OnLine Journal of Biological Sciences

Original Research Paper

Sequence Analysis of Scaffold/Matrix Attachment Regions (S/MARs) From Human Embryonic Kidney and Chinese Hamster Ovary Cells

Nur Shazwani Mohd Pilus, Azrin Ahmad and Nurul Yuziana Mohd Yusof

Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Article history
Received: 21-08-2018
Revised: 05-09-2018
Accepted: 25-09-2018

Corresponding Author:
Nur Shazwani Mohd Pilus
Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia
Email: shazwani.pilus@gmail.com

Abstract: Binding of intergenic Scaffold/Matrix Attachment Regions (S/MARs) to nuclear matrix proteins is believed to poise adjacent genes for transcription by forming chromatin loops. Vector constructs containing Scaffold/Matrix Attachment Regions (S/MAR) flanking the gene of interest, therefore, are able to enhance recombinant protein expression in mammalian cells. We compared two methods that are based on buffers containing 2M NaCl and Lithium-3,5-diidosalicylate (LIS) to isolate S/MARs from HEK293 and CHO DG44 cell lines. Isolated S/MARs were sequenced using the Illumina HiSeq platform and mapped against CHO DG44 genome contigs and the human genome GRCh37.p13 respectively (Sequence raw data from this article have been deposited at the EMBL Data Libraries under Study ID PRJEB26090 (ERP108063)). The 2M NaCl method produced 16 million S/MAR consensus sequences which included nine million and seven million from HEK293 and CHO DG44 respectively. LIS method, on the other hand, generated thirteen million S/MAR consensus containing 8.4 million and 4.7 million from HEK293 and CHO DG44, respectively. In order to compare all sets of S/MAR consensus, BLASTN analyses were performed based on exact matches. The number of perfect matches between S/MAR sequences produced by both methods was quite low: 0.46% and 0.07% for HEK293 and CHO DG44 cells respectively, indicating that the two methods isolate different sets of S/MARs. Comparison between the two cell lines found six S/MARs in common, with average coverage of 82%, obtained by the 2M NaCl method, but none of these are intergenic. The LIS method gave 38 S/MARs with average coverage of 85%, common to both cell types; of these, 13 were intergenic. We hypothesize that S/MARs from HEK293 and CHO DG44 isolated using the LIS method have the potential to be universal vector expression elements that can overcome the problem of low production yield.

Keywords: 2M NaCl, LIS, Intergenic, Gene Transcription, Biopharmaceutical, Expression Vector

Introduction

In the era of modern medicine, recombinant protein therapeutics are contributing significantly to innovative and effective therapies for treatment of numerous human diseases (Agarwal *et al*., 1998). Therapeutic proteins such as antibodies and enzymes, produced in mammalian cells, have been successfully utilized in the treatment of diseases in the past decade. Gene transfer technology in mammalian cells - particularly in cells engineered for production of proteins - requires sustainable and high level expression. However, the positional effect of transgene integration sites might hinder the effectiveness of the recombinant protein expression in the transfected cells. This factor could be due to the influences of the chromatin structure effects and/or dominant regulatory elements flanking the integration sides of the gene (Feng *et al*., 2001). One of the strategies to overcome this effect is

by adapting scaffold or matrix attachment regions, S/MARs, in the expression vectors (Allen *et al*., 2005; Argyros *et al*., 2011).

The terms scaffold and matrix referring to the same biological entities, which are proteins structuring the nucleus, but these two were differentiated by two different isolation methods (Bode and Maass, 1988; Donev, 2000). The Matrix Attachment Region (MAR), was introduced by Berezney and Coffey (1974), after their discovery of fibrous protein structures in the nucleus known as Matrix Proteins (MPs). Both MARs and MPs were attained by isolating nuclei using a buffer of NaCl, detergent and enzymes; the NaCl helps to disrupt histone/DNA interactions by competing for binding sites on the DNA (Earnshaw and Laemmli, 1983), but some have argued that this creates artifacts due to precipitation under high salt conditions (Berezney and Coffey, 1974). Later, Mirkovitch *et al*. (1984) introduced an isolation method that reduced the artifacts using a low concentration of Lithium-3, 5-diiodosalicylate (LIS) in place of the high salt (Mirkovitch *et al*., 1984); the LIS acts as an anionic salt that lowers the ionic strength and reduces the flexibility of DNA which, along with charge repulsions, displaces it from the histones (Marky and Manning, 1991). Hence, the DNA fractions attached to a scaffold protein isolated using LIS method were referred as Scaffold Attachment Region (SAR). As S/MARs bind to nuclear proteins, they are associated with important biological roles particularly in genome organization (Berezney *et al*., 1995; Bode *et al*., 2006; Manuelidis, 1990), gene transcription stabilization (Cockerill and Garrard, 1986) and assisting genome replication (Bode *et al*., 1996).

S/MARs have been implicated in the regulation of gene expression due to their co-localization with the transcription units and regulatory elements in genomes (Bode *et al*., 2000). S/MARs are believed to regulate gene expression by initiating interactions between DNA activating complexes and genes and also by controlling chromatin accessibility (Heng *et al*., 2004). They act by forming loops that poise specific regions of the genome for transcription (Bode *et al*., 1996; Jackson, 1997; Razin, 2001). As S/MARs could be directly involved in the regulation of the gene expression at the chromatin structure level, it is believed that the use of these elements in expression vectors might aid high level production of protein in host cells (Girod *et al*., 2005). However, it is crucial to investigate the function of S/MARs, which have the potential to either up- or down regulate gene expression and such information on the relationship between S/MARs and gene regulation is still in deliberation.

Based on the 2009 chromosome-level study by Linnemann *et al*. on HeLa cells' chromosome 16, SARs located at 5’ of a gene are associated with the expressed transcripts while MARs positioned within a gene are related to gene silencing (Linnemann *et al*., 2009). These varied functions of S/MARs were discovered based on comparison of two different extraction methods. LIS extraction will disrupt binding mediated through transcription complexes to yield nuclear scaffold (Bode *et al*., 1996), whereas 2M NaCl extraction is suggested to isolate a nuclear matrix that is interwoven with newly synthesized RNA (Ma *et al*., 1999). Integrated information based on analysis of the DNA regions from these two methods and gene expression profiling demonstrated that SARs at 5’ of genes are related to highly expressed transcripts and genes attached to the intergenic MARs are silent (Linnemann *et al*., 2009). Thus, it is feasible that S/MARs could enhance the expression of a gene which they flank. However, not many studies have been carried out to assess regions of matrix association throughout the genome.

At the beginning era of high-throughput sequencing technology using combination of DNA library construction and Sanger’s sequencing method, a genomic array-based analysis using large insert library clones from a human genomic library was performed to identify S/MARs extracted by LIS method. A total of 2.5 Mbp S/MARs were mapped to a human neocentromer imparted the centromer’s function in nuclear organization during mitosis and meiosis (Sumer *et al*., 2003). As the sequencing technology advanced to next generation approaches, an improved method of identifying S/MARs from *Drosophila melanogaster* was accomplished. A total of 7353 S/MARs were isolated using LIS method and were sequenced by SOLiD platform (LifeTech, USA). Through intensive genome wide analysis, these S/MARs were found to represent 2.6% of the genome and were recognized as DNA elements associated with transcription sites of highly expressed genes (Pathak *et al*., 2014).

While the information for a genome-wide study has not been established for mammalian cells, our study aims to identify S/MAR sequences based on both LIS (Keaton *et al*., 2011) and NaCl (Krawetz *et al*., 2005) extraction methods for two different mammalian cell lines, CHO DG44 and HEK 293, at genome level using the Solexa sequencing platform (Illumina Incorporation, USA). Sequences of the isolated S/MARs were generated and mapped to respective genome data. Clustering analysis between the two datasets of S/MAR sequences from the two mammalian cell lines was performed to narrow down the S/MAR dataset based on sequence similarities. These shortlisted S/MAR sequences were identified for their location in the genome, either intergenic or intragenic. Hopefully, such information could provide a better understanding of S/MARs, to enable a strategy for genetic intervention to produce a better host cell line, better downstream culture environments or a better expression vector. Such improvements may lead to higher yields and this greater affordability, of therapeutic proteins.

