UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Master's thesis in medecine

TALEN TECHNOLOGY USED TO PRODUCE A SYNOVIAL SARCOMA CELL MODEL

Student
Shaima El Chammah

Tutor
Prof. Ivan Stamenkovic
Institut of Pathology, CHUV

Co-tutor
Dr Carlo Fusco
Institut of Pathology, CHUV

Expert
Prof. Michel Aguet
Institut of Pathology, CHUV

Lausanne, Avril 2015

INDEX

ABSTRACT

Synovial sarcoma (SS) is to this day a devastating malignancy, of which we have very little mechanistic understanding. Two elements characterize SS: the cell of origin, a multipotent myoblast precursor cell, and its signature chromosomal translocation t(X:18) that generates the SYT-SSX fusion gene and corresponding fusion protein. The aim of this project was to produce a TALEN, a tool that could be used to produce SS cells *de novo* by introducing the translocation into a selected cell line (C3H cells). This may give us insight into the translocation's impact on the cell's physiology early in transformation. The TALEN, composed of a nuclease (Folk I) and an amino acid repeat module array, responsible for the specificity of the nuclease's cut point, was generated using a basic cloning technique. The targeted break point in the murine genome was determined following several criteria including its resemblance to human SS break point and the limitation of off-target effect. Damaged DNA produced by the transfection of C3H cells with the TALEN triggered homologous recombination (HR) thereby introducing the LoxP site into the targeted sequence. The lox P sites were then recognized and recombined by CRE thus producing the translocation between chromosome X and 18. Efficiency and specificity of TALEN cleavage must now be demonstrated by PCR screening. HR's specificity and efficiency, must also be established using T7 Endonuclease I mutation detection assay.

INTRODUCTION

Sarcomas are a rare form of cancer that account for 2-3% of all malignancies[i]. They are, however, common in children where they represent 15% of malignant tumors[ii] and are often very aggressive making their impact all the more important[iii]. Indeed, sarcomas have a strong tendency to metastasize and are frequently refractory to conventional anti-cancer therapies. From a genetic point of view, sarcomas can be divided into two categories: the first is characterized by a signature mutation, often a translocation leading to the formation of a fusion gene that encodes a chimeric protein with oncogenic properties; the second, by contrast, has no particular mutation pattern[iv]. A member of the former group, Synovial Sarcoma (SS), will be the subject of this study.

Clinically, sarcomas are classified according to their anatomical site of origin, i.e. bone or soft tissues. Soft tissue sarcomas, or STS, can be found in virtually all connective tissue other than bone including muscle, fat, vessels, the peripheral nervous system and fibrous tissue[v]. Genetic analysis of STS has demonstrated that many are characterized by a specific chromosomal translocation[vi]. STS are rare, with less than 5000 cases yearly in the U.S.A[vii]. Synovial sarcoma represents 10% of all sarcomas[viii] and is slightly more frequent in men (55%) than in women (45%), with an average patient age of 33±16 years[ix]. Despite its name, synovial sarcoma does not seem to be derived from synovial cells or even the synovial lining of a joint. Although the tumor is most frequently encountered in acral parts of the body, in joints, and other associated structures such as ligamentous tissue, it can also be found in visceral organs. While its cell of origin remains unknown, it does seem to derive from a multipotent stem cell capable of mesenchymal and neuroectomdermal differentiation[x]. Recently, Haldar et al. have demonstrated that SYT/SSX expressed under the myf5 promoter in a targeted mouse line induces the formation of SS with 100% penetrance[xi] [xii]. This discovery confirmed suspicions that the cell of origin is somewhere in the myoblast lineage. There are three histological types of synovial sarcomas: biphasic, monophasic fibrous or monophasic epithelial; however this does not influence its level of malignancy or the lesions' location[xiii]. In fact most synovial sarcomas are characterized as high grade tumors: the five year survival rate approximates 70%, but the 10 year survival rate plummets to 45%[xiv].

