Hes·so
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts
Western Switzerland

HES SO, Sion

Literature approach, analysis and comparisons of certain genetic elements that are common to recombinant mammalian expression vectors, including matrix attachment region as epigenetic regulator

X01, Master in life science, applied bioscience

Quentin Aeberli
01/02/2018

Table of content

Introduction

The aim of this master thesis introduced here is to develop and characterize one or several expression vectors to be used to create recombinant mammalian cells expressing stably a gene of interest. One of the key conditions for this work is that the genetic elements contained in the plasmid to be created must not infringe any patents or licences. The work will result in a vector that can be used also commercially like e.g. for industrial cell lines development through an HES-SO service.

The master thesis is organized in three phases. The first one, summarized in this document, contains a literature approach and the analysis and comparisons of certain genetic elements that are common to all mammalian expression vectors and will also appear on the vector to be created. The purpose of this first part is to analyse existing methods, vectors and particular genetic elements and to decide which ones to choose for our development. The second phase will deal with vector design *in silico* followed by the development of a strategy concerning the subcloning of genetic elements and their positioning within the vector and relative to a gene of interest. In the third part the above will be applied and the designed plasmid(s) will be transfected into a CHO (Chinese Hamster Ovary) host cell line and eventually also into the Human Embryonic Kidney (HEK293) host cell line. The gene of interest to be used will be the Green Fluorescent Protein that will allow for a rapid analysis of protein expression *in vivo* and determination of stability over several cell generations.

Literature approach

One challenging problem of stable recombinant protein expression is the genetic stability of the newly introduced gene of interest (GOI) in the mammalian cell genome. During cultivation of recombinant cell lines for many generations, several events like silencing or variable gene transcription can reduce or even stop expression of the GOI. This is due to the intrinsic dynamics of genomic DNA, turning transcriptionally active euchromatin into “silent” heterochromatin or silencing transgene by genetic rearrangements. In stable mammalian cell lines development this is usually addressed by using transposons or other epigenetic elements to keep entire expression cassettes active in terms of transcription. Some systems, such as the GS (glutamine synthetase) or DHFR (dihydrofolate reductase) system try to overcome silencing of the integrated GOI by applying a selective pressure that leads to the amplification of the GOI. However, these systems are long and expensive processes.
A particular epigenetic tool has been discovered in the late 80s, named scaffold or matrix attachment regions (S/MARs). MARs are sequences, between 0.5 and 6kb, which contain an A/T rich nucleotide region in the middle of it. These sequences have the property to bind to nuclear protein structures, such as histones, and create DNA loops. Figure 1

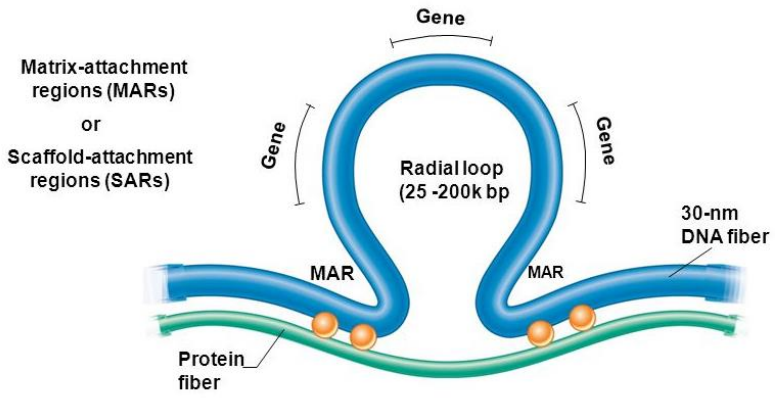

Figure 1 MARs loop structure

The discovery of these MAR-elements led in recent years to the use these elements in stable mammalian cell lines development. Indeed, the structure resulting from the interaction of MARs with nuclear proteins prevents DNA from silencing and even seems to have promoting and enhancing properties in terms of transcription. Since their first discovery, many MAR-sequences have been revealed and patented in order to create commercial mammalian expression vectors, including ones for transient expression. Also, many papers on applications for MARs have been published. The main academic actor in the development of MARs for commercial stable mammalian cell line creation is the UNIL-EPFL in Switzerland. Starting with the evaluation of the well-known chicken lysozyme MAR (lys MAR), then human and mice MARs discovered by in silico analysis, the work allowed to establish MARs as important epigenetic elements in mammalian expression vectors.
Is has been shown that the implementation of one or several MAR sequences in transfection vectors mediate high expression level, stable expression with low unstable transgene and clonal variability.[1–5] Understanding MARs and their use in the creation of stable mammalian recombinant cell lines is the main part of the literature search done in this project.

