The impact of intragenic CpG content on gene expression

Asli Petra Bauer¹, Doris Leikam¹,², Simone Krinner¹, Frank Notka², Christine Ludwig¹, Gernot Längst³ and Ralf Wagner¹,²,*

¹Institute of Medical Microbiology and Hygiene, Molecular Microbiology & Gene Therapy Unit, University of Regensburg, ²Geneart AG, Biopark II, Josef-Engert-Straße 11 and ³Institute for Biochemistry, Genetics and Microbiology, University of Regensburg, 93053 Regensburg, Germany

Received November 23, 2009; Revised February 3, 2010; Accepted February 8, 2010

ABSTRACT

The development of vaccine components or recombinant therapeutics critically depends on sustained expression of the corresponding transgene. This study aimed to determine the contribution of intragenic CpG content to expression efficiency in transiently and stably transfected mammalian cells. Based upon a humanized version of green fluorescent protein (GFP) containing 60 CpGs within its coding sequence, a CpG-depleted variant of the GFP reporter was established by carefully modulating the codon usage. Interestingly, GFP reporter activity and detectable protein amounts in stably transfected CHO and 293 cells were significantly decreased upon CpG depletion and independent from promoter usage (CMV, EF1α). The reduction in protein expression associated with CpG depletion was likewise observed for other unrelated reporter genes and was clearly reflected by a decline in mRNA copy numbers rather than translational efficiency. Moreover, decreased mRNA levels were neither due to nuclear export restrictions nor alternative splicing or mRNA instability. Rather, the intragenic CpG content influenced de novo transcriptional activity thus implying a common transcription-based mechanism of gene regulation via CpGs. Increased high CpG transcription correlated with changed nucleosomal positions in vitro albeit histone density at the two genes did not change in vivo as monitored by ChIP.

INTRODUCTION

The increasing demand for pharmaceutical-relevant proteins used in vaccine development, clinical therapy and gene therapy approaches requires improved strategies of production with regard to a sustained and high transgene expression. The recombinant expression of transgenes is often complicated by several factors such as bad solubility and missing posttranslational modifications in a prokaryotic expression background (1) or low-expression yields when using eukaryotic expression systems (2–4). Much effort has been spent on the development of optimized expression platforms, as high-level expression of recombinant proteins strongly depends on the diligent design of the expression cassette (5). The fusion of intronic sequences upstream of the open reading frame and the use of strong hybrid promoters have been demonstrated to be successful in enhancing transgene expression (6,7). The adaptation of the codon usage to avoid rare codon was another progress in yielding higher expression levels (8). With regard to viral sequences it was shown for the human immunodeficiency virus type 1 gag gene that the adaptation to the human codon usage gave a CG-rich sequence and resulted in the constitutive and high-level expression of the otherwise strictly regulated gene in mammalian cell lines (9,10). Optimization of codon usage has therefore substantially encouraged the development of safe lentiviral vectors (11) and highly efficient packaging cell lines for vaccine design.

Yet there is need for further progress considering protein amount and stability of transgene expression to surpass vector-based expression vectors relieving handling and safety issues. The optimization of codon usage has also been applied to plasmid DNA-based gene therapy

*To whom correspondence should be addressed. Tel: +49 941 944 6452; Fax: +49 941 944 6455; Email: ralf.wagner@klinik.uni-regensburg.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2010. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
approaches to take advantage of their safety issues and simultaneously improve protein expression, low expression levels being one of the most prominent disadvantages of plasmid DNA (12). However, plasmid DNA applications using optimized expression cassettes combined with strong viral promoters have been shown to be impaired by downregulation of gene expression via transcriptional silencing in vivo and in vitro (13). The repression of transcription often correlates with methylation of CpG dinucleotides representing an epigenetic control mechanism involved in development, imprinting, X-chromosomal gene inactivation and cancerogenesis (14–16). CpG dinucleotides are statistically under-represented within the eukaryotic genome as CpG dinucleotides methylated at the carbon 5-position of the cytosine residue are readily deaminated thereby causing a transition mutation from cytosine to thymine (17). However, there are CpG dense genomic regions which are virtually protected from CpG methylation likely due to occupation by nuclear factors or less efficient binding of methyltransferases to CpGs clustering at narrow intervals (18). These CpG islands are almost exclusively either located within or in the proximity of promoter regions and sometimes extend into transcribed areas (19). Whereas unmethylated CpGs are associated with transcriptional activity, their hypermethylated state is often observed in carcinogenesis and mediates gene silencing directly by sterically hindering transcription factor binding or indirectly by interacting with methyl-binding domain repressor proteins which in turn recruit chromatin modifying complexes to achieve a condensed chromatin state (20). The disposition of CpG dinucleotides to cause mutations (21) and gene silencing (22) led to the avoidance of CpGs when designing expression constructs for recombinant protein production (23) and gene therapy (24). So far, it is still unclear whether DNA methylation is a prerequisite or rather a consequence of gene silencing (25). Apart from hypermethylation and silencing of endogenous promoters during cancer progression (26), strong heterogeneous and ubiquitously expressing promoters such as the commonly used immediate-early gene promoter from cytomegalovirus (CMV) have been reported to be susceptible to CpG methylation, in contrast to those derived from housekeeping genes (27). Moreover, methylation of promoter regions is not solely responsible for transcriptional repression. Flanking methylated vector elements including intragenic regions have been shown to contribute to silencing of gene expression, potentially by promoting methylation of adjacent CpGs and establishing transcriptionally inactive chromatin structures (28,29). Especially CpG-rich vector sequences of bacterial origin, which are prone to de novo methylation, have been proven to be significantly involved in silencing of codcelivered expression cassettes (30). Accordingly, recent strategies for efficient transgene expression have preferred the usage of CpG depleted vector backbones (24,31). In addition to regulatory vector elements, the transgene encoding region has also been examined for its contribution to transcription in vitro and in vivo. The in vitro methylation of the intragenic region was demonstrated to be sufficient to silence gene expression in the absence of promoter methylation (32,33). In support of these findings, CpG depleted transgenes not susceptible to methylation have been proven to be beneficial for high and sustained levels of protein expression in vivo (22,34). However, gene silencing via de novo methylation has also been described in stably transformed cell lines (35) and the influence of intragenic CpGs on the expression of a transgene might deserve further study in view of efficient production of recombinant proteins. Recent work showed that a CpG-rich reporter gene was silenced in murine cells following in vitro-methylation, whereas the unmethylated counterpart was capable of long-term and high-level expression (36). Indeed, other aspects contributing to efficient protein expression in cell culture such as improved mRNA stability or prolonged mRNA half-life as a consequence of increased GC content might put the proposed negative effects of methylation dependent silencing into perspective (37,38).