As mentioned above, about one third of all sarcomas have been linked to specific translocations. Ewing's sarcoma and the EWS/FLI translocation are clearly associated, whereas the FUS/CHOP aberrant protein is related to myxoid liposarcoma and can generate these tumors in mice[xv]. Synovial sarcoma is itself associated 95% of the time to a specific translocation between chromosome X and chromosome 18 (p11;q11). The result of this translocation is a chimeric protein composed of SYT and SSX[xvi]. The former, a ubiquitously expressed protein, contains a N-terminal SNH domain and C-terminal QPGY domain important in transcriptional regulation. Indeed, SNH interacts with BRM and forms part of the SWI/SNF complex involved in chromatin structure remodeling. It is also important to note its role as a transcriptional activator[xvii]. In contrast, SSX exists in several forms, but is only expressed in testis and thyroid[xviii]. It associates with the polycomb group and acts as a co-repressor. What has been elucidated thus far by different research groups regarding synovial sarcoma, may be summed up as follows: the chimeric SYT-SSX protein can produce two opposite functions: transcription activation and repression. The SYT subunit has a major role in chromatin remodeling, one of the key ways epigenetics may regulate gene expression[xix]. However, it is important to note that the SYT-SSX protein cannot directly bind to DNA and therefore requires interactions with other proteins which act as mediators[xx].

These last two characteristics have up until now hindered research on the specific pathogenesis of synovial sarcoma. How does the chimeric protein contribute to the malignant character of the tumor? How and what signaling pathways are activated or inhibited by this translocation? Several pathways have been associated with the SYT-SSX protein including Hedgehog, Notch, RAR, TGF beta and Wnt[xxi]. What has yet to be elucidated is how these different pieces of the puzzle fit together to produce the final result. This is particularly baffling when we take into account the fact that synovial sarcomas may remain dormant for years before they become aggressive[xxii].

To answer these questions a mutant cell would have to be analyzed from the very beginning of synovial sarcoma's pathogenesis: immediately after the translocation. This would enable us to observe the entire evolution of the transformed cell. To this date, this remains a challenging task because the only available biological model is synovial sarcoma primary cells cultivated from surgical samples from patients, which clearly have already evolved to a cancer cell state.

The aim of this work is to produce primary mesenchymal cells, with some degree of pluripotency, containing the SYT-SSX translocation in a way that may mimic the natural evolution of SS cells in humans. There are two considerations to be made here; first, it is crucial to avoid the non-physiological promoters that are usually used in expression vectors as this may aberrantly modulate protein levels in the cell, possibly altering the oncogenic signal. Second, we need to use a cell type that should be similar to the cell of origin of SS: therefore we have chosen the C3H cell line, mainly because they are pluripotent mesenchymal cells with myoblastic features which fit the observations made by Haldar et al. as mentioned above[xxiii xxiv].

The genome-editing techniques developed in the past decade together with the CRE/LOX technology have made it possible to artificially generate chromosomal translocations in a cell line in vitro[xxv].

The CRE-LoxP technology can be used to introduce two loxP sites into a corresponding break point by taking advantage of the homologous recombination (HR) machinery. Once the LoxP is placed into the genome it can be recognized by the CRE recombinase to generate the translocation between chromosome 18 and chromosome X. CRE-LoxP recombination may be also applied for other uses and is further detailed below (Results, page 8).

Up until recently, the use of CRE-Lox was limited by the fact that naturally occurring DNA repair, which was mostly used for mouse gene targeting, relied upon double stranded break (DSB)-dependent homologous repair. However, the latter is quite rare and occurs only in a limited array of cell types and a restricted number of organisms. HR however could be promoted by inducing DSB with a site directed nuclease thus highly increasing the chances of a successful HR-guided insertion of the LOX sites. This procedure ultimately leads to targeted genome editing which allows almost any genomic modification, with very few remaining limitations.

Several genome-editing tools to introduce DSB were developed these past few years, including Zinc Finger Nuclease[xxvi] (ZFN), CRISPR[xxvii] and TALENS. TALs are proteins secreted by Xanthomonas bacteria; the elucidation of their mechanism broadened the horizons of genome editing. Indeed, TALs are composed by unique subunits capable of recognizing any of the four nucleotides (A, T, C or G). By rearranging the order of the individual subunits any DNA sequence of 32 nucleotides can be recognized by an ad hoc engineered TALEN array[xxviii]. In

comparison to the other techniques, it is a good compromise between feasibility and specificity. Indeed, while ZFN is as highly specific as TALEN, this tool is more challenging to create; on the other hand, CRISPR is more efficient when it comes to the production process, but unfortunately, it does have a greater probability of producing off target effect. In our project, TALENs will be used together with the CRE-LOX recombination tool to induce the SYT-SSX translocation responsible for synovial sarcoma.

In summary, TALENs have a wide range of possible uses, and can be exploited in a large array of cell types and organisms. By its high specificity and extensive modularity TALENs could enable us to take advantage of the CRE/LoxP system to introduce an artificial translocation mimicking t(X;18)(p11;q11) in a cell model resembling the one of the cells of origin of SS. There is hope this new technology will give us insight into the mechanisms that underlie primary tumor development.

