Available online at www.sciencedirect.com
ScienceDirect
Protein Expression and Purification 52 (2007) 210–218
Protein Expression & Purification
www.elsevier.com/locate/yprep

Comparative expression study to increase the solubility of cold adapted *Vibrio* proteins in *Escherichia coli*

Laila Niiranen [a], Sigrun Espelid [a], Christian R. Karlsen [a], Milla Mustonen [b], Steinar M. Paulsen [a], Pirkko Heikinheimo [b], Nils P. Willassen [a,*]

[a] *The Norwegian Structural Biology Centre (NorStruct) and Department of Molecular Biotechnology, Institute of Medical Biology, Faculty of Medicine, University of Tromsø, N-9037 Tromsø, Norway*
[b] *Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, FIN-00014 Helsinki, Finland*

Received 10 August 2006, and in revised form 7 September 2006
Available online 15 September 2006

Abstract

Functional and structural studies require gene overexpression and purification of soluble proteins. We wanted to express proteins from the psychrophilic bacterium *Vibrio salmonicida* in *Escherichia coli*, but encountered solubility problems. To improve the solubility of the proteins, we compared the effects of six N-terminal fusion proteins (Gb1, Z, thioredoxin, GST, MBP and NusA) and an N-terminal His_6-tag. The selected test set included five proteins from the fish pathogen *V. salmonicida* and two related products from the mesophilic human pathogen *Vibrio cholerae*. We tested the expression in two different expression strains and at three different temperatures (16, 23 and 37 °C). His_6-tag was the least effective tag, and these vector constructs were also difficult to transform. MBP and NusA performed best, expressing soluble proteins with all fusion partners in at least one of the cell types. In some cases MBP, GST and thioredoxin fusions resulted in products of incorrect size. The effect of temperature is complex: in most cases level of expression increased with temperature, whereas the effect on solubility was opposite. We found no clear connection between the preferred expression temperature of the protein and the temperature of the original host organism's natural habitat.

Keywords: Cold adaptation; Fusion tag (Gb-1, Z, thioredoxin, GST, MBP, NusA); Protein expression; Solubility; *Vibrio*

The field of protein X-ray crystallography is more popular than ever as protein scientists and pharmacology companies are searching for new folds and novel targets for therapy. In this quest the focus has necessarily turned from easily approachable organisms and proteins to the more exotic and problematic targets, the so-called "high-hanging fruits". At the same time, sequence information accumulates at increasing speed through various genome projects all over the world. The easily available sequence data has paved the way for what is known as structural proteomics or structural genomics, where the goal is to determine the three-dimensional structures of all proteins [1].

Functional and structural studies require large amounts of pure recombinant protein and an ideal study target would be not only soluble and available in high yields but also easy to purify. Affinity chromatography is a powerful tool in extracting a desired protein from crude extract to a nearly homogenous preparation in a single step. Most natural proteins lack properties which allow affinity extraction and require addition of N- or C-terminal fusion tags. The most commonly used tag is six or more histidines (His_6)[1]; this tag is small and easy to add, and it is less likely to interfere with protein function or crystal packing. For the His_6-tag, there are also several well established purification protocols

* Corresponding author. Fax: +47 77 64 5350.
E-mail address: nilspw@fagmed.uit.no (N.P. Willassen).

[1] *Abbreviations used:* His_6, histidines; MBP, maltose-binding protein; GST, glutathione-*S*-transferase; Gb1, Gβ-1 domain of protein G; NusA, N-utilizing substance A.

1046-5928/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2006.09.005

and matrixes [2]. Other often used fusion partners are the maltose-binding protein (MBP), glutathione-*S*-transferase (GST), Gβ-1 domain of protein G (Gb1), thioredoxin, and N-utilizing substance A (NusA) [3,4]. MBP, GST and the Z-domain of protein A were originally developed as affinity tags, but they also increase solubility [5–7]. This is a highly preferable property, since the use of strong promoters and high inducer concentrations often leads to accumulation of insoluble protein aggregates (inclusion bodies) in *Escherichia coli* [8]. This aggregation is possibly due to overload of the folding pathway during overexpression of the protein (sometimes up to 50% of total cellular protein), which increases the probability of misfolding [9]. The problem is even more severe with extremophilic proteins which do not fit well into the *E. coli* folding machinery. Nonetheless, *E. coli* is still the favored host organism because of its easy manipulation, low cost and well known characteristics.

