SCIENCE CHINA
Life Sciences

• RESEARCH PAPER •

January 2010 Vol.53 No.1: 94–100
doi: 10.1007/s11427-010-0003-x

# A "GC-rich" method for mammalian gene expression: A dominant role of non-coding DNA GC content in the regulation of mammalian gene expression

JIA Qian[1*], WU HongTao[1], ZHOU XingJun[1], GAO Jian[1], ZHAO Wei [1], Aziz JouDi[2,4], WEI JingShuang[1], HOU LiHua[5], WU ShuYin[4], ZHANG Ying[3], DONG XiangFeng[1], HUANG YanMin[1], JIN WeiYuan[4], ZHU HongJie[3], ZHAO XinHui[5], HUANG ChunHua[6], XING LiPing[4], LI LiWen[4], MA Jun[4], LIU XiYan[4], TAO Ran[4], YE ShuaiDong[4], SONG YiGao[4], SONG LingLing[4], CHEN GuanPing[4], DU ChunLing[4], ZHANG XueTing[3], LI Bo[3], WANG YanTao[3], YANG Wei[3], Gilbert Rishton[2], TENG YuYang[3], LENG GouQing[3], LI LuanFeng[4], LIU WenXian[1], CHENG LiJun[1], LIANG QiuBo[3], LI ZhengWu[3], ZHANG XiuQin[3], ZUO YaJun[6], CHEN Wei[5*], LI HuiCheng[3] & Matthew (Mizhou) HUI[1,2,3,4*]

[1]*NCPC R&D Center, Shijiazhuang 050015, China;*
[2]*AmProtein Inc, CA 93010, USA;*
[3]*Harbin Pharmaceutical Group R&D Center, Harbin 150086, China;*
[4]*AmProtein China, Hangzhou 310019, China;*
[5]*Institute of Microbioloy and Epidemiology, Academy of Military Medical Sciences, Beijing 100071, China;*
[6]*Shanghai Huayi Biotech, Shanghai 200233, China*

Received July 22, 2009; accepted August 24, 2009

High mammalian gene expression was obtained for more than twenty different proteins in different cell types by just a few laboratory scale stable gene transfections for each protein. The stable expression vectors were constructed by inserting a naturally-occurring 1.006 kb or a synthetic 0.733 kb DNA fragment (including intron) of extremely GC-rich at the 5′ or/and 3′ flanking regions of these protein genes or their gene promoters. This experiment is the first experimental evidence showing that a non-coding extremely GC-rich DNA fragment is a super "chromatin opening element" and plays an important role in mammalian gene expression. This experiment has further indicated that chromatin-based regulation of mammalian gene expression is at least partially embedded in DNA primary structure, namely DNA GC-content.

**GC-rich, GC content, DNA structure, stable expression, mammalian cell, expression mechanism**

**Citation:** Jia Q, Wu H T, Zhou X J, *et al*. A "GC-rich" method for mammalian gene expression: A dominant role of non-coding DNA GC content in the regulation of mammalian gene expression, Sci China Life Sci, 2010, 53: 94–100, doi: 10.1007/s11427-010-0003-x

Prokaryotic cell gene expression is mainly biochemically regulated by the binding of transcriptional factors and other DNA binding proteins to gene regulatory sequences, such as the promoter and enhancer regions. Eukaryotic cells have a nucleus each, containing condensed chromatin that may prevent access of the transcriptional factors and other DNA binding proteins to the gene regulatory sequences embedded inside the condensed chromatin. Therefore, eukaryotic cells have a non-open genome. The openness of the chromatin plays an important role in regulation of eukaryotic gene

*Corresponding author (email: qianjiazh@hotmail.com; chenwei0226@yahoo.com.cn; mhui@amprotein.com)

expression. What is the material foundation that determines chromatin's openness? Is this foundation embedded in primary structure of DNA, namely the GC content of the DNA? In this article, we will address these questions, at least in part.

Going back in time, we know that dinosaurs were extremely large, muscular organisms. Achieving this huge frame must have required super gene expression strength of its major body components, such as beta-actin. Indeed, dinosaurs have been linked to birds including *Gallus gallus,* the domestic chicken [1,2]. Surprisingly, the chick *beta-actin* gene contains a 1.006 kb extremely GC-rich, non-coding sequence, Intron-1 [3] (average 75.3% GC content) (Figure 1). The segment with the highest GC content identified in this sequence was composed of 90.8% GC (Table 1). This extremely high GC content is unusual in mammalian and other vertebrates' genomes [4,5].