METHODS

Construction of TAL arrays

The construction of the four protein subunit arrays was based on the following protocol: *Protocol for REAL and REAL-Fast Assembly of TALEN Expression Plasmids*, which was developed by Deepak Reyon, Cyd Khayter, Maureen R. Regan, J. Keith Joung, & Jeffry D. Sander (http://talengineering.org/platforms.html).

-MaxiPrep

Plasmids encoding the TAL subunits which were obtained from ADDGENE were extracted from bacteria using a maxiprep kit according to the manufacturer protocol (QIAGEN). The procedure is the following:

The stubs received from the manufacturer where streaked on agar/ampicillin plates in order to re-amplify the strain. The plate was incubated overnight. A single colony was picked and put into a bacteria tube containing 3ml of LB medium (10 g for 500 ml) and ampicillin (100 µg/ml) and left to incubate for one night. 2 ml of this culture and 500 µl of ampicilline were added to an autoclaved erlenmeyer flasks containing 500 ml LB medium. The culture was grown over night at 37°C and plasmid DNA was purified using the Qiagen maxiprep kit according to the according to the manufacturer's protocol.

-Competent cells

Bacteria cells where transformed using electrocompetent Top10 cells (Invitrogen).

-Plasmid Digestion

1 µg of each 5'end TAL was digested by BamH1 and Bsa1 restriction enzymes using New England Biolabs enzymes, according to manufacturer specifications.1.5 µg of each 3'end TAL was digested with New England Biolabs BamH1 and Bbs1 restriction enzymes, according to manufacturer specifications.

-Agarose gel

Agarose gels were used to run the digested DNA samples and verify the integrity of the digestions.

For the 5'end TALs, a 1.2% agarose/TAE gel was used to run the samples. The TAE/agarose mixture was heated using a microwave until the agarose had dissolved completely. Gel red (Biotium) was added to the melted agarose (1:10000). The samples were compared to the BTSD II lambda reference (Biolabs). The gel was run at 70 V and 500 mA until the bands were properly separated.

The same procedure was repeated with a 2.4% agarose gel for the 3'end plasmids.

-Extraction of DNA from the agarose gel

Plasmids were extracted from the agarose gel using a kit according to the manufacturer protocol (QIAGEN). Each DNA was eluted in 50 µl of TE (10 mM Tris pH 8, 1 mM EDTA).

-Ligation of the digested DNA

For the first and second round of ligations, 10 ng/µl of insert (3'end) and 3 to 4 ng/µl of backbone (5'end) were mixed to 1.5 µl T4 DNA ligase buffer and 0.5 µl of T4 DNA ligase. This was left overnight at 15°C.

-Electroporation and screening

The ligation was precipitated in 10 volumes of 5-butanol and the pellet was washed and resuspended in 4 µl of water.

2 µl of this resuspended DNA was added to the 80 µl of Top10 competent cells (Clontech) and electroporated at 2.5KV, 400 microFarad in 0.2 cm cuvets..

The transformed cells were left to incubate 40 minutes at 37°C in SOC medium. 150 µl of each tube was plated on an ampicillin plate and grown at 37°C overnight.

-PCR Screening

A PCR screening method was then used to identify the positive clones. 5 Colonies per clone were picked and washed in a PCR tube containing: water, 1 ng of dNTP, 1 ng of 5'primer (M13 forward), 1 ng of 3'primer (M13 reverse), 5x coloured TAQ buffer, and 0.2 µl of TAQ enzyme (the total volume= 50 µl). A standard PCR programme was run.

After amplification, 20 µl of each PCR tube was run on a 1.5% agarose gel. The expected size of the band was of 210bp for the first round (each insert measures 100bp, the first round measured the ligation of two inserts). The samples containing this band were considered positive clones. The positive clones which were grown on the plate were picked and grown overnight in a bacteria tube containing 3 ml of LB and ampicilline (100 µg/ml).

The plasmids were then extracted with a minipreparation kit according to kit instructions (QIAGEN) and the DNA was re-suspended in 50 µl of TE.

The proper assembly of the array was controlled a second time by digesting a sample of the DNA with New England Biolabs restriction enzymes Bamh1 and Xba1, according to manufacturer instructions. The digested plasmids were run on a 1.5% agarose gel. By estimating the length of the inserts we were able to confirm that ligation had been successful.

Once the correct assembly of the array was confirmed, the DNA was digested either as a 3' or 5' end, run on a gel, extracted and ligated as it had been done previously.

These steps were repeated until the entire array had been assembled.