Beside MARs other key elements are necessary to construct a transfection vector for stable recombinant mammalian cell line development. Surrounding the GOI are e.g, a promotor and a terminator, responsible for the initiation and termination of gene transcription, enhancer elements help recruiting transcription initiator elements. The same expression regulation elements are used to provide a resistance gene to a selection marker, such as an antibiotic. And finally the plasmid contains a bacterial resistance and origin of replication in order to amplify and manipulate it in *E.coli*.

Another important element is the leader peptide sequence, which is an amino acids sequence (10-25 amino acids (aa)) placed on the N-terminal part of the protein. This peptide sequence induces the transport of the protein out of the cytoplasm and is cleaved by specific enzymes before secretion. Studies have shown that the optimal signal peptide sequence is strongly dependent of the target protein, as exemplified for IgG expression.[6]

The MARs elements

One of the most extensively analysed MAR elements in terms of transcription is the 5' chicken lysozyme MAR (Lys MAR).[3] These investigations led to a publication describing this MAR transfected together with the GOI (all in one vector, cis-mode) or co-transfected (MAR and GOI on separate vectors, trans-mode)[15]. By using the BLAST N database the published Lys MAR sequence was compared to BLAST N-deposited MAR sequences. It was then investigated whether patents protecting these sequences could be found within the patent databases at www.lens.org and at NCBI.[16]
Research in the MAR domain continued resulting in many discoveries of MAR regions in human and mouse genomes by in-silico analysis. Some large screening experiments have been then done to find out optimal MAR-sequences, leading to some very efficient transfections vectors.[4,7]
The two main criteria of a successful stable recombinant mammalian cell line are the production of the protein of interest per cell per day and the stable integration of the transgene (eventually in several copies) into the host cell genome. These criteria are the main focus in MAR vector studies performed by Girod et al.[3,4,7] The resulting sequences have then been optimized by remodelling and shortening and most of these new sequences have been patented. Girod et al. and Kostirko et al. then demonstrated that only two sequences have performed better than the Lys MAR regarding the two criteria mentioned above; the S4 (mice) and the 1-68(human) MAR-elements, both patented. [4,5,8]

In this literature search we identified two other well-known sequences that could be used as MAR elements in transient AND stable expression vectors. The human β-interferon and the β-globin MAR (iMAR and gMAR, respectively). These MARs have been used successfully in vector construction to produce eGFP. As shown by Zhao et al. using two copies of β-globin MAR with an SV40 promotor resulted in a 14-fold increase in production of eGFP, compared to control without MAR[1]. This is very similar to the most recent study using S4 and 1-68 MARs which claims to improve eGFP production by 20-fold.[9] One of these MAR sequences could be used as positive control during a MAR selection phase in this project.

A search to find some MAR sequences that could be useful as epigenetic regulator in stable expression has been done in order to use them in our expression vector. We first gave a look at viral MAR sequences, especially the Simian Virus 40 (SV40) the Human Papilloma Virus Type 16 (HPV) and the Ebstein-Barr Virus (EBV). [10–12]
These sequences have been discovered in the late 80s or in the 90s, by enzymatic digestion and testing fragments attaching to the nuclear matrix. A variety of mammalian host cell lines have been tested for integration of these MARs into the genome and it seems that integration strongly depends on the presence of some recombinases, isomerases and other enzymes. As these enzymes are not expressed in all cells equally this also means that the MAR attachment is dependent on the type of host cell used. In this project these viral MARs would be used in CHO and other production cell lines. As we can't predict their affinity without testing them in expression vectors it might be too risky to work with such viral MARs. In addition, the SV40 et HPV MARs have been already used as epigenetic regulators in expression vectors and this use has been patented [10,11]. Therefore, it was decided not to investigate further applications of these viral MARs. The EBV MAR is free of right but its size (5000bp) could be problematic during the plasmid construction.

