Project no.: 037226

Project acronym: MEGATOOLS

Project title: New tools for functional genomics based on homologous recombination induced by double strand break and specific meganucleases

SPECIFIC TARGETED RESEARCH OR INNOVATION PROJECT

Final activity report

Period covered: from 01/10/06 to 30/09/09 **Date of preparation:** 10/11/09

Start date of project: 01/10/06 **Duration:** 36 months

Project coordinator name: F. PAQUES

Project coordinator organization name: CELLECTIS SA **Revision:v0**

1. PROJECT EXECUTION

The genome sequence programmes have contributed a huge amount of information, and opened even more possibilities. An exhaustive catalogue of genes is now available for many organisms, but the real meaning of this information remains to be deciphered. For example, most identified genes have no known function. Sometimes, homologies with other identified genes might give insight into the function, although this exercise can prove relatively perilous. In fact, a large amount of experimental work will be necessary to address the complexity of functional genomics, and the classical approaches might prove vain. Two kinds of large scale approaches have been envisioned so far. On one hand DNA chips have provided temporal information about the expression of all genes in an organism. On the other hand, knock-out and RNAi knock-down experiments provide information on the effect of loss-of-function of a particular gene. Ideally, one will like to have other possibilities like modifying, adding or subtracting genetic material in a controlled manner and target not only protein-coding sequences but any kind of DNA sequence like regulatory regions, introns etc… The way to achieve this is through meganuclease-induced recombination, a technology that allows for very highly efficient homologous gene targeting. To work properly, this approach requires enzymes that can target and cleave genes at precise locations within the chromosomes of selected cells. Very specific endonucleases should be used that ideally would target only one nucleotide stretch in a whole genome. The aim of the MEGATOOL project was to develop a large number of new sequence-specific endonucleases to recognize and target almost any possible DNA sequence in any living cell or organism, as well as to optimize homologous recombination, to provide scientists with a powerful tool to do functional genomics. In this programme, we have focused on rodent genomes, and more specifically on the mouse genome.

The MEGATOOL consortium is a partnership of academic and industrial units which provide all the necessary expertise in biochemistry, molecular-, structural-, cell-, and computational-biology to achieve its ambitious goal.

In conclusion, in a research area that remains challenging and has numerous implications not only for basic research but for human health, MEGATOOLS aimed at developing new tools, significantly advancing our knowledge in homologous recombination beyond current state-of-the-art, and feeding back immediately its discoveries into the community giving it strong leverage with an enabling breakthrough, in ways that should be scientifically and economically relevant.

Since meganuclease-induced recombination represents an extremely powerful tool for gene alteration, we focused on the generation of four kinds of results:

(i) A large collection of novel meganucleases. This collection of novel proteins should greatly enhance the repertoire of natural meganucleases, and thus, allow for the targeting of a large number of genes in organisms whose genome has been sequenced, with a strong focus on rodent genomes.

(ii) The means to exponentially increase this collection. The collection of novel meganucleases should provide a unique database of characterized DNA binders. Structural and statistical studies should reveal the laws governing these interactions, and these data could in turn be used in a predictive way, for the design of novel meganucleases.

(iii) The methods, procedures and quality standards to make these meganucleases widely usable as research tools.

(iv) A refined method to use these meganucleases in cells. The focus as on mouse cells for functional genomics, providing a direct validation.

During the 36 month of the project major progress has been made towards achieving its ambitious goals:

- 3D-structural information for two I-CreI based heterodimeric meganucleases and for I-DmoI in complex with their respective DNA targets has been obtained by crystallography and provides a rational basis in the development of gene replacement tools.
- FOLD-X, a computer algorithm for the structure-based prediction of the stability of proteins and protein complexes has been improved for the quantitative prediction of protein-nucleic acid interactions.
- New locally altered I-CreI derivates has been generated, thus the theoretical targetable DNA sequences has been expanded and the combinatorial rate of success increased.
- I-CreI based meganucleases have been improved by abolishing the unwanted homodimers subspecies formation.
- New engineered meganucleases derived from the chimeric DmoCre meganuclease have been produced.
- Highly specific engineered Meganucleases have been produced against targets in ROSA26 and GLUTAMATE SYNTHETASE (GS), Hypoxanthine ribosyl transferase (HPRT) mouse genes.
- Using a new approach, new meganucleases targeting 6 new mouse genes (figure below) could be produced in few months.

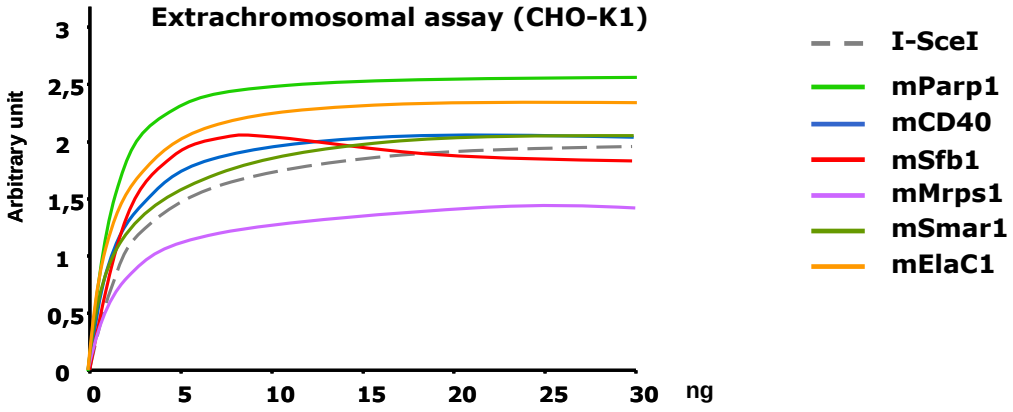

Figure : Cleavage activity in mammalian cells of the meganucleases targeting mouse genes. Activities of different meganucleases were tested inour SSA extra-chromosomic assay in CHO-K1 cells. genes targeted by the meganucleases are indicated.