Cloning TAL arrays into the nuclease backbone

The plasmids of backbones JDS70, JDS 71, JDS 74, JDS 78 that had been grown and purified by the maxiprep protocol (Qiagen) were digested by New England Biolabs restriction enzyme Bsmb1 according to manufacturer instructions.

Each array was digested by Bbs1 and Bsa1 New England Biolabs restriction enzymes according to manufacturer instructions. All digested DNA was run on a 1.5% agarose gel, extracted with a plasmid extraction kit (QIAGEN), according to the manufacturer's instructions. The appropriate plasmids were ligated using T4 DNA ligase as described above. The rest of the procedure followed the same principles as those described for the production of the arrays.

The correct clones were identified by digesting with KpnI and BamHI and their sequence was verified using the following primers:

JDS2978 Forward 5' TTGAGGCGCTGCTGACTG 3'

JDS2980 Reverse 5' TTAATTCAATATATTCATGAGGCAC 3'

Testing TALEN's cleavage efficiency and specificity

In order to test our arrays' capacity to cleave DNA we will transfect C3H cells with the TALENs. 24-48 hours post-transfection, C3H cells will be lysated and the DNA extracted. This DNA will be run on an agarose gel by eletrophoresis.

Testing HR's efficiency and specificity

Mismatch cleavage essay using T7 endonuclease I will be used to ensure HR was efficacious (meaning that the LoxP site had successfully been inserted into the cleaved DNA). This method relies on specific primers. Mismatch cleavage essay consists of an amplification of a specific sequence of DNA and a digestion by T7 endonuclease I. This method will be adapted from the following reference: https://www.neb.com/tools-and-resources/feature-articles/crispr-cas9-and-targeted-genome-editing-a-new-era-in-molecular-biology.

RESULTS

The aim of my project was to create a tool to enable further research to shed light on SS pathogenesis. In order to better understand SS pathogenesis, two elements had to be considered; the first is the cell of origin. Indeed, as mentioned earlier, the cell of origin has to be permissive to the oncogeneic properties of a mutation or translocation, in order for the tumor to form. To respect as closely as possible the properties of what is believed to be the cell of origin, we used C3H cell line. The second necessary element is the translocation itself. Indeed as mentioned in the introduction, synovial sarcoma's aberrant protein interacts doesn't have a DNA binding domain and therefore affects the transcriptome indirectly by participating in protein-DNA complexes with other proteins; as a consequence, protein levels will likely affect the events leading to synovial sarcoma pathogenesis. Synovial sarcoma would therefore be better studied in a cell model where transcription of SYT-SSX closely resembles the one occuring during the evolution of the disease.

The tool I have created was used to induce the SYT-SSX translocation in the C3H cell line. In order to accomplish this, we have had to rely on two important biological technologies: the CRE-LoxP recombination system and targeted genome editing, in this case with TALEN technology.

CRE-LoxP

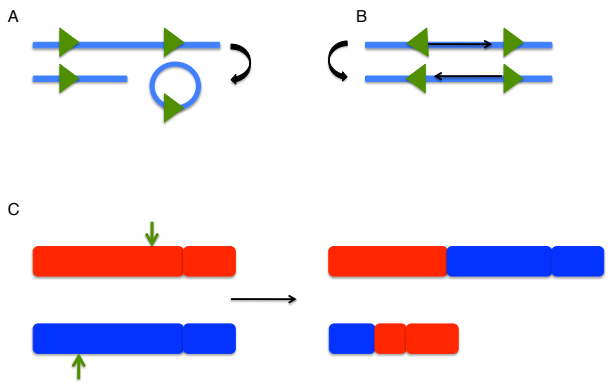

Figure 1. Different uses of CRE-LOX recombinase system: A) deletions, B) Inversions of DNA sequences, C) Translocations

The CRE-LoxP recombination tool is composed of two counterparts. The first is CRE which is the enzyme implicated in the recombination process. This enzyme recognizes specifically the loxP DNA sequence and thereby recombines two sites with each other. This technology has a broad range of uses. It can be implemented to produce insertion or deletion; in other configurations it may be used to induce inversions of DNA

sequences (figure 1). In our case, the LoxP site will be placed on the mouse's homologous regions of the human break points in chromosome 18 and chromosome X, thereby inducing a translocation (figure 1 and figure 2). Once the CRE recombinase is expressed, the introduction of the loxP site in the DNA relies on two elements: homologous recombination (HR) and its template, the oligodesign. Placing the LoxP sites with precision requires a very precise editing of the genome. This can be accomplished only by the homologous recombination process that is often activated when there is damaged DNA to be repaired.