Materials and Methods

Cell Culture

CHO DG44 and HEK293 cell lines were obtained courtesy of Inno Biologics Sdn. Bhd. and cell preparation subjected to human is conformed to the principles outlined in the Declaration of Helsinki. CHO cells were cultured in HyClone^{TM} SFM4CHO^{TM} (Thermo Scientific, USA) while HEK293 cells were cultured in 293 SFM II (Invitrogen, USA). Both cells were cultured in spinner flask with agitation at 45 rpm until they reached log phase. An amount of 7×10^6 cells/ml were harvested for S/MAR isolation using 2M NaCl and another 1×10^6 cells/ml for isolation using lithium-3,5-diiodosalicylate (LIS). Medium was removed and cell pellet were washed using 1X PBS buffer pH7.4 supplemented with protease inhibitor (Roche, USA). One tablet of protease inhibitor was added into each 10 mL of PBS buffer.

2M NaCl Isolation Method

Halo in Gel

S/MAR isolation using 2M NaCl was done in two parts. The first part known as halo in gel is required to determine the minimum time taken for cell nuclei to form the largest halo size within incubation period between 1 and 10 min, at one-minute interval. Nuclear halo is characterized as an overlapping chromatin strand anchored to matrix protein by means of S/MAR after depletion of histones (Krawetz *et al*., 2005).

A total of 11 slides containing a layer of 0.5% (w/v) low-melting agarose gel mixed with approximately 6×10^4 cells per slide were prepared to test the incubation time with halo buffer. Encapsulated cells were treated with nuclei buffer for 1 hour on ice to isolate nucleus. Cell nuclei were washed using PBS buffer pH 7.4 supplemented with protease inhibitor for 1 min. Each of 10 slides was dedicated for incubation in halo buffer containing 2M NaCl for every 1 min starting from 1 to 10 min. One slide is reserved for negative control. To stop the reaction of halo buffer, slides were dipped in 1X PBS buffer pH 7.4 supplemented with protease inhibitor for 1 min. To fix the nuclei on gel, a cold absolute ethanol were applied and slides were dried at 55°C for 30 min. Halo image were visualized using fluorescent microscope after staining with 100 µg/ml ethidium bromide. Six halos were randomly captured for each incubation time to get an average halo size. The size is obtained after subtracting the outmost area with the inner area using ImageJ V1.50i software (Fig. 1). A time point with largest area difference was the most convenient incubation time to induce halo structure for a particular cell type.

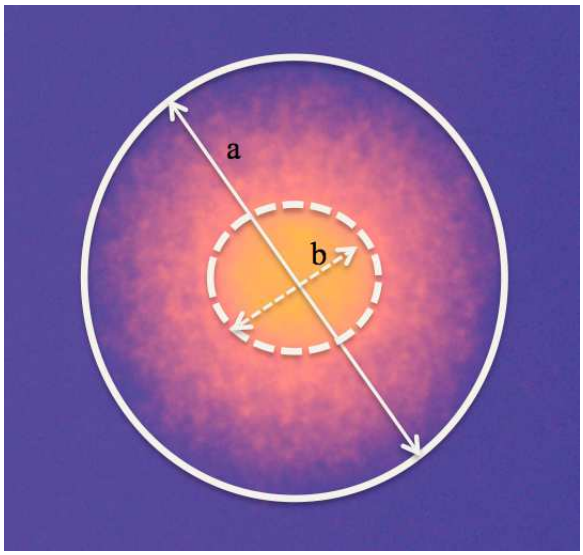

Fig. 1: Halo size is determined by subtracting the outer area (a) with inner area (b)

Halo in Solution

The procedure of nuclei isolation is repeated in this second part of isolation but extraction is done in solution. The time obtained from halo in gel method is applied for incubation in buffer containing 2M NaCl to induce nuclei halo formation. An amount of 7×10^6 cells/mL was harvested and medium was removed by centrifugation at 65 x g for CHO DG44 and 200 x g for HEK293 cells for 7 min at 4°C. Pellet was resuspend in 2 mL 1X PBS buffer pH 7.4 supplemented with 1 mg/ml Bovine Serum Albumin (BSA) (Amresco, USA) and protease inhibitor and centrifuged at respective speeds for 7 min at 4°C. Pellet was resuspend in 2 mL nuclear buffer (10 mM Tris-HCl pH7.7, 100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 0.5% (v/v) Triton-X 100, protease inhibitor) for 1 hour on ice to isolate nucleus. Cell nuclei were collected by centrifugation at respective speeds for 7 min at 4°C. Pelleted nuclei were washed with 2 mL 1X PBS pH 7.4 supplemented with protease inhibitor and centrifuged at respective speeds for 7 min at 4°C. Pellet was resuspend with 2 mL halo buffer (10 mM Tris-HCl pH7.7, 10 mM EDTA, 2 M NaCl, 1 mM DTT) and incubation was done on ice with the duration of 8 min for CHO DG44 and 7 min for HEK293 as per determined from halo in gel procedure. A total of 40 mL restriction enzyme buffer (50 mM Tris-HCl pH8.0, 10 mM $MgCl_2$) were added to the nuclei solution and centrifuged at 200 x g for both cell lines for 7 min at 4°C. An amount of 1 mL supernatant was reserved in tube for digestion with 100 U *Eco*RI and 100 U *Bam*HI. Incubation was done at 37°C for 4 h with agitation at 110 rpm. To separate S/MAR from genomic DNA, nucleus was centrifuged at 16,000 x g for 5 min at 4°C. Supernatant was labeled as loop fraction. A volume of 300 µL proteinase K

buffer (50 mM Tris-HCl pH8.0, 50 mM NaCl, 25 mM EDTA, 0.5% (v/v) SDS, 120 µg proteinase K enzyme) were added to each pellet and supernatant fractions. S/MAR were recovered after overnight incubation with proteinase K enzyme at room temperature.

LIS Isolation Method

A total of 7 x 10^6 cells/ml CHO DG44 and HEK293 were harvested and medium were removed by centrifugation at 65 x g for CHO DG44 and 200 x g for HEK293 for 5 min at 4°C. Pellet was washed with 2 mL PBS buffer supplemented with 0.1 mM PMSF. Centrifugation was done as previously mentioned speed for 5 min. Cell pellet were resolved with 2 mL lysis buffer (50 mM KCl, 0.5 mM EDTA, 0.05 mM spermine, 0.125 mM spermidine, 1 mM DTT, 0.1% (w/v) digitonin, 0.5 mM Tris-HCl, 0.1 mM PMSF) and 1.25 volume of stabilization buffer (50 mM KCl, 0.625 mM Cu_2SO_4, 0.05 mM spermine, 0.125 mM spermidine, 1 mM DTT, 0.1% (w/v) digitonin, 0.5 mM Tris-HCl, 0.1 mM PMSF) prior to incubation on ice for 20 min. After 20 min, 10 ml LIS buffer (10 mM LIS, 100 mM $C_2H_3LiO_2$, 0.05 mM spermine, 0.125 mM spermidine, 1 mM DTT, 0.05% (w/v) digitonin, 20 mM HEPES-KOH pH7.4) was added and mixture was left to stand at room temperature for 10 min. To separate nuclei, mixture was centrifuged at 2620 x g for 35 min. Supernatant was carefully removed and 2 mL of matrix washing buffer (20 mM KCl, 70 mM NaCl, 10 mM $MgCl_2$, 20 mM Tris-HCl pH7.4) were added resolved pellet. Mixture was centrifuged at 2620 x g for 35 min. Pellet was washed twice with restriction buffer (50 mM NaCl, 10 mM $MgCl_2$, 100 mM Tris-HCl 7.4) and centrifuged at 2620 x g for 35 min at each wash. To separate S/MAR from genomic DNA, 1 mL of restriction buffer (50 mM NaCl, 10 mM $MgCl_2$, 100 mM Tris-HCl pH7.4, 0.025% (v/v) Triton X-100) was added to nuclei pellet. A concentration of 100 U E*co*RI and 100 U *Bam*HI were added to the solubilized pellet and incubated at 37°C for 1.5 h with 110 rpm agitation. Mixture was centrifuged at 2620 x g for 10 min and supernatant were saved as loop fraction. Another 1 ml of restriction buffer with *Eco*R1 and *Bam*HI were added to nuclei pellet and incubation was continued for another 1 hour. At minutes 45, 20 µg/ml RNase A was added and incubation continued until minutes 60. Mixture was centrifuged at 2620 x g for 10 min and supernatant was mixed with 300 mM NaCl and 27 mM EDTA to preserved the DNA. To digest the bounded protein on S/MAR, pellet was solubilized with 1 mL of K1 buffer (300 mM NaCl, 2.5 mM EDTA, 10 mM Tris-HCl pH8.0) followed by 2 mL of proteinase K buffer (1% Nlaurylsarcosine, 450 mM NaCl, 45 mM EDTA, 60 mM Tris-HCl pH8.0, 120 µg/ml Proteinase K enzyme) and incubated overnight at room temperature. S/MAR in pellet and loop fractions collected from 2M NaCl and LIS methods were purified using phenol:chloroform:isoamylalcohol (25:24:1) (Chomczynski and Sacchi, 1987).