Statistical models for the solubility or insolubility prediction of *E. coli* expressed proteins [10,11] are inaccurate with extremophilic proteins and predict insolubility better than solubility [12]. Hence, the solubility problem has been tackled by testing the effect of different culture media, additives and temperatures [13,14], but also by using different *E. coli* strains with altered codon usage or with additional heat- or cold-shock proteins. These proteins often act as chaperones and have a dual role: they assist in proper folding of the expressed proteins as well as help the host to grow at suboptimal temperatures [15,16]. High-throughput protein production units have intensively explored strategies to follow when solubility problems are encountered [1,4,12,17–19]. The most extensive studies are by Hammarström et al. [3], Dyson et al. [20] and Shih et al. [4]. Even so, only a few studies compare the effects of different approaches in a parallel and systematic way. The above mentioned three screens compare the expression of mammalian and eukaryotic genes using different tags at one single or two temperatures but mainly in one expression strain only. Screens of the expression of extremophilic bacterial genes are limited both in number and span: Kataeva et al. [12] tested the effect of MBP-fusion and three different induction temperatures on mesophilic *Shewanella oneidensis* and thermophilic *Clostridium thermocellum* proteins, whereas Donnelly et al. [18] expressed *Salmonella typhimurium* proteins with MBP-tag. No screens have been done with a set of cold adapted proteins.

The results from the expression and solubility studies cannot be directly transferred to other species or sets of genes, and comparisons are often relevant only within closely related species or a particular family of proteins. We encountered solubility problems when expressing proteins from the psychrophilic fish pathogen *Vibrio salmonicida* in *E. coli*. A screening system was therefore designed to improve the solubility of *Vibrio* proteins. We selected five different size target proteins involved in DNA modification or transcription regulation in *V. salmonicida* and, for comparison, two related proteins from the mesophilic human pathogen *Vibrio cholerae*. With these genes, we compared the effect of seven different N-terminal fusion partners, two *E. coli* strains and three culture temperatures on protein expression and solubility. We found that in addition to the fusion partner, also expression host and culture temperature affect the levels of expression and the solubility of the produced proteins.

Materials and methods

Cloning

The genes chosen from both *V. salmonicida* (vs) and *V. cholerae* (vc) were *dam* (DNA adenine methylase) and *recJ* (recJ exonuclease), and from *V. salmonicida fur* (ferric uptake regulator), *soxR* and *oxyR*. The sizes of the respective proteins are 17–65 kDa (Table 1). The *V. salmonicida* genes were cloned from chromosomal DNA or a BAC-library clone with specific primers (designed on the basis of a predicted gene from genome sequencing project). The *V. cholerae* genes were cloned from chromosomal DNA (strain ATCC 14035 serovar O:1) with specific primers based on the O1 biovar Eltor strain N16961 sequence (NCBI genome database, www.ncbi.nlm.nih.gov). Enzymes used in PCR reactions were from Invitrogen (Platinum *Pfx* polymerase) and New England Biolabs (*Taq* polymerase).

The Gateway system (Invitrogen) was used for cloning the genes into pDONR221 donor vector and the seven destination vectors (Table 2). Cloning was done as described in the Gateway manual. Entry and expression clones were transformed into competent *E. coli* DH5α cells and correct construct architecture confirmed by colony-PCR. Plasmids

Table 1
Description of the target genes

Gene		Protein size (kDa)	
		Vs	Vc
fur	Ferric uptake regulator	16.6	
soxR	Oxidative stress transcription factor	16.8	
dam	DNA adenine methylase	31.9	31.6
oxyR	Oxidative stress transcription factor	33.8	
recJ	Recombination pathway exonuclease	63.7	64.8

Table 2
Description of the Gateway destination vectors

Vector	Fusion protein	Promoter	Fusion size (kDa)
pDEST15[a]	GST	T7	26.2
pDEST16[a]	Thioredoxin	T7	14.2
pDEST17[a]	6*His	T7	2.4
pDEST-TH1[b]	MBP	tac	44.2
pDEST-TH3[b]	Gb1	T7lac	8.6
pDEST-TH7[b]	NusA	T7lac	56.4
pDEST-TH10[b]	Z	T7lac	11.1

[a] Invitrogen.
[b] Hammarström et al. [3].

were extracted by alkali lysis and isopropanol precipitation [21] or with Wizard Plus SV Miniprep kit (Promega). All entry clones were checked by sequencing with M13 primers using the BigDye chemistry (Applied Biosystems) before proceeding to make expression clones.

Expression and solubility tests

Two different *E. coli* strains were selected for expression. The BL21(DE3)RIL CodonPlus® strain (RIL) from Stratagene carries a plasmid that supplies extra tRNA for codons that are common in organisms with AT-rich genomes, but rare in *E. coli*. BL21-AI™ (AI) from Invitrogen is a protease deficient strain where expression from T7-based expression vectors is tightly regulated by the *araBAD* promoter.

Expression cultures of 2 ml 2xYT broth with the appropriate antibiotics were inoculated with 100 μl of overnight cultures or with 5–10 colonies from transformation plate. Carbenicillin was used instead of ampicillin for pDEST-TH7 constructs. Expression cultures were incubated at 37 °C for 2 h (16 °C cultures) or $1\frac{1}{2}$ h (23 and 37 °C cultures), after which the cultures were transferred to 16, 23 or 37 °C and allowed to acclimatize for 30 min. Expression was then induced by adding isopropyl-β-D-galactoside (IPTG) (Promega) to a final concentration of 0.5 mM or L-arabinose (Sigma) to a final concentration of 0.2%, or both as needed. Cells were harvested by centrifugation 4 h after induction and pellets frozen.