GC and AT bonds can be thought of as different building materials used in constructing DNA secondary structure, the double-helix. The Guanine-Cytosine (GC) bond is comprised of three hydrogen bonds while two hydrogen bonds secure the Adenine-Thymine (AT) bond. The difference in the number of hydrogen bonds between these pairs imparts differential rigidity to the DNA double-helix (secondary structure). Due to a higher number of hydrogen bonds, the GC union is stronger than that of Adenine and Thymine. Likely, a DNA double-helix composed of GC bonds is more rigid than made of AT bonds and form different chromatin structures that may interact with DNA methylation, histone methylation & accetylation, and non-histone proteins differently [6–8] and may make an adjacent structural gene more or less open to transcriptional factors and other DNA binding proteins.

We hypothesized that genomic DNA containing highly GC-rich content might hold the secret of a high constitutive level of mammalian or possibly all eukaryotic gene expression through regulating chromatin openness.

## 1 Materials and methods

### 1.1 Plasmids, cell lines and culture conditions

pBR322 backbone containing an Ampicillin selection marker was used to construct the pMH3 expression vector (Figure 2). A *Pvu*I site in the middle of Ampicillin selection marker was used for linearization of the pMH3 expression vector for stable gene transfection into mammalian and chicken cells. CHO-S (SKU# 11619-012) was purchased from Invitrogen. CHOK1 (ATCC CRL 9618) was purchased from ATCC. R1610 (ATCC CRL 1657) was purchased from ATCC. DF-1 (chick) was a gift from Dr. Bu ShiGao (Harbin Veterinary Research Institute, Chinese Academy of Agriculture Science).

CTGCAGTGACTCGAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCC
GCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTC
CTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGC
CTTAAAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGT
GTGTGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGG
CGCGGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGC
CCCGCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGG
GGTGAGCAGGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCCTGCACCCCCCTCCCCG
AGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCG
CCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGG
CCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGC
GGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC
CAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCGA
AGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGC
CGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGGGGAC
GGGGCAGGGCGGGGTTCGTCGGCGCCGGCGGGGTTTATATCTTCCCTTCTCTGTTCCTCCGCA
GCCCCCAAGCTT

**Figure 1** GC-rich chick beta-actin intron-1 (1.006 kb). There are seven GGCGGG box, one CCCGCC box [9], and 136 CG sites in the above sequence.

**Table 1** Analysis of GC content of chick beta actin gene intron-1

| Positions | 1–100 bp | 200–300 bp | 330–430 bp | 520–650 bp | 750–830 bp |
|---|---|---|---|---|---|
| GC content | 78.0% | 82.0% | 80.0% | 90.8% | 80.0% |

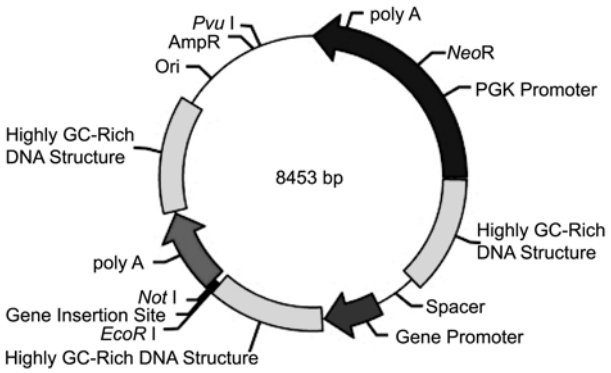


**Figure 2** pMH3 expression vector.

### 1.2 Construction of the expression vectors

A 1.494 kb chick actin gene promoter fragment was digested by *Pst*I and *Hind*III from 5′-regulatory sequence (a gift from Dr. N Fregien, University of Miami), and was used as a promoter for the vector construction. Rabbit globulin poly A was employed for the construction. The 1.006 kb chick *beta-actin* gene intron-1 modified with *Pst*I at 5′-end and *Hind*III at 3′-end was used as GC-rich intron-1 to construct the pMH3 expression vector (Figure 2). The PGK promoter driven *neo* resistance gene was employed as a selection marker for stable mammalian and chicken gene transfection. G418 was used for selecting stable clones. Detailed structure of pMH3 expression vector is shown in Figure 2.