- New methods standard procedures were developed to monitor meganucleases activity at the endogenous locus.
- Tools to monitor the potential toxicity of the engineered meganucleases were developed to ensure the safety of the meganucleases producted
- Meganuclease targeting the HPRT gene was successfully used in gene targeting experiments. Expression of the gene inserted in the HPRT locus appears to be stable over time.

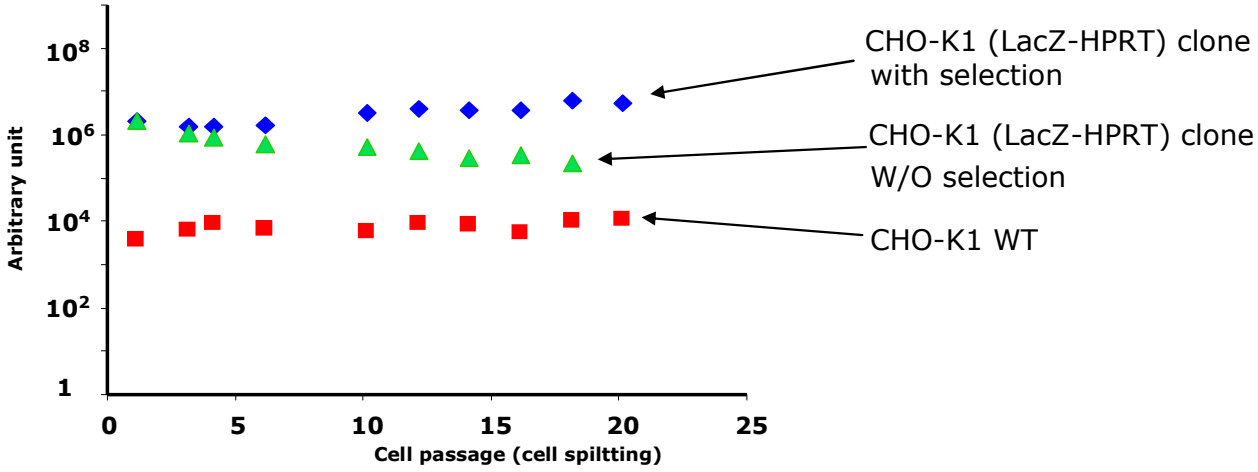

Figure: targeted insertion at the HPRT locus allows stable expression The LacZ gene wasinserted in the HPRT locus by targeted insertion and ernzymatic activities were quantified at each cell passage. Cells were cultured with or without selection pressure.

- The meganuclease targeting the HPRT gene is now part of a product commercialized by CELLECTIS SA.

Figure: Meganucleases can be engineered to change their recognition site sequence. This breakthrough in protein engineering alleviates the need to pre-engineer cell lines with a natural meganuclease recognition site. Engineered meganucleases achieve targeted integration on "wild type" cell lines.

- **targeted gene integration in your lab's cell line**
- **guaranteed gene expression**
- **sustained protein production**
- **clone homogeneity**
- **fast & effortless**

cGPS® and cGPS®Custom are the unique ways to stably express a protein from your cell line!

- The tremendous amount of knowledge produced during this project are protected by 5 patents

- **11** articles have been published in international publication journal

Crystallization and preliminary X-ray diffraction analysis on the homing endonuclease I-Dmo-I in complex with its DNA target.
P. Redondo, J. Prieto, E.Ramos, F. Blanco & G. Montoya (2007) Acta Cryst F Struct. Biol. 63(Pt 12):1017-20.

Generation and analysis of mesophilic variants of the thermostable I-DmoI homing endonuclease.
J. Prieto, J-C Epinat, P. Redondo, E. Ramos, D. Padró &, G. Montoya, F. Pâques and F. Blanco. J. Biol. Chem. (2008) Feb 15;283(7):4364-74.

Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences.
Fajardo-Sanchez E, Stricher F, Pâques F, Isalan M, Serrano L.
Nucleic Acids Res. 2008 Apr;36(7):2163-73.

Molecular basis of recognition and cleavage of the human Xeroderma pigmentosum group C gene by engineered homing endonuclease heterodimers.
P. Redondo , J. Prieto , I. Muñoz , A. Alibés , F. Strichter , L. Serrano, S. Arnould , C. Perez , J.P. Cabaniols , P. Duchateau, F. Paques , F. Blanco & G. Montoya. (2008) Nature Nov 6;456(7218):107-11

Crystal structure of I-DmoI in complex with its target DNA provides new insights into meganuclease engineering.
P. Redondo, M. Marcaida, J. Prieto, E. Ramos, S. Grizot, P. Duchateau, F. Paques, F. Blanco & G. Montoya (2008) Proc Natl Acad Sci U S A. 2008 Nov 4;105(44):16888-93.

Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas S, Lemaire L, Montoya G, Blanco FJ, Pâques F, Duchateau P. Nucleic Acids Res. 2009 Sep;37(16):5405-19

Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds
Sylvestre Grizot, Jean-Charles Epinat, Séverine Thomas, Sandra Rolland, Frédéric Pâques, Philippe Duchateau. Nucleic Acids Res. 2009 (accepted for publication)

Structure-based DNA binding prediction and design Alibés, A., Serrano, L., Nadra, A.D. Zinc Finger Proteins: Methods and Protocols. Humana Press. Eds. Joel Mackay and David Segal. *In press*.

Structure-based prediction of Protein-DNA binding specificities Alibés, A., Nadra, A.D., De Masi, F., Bulyk, M.L., Stricher, F., Serrano, L " *Submitted*.