Cell pellets were resuspended in 20 mM Tris–HCl (pH 7.0) and an amount of the bacterial suspension corresponding to 1.5 ml of culture with OD_{600nm} of 2.0 was spun down to collect a uniform number of cells from each sample. These cell pellets were lysed with 100 μl CelLytic B-II extraction reagent (Sigma) according to the manufacturer's instructions to yield soluble and insoluble fractions. Equal amounts of the samples were analyzed on 4–12% Bis–Tris NuPage gels (Invitrogen) and stained with SimplyBlue SafeStain (Invitrogen). Expression levels of the fusion proteins were assessed by comparing the band strength of both soluble and insoluble sample to those of neighboring samples with different constructs, and solubility levels were determined by comparing bands in the insoluble and soluble fractions.

Tandem MS analysis

Protein identification was done by tandem MS analysis after in-gel trypsin digestion of gel plugs of SDS–PAGE separated proteins. Protein spots were excised, reduced and alkylated using a modified method of Shevchenko et al. [22]. The eluted trypsin-generated peptides were concentrated and desalted on OMIX C18 pipette tips (Varian Inc.) according to the manufacturer's instructions prior to MS-analysis. Peptide mixtures were analyzed on a Q-TOF UltimaGlobalTM mass spectrometer with a nanospray ion source interface (Micromass). MS/MS ion data were searched against NCBInr database using the publicly available Mascot search engine (Matrix Science Ltd). The searches were taxonomically restricted to bacteria. Tolerances were set to 0.6 and 0.2 Da for the peptide precursors and the fragment ions, respectively.

Results

We produced a total of 49 different expression constructs by recombining the set of seven entry clones with seven Gateway vectors with N-terminal fusions of His_6-tag, Gb1, Z, thioredoxin, GST, MBP or NusA. Transformation of some pDEST17 constructs into expression strains was problematic especially with the RIL strain, and only a few colonies could be produced even after repeated attempts. In the case of pDEST17-*vcdam* we did not obtain any transformants of RIL.

Effects of N-terminal tags

Fig. 1 shows two examples of SDS–PAGE gels with soluble and insoluble fractions of SoxR expression in RIL strain at two different temperatures. Expression levels were high for Gb1, Z, MBP and NusA tagged proteins at both

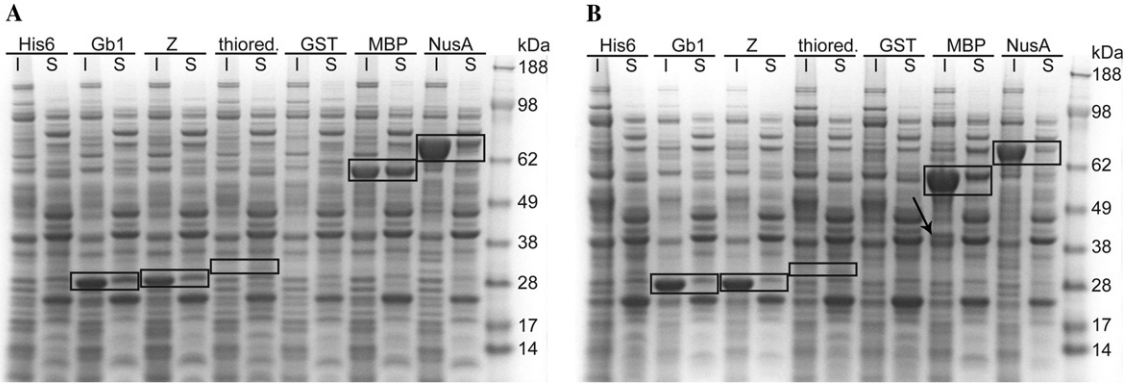

Fig. 1. SDS–PAGE gels showing insoluble (I) and soluble (S) fractions of expressed SoxR in RIL strain at 23 °C (A) and at 37 °C (B) using seven different tags. Expressed fusion proteins are denoted with black boxes. The arrow in gel B points out a product band of incorrect size that was subjected to tandem MS analysis.

temperatures, but product solubility was better at 23 °C. Expression with thioredoxin tag was weak, but all product was soluble. No expression was detected with His_6 or GST tag. Table 3 summarizes the effects of used tag, expression strain and temperature on protein expression and solubility. In short, all fusion proteins were expressed, and expression levels were in most cases moderate to high with all tags. An exception was the Dam proteins where expression was weak to moderate with tags other than MBP and NusA. These two tags had the highest overall expression in our study and were the only ones able to express the gene products at all temperatures in both strains. With other tags, expression of all proteins was not obtained in both strains or at every temperature. All target proteins could be expressed in soluble form with at least two tags, although in some cases solubility was very low. MBP and NusA were the best fusion partners with respect to protein solubility. With these tags all seven fusion proteins were soluble under at least some conditions, totaling 79% and 69% of conditions, respectively. The percentage was under 50% for other tags. The ranking of tags based on our results is MBP > NusA ≫ Gb-1 > thioredoxin > GST > Z > His_6.