The 0.733 DNA fragment was assembled by ligation of five different GC-rich fragments synthesized by Genescript Inc. (www.genescript.com). The sequences of these five GC-rich fragments were derived from a GC-rich DNA fragment of a chicken genome (part of chicken cardiac alpha actin intron-1) and a GC-rich DNA fragment of a rice genome. This GC-rich DNA was used to replace 1.006 kb chick beta-actin GC-rich intron-1 and construct the "sequence" non-specific GC-rich expression vector.

### 1.3 Transfections, cell lines screening and evaluations

Stable gene transfections were conducted by a gene pulser (Bio-rad). The high expressors were selected by a dot blot (Figure 3) as well as ELISA methods for the 1st and 2nd or 3rd cloning in order to get pure cell population. The high expressors were adapted to suspension serum-free medium B001 (www.amprotein.com) and scaled up by using non-optimized fed-batch culture in AmProtein's 50-liter "Current" bioreactors. Purification used affinity protein-A method (Pharmacia), followed by an ion exchange method.

## 2 Results

Gene expression in prokaryotic cells such as *E-coli* is frequently strong and normally requires 3-hour or shorter expression periods for protein detection. Mammalian gene expression is always weak and difficult to detect even by sensitive ELISA or the less sensitive, semi-quantitative dot blot detection method. Thus, a 24–48 hour expression period is often required for mammalian protein detection.

To examine our hypothesis, we inserted this 1.006 kb intron-1 (average 75.3% GC content) from the chick *beta-actin* gene (a gift from Dr. N Fregien, University of Miami) (ATCC 37507) at the 5′ and 3′ flanking regions of more than 20 protein or antibody genes (Figure 4). We achieved extremely high 3-hour stable gene expression in one chicken and four mammalian cell lines by using just a

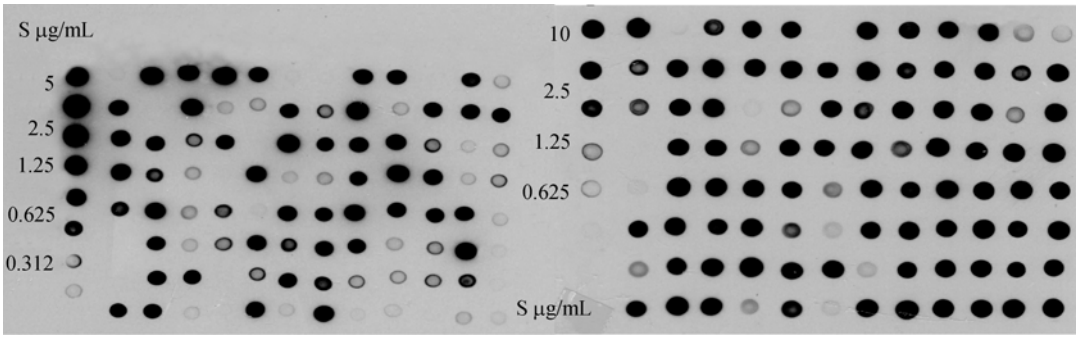


**Figure 3** Dot blot detection of protein production in a 3-hour period for a human antibody and VEGF trap (an IgG1 Fc-fused protein). The furthest left lane is standard human anti-VEGF antibody (Avastin) or TNFR2-Fc (Enbrel) 10 mg/L (10 μg/mL) series dilution.