Homing endonucleases: from basics to therapeutic applications (2009) Maria J. Marcaida, Ines G. Muñoz Francisco J. Blanco • Jesus Prieto Guillermo Montoya. Cellular and Molecular Life Sciences In press

Molecular basis of SCID DNA recognition by engineered heterodimers and single chain meganucleases. (2009) Ines G. Muñoz, S. Subramanian, J. Coloma, Nekane Merino, Maider Villate Frédéric Pâques, sylvestre Grizot, Philippe Duchateau, Francisco J. Blanco, Jesus Prieto, Guillermo Montoya. Submitted.

Project objectives and major achievements

The specific objectives of the project are:

A) To combine state-of-the-art technologies in biochemistry, 3D structure analysis and modeling, and protein engineering in order to develop target-specific meganucleases.

B) To explore the development of meganucleases those are suitable for genome engineering in mouse cells.

Our major achievements are:

A) Improvement of a technology platform to provide tailor-made meganucleases for the purpose of precise "genome surgery"

B) Crystallization and structure resolution of several meganucleases in complex with their cognate target

C) Improvement of the computer algorithm FOLD-X for quantitative prediction of protein-DNA interactions.

D) Implementation of the MEGATOOLBASE

E) Production of custom-made meganucleases targeting the ROSA26, GLUTAMATE SYNTHETASE (GS), Hamster's Hypoxanthine ribosyl transferase (HPRT) mouse genes.

F) Establishment of standard procedure for genome engineering

G) Commercialization of the first meganuclease-based genome engineering tools

Workpackage 1: structure determination

Lead Partner: **CNIO (Guillermo Montoya)**
Partner: **CRG**
People involved: Jesús Prieto
Pilar Redondo
Francisco Blanco
Guillermo Montoya
Inês Muñoz
Maria Marcaida

Workpackage objectives and starting point of work at beginning of reporting period

The project started with the production of crystals of the variants that targeted mutations on XPC and SCID. During these 3 years we obtained the crystals, data were collected and the structures solved and refined. At the moment we are finishing to write some articles about the new variants V2V3 directed against the RAG target generated by Cellectis

Workpackage progress

In this period we have succeeded to isolate and crystallize some of heterodimeric meganucleases based on the I-CreI scaffold. We have obtained crystals of the Amel3-4 and Ini3.-4 derivatives in complex with the Xeroderma pigmentosum group C target DNA (Figure 1.1). The crystals contain cleaved and bound DNA because they have been grown in the presence of Mg^{2+} and Ca^{2+} respectively.