We observed expression products of incorrect size either as the only product or in addition to the correct size product with three of the tags: thioredoxin, GST and MBP. Expression of Fur with thioredoxin and GST tags yielded insoluble high molecular weight products of incorrect size in RIL at 23 °C and 37 °C, but the correct size product in AI strain and at 16 °C in RIL. MBP fusion led to additional insoluble products of tag or target protein size in 36% of all cases. The presence of these bands was not linked to a particular expression strain. Three incorrect size protein bands, two about 44 kDa bands from pTH1-*soxR* (Fig. 1B) and pTH1-*oxyR* expression and one about 32 kDa band from pTH1-*vsdam* expression, were subjected to tandem MS analysis. Based on the significant hits list the proteins are MBP or its degradation products. There were no hits pointing to the actual target proteins.

Effect of strain

The AI strain was able to express more proteins (100% *vs*. 86% total for RIL), but was only slightly better (65% *vs*. 57% total) when solubility was compared. There were six cases where only AI produced soluble protein against three cases where only RIL succeeded. However, when there was expression in both strains, we could not see any major differences in the expression levels of the target protein. The pDEST17 vector and RIL strain did not function well together, since in the majority of cases transformants were difficult to obtain.

Effect of temperature

In general, lower temperature resulted in lower expression level, but there were also cases where expression increased as temperature was decreased. Increased expression did not necessarily lead to increased solubility. The preferred temperature for obtaining soluble product was 16 °C or 23 °C (product soluble in 51% and 48% of cases, respectively), but if the product was soluble at both temperatures the amount was usually higher in 23 °C extracts. There was no difference in preference between 16 °C and 23 °C for vs proteins and vc proteins. The optimal temperature depends more on the gene than on the temperature adaptation of the protein.

Effect of gene size and origin

There was a certain gene size dependency regarding solubility but not expression. The smallest protein, Fur, was to some extent soluble even with His_6 tag, whereas larger tags were needed for solubility as the protein size increased. There were no clear differences in detectable expression levels between the two *V. salmonicida* and *V. cholerae* derived genes, although *V. cholerae* proteins were easier to produce in soluble form (soluble product for vs in 18% and for vc in 43% of all conditions). In general, same tags were preferred irrespective of species.

Discussion

Large quantities of recombinant protein are needed for functional and structural characterization studies. For each protein a suitable expression system must be found where the protein is produced in a soluble and correctly folded conformation in high yields. This process of expression system selection is based primarily on trial and error, and bears often more resemblance to lottery than logical scientific research. An overview of previous studies shows that it is hard to find generalized expression guidelines which would apply to all species and proteins. Working with recombinant proteins from unconventional species is therefore particularly challenging as little information on (successful) protein overexpression is available. Our main objective was to find any preferences on fusion tags, expression strains or temperature conditions for overexpression of cold adapted proteins. N-terminal fusion was chosen based on vector availability and in order to keep the screen simple. The tags in our study represent seven commonly tested or promising tags and range in size from small to large (from 2.4 kDa for His_6 to 56.4 kDa for NusA). Similarly, the selected target proteins vary in size. We chose three different expression temperatures (16, 23 and 37 °C) because our main interest is in cold-adapted proteins, and slower protein production due to lower temperature [23] might give more time for proper protein folding. The optimum growth temperature for *V. salmonicida* is 15 °C [24] whereas *V. cholerae* is a mesophile. The AI strain was selected because its tight control of expression might prevent toxicity or aggregation problems arising from basal level expression at pre-induction temperature. The RIL strain was chosen because it is modified to better translate genes with codon usage differing from that of *E. coli*'s.

Table 3
Expression and solubility levels of the *Vibrio* proteins when expressed as seven different gene fusions in two expression strains at 16, 23 and 37 °C[a,b]

