Biotechnology Reports 11 (2016) 62–69

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

Incorporation of a tag helps to overcome expression variability in a recombinant host

Saikat Bhattacharya[a], Divya Reddy[a], Raja Reddy[b], Asmita Sharda[a], Kakoli Bose[b], Sanjay Gupta[a,*]

[a] Epigenetics and Chromatin Biology Group, Gupta Lab, India
[b] Integrated Biophysics and Structural Biology Lab, Cancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Kharghar, Navi Mumbai 410210, MH, India

ARTICLE INFO

Article history:
Received 8 March 2016
Received in revised form 8 June 2016
Accepted 27 June 2016
Available online 18 July 2016

Keywords:
Rare codons
Histones
Truncated GST
Misfolding
Translation

ABSTRACT

Epigenetics have witnessed a renewed interest over the past decade and assays with recombinant histones has become an important tool for uncovering various aspects of histone biology. However, at times absence of recombinant histone accumulation in bacteria is encountered which is also commonly observed for many eukaryotic proteins in general. In this study, we have investigated the effect of multiple parameters on heterologous expression of proteins. We show that there is marked variability in the accumulation of H2A.2, H2B.1, H3.2 and H4 in the recombinant host, possibly owing to translational variability and degradation by the host proteases. We found that the variability could be overcome by incorporation of the commonly used purification tags, like GST or MBP, of appropriate size and position. Our results provide compelling evidence that transcript parameters like rare codon and GC content, mRNA secondary structure etc. together modulate translation kinetics and govern recombinant protein accumulation.
© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Histones are highly basic proteins with the primary function of packaging the DNA of an organism in a very organised fashion inside the nucleus. Histones undergo various post-translational modifications (PTMs) and have sequence divergent forms known as histone variants. The functional importance of PTMs and variants has led to massive interest in uncovering the outcome of the biochemical changes in the composition of nucleosomes and how these changes are brought about. For conducting *in vitro* experiments recombinant expression has become a very powerful tool for obtaining abundant amounts of highly purified histones with specific PTMs or amino acid composition.

E. coli has been popularly used as heterologous host for robust recombinant production of histones like for most other proteins. However, at times, drastic differences in expression levels of histones is observed in bacteria with complete absence of recombinant histone expression in the worst scenarios [16,28]. This has been attributed to the presence of rare codons in the coding sequence of histones. As per the codon usage bias hypothesis, rare codons may play a role in determining levels of protein expressed in a heterologous host [3]. Apart from the absolute number of such codons, parameters like rare codon clusters [12], their number in the 5′-end of a transcript [7,29] and first six codons [6] and the ones coding for arginine [4,22] are considered important determinants. Likewise, high frequency of arginine residues is suspected to negatively affect expression of histone H1 and H5 [8]. Although, codon optimization proves to be useful in some cases, however, at times, the lack of expression persists [1,28] suggesting that other factors might be involved [reviewed by Ref. [31]].

Abbreviations: HAX-1, human protein HCLS-1 associated protein X-1; NAP1, nucleosome assemble protein 1; DUSP1, dual specificity phosphatase 1; PP1, protein phosphatase 1; IPTG, Isopropyl β-D-1-thiogalactopyranoside; RT-PCR, reverse transcriptase polymerase chain reaction; CAI, codon adaptation indexes; RBS, ribosome-binding site; TMAO, trimethylamine oxide; GST, glutathione-*S*-transferase; MALDI, matrix-assisted laser desorption/ionization; MBP, maltose binding protein; GAPDH, glyceraldehyde phosphate dehydrogenase.

* Corresponding author at: Epigenetics and Chromatin Biology Group, Gupta Lab, Cancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Kharghar, Navi Mumbai 410210, MH, India.

E-mail addresses: sbhattacharya@actrec.gov.in (S. Bhattacharya), dvelga@actrec.gov.in (D. Reddy), rkuppili@actrec.gov.in (R. Reddy), asharda@actrec.gov.in (A. Sharda), kbose@actrec.gov.in (K. Bose), sgupta@actrec.gov.in (S. Gupta).

http://dx.doi.org/10.1016/j.btre.2016.06.002
2215-017X/© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

We also encountered lack of accumulation of few of the histone isoforms on attempting heterologous expression in *E. coli* but did not find a stringent correlation with rare codon content. Interestingly, the transcripts for these histone proteins were being formed in the host. Our results provide compelling evidence that the codon bias usage, mRNA secondary structure and GC content of the histone transcript together modulate the post-transcriptional steps and govern recombinant protein accumulation. Notably, we found that the variability in accumulation could be overcome by incorporation of the purification tags, like GST or MBP, of appropriate size and position. Further, we extended these correlations to six non-histone proteins and show that this probably, can be used as a general strategy to obtain heterologous expression.

2. Materials and methods

2.1. Construct preparation

The coding sequences of the genes were amplified from the cDNA synthesized from RNA (treated with DNaseI) isolated from cell lines. The amplicons were cloned into pTZRT57 vector (Thermo scientific). The cloned fragments were sequenced. For subsequent cloning into different expression vectors, the coding sequences were amplified with primers incorporating the appropriate restriction sites and were subcloned, maintaining the correct reading frame. More details pertaining to cloning are available on request. Please see Supplementary Fig. S8 and S9 for a few of the construct maps.