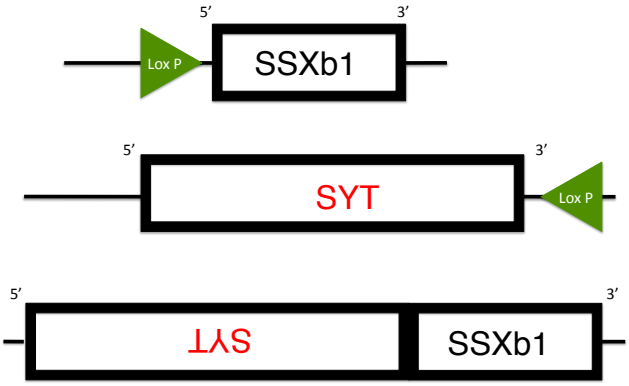

Figure 2. Translocation of SYT gene on chromosome 18 and SSX gene on chromosome X using LOX P.

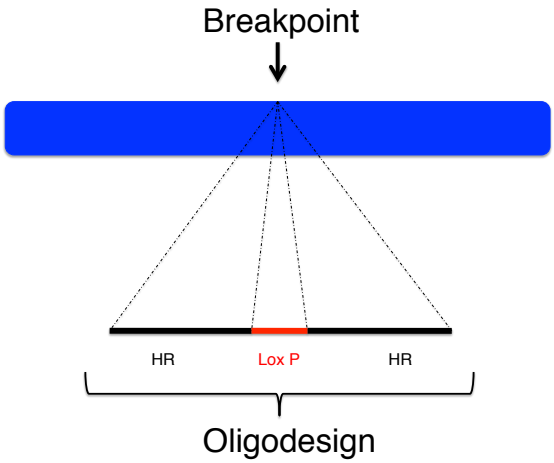

Figure 3. Oligodesign example: the LOX P site is centred by two Homologous recombination (HR) sequences.

Homologous recombination and the CRE-LoxP system are often combined in ES cells to produce KO mice. However, spontaneous HR occurs extremely rarely in almost all differentiated cell lines and adult species. Because of this, researchers have recently developed new methods that can

induce HR; these methods include the TALEN. What the TALEN does is damage the DNA at a specific site. This lesion trigger the repair machinery of the nucleus: HR. Previous studies (REF) have shown that if a short single strand sequence (in our case the LoxP site) flanked by two homologous regions of about 60 BP each (the oligodesign, shown figure 3) is transfected into a cell, it will induce a crossing over around the damage of DNA. This will introduce the desired sequence in the targeted position of the chromosome.

TALEN

As mentionned above, the CRE-LoxP system is used in association with the newly developed TALEN technology. TALEs, Transcription Activator-Like Effectors, may be assembled and synthesized into specific sequences: an array. Each TALE is composed of 33 to 35 amino acid tandem repeat modules. Each module contains on the 12th and 13th position highly variable amino acids. There is a simple one to one code between these two amino acids and a DNA base. In sum, each amino acid repeat sequence will specifically recognize one nucleotide and the specificity of this recognition is ensured by the two amino acids on positions 12 and 13 (figure 4).

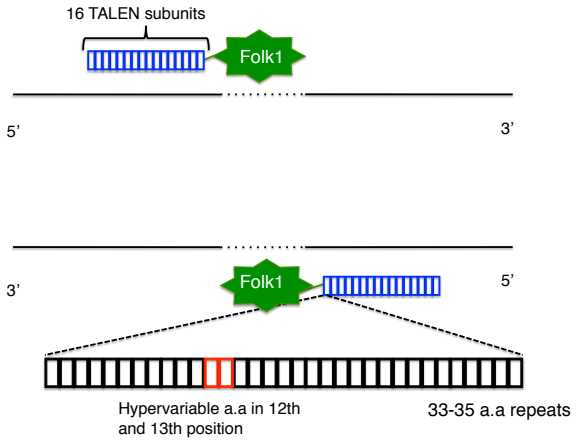

Figure 4. TALE Nuclease model: the TALE is composed 16 subunits, each composed of 33-35 amino acid (a.a) repeats.

The array is associated to a nuclease, FolkI, which cleaves a single strand of DNA. The TALEN therefore works as a dimer: one on the positive strand and one on the negative strand, to produce a highly specific double stranded break. This, in association with the CRE-LoxP recombinase, will be used to produce a translocation[xxix].