Gene	Host strain	Temperature (°C)	6*His		Gb-1		Z		Thioredoxin		GST		MBP		NusA	
			Exp	Sol	Exp	Sol	Exp	Sol	Exp	Sol	Exp	Sol	Exp	Sol	Exp	Sol
vsfur	BL21-AI	16	+++	+	++	++	+	+	+++	+++	+	0	+++	+++	++	+
		23	+++	+	+++	+++	++	+++	+++	++	+++	++	+++	+++	+++	++
		37	+++	0	+++	+++	+++	++	+++	++	+++	+	+++	++	+++	+
	BL21(DE3)RIL	16	–	–	++	+	++	0	+	++	–	–	+++	++	+++	++
		23	–	–	+++	+++	+++	+++	§	–	§	–	+++	+++	+++	+++
		37	–	–	+++	+++	+++	+++	§	–	§	–	+++	++	+++	++
vssoxR	BL21-AI	16	+++	0	++	+	+	+	+++	+	+	0	++	++	+++	+
		23	+++	0	++	+	+	+	+++	+	++	+	+++	+++	+++	+
		37	+++	0	+++	+	+++	+	+++	+	+++	0	+++	+	+++	0
	BL21(DE3)RIL	16	–	–	+++	+	++	0	–	–	–	–	+++°	++	+++	+
		23	–	–	+++	++	++	+	+	+++	–	–	+++°	++	+++	++
		37	–	–	+++	+	+++	0	+	+++	–	–	+++°	++	+++	+
vcdam	BL21-AI	16	+	0	++	0	+	0	+	0	+	0	+++	0	+	+
		23	++	0	+	+	+	0	++	+	+	0	++	++	++	0
		37	+	0	++	0	++	0	++	+	+++	0	+++°	++	+++	0
	BL21(DE3)RIL	16	nt		+	0	–	–	–	–	–	–	+++	++	++	+
		23			+	0	+	0	–	–	–	–	+++	++	+++	+
		37			+	0	++	0	–	–	+	0	+++	+	+++	0
vsdam	BL21-AI	16	+	0	+	0	+	0	++	0	+	0	++	+	++	+
		23	++	0	+	0	+	0	+	0	+	0	+°	+	+	0
		37	++	0	+++	0	++	0	+++	0	++	0	+°	0	+++	0
	BL21(DE3)RIL	16	++	0	++	0	+	0	–	–	+	+	++	+	++	+
		23	–	–	–	–	–	–	++	0	++	0	++°	0	+++	0
		37	–	–	+	0	+	0	–	–	+	0	++°	0	+++	0
vsoxyR	BL21-AI	16	+++	0	++	0	+	+	+++	+	+	+	+++°	0	++	+
		23	+++	0	+++	0	++	0	+++	+	++	+	+++°	0	+++	+
		37	+++	0	+++	0	+++	0	+++	+	+++	+	+++°	0	++	0
	BL21(DE3)RIL	16	–	–	+++	0	++	0	+	0	++	+	+++°	+	+++	+
		23	–	–	+++	0	++	0	–	–	++	+	+++°	+	+++	+
		37	–	–	+++	0	+++	0	++	+	++	+	+++°	+	++	0
vsrecJ	BL21-AI	16	+++	0	++	+	++	0	+++	0	++	0	+++	+	++	+
		23	+++	0	+++	0	+++	0	+++	0	+++	0	+++	+	+++	0
		37	+++	0	+++	0	+++	0	+++	0	+++	0	+++	0	+++	0
	BL21(DE3)RIL	16	++	0	+++	+	++	0	+++	0	++	0	+++	+	+++	+
		23	+++	0	+++	0	+++	0	+++	0	+++	0	+++	++	+++	+
		37	+++	0	+++	0	+++	0	+++	0	+++	0	+++	0	+++	0
vcrecJ	BL21-AI	16	+++	+	++	++	+	+	+++	0	+	0	+++	++	++	+
		23	+++	++	+++	+	++	+	+++	0	+++	+	++	++	++	++
		37	+	0	++	0	++	0	++	0	++	0	++°	+	++	+

	BL21(DE3)RIL	16	+	0	++	++	+	++	++	+	+++	++	++	++	+
		23	+	0	+++	++	+++	++	+++	0	–	–	+++	+++	++
		37	+	0	++	0	++	0	+++	0	–	–	+++	+++	+
Yield (%)[c]	BL21(DE3)AI	Total	100	29	100	71	100	57	100	57	100	57	100	86	100	100
		16	100	**29**	100	**57**	100	**57**	100	43	100	14	100	71	100	**100**
		23	100	**29**	100	**57**	100	43	100	**57**	100	**57**	100	86	100	57
		37	100	0	100	29	100	29	100	**57**	100	29	100	57	100	29
	BL21(DE3)RIL	Total	43	0	100	57	100	43	86	57	71	43	100	100	100	100
		16	43	0	100	57	86	14	57	**29**	57	**43**	100	**100**	100	**100**
		23	29	0	86	43	86	43	57	14	43	14	100	86	100	86
		37	29	0	100	29	100	14	57	29	57	14	100	71	100	43
	Combined		100	**29**	100	**71**	100	**57**	100	**71**	100	**71**	100	**100**	100	**100**
	All conditions		67	10	98	45	95	33	79	38	76	29	100	79	100	69

[a] Expression levels given as: +++ = strongest band on SDS–PAGE gel, ++ = among the strongest bands, + = visible band, – = no expression, § = incorrect size, ° = additional bands, nt = nontransformable.

[b] Solubility levels given as: +++ = majority in soluble fraction, ++ = minority in soluble fraction but among the strongest bands, + = visible band, 0 = nothing in soluble fraction, – = no expression.

[c] Percentage of expressed genes and soluble gene products. Expression of incorrect size only was counted as no expression. Bolded conditions for each tag represent the minimal amount of different conditions needed for producing the maximal percentage of soluble products described in "combined". If two conditions can be used to give same effect, both are in bold. "All conditions" is the percentage of conditions producing soluble product.

The selection of cell lysis reagent CelLytic II-B was based on the rapidity of the lysis procedure. It should be noted that there is a difference in protein composition of soluble and insoluble fractions obtained by CelLytic or by traditional methods such as sonication. It is possible that a greater proportion of proteins would have been soluble if another extraction method had been used. Nevertheless, this should not affect the trends we saw in protein solubility.