2.2. Growth and IPTG induction of transformed bacterial expression hosts

A single colony was inoculated from the plates of transformed bacteria in 5 mL or 20 mL LB media and incubated at 37 °C until the OD_{600} reached between 0.3 and 0.6. Induction of recombinant protein expression were carried out with 0.2 mM IPTG. When lack of expression was encountered, titration of IPTG concentration was carried out ranging 0.2 mM–2 mM. The cultures were induced for 3 h at 37 °C or overnight at 18 °C. To investigate the effect of trimethylamine oxide (TMAO), it was added at a working concentration of 60 mM, 1.5 h post IPTG induction and the cells were harvested 1.5 h post-addition. The soluble and the insoluble fractions of proteins were separated by resuspending the cells in buffer containing 50 mM Tris-Cl pH 8.0, 0.5% Triton X-100 and 100 μg/ml lysozyme followed by three rounds of sonication, each for 30 s at 30% amplitude. The lysate was then centrifuged at 27000*g* for 30 min at 4 °C. The supernatant and pellet, thus obtained contains the soluble proteins and the insoluble proteins respectively. The proteins were resolved by 18% SDS-PAGE followed by Coomassie staining (Brilliant Blue R250).

2.3. RT-PCR and real-time PCR

Total RNA was extracted and treated with DNaseI from previously collected bacterial cells as per the manufacturer's (Macherey-Nagel) instructions. RNA (2 μg) was subjected to reverse transcription using M-MLV Reverse Transcriptase and random hexamer primers according to the manufacturer's instructions. cDNAs were then amplified with the corresponding gene-specific primer sets, designed to amplify the total coding sequence. The PCR products were analyzed on a 1% agarose gels. The cDNA synthesized was further used for Real-time PCR experiments with Syber green dye. The expression levels were plotted as relative fold change with respect to GAPDH (glyceraldehyde 3-phoshphate dehydrogenase). Similar results were obtained when fold change was plotted with respect to Ampicillin expression which is expressed from the vector backbone (data not shown).

2.4. Rare codon, CAI and RNA secondary structure prediction

Rare codons were predicted using the Caltor Prediction tool (http://people.mbi.ucla.edu/sumchan/caltor.html). CAI (Codon Adaptation Index) for the coding regions was calculated using CAIcal server (http://genomes.urv.es/CAIcal/) and RNA structures were analyzed by Vienna RNA Websuite program on the web at http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi.

2.5. Accession numbers

H2A.1 (GenBank: JX661508.1), H2A.2 (GenBank: JX661509.1), H2B.1 63.1 (XM_002725263.1), H2B.1 68.1(XM_002725268.1 updated to XM_002725268.2), H4 (NM_001123469.1), H3.3 (GenBank: BC006497.2), HAX-1 (GenBank: AK290626.1), DUSP1 (NM_004417.3), PP1α (NM_002708.3), HLA (GenBank: CAA59215.1), β2M (NM_004048.2). H2B.1 XM_002725263.1 and H2B.1 XM_002725268.1 are referred to as H2B.1 63.1 and H2B.1 68.1 respectively in this article.

3. Results

3.1. Lack of recombinant histone expression is a commonly encountered phenomenon

During purification of rat histone proteins using recombinant methods, we encountered lack of expression of few of the transcripts. There was lack of accumulation of proteins H2A.2 (lane 6), H2B.1 68.1 (lane 10) and H4 (lane 16) even though we used the *E. coli* BL21 (DE3) pLysS strain [Fig. 1a]. No accumulation was observed even at lower growth temperatures post-induction and at varied IPTG concentrations (data not shown). RT-PCR [Fig. 1b (i)] and real-time PCR data demonstrates that not only were the transcripts produced, but also, the relative transcript levels of H2A.2, H2B.1 68.1and H4 were comparable to H2A.1 [Fig. 1b (ii)]. The expression of histones was next attempted in Rosetta (DE3) pLysS, a codon-optimized strain, as the presence of rare codons is speculated to interfere with translation. However, the lack of accumulation of H2A.2 (lane 6) and H4 (lane 16) [Fig. 1a lower image] persisted.

3.2. Recombinant expression does not stringently correlate with rare codon parameters

Analysis of the rare codon parameters of the histone transcripts was carried out to look for possible correlation between accumulation level of proteins and the proposed parameters. Several inconsistences with the codon bias hypothesis were noted. For example, the H3.3 transcript, in addition to 12 rare codons for arginine, contains the highest percentage (16.91%) and number of rare codons in a cluster (four) [Table 1]. Robust accumulation of H3.3 in *E. coli* and lack of accumulation of H2A.2 with the least percentage (6.15%) and rare codons in a cluster (zero), is contradictory to codon bias hypothesis [Table 1]. Further, no correlation between the presence of rare codons at the 5′-end of a transcript with expression levels was observed ["First 6 codons", Table 1]. These contradictions were also reflected in codon adaptation indexes (CAI), which is a proposed important parameter [25] [Table 1]. For instance, H2A.2 having the highest CAI of 0.706 amongst the histone transcripts didn't express in either of the strains [Fig. 1a,c].