Quantity and Quality Analysis of S/MAR

Quantity was measured using spectrophotometer NanoDrop ND (Thermo Fisher Scientific, USA) for S/MAR extracted using both methods. Purified S/MARs extracted using 2M NaCl were analyzed using Bioanalyzer (Agilent Technologies, USA) while purified S/MARs extracted using LIS were analyzed using 1% electrophoresis agarose gel.

Sample Preparation for NGS Sequencing

Sample preparation for sequencing using NGS HiSeq 2000 platform was performed according to Nextera XT DNA kit manual (Illumina Incorporation, USA). A total of 1 ng of S/MAR sample was used as starting material for paired-end sequencing. Sequencing was outsourced to Malaysia Genome Institute and completed after 2 weeks.

Data Analysis

Trimming Sequencing Reads

Post NGS sequencing data was trimmed using SolexaQA software package to eliminate any low quality reads. A cut off value of Qphred 20 was set to obtain at least 99% sequence target using DynamicTrim and any reads with length lower than 50 bp were removed using LengthSort. Every sequencing pair was determined and any unpaired reads were kept separately as singletons. Both paired reads and singletons of HEK293 S/MAR were mapped against human genome GRCh37.p13 (www.gencodes.org/releases/19.html) using CLC Genomic Workbench 7.0 to generate consensus sequences. Meanwhile, CHO DG44 S/MAR was mapped against CHO DG44 contigs since the genome is currently developing.

S/MAR Matched Sequence Search Against Loop Fractions

Both isolation methods have produced loop fractions resulting from the restriction enzyme degradation that separated them from matrix or scaffold fractions. All four loop-fractions were sequenced together with respective S/MAR fractions. The loop consensus was then BLAST with respective S/MAR fractions using BLASTN 2.2.28 program. E value cut off was set to zero to limit search for only exact sequence hit.

S/MAR Matched Sequence Search Across Two Methods

S/MAR consensus sequence of HEK293 and CHO DG44 isolated using 2M NaCl were BLAST against S/MAR isolated using LIS to search for any matched

sequence obtained by both methods. Sequence comparison was done according to cell line using the same program and parameter settings.

S/MAR Matched Sequence Search Across Two Cell Lines

S/MAR isolated from HEK293 were BLAST using the same program and parameter settings against S/MAR isolated from CHO DG44 to search for any shared sequences across these two cell lines.

Mapping of HEK293-CHO DG44 S/MAR against Annotated Human Genome

Matched sequence from BLAST result across two cell lines were mapped against annotated human genome Patch 13 (NCBI) using CLC Genomic Workbench 6.0.2 to locate the position of shared HEK293 and CHO DG44 S/MAR. To get the detailed identity of mapped S/MAR, the consensus sequences were BLAST using nonredundant (nr) database with zero E value cut off.

S/MAR Matched Sequence Search against Annotated Protein

BLASTX analysis among S/MAR consensus were performed against annotated protein database from CHO K1 (www.chogenome.org) since our CHO DG44 database is still under construction. Meanwhile, S/MAR consensus from HEK293 were analyzed based on human genome database GRCh37.p13 (www.gencodes.org/releases.19.html). Both BLAST analysis was set to E-value cut off at 10^{-10} to produce more stringent result using BLASTX 2.2.30+.

Results

Isolation of S/MARs

Potential S/MARs isolated using both the 2M NaCl and LIS methods were quantified using a Nanodrop spectrophotometer. Quality assessment was determined by 1%(w/v) agarose gel electrophoresis for DNA samples extracted using LIS method. Due to the low yield, DNA samples obtained from the 2M NaCl method were analysed on a Bioanalyzer (Agilent, USA) (refer Supplementary Materials). The percentages of DNA recovered in the S/MAR (attached) and loop (nonattached) fractions were determined in order to evaluate the distribution of both fractions after being isolated by these two methods (Table 1). For the 2M NaCl method, the fraction of DNA in the S/MAR fraction (26% and 37% for HEK and CHO cells, respectively) was similar to a previous study which found about 30% to 40% of DNA is recovered in this fraction using 2M NaCl (Boulikas, 1995). LIS extraction has lower percentage of S/MAR, 3.2% and 2.7% for HEK293 and CHO DG44, respectively. This result might due to cleavage by restriction enzymes being more efficient in the LIS method compared to NaCl method. This, in turn, may be because of DNA structural changes in high salt, affecting the site-recognition of *Eco*RI and *Bam*HI (Travers, 1993).

Next Generation Sequencing of S/MARs

This study is the first report of sequencing of the S/MAR fragments from both LIS and NaCl isolation methods from mammalian cell lines using the Solexa platform (Illumina Incorporation, USA). The libraries were prepared using the Nextera XT DNA kit (Illumina Incorporation, USA) and the sequencing was performed as paired-end, which is an advantage for alignment accuracy (Quinlan *et al*., 2010). From the total of all S/MARs' reads, at least 79% of them were considered as high quality reads (data not shown). The percentage of S/MAR reads mapped against the respective human genome and CHO DG44 genome contigs (Ahmad, 2016) using CLC Genomic Workbench software analysis ranges between 91.8% and 99.5% (Table 2).

Sequence Analysis of S/MAR Data

The loop DNA fractions were sequenced together with respective S/MAR fractions to examine the effectiveness of both methods in capturing S/MARs that were interacting with matrix protein hence, being least contaminated by loop DNA. Although S/MARs are present throughout the genome, not every S/MAR will interact with matrix protein all of the time to form loop: Interactions depend on cell cycle stage and cell type at the time S/MAR isolation was performed (Barboro *et al*., 2012; Boulikas, 1995). BLASTN analyses were performed to compare sequences between S/MAR fractions and loop DNA fractions. For NaCl-isolated S/MAR from HEK293, BLASTN analysis showed a low percentage (0.13%; 22,839/17,215,861) of matched sequences while for CHO DG44 0.32% (45,828/14,157,742) were matched. For LIS-isolated material, matches between S/MAR and loop fractions were even lower; that is 0.07% (10,083/14,960,547) from HEK cells and 0.13% (14,837/11,607,655) from CHO (Fig. 2). These results suggest that both methods achieve a very clean partitioning between S/MAR and loop DNA.

It has been claimed that some S/MARs may involved in either gene silencing or gene activating, depending on the isolation method used (Donev, 2000; Linnemann *et al*., 2009). The location of S/MAR in the genome, whether they flanked a gene or located in between genes, influenced its property and it also has close relationship with the isolation method used (Dijkwel and Hamlin, 1988). Thus, in this study, BLASTN analysis was performed between 2M NaCl and

LIS S/MAR consensus sequences, to see whether the same sequences were isolated by both methods. There were only 0.46% (80,807/17,407,528) and 0.07% (8340/11,877,261) sequence matches between the NaCl and LIS datasets for HEK293 and CHO DG44 cells, respectively (Fig. 3). This indicates that the two methods isolate different and largely non-overlapping, populations of sequence.

We also used BLASTN analysis to look for S/MARs, which were common to both cell types, using each of the two methods. There are only six and 38 consensus sequences in common between the two cell types, the NaCl and LIS isolation methods, respectively (Fig. 4). This is not unexpected, given that the two cell lines are from different species (Chinese hamster and human) and there are also different cell types with different expression profiles. However, the 44 S/MARs shared between the two cell lines could become a potential element in expression vectors to be applied across different types of mammalian cell line (Table 3 to 5).