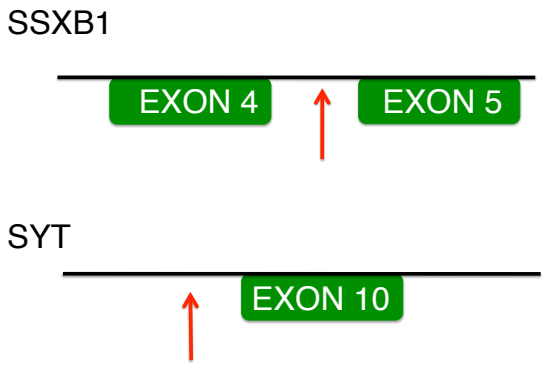

Figure 5. The red arrow indicates the translocation location of gene SSXb1 on chromosome X and of SYT on chromosome 18

In our case, we aim to produce the mutation implicated in synovial sarcoma pathogenesis: a translocation between SYT on chromosome 18 and SSX1b, on chromosome X (mice gene), more specifically, between exon 5 of SSX1b and exon 10 of SYT[xxx] (figure 5).

SYT-SSX Translocation

In order to accomplish this, we had to identify in the murine genome the gene sequences that correspond to those implicated in the human SYT-SSX translocation. NCBI's "Gene" directory was used to find the genomic sequence of SYT and SSX1b, distinguish exons from introns, and identify translated from non-translated regions. According to literature, the most similar region between the mouse and human SSX gene can be found in the C-terminal part of the murine SSX1b gene[xxxi]. However, the organization of exons in humans and mice is not identical. This rendered identification of viable break points on SSX1b and SYT all the more challenging. Indeed, exon 4 of SSX1b contains an extra nucleotide, a guanine, which, after splicing occurs, is reattached to the first two nucleotides of exon 5, thereby producing a triplet for translation. This important element had to be considered when designing the oligodesign and to ensure that the frame of translation is respected.

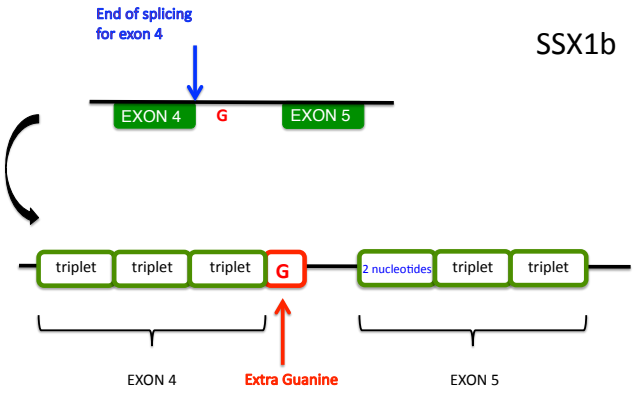

Figure 6. In SSX1b exon 4 contains an extra nucleotide: a guanine (G) which is separated from the exon after splicing and attached to the first two nucleotides of exon 5.

Another important aspect in the identification of a break point is to avoid repetitive sequences as they increase off target risk. This can easily be accomplished by using a program designed to identify repetitive sequences in the genome. SSX1b had no repetitive regions, but the design was limited by the presence of the extra guanine at the end of exon 4. For SYT, however, it seems that most naturally occurring break points were at Topoisomerase II cleavage sites[xxxii]. This was taken into account when selecting the cleavage site on chromosome 18.

Finally, another challenging aspect of the SYT-SSX translocation is that the SYT gene in mice is found on the negative strand, whereas in humans it is on the positive strand. For that reason, the LoxP site will have to be introduced in a design that would invert the SYT gene after recombination (as shown in figure 2). This was essential to produce the closest model possible to the human translocation.

Once the specifics for the CRE-LoxP recombinase had been determined, the sequences containing the breaking points were copied and pasted into the ZiFit web site. This website contains a program which generates the schematics for TALEN production (see Appendix, p 21). TALEN construction must also follow specific rules. The array length must be between 16 and 18bp and must be separated by a 15-30bp spacer. Moreover, the selected DNA sequences were also run on http://galaxy2.informatik.uni-halle.de:8976 for off-target prediction.

Generating the TALEN arrays

Once the position of the break point was determined and the arrays designed, the TAL subunits were cloned as described above (see Materials and Methods).