Another idea is to use the BLASTN database and to compare known MAR sequences against other species and try to find similar sequences which could be used in vectors.

The following table summary every MAR sequences selected that could be used with respect to the found patents.

MAR name	Similar to	Characteristics
Chicken Lysozyme (2950bp)		Already used as epigenetic regulator.[3] discovered in 1996
Chicken Lysozyme like(~3000bp)	Chicken Lysozyme MAR	90-95% identic, from close species
Pig/Sheep X-29 like (~2900 bp)	X-29 MAR(human)	80% identic, Pig/sheep X chromosome, central AT region missing
Beta Interferon like (~2200bp)	Beta Interferon MAR(human)	95% identic, chimpanzee
Beta globulin like (~2300bp)	Beta globulin (human)	70% identic, whales

Figure 2 Available MARs sources, summary

We face many challenges by using new MAR sequences in this work. The main one being that the chosen sequences have not have been used in stable transfections before. So they are no way to predict their effect as epigenetic regulator, even if they are close to some known as efficient MARs. This singularity of MAR is not yet well understood, al lot of sequences have been tried in vectors but the underlying mechanisms are not well defined. This means that one can easily recognize and isolate MAR sequences but hardly predict, *in silico*, their ability to be used as epigenetic regulator.[5]
The first blast we performed was with the beta-Interferon-specific MAR-element, and some sequences from chimpanzee have been identified with on average 95% similarity, which means that this sequence is coming from another species and might be used for our purposes.
The blasting of human β-globulin MAR resulted in an interesting sequence from the Pseudorca *crassidens,* the false killer whale. The obtained sequence matches 70% of the human β-interferon MAR sequence and therefore might be used as a MAR for our purposes.
Some sequences on the X chromosome of some non-human species show similarities with the X-29 MAR, one of the today's most efficient MAR in stable cell line development.[7] But a key central part is missing in these sequences. This sequence contains a 400bp AT rich fragment, proved to be essential for MAR functionality in stable expression.[4] We could also use the Lys MAR or a similar (>95%) one from close species, which is well studied and has been discovered in the late 80s.

There are too many MARs to screen to cover every genomic sequence available, and MARs from non-mammalian species have also been proven to work as epigenetic tool in mammalians cell line development. This could open a much bigger range of target species for MARs selection. By choosing MAR sequences close to existing efficient MARs we lower the risk of choosing a non-functional epigenetic regulator.