Nonetheless, the exact contribution of CpG dinucleotides onto transcription is discussed controversially. In this study, the influence of CpG dinucleotides within the open reading frame was addressed in more detail to elucidate their impact on expression and to have the ability to predict essential sequence elements for an appropriate transgene design.

Our results showed a severe reduction in reporter expression following CpG depletion from the reporter encoding region, which clearly correlated with decreased levels of newly synthesized mRNA transcripts thus implying a transcriptional-based regulation of gene expression via CpG dinucleotides in the absence of methylation.

MATERIALS AND METHODS

Plasmid constructions

A synthetic version of the green fluorescent protein (GFP) encoding gene of Aequorea victoria adapted to human codon-usage (huGFP0) and containing 60 CpG dinucleotides has been described previously (39). For construction of GFP expression vectors, the huGFP60 gene was amplified by PCR using oligonucleotides 5′-TACGAAGCCTTGCCACCATTGGTGACGCAAGGG-3′ and 5′-AGCAGCTGTACAAGTAATAGGATCCTA CT-3′. To obtain the pS/huGFP60 construct, the resulting PCR fragment was digested with HindIII and BamHI inserted into the prokaryotic expression plasmid pPCR-Script (Stratagene) providing expression control by a T7 promoter. Likewise, the fragment was cloned into the eukaryotic expression vector pcDNA5/FRT (Invitrogen) to obtain pc5/huGFP60. A synthetic huGFP gene lacking CpG dinucleotides (huGFP0) was generated by Geneart AG (Regensburg, Germany) via stepwise PCR amplification and inserted into the corresponding expression vectors resulting in plasmids pS/huGFP0 and pc5/huGFP0. Analogously, both GFP genes were cloned into a pcDNA3.1 (Invitrogen)-derived vector devoid of the neomycin resistance gene (pc3.1100) and a derivative thereof exhibiting a 53% reduction in CpG content (pc3.147). Diminishment of CpG numbers was achieved.
by depleting CpG dinucleotides from the β-lactamase gene, the multiple cloning site, parts of the pUC origin of replication and other non-coding sequences. To lower CpG content further, the conventional CMV promoter within pc3.147 was replaced by a previously described CpG free promoter version (24) resulting in pc3.132 with an overall CpG content of 32% compared to the original pc3.1100 vector. The gene encoding HIV-1 capsid protein was PCR-amplified from the formerly described synagene, which is adapted to the codon usage of highly expressing mammalian genes (9). The cytokine genes murine MIP-1α (GenBank® accession number M23447) and human GMCSF (GenBank® accession number G15899) were RNA- and codon-optimized for expression in mammalian cells with respect to codon adaptation index [CAI,(40)] and synthesized via stepwise PCR from oligonucleotides (Geneart AG, Regensburg, Germany). Based on the resulting codon-optimized genes huCA38, muMIP13 and huGM-CSF12, gene variants completely devoid of CpG dinucleotides were synthesized and designated huCA0, muMIP0 and huGM-CSF0, respectively. All genes were inserted into the eukaryotic expression vector pcDNA3.1 for transient expression experiments.

Calculation of codon frequencies

The ‘Codon Adaptation Index’ CAI represents a mean value that is calculated based on the ‘Relative Adaptiveness’ of each individual codon used for the two CpG high and CpG low alleles encoding GFP, MIP1α, HIV Capsid and GM-CSF, respectively. The Relative Adaptiveness reflects the frequency of individual codons encoding a given amino acid, whereas the most frequently used triplet is put to 1.0 and less frequently used codons are scaled down accordingly: e.g. for His there are two codons (CAC and CAT) used with a frequency of 59 and 41%, respectively, resulting in a Relative Adaptiveness of 1.0 (CAC) and 0.69 (CAT). In the case of the 2 GFP alleles, the Relative Adaptiveness was calculated and plotted on a linear scale for each individual codon position (Figure 1C). To calculate the CAI values, the individual Relative Adaptiveness values determined for each codon within e.g. huGFP0 (modified and non-modified codons) and huGFP60 were added up and divided by the number of amino acid.

Cell culture, transient transfections and infections

Adherent human lung carcinoma H1299 cells and human embryonic kidney 293 cells were grown in Dulbecco’s Modified Eagle’s Medium (DMEM) supplemented with 10% (v/v) heat inactivated fetal calf serum (FCS), 2 mM l-glutamine and 1% penicillin/streptomycin. Flp-In 293 and Flp-In CHO cells (Invitrogen) were cultivated in DMEM or HAM’s F12 medium (Invitrogen) supplemented the same way plus 1% zeocin (Invitrogen). All cell lines were maintained in a 5% CO2 atmosphere at 37°C.

For transient transfections, 2.5 x 10⁵ 293 cells or 1.5 x 10⁵ CHO or H1299 cells were plated on six-well culture dishes and transfected after 24 h with 15 μg of...
plasmid DNA using the calcium phosphate precipitation technique. For infections, 1.5 × 10^7 293 cells were plated on 12-well dishes. After 24 h, cells were washed with serum-free DMEM and were subsequently infected with a Modified Vaccinia Ankara virus strain expressing the T7-polymerase (MVA-T7) at an MOI of 10. One-hour postinfection, cells were washed and transfected with pS/huGFP_{60} and pS/huGFP_{0} constructs, respectively, using FuGENE 6 (Roche). Cells were cultured for another 12 h, and protein expression was analysed by fluorescent-activated cell sorting (FACS) as described.