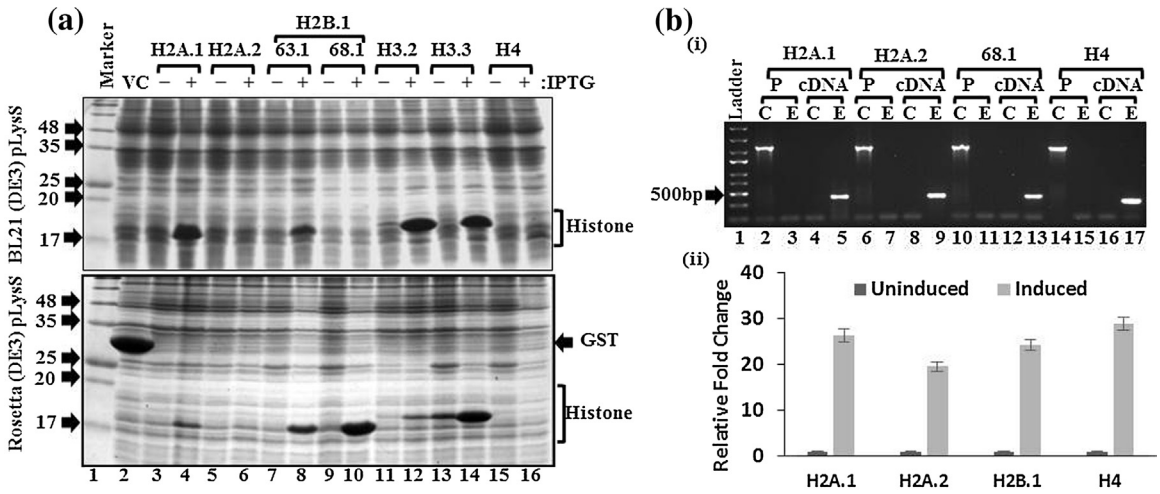

Fig 1. Expression status of recombinant proteins in *E. coli*. (a) SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS (upper image) and Rosetta (DE3) pLysS (lower image) harbouring the pET3a histone constructs, induced with 0.2 mM IPTG at 37 °C for 2 h. pET3a empty vector was used as vector control in BL21 (DE3) pLysS (marked as VC). pGEX2 T empty vector expressing GST was used as an induction control in Rosetta (DE3) pLysS (marked as VC) . The region marked as Histone spans the molecular weight range of the histones being expressed. (b) For checking expression status at the transcript level (i) semi-quantitative RT-PCR was done. 'cDNA' denotes cDNA synthesized by RT-PCR using random hexamer primers followed by PCR amplification with respective gene specific primers. For control (C) cDNA from bacteria having empty pET3a vector was amplified using corresponding gene specific primers and for the experimental sets (E) cDNA from bacteria harbouring pET3a histone constructs was amplified using respective histone primers. H2A.1 was used as a positive control. 'P' denotes experiment to validate DNA free preparation of RNA, in which C is control for amplification where respective plasmids were used as a template. E is the experimental set in which DNA removal by DNaseI from RNA preparation was validated. (ii) Real time PCR to denote the relative fold change upon induction for respective constructs.

3.3. mRNA secondary structure and GC content of transcripts does not always correlate with protein expression levels

As an alternative explanation for lack of recombinant protein accumulation the importance of factors like strong mRNA secondary structure and the 'GC' content of the transcript have been emphasized in relation to transcription and translation [8,14,15]. Therefore, we investigated the possible correlation of mRNA secondary structure and GC content of the histone transcript on protein expression. Vienna RNA web suite [10] was used to maintain consistency with the previously published literature [32] for predicting a possible long-range secondary structure spanning the ribosome-binding site (RBS), translation initiator AUG and the first seven codons. Again, a general pattern did not emerge from our analysis, considering the complexity of the overall secondary structure, the structure near the ATG translation initiation codon, free energy of the structure (see Supplementary Fig. S1) and the% GC of the full-length mRNA transcripts as well as the first six codons (see Supplementary Fig. S1). Most strikingly, the mRNA secondary structure, the free-energy values (−14.20 kcal/mol) and GC content were identical in the case of the three transcripts of H2B.1 (see Supplementary Fig. S1). Thus, their differential expression pattern strongly suggests that the mRNA secondary structure and/or GC content may not be the primary cause behind the results obtained. Although, the secondary structures are predicted, nonetheless, the differential expression pattern observed for the same transcript in different strains cannot be explained based on mRNA secondary structure differences.

3.4. Incorporation of N-terminus tag affects recombinant protein accumulation

Tags are frequently used in purification of proteins. We wanted to test the effect of tags on recombinant protein expression. Interestingly, with 6xHis tag accumulation of H2A.2 protein (lane 6) was seen even without using a codon optimized strain [Fig. 2a]. This further highlights the lack of direct correlation between rare

Table 1
In silico rare codon analysis of histone transcripts.

Amino Acid	Rare Codon	H2A.1	H2A.2	H2B.1 63.1	H2B.1 68.1	H3.2	H3.3	H4
Arg	CGA	0	0	0	0	0	2	1
	CGG	2	2	1	1	5	3	2
	AGG	2	0	0	0	1	5	0
	AGA	0	0	0	0	0	2	1
Gly	GGA	1	1	3	1	0	1	2
	GGG	0	1	1	1	3	3	1
Iso	AUA	0	0	0	0	0	1	0
Leu	CUA	1	0	0	0	0	0	1
Pro	CCC	3	2	3	3	3	5	1
Thr	ACG	1	2	2	2	4	1	0
	Consecutive	0	0	2	2	2	4	2
	First 6 Codons	1	0	1	1	0	1	2
	Frequency	7.69%	6.15%	7.93%	6.34%	11.76%	16.91%	8.73%
	CAI	0.673	0.706	0.625	0.654	0.666	0.644	0.617
	Total	10	8	10	8	16	23	9
	Expression	√	X	√	X	√	√	X

