

Molecular Chaperone Assisted Expression Systems: Obtaining Pure Soluble and Active Recombinant Proteins for Structural and Therapeutic Purposes

Xolani Henry Makhoba¹, Ofentse Jacob Poe² and Mathews Simon Mthembu^{1*}

¹Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, Private Bag X1001, Kwadlangezwa 3886. South Africa

²Biosciences, Council for Scientific and Industrial Research, Pretoria 0001, South Africa

Abstract

For many years recombinant protein production has been at the center of biosciences used for structural and therapeutic purposes. The production of recombinant proteins in foreign host system such as *E. coli* has been a biggest challenge. This has brought negative impact on the ongoing search for alternative drugs and vaccines for diseases such as schistosomiasis, African sleeping sickness and malaria. At present, no system is available to produce various recombinant proteins from different sources. Hence, the search for a universal system that can be utilized for any recombinant protein is crucial. The basic requirements for high quality protein production in a pure active and soluble form have led to the use of molecular chaperones as supplement during recombinant proteins production in *E. coli*. Molecular chaperones are proteins that are known to assist the newly synthesized proteins to complete their folding stages. This system has improved various proteins that are difficult to produce in *E. coli* host. However, some proteins have not been produced as pure, soluble or active proteins. For this reason this has led to some researchers to suggest the use of engineered host systems that do not produce endogenous molecular chaperones such as Hsp70 or Hsp40. However, these systems are not easy to work with because of their fragileness. Therefore, this work explores some progresses made as well as challenges faced in coming up with the desired system that can meet all the requirements to obtain the desired protein product.

Keywords: Molecular chaperones; Universal system; Recombinant proteins; Inclusion bodies and diseases

Introduction

Numerous proteins have previously reported to fold on their own *in vitro* and work done by [1], shows that linear polypeptides chain poses the required information which determines the three-dimensional structure of the protein. All the processes involved in protein folding have drawn a lot of attention for many years in the area of molecular biology. Amongst the other reasons is that most proteins lack the capability to fold on their own, thus aggregates and form inclusion bodies. Heterologous protein expression using the Gram-negative bacterium, *E. coli*, is an important and frequently used for genetic manipulations and for expressing recombinant proteins. Bacterial expression systems are widely used and favored by industrial and pharmaceutical companies because of their ability to rapidly produce recombinant proteins using low cost substrates [2]. Nonetheless, the overexpression of some recombinant proteins in *E. coli* may result in the formation of insoluble proteins [3] Most proteins including those from *Plasmodium* are expressed as insoluble proteins in inclusion bodies [4] Inclusion bodies primarily contain over-expressed non-native proteins and occurs as a result of differences between the codon usage of the *E. coli* and the protein of interest [5]. Mehlin [4] selected and attempted to express 1000 *P. falciparum* open reading frames using *E. coli* BL21 StarTM (DE3) cells. Only 6.3% of these cells were expressed as soluble proteins. The same group attempted to improve expression by a technique called “codon-optimisation” an approach where codons of the target gene were altered to suit the codon preference of expression host [2] Codon-optimisation had no significant influence on the production of *P. falciparum* proteins from *E. coli* [4] Consequently, less than 0.5% of all protein structures available on the Protein Data Bank were malarial proteins [6,7]. Furthermore, the production of misfolded or partially misfolded proteins should also be avoided as they often exhibit a lower activity as compared to monomeric folded proteins and consequently the expression of misfolded protein adds to

the production costs since they will require more purification steps and refolding procedures.

Conventional ways used to overcome inclusion bodies

Various buffer formulations using surfactants and amino acids (such as L-arginine) have previously been reported to suppress protein unfolding or aggregation [2]. However, the major restriction with the use of surfactants is that they tend to co-purify with the protein of interest. Although some studies have shown that the use of L-arginine suppresses protein aggregation [8], other researchers still argue that L-arginine may play a significant role on the protein thermodynamic stability [5,9]. The use of low incubation temperatures ($\leq 20^{\circ}\text{C}$) have also been identified as important factors in the expression of soluble proteins [10] Various studies have demonstrated that the proteins expressed at low temperatures are likely to be more soluble [11,12]. Wang and his colleagues [12] hypothesized that at low temperatures the expression rate of proteins is greatly reduced, thus increasing the folding efficiency of protein. It was proposed that endogenous proteases may be less-active at low temperatures, thus, allowing the expressed protein properly fold [2].