**Generation of stable cell lines**

Flp-In CHO and Flp-In 293 cell lines were cotransfected with 1.5 μg of either pc5/huGFP_{60} or pc5/huGFP_{0} and 13.5 μg of pOG44 plasmid expressing the Flp recombinase gene, as described in the manufacturer’s instructions (Invitrogen). Cells were incubated for 24 h to allow for expression of the hygromycin resistance gene [hygromycin B phosphotransferase gene (hph)], and cells expressing GFP were selected by increasing concentrations of hygromycin B (PAA) in the medium. Recombinant Flp-In CHO and Flp-In 293 cell lines were maintained under selective pressure in the presence of 1 and 0.2% hygromycin B, respectively. Based upon polyclonal cell cultures expressing either of the huGFP variants, 20 monoclonal cell lines were established by singularizing GFP positive cell clones under the microscope.

**Western blot analysis and ELISA**

Cells were harvested 48-h posttransfection, washed two times in phosphate buffered saline (PBS) and lysed in RIPA buffer [50 mM Tris–HCl [pH 8.0], 150 mM NaCl, 0.1% sodium dodecyl sulfate (SDS), 1% Nonidet P-40 (w/v), 0.5% sodium deoxycholate (w/v)] supplemented with protease inhibitors (Complete Mini, Roche). Insoluble components were removed by centrifugation (10,000g, 30 min), and the total amount of protein was quantified using the Bio-Rad Protein Assay centrifugation (10,000g, 30 min), and the total amount of protein was quantified using the Bio-Rad Protein Assay following the manufacturer’s instructions. For Western blot analysis, 50 μg of total protein were separated by 12.5% SDS polyacrylamid gel electrophoresis (PAGE) and transferred to nitrocellulose. GFP expression was analysed using the GFP-specific polyclonal A. v. peptide antibody sc8334 (Santa Cruz Biotechnology) and a horse-radish peroxidase-labelled anti-rabbit antibody (Dako, D0306). Actin expression was detected using a mouse monoclonal antibody (Sigma, A5441) and a horseradish-labelled anti-mouse antibody (Dako, P0260). Cell-associated expression levels of huCA were determined by ELISA using capsid-specific monoclonal antibodies and a capsid protein standard from Polymun (Vienna, Austria). Culture supernatants were collected 48 h posttransfection and cleared from cell debris (300g, 5 min). For quantification of cytokine secretion, 1 μg of total protein diluted in growth medium was used in ELISA development systems from R&D Systems (Minneapolis, USA) for analysis of murine MIP-1α and from BD Biosciences Pharmingen (BD OptEIA Set, San Diego, USA) for analysis of human GM–CSF according to the manufacturer’s protocol. Absorbance of capsid protein and cytokine ELISA was measured at 450 nm on a Micro plate reader 680 from Bio-Rad (Hercules, USA).

**FACS analysis**

Transiently transfected cells were harvested 48-h posttransfection and resuspended in PBS/1% FCS. Stable cell lines were harvested at different times posttransfection and treated analogously. Fluorescent-activated cell sorting (FACS) was performed with a FACS Calibur from Becton Dickinson or a FACS Canto (BD), respectively. The obtained FACS data were processed using WinMDI 2.8 software (Josef Trotter, La Jolla, USA) or FACSdiva software (BD).

**Isolation of genomic DNA and DNA sequencing**

Genomic DNA from 5 × 10^6 recombinant CHO Flp-In cells was isolated with the QIAamp DNA Mini Kit according to the manufacturer’s instructions (Qiagen). 100 ng of genomic DNA were used in a PCR reaction for amplification of the complete expression cassette of the reporter genes stably integrated into the host cells using oligonucleotides 5'-ACGGCGTTGACATTGATTTGAC-3' and 5'-AATGCGATGCAATTTCCTCA-3'. The resulting products comprised 1632 bp and were sequenced subsequently.

**Bisulphite modification and DNA sequencing**

Genomic DNA from 5 × 10^6 recombinant CHO Flp-In cells was isolated with the QIAamp DNA Mini Kit according to the manufacturer’s instructions (Qiagen), and 1 μg of genomic DNA was subsequently subjected to sodium-bisulphite modification using the EpiTect Bisulitle Kit from Qiagen. Ten nanograms of the bisulphite-treated DNA were PCR-amplified in GoTaq® Green Master Mix (Promegu) together with the respective oligonucleotides designed to anneal to CpG free regions allowing for equal amplification of methylated and unmethylated DNA. To cover the entire gfp open reading frame, two separate PCR reactions were carried out using oligonucleotides 5'-TTGGTTATTATGTGAGTAAGGG-3' and 5'-TAATTATACTTCAACATTAC-3' to amplify the 5' region of huGFP_{60} and oligonucleotides 5'-AGGAGYGTATTATTTTTTAA GGATTTTTTTATTTGG-3' and 5'-AATATCTCACAATAATTCACCA CACT-3' to generate the 3' region. Likewise, the huGFP_{0} sequence was amplified using oligonucleotides 5'-AGTTTTGTATTATGTGGTTTAAGG-3' and 5'-TAAATCAATACCTCCTAATCTA-3' to generate the 5' moiety and oligonucleotides 5'-AGAGGAGTATTATTTTTTAA GGA-3' and 5'-ATATATTATACTCACAATAATTCACCA CACT-3' to amplify the 5' region and oligonucleotides 5'-GGATTTTTTTATTTG GAGATTATTTTTTAAGGA-3' and 5'-TAAATATCT TACAAATCTCACCAC A-3' to obtain the 3' region. To cover the entire CMV promoter region primers 5'-TTG TTGAGGAGATTATTTTTTTAGGA-3' and 5'-TAAATATCTTACAAATCTCACCAC A-3' were used to amplify the 5' region and oligonucleotides 5'-GGATTTTTTTATTTGAGATTATTTTTTTAGGA-3' and 5'-TAAATATCTTACAAATCTCACCAC A-3' were used to amplify the 3' region. PCR products were gel-purified (QIAquick Gel...
RNA isolation, cDNA synthesis and real-time PCR

Nuclear and cytoplasmic RNA species were prepared using the RNeasy-Kit following the manufacturer’s instructions (Qiagen). RNA samples were then treated with RNase-free DNase (Roche) and subsequently used in a PCR with transgene-specific primers to exclude DNA contamination. One microgram of DNA-free RNA was reverse transcribed using the first strand cDNA synthesis kit for real-time PCR (RT-PCR) (AMV) from Roche together with 3.2 μg p(dN)6 random primer in the presence of RNase inhibitor. To detect alternative splice products, 1 μl of resulting cDNA samples was used in a PCR reaction together with primers 5'-AAGCT TGCCACATGGTG-3' and 5'-GAGCTGTCAAGGT ATAGGATCC-3' amplifying the complete open reading frame of huGFP0 and huGFP0. RT-PCR was carried out in a LightCycler (Roche) with the LC Fast Start DNA Polymerase (Roche) with 10-fold serial dilutions (10^2–10^10 copies/μl) of resulting cDNA samples was used for the generation of according PCR fragments. The relative transcription level of the gfp transgenes was related to hph transcripts, and resulting error values ranged between 0.01 and 0.03.