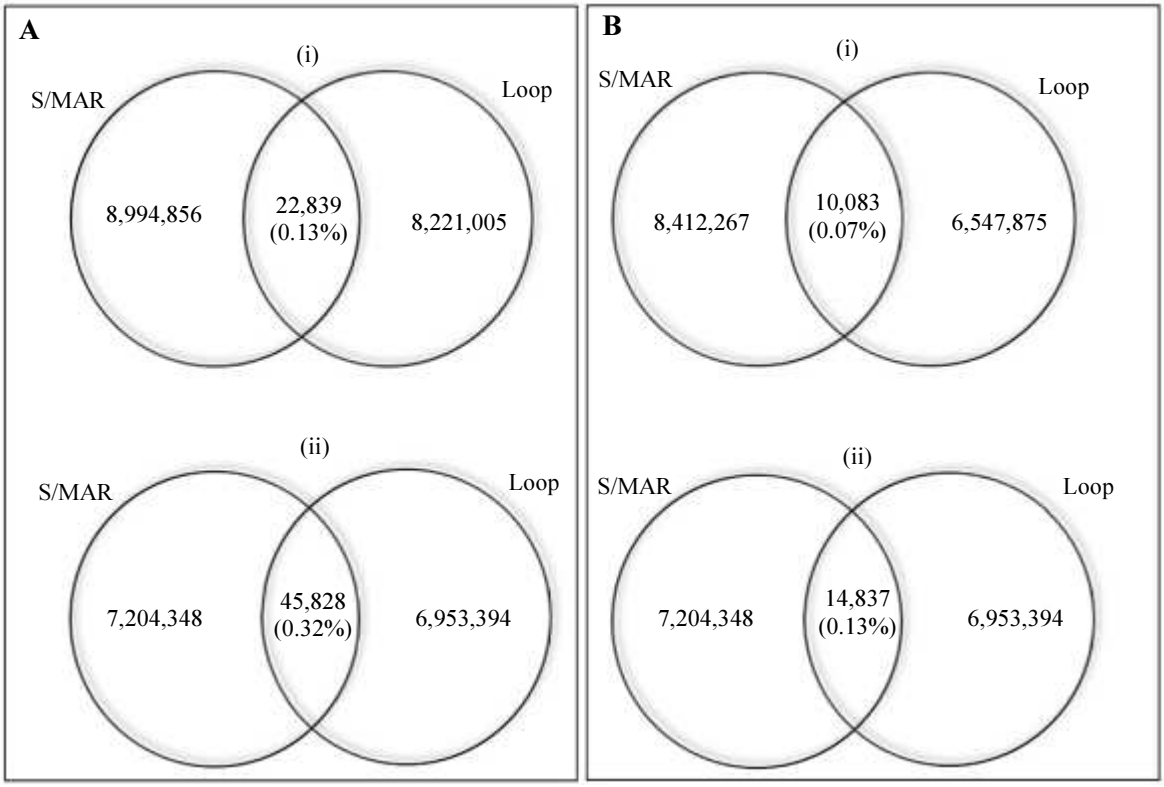

Fig. 2: BLAST result for matched sequence search between S/MAR fractions and loop fractions. A(i) is HEK293, A(ii) is CHO DG44 both from NaCl method and B(i) is HEK293, B(ii) is CHO DG44 both from LIS method

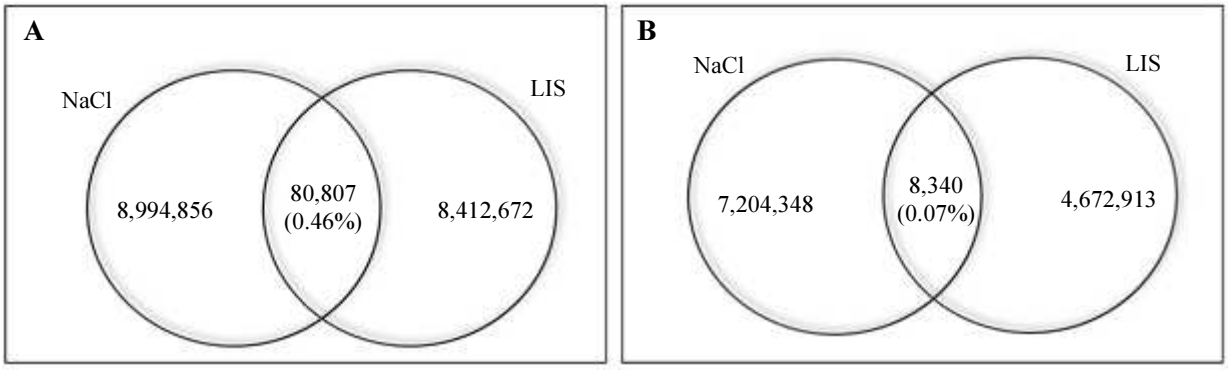

Fig. 3: BLAST result for matched sequence search between two different methods. (A) S/MAR isolated from HEK293 and (B) S/MAR isolated from CHO DG44

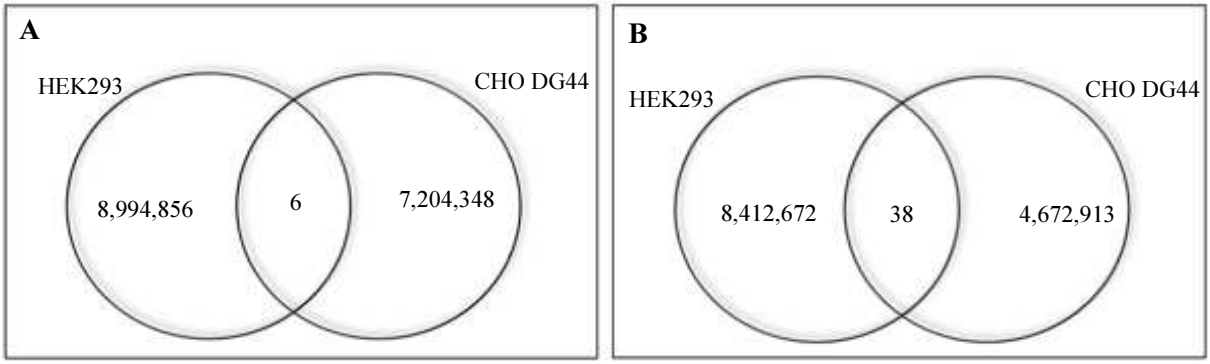

Fig. 4: BLAST result for sequence similarity search between two different cell lines. (A) NaCl-isolated S/MAR and (B) LISisolated S/MAR

Table 1: Quantitative analysis of HEK293 and CHO DG44 S/MAR using spectrophotometer

Isolation method	Cell line	Fractions	Quantity (µg)	Percentage of DNA fraction (%)[a]	$A_{260/280}$[b]	$A_{260/230}$[c]
NaCl	HEK293	S/MAR	0.98	26.1	1.91	2.17
		Loop DNA	2.76	73.9	1.89	2.10
	CHO DG44	S/MAR	1.31	37.1	1.88	2.09
		Loop DNA	2.22	62.9	2.05	2.32
LIS	HEK293	S/MAR	1.50	3.2	1.95	2.16
		Loop DNA	45.73	96.8	1.82	2.36
	CHO DG44	S/MAR	0.66	2.7	1.67	1.78
		Loop DNA	23.70	97.3	1.88	2.38

[a]Percentage of DNA fraction was obtained based on below formula: Quantity of S/MAR x 100% (quantity of S/MAR fraction + quantity of loop DNA fraction)
[b]Ratio of light absorption at 260 nm wavelength to light absorption at 280 nm wavelength
[c]Ratio of light absorption at 260 nm wavelength to light absorption at 230 nm wavelength

Table 2: Number of S/MAR consensus sequences generated after mapping

Isolation method	Cell line	Mapped reads	Percentage of mapped reads (%)	Total consensus sequences
NaCl	HEK293	25,091,274	99.5	8,994,856
	CHO DG44	24,100,613	91.8	7,204,348
LIS	HEK293	74,264,342	92.7	8,412,672
	CHO DG44	32,214,332	97.6	4,672,913

Table 3: Genes associated with NaCl-isolated S/MAR (CHO DG44-HEK293)