***Corresponding author:** Mathews S Mthembu, Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, South Africa, Tel: +2735 902 6098; Fax: +2735 902 6568; E-mail: mthembums@unizulu.ac.za

Received August 27, 2015; **Accepted** September 24, 2015; **Published** September 28, 2015

Citation: Makhoba XH, Poe OJ, Mthembu MS (2015) Molecular Chaperone Assisted Expression Systems: Obtaining Pure Soluble and Active Recombinant Proteins for Structural and Therapeutic Purposes. J Proteomics Bioinform 8: 212-216. doi:10.4172/jpb.1000371

Copyright: © 2015 Makhoba XH, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Codon bias refers to the high frequency and preferential use of a particular codon coding for an amino acid within an organism [13]. There is a high level of codon mismatch between the protein of interest and expression hosts, such as *E. coli* and two different strategies have been used to minimize codon bias. Firstly, the intracellular tRNA pool can be expanded by using plasmids which encode rare tRNAs used in certain organisms, for example *E. coli*. Secondly, altering the codons of the target genes to suit the codon preference of the non-natural expression host is a process referred to as 'codon-optimization'. This approach has met with some success, especially in antimalarial vaccine targets, but unfortunately the expression of several other malaria proteins was not improved through codon-optimization alone (Figure 1). As an example, Mehlh [4] reported that nine out of the twelve *P. falciparum* genes that had been synthesized with optimized codons could still not be expressed in *E. coli*, and the three proteins that were expressed were in inclusion bodies. This led them to conclude that codon usage did not have a significant impact on protein expression.

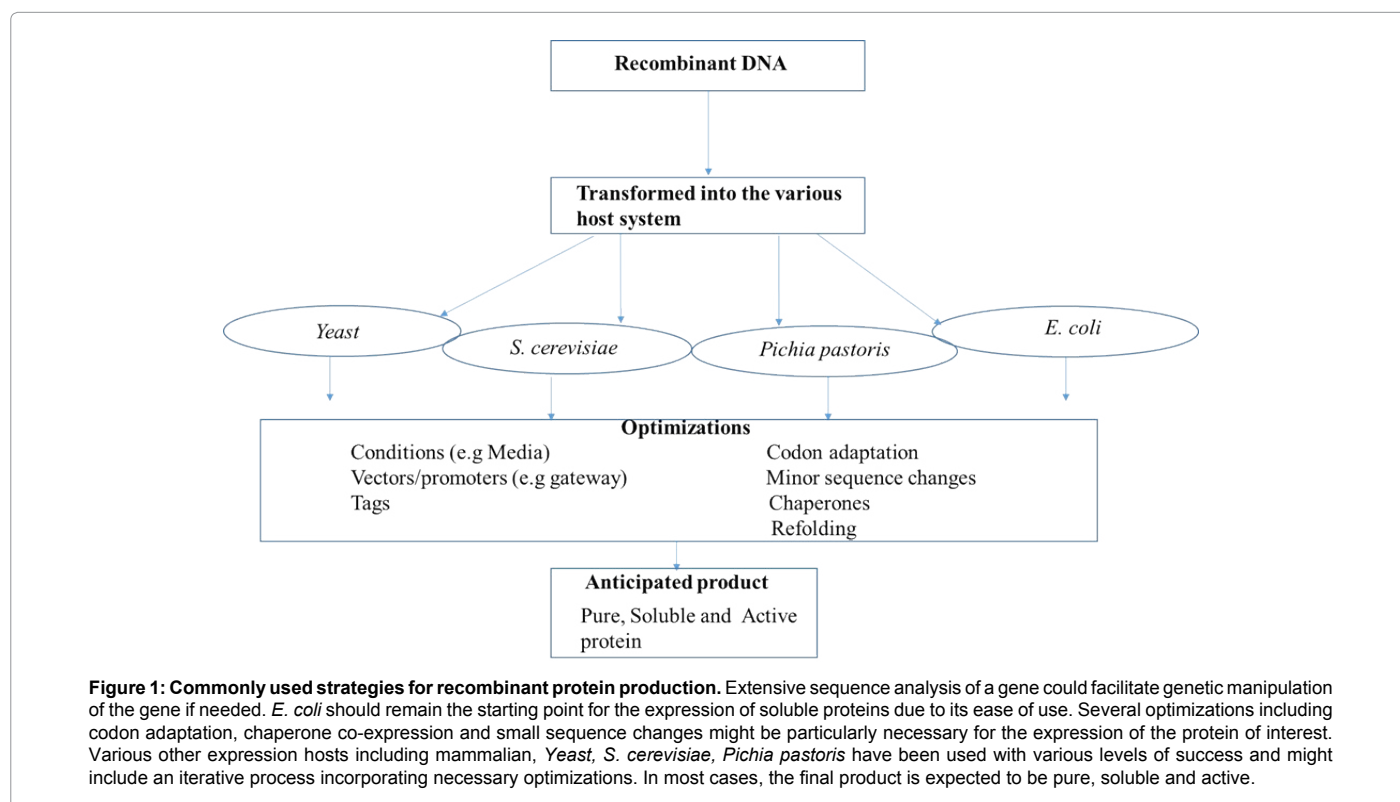
Engineered cells to improve protein production