Analysis of mRNA half-life

Decay of respective mRNA transcripts was measured as described previously (41): For absolute quantification, mRNA amounts were extrapolated from an external standard curve generated with 10-fold serial dilutions (10^2–10^10 copies/μl) of the respective plasmid DNA template. Transcript copy numbers were then calculated from the external standards via an internal calibrator by LightCycler Software 3.5 (Roche).

Chromatin immunoprecipitation

Exponentially growing CHO cells were cross-linked using 1% formaldehyde for 5 min. Cross-linking reactions were quenched with 0.125 M glycine. Cells were washed three times in 1× PBS. Hypotonic lysis of cells was performed by 15 min incubation on ice in 1% SDS, 10 mM EDTA, 50 mM Tris–HCl pH 8.1, EDTA-free complete protease inhibitor cocktail (Roche) containing buffer. Samples were sonicated using a Bioruptor sonicator (Diagenode) to approximately 500 bp DNA fragments. Cell debris was spun at 16100 g for 5 min at room temperature and the supernatant fraction was transferred into a new tube. This chromatin fraction was stored at 4°C for 3.5 days and used as described above. After determination of decay constant k, the respective mRNA half-life was calculated.
input for immunoprecipitations. Chromatin samples were diluted 1:10 with 150 mM NaCl, 20 mM Tris–HCl pH 8.1, 2 mM EDTA, 1% Triton X-100, EDTA-free complete protease inhibitor cocktail (Roche) containing buffer. 5 μg of a polyclonal histone H3 specific antibody (Abcam, ab1791) was added to 500 μg chromatin and incubated in chilled buffer overnight at 4°C with gentle rotation. Beads were washed subsequently with 1 ml each of the immunoprecipitation buffer (5x), in 250 mM LiCl, 5% NP-40, 0.5% deoxycholate, 1 mM EDTA, 10 mM Tris–HCl, pH 8.1 containing buffer (1x), and in 1×TE (1x). Elution was performed using 1% SDS, 0.1 M NaHCO₃ at 37°C for 2×15 min in a shaker. Eluted chromatin fractions were subjected to RNaseA treatment for 1 h at 37°C followed by proteinase K digestion at 37°C, 8 h and reversal of cross-linking at 65°C for a further 8 h. DNA was precipitated and dissolved in 50 μl of H₂O. DNA samples were then diluted 1:10 with water and used as a template in quantitative PCR reactions. Chromatin immunoprecipitations (ChIPs) were performed from different chromatin preparations and chromatin samples incubated with normal rabbit IgG (SantaCurz, sc-2027) served as negative control.

RESULTS

The GFP reporter genes

Although CpG dinucleotides have been shown to affect efficiency of transgene expression in vivo, the consequences of CpG-mediated regulation on high-level expression of foreign genes in established cell lines still remain unknown. In order to analyse the influence of intragenic CpG content on transgene expression, a humanized version of the GFP (huGFP₀) was used as reporter gene (39). Apart from quantitative determination of reporter expression, GFP was selected to allow evaluation of transcriptional control of the CMV major immediate early promoter or the human elongation factor 1α promoter, respectively, and the bovine growth hormone polyadenylation signal (BGH) (Figure 1D).

The CpG quantity in the open reading frame influences the reporter activity in transiently transfected mammalian cells

To investigate the influence of intragenic CpG content on protein expression in cell lines, H1299 cells were transiently transfected with the respective huGFP₆₀ and huGFP₀ constructs, and GFP expression was determined by FACS analysis. Interestingly, CpG depletion in huGFP₀ consistently resulted in a 1.6 or 2-fold decrease in reporter activity as compared to huGFP₆₀, likewise in the context of the commercial plasmids pc3.1₀₀₀ or pc₅, both of which support transgene expression via the CMV immediate early promoter (Figure 2A). Since transfection efficiencies were comparable for both GFP variants irrespective of the vector backbone used (Figure 2B), the huGFP₀-derived reduction of auto-fluorescence presumably results directly from the specific depletion of intragenic CpGs or the general alterations in overall plasmid CpG content. To address this question, we analysed expression of the huGFP gene variants within vectors pc₃.₁₀₀₀ and pc₃.₁₇₇ comprising CpG contents in the backbone of 100 and 47%, respectively. As displayed in Figure 2C, reporter activities of both huGFP variants were not altered significantly when expressed from different vector backbones. Beyond the CpG reduction within the vector backbone, the conventional CMV promoter within pc₃.₁₇₇ was replaced by a previously described CpG free CMV promoter version (24) resulting in an overall CpG content within the complete plasmid of 32%. This promoter-specific CpG reduction slightly decreased the absolute expression levels of both reporter genes (data not shown) while the relative difference in expression was maintained. Thus, the decrease in huGFP₀-derived fluorescence apparently is directly associated with intragenic CpG depletion. To examine whether the observed effect was universally valid rather than restricted to the gfp gene, three more RNA- and codon-optimized reporter genes and their CpG-depleted counterparts were tested. The viral gene encoding the capsid protein of human immunodeficiency virus type-1 (huCA; 699 bp) and two endogenous cytokine genes encoding murine macrophage inflammatory protein 1z (muMIP; 279 bp) and human granulocyte–macrophage
Intragenic CpG depletion reduces expression in stably transfected mammalian cells