Effects of N-terminal tags

Some of the clones did not produce visible amounts of recombinant protein. One cause of this lack of fusion protein expression could be degradation by proteases. Prematurely terminated polypeptides, trapped folding intermediates and partially folded proteins are targeted for degradation to avoid their accumulation in cells [9]. The insoluble products of incorrect size with thioredoxin, GST and MBP fusions were possibly produced in high amounts and resulted in aggregates. Interestingly, the bands analyzed with tandem MS consisted of the MBP tag or fragments of it, and no actual target proteins could be detected although fusion tags should not interfere with target protein identification by MS [19]. All constructs were confirmed correct by sequencing before the expression studies. It is therefore suggested that the tags have folded or started to fold correctly and thus are resistant to degradation, whereas the target proteins have either not been translated in full or have not folded properly and as such are susceptible for degradation. The fact that no degradation products between full length fusion protein and tag protein size were observed implies that translation has stopped before the target proteins. Hammarström et al. [3] also observed expression products of much smaller size than expected and speculated that these could be stable degradation products. In their study this was seen with GST, Gb-1, MBP and NusA tags. Expression of incorrect size gene products is probably not restricted to any particular tags. Observation of such products in our study and in the study by Hammarström et al. [3] suggests, however, that thioredoxin, GST, MBP, NusA and Gb-1 proteins are more stable or resistant to degradation than other tags.

Large tags led to better expression and more soluble products in our study, rendering MBP and NusA the best tags. MBP was superior to NusA at 37 °C in both strains and at 23 °C in AI strain, whereas NusA was better at 16 °C in AI. Our results agree with the results of Kataeva et al. [12] who found that MBP and NusA increased the solubility of thermo- and mesophilic proteins better than the GST tag when expressed at 37 °C. In studies on eukaryotic genes, MBP and NusA score well but are not outstanding in comparison to thioredoxin [3] or GST [17,4]. This suggests that thioredoxin and GST are better suited for mammalian proteins whereas MBP and NusA work well with bacterial proteins. MBP is by far the most studied solubility enhancing tag, and several groups have reported its

superior performance compared to other tags [5,25]. It has been suggested that MBP can act as a general (passive) molecular chaperone preventing the aggregation of its fusion partner's folding intermediates [5]. Douette et al. [26] propose that the solubility enhancing properties of MBP and NusA are in part related to interactions with the GroEL chaperone pathway.

The high solubilising effect of MBP and NusA combined with their large size raises the question of "false" solubility: are both fusion partners correctly folded and soluble or is the soluble tag attached to a misfolded protein? It is particularly suggestive that solubility levels of our MBP fusions are higher in smaller proteins or "easier" (*V. cholerae*) proteins, although there is no size effect if only the Dam and RecJ proteins are considered. Nominé et al. [27] reported that their MBP-fusion protein preparations contained high-molecular-weight aggregates of remarkably homogenous size, and hypothesized that these structures were micelles of misfolded target protein shielded from solvent by the MBP tags. They detected some target protein activity in the samples, but assumed the percentage of folded protein to be very low. Activity tests (results not shown) of our MBP-vsDam and MBP-vsRecJ confirmed that both these proteins were enzymatically active in their fusion forms. It is curious that expression is not affected by the target gene size but solubility is. This size dependency was also noticed by Kataeva et al. [12]: in their experiment there was a noticeable decrease in solubility of target protein expressed with MBP tag when the protein molecular mass exceeded 60 kDa. In a systematic comparison of MBP and NusA [28] little difference was observed between solubility enhancing effects of these two tags. Moreover, there was no difference in their ability to promote passenger protein folding, and it was concluded that solubility or folding efficiency of the passenger protein depends primarily on the passenger protein itself. As Cabrita et al. [17] point out it is important to do parallel cloning with several different tags in order to find the best tag for a particular target protein. Although in our study MBP and NusA are overall the best tags in improving solubility, the amount of soluble product is in some cases higher with thioredoxin, Gb1 or Z.

Effect of strain

It seems that the AI strain with its tighter control of the expressed gene is better suited for expression of "difficult" proteins. No problems with transformation were encountered with this strain contrary to the RIL strain. Insoluble proteins that were poorly expressed in AI were often not expressed at all in RIL. This agrees well with the results obtained by Studier [13], who found that clones expressing the toxic T7 gene *5.3* protein could be more readily established in AI than in ordinary BL21(DE3) cells. Possibly RIL rapidly loses plasmids with genes exerting adverse effects whereas AI with tighter control of basal level expression does not have the need to do so before induction phase. In our study, proteins that were expressed in both strains often had similar solubility as well. Studier [13] also found equal expression and solubility levels in the two hosts for most target proteins, and found no difference in expression levels between RIL and a BL21-AI/RIL strain with genes that were well expressed in RIL. It is therefore suggested that host strain selection affects only the expression of protein and has little influence on product solubility. The studied AI and RIL strains are, however, both BL21 derivatives, and it is possible that hosts from other *E. coli* lineages would perform differently.