Genes associated with S/MAR	S/MAR ID	Protein coded	Protein function	Source
ARHGAP5	S/MAR_HC_N_TRA01	*Homo sapiens* Rho GTPase 5 activated protein	GTPase-activating protein for Rho family members	www.uniprot.org
MATR3	S/MAR_HC_N_TRA02	Matrin-3	Play a role in transcription or interact with nuclear matrix protein to form internal fibrogranular network	www.uniprot.org
EHBP1	S/MAR_HC_N_TRA03	EH domain-binding protein 1	Play a role in actin reorganization	www.uniprot.org
SOX6	S/MAR_HC_N_TRA04	Transcription factor SOX-6	Transcriptional activator and play a role in several developmental process	www.uniprot.org
ENAH	S/MAR_HC_N_TRA05	Protein enabled homolog	Induces the formation of actin rich outgrowths in fibroblast	www.uniprot.org
ZFP62	S/MAR_HC_N_TRA06	Zinc finger protein 62 homolog	Play a role in differentiating skeletal muscle	www.uniprot.org

Table 4: Genes associated with CHO DG44-HEK293 S/MAR isolated using LIS that located at intergene

Flanking genes	S/MAR ID	Protein coded	Protein function	Source	Position
MEIS2	S/MAR_HC_L_TER01_a	Homeobox protein Meis2	Interacting with DNA sequence that is in cis with and relatively close to a core promoter for RNA polymerase II	Yang *et al*. (2000)	60 kb from 5' end of MEIS2 and 150 kb from 3' end of RPS15P8
RPS15P8	S/MAR_HC_L_TER01_b	Transmembrane and coiled-coil domain containing protein 5A			
ZFHX4-AS1	S/MAR_HC_L_TER02	Antisense RNA1	Non-protein coding and belongs to class of antisense RNA	NCBI	16 kb from ZFHX4-AS1
SUPTH16HP	S/MAR_HC_L_TER03_a	Ty16 homolog *S. Cerevisiae* pseudogene	Interact specifically with histone H2A/H2B to effect nucleosome disassembly to facilitate transcription elongation	Orphanides *et al*. (1999)	50 kb from 3' end of SUPTH16HP and 50 kb from 3' end of GOT2P4
GOT2P4	S/MAR_HC_L_TER03_b	Glutamicoxaloacetic transaminase 2 pseudogene 4	Pseudogene		
Putative gene (LOC100505985)	S/MAR_HC_L_TER04	-	Belongs to class of lncRNA	www.genecards.com	5' end of LOC100505985 (putative gene)
MIR2113	S/MAR_HC_L_TER05_a	MicroRNA 2113	Short nucleotide (20-24 nt) that involved in Posttranscriptional regulation and gene expression in multicellular organism by affecting stability and translational mRNA	NCBI	206 kb from 3' end of LOC101927314 and 109 kb from 5' end of MIR2113
Putative gene (LOC101927314)	S/MAR_HC_L_TER05_b	-	-		
Putative gene (LOC100505498)	S/MAR_HC_L_TER06_a	-	-		36 kb from 3' end of RPL6P5 and 118 kb from 5' end of LOC100505498
RPL6P5	S/MAR_HC_L_TER06_b	Ribosomal protein L6 pseudogene 5	Pseudogene	NCBI	
Putative gene (LOC101926956)	S/MAR_HC_L_TER07_a	-	-		202 kb from 5' end of LOC101926956 and 323 kb from 5' end of RPS26P10
RPS26P10	S/MAR_HC_L_TER07_b	Ribosomal protein S26 pseudogene 10	Specifically binds to domain C of Taxresponsive enhancer element in long terminal repeat of human T-cell leukemia virus type I (HTLV-I)	Morita *et al*. (1993)	
POLD2PI	S/MAR_HC_L_TER08_a	Exon and gene polymerase (DNA directed) delta 2 accessory subunit pseudogene 2	Pseudogene	NCBI	289 kb from 5' end of POLD2PI and 87 kb from 5' end of CCT7P2
CCT7P2	S/MAR_HC_L_TER08_b	Chaperonin containing TCP1 subunit pseudogene 2	Pseudogene	NCBI	
Putative gene (LOC100419816)	S/MAR_HC_L_TER09_a	Orthodontic homeobox 2 pseudogene	Influence proliferation and differentiation of dopaminergic neuronal progenitor cell during mitosis	NCBI	161 kb from 5' end of LOC100419816 and 121 kb from 3' end of RPL17P12
RPL17P12	S/MAR_HC_L_TER09_b	Ribosomal protein pseudogene 12	Pseudogene		
RAB5CP2	S/MAR_HC_L_TER10_a	RAS oncogene family pseudogene 2	Pseudogene		150 kb from 3' end of RAB5CP2 and 235 bp from 3' end of PCBP2P3
PCBP2P3	S/MAR_HC_L_TER10_a	Poly(rC) binding protein2 pseudogene 3	Pseudogene		
Putative gene (LOC100419816)	S/MAR_HC_L_TER11_a	Orthodonticle homeobox 2 Pseudogene	Influence proliferation and differentiation of dopaminergic neuronal progenitor cell during mitosis	NCBI	101 kb from 5' end of LOC100419816 and 61 kb from S/MAR_HC_L_TER 09_a kb
HIG1	S/MAR_HC_L_TER12_a	Hypoxia inducible domain family, member 1A	Involve in the assembly of respiratory supercomplex	NCBI	38 kb from 5' end of HIGI
Putative gene (LOC100190924)	S/MAR_HC_L_TER13_a	Eukaryotic translation initiation factor 4E binding protein pseudogene	Pseudogene	NCBI	175 kb form 5' end of LOC100190924 and 379 kb from 5' end of POU3F2
POU3F2	S/MAR_HC_L_TER13_b	Pou class 3 of neural transcription factor	Involved in neuronal differentiation	www.uniprot.org	

Table 5: Genes associated with CHO DG44-HEK293 S/MAR isolated using LIS that located at intragene

Genes associated with S/MAR	S/MAR ID	Protein coded	Protein function	Source
RBM12B2	S/MAR_HC_L_TRA01	RNA binding protein 12B-B	A protein that exhibits poly(A) RNA binding (ortholog)	
KDM6A	S/MAR_HC_L_TRA02	Lysine(K)-specific demethylase 6A	Act as histone demethylase that play a role in histone code	www.uniprot.org
SYNCRIP	S/MAR_HC_L_TRA03	Synaptotagmin binding cytoplasmic RNA interaction protein	Play a role in mRNA maturation	NCBI
CPSF2	S/MAR_HC_L_TRA04	*Homo sapiens* cleavage and polyadenylation specific factor 6	Involve in pre-mRNA 3'- end formation	www.uniprot.org
ZCCHC7	S/MAR_HC_L_TRA05	Zinc finger-CCHC domain containing protein 7	Involve in gene transcription, translation, mRNA trafficking, cytoskeleton organization, epithelial development, cell adhesion, protein folding, chromatin remodeling, zinc sensing	www.ebi.ac.uk (Interpro)
RSBN1	S/MAR_HC_L_TRA06	Round spermatid basic protein	Play important role in transcriptional regulation in haploid germ cells	www.ebi.ac.uk
NRXN1	S/MAR_HC_L_TRA07	Neuroxin 1	Involved in cell-cell interactions, exocytosis of secretory granules and regulation of signal transmission	www.uniprot.org
RANBP17	S/MAR_HC_L_TRA08	RAN binding protein	Transport of protein and large RNAs Through nuclear pore complex	NCBI
MIPOL1	S/MAR_HC_L_TRA09	Mirror-image polydactyly gene 1	May function as tumor suppressor	NCBI
FAM172A	S/MAR_HC_L_TRA10	Family with sequence similarity 172, member A	(Function not characterized yet)	-
MATR3	S/MAR_HC_L_TRA11	Matrin 3	May play a role in transcription or interact with nuclear matrix to form internal fibrogranular network	www.uniprot.org
POLA1	S/MAR_HC_L_TRA12	DNA polymerase alpha catalytic subunit isoform	Initiation in DNA replication	www.uniprot.org
LRBA	S/MAR_HC_L_TRA13	LPS-responsive vesicle trafficking, beige and achor containing protein	Maybe involve in intracellular vesicles to activated receptor complex, which aids in secretion and membrane deposition of immune effector molecules	NCBI
RASAL2	S/MAR_HC_L_TRA14	*Homo sapiens* RAS protein activator like 2	Activator of Ras superfamily of small GTPase	www.genecarads.org
EHBP1	S/MAR_HC_L_TRA15	EH-binding protein 1, isoform X4	(As described in S/MAR_HC_N_TRA03)	
SRSF10	S/MAR_HC_L_TRA16	*Homo sapiens* serine/arginine-rich splicing factor 10	Splicing factor or repressor of pre-mRNA splicing	www.uniprot.org
RANBP17	S/MAR_HC_L_TRA17	*Homo sapiens* RAN binding protein 17	(As described in S/MAR_HC_L_TRA08)	
ANKHD1-EIF4EBP3	S/MAR_HC_L_TR A18	*Homo sapiens* ANKHD1-EIF4EBP3 readthrough	Naturally occurring readthrough transcript of neighbouring ANKHD1 and EIF4EBP3 genes	
SKAP2	S/MAR_HC_L_TRA19	Src kinase-associated phosphoprotein 2	Substrate of Src family kinase and involved in Src signaling pathway and regulate activity of immune system	NCBI
LRBA	S/MAR_HC_L_TRA20	-	As described in S/MAR_HC_L_TRA13)	
ASCC3	S/MAR_HC_L_TRA21	Activating signal cointegrator1 complex subunit 3	Member of helicase and involved in repair of alkylated DNA	NCBI
TEX41	S/MAR_HC_L_TRA22	A non-coding protein known as testis expressed 41	A RNA gene affiliated with long non-coding RNA class	www.uniprot.org
DACH1	S/MAR_HC_L_TRA23	Dachshund homolog 1 isoform a	Transcription factor involved in regulation of organogenesis	www.uniprot.org
EHBP1	S/MAR_HC_L_TRA24	EH domain binding protein 1	(As described in S/MAR_HC_N_TRA03)	NCBI
UCKL1	S/MAR_HC_L_TRA25	*Homo sapiens* uridinecytidine kinase 1-like 1	Catalyze phosphorylation of urudine to uridine monophosphate as part of ribonucleotide salvage pathway	