Due to the apparent advantages concerning information content of GFP read out, the gfp reporter genes were selected to further evaluate potential effects of CpG depletion on reporter expression. For this purpose we used CHO and 293 cell lines, both being widely employed for recombinant protein production (48). Corresponding cell lines stably expressing the reporter gene variants within otherwise identical genetic backgrounds were established using the Flp-In recombination system. The integrity of the reporter gene expression cassettes was verified by PCR of isolated genomic DNA and sequencing. Flp-In CHO and 293 cell clones expressing single copies of either huGFP$_{60}$ or huGFP$_{60}$ genes were selected in the presence of hygromycin and analysed by fluorescence microscopy (data not shown) and FACS analysis (Figure 4A). In accordance with data obtained from transient expression,
Figure 3. Influence of intragenic CpG depletion on transient expression of viral and cytokine reporter genes. (A) H1299 cells were transiently transfected with 15 μg of the indicated constructs in pcDNA3.1. Forty-eight-hours posttransfection, normalized cell lysates (CA) and supernatants (muMIP and huGM) were screened for protein content by ELISA as described in materials and methods. Protein amounts obtained from transfections with the CpG-rich genes were set to 100%, and values for the CpG lacking variants were related accordingly. The mean of four independent transfection experiments is shown, and standard deviations are indicated. (B) Comparison of sequence-specific parameters. CpG: CpG dinucleotide; CAI: codon adaption index; GC content: percentage of bases C and G within the overall sequence; TpA: TpA dinucleotide. (C) Relative Adaptiveness distribution of the CA, muMIP and huGM gene variants. The Relative Adaptiveness reflects the frequency of individual codons, whereas the most frequently used triplet encoding a given amino acid is set to 1.0 and less frequently used codons are scaled down accordingly.
Figure 4. Influence of intragenic CpG content on long-term stable GFP expression in mammalian cells. (A) CHO (left) and 293 cells (right) stably expressing huGFP₀ or huGFP₆₀ were subjected to FACS analysis counting 10,000 events per gate. The number of scored events (y-axis) versus the respective fluorescence intensity of GFP (x-axis) is shown for huGFP₀ and huGFP₆₀ as well as for untransfected cells (mock) to monitor background fluorescence. (B) Stably transfected CHO cells were harvested and 50 μg of total protein were subjected to western blot analysis using a GFP-specific antibody. Two monoclonal and one polyclonal CHO cell line expressing either huGFP₀ (lanes 2–4) or huGFP₆₀ (lanes 5–7) were analysed. The monoclonal cell lines for huGFP₀ are shown in lanes 2 and 3, the monoclonal cell lines for huGFP₆₀ are depicted in lane 5 and 6. The positions of the 27-kDa GFP protein and the 42-kDa β-Actin protein are indicated by arrows. Untransfected cells (mock, lane 1) served as negative control. (C) Polyclonal (poly) and monoclonal (mono) CHO cell lines (left) and polyclonal 293 cells (right) stably expressing either huGFP₀ or huGFP₆₀ were quantified for expression of the GFP reporter genes by FACS analysis over a period of 56 weeks (x-axis) indicated as mfi values (y-axis). Mocktransfected cells were used for background subtraction. (D) CHO cells stably expressing huGFP₀ or huGFP₆₀ under control of the EF-1α promoter were subjected to FACS analysis counting 10,000 events per gate. The number of scored events (y-axis) versus the respective fluorescence intensity of GFP (x-axis) is shown for huGFP₀ and huGFP₆₀ as well as for untransfected cells (mock) to monitor background fluorescence. (E) Polyclonal CHO cells stably expressing either huGFP₀ or huGFP₆₀ under control of the EF-1α promoter were quantified for expression of the GFP reporter genes by FACS analysis over a period of 15 weeks (x-axis) indicated as mfi values (y-axis). (F) Genomic DNA from stably transfected CHO cells was isolated, bisulphite-treated and amplified via PCR with huGFP specific oligonucleotides. PCR products were sequenced, and chromatograms were analysed as described in materials and methods. A section of huGFP₆₀ sequence prior to (upper panel) and post bisulphite treatment (lower panel) is depicted, and cytosine residues potentially available for methylation are underlined. Unmethylated cytosines are represented by thymines the chromatogram, whereas methylated cytosines are not converted. The first nucleotide (G) of the extracted sequence corresponds to position 138 of the gfp open reading frame.
cells stably expressing huGFP\textsubscript{0} showed a significant loss in GFP-mediated fluorescence with mean fluorescence intensity (mfi) values determined to be 7-fold (CHO) and 10-fold (293) lower than measured for the huGFP\textsubscript{60} expressing counterparts. This effect was observed in all established monoclonal cell lines (15 monoclonal cell lines for each gfp gene variant) irrespective of the cell type analysed (data not shown), indicating that cell-specific effects did not play a role within this context. To further exclude that compromised expression of huGFP\textsubscript{0} was due to incomplete production of the full-length gene product or rapid protein degradation, various cell samples derived from stably transfected polyclonal and monoclonal CHO cell lines were subjected to western blot analysis with a GFP-directed antibody. As expected, detectable GFP amounts differed significantly in cells expressing huGFP\textsubscript{60} or huGFP\textsubscript{0} while truncated proteins could not be detected in corresponding cell lysates (Figure 4B), suggesting correct protein synthesis of the CpG-depleted GFP variant.

To prove the promoter independency of this CpG-mediated expression effect, we also tested EF-1\textzeta-driven GFP expression. The choice for the human elongation factor 1\textzeta promoter was based on its mammalian origin and its strong and at the same time gene silencing evading nature. Expression analysis confirmed the CMV derived results and showed a clear correlation between intragenic CpG content and protein expression levels (Figure 4D).

As silencing of gene expression is rather a gradual process manifested in the course of time (49), we monitored whether depletion of CpGs from the GFP coding region is beneficial for long-term protein expression in mammalian cells. Therefore, CHO cells stably expressing the huGFP variants under control of the CMV and the EF-1\textzeta promoter, respectively and under constant selective pressure were weekly subjected to FACS analysis. During this observation period, all cell lines showed relatively constant huGFP\textsubscript{60} expression levels with average mfi values of 750 for the CMV-driven GFP expression in CHO and 293 cells and 250 for the EF-1\textzeta-driven protein expression in CHO cells, respectively (Figure 4C and E). Likewise, expression of CpG-depleted huGFP\textsubscript{0} reporter genes remained constant, however, at lower levels which were reduced by 6- to 9-fold when compared with huGFP\textsubscript{60}. Although a steady expression under selective pressure is not surprising, these data assured unchanging experimental conditions for further studies, as neither huGFP\textsubscript{60} nor huGFP\textsubscript{0} reporter expression was silenced over the observed period.