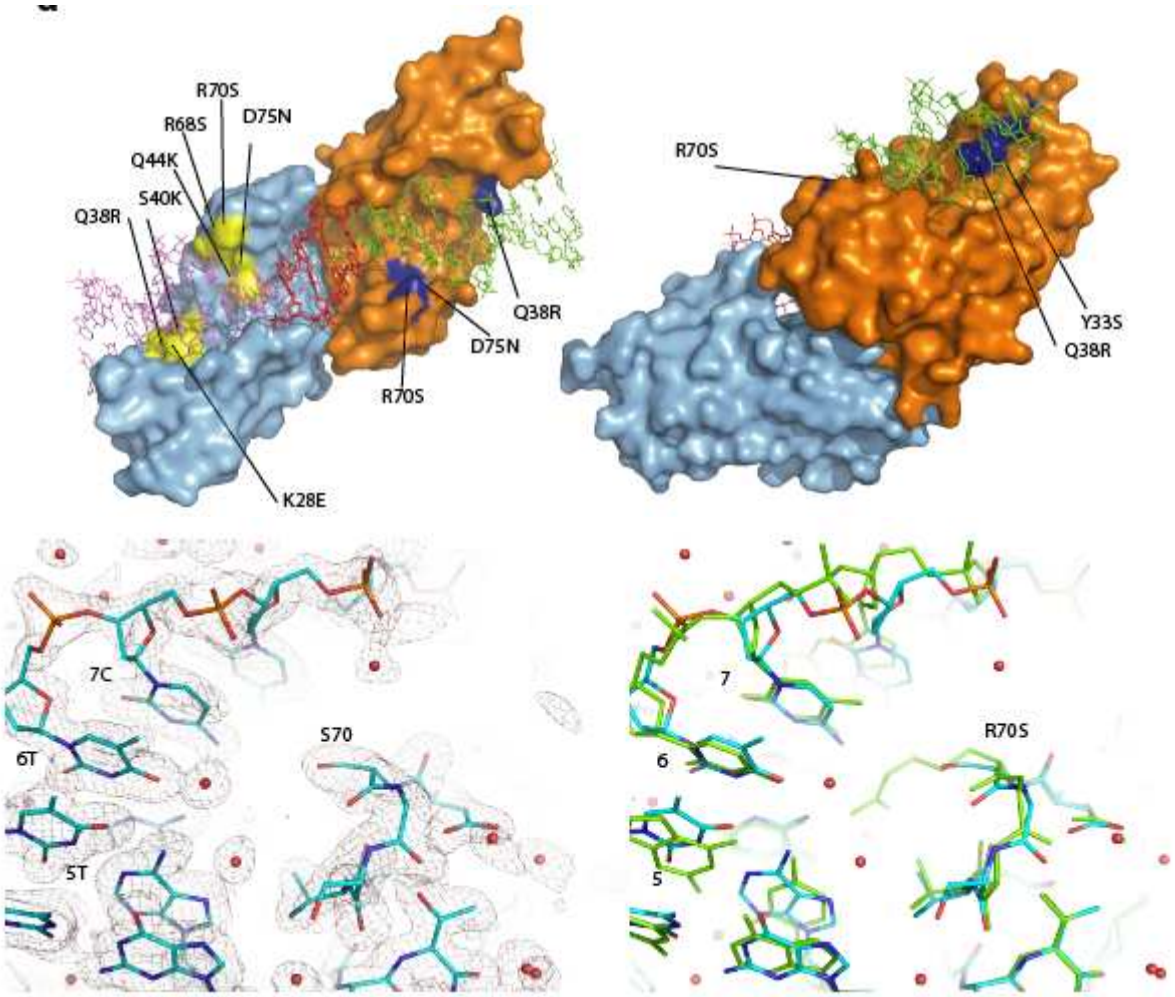

Figure 1.1.- Crystal structures of the Ini3-4 heterodimer. Surface models of the Ini3-4 structure; the different monomers are shown in blue (Amel4 and Ini4) and orange (Amel3 and Ini3). The DNA is colored according to the binding regions of each monomer (violet for Amel4 and Ini4, green for Amel3 and Ini3, and red for the region corresponding to the cohesive ends of the DNA after cleavage). Mutations are mapped on the protein surface in yellow and blue for monomers 3 and 4, respectively. 2Fo-Fc omit maps, with a contour level of 1.25 □ of the Mg^{2+}-bound Ini3-4 structure in the region of the Q38R and the R70S mutations. Superimposition of the heterodimer structures with I-CreI (green), showing the local differences produced by the mutations in the DNA-binding area. WT structures correspond to pdb entries (1g9y and 1g9z).

We have also solve the structure of heterodimeric variants in complex with the DNA of their RAG gene target (Fig1.2). Mutations on the RAG gene generate the disease called severe combined immunodeficiency (SCID). In addition we have recently solved the structure of the single chain variants that target the DNA sequence of the RAG gene in its locus in presence of Mg 2+ and Ca2+ to deepen our understanding on the molecular basis of these variants (Fig1.3).

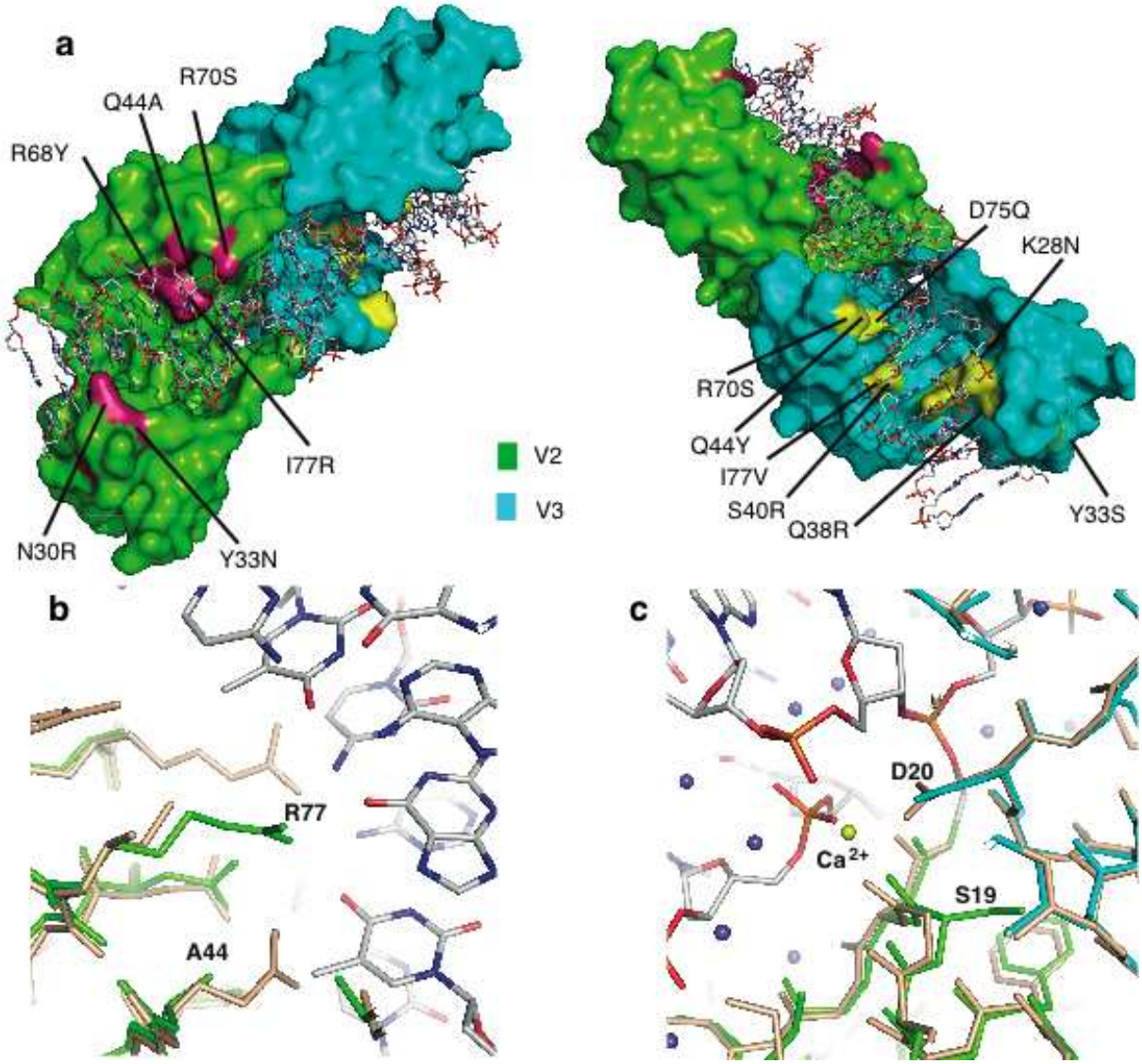

Figure 1.2.- Crystal structure of the V2V3 heterodimer. **A**. Surface model of the V2V3 structure in complex with RAG1 DNA. Mutations are mapped on the protein surface in magenta and yellow for each monomer. **B.** Detail of the I77R and Q44A mutations involved in protein DNA interactions compared to I-CreI. **C**. Detailed view of the G19S mutation in the V2K7E(G19S) / V3E8K heterodimeric protein. The RAG1 DNA target is depicted below with the 5NNN and 10 NNN base regions in bold (stand B correspond to the coding strand)