To characterize these 44 HEK/CHO shared S/MARs in respect of their position in the genome, the consensus sequences were mapped against the human reference genome using CLC Genomic Workbench software. All six of the shared S/MARs isolated using the NaCl method are intragenic (Table 3), while 13 out of 38 HEK293-CHO DG44 S/MAR isolated using LIS method were at intergenic positions or located in between genes (Table 4) and the rest are intragenic (Table 5). As reported by several studies, intergenic S/MAR are usually involved in gene activation, particularly for those positioned at the upstream of a gene (Agarwal *et al*., 1998). Overall, most of the 44 shared S/MAR sequences have sequence similarities with protein binding, RNA binding protein, transcription factor, DNA polymerase, matrin and microRNA, in which putatively, they may involved in gene transcription. For example, three of the genes encodes for EH domain binding protein and another two encodes for RAN binding protein. Two genes that are related to RNA binding protein are SYNCRIP gene (synaptotagmin binding cytoplasmic RNA interacting protein) and RBM12B2. Two of the

LIS-isolated S/MARs are located intergenically with poly(RC) binding protein pseudogene and initiation factor 4E binding protein. Other than protein binding property, three of the S/MAR HEK293-CHO DG44 has matched sequence with genes coding for transcription factor such as SOX6, DACH1 and POU3F2. ZFP62 and CCHC are two genes that codes for zinc finger protein. Two of the shared S/MARs have sequence matching the gene for matrin, one of the major components in nuclear matrix protein that play a role in transcription or binding of S/MAR to nuclear matrix (Lewis and Laemmli, 1982). Two shared S/MARs that are associated with initiation of replication because one of them has matches the DNA Polymerase (POLA1) gene and another is located adjacent to polymerase delta 2 gene.

We further analyzed all consensus S/MARs (from both cell types and from both methods) by performing BLASTX analysis to determine whether our S/MARs are protein-coding genes based on their sequences. S/MARs from HEK293 cells were BLASTXed against an annotated human genome database (GRCh37.p13); those from CHO DG44 cells were BLASTXed against the CHO K1 genome, since the CHO DG44 database is still under development. HEK293 S/MARs by the NaCl method produced 6.7% (624,001/9,291,331) sequences that fall in coding regions, while for CHO DG44 NaCl S/MARs the figure was only 1.2% (88,859/7,204,348). For LIS-generated S/MARs, the corresponding figures were 9.6% (836,758/8,736,261) and 0.9% (44,049/4,672,913), for HEK and CHO cells, respectively.

Discussion

Since 1974, scientists have been trying to isolate interacting complexes between DNA and nuclear proteins, but it has been shown that the choice of isolation method used greatly affects the protein composition of the recovered (matrix or scaffold) material (Earnshaw and Laemmli, 1983). Both LIS and NaCl are the main component to dissociate histone from holding the densely packed chromatin causing the chromatin to loose up while leaving a halo structure poised by S/MAR that interacting with protein matrix. The principle behind 2M NaCl or high salt isolation method is to alter ratio of anion to cation concentration between DNA and histone. The increased amount of positive ions by NaCl has created a competitive binding with negatively charged H1 histone against binding site on the chromation causing the H1 histone to dissociate from chromatin (Guo and Cole, 1989). LIS, on the other hand, acts as lithium acid salt that created a low ionic strength to cell environment (Gavin *et al*., 1998). Low ionic strength ambient has influenced DNA structure to become stiff because the presence of negative ions stimulate repulsive force among molecules of phosphate group on DNA chain until the chain stretched up thus changed the structure causing histone to dissociate from chromatin (Marky and Manning, 1991).

Linnemann *et al*. (2009) compared the two NaCl and LIS methods for S/MAR isolation to study any differences of the isolated S/MAR on the aspect of function and their role in changing genome structure associated with gene expression. The study reported that NaCl isolated-S/MARs were likely to be apart from the genes condensed regions, which most of them were located at the telomeric regions, whereas LIS isolated-S/MARs are mostly at the 5' end of active genes. However, the S/MAR distribution in the study only focused on five chromosomes of HeLa S3 cell line.

To explore the distribution and sequence features of S/MARs within a genome, several studies have been conducted involving different techniques such as Southern blotting, MAR-PCR array and *in silico* prediction by computational software (Dijkwel and Hamlin, 1988; Rudd *et al*., 2004; Tachiki *et al*., 2009). A total of 7,535 S/MAR sequences have been generated using SOLID sequencing and, of these, 95% contain the ORI sequence motif and 3% are located within 100 bp downstream of a transcription initiation site (Pathak *et al*., 2014). The experimental design is almost the same as our study except that the S/MARs were obtained from *Drosophila melanogaster* embryonic cells through a modified method using combination of DNase I, detergent and salt to extract the nuclear matrix prior to high salt treatment to isolate S/MARs.

In our study, we used the two original methods for S/MAR isolation, the LIS and the NaCl methods. In order to provide information on S/MAR sequences that are interacting with matrix protein, we have run BLASTN analysis of S/MAR and loop DNA fractions. For both HEK293 and CHO DG44 extracted using both methods, we found a very low proportion (0.07-0.32%) of sequences co-present in both S/MAR (attached) and loop (non-attached) fractions. These co-present sequences are probably due to differences in cell cycle stages when the isolation procedure was done and reflect the dynamic nature of matrix attachments (Barboro *et al*., 2012; Berezney *et al*., 1995).

The ability of S/MARs to increase transgene expression makes them potentially useful to the biotechnology industry, particularly in biopharmaceutical applications. Thus, the choice of HEK293 and CHO DG44 cell lines used is based on their application as "workhorses" for mammalian-based biofactory production of vaccines and therapeutic proteins (Jayapal *et al*., 2007). Interestingly, though, S/MAR characteristics are conserved across species (Bode *et al*., 2006). For example, S/MARs from human showed the same insulating effect on transgene expression in other organisms such as *Drosophila melanogaster*. Other research has shown that a κ intronic

S/MAR can be replaced by another S/MAR from genomic location yet still show the same methylation pattern and normal gene expression (Namciu *et al*., 1998). These findings suggest that the same S/MARs may be applicable across multiple cell types, species and genes. Our study, found a total of 17.4 and 11.9 million S/MARs from HEK293 and CHO DG44, cells respectively, but, we have focused on the 44 sequences in common between the two cell types. The presence of these sequences across two different cell types from two species suggests that they might be usefully incorporated in expression vectors in a variety of mammalian cell systems.