Biologically active impurities can jeopardize research or therapeutic applications even if present in trace amounts. One consequence during purification of proteins expressed in *E. coli* is DnaK contamination [6]. One approach developed to circumvent DnaK contamination is extensive washing of columns with ATP since DnaK in its ATP-bound state has low affinity for protein. However, this strategy lengthens the purification procedure and is immensely expensive for large scale purifications [14]. In an attempt to eliminate DnaK contamination it was investigated whether recombinant proteins could be produced in the absence of DnaK. Towards that end, the $\Delta dnaK$ derivative of the extensively employed *E. coli* B host strain BL21 (DE3) was constructed. The consequences of the absence of DnaK for the production, solubilize,

correctly assembly, and activities of several recombinant proteins in BL21 (DE3) have been studied. Obtaining a BL21 (DE3) $\Delta dnaK$ strain has allowed to elucidate to what extent such an *E. coli* chaperone is indispensable to protein overproduction in the particular genetic context of an *E. coli* strain that lacks Lon, an ATP dependent protease responsible for degrading unfolded proteins [15]. However, this strain grows at 30°C. This means that any decrease or increase in temperature may result to the death of this cell line. This makes working with this expression system extremely difficult and part of the reasons could be due to the absence of the endogenous molecular chaperones to protect the system from adverse temperatures.

Other protein expression systems such as mammalian and insect cells are very laborious, time-consuming and difficult to maintain [10]. Most pharmaceutically produced proteins such as cytokines [12] and enzymes [4] have previously been produced using Gram-positive expression strains. *Bacilli* strains are attractive hosts due to their ability to express heterologous protein inside cultured cells and secrete the expressed protein into the extracellular medium [10,15]. Some bacterial strains such as *Bacillus brevis* (*Brevibacillus choshinensis* (Takara-Bio, Japan) have intracellular protease gene (*imp*) and extracellular protease gene (*emp*) mutation in order to protect the structural integrity of expressed protein.

According to [14] in living cells newly synthesized and pre-existing proteins are at high risk to become misfolded and form aggregation, therefore need molecular chaperones for assistance. Molecular chaperones are proteins that help both new and pre-existing polypeptide to fold properly [14]. Molecular chaperones may act as "holdases" by stabilizing non-native protein conformations from further aggregation or as "foldases" by helping unfolded or partially unfolded proteins to fold to their native state. The expression of heat



shock proteins (Hsps) due to heat-stress was first reported by [16] using *Drosophila melanogaster* cells. Hsps are ubiquitous proteins that are needed to ensure the appropriate folding and conformation of other proteins in the cell [13,17]. Hsps are classified according to their average molecular weight, with major classes consisting of small Hsps (sHsp), Hsp40, Hsp60, Hsp70 and Hsp100 [10,16] (Table 1).