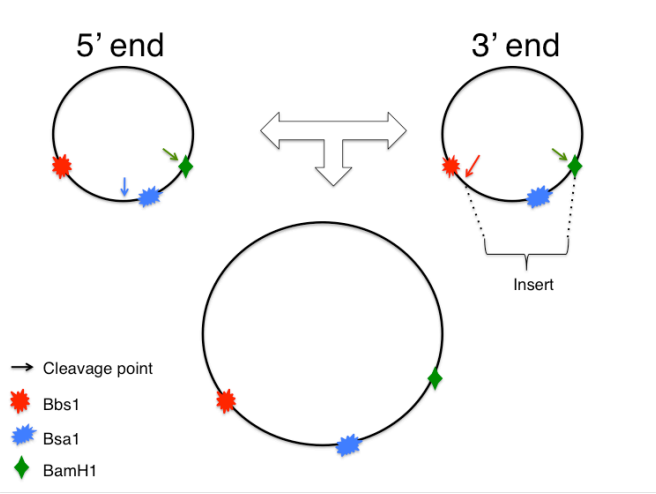

Figure 7. BamH1 and Bsa1 were used to digest 5'ends, while BamH1 and Bbs1 were used for the 3'ends. The clone produced by the ligation boasts all three binding sites.

Briefly, to produce the TALENs standard DNA molecular biology cloning techniques were used. Bacteria stubs containing the plasmids coding for specific amino acid module repeats were

purchased from ADDGENE. In each plasmid, the sequence of interest was excised using the corresponding restriction enzymes. While only the insert was isolated from the 3'ends, both backbone and insert were isolated from the 5'ends. For each cloning step the 3'end was inserted and ligated into the 5'end backbone (figure 7). The backbones were cleaved using BamH1 and Bsa1 whereas inserts were cleaved with BamH1 and Bbs1. While BamH1 cleaves DNA at its binding site, it is important to note that this is not the case for neither Bsa1 nor Bbs1. As shown in figure 7, this allows for each enzyme binding site to be maintained through the production of the next generation of clones, thus permiting the next round of cloning to continue. Finally, once the four arrays were generated, they had to be cloned in frame with the TAL nuclease to generate the final four plasmids, ArrayI_JDS71, ArrayII_JDS74, ArrayIII_JDS78 and ArrayIV_JDS78 (sequences are shown in GeneBank format in the Appendix, pages 27-42). The arrays were checked by digestion (KpnI and BamHI) and agarose-gel electrophoretic analysis, which correctly shows two bands, the backbone of 4763 bp and the Array of 2295 bp (figure 8). Subsequently the plasmids were sequenced using the JDS2978 Forward and JDS2980 Reverse primers to verify that the plasmid sequence conforms to the expected assembled subunit arrays (Appendix pages 26-41).

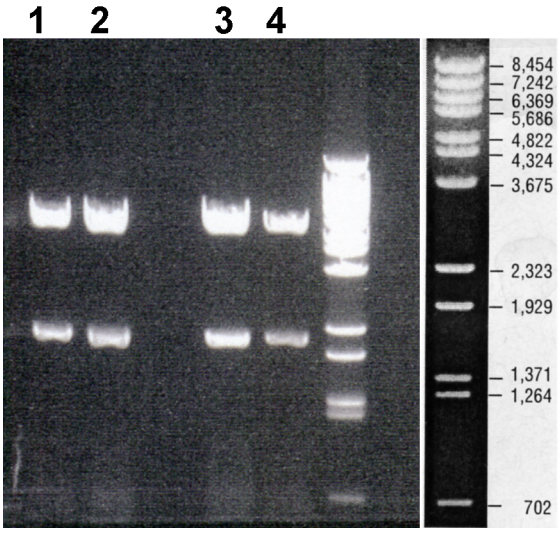

Figure 8. BamHI and KPNI were used to digest ArrayI_JDS71 (lane 1), ArrayII_JDS74 (lane 2), ArrayIII_JDS78 (lane 3), ArrayIV_JDS78 (lane 4) and run on 0.6% agarose gel. The backbone molecular size is 4763 bp and the array size is 2295 bp. The molecular weight markers (right) are Lambda BsteII (New England Biolabs).

DISCUSSION AND CONCLUSION

Synovial sarcoma is a highly malignant tumor that often affects adolescents and young adults and of which we have very little mechanistic understanding. The aim of this project was to produce a tool that would give impetus to future research on synovial sarcoma. Two key elements that characterize SS and that are important for rational therapeutic design are the cell of origin, which, although unknown is most likely some multipotent myoblast precursor cell (as shown in Haldar et al.'s article, see the introduction); and the SYT-SSX fusion protein that is a consequence of the (tX:18) chromosomal translocation. The tool we produced, a TALEN, will be used to introduce the signature translocation in a C3H cell. The latter displays mesenchymal pluripotent cell properties and has a tendency to differentiate into myoblasts, thereby sharing many of the characteristics of what is believed to be the SS cell of origin. The importance of introducing the translocation is explained by the chimeric protein's behaviour. Indeed, synovial sarcoma's aberrant protein affects the transcriptome indirectly by participating in protein-DNA complexes with other proteins; as a consequence, fusion protein expression levels will likely affect the events leading to synovial sarcoma pathogenesis. Moreover, producing a SS cell *de novo* will hopefully give us insight into the impact of the translocation on the cell's physiology early in transformation.