S/MARs are believed to act by controlling transcription of the gene (or transgene) regardless of its position in the host genome (Poljak *et al*., 1994). A strong interaction between S/MARs matrix proteins results in the formation of a chromatin loop, which isolates the gene flanked by the S/MARs from adjacent silencing regions (Wang *et al*., 2010). The S/MAR-matrix binding site becomes the assembly site for transcription machinery including transcription factors and DNA polymerase (Heng *et al*., 2004; Ottaviani *et al*., 2008). S/MARs isolated using LIS usually occur either upstream or downstream of a gene and are involved in active gene expression. Conversely, S/MARs isolated using NaCl tend to lie in gene-poor regions and are usually associated with gene silencing (Linnemann *et al*., 2009). In order to investigate the function and potential utility of the 44 S/MARS shared between HEK CHO cells, we mapped them against the human genome to determine their positions relative to nearby genes. All six shared sequences isolated using the NaCl method are intragenic, which corresponds with the findings of with Agarwal *et al*. (1998). This result shows that the way S/MARs bind to matrix protein are not certain to specific sequences but rely mostly on cell type and cell cycle stage (Boulikas, 1995). Even if S/MARs are located away from the flanked gene, S/MARs are able to bind to matrix protein and could regulate gene expression in sequential manner (Forrester *et al*., 1994).

Hence, in this study, we have narrowed our focus to the intergenic S/MARs. Of the 13 intergenic S/MARs isolated using LIS method, we have identified 22 neighbouring genes that reside within 0 to 382 kb of the S/MAR. Four of the intergenic S/MARs are not shown in pair of their neighbouring genes because they are located too far from S/MARs (above 400 kb away). From the 22 neighbouring genes, seven of them lie within 100 kb of the S/MAR; of these seven, two are involved in transcription, three are pseudogenes, one encodes and antisense RNA and one is immediately adjacent to a putative non-coding RNA gene. A further eight of the 22 neighboring genes are lie between 100 and 200 kb from the S/MAR; six of these are pseudogenes, one encodes a transmembrane and coiled-coil domain protein and one encodes a microRNA (Table 4). Lastly, there are seven neighbouring genes that located between 200 and 300 kb from S/MAR; five of them are pseudogenes, one is involved in respiratory supercomplex assembly and another is a transcription factor involved in neruronal differentiation. Those S/MARs, which are adjacent to genes involved in transcription, might be of interest as gene expression regulators; however, further experiments are needed to test this.

In agreement with the previously reported characteristics of S/MARs (Girod *et al*., 2005), we found very few that lay within coding sequences (through the BLASTX analysis). This agrees with the supposed role of S/MARs in creating gene-containing chromatin domains to either facilitate or repress transcription (Namciu *et al*., 1998; Ma *et al.*, 1999).

Future studies will benefit from the completion of the CHO DG44 genome database, against which we will be able to map the S/MARs isolated in this study and we hope to create a S/MAR database once the DG44 database is available. We also hope to test the function of some of these S/MARs in promoting transgene transcription from vectors in mammalian cells.

Conclusion

S/MARs, as DNA elements that determine chromatin organization and regulate gene expression, have been exploited as vector expression elements that can stabilize expression in mammalian cell host systems. However, the S/MARs we have isolated in this study need further evaluation in this capacity. An interaction study between our S/MAR sequences and matrix proteins, especially in vivo, would help us to understand the function of individual S/MARs in cells and how this relates to cell cycle and gene expression. We plan to further investigate the behaviour of our S/MARs via molecular docking with matrix proteins as well as by performing biophysical and biochemical analysis of the cells during the interactions. Finally, our main target is to capture the interaction of S/MARs and matrix proteins and to perform the analysis through live cell imaging. In the meantime, the effect of incorporating the S/MARs we have identified into expression vectors is our main priority in order to confirm their capability in enhancing recombinant protein production by overcoming transgene silencing caused by positional effect.

Acknowledgement

We would like to thanks Ministry of Higher Education Malaysia (FRGS/2011/SG/UKM/03/7) for funding this research. Thanks to Dr. Heilly Chong and the staff of Protein Science Department, Inno Biologics Sdn. Bhd. for providing the cell lines. Thanks to the School of Bioscience and Biotechnology, Faculty of Science and Technology,

Universiti Kebangsaan Malaysia and Malaysia Genome Institute for materials, facilities and technical supports. Thank you to Mr. Mohammad Faizal for assisting in bioinformatics analysis, Dr. Norazfa for technical assistance in halo in gel method and also to Dr. Paul H. Dear for proofreading this paper.

Funding Information

This manuscript is funded by Ministry of Higher Education Malaysia (FRGS/2011/SG/UKM/03/7).

Author's Contributions

Nur Shazwani Mohd Pilus: Drafted and wrote the manuscript, performed the experiment and result analysis.

Azrin Ahmad: Assisted in bioinformatic analysis on the CHO genome annotation.

Nurul Yuziana Mohd Yusof: Initiated the idea of S/MAR study, supervised the experiment's progress, result interpretation and helped in manuscript preparation.

Ethics

There are no ethical issues after the publication of this manuscript.

References

Agarwal, M., T.W. Austin, F. Morel, J. Chen, E. Böhnlein and I. Plavec, 1998. Scaffold attachment region-mediated enhancement of retroviral vector expression in primary T cells. J. Virol., 72: 3720-3728. PMID: 9557654

Ahmad, A., 2016. Genomic and transcriptomic analysis of CHO DG44 cell lines using next generation sequencing approach. MSc Thesis, School of Bioscience and Biotechnology, Universiti Kebangsaan Malaysia.

Allen, G.C., S. Spiker and W.F. Thompson, 2005. Transgene Integration: Use of Matrix Attachment Regions. In: Transgenic Plants: Methods and Protocols, Peña, L. (Ed.), Springer, Totowa, ISBN-10: 1592598277, pp: 313-326.

Argyros, O., S.P. Wong, C. Fedonidis, O. Tolmachov and S.N. Waddington *et al*., 2011. Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver. J. Molecular Med., 89: 515-529. DOI: 10.1007/s00109-010-0713-3

Barboro, P., E. Repaci, C. D'arrigo and C. Balbi, 2012. The role of nuclear matrix proteins binding to matrix attachment regions (Mars) in prostate cancer cell differentiation. PLoS One, 7: e40617-e40617. DOI: 10.1371/journal.pone.0040617

Berezney, R. and D.S. Coffey, 1974. Identification of a nuclear protein matrix. Biochem. Biophys. Res. Commun., 620: 1410-1417. DOI: 10.1016/0006-291X(74)90355-6

Berezney, R., M.J. Mortillaro, H. Ma, X. Wei and J. Samarabandu, 1995. The nuclear matrix: A structural milieu for genomic function. Int. Rev. Cytol., 162A: 1-65. PMID: 8575878

Bode, J., C. Benham, E. Ernst, A. Knopp and R. Marschalek *et al*., 2000. Fatal connections: When DNA ends meet on the nuclear matrix. J. Cell Biochem. Suppl. Suppl., 35: 3-22. PMID: 11389527

Bode, J. and K. Maass, 1988. Chromatin domain surrounding the human interferon-. beta. Gene as defined by scaffold-attached regions. Biochemistry, 27: 4706-4711. DOI: 10.1021/bi00413a019

Bode, J., T. Schlake, M. Rios-Ramirez, C. Mielke and M. Stengert *et al*., 1996. Scaffold/matrix-attached regions: Structural properties creating transcriptionally active loci. Int. Rev. Cytol., 162: 389-454. DOI: 10.1016/S0074-7696(08)61235-8

Bode, J., S. Winkelmann, S. Götze, S. Spiker and K. Tsutsui *et al*., 2006. Correlations between Scaffold/Matrix Attachment Region (S/MAR) binding activity and DNA duplex destabilization energy. J. Molecular Biol., 358: 597-613. DOI: 10.1016/j.jmb.2005.11.073

Boulikas, T., 1995. Chromatin domains and prediction of MAR sequences. Int. Rev. Cytol., 162A: 279-388. DOI: 10.1016/S0074-7696(08)61234-6

Chomczynski, P. and N. Sacchi, 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochem., 162: 156-159. DOI: 10.1016/0003-2697(87)90021-2