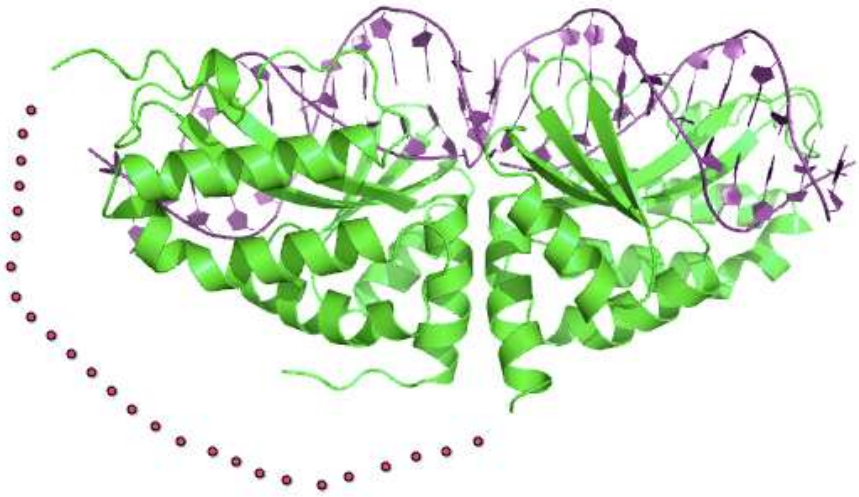

Figure 1.3.- Crystal structure of the single chain engineered meganuclease the loop that joins the heterodimers is depicted with a line of magenta points

Furthermore we have solved the structure of a new meganuclease (I-DmoI) in complex with its target DNA (see figure 1.4). The enzyme has been solved with the bound and cleaved target and suggesting the possibility of a sequential DNA cleavage mechanism.

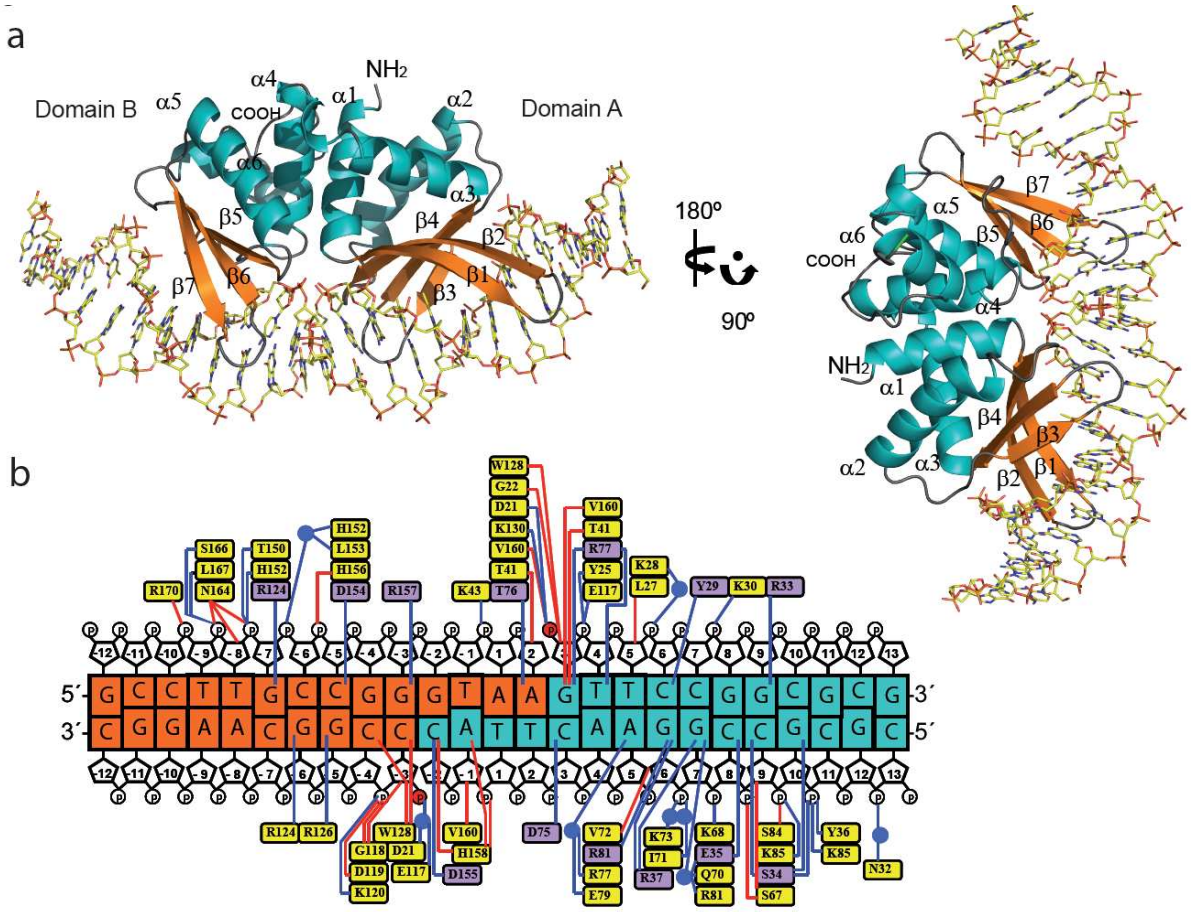

Figure 1.4.- Crystal structure of I-DmoI in complex with its target DNA. **a)** The protein moiety is colored according to its secondary structure (□-helices in blue, □-strands in orange and loops in grey), the complex is shown in two different orientations. The calcium ion is shown as a red sphere. Whereas domain A contains two loops that contact the DNA (L1a and L2a), the domain B only has one loop (L2b) engaged in contacts with the nucleic acid. L2a and L2b are primarily associated with the central bases of the target site, and L1a is associated with bases outside that region —reflecting the asymmetry of the target recognition by I-DmoI. The crystallization oligonucleotide construct is shown below **b)**. Throughout the text the individual bases are named with a subindex strandA (coding strand) or strandB (non-coding strand) indicating the DNA strand where they are located. The protein-DNA contacts are show in the scheme. The cleavage sites are indicated by the red phosphates and each cleavage product is colored in blue (domain A) and orange (domain B) according to the protein domain that binds. Contacts colored in blue and red are associated with polar contacts and van der Waals interactions respectively. Blue dots represent water molecules involved in the interaction. Amino acids depicted on a yellow or violet background represent interactions with the bases or the DNA skeleton (ribose or phosphates) respectively.