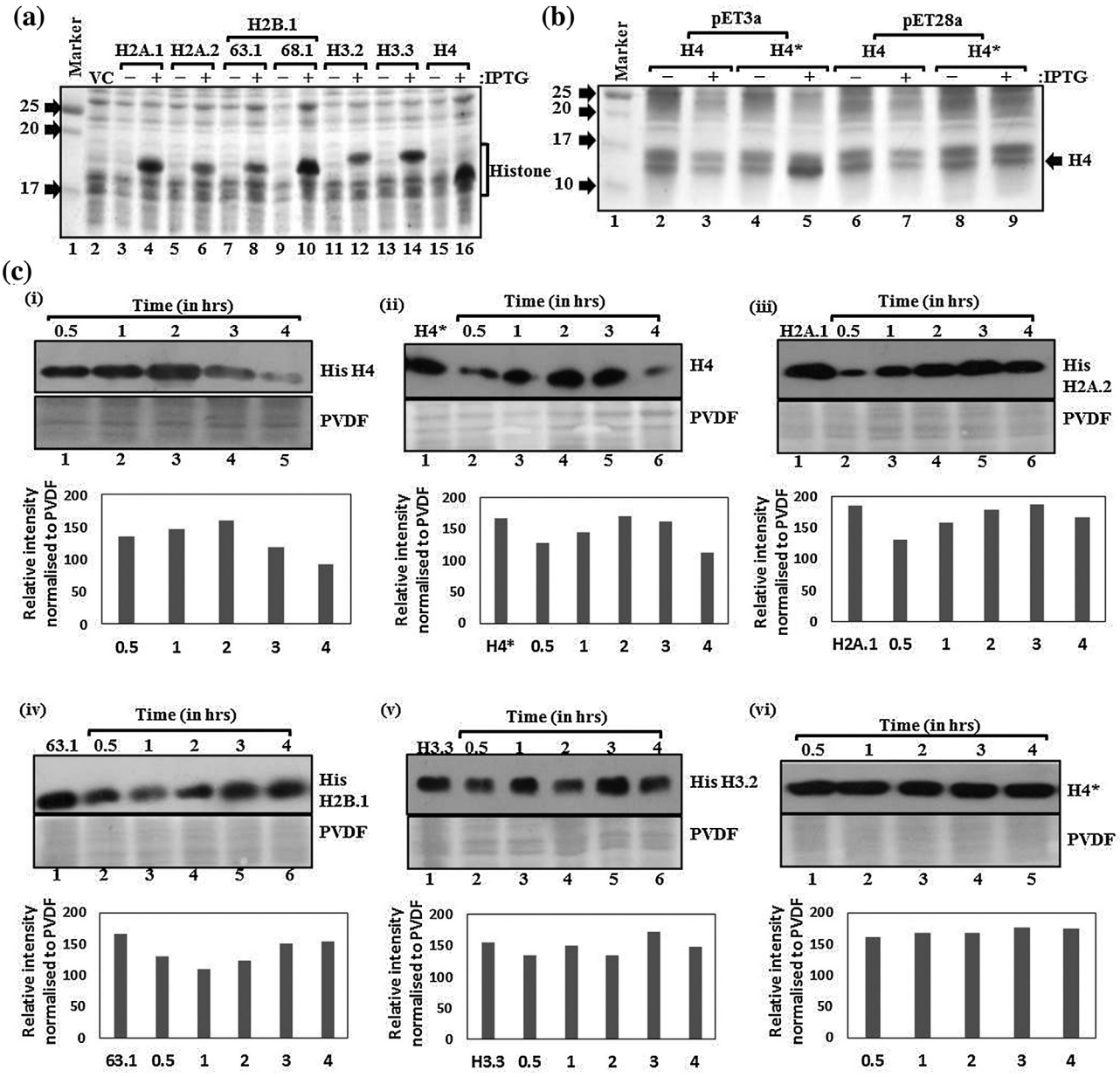

Fig. 2. Effect of incorporation of N-terminal 6X His tag: (a) SDS-PAGE analysis of total cell lysates of BL21 (DE3) pLysS harbouring 6X his tag histone constructs, induced with 0.2 mM IPTG at 37 °C for 2 h (b) 18% coomassie stained gel showing the expression status of histone H4 with and without 6x His tag, induction done at 37 °C for 2 h with 0.2 mM IPTG. H4* denotes codon optimized H4. (c) Western blotting analysis of total cell lysates of BL21 (DE3) pLysS harbouring various histone constructs at different time points post-induction with 0.2 mM IPTG, with anti-His antibody for (i)(iii)–(v) and H4 antibody for (ii)(vi). H4* was selected as a negative control for the experiment, as it does not undergo degradation. The intensity quantification of the bands in the western blots is also depicted below the blots.

codons and protein accumulation. Expression was also seen in case of pET28a H2B.1 68.1 (lane 10) and H4 (lane 16) but the results were "inconsistent". Importantly, this clearly suggested translational variability or the possibility of involvement of protein degradation or both behind the varied accumulation pattern observed for different proteins with changing parameters.

We wanted to investigate this aspect further. We included a codon optimized H4 transcript, in our studies ($H4^*$) that expresses robustly in bacteria. Notably, expression was not obtained in pET28a vector [Fig. 2b]. A time-course experiment was carried out post-induction to monitor H4 accumulation with non-codon optimized H4 in pET28a [Fig. 2c (i)] and pET3a [Fig. 2c (ii)] vector backbone. A decrease in the intensity of band in the immunoblot analysis was observed. Similar decrease was observed in previous report [30] although the kinetics observed was different. Comparable observations were made for H2A.2, H2B.1 68.1 and H3.2 proteins as well [Fig. 2c (iii)–(v)], again, with differing kinetics. Changing growth medium or incorporation of protein stabilising osmolytes like trimethylamine oxide (TMAO) and glycerol did not significantly improve recombinant protein accumulation (data not shown).

3.5. Incorporation of GST tag overcomes variability in recombinant protein accumulation

Strains like BL21 (DE3) pLysS and Rosetta (DE3) pLysS, that are deficient in *OmpT* spare GST fusion proteins from degradation [9]. Hence, GST was cloned upstream of the histones already cloned in