**Expression of the CpG-rich gfp gene is not impaired by intragenic CpG methylation**

It is now firmly established, that deficient gene expression often correlates with CpG hypermethylation within and around promoter regions. However, long-term effects of intragenic CpG methylation have been explored less extensively. Since the established cell lines were under constant selective pressure and huGFP\textsubscript{60} showed a constant expression profile over >56 weeks, silencing via CpG methylation is unlikely. Nonetheless, we found it intriguing to what extent CpG dinucleotides within huGFP\textsubscript{60} might trigger cytosine hypermethylation in the genomic background. Thus, huGFP\textsubscript{60} DNA from stably transfected CHO cells cultivated for >50 weeks was subjected to bisulphite genomic sequencing. Sodium bisulphite selectively deaminates unmethylated cytosines to uraciles, whereas methylated cytosines are not modified (50). In the subsequent PCR reaction, uraciles are replaced by thymines resulting in a final C to T conversion. To validate the method, the pc3.1\textsubscript{0}/huGFP\textsubscript{60} plasmid was subjected to quantitative in vitro methylation prior to bisulphite sequencing and resulting chromatograms indicated that in vitro-methylated cytosines were not affected by bisulphite treatment (data not shown). In contrast, analysis of the genome-derived huGFP\textsubscript{60} sequence revealed a quantitative conversion of all cytosines to thymines following the bisulphite reaction. This clearly shows that none of the CpG dinucleotides within the huGFP\textsubscript{60} open reading frame was originally methylated (Figure 4F). As methylation of CpG dinucleotides within promoter regions is also known to diminished transgene expression we analysed the methylation status of the integrated CMV promoter of the stable cell lines, promoting transcription of either huGFP\textsubscript{0} or huGFP\textsubscript{60}. The results clearly showed that under the tested assay conditions neither the promoter of huGFP\textsubscript{0} nor of huGFP\textsubscript{60} exhibited CpG methylation, thus excluding a promoter-methylation specific influence on the transcription of huGFP\textsubscript{0} and huGFP\textsubscript{60} (Supplementary Data; Figure 3).

As expected, these results imply that under selective conditions the genome-integrated CpG-rich reporter gene was not susceptible for intragenic hypermethylation within the observed period.

**Intragenic CpG depletion has no effect on translational efficiency**

Since CpG depletion from the huGFP reading frame was associated with a clear loss of reporter expression, we next examined, to what extent this phenomenon might result from decreased efficiency of protein translation due to the changes in codon usage. In order to exclude potential effects of the modified reporter gene sequence on posttranscriptional events like RNA export, we infected 293T cells with a modified vaccinia ankara strain known to replicate exclusively in the cytoplasm and expressing the T7 polymerase (MVA-T7). Upon infection this system allows the cytoplasmic transcription of a transfected reporter gene controlled by the T7 promoter. The translation efficiency of the pS/huGFP\textsubscript{60} and pS/huGFP\textsubscript{0} constructs following transfection of the MVA-T7-infected cells was quantified by FACS analysis (Figure 5A) while the amount of gfp-specific transcripts was assessed by quantitative PCR (Figure 5B). Both reporter genes produced almost equal amounts of gfp mRNA copy numbers and also comparable amounts of GFP in this artificial system, widely excluding a major impact of CpG content on translation although possible saturated translation conditions in the artificial T7/vaccinia system cannot be completely ruled out.
Intragenic CpG content correlates with levels of steady-state RNA

To further elucidate the mechanism underlying the significant decrease in huGFP\(_0\) expression, we asked, whether CpG depletion had a major influence on related mRNA amounts. To quantify the steady-state level of huGFP transcripts, cytoplasmic and nuclear RNA fractions from stably transfected CHO and 293 cells were subjected to reverse transcription and quantified via RT-PCR. Whereas the overall number of hph transcripts within cytoplasmic and nuclear fractions of CHO cells remained constant (Figure 6A), relative quantification analysis, the most accurate method to quantify cDNA (41), yielded a 7-fold cytoplasmic and 13-fold nuclear reduction of huGFP\(_0\) transcripts compared to huGFP\(_{60}\) mRNA levels. Even more striking effects were obtained from stably transfected 293 cells, where the total amount of huGFP\(_{60}\) transcripts exceeded those of huGFP\(_0\) by 29-fold (data not shown). The specificity of obtained PCR products was evaluated by sequence verification and melting curve analysis. To confirm the data gained by relative quantification, absolute amounts of gfp and hph transcripts in corresponding fractions of CHO cells were determined using external standard curves (Figure 6B). According to this method, the CpG depleted mRNAs were decreased by 5-fold in the nucleus and 8-fold in the cytoplasm compared to the CpG containing counterparts. In sum, these data provide evidence that the observed reduction of huGFP\(_0\) reporter expression strictly correlates with decreased steady-state mRNA levels. Since ratios of cytoplasmic versus nuclear transcript numbers calculated from absolute quantification data were similar for both reporter species (12- and 9-fold for huGFP\(_{60}\) and huGFP\(_0\), respectively), we can exclude that effects associated with mRNA nuclear export contribute to differential GFP expression.

CpG variations do not affect stability or splicing of reporter transcripts

Next we examined, whether CpG reduction in the huGFP\(_0\) reporter gene influenced the stability of corresponding RNA messages. Thus, we determined the half-lives of huGFP\(_{60}\) and huGFP\(_0\) transcripts in stably transfected CHO cells. For this purpose, de novo RNA synthesis was blocked by Actinomycin D for different time periods, and total RNA was subjected to reverse transcription and quantified (Figure 7A). As further depicted in Figure 7B, \(\beta\)-actin mRNA isolated from the huGFP expressing cells exhibited consistent half-life values of 3.5 h, whereas half-lives of \(\sim\)3 h were determined for both gfp reporter transcript species, indicating that codon substitutions within huGFP\(_0\) did not impact transcript stability. Despite diligent design of the huGFP\(_0\) variant, CpG depletions by means of introducing silent mutations generated four additional cryptic splice sites [huGFP\(_{60}\) (8) versus huGFP\(_0\) (12)], which might in theory cause covert alternative splicing and thereby reduce the amount of detectable full-length transcripts. To deal with this issue, respective cytoplasmic and nuclear mRNA fractions were subjected to reverse transcription, and resulting cDNA samples were used in qualitative PCR with oligonucleotides amplifying the whole GFP coding region. Resulting PCR products were analysed and sequence-verified. As indicated in Figure 7C, only full-length messenger transcripts were detected for either GFP variant, which could be confirmed in northern blot analyses using a BGH-specific probe (data not shown). The corresponding results strongly argue against alternative splicing events as a cause for decreased huGFP\(_0\) transcript levels. In sum, these data imply that reduced protein and mRNA levels observed for a CpG depleted GFP version in vitro cannot be ascribed to reduced mRNA stability or erroneous splicing.