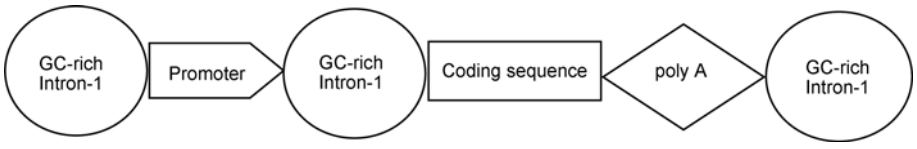


**Figure 4** Insert position of GC-rich chick *beta-actin* intron-1.

**Table 2** GC-rich intron-1 and protein expression

| Cell types | Protein types | Protein level of the best clone in a 3-hour expression period detected by dot blot (mg/L) | Protein level in a 3-hour expression period of the best clone detected by ELISA (mg/L) | Estimated production speed of the best clone (pg $cell^{-1}$ $d^{-1}$) | Purification yield of the best clone (mg/L) |
|---|---|---|---|---|---|
| CHO-S (Invitrogen) | Enbrel (TNFR2-Fc) | | 7 | 179 | |
| CHO-S (Invitrogen) | CD52 antibody | | 10 | 118 | |
| CHO-S (Invitrogen) | HBsAg | | 0.7 | 108 | |
| CHO-S (Invitrogene) | Erbitux (anti-EGF antibody) | | 15 | 60 | 450 |
| CHO-S (Invitrogen) | IL-1ra-IgG1Fc-IL-18bp | | | | 550 |
| CHO-S (ATCC) | IgG1 Fc-IL-1ra | | 4.42 | 150 | |
| CHO-S (Invitrogen) | GLP1-IgG1Fc-leptin | | | | |
| CHO-S (Invitrogen) | IgG1 Fc-Trail | | 11.3 | 226.3 | |
| CHO-S (Invitrogen) | PYY-IgG1Fc-leptin | | 6.8 | 172 | |
| CHO-S (Invitrogen) | Symlin-IgG1Fc-leptin | | 14 | 180 | |
| CHO-S (Invitrogen) | GLP1-IgG1Fc | | 2.37 | 33.5 | |
| CHO-S (Invitrogen) | HSA | | 2.03 | 270 | |
| CHO-S (Invitrogen) | EPO-CTP(hyperglycosylated EPO) | | 14 | 300 | 100 |
| CHO-S (Invitrogen) | Anthrax antibody (murine) | 12—15 | | 70—90 | |
| CHOK1 (ATCC) | Enbrel (TNFR2-Fc) | 8—10 | 5 | 80—90 | |
| CHOK1 (ATCC) | A human antibody (confidential) | 3—4 | 2.67 | 60—80 | |
| CHOK1 (ATCC) | CD52 antibody (Campath) | 5—6 | 12 | 32—48 | |
| CHOK1 (ATCC) | Her2 antibody | 9—10 | | 70—90 | 450 |
| CHOK1 (ATCC) | IgE antibody | 8—10 | | 55—70 | |
| CHOK1 (ATCC) | VEGF trap (IgG1 Fc fused protein) | 8—10 | | 60—80 | |
| CHOK1 (ATCC) | HSA-EPO | | 1.7 | 139 | |
| CHO-S (Invitrogen) | FSH | | 2.2 | 60 | |
| CHOK1 (ATCC) | IgG1 Fc-Endostatin | 2.5—5 | | 30—40 | |
| CHOK1 (ATCC) | CTP-Endostatin （hyperglycosylated） | 2—3 | | 15—20 | |
| CHOK1 (ATCC) | TNFR1-Fc-IL-18bp | 2—3 | | 10—15 | |
| CHOK1 (ATCC) | IL-1ra-IgG1Fc-IL-18bp | 5—8 | | 40—60 | |
| CHOK1 (ATCC) | IL-4R-IgG1CH2-IL-1ra | 2—2.5 | | 15—20 | |
| R1610 (ATCC) | Enbrel (TNFR2-Fc) | | 8 | 147 | |
| R1610 (ATCC) | IgG1 Fc-Trail | | 6 | 113 | |
| R1610 (ATCC) | GLP1-IgG2 Fc | 5—7 | | 30—40 | |
| DF-1 (chick) | Enbrel (TNFR2-Fc) | 8—10 | | 35—40 | |

few laboratory scale gene transfections for each protein (Table 2). By contrast, no protein expression (TNFR2-Fc, commercial name Enbrel) was detected in 3 and 6-hour production periods without the GC-rich intron-1 at 5′ or/and 3′ flanking regions of the genes of interest under the native chick *beta-actin* gene promoter or CMV promoter sequence alone [10].

We also visualized expressed proteins from a hyperglycosylated EPO production clone by direct protein staining of the culture supernatant after a 24-hour culture period. Amazingly, we clearly saw the protein band (Figure 5).