Deviations from the project workprogram

none

Status of Deliverables and Milestones

Status of the deliverables for the period and forecast for the other deliverables of the WP.

Del. no.	Deliverable name	Original delivery dates (**)	Revised delivery dates (**)	Status (***)	Nature
D1.1	Expression & purification	Month 12	Month 12	**finished**	**R**
D1.2	Biophysical characterization	Month 12	Month 12	**finished**	**R**
D1.3	Enzyme-DNA characterization	Month 24	Month 24	**finished**	**R**
D1.4	Crystal structure	Month 36	Month 36	**finished**	**R**

Status of the milestones for the period and forecast for the other milestones of the WP.

Milest. no.	Description	Original target date (**)	Revised target date (**)	Status (***)
M1.1	Purification & biophysical characterization	Month12	Month12	finished
M1.2	Enzyme-DNA characterization method	Month24	Month24	finished
M1.3	Protein coordinates	Month36	Month36	Finished

(**) Dates are expressed in project month (1 to 36).
(***) Status = Not started – In process – Finished
R= report, P= prototype, D= demonstrator, O= Other

Publications

Crystallization and preliminary X-ray diffraction analysis on the homing endonuclease I-Dmo-I in complex with its DNA target. P. Redondo, J. Prieto, E.Ramos, F. Blanco & G. Montoya (2007) ***Acta Cryst F Struct. Biol***. *63(Pt 12):1017-20. Epub 2007 Nov 21*.

Generation and analysis of mesophilic variants of the thermostable I-DmoI homing endonuclease. J. Prieto, J-C Epinat, P. Redondo, E. Ramos, D. Padró &, G. Montoya, F. Pâques and F. Blanco. ***J. Biol. Chem***. (2008) Feb 15;283(7):4364-74. Epub 2007 Nov 12

Molecular basis of recognition and cleavage of the human Xeroderma pigmentosum group C gene by engineered homing endonuclease heterodimers. P. Redondo , J. Prieto , I. Muñoz , A. Alibés , F. Strichter , L. Serrano, S. Arnould , C. Perez , J.P. Cabaniols , P. Duchateau, F. Paques , F. Blanco & G. Montoya**.** (2008) ***Nature*** Nov 6;456(7218):107-11.

Crystal structure of the homing endonuclease I-Dmo-I in complex with DNA reveals the target cleavage and recognition mechanisms. P. Redondo, M. Marcaida, J. Prieto, E. Ramos, S. Arnould, F. Paques, F. Blanco & G. Montoya (2008) **Proc Natl Acad Sci U S A.** Nov 4;105(44):16888-93**.**

Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease (2009) Sylvestre Grizot, Julianne Smith, Fayza Daboussi, Jesús Prieto, Pilar Redondo, Nekane Merino, Maider Villate, Séverine Thomas, Laetitia Lemaire, Guillermo Montoya, Francisco J. Blanco, Frédéric Pâques and Philippe Duchateau. **Nucleic Acids Res** 37, 5405-5419.

Homing endonucleases: from basics to therapeutic applications (2009) Maria J. Marcaida, Ines G. Muñoz Francisco J. Blanco • Jesus Prieto Guillermo Montoya. **Cellular and Molecular Life Sciences** In press

Molecular basis of SCID DNA recognition by engineered heterodimers and single chain meganucleases. (2009) Ines G. Muñoz, S. Subramanian, J. Coloma. Nekane Merino, Maider Villate Frédéric Pâques and Philippe Duchateau Francisco J. Blanco, Jesus Prieto, Guillermo Montoya. Submitted.

Workpackage 2: Design of new specificities

Lead Partner: **CRG (Luis Serrano)**

People involved: Luis Serrano
Alejandro Nadra
Andreu Alibés
François Stricher
Michela Bertero (Management)

Workpackage objectives and starting point of work at beginning of reporting period

The workpackage objectives are: a) the improvement of FoldX software (software developed by Partner3) in order to make it a tool that reliably predicts quantitative data on protein-DNA binding; b) generation of libraries of meganucleases so any DNA sequence can be targeted ; and c) i*n vitro* validation of some of the *in silico* predictions.

Progress towards objectives

New FoldX version

To improve the integration of DNA inside FoldX, Partner 3 first changed the partial charges of all atoms forming the bases aromatic ring to homogenize the electrostatic field and prevent extra interactions on the edges. At the same time, Van der Waals clashes were increased for heavy atoms of equivalent partial charges to mimic dipole-dipole repulsion (which is normally not taken into account on FoldX). To be able to correctly assign to each DNA base its pairing partner, all inter-chain contacts of DNA strands were analyzed in all high resolution structures (less than 2.1 Å) from the PDB database. This gave a set of parameters defining precisely the geometry of pairing bases, independent of the global “backbone” conformation of the strands.