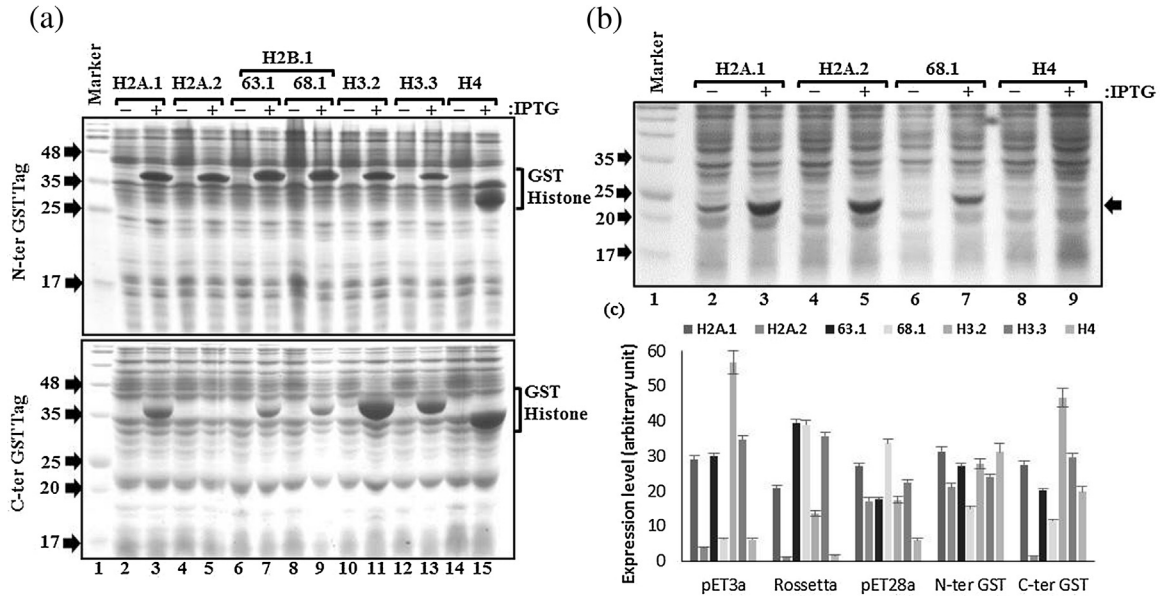

Fig. 3. Size and position of the tag are critical: (a) SDS-PAGE (18% coomassie stained) analysis of cell lysates of BL21 (DE3) pLysS harbouring N-terminal GST tagged (upper image) and with C-terminal GST (lower image) histone constructs induced with 0.2 mM IPTG 37 °C for 2 h. (b) Expression status of histones tagged with N-terminal domain of GST (truncated GST), protein production induced by adding 0.2 mM IPTG at 37 °C for 2 h. (c) Graph showing the relative levels of various histones cloned in different vectors and expressed in different strains.

pET3a (*construct details available on request*). This rules out contribution of difference in vector backbone and copy number towards expression level. After incorporation of GST, indeed, consistent levels of accumulation were seen for all the constructs, including H2B.1 68.1 and H4, as judged by the presence of an intense band around ~35 kD in the induced samples in coomassie stained gels [Fig. 3a].

To test whether the location of the tag is also a determinant, the GST tag was cloned at the C-terminus of histones. Apart from H2A.2 (lane 5), all constructs expressed robustly, although, there was a considerable difference in expression levels of different proteins [Fig. 3a lower image]. Importantly, this rules out the possibility that the expression achieved was due to altered stability of the 5′ end mRNA or 5′ end amino acid residues or enhanced translation initiation resulting from tag incorporation. Quantification depicted in Fig. 3c indicates that with N-ter GST tag most consistent level of recombinant protein accumulation is achieved.

Next, we asked whether the size of the tag influences accumulation levels. To address this, the N-terminal domain of GST comprising of 86 amino acid residues was cloned upstream of the histones in pET3a vector and their expression was attempted. The results were similar to that obtained with the 6xHis tag with stable expression of H2A.2 and H2B.1 and the H4 protein being expressed inconsistently [Fig. 3b]. This clearly suggested that the size of the tag with respect to that of protein of interest is an important determinant.

3.6. Incorporation of a suitable tag can possibly be used as a general procedure to obtain recombinant protein expression

We included a few non-histone proteins, HAX-1, DUSP1, DUSP1 Mutant, HLA, β2 M and NAP1 for which correlation between protein accumulation and the different proposed parameters was not observed (see Supplementary Fig. S2). The proteins HAX-1, DUSP1, DUSP1 Mutant and NAP1 were tagged with 6xHis, GST and MBP and their expression was checked in BL21 (DE3) pLysS. Quantitative depiction of the three independent experiments is depicted in Fig. 4 and representative gel pictures can be seen in Supplementary Fig. S3. Indeed, the expression of HAX1 and DUSP1 with GST tag was achieved, however, the level was very low. For DUSP1 mutant no expression was observed in BL21 (DE3) pLysS.

Earlier, the result with N-terminal truncated GST tag upstream of histones suggested that the ratio of the size of the tag to that of the protein might be important. Thus, the expression of the non-histone proteins was attempted with a bigger tag, MBP. Comparison of lane 7 and lane 9 of Supplementary Fig. S3 a(i) clearly suggests that yield was significantly enhanced when MBP tag was used instead of GST. Notably, from the quantification plots it is apparent that once a tag of appropriate size has been incorporated then the expression in Rosetta (DE3) pLysS is higher than BL21 (DE3) pLysS [Fig. 4], probably, owing to the more robust translation resulting from increased tRNA supplies for the rare codons.

3.7. Solubility of proteins doesn't necessarily correlate with accumulation

Solubility of proteins is proposed to be inversely correlated with accumulation due to increased accessibility to host proteases [13,20,23,27]. The solubility profile of histones expressed using pET28a vector in BL21 (DE3) pLysS strain at 37 °C is depicted in Fig. 5a. All the histones were mostly insoluble. Further, we tested the solubility of H2A.1 with different tags. The experiments were carried out at 18 °C, as lower temperature may lead to lower metabolic rate allowing better folding and solubility. The codon optimized Rosetta (DE3) pLysS strain was also used as codon optimisation have been shown to alter protein solubility [21]. We did find altered solubility for 6x His-tagged H2A.1 at 18 °C in Rosetta (DE3) pLysS strain. Major alteration in solubility profile was observed with GST and MBP tag [Fig. 5b]. Therefore, we checked the solubility of the non-histone proteins with GST and MBP tag at 37 °C and 18 °C in both the strains [Fig. 5c]. However, we didn't find any correlation between protein solubility and expression, as irrespective of solubility, we could achieve robust accumulation.