The results in Table 2 clearly show that inserting the GC-rich intron-1 at 5′ and 3′ flanking regions of the gene of interest greatly enhanced expression in different cell lines for different proteins. The expression has reached the levels above the current industrial average levels (20–60 pg cell$^{-1}$ d$^{-1}$) [11] that were achieved using the labor-intensive gene copy amplification method employed elsewhere.

To further prove that these high-expressing cells from Table 2 were truly producing recombinant proteins at the industrial level, we have selected some stable clones for non-optimized scale up cultures using the 5-, 50- and 300-liter working volume bioreactors [12]. The protein purification yields are demonstrated in Table. 2.

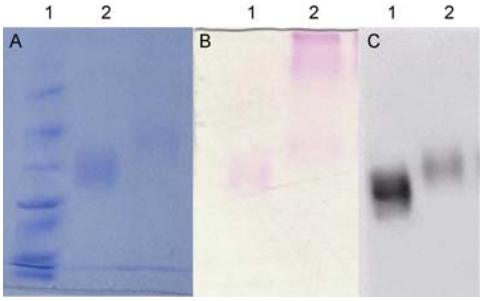


**Figure 5** Visualization of 24-hour expressed proteins in 96-well plates by the direct culture supernatant. A, Commassie blue staining; B, sugar staining; C, Western blot. 1, 5 μg purified EPO; 2, 20 μL EPO-HyperG enhanced generic supernatant.

The above results indicated that a simple industrial "GC-rich" method for mammalian protein and antibody expression has been successfully established at this point.

The generated cell clones from Table 2 have the potential to be optimized into gram/liter production yields although the production speed (pg cell$^{-1}$ d$^{-1}$) defined in Table 2 was measured in 96-well plates (0.1 mL/well) after a 3 or 24-hour culture period. However, scale up protein purification yields in Table. 3 did not reach gram/liter. It indicates that medium optimization and scale up optimization are further required for commercial scale protein production.

In order to determine if the GC-rich DNA fragment was the key for the extremely high mammalian gene expression shown in Tables 2 and 3, we synthesized a 0.733 kb "sequence" non-specific GC-rich DNA fragment (average of 71.9% GC content shown in Table 4) with no GGCGGG and CCCGCC boxes (Figure 6) to replace the 1.006 kb GC-rich chick *beta-actin* intron-1 (Figure 1). This 0.733 kb "sequence" non-specific GC-rich DNA (Figure 6) was derived respectively from GC-rich DNA fragments of a chicken genome and a rice genome. This result in Table 5 and Figure 7, taken together with other results shown in Tables 2 and 3 provided the first experimental evidence indicating that a non-coding GC-rich DNA fragment is a super "chromatin opening element" and plays a dominant role in the regulation of mammalian gene expression.

We have also inserted the GC-rich DNA at either 3′ or 5′ flanking regions of genes of interest alone. Our results in Figure 7 indicated similar expression, but with fewer high expressors, suggesting the requirement of more cell transfections and clone screening. The constructs illustrated in Figs. 2 and 6 had the best results.

## 3 Discussion

In order to examine our hypothesis, we successfully em-

**Table 3** Production rate (pg cell$^{-1}$ d$^{-1}$) of different cell clones (CHOK1 E1 and CHO-S A3) from AmProtein-China and NCPC in different media and production periods (ELISA)

| Clone name | Enbrel E1 (AmProtein-China) | Enbrel A3 (NCPC) | Enbrel E1 (AmProtein-China) | Enbrel A3 (NCPC) | Enbrel E1 (AmProtein-China) | Enbrel A3 (NCPC) |
|---|---|---|---|---|---|---|
| Medium | DMEM/F12 | DMEM/F12 | B001 (rich medium) | B001 (rich medium) | DMEM/F12/M199 | DMEM/F12/M199 |
| 3-hour production period (pg cell$^{-1}$ d$^{-1}$) | 12.7 | 9.5 | 44.9 | 29.2 | 12.0 | 9.2 |
| 24-hour production period (pg cell$^{-1}$ d$^{-1}$) | 70.1 | 55.2 | 165.0 | 94.5 | 57.9 | 38.8 |
| Cell density | 0.65×10$^6$/well | 0.65×10$^6$/well | 0.65×10$^6$/well | 0.65×10$^6$/well | 0.65×10$^6$/well | 0.65×10$^6$/well |