DNA mutation is a combinatorial problem involving nine dihedral angles moves and several angles distortion per base. To simplify, and as FoldX does not enable backbone movement yet, Partner 3 decided to take into account only one dihedral angle and one angle distortion, defining the base position. In order to achieve this, we first superimpose the new base on the old one and we then move it slightly around, based on constraints derived from the whole pdb analysis previously done. Each

one of this movement is made in combination with the pairing base and we iteratively move neighboring bases and residues to adapt to the new environment.

A test set of 99 protein mutants binding DNA were created (including 46 conservative mutations) taken from the ProNIT database. For each of these mutants, the changes in affinity for the nucleic acid were experimentally determined and the comparison with the FoldX predictions is shown in Figure 2.1.

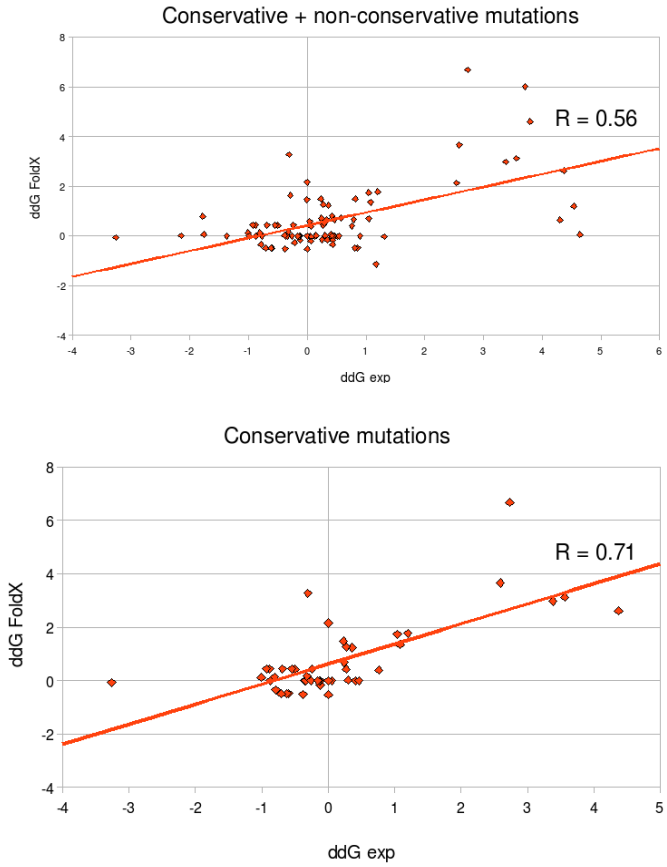

Figure 2.1. FoldX vs experimental data, validation set.

Meganucleases' libraries

We have created mutant libraries for I-DmoI (based on the crystal structure solved by the CNIO group) and E-DreI (based on the structure with pdb 1MOW). For each of the two meganucleases, and using the last version of FoldX, we have searched for those that could interact specifically with DNA bases. We have followed two methods for creating the libraries. The first one is simply mutating each of those residues to all other amino acids and combine this with each of the 64 possibilities of the neighboring DNA triplet. The second method was done by first mutating all residues around the residue contacting DNA to Alanine, to give more freedom to the mutants, and then proceeded the same way as the first method.

The results are a series of matrices describing the effect (in difference of interaction energy versus the WT structure) for each mutation and each DNA triplet. A graphical example is shown in Figure 2.2.

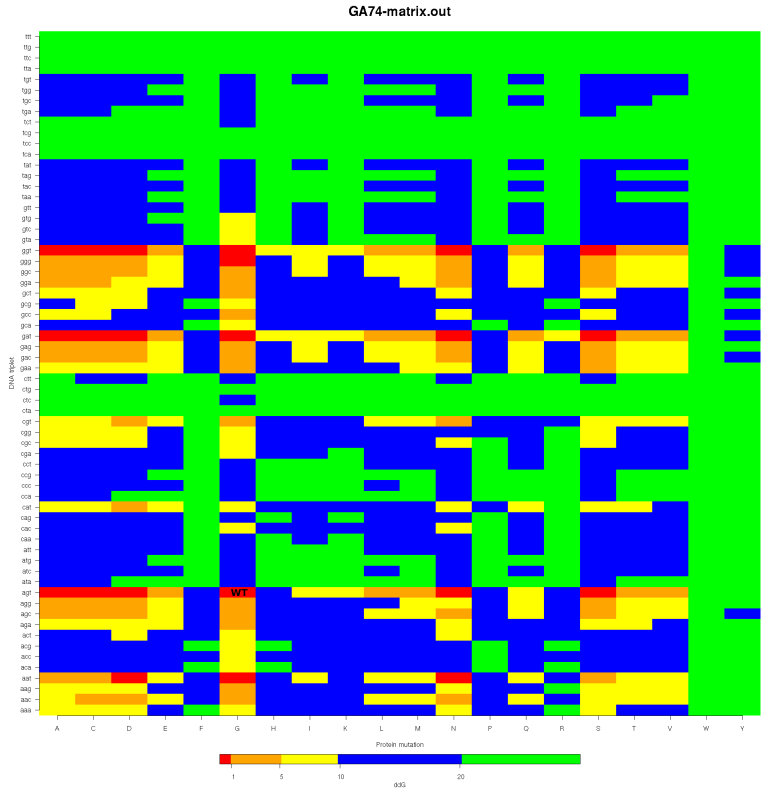

Figure 2.1: Matrix graphically describing the changes in interaction energy in E-DreI due to *mutations in the residue G74.*

Improvement of DmoCre

In order to improve the stability of DmoCre reported by Cellectis, Partner 3 analyzed its structure by mean of FoldX. Starting with E-DreI (PDB:1MOW) crystal structure as template, the corresponding residues were mutated to construct the DmoCre protein. Each residue was then mutated to the remaining 19, excluding those participating in the active site or located at the DNA interface. As a result, Partner 3 proposes fourteen mutations (G9N, I10V, W19E, L47W, F51Y, I52M, R55H ,I98F ,R105M ,A108S ,I113L ,E158K ,N193M and F194L), that should improve dimer interface up to ~7 kcal / mol (compared to E-DreI) without affecting the folding stability significantly. We should keep in mind that maximizing the interaction energy is not necessarily good for the activity. The list of mutations for the engineered DmoCre has been sent to Cellectis for experimental characterization and show a good cleavage activity. An experimental validation of the increase in stability is under way.