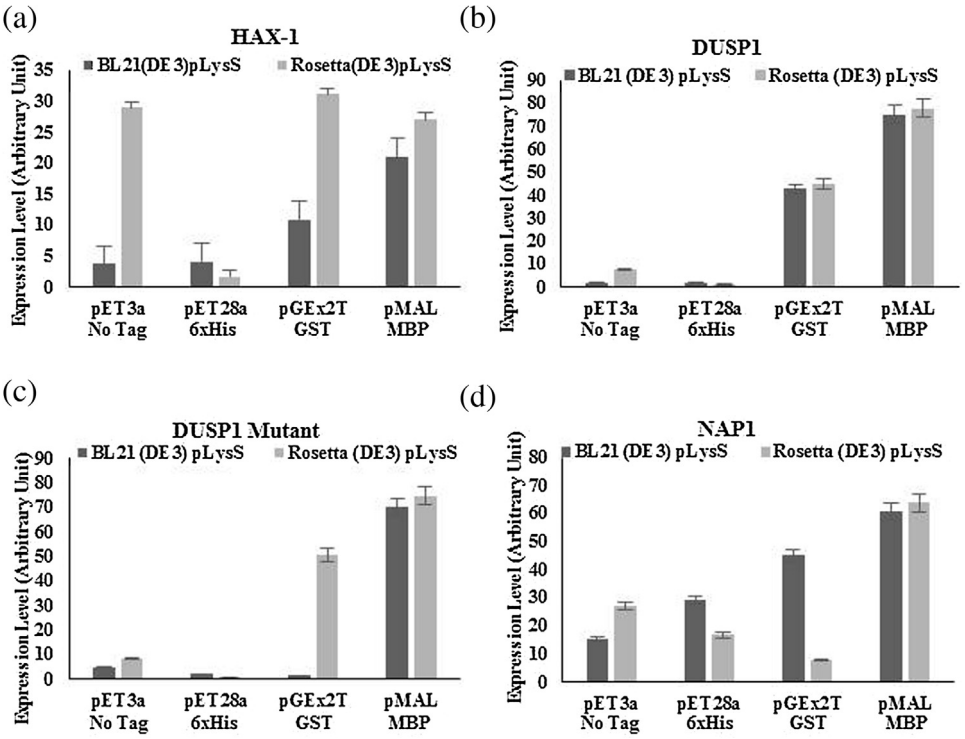

Fig. 4. Testing the hypothesis on Non-Histone proteins. Quantitative data of the constructs cloned in various vectors and expressed either in BL21 (DE3) pLysS or Rosetta (DE3) pLysS strain. (a) is for Hax1, (b) DUSP1, (c) DUSP1 mutant, (d) NAP1. See Supplementary Fig. S3 for representative gel pictures.

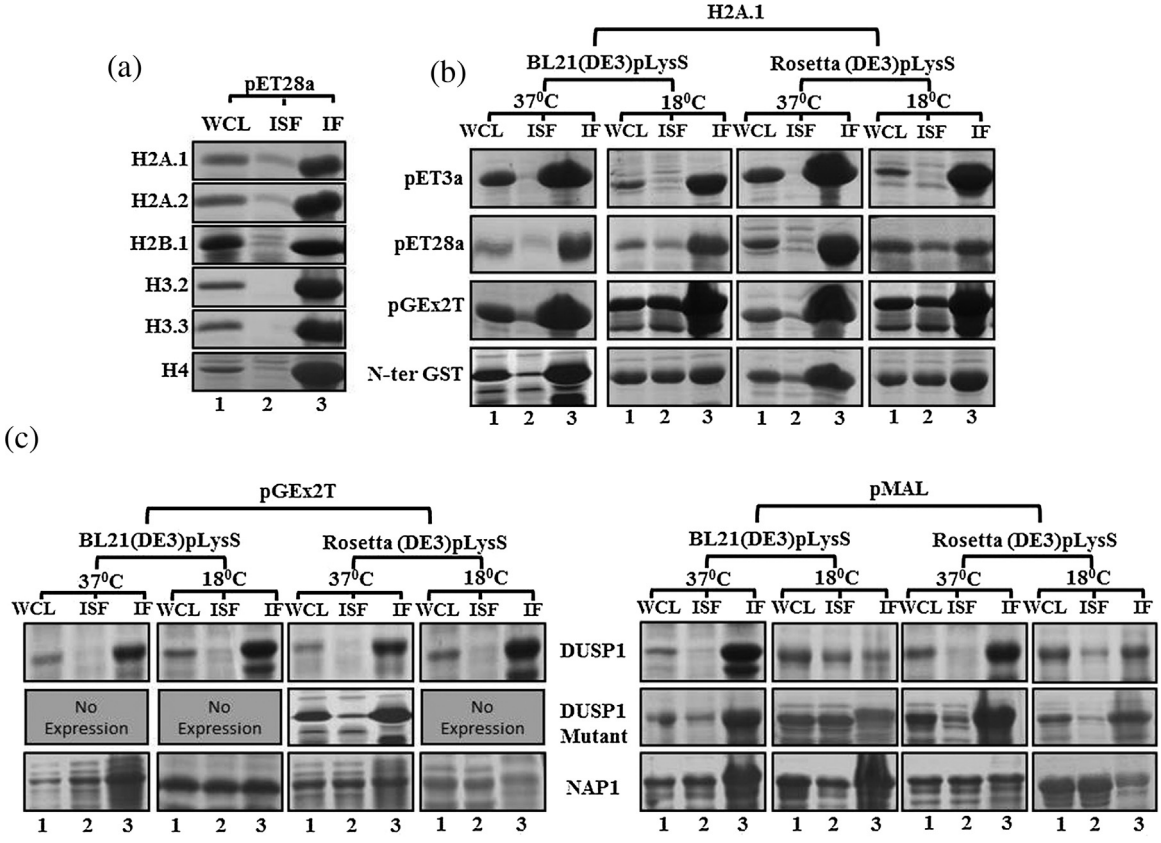

Fig. 5. Solubility analysis of various proteins. Various fractions, WCL—Whole cell lysate, ISF—Induced soluble fraction and IF—Insoluble fraction, post induction with IPTG were collected and separated on SDS gel. (a) Solubility profile of histones cloned in pET28a vector and expressed in BL21 (DE3) pLysS at 37 °C, induction by 0.2 mM IPTG for 2 h. (b) SDS gel with comparative analysis of H2A.1 when cloned and expressed in various vectors, in BL21 (DE3) pLysS and Rosetta (DE3) pLysS at 37 °C and 18 °C. (c) Comparative solubility profile of Non-Histone proteins DUSP1, DUSP1 mutant and NAP1 when expressed as GST and MBP fusions (cloned in pMAL) in both the bacterial strains at 37 °C and 18 °C.