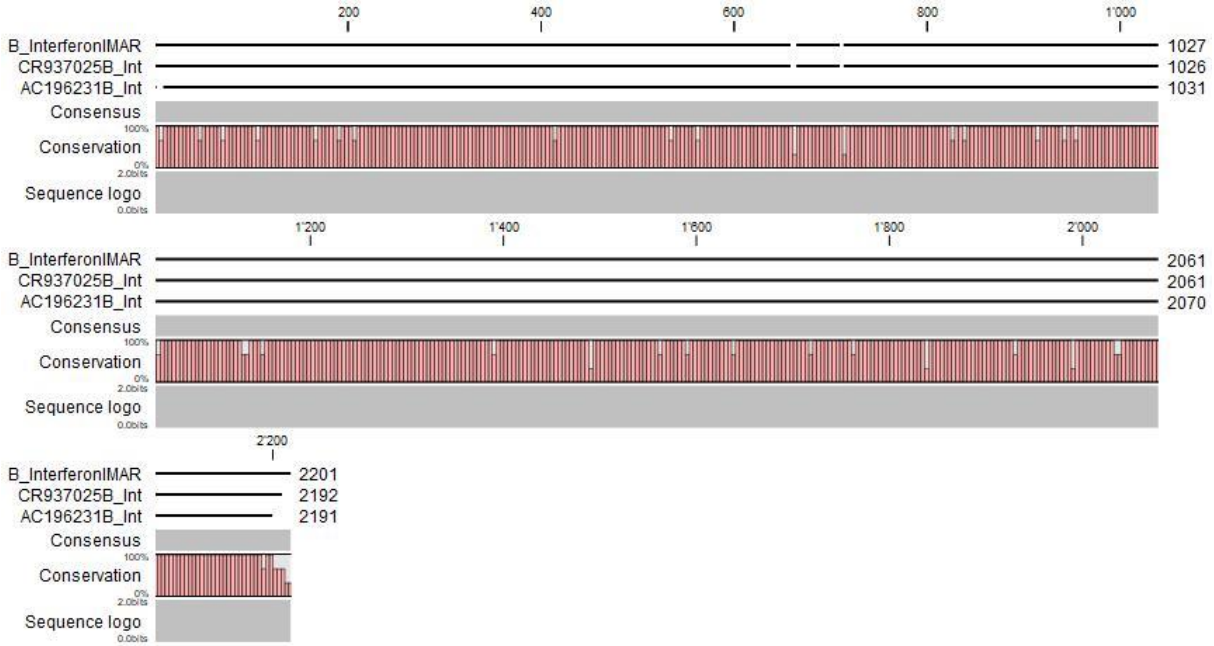

Figure 2 Blasting results from Human Beta Interferon MARs, from chimpanzee and Rhesus Monkey

Plasmid construction

The purpose of this project is to design, construct and transfect several plasmids into a mammalian host cell line. The base for the selection of the best plasmid will be mainly done by analysing eGFP expression levels. Several elements have to be put together in order to structure a plasmid for stable transfection. The skeleton that we will use is based on pcDNA 3.1 +, an expression vector that is not infringing any patents. After looking closely to several plasmid sequences and maps from other expression vectors described in literature, we can summarize the main vectors elements as follow:

Plasmid	**Potential Construction**	**pcDNA.3.1+**
Promoter/terminator of selection marker	SV40\CMV	SV40
Enhancer GOI	Human\CMV\IE\SV40	CAG
Promoter GOI	SV40, CMV, UBC, EF1A, PGK, CAGG	CMV
Terminator GOI	BGH, human gatrin	BGH_PA
MAR	X-29, chicken lysozyme(Lys MAR), 1-68, β-interferon, β-globin	
Selection marker	Hygromycin, puromycin, neomycin	Neomycin

Figure 3 Plasmids genetic elements summary

As mentioned the backbone for this work will be the plasmid pcDNA3.1, used generally for transient transfection in mammalian cells.