Due to the significant role molecular chaperones play in maintaining proteome stability, it is therefore not surprising that intracellular Hsps are found in most subcellular compartments such as the nucleus, cytosol, mitochondria and the endoplasmic reticulum (ER) [13,18]. Other molecular chaperones such as Trigger factor, DnaJ/Hsp40, DnaK/Hsp70, GroEL/Hsp60 and GroES/Hsp10 have also been shown to have assisted numerous proteins to fold inside the cellular system [10]. Trigger factor is known to interact with the newly synthesized polypeptide upon its exit from ribosomes [16]. The ribosome associated chaperone Trigger factor is believed to interact with all newly synthesized protein upon their exit from ribosomes [19]. The interaction of these newly synthesized proteins with Trigger factor is to prevent the formation of inclusion bodies because of misfolding of the newly synthesized proteins [13]. However, Trigger factor does not hold onto the substrate for long as the protein continues to be

synthesized. Therefore, newly synthesized protein become larger and thus increases the chances for the protein to become misfolded and aggregated. This means Trigger factor could not prevent newly synthesized protein at this stage, and therefore hands the protein over to DnaK/Hsp70 which cooperates with DnaJ/Hsp40 during the folding of the substrate protein [14,18]. Some of the proteins at this stage could have completed their folding processes, therefore do not require further assistance from molecular chaperones. However, other proteins do require further assistance and therefore are being handled by GroEL in combination with GroES to complete the folding process (Figure 2). This work also reports on some of the proteins that have benefited from molecular chaperones system during their folding stages. This includes the features that they probably share in order to be recognized by molecular chaperones and some of the system developed to improve protein production.

Advantages of co-expression with molecular chaperones

The recombinant protein production is required in high quantity, pure and active form for structural studies, however obtaining a pure and active protein has proven to be a challenge. In order to get a required product some of the proteins need to be supplemented with molecular chaperones for proper folding. A number proteins have been successfully produced in pure, soluble and active form, but some proteins could only be produced in either pure but not soluble or active. For example, five different plasmids with 5 different combinations of 6 chaperones molecule transformed into *E. coli* along with human basic Fibroblast Growth Factor expression plasmid. Each transformant that contain both plasmids for expression of hbFGF and chaperone combinations was induced with proper concentration of related inducers [10]. Subsequently, total amount of produced hbFGF was analyzed based on SDS-PAGE and ELISA. It was indicated that “TF” and “DnaK/DnaJ/GrpE” destabilized hbFGF, while “DnaK/DnaJ/GrpE/GroEL/GroES” and “GroEL/GroES” combinations were able to

Molecular chaperone	Target proteins	Size (kDa)	Function
TF Chaperone	8 amino acids motif enriched with aromatic residues	48	Holding
DnaK Chaperone	Segments of 4 to 5 hydrophobic amino acids	70	Folding
GroEL chaperone	Folds enriched in hydrophobic residues	60	Folding
ClpB	Segments enriched with aromatic residues	100	Disaggregation

Table 1: Major groups and functions of molecular chaperones.