Effect of temperature

The effect of temperature on protein expression level was not as straightforward as was expected, and this applies to the solubility level as well. Although there was a general trend for less expression and more soluble product at lower temperature, there were several exceptions. Our experiments demonstrated that only in 2% of all cases 37 °C was the best temperature for protein solubility. Studier [13] found that target protein solubility could occasionally decrease after prolonged incubation at 37 °C. In such cases product was soluble in parallel cultures induced at 20 °C, indicating that lower temperature could be beneficial for product solubility. There are several factors that may contribute to the greater percentage of soluble product at temperatures below 37 °C: Culturing *E. coli* at suboptimal growth temperatures leads to lower general production rate. The nascent proteins will therefore have more time to fold and more chaperones available to help them in this process. Transferring the cultures to 16 °C or 23 °C induces expression of chaperones and cold-shock proteins, which help to maintain proteins soluble but may also assist folding. It is also possible to use too low an expression temperature. Our finding that more products were soluble at 23 °C than at 16 °C is interesting, particularly because it was observed with proteins from both *Vibrio* species. Also for meso- and thermophilic proteins [12] reducing induction temperature from 37 to 28 °C had a positive effect on the solubility, but reducing temperature further to 18 °C had no additional advantage. Despite the overall increased solubility, thermophilic proteins were more soluble than mesophilic ones. Idicula-Thomas and Balaji [29] have shown a strong positive correlation between the thermostability and solubility of proteins. It is therefore not surprising that psychrophilic proteins have poor solubility. An expression and folding temperature of 16 °C should nevertheless not be a problem for proteins from psychrophilic species, but it may be that these proteins are marginally stable and need to be extensively stabilized in their native organisms. DNA itself may be a stabilizing factor, which would help to explain the phenomena in our set of DNA associated proteins. Overproduction of unstable proteins will lead to protein aggregation if their numbers exceed those of the stabilizing components.

Conclusions

Our aim was to find expression preferences for high throughput production of psychrophilic *Vibrio* proteins. Interestingly, we found the preferred fusion tag and expression temperature for obtaining soluble product to depend more on the gene than on the temperature adaptation of the protein. In conclusion, small proteins were easier to express, whereas large tags (MBP and NusA) were best at promoting solubility. Using the AI strain with tight control of expression was beneficial, as well as lowering induction temperature to 23 °C, but not to 16 °C. The recommended experiment for high yields of soluble protein would therefore be to express MBP tagged protein in the AI strain with induction at 23 °C. These conditions are easy and fast to test, but are likely to fail especially with larger proteins. Trying other tags can be useful, but the benefits of testing other host strains or culture temperatures are in our view questionable. For difficult cases there are several ways to proceed further in the search for soluble products: using weaker promoters, decreasing the inducer concentration, using cold-inducible promoters together with lower transcription temperatures or co-expressing molecular chaperones [9,30,31]. One exciting finding is that some maltodextrin-binding proteins from diverse micro-organisms are even better solubility enhancers than *E. coli* MBP [25]. Among them are two very thermostable variants and, interestingly, also the *V. cholerae* MBP ortholog. It would be intriguing to test if a *Vibrio* derived tag combined with expression from cold-induced promoter at expression temperatures in the range of the natural *V. salmonicida* environment would further promote the solubility of *V. salmonicida* proteins.

Acknowledgments

We thank Dr. Helena Berglund for contributing the fusion vectors pDEST-TH1, pDEST-TH3, pDEST-TH7 and pDEST-TH10. This work was supported by the national Functional Genomics Program (FUGE) of the Research Council of Norway, Biocentrum Helsinki Consortium (P.H.) and Sigrid Juselius Foundation (P.H. and M.M.).

References

[1] D. Christendat, A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J.R. Cort, V. Booth, C.D. Mackereth, V. Saridakis, I. Ekiel, G. Kozlov, K.L. Maxwell, N. Wu, L.P. McIntosh, K. Gehring, M.A. Kennedy, A.R. Davidson, E.F. Pai, M. Gerstein, A.M. Edwards, C.H. Arrowsmith, Structural proteomics of an archaeon, Nat. Struct. Biol. 7 (2000) 903–909.

[2] J. Porath, Immobilized metal ion affinity chromatography, Protein Expr. Purif. 3 (1992) 263–281.

[3] M. Hammarström, N. Hellgren, S. van den Berg, H. Berglund, T. Härd, Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*, Protein Sci. 11 (2002) 313–321.

[4] Y.P. Shih, W.M. Kung, J.C. Chen, C.H. Yeh, A.H.J. Wang, T.F. Wang, High-throughput screening of soluble recombinant proteins, Protein Sci. 11 (2002) 1714–1719.

[5] R.B. Kapust, D.S. Waugh, *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, Protein Sci. 8 (1999) 1668–1674.

[6] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase, Gene 67 (1988) 31–40.

[7] E. Samuelsson, T. Moks, B. Nilsson, M. Uhlen, Enhanced in vitro refolding of insulin-like growth factor I using a solubilizing fusion partner, Biochemistry 33 (1994) 4207–4211.

