit protein complexes or even high molecular weight virus-like particles [42–44]. A potential disadvantage of the baculovirus-insect cell system is that it is a high-cost and time-consuming method of recombinant protein expression.

So far, there are no reports of globins recombinantly expressed in insect cells. Since the baculovirus-insect cell expression system is widely used for the recombinant expression of high molecular weight proteins, we used it to express the 190 kDa full length ADGB protein.

Expression of NGB, CH-IQ-AB and full length ADGB in Sf9 cells

NGB, CH-IQ-AB and full length ADGB cDNA were cloned into the pDEST10 vector, which possesses the strong transcriptional promoter derived from the very late baculovirus gene *polyhedrin*, able to produce milligrams of proteins. Sf9 insect cells were infected with recombinant NGB, CH-IQ-AB and ADGB baculovirus particles. 72 h after infection, the expression of recombinant protein was analyzed with western blot (Fig. 3). Expression was detected for NGB and ADGB. The expression level of CH-IQ-AB was however very low, again possibly reflecting that the globin domain would require the context of the full length ADGB protein for proper folding.

Surprisingly, for NGB multiple bands were detected with western blot analysis (Fig. 3A). It seems that NGB a) can be detected as dimer and b) appears in three apparent molecular mass forms, in Sf9 insect cells. The dimerization could not be prevented by heat denaturation or addition of reducing agents, like β-mercaptoethanol or DTT. Dimerization was also seen for NGB when it was expressed in *E. coli*, but not in *P. pastoris* (Fig. 4). In *P. pastoris* NGB is secreted into the medium, and is
known for their capabilities of protein N-glycosylation, we analyzed the cofactors or phosphorylation. Indeed, it has previously been displayed that ADGB can give rise to different apparent molecular mass forms after expression in the insect cell system. It seems that NGB only forms these different molecular mass forms after expression in the higher eukaryotic baculoviral insect cell system and not in E.coli or P.pastoris, assuming that the latter do not provide the cellular environment for formation of these different conformations or post-translational modifications.

In contrast to the CH-IQ-AB domain, the full length ADGB was significantly expressed in S99 cells, confirming the hypothesis that the full length protein is required for stable expression. However, we observed a time-dependent truncation of the full length protein (Fig. 3C) and even after scaling-up the expression, the amount of full length ADGB was insufficient for further biophysical characterization. This truncation is most probably a result of proteolysis by intracellular proteases or by auto-cleavage by the N-terminal calpain-like domain, a phenomenon commonly observed among calpains [53-56]. To further evaluate if ADGB autolytic activity would be the source of this potential auto-cleavage we added three different protease inhibitors to the growth medium; the cysteine protease inhibitor E64, the acid protease inhibitor Pepstatin A and the calpain specific Calpain II inhibitor. However, administration of these inhibitors did not prevent the truncation of ADGB. The ADGB calpain-like domain is homologous to calpain 7, for which no specific protease inhibitors exists. Interestingly, substantial autolytic activity has also been demonstrated for calpain-7 [57]. Further research is necessary to understand this putative auto-cleavage of ADGB.

Conclusion

In this paper we explored three different expression systems for the recombinant expression of globins; the prokaryote expression system E.coli and the two eukaryote expression systems P.pastoris and baculovirus infected S99. In general, globins are small globular proteins, which do not need post-translation modification for a proper folding and in this regard E.coli is the obvious first choice for recombinant expression. However, the major disadvantage of recombinant expression in E.coli is the formation of insoluble aggregates, inclusion bodies.

We attempted to express the circular permuted globin domain of ADGB in E.coli, P.pastoris and S99 cells, however, a soluble expression was not possible. This suggest that the globin domain of ADGB needs the full context of the full length protein in order to form a stable globin fold. Furthermore, we displayed that ADGB was truncated, possibly due to auto-cleavage by the N-terminal calpain-like domain.
to auto-cleavage by the internal calpain-like domain. Our study demonstrates that, in general, it is always possible to express globins in E.coli and that switching to a (higher) eukaryotic expression system might represent a valid alternative for more complex, multi-domain globins, like the chimeric globin ADGB.

When NGB was recombinantly expressed in insect cells, three apparent molecular mass forms and dimerization were observed, this was not observed in E.coli or P. pastoris. Interestingly, the same multiple band pattern on western blot has been observed in mitochondrial enriched fractions of mouse retina and in brain tissue. It seems that the baculoviral insect cell expression system provides a cellular environment for folding and processing of proteins that is more similar to the natural context of Ngb (brain and retina) compared to the cellular environments of E.coli or P. pastoris. In this regard it might be worthwhile to switch to a (higher) eukaryotic expression system for eukaryotic proteins in general, even if they are easily expressed in a bacterial or yeast expression system.

Funding

This work was supported by the Fund for Scientific Research (FWO) (11N5316N) and by the German Research Foundation (HO5837/1-1) and the Swiss National Science Foundation (31003A_173000).

Conflicts of interest

The authors do not have any conflict of interest.

Acknowledgments

M. L. Van Hauwaert and H. Berghmans (University of Antwerp) are gratefully acknowledged for technical assistance. We thank S. Santambrogio (University of Zürich) for providing the pENTR4 ADGB plasmid.

References