A loss of de novo transcriptional activity is responsible for moderate expression of the CpG depleted GFP variant

Apart from posttranscriptional processes modulating transgene expression, transcriptional dynamics crucially contribute to transgene-specific RNA synthesis. To investigate, whether differential transcriptional activity may account for decreased huGFP\(_0\) mRNA amounts, a nuclear run-on assay was performed using nuclei from

![Figure 5](https://academic.oup.com/nar/article-abstract/38/12/3891/2409319/3901)
stably transfected CHO cells. Captured de novo synthesized RNA molecules were subjected to reverse transcription and quantified by qPCR analysis and referred to de novo synthesized β-actin transcripts (Figure 8). Interestingly, the CpG depleted huGFP<sub>0</sub> construct yielded a 7-fold reduced amount of de novo synthesized GFP mRNA as compared to huGFP<sub>60</sub>. This thoroughly reflects the
observed discrepancies in RNA and protein expression levels. Taken together, our data propose a clear correlation of intragenic CpG content and de novo transcriptional activity, implying a transcription-based regulation of transgene expression via CpG dinucleotides.

Intragenic CpG dinucleotides affect nucleosome positioning

Having experienced the relation between CpG dinucleotides and RNA de novo synthesis, we assumed a CpG mediated effect on chromatin structure, which affects the activity of gene expression. To initially test the impact of CpG dinucleotides and sequence variation on nucleosome positioning, we performed in vitro nucleosome reconstitution assays with the variant huGFP DNA sequences. As the open reading frame of the huGFP gene variants comprises 720 bp and hence is too large to reveal the positions of individual nucleosomes by native electrophoretic mobility shift assays (EMSAs), the huGFP coding sequence was partitioned into three overlapping DNA fragments of 280–300 bp and prepared by PCR amplification. Mononucleosomes were reconstituted by the salt dialysis method and comparative analysis of the nucleoprotein complexes revealed distinct band patterns for each of the analysed DNA fragments of the two gene variants. This suggests that changing the level of intragenic CpG content directly affects sequence dependent nucleosome positioning characteristics within the huGFP constructs in vitro (Figure 9). Whereas the proximal DNA element of the huGFP0 construct shows a decreased number of different nucleosome positions...
Considering that nucleosome positions in vitro do not alter nucleosome density in vivo, we examined whether intragenic CpG dinucleotides affect active chromatin assembly in stable cell lines (6,23,54). The observed changes in in vitro nucleosome positioning could affect chromatin structure and gene expression in vivo.

CpG dinucleotides do not alter nucleosome density

We investigated whether intragenic CpG dinucleotides affect active chromatin assembly in stable cell lines. We used a ChIP experiment to compare the nucleosome density around the centre of the DNA fragment as compared to the huGFP_{60} construct, the central and the terminal DNA elements exhibit a more distributed pattern with additional nucleosome positioning sites. Changes in the DNA sequences do directly affect the histone-binding properties of the DNA which were shown to contribute to nucleosome positioning in vivo (51). The observed changes in in vitro nucleosome positioning could affect chromatin structure and gene expression in vivo.

DISCUSSION

Various strategies have been exploited to improve and optimize efficacy of plasmid DNA vectors for recombinant protein production in vitro. Amongst others, fusion of intronic sequences S' to the coding region, usage of strong hybrid or cell-specific promoters or insertion of elements promoting the formation of transcriptionally active chromatin-like structures may serve as potent tools to increase protein synthesis in established cell lines (6,23,54).

Beyond, adaptation of the codon usage to highly expressing genes of the respective organism has been shown to significantly augment transgene expression (8). However, conflicting results have been reported for the endogenous content of CpG dinucleotides potentially modulating transgene expression in vivo as well as in cell culture. Whereas usage of CpG depleted expression cassettes has clearly shown to boost expression levels in transgenic mice and somatic cells, arguing that corresponding sequences are not prone to methylation-dependent gene silencing (24,28,31), the presence of CpG-rich regions might otherwise be beneficial for postranscriptional processes based on the significant increase in RNA stability (9,38,39). Interestingly, expression of transgenes in stable cell lines has been shown to be...
critically influenced by differentiation status of the used cell type, integration locus and orientation, the usage of weak or strong promoter/enhancer units or simply maintenance of selective pressure (34,35,55,56) which altogether might result in controversial interpretation of CpG-mediated effects.

To explore whether the intragenic CpG content directly influences transgene expression in established mammalian cells, we synthesized a codon-optimized GFP reporter gene completely devoid of CpG dinucleotides and compared expression efficiency to its CpG-rich equivalent. Our data demonstrated a clear and direct correlation of intragenic CpG content with reporter activity in vitro, resulting in significantly reduced expression and fluorescence of the CpG lacking GFP variant. The associated phenotype was observed irrespective of the used cell type and promoter, and could not be ascribed to aberrant protein production. Furthermore, this CpG-specific effect was successfully transferred to unrelated genes of viral or mammalian origin which points to a general mechanism driven by CpG dinucleotides.