[8] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1992) 411–421.

[9] F. Baneyx, M. Mujacic, Recombinant protein folding and misfolding in *Escherichia coli*, Nat. Biotechnol. 22 (2004) 1399–1408.

[10] D.L. Wilkinson, R.G. Harrison, Predicting the solubility of recombinant proteins in *Escherichia coli*, Biotechnology 5 (1991) 443–448.

[11] S. Idicula-Thomas, A.J. Kulkarni, B.D. Kulkarni, V.K. Jayaraman, P.V. Balaji, A support vector machine-based method for predicting the propensity of a protein to be soluble or to form inclusion body on overexpression in *Escherichia coli*, Bioinformatics 22 (2006) 278–284.

[12] I. Kataeva, J. Chang, H. Xu, C.-H. Luan, J. Zhou, V.N. Uversky, D. Lin, P. Horanyi, Z.J. Liu, L.G. Ljungdal, J. Rose, M. Luo, B.-C. Wang, Improving solubility of *Shewanella oneidensis* MR-1 and *Clostridium thermocellum* JW-20 proteins expressed into *Escherichia coli*, J. Proteome Res. 4 (2005) 1942–1951.

[13] F.W. Studier, Protein production by auto-induction in high-density shaking cultures, Protein Expr. Purif. 41 (2005) 207–234.

[14] J.G. Thomas, F. Baneyx, Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant *Escherichia coli*, Protein Expr. Purif. 11 (1997) 289–296.

[15] S.C. Makrides, Strategies for achieving high-level expression of genes in *Escherichia coli*, Microbiol. Rev. 60 (1996) 512–538.

[16] M. Ferrer, T.N. Chernikova, M.M. Yakimov, P.N. Golyshin, K.N. Timmis, Chaperonins govern growth of *Escherichia coli* at low temperatures, Nat. Biotechnol. 21 (2003) 1266–1267.

[17] L.D. Cabrita, W. Dai, S.P. Bottomley, A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production, BMC Biotechnol. 6 (2006) 12.

[18] M.I. Donnelly, M. Zhou, C.S. Millard, S. Clancy, L. Stols, W.H. Eschenfeldt, F.R. Collart, A. Joachimiak, An expression vector tailored for large-scale, high-throughput purification of recombinant proteins, Protein Expr. Purif. 47 (2006) 446–454.

[19] U. Korf, T. Kohl, H. van der Zandt, R. Zahn, S. Schleeger, B. Ueberle, S. Wandschneider, S. Bechtel, M. Schnolzer, H. Ottleben, S. Wiemann, A. Poustka, Large-scale protein expression for proteome research, Proteomics 5 (2005) 3571–3580.

[20] M.R. Dyson, S.P. Shadbolt, K.J. Vincent, R.L. Perera, J. McCafferty, Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression, BMC Biotechnol. 4 (2004) 32.

[21] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.

[22] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, Anal. Chem. 68 (1996) 850–858.

[23] A. Farewell, F.C. Neidhardt, Effect of temperature on in vivo protein synthetic capacity in *Escherichia coli*, J. Bacteriol. 180 (1998) 4704–4710.

[24] E. Egidius, R. Wiik, K. Andersen, K.A. Hoff, B. Hjeltnes, *Vibrio salmonicida* sp. nov., a new fish pathogen, Int. J. Syst. Bacteriol. 36 (1986) 518–520.

[25] J.D. Fox, K.M. Routzahn, M.H. Bucher, D.S. Waugh, Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers, FEBS Lett. 537 (2003) 53–57.

[26] P. Douette, R. Navet, P. Gerkens, M. Galleni, D. Lévy, E.F. Sluse, *Escherichia coli* fusion carrier proteins act as solubilizing agents for recombinant uncoupling protein 1 through interactions with GroEL, Biochem. Biophys. Res. Commun. 333 (2005) 686–693.

[27] Y. Nominé, T. Ristriani, C. Laurent, J-F. Lefèvre, É. Weiss, G. Travé, Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein, Protein Expr. Purif. 23 (2001) 22–32.
[28] S. Nallamsetty, D.S. Waugh, Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners, Protein Expr. Purif. 45 (2005) 175–182.
[29] S. Idicula-Thomas, P.V. Balaji, Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in *Escherichia coli*, Protein Sci. 14 (2006) 582–592.
[30] M. Ferrer, T.N. Chernikova, K.N. Timmis, P.N. Golyshin, Expression of a temperature-sensitive esterase in a novel chaperone-based *Escherichia coli* strain, Appl. Environ. Microbiol. 70 (2004) 4499–4504.
[31] G. Qing, L.C. Ma, A. Khorchid, G.V. Swapna, T.K. Mal, M.M. Takayama, B. Xia, S. Phadtare, H. Ke, T. Acton, G.T. Montelione, M. Ikura, M. Inouye, Cold-shock induced high-yield protein production in *Escherichia coli*, Nat. Biotechnol. 22 (2004) 877–882.