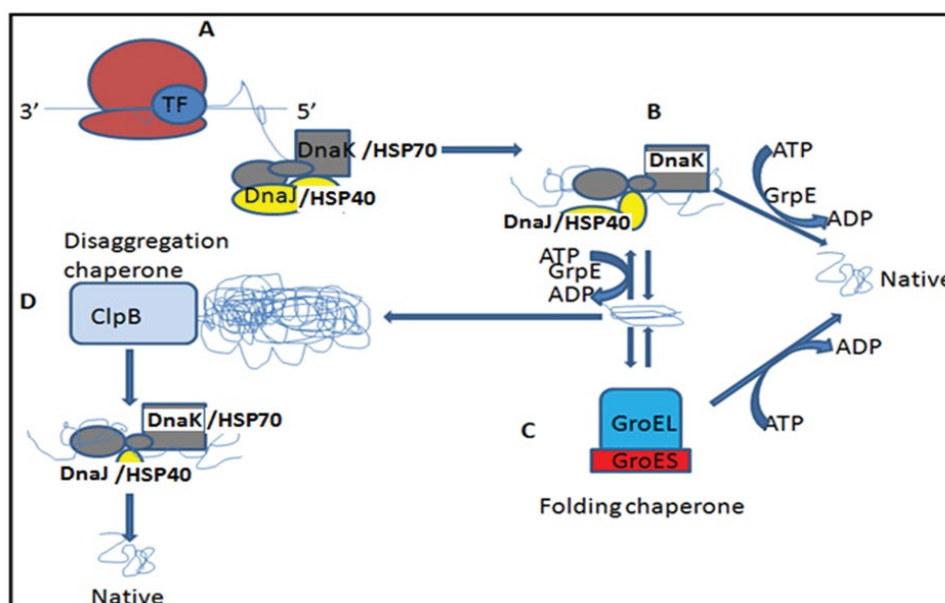


Figure 2: Folding pathways of the newly synthesized protein. (A) TF (Trigger factor) binds emerging nascent chain hand it over (B) Hsp70 homolog DnaK and its cofactors (GrpE, ATP) system, and further hand it over (C) GroEL/Hsp60 chaperonin system for further folding purposes (D) ClpB/Hsp100 is a disaggregating chaperone (Modified from Hartl and Hayer-Hartl, 2002).

stabilize the target protein. It was also revealed that “GroEL/GroES/TF” combination negatively affected the hbFGF production [10,17]. In addition, individual chaperones were reported to have successfully improved some target proteins. Work done by Stephens and co-workers showed that one of the extensively characterized molecular chaperon, heat shock protein 70 from *Plasmodium falciparum* improved the production of the target protein from the same origin produced in *E. coli* [20-22]. Nitrile hydratase (NHase) is an industrially important protein that catalyzes the hydration of nitriles to produce amides. This enzyme has previously been used in the manufacturing of industrial products such as acrylamide and nicotinamide, NHase has also been used in wastewater treatment [10]. Pei and colleagues [8] recently demonstrated that co-expression assays using molecular chaperones such as GroEL/ES, DnaK/J-GrpE and trigger factor significantly promote the soluble expression of recombinant co-type NHase in *E. coli*. Furthermore, GroEL/ES and DnaK/J-GrpE chaperones-assisted expression systems were shown to produce activated and soluble NHase protein, thus, further reducing the industrial production cost of NHase co-type protein.

Chaperone-assisted expressions of Mab and VLPs

Chaperone-assisted folding expression systems are not only beneficial for industrial applications, but can also be adapted for therapeutic proteins such as antibody and vaccine production. Deuerling and colleagues [18] demonstrated that the co-translation of trigger factor or DnaK with an aggregation-prone antibody efficiently increased the level of production of soluble protein produced. Co-expression with molecular chaperones has also been documented to have improved the solubility and affinity of Single-chain variable fragment (fused Heavy and Light chain) antibodies [11,12,23]. The advantages of chaperone expression have also been documented in the expression of Human papillomavirus 16 E7 oncoprotein using molecular chaperones (DnaK–DnaJ–GrpE, GroEL–GroES) [10]. In the presence of molecular chaperones the expression of soluble proteins is significantly increased in *E. coli* [14].

Conclusion

The universal system to produce recombinant proteins is still a bottleneck in order to produce large quantities of proteins for structural studies towards designing compounds or drugs for various diseases. Different systems have been proposed to address this challenge. Amongst the challenges is the production of insoluble proteins or some of the proteins are not produced at all because of the system used. *E. coli* is used as a host system for protein production, due to the presence of its own proteins it is very difficult to bring in foreign proteins in the system. However, molecular chaperones have been studied and are the promising tools to solve this problem [14,24,25]. Their role is to bind to newly synthesized proteins until the fold properly. Molecular chaperones from *E. coli* and *Plasmodium falciparum* are well documented, also have been shown to improve some of the recombinant proteins produced in *E. coli* system [10,22]. However, the question of getting a system that could accommodate all recombinant proteins of different sizes still remains. Therefore, the search for the system that can be used to produce all forms of proteins from different sources is still in progress.

Acknowledgements

This material is based on the work in progress and is supported financially by the National Research Foundation. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the NRF does not accept any liability in regard thereto. Financial support is also

being acknowledged from the University of Zululand Research Committee. Without their support, this work would not have been started nor finished.