Transcriptional silencing via extensive CpG hypermethylation associated with chromatin condensation is a well-documented phenomenon in therapeutic DNA applications but has also been reported to interfere with cell culture-based production of recombinant proteins over time (35). Different strategies such as inserting CpG island fragments into expression vectors (57) or supplementing culture medium with demethylating agents (58) have been applied to avoid methylation-dependent silencing thereby improving expression efficiencies of foreign transgenes in CHO cells. To further reveal any chromatin structure-dependent consequences of DNA methylation, we compared transient GFP production with expression in stably transfected cell lines, where transgenes are integrated at defined sites and may be potentially influenced by adjacent genomic CpG hypermethylation patterns. As compared to transient experiments, differential expression of the gfp genes was even more pronounced upon stable transfection, where reporter activity of the CpG-rich GFP variant exceeded the CpG lacking counterpart by 6–9-fold and 10–20-fold in the analysed cell lines, respectively.

Interestingly, decreased reporter activity of similar CpG free gfp genes expressed in cultured cells or in transduced mice has also been observed by others suggesting that deficiencies in reporter activity might originate from either posttranscriptional events or less efficient translation of the CpG-depleted transcript due to modified codon usage (34,59). In order to avoid effects of codon quality on protein production, we have used gfp genes comprising similar CAI values as indicators of codon quality (40) and comparable codon distributions. Indeed, equal gfp-specific transcripts and GFP amounts were obtained for both reporter genes in a MVAT7-based cytoplasmic transcription/translation system, widely excluding a contribution of differential translational activities to the loss of reporter expression seen in huGFPα stable cell lines.

In contrast we could provide clear evidence that diminished levels of the CpG-depleted GFP strictly correlated with a significant decrease in steady-state RNA copy numbers. Since downmodulation of reporter gene expression as reflected by decreased RNA amounts might yet become manifest at the posttranscriptional level, we analysed potential effects of CpG depletion on splicing, translocation or stability of the respective RNA messages. Our results, however, clearly demonstrated that reduced RNA levels could not be ascribed to erroneous splicing events, nuclear export restrictions or diminished RNA half-life, indicating that corresponding mechanisms were not significantly influenced by alterations of the intragenic CpG content. In contrast to this finding, compromised RNA stability was formerly shown to be responsible for the low levels of steady-state RNA detected upon CpG depletion (38). Noteworthy the study at hand was performed with GFP variants comprising equal TpA numbers, while the changed stability was detected upon maximization of TpA numbers in the CpG-free gene variant. Since UpA sites emerging in the corresponding transcripts serve as preferential targets for cellular endonucleases (37,60), an excess of these motifs is likely associated with rapid RNA degradation, thus explaining the reported effects.

Since we could not provide any evidence for altered posttranscriptional regulation of CpG-depleted genes, decreased RNA levels rather pointed to altered transcriptional events. This hypothesis was confirmed by nuclear run-on analysis revealing a clear positive correlation of intragenic CpG content and the amount of de novo synthesized mRNA transcripts. The in vitro analysis of nucleosome positioning gave first hints, that the variations of CpG dinucleotide content has an influence on the chromatin structure along the open reading frames, since nucleosome reconstitution exhibited distinct nucleosome positioning patterns on the related DNA elements. This observation together with the finding that DNA sequence is sufficient to direct nucleosome positioning in vivo (51) it may well be that sequence dependent nucleosome positioning and stability may contribute to the observed transcriptional disparity. Yet, ChIP experiments could not reveal differences in the overall histone H3 occupancy of the reporter gene variants. However, ChIP experiments only resolved histone H3 density, but did so far not address the presence of histone modifications, histone variants or even the presence of transcription factors. Hence, future experiments have to deal with the detection of histone modifications as well as the analysis of transcription factors present at the transcription start sites of the gfp gene variants and RNA Polymerase II at the coding region.

In light of these findings it is worth mentioning that the ubiquitously expressed and CpG binding protein CFPI (CXXC finger protein 1) has previously been shown to transactivate exclusively CpG containing promoter regions (61). This nuclear transcriptional activator specifically binds to unmethylated CpG dinucleotides via a CXXC zinc finger motif, the affinity for target sequences increasing with accumulative numbers of CpGs (62). Hence, CFPI might play a role as an epigenetic regulator in modulating gene expression via CpG dinucleotide methylation and histone modifications (63–65).
Furthermore, consensus CFP1 binding motifs specified as (A/C)CpG(A/C) (62) could be mapped within the CpG rich gfp reporter gene, promoting a possible interaction with CFP1. Electrophoretic mobility supreraseph assays of a recombinant CFP1 protein suggest binding to the CpG-containing gfp gene but not to the CpG-free gfp allele (data not shown). However, upon the availability of appropriate antibodies ChiP experiments may be well suited to elaborate different gene occupancies of CFP1 between CpG high and CpG low reporter gene containing cell lines. Besides, it remains to be examined whether high-level transcription of a CpG-rich gene as compared to its CpG depleted counterpart is directly associated with efficient intragenic binding of another corresponding trans-activator or a related cellular protein.

Herein we have shown that depletion of CpGs from the coding region of a reporter gene correlated with a clear loss of expression originating from attenuated transriptional activity. In further support of these findings, reduced de novo transcription could also be observed for other CpG-free gene variants (unpublished data) indicating that this effect was not gene specific. Concerning the interaction of the huGFP gene variants with chromatin related factors, it could be shown that site-specific preferences for nucleosome formation were determined among the huGFP variants. It is suggested that the distribution of CpG dinucleotides, which differs in the huGFP fragments is responsible for variable binding efficiencies and positions. The presence of a certain sequence pattern, predicted to enable efficient nucleosome binding, however have yet to be identified. A 3 bp periodicity of CG and GC dinucleotides has been found to be a highly nucleosome favoured sequence (66). It is suggested that different nucleosome preferences displayed by nucleosome reconstitutions are at least partly due to altered bending flexibility.

It remains to be seen whether transgene expression can be further increased by introduction of additional CpGs into otherwise moderately expressing genes. In sum, these findings provide a new perspective for the application of transgenes comprising a high intragenic CpG density for the rational design of recombinant expression cassettes.

ACKNOWLEDGEMENTS

The authors thank Gerd Sutter (Paul Ehrlich Institute, Langen, Germany) for kindly providing the MVA-T7 virus strain and the infection protocol. They also thank Marcus Graf (GENEART AG) for calculation of codon frequencies.

REFERENCES


FUNDING

BMBF grant 01K10211 (to R.W.). Funding for open access charge: BMBF grant 01K10211 (to R.W.).

Conflict of interest statement. None declared.