TALEN technology, engineered arrays composed of amino acid repeat subunits capable of recognizing any of the four DNA bases, was concluded to be a very efficient way to generate a cell containing both the translocation and properties of the SS cell of origin. This may ensure production of cells that resemble SS cancer cells in humans. Nonetheless, the cell model produced will distinguish itself from its human counterpart. As a direct consequence what we may observe following our research may not be identical to reality. First, as mentioned above, the cell of origin is strongly believed to be a myoblastic precursor cell[xxxiii], but this has yet to be fully demonstrated. If the selection in the cell line is erroneous the processes that will be analyzed may lead to nothing or worse, wrong conclusions. The second obvious issue is the animal model. C3H cells are mouse cells, but mice do not naturally develop synovial sarcoma unlike cats and dogs for example. Choosing a cell-line from an animal that does not develop the disease naturally may also lead to inconclusive results.

Furthermore, the aberrant protein produced by the translocation will be similar but not identical to its human counterpart: in reality, exons which make up SYT on chromosome 18 and SSX on chromosome X, are not identical in human and mouse, despite our selection of the closest resembling sequences (the C terminal of SSXb1). Neither is exon reorganization after splicing. Moreover, the fact that SSX is a paralogue (one ancestral gene was duplicated and the homologues produced by this process evolved independently from one another), further complicates the matter.

Now that the four arrays have been sequenced, confirming that they correspond to the targeted DNA, it is important to evaluate and demonstrate TALEN and HR's specificity and efficiency. For the former C3H cells will be transfected with the TALEN and allowed to grow. The transfected genomic DNA will later be extracted and PCR amplified using primers specific to the targeted cleaved DNA. If the corresponding DNA sequence is absent from the DNA sample, the primers will not able to bind to the DNA. As a result, amplification will not occur. Following amplification, DNA will be run on an agarose gel; if the TALEN cuts specifically three bands of three different sizes should be visible on the gel. The highest molecular weight band would represent un-cleaved genomic DNA (meaning that the method is not 100% effective as only a portion of the DNA was cleaved). The two lower bands would represent the cleaved DNA. The negative control for this experiment will be genomic DNA extracted from C3H cells which will not have been transfected with the TALEN and therefore produce only un-cleaved genomic DNA. The efficiency of the TALEN will be determined by the strength of the band signal.

A similar method to the one described above will also be used for assessing HR: T7 Endonuclease I mutation detection assay. In order to verify that HR functions properly, we have to ensure that the loxP site is in fact introduced into the TALEN-damaged DNA. When HR is activated by DNA damage, it uses the oligodesign (figure 3) as a template to introduce the LoxP site. This should create three groups of DNA: heterozygous DNA containing one copy of the lox P site, homozygous wild type (WT) DNA and homozygous DNA for the lox P site. DNA from C3H cells which have been damaged by TALEN and been exposed to the oligodesign will be extracted and PCR amplified using primers specific for the transition between SYT/LoxP and SSX/LoxP. The amplicons produced will then be digested with T7 endonuclease I, an enzyme that recognizes specifically the indel caused by the presence of a mismatch between the two DNA strands (heterozygote for LoxP). Finally, the DNA will also be run on an agarose gel. If

HR is successful, this gel should contain three separate bands for both chromosome X and 18. The first one represents the largest sample of DNA, the DNA which was not cleaved by endonuclease I. This can either be the homozygous WT DNA or homozygous DNA for the LoxP site. The two lower bands represent the cleaved DNA which was cleaved at the loxP site. The negative control used in this case would be TALEN transfected C3H DNA to which the oligodesign will not have been added.

Although many obstacles have made the project challenging, and despite the fact the end product may not be identical to the human reality, this project will provide a unique opportunity to closely study the different steps and the entire pathogenic pathway driven by the SYT-SSX translocation, in a cell model which, in many aspects, closely resembles the natural state. By producing the SYT-SSX translocation in a multipotent myoblast precursor cell, we are one step closer to a better understanding of the primary cell environment that allows the fusion protein to express its full oncogenic potential. This is crucial to acquire a stronger grasp of sarcoma pathogenesis and could allow for improved diagnosis, treatment and therefore prognosis of SS, a malignancy that can still leave patients with a dire quality of life and loss of autonomy[xxxiv].