**Table 4** Analysis of GC content of the synthetic 0.733 kb DNA fragment

| Positions | 1–147 bp | 148–294 bp | 295–442 bp | 443–590 bp | 591–733 bp |
|---|---|---|---|---|---|
| GC content | 74.8% | 78.1% | 78.9% | 70.7% | 58.1% |

TTACTAGACAGGTGAGTGTCGCCGCGTCGT
CGCCGCGTCCCCCGCGCACGCGTCGTCGTC
GTCGCCGCGGCCCCGCGCCGTAGCCGCCG
CCGCCGGCCTCCCCCCTCCCTCCTCCGGCC
AGATCTGGGTCTAAGTATGCCCGGGGCAAG
GGAGGGGGGAGGGGAGCCGCCGCCGCGG
CCCCCGCGCGCGCGACGTCGTCGTCACCGC
CGCCCCGCGCCGTAGCCGCTGCCGCCGGC
CTTCCCCCTCCCTCCTCCAGCCAGATCTGGG
AGGGAGGGGGTAAGAATGCCGCGGGAGGG
AGGGGAGCCGCCGCCGCGCCCCCCGCGCG
TCGTCGTCGCCGTGGCCCTGCGCCGTCGTC
ATCGTCGCCACCGCCGCCGGCGGCCTCAGC
TTGTGCGGTGTTACCGGCGTCCGCCTGCTC
GCCCTTTGTGCGGCTCTAGAGCATCTGCCG
CGAGCACGGGACGGGACACCGCCCTCTGCC
CTTCTCGCCCTCGGAGCGGCTGCGGCTCGA
CGGGGCGCGGTGCGCCGGAGGCGGCGTGC
GCTGGCAGAGCCGGACTCTGCGATACCATT
TCGCTTAAGCTCGAGACTCATCCATTTTGATC
TGTGCCCATAGGGTACCGCTAGTGCCCAAA
GTTAGCGTTCTGGGAGCCTTCTGAATCGTGA
ACGGCAGCGGCGACGGGTCCCTCAGAGCC
GTGCCCGAGCGCCGCCGCCTTCTTCTTTTTC
CTACAGCTCAAGCTTG

**Figure 6** A synthetic 0.733 kb "sequence non-specific" GC-rich DNA fragment. There are no GGCGGG and CCCGCC boxes [9]. There are 119 CG sites in the above sequence.

**Table 5** Application of the synthetic 0.733 kb DNA fragment and corresponding protein expression[a)]

| Cell types | Protein types | a | b | c |
|---|---|---|---|---|
| CHO-S (Invitrogene) | Enbrel (TNFR2-Fc) | >10 | 14.0 | 75 |
| CHOK1 (ATCC) | Enbrel (TNFR2-Fc) | >10 | 14.5 | 95 |

a) a, the detected protein level in the 3-hour production period of the best clone (mg/L); b, the detected protein level in the 3-hour expression period of the best clone detected by ELISA (mg/L); c, the estimated production speed of the best clone (pg cell$^{-1}$ d$^{-1}$).

ployed 3-hour expression periods for our protein detection (Table 2). To further prove that these high-expressing cells from Table 2 were truly producing recombinant proteins at the industrial level, we have selected some stable clones for non-optimized scale up cultures (Table 3). Thus, a simple industrial "GC-rich" method for mammalian protein and antibody expression has been successfully established.

Based on the facts that GC linkages have three hydrogen bonds and that a GC-made DNA double helix is different from an AT-made one, it is likely that DNA GC content