4. Discussion

Heterologous expression is a very powerful tool for producing recombinant proteins. Many reports provide guidelines to optimise recombinant protein expression, however, not many studies address why many a times the proposed hypotheses fail to explain the lack of recombinant protein expression. Factors like rare codons need not be detrimental to protein accumulation and codon-optimisation, at times, may be unfavourable for protein expression. Possibly, rather than a barrier to the translation machinery, the various transcript parameters like rare codons or mRNA secondary structure (to which GC content also contributes), are modulators of translation kinetics. The overall translation speed can be modulated through tRNA recycling at the ribosome [5] and attenuation of *trp* operon presents a classic example of this [19].

In such scenario, the question arises that why sometimes there is complete lack of expression? Our immunoblot data provides interesting insights that helps us to understand the underlying reason. The intensity of bands did not show a unidirectional pattern suggesting translational variability or degradation or both during the course of induction of recombinant protein expression. In fact, translational variability may itself lead to degradation as translation and protein folding are coupled in bacteria. Presence of rare codons can affect the rate of translation and would influence the folding, thus local protein structure [24]. The local protein structure probably in turn may define the response of targeting and degradation by the host protease. Therefore, altering the "translation kinetics code" inherent in the sequence of the transcript may lead to complete lack of accumulation or may result in improved accumulation. This also emphasizes that the position of a rare codon should be more important than its absolute number. Subsequently, coding sequences having more rare-codons, even though would have slower translation kinetics, may lead to better accumulation if present at positions which promote "better" folding. This hypothesis is consistent with the proposition that rare codons may provide the genetic instruction for the regulation of rate of protein synthesis to allow proper folding of the nascent polypeptide [17].

The challenge remains to establish a predictable correlation between any of the proposed parameters and expression, which can help us to engineer means of circumventing expression variability. Unfortunately, it appears it is not going to be very straight forward. An earlier study showed that for the same transcript replacing the rare codons with the major ones or changing the GC content can independently bring about improved expression [26]. Clearly, not only are these parameters independently important, but correlations amongst secondary structure, the GC content and the codon usage of the transcript are significant [11]. As an alternative, we found that incorporation of a tag leads to much more consistent and predictable outcomes of protein expression and the size of the tag with respect to that of proteins is also important. The above hypothesis also gives us the reason as to why some proteins, which do not efficiently express without a tag, are sometimes robustly expressed as fusion with known tags like GST and MBP. Importantly, an N-terminus tag appears to be better suited. Probably, at the N-terminus of a fusion, the folding of GST/MBP precedes that of the downstream unit. However, when GST is present downstream of a protein, its folding might be influenced by its fusion partner leading to the attainment of a misfolded form.

5. Conclusions

To summarise, translational variability arising due to the combined effect of rare codon and GC content, mRNA secondary structure etc. leads to inefficient accumulation of heterologous proteins, possibly by inducing degradation of the protein. The expression variability may be averted by incorporating an N-terminal tag of suitable size. Finally, the steps depicted in Supplementary Fig. S4 should allow a researcher to troubleshoot expression problems in minimum time with maximum probability of success.

Competing interests

No competing interests.

Author's contributions

SB prepared constructs, performed expression analysis, purified proteins, and contributed to manuscript preparation. DR prepared constructs, performed expression analysis, carried out cell line maintenance and transfections. RR prepared constructs and performed expression analysis. AS prepared constructs. KB and SG contributed to experimental designing and manuscript preparation. All authors read and approved the final manuscript.

Acknowledgments

The authors are grateful to Dr. Karolin Luger, Colorado State University for providing codon optimized construct of H4. We are also thankful to Dr. Francoise Carlotti, LUMC-Netherlands for providing DUSP-1 wild type and mutant constructs in mammalian expression vector from which they were subcloned. SB and DR are supported by CSIR India fellowship, AS by UGC India fellowship and RR by ACTREC fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.btre.2016.06.002.