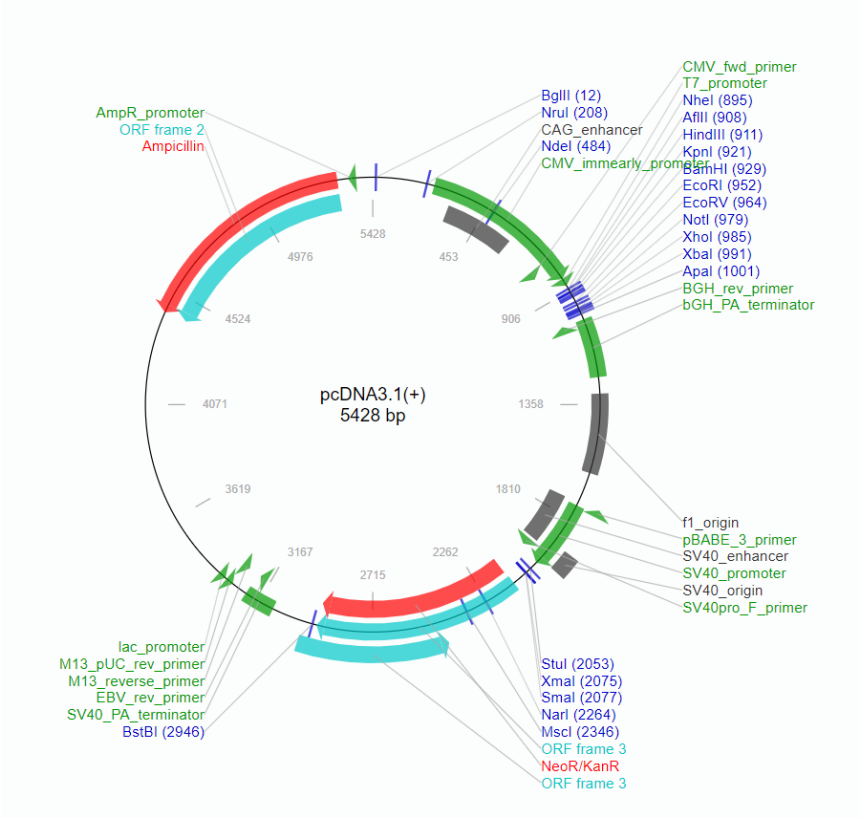

Figure 3 pcDNA 3.1 (+), plasmid map AddGene

The different genetic elements are shown on Figure 3, the MARs chosen will be inserted upstream the CMV promotor and maybe also a second time downstream the SV40 terminator. The combined interactions of GOI promotor and MARs have been studied and led to controversial results[1], the two promotors mainly used in association with MARs are the SV40 and the CMV, both seem to have a similar effect in terms of transcription as soon as they are placed directly downstream the MAR sequence.
The advantage of CMV is the strong expression in a large spectrum of cell type, and independent from the stable expression of SV40 large T antigen.
It is coupled in pcDNA3.1 with a Human β globin terminator (bGH_PolyA) which is a classic eukaryotic and mammalian terminator. It contains six purine-rich zones, which are shown to be essential for transcription termination.[13] The promoter part of the plasmid is designed to be easily replaced, being between two restriction sites, this means that we could also try an SV40 promoter on the GOI.
A short analyse of leader peptides sequences has been done, as mentioned this would concern further steps of the project, namely the IgG stable expression and secretion. Many sequences have been found, for heavy and light chain leader peptides. Fact is that the efficiency of such sequences strongly depends on the target protein and also differs strongly for different IgGs. Which means that these particular sequences will need to be specific for the IgG we may try to produce.[6,14]

For transfections with the created vectors we may also have to consider linearizing the vector before transfection.

A study pointed out the importance of consecutive transfections with exactly the same vector, after one day (21 hours)[9]. They propose several explanations of this technique, but it seems that multiple transfections improve homologous recombination events. This leads to a better transgene expression and integration, by integrating several vector copies at a same chromosomal spot.

Others also demonstrated that the transfection of MAR-based expression vectors done in CIS and TRANS, (meaning that the MAR is on the main plasmid and on a second one, co-transfected, without the GOI and the mammalian selection marker), also significantly increase the expression level. This may be due to the same effect as consecutive transfection and therefore an alternative to that method.[3]

To summary the envisaged modifications that will be done on pcDNA3.1(+) to create a new mammalian stable transfection vector, are summarized below.

Plasmid name	promoter	MAR1	MAR2
pCMV_ChickLys	CMV	Chicken Lysozyme	
pSV40_ChickLys	SV40	Chicken Lysozyme	
pCMV_BInt	CMV	B-Interferon	
pCMV_BIntII	CMV	B-Interferon	B-interferon

Conclusion

To summarize the search done, there have been three phases in the X01 project.
Firstly, to understand the stable mammalian transfection and find out what are the existing methods. This included the choice of the MAR-system as epigenetic regulator, because it allows us to avoid the time and cost consuming methods such as GS or DHFR-based expression systems.
The second phase, which was the longer one, was a dive into MAR-based stable transfection vector studies and understanding of this method that progresses since the early 2000s. This part also included the search for existing patents on sequences or methods, which is not completely achieved, because of the complexity of the patent evaluation and relevance. However, a selection of public MAR-elements has been obtained by this search as well as by discussing with experts. Then we identified different MAR-sequences from different species by performing a BLASTN comparison using known MAR-sequences. Certain interesting MAR-sequences have been identified and these could be used for vector construction.
Finally the pcDNA3.1 plasmid was looked at and the search for promoters and other genetic elements that we will include in this vector was also investigated. A lot of research are focussing on the MAR sequences and positioning of the sequences within the vector, a few of the publications read go into detailed screenings of promotors or selection markers.

Taken together, this report summarizes the diverse possibilities for the creation of a MAR-based expression vector and represents a base for the X02 project, which will include the final choice of plasmid design.