References

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181: 223-230.
2. Hamada H, Arakawa T, Shiraki K (2009) Effect of additives on protein aggregation. *Curr Pharm Biotechnol* 10: 400-407.
3. Bukau B, Walter GC (1990) Mutations altering heat shock specific subunit of RNA polymerase suppress major cellular defects of *E. coli* mutants lacking the DnaK chaperone. *EMBO J* 9: 4027-4036.
4. Mehlin C, Boni E, Buckner FS, Engel L, Feist T, et al. (2006) Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. *Mol Biochem Parasitol* 148: 144-160.
5. Baynes BM, Wang DI, Trout BL (2005) Role of arginine in the stabilization of proteins against aggregation. *Biochemistry* 44: 4919-4925.
6. Carrió MM, Villaverde A (2002) Construction and deconstruction of bacterial inclusion bodies. *J Biotechnol* 96: 3-12.
7. Flick K, Ahuja S, Chene A, Bejarano MT, Chen Q (2004) Optimized expression of *Plasmodium falciparum* erythrocyte membrane protein 1 domains in *Escherichia coli*. *Malar J* 3: 50.
8. Pei X, Wang Q, Meng L, Li J, Yang Z, et al. (2015) Chaperones-assisted soluble expression and maturation of recombinant Co-type nitrile hydratase in *Escherichia coli* to avoid the need for a low induction temperature. *J Biotechnol* 203: 9-16.
9. Das U, Hariprasad G, Ethayathulla AS, Manral P, Das TK, et al. (2007) Inhibition of protein aggregation: supramolecular assemblies of arginine hold the key. *PLoS One* 2: e1176.
10. Alibolandi M, Mirzahoseini H, Nehi FM, Tabatabaian G, Amini H, et al. (2010) Improving recombinant protein solubility in *Escherichia coli*: Identification of best chaperone combination which assists folding of human basic fibroblast growth factor. *Afri J Biotechnol* 9: 8100-8109.
11. Sonoda H, Kumada Y, Katsuda T, Yamaji H (2011) Effects of cytoplasmic and periplasmic chaperones on secretory production of single-chain Fv antibody in *Escherichia coli*. *J Biosci Bioeng* 111: 465-470.
12. Wang R, Xiang S, Feng Y, Srinivas S, Zhang Y, et al. (2013) Engineering production of functional scFv antibody in *E. coli* by co-expressing the molecule chaperone Skp. *Front Cell Infect Microbiol* 3: 72.
13. Nishihara K, Kanemori M, Yanagi H, Yura T (2000) Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in *Escherichia coli*. *Appl Environ Microbiol* 3: 884-889.
14. Dobson CM (2004) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* 15: 3-16.
15. Ratelade J, Miot MC, Johnson E, Betton JM, Mazodier P, et al. (2009) Production of recombinant proteins in the lon-deficient BL21(DE3) strain of *Escherichia coli* in the absence of the DnaK chaperone. *Appl Environ Microbiol* 75: 3803-3807.
16. Agashe VR, Guha S, Chang HC, Genevaux P, Hayer-Hartl M, et al. (2004) Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* 117: 199-209.
17. Gupta R, Lakshminpathy SK, Chang HC, Etschells SA, Hartl FU (2010) Trigger factor lacking the PPLase domain can enhance the folding of eukaryotic multidomain proteins in *Escherichia coli*. *FEBS Lett* 584: 3620-3624.
18. Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A, Bukau B (1999) Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* 400: 693-696.
19. Kramer G, Rutkowska A, Wegrzyn RD, Patzelt H, Kurz TA, et al. (2004) Functional dissection of *Escherichia coli* trigger factor: unraveling the function of individual domains. *J Bacteriol* 186: 3777-3784.
20. Matambo TS, Odunuga OO, Boshoff A, Blatch GL (2004) Overproduction, purification, and characterization of the *Plasmodium falciparum* heat shock protein 70. *Protein Expr Purif* 33: 214-222.

21. Shonhai A, Boshoff A, Blatch GL (2005) *Plasmodium falciparum* heat shock protein 70 is able to suppress the thermosensitivity of an *Escherichia coli* DnaK mutant strain. *Mol Genet Genomics* 274: 70-78.
22. Stephens LL, Shonhai A, Blatch GL (2011) Co-expression of the *Plasmodium falciparum* molecular chaperone, PfHsp70, improves the heterologous production of the antimalarial drug target GTP cyclohydrolase I, PfGCHI. *Protein Expr Purif* 77: 159-165.
23. Ow DS, Lim DY, Nissom PM, Camattari A, Wong VV (2010) Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production. *Microb Cell Fact* 9: 22.
24. Fink AL (1999) Chaperone-mediated protein folding. *Physiol Rev* 79: 425-449.
25. Bukau B, Weissman J, Horwich A (2006) Molecular chaperones and protein quality control. *Cell* 125: 443-451.