References

[1] M.F. Alexeyev, H.H. Winkler, Gene synthesis, bacterial expression and purification of the *Rickettsia prowazekii* ATP/ADP translocase, Biochim. Biophys. Acta Biomembr. 1419 (1999) 299–306, doi:http://dx.doi.org/10.1016/S0005-2736(99)00078-4.
[3] S.K. Behura, D.W. Severson, Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes, Biol. Rev. 88 (2013) 49–61, doi:http://dx.doi.org/10.1111/j.1469-185X.2012.00242.x.
[4] T.L. Calderone, R.D. Stevens, T.G. Oas, High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*, J. Mol. Biol. 262 (1996) 407–412, doi:http://dx.doi.org/10.1006/jmbi.1996.0524.
[5] G. Cannarozzi, G. Cannarrozzi, N.N. Schraudolph, M. Faty, P. von Rohr, M.T. Friberg, A.C. Roth, P. Gonnet, G. Gonnet, Y. Barral, A role for codon order in translation dynamics, Cell 141 (2010) 355–367, doi:http://dx.doi.org/10.1016/j.cell.2010.02.036.
[6] G.F. Chen, M. Inouye, Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes, Nucleic Acids Res. 18 (1990) 1465–1473, doi:http://dx.doi.org/10.1093/nar/18.6.1465.
[7] M.H. de Smit, J. van Duin, Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 7668–7672, doi:http://dx.doi.org/10.1073/pnas.87.19.7668.
[8] S.E. Gerchman, V. Graziano, V. Ramakrishnan, Expression of chicken linker histones in *E. coli*: sources of problems and methods for overcoming some of the difficulties, Protein Expr. Purif. 5 (1994) 242–251, doi:http://dx.doi.org/10.1006/prep.1994.1037.
[9] S. Gottesman, Proteases and their targets in *Escherichia coli*, Annu. Rev. Genet. 30 (1996) 465–506, doi:http://dx.doi.org/10.1146/annurev.genet.30.1.465.
[10] A.R. Gruber, R. Lorenz, S.H. Bernhart, R. Neubock, I.L. Hofacker, The Vienna RNA websuite, Nucleic Acids Res. 36 (2008) W70–W74, doi:http://dx.doi.org/10.1093/nar/gkn188.
[11] W. Gu, T. Zhou, C.O. Wilke, A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes, PLoS Comput. Biol. 6 (2010) e1000664, doi:http://dx.doi.org/10.1371/journal.pcbi.1000664.

[12] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, Curr. Opin. Biotechnol. 6 (1995) 494–500, doi:http://dx.doi.org/10.1016/0958-1669(95)80082-4.
[13] J.F. Kane, D.L. Hartley, Formation of recombinant protein inclusion bodies in *Escherichia coli*, Trends Biotechnol. 6 (1988) 95–101.
[14] T. Keller, D. Mis, K. Jia, C. Wilke, Reduced mRNA secondary-structure stability near the start codon indicates functional genes in prokaryotes, Genome Biol. Evol. 4 (2012) 80–88, doi:http://dx.doi.org/10.1093/gbe/evr129.
[15] M. Kubo, T. Imanaka, mRNA secondary structure in an open reading frame reduces translation efficiency in *Bacillus subtilis*, J. Bacteriol. 171 (1989) 4080–4082.
[16] K. Luger, T.J. Rechsteiner, T.J. Richmond, Expression and purification of recombinant histones and nucleosome reconstitution, in: P. Becker (Ed.), Methods in Molecular Biology (Clifton, N.J.), Methods in Molecular Biology™, Humana Press, 1999, pp. 1–16, doi:http://dx.doi.org/10.1385/1-59259-681-9:1.
[17] M. Marin, Folding at the rhythm of the rare codon beat, Biotechnol. J. 3 (2008) 1047–1057, doi:http://dx.doi.org/10.1002/biot.200800089.
[19] D.E. Morse, R.D. Mosteller, C. Yanofsky, Dynamics of synthesis translation, and degradation of trp operon messenger RNA in *E. coli*, Cold Spring Harb. Symp. Quant. Biol. 34 (1969) 725–740, doi:http://dx.doi.org/10.1101/SQB.1969.034.01.082.
[20] M. Murby, M. Uhlén, S. Ståhl, Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*, Protein Expr. Purif. 7 (1996) 129–136.
[21] G.L. Rosano, E. a Ceccarelli, Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain, Microb. Cell Fact. 8 (2009) 41, doi:http://dx.doi.org/10.1186/1475-2859-8-41.
[22] A.H. Rosenberg a, E. Goldman, J.J. Dunn, F.W. Studier, G. Zubay, Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system, J. Bacteriol. 175 (1993) 716–722.
[23] Rudolph, R., 1996. Successful protein folding on an industrial scale. Princ. Pract. protein Fold.
[24] R. Saunders, C.M. Deane, Synonymous codon usage influences the local protein structure observed, Nucleic Acids Res. 38 (2010) 6719–6728, doi:http://dx.doi.org/10.1093/nar/gkq495.
[25] P.M. Sharp, W.-H. Li, The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications, Nucleic Acids Res. 15 (1987) 1281–1295, doi:http://dx.doi.org/10.1093/nar/15.3.1281.
[26] G. Sinclair, F.Y.M. Choy, Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*, Protein Expr. Purif. 26 (2002) 96–105, doi:http://dx.doi.org/10.1016/S1046-5928(02)00526-0.
[27] P. Suthakaran, J. Balasubramanian, M. Ravichandran, V. Murugan, L.N. Ramya, K.K. Pulicherla, Studies on lipidification of streptokinase: a novel strategy to enhance the stability and activity, Am. J. Ther. 21 (2014) 343–351.
[28] Y. Tanaka, M. Tawaramoto-Sasanuma, S. Kawaguchi, T. Ohta, K. Yoda, H. Kurumizaka, S. Yokoyama, Expression and purification of recombinant human histones, Methods 33 (2004) 3–11, doi:http://dx.doi.org/10.1016/j.ymeth.2003.10.024.
[29] T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, Y. Pilpel, An evolutionarily conserved mechanism for controlling the efficiency of protein translation, Cell 141 (2010) 344–354, doi:http://dx.doi.org/10.1016/j.cell.2010.03.031.
[30] L. Vergani, F. Canneva, P. Ghisellini, C. Nicolini, Expression, purification, and structural characterization of human histone H4, Protein Expr. Purif. 24 (2002) 420–428, doi:http://dx.doi.org/10.1006/prep.2001.1601.
[31] M. Welch, A. Villalobos, C. Gustafsson, J. Minshull, You're one in a googol: optimizing genes for protein expression, J. R. Soc. Interface 6 (2009) S467–S476, doi:http://dx.doi.org/10.1098/rsif.2008.0520.focus.
[32] W. Zhang, W. Xiao, H. Wei, J. Zhang, Z. Tian, mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in *Escherichia coli*, Biochem. Biophys. Res. Commun. 349 (2006) 69–78, doi:http://dx.doi.org/10.1016/j.bbrc.2006.07.209.