Cockerill, P.N. and W.T. Garrard, 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell, 44: 273-282. DOI: 10.1016/0092-8674(86)90761-0

Dijkwel, P.A. and J.L. Hamlin, 1988. Matrix attachment regions are positioned near replication initiation sites, genes and an interamplicon junction in the amplified dihydrofolate reductase domain of Chinese hamster ovary cells. Mol. Cell Biol., 8: 5398-5409. DOI: 10.1128/MCB.8.12.5398

Donev, R.M., 2000. The type of DNA attachment sites recovered from nuclear matrix depends on isolation procedure used. Mol. Cell Biochem., 214: 103-110. DOI: 10.1023/A:1007159421204

Earnshaw, W.C. and U.K. Laemmli, 1983. Architecture of metaphase chromosomes and chromosome scaffolds. J. Cell Biol., 96: 84-93. DOI: 10.1083/jcb.96.1.84

Feng, Y.Q., M.C. Lorincz, S. Fiering, J.M. Greally and E.E. Bouhassira, 2001. Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. Mol. Cellular Biol., 21: 298-309. DOI: 10.1128/MCB.21.1.298-309.2001

Forrester, W.C., C. Van Genderen, T. Jenuwein and R. Grosschedl, 1994. Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. Science, 265: 1221-1225. DOI: 10.1126/science.8066460

Gavin, I.M., S.I. Usachenko and S.G. Bavykin, 1998. Nucleosome structural transition during chromatin unfolding is caused by conformational changes in nucleosomal DNA. J. Biol. Chem., 273: 2429-2434. DOI: 10.1074/jbc.273.4.2429

Girod, P.A., M. Zahn-Zabal and N. Mermod, 2005. MAR elements as tools to increase protein production by CHO cells. In: Animal Cell Technology Meets Genomics, Gòdia, F. and M. Fussenegger (Eds.), Springer, Dordrecht, pp: 411-415.

Guo, X. and R.D. Cole, 1989. Chromatin aggregation changes substantially as pH varies within the physiological range. J. Biol. Chem., 264: 11653-11657.

Heng, H.H., S. Goetze, C.J. Ye, G. Liu and J.B. Stevens *et al*., 2004. Chromatin loops are selectively anchored using scaffold/matrixattachment regions. J. Cell Sci., 117: 999-1008. DOI: 10.1242/jcs.00976

Jackson, D.A., 1997. Chromatin domains and nuclear compartments: Establishing sites of gene expression in eukaryotic nuclei. Molecular Biol. Rep., 24: 209-220. DOI: 10.1023/A:1006873614521

Jayapal, K.P., K.F. Wlaschin, W. Hu and M.G. Yap, 2007. Recombinant protein therapeutics from CHO cells-20 years and counting. Chem. Eng. Progress, 103: 40-40.

Keaton, M.A., C.M. Taylor, R.M. Layer and A. Dutta, 2011. Nuclear scaffold attachment sites within ENCODE regions associate with actively transcribed genes. PLoS One, 6: e17912-e17912. DOI: 10.1371/journal.pone.0017912

Krawetz, S.A., S. Draghici, R. Goodrich, Z. Liu and G.C. Ostermeier, 2005. In silico and wet-bench identification of nuclear matrix attachment regions. Meth. Mol. Med., 108: 439-458. PMID: 16028699

Lewis, C.D. and U.K. Laemmli, 1982. Higher order metaphase chromosome structure: evidence for metalloprotein interactions. Cell, 29: 171-181. DOI: 10.1016/0092-8674(82)90101-5

Linnemann, A.K., A.E. Platts and S.A. Krawetz, 2009. Differential nuclear scaffold/matrix attachment marks expressed genes. Hum. Mol. Genet., 18: 645-654. DOI: 10.1093/hmg/ddn394

Ma, H., A.J. Siegel and R. Berezney, 1999. Association of chromosome territories with the nuclear matrix. Disruption of human chromosome territories correlates with the release of a subset of nuclear matrix proteins. J. Cell Biol., 146: 531-542. DOI: 10.1083/jcb.146.3.531

Manuelidis, L., 1990. A view of interphase chromosomes. Science, 250: 1533-1540. PMID: 2274784

Marky, N.L. and G.S. Manning, 1991. The elastic resilience of DNA can induce all-or-none structural transitions in the nucleosome core particle. Biopolymers, 31: 1543-1557. DOI: 10.1002/bip.360311310

Mirkovitch, J., M.E. Mirault and U.K. Laemmli, 1984. Organization of the higher-order chromatin loop: Specific DNA attachment sites on nuclear scaffold. Cell, 39: 223-232. DOI: 10.1016/0092-8674(84)90208-3

Morita, T., T. Sato, H. Nyunoya, A. Tsujimoto and J. Takahara *et al*., 1993. Isolation of a cDNA clone encoding DNAbinding protein (TAXREB107) that binds specifically to domain C of the tax-responsive enhancer element in the long terminal repeat of human T-cell leukemia virus type I. AIDS Res. Human Retroviruses, 9: 115-121. DOI: 10.1089/aid.1993.9.115

Namciu, S.J., K.B. Blochlinger and R.E. Fournier, 1998. Human matrix attachment regions insulate transgene expression from chromosomal position effects in Drosophila melanogaster. Mol. Cell. Biol., 18: 2382-2391. DOI: 10.1128/MCB.18.4.2382

Orphanides, G., W.H. Wu, W.S. Lane, M. Hampsey and D. Reinberg, 1999. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature, 400: 284-288. DOI: 10.1038/22350

Ottaviani, D., E. Lever, P. Takousis and D. Sheer, 2008. Anchoring the genome. Genome Biol., 9: 201-201. DOI: 10.1186/gb-2008-9-1-201

Pathak, R.U., A. Srinivasan and R.K. Mishra, 2014. Genome-wide mapping of matrix attachment regions in Drosophila melanogaster. BMC Genom., 15: 1022-1022. DOI: 10.1186/1471-2164-15-1022

Poljak, L., C. Seum, T. Mattioni and U.K. Laemmli, 1994. SARs stimulate but do not confer position independent gene expression. Nucleic Acids Res., 22: 4386-4394. DOI: 10.1093/nar/22.21.4386

Quinlan, A.R., R.A. Clark, S. Sokolova, M.L. Leibowitz and Y. Zhang *et al*., 2010. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res., 20: 623-635. DOI: 10.1101/gr.102970.109

Razin, S.V., 2001. The nuclear matrix and chromosomal DNA loops: Is their any correlation between partitioning of the genome into loops and functional domains? Cellular Mol. Biol. Lett., 6: 59-70. PMID: 11544631

Rudd, S., M. Frisch, K. Grote, B.C. Meyers and K. Mayer *et al*., 2004. Genome-wide in silico mapping of scaffold/matrix attachment regions in Arabidopsis suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. Plant Physiol., 135: 715-722. DOI: 10.1104/pp.103.037861

Sumer, H., J.M. Craig, M. Sibson and K.H.A. Choo, 2003. A rapid method of genomic array analysis of Scaffold/Matrix Attachment Regions (S/MARs) identifies a 2.5-Mb region of enhanced scaffold/matrix attachment at a human neocentromere. Genome Res., 13: 1737-1743. DOI: 10.1101/gr.1095903

Tachiki, K., Y. Kodama, H. Nakayama and A. Shinmyo, 2009. Determination of the *in vivo* distribution of nuclear matrix attachment regions using a polymerase chain reaction-based assay in Arabidopsis thaliana. J. Biosci. Bioeng., 108: 11-19. DOI: 10.1016/j.jbiosc.2009.02.010

Travers, A., 1993. DNA-Protein Interactions. 1st Edn., Springer, London, ISBN-10: 0412259907, pp: 180.

Wang, T.Y., J.H. Zhang, C.Q. Jing, X.J. Yang and J.T. Lin, 2010. Positional effects of the matrix attachment region on transgene expression in stably transfected CHO cells. Cell Biol. Int., 34: 141-145. DOI: 10.1042/CBI20090017

Yang, Y., C.K. Hwang, U.M. D'souza, S.H. Lee and E. Junn *et al*., 2000. Threeamino acid extension loop homeodomain proteins Meis2 and TGIF differentially regulate transcription. J. Biol. Chem., 275: 20734-20741. DOI: 10.1074/jbc.M908382199