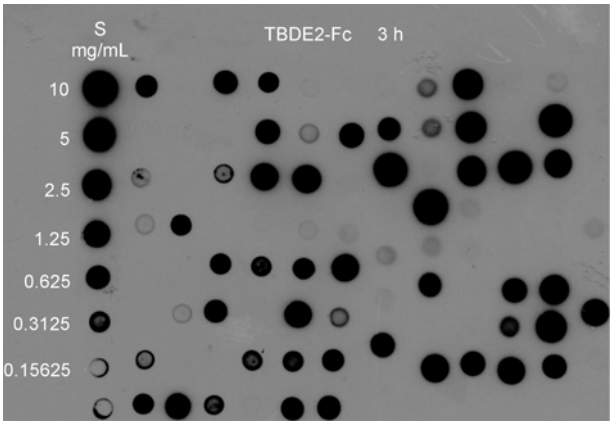


**Figure 7** Dot blot detection of protein production. A 0.733 kb GC-rich DNA was inserted in the 3′-flanking region of genes of interest (Promoter-TNFR2-Fc-Poly A-0.733 kb GC-rich fragment) alone. Dot blot detection of protein production in the 3 hour period for TNFR2-Fc (Enbrel). The furthest left lane is standard Enbrel (Amgen) 10 mg/L (10 ug/mL) series dilution.

makes a difference in forming DNA secondary structure that further affects the chromatin openness.

Giorgio Bernardi and Vinogradov reported that GC-rich genes were associated with the expression of certain genes, such as house-keeping genes [4,5]. Meissner *et al*. [13] recently reported that a core region of unmethylated CpGs in high-CpG-density promoters [14,15] was associated with ubiquitous "house-keeping genes" and highly-regulated key developmental genes where the transcriptional initiation mark H3K4me3 (univalent) is generally highly expressed. Meissner's study also identified 25051 sites of the enriched transcriptional initiation mark H3K4me3 (univalent) at locations from 1 kb to >100 kb away from known promoters. CpGs sampled at these sites (outside promoters and CpG islands) had significantly lower methylation levels. This relationship was particularly strong for CpGs located in highly conserved non-coding elements. These reports indicated a link between "sequence" non-specific GC content, including CpG islands, and DNA methylation that might further affect gene expression.

Together with our experimental evidences reported here, an immediate study to link the genomic non-coding GC content with adjacent gene expression activities was indicated. These results also suggested studies to understand the role of a non-coding GC-rich DNA fragment in the regulation of gene expression in other eukaryotic cells such as plant cells, which may have a significant impact on molecular farming.

## 4 Conclusions

We concluded that a simple industrial "GC-rich" method for mammalian protein and antibody expression has been successfully established. We also conclude that flanking GC-rich non-coding DNA plays the key role in the regula-

tion of gene expression in mammalian and chicken cells. Importantly, our discovery indicates that the chromatin-based regulation of gene expression is at least partially embedded in DNA primary structure, namely DNA GC-content.


1 Cyranoski D. Giant bird-like dinosaur found: Chinese researchers unearth a surpising find. Nature. Published online 13 June 2007.
2 Fucheng Z H, Zhonhe Z H, Xing X, *et al*. A bizarre Jurrassic maniraptoran from China with elongate ribbon-like feathers. Nature, 2008, 455: 1105–1108
3 Kost T A, Theodorakis N, Hughes S H. The nucleotide sequence of the chick cytoplasmic beta-actin gene. Nucleic Acids Res, 1983, 11: 8287–8301
4 Bernardi G. Isochores and the evolutionary genomics of vertebrates. Gene, 2000, 241: 3–17
5 Vinogradov A E. DNA helix: the importance of being GC-rich. Nucleic Acids Res, 2003, 31: 1838–1844
6 Cao H, Widlund H R, Simonsson T, *et al*. TGGA repeats impair nucleosome formation. J Mol Biol, 1998, 281: 253–260
7 Lowary P T & Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol, 1998, 276: 19–42
8 Levitsky V G. RECON: A program for prediction of nucleosome formation potential. Nucleic Acids Res, 2004, 32: 346–349
9 Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol, 1987, 196: 261–282
10 Fregien N, Davidson N. Activating elements in the promoter region of the chicken beta-actin gene. Gene, 1986, 48: 1–11
11 Anderson D C, Krummen L. Recombinant protein expression for therapeutic applications. Curr Opin Biotechnol. 13: 117–123
12 Jia Q, Li H C, Hui M, *et al*. A Bioreactor system based on a novel oxygen transfer method. Bioprocess Int, 2008, 6: 66–78
13 Alexander M, Tarjei S M, Hongcang G, *et al*. Lander Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature, 2008, 454: 766–770
14 Saxonov S, Berg P, Brutlag D L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci USA, 2006, 103: 1412–1417
15 Takai D, Jones P A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci USA, 2002, 99: 3740–3745