Bibliography

1. Zhao, C. P. Chun-Peng Zhao1, Xiao Guo2, Si-Jia Chen1, Chang-Zheng Li, Yun Yang, Jun-He Zhang, Shao-Nan Chen, Yan-Long Jia & Tian-Yun Wang. Matrix attachment region combinations increase transgene expression in transfected Chinese hamster ovary cells. *Sci. Rep.* **7,** 1–7 (2017).

2. Zahn-Zabal, Monique Zahn-Zabal, Michel Kobr, Pierre-Alain Girod, Markus Imhof, Philippe Chatellard, Maria de Jesus, Florian Wurm b, Nicolas Mermod M. Development of stable cell lines for production or regulated expression using matrix attachment regions. *J. Biotechnol.* **87,** 29–42 (2001).

3. Girod, P. A., Zahn-Zabal, M. & Mermod, N. Use of the chicken lysozyme 5′ matrix attachment region to generate high producer CHO cell lines. *Biotechnol. Bioeng.* **91,** 1–11 (2005).

4. Harraghy, N., Regamey, A., Girod, P. A. & Mermod, N. Identification of a potent MAR element from the mouse genome and assessment of its activity in stable and transient transfections. *J. Biotechnol.* **154,** 11–20 (2011).

5. Kostyrko, K. Kaja Kostyrko, Samuel Neuenschwander, Thomas Junier, Alexandre Regamey, Christian Iseli, Emanuel Schmid-Siegert, Sandra Bosshard, Stefano Majocchi,Valérie Le Fourn, Pierre-Alain Girod, Ioannis Xenarios, Nicolas Mermod, MAR-Mediated transgene integration into permissive chromatin and increased expression by recombination pathway engineering. *Biotechnol. Bioeng.* **114,** 384–396 (2017).

6. Klatt, S. & Konthur, Z. Secretory signal peptide modification for optimized antibody-fragment expression-secretion in Leishmania tarentolae. *Microb. Cell Fact.* **11,** 1 (2012).

7. Pierre-Alain Girod, Duc-Quang Nguyen, David Calabrese, Stefania Puttini, Mélanie Grandjean, Danielle Martinet, Alexandre Regame, Damien Saugy, Jacques S Beckmann, Philipp Bucher & Nicolas Mermod, Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat. Methods* **4,** 747–753 (2007).

8. Arope, S., Harraghy, N., Pjanic, M. & Mermod, N. Molecular characterization of a human matrix attachment region epigenetic regulator. *PLoS One* **8,** (2013).

9. Me´ lanie Grandjean, Pierre-Alain Girod1, David Calabrese, Kaja Kostyrko, Marianne Wicht, Florence Yerly, Christian Mazza, Jacques S. Beckmann, Danielle Martinet and Nicolas Mermod, High-level transgene expression by homologous recombination-mediated gene transfer. *Nucleic Acids Res.* **39,** (2011).

10. MAR_EBV_Jankelevich et al_EMBO_1992.pdf.

11. MAR_SV40_Pommier et al_JVirol_1990.pdf.

12. Tan, S. H., Bartsch, D., Schwarz, E. & Bernard, H. U. Nuclear matrix attachment regions of human papillomavirus type 16 point toward conservation of these genomic elements in all genital papillomaviruses. *J. Virol.* **72,** 3610–22 (1998).

13. Map, P. IP-Free © Mammalian Transient Expression Vectors IP-Free © Mammalian Transient Expression Vectors : CMV Promoter. **8646,** 1–4

14. Ryan Haryadi, Steven Ho, Yee Jiun Kok, Helen X. Pu, Lu Zheng, Natasha A. Pereira, Bin Li, Xuezhi Bi, Lin-Tang Goh, Yuansheng Yang, Zhiwei Song. Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. *PLoS One* **10,** 1–16 (2015).

15. US 2003/0087342 A1, The uses of MAR elements to increase stable and transient transfection efficiency, by Nicolas Mermod, Buchillon (CH); Monique Zahn-Zabal, Lausanne (CH); Markus Imhof, Ecublens (CH); Philippe Chatellard, Lausanne (CH); Pierre-Alain Girod, Lausanne (CH), (2003)

16. https://www.lens.org/lens/patent/US_2003_0087342_A1/fulltext checked 01.02.2018