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.

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

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].