Specificity of I-CreI and their mutants

In collaboration with the group of Dr. Montoya and Cellectis we have analyzed the specificity of I-CreI and the two mutants solved by the CNIO group (Ini3-4 and Amel3-4) as shown in Figure 2.3. A structural characterization of the differences between the three structures has been done. This analysis is part of a join paper by the three institutions (Redondo et al., 2008).

An in silico analysis with FoldX of the sequence-structure relationship of the DNA from I-CreI, Ini3-4/XPC and Amel3-4/XPC structures in the absence of the protein moieties have been performed by modeling the WT sequence onto the XPC structure and vice versa. The calculated differences in

overall energy showed that the WT sequence is energetically more compatible with the XPC DNA structure than the XPC sequence with the WT DNA structure. These results suggest that some sequences may force the DNA to adopt a conformation energetically unfavorable for binding to a given I-CreI structure corresponding to another DNA target.

These results suggest that the use of computer assisted protein design to generate meganucleases with customized specificity for very different DNA sequences would require the support of different protein-DNA structures to manage the conformational diversity of protein and DNA.

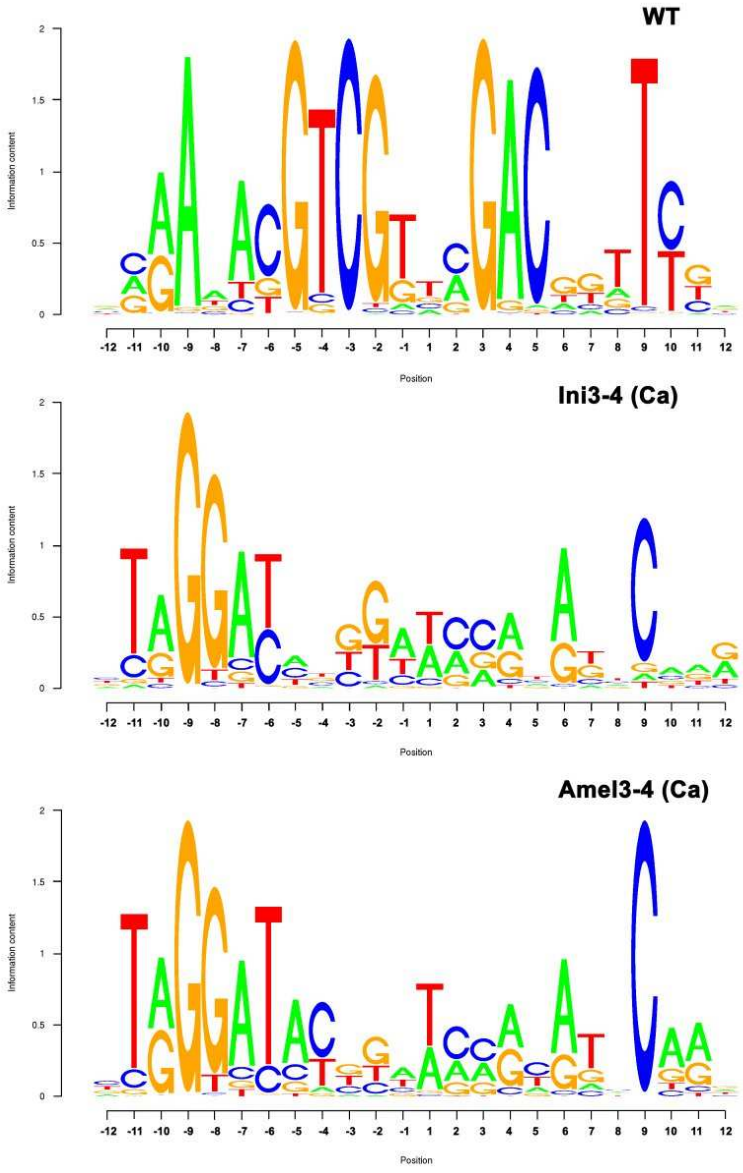

Figure2.2: Specificity logos for the WT and the two proposed mutants.

DNA position 11 in I-CreI

Libraries constructed by Cellectis appear to require pyrimidines at position eleven. With the aim of designing mutants more permissive at this position, we evaluated mutations for four different scaffolds (Wt, Ini3, Ini4, Amel3 and Amel4). By mutating one or two residues we were able to target any base at position 11 for all the templates with energies ranging from -4.2 to 0.3 kcal / mol, being most of them favorable. Amel4 scaffold seems to be the best one for allowing any of the bases at this position since single mutants for serine 32 gave energies around -3 kcal / mol for any of the bases, having a lot of fabourable variants for tuning discrimination. In WT, Ini4 and Amel4 one residue appears to be responsible for the discrimination at position 11: Y33. When mutated to Serine (Ini3 and Amel3) it allows and prefers a T. Mutations preferring a base at position 11 appear to be only slightly affected by the identity of base 10, except variants for. Y33 which makes specific contacts with that base.

Mutation landscape of existing I-CreI structures

20 I-CreI mutants were provided by Cellectis together with the DNA sequence they are known to bind. As a test to know how much can be achieved with the available structures of I-CreI and a protein design algorithm, we have modeled these 20 complexes using the wild type, amel3-4, ini3-4 and RAG1 structures. We considered that binding was predicted when the interaction energy for the mutated structure was within 2 kcal/mol of the interaction energy of the original crystal structure and no single residue had large intraclashes (larger than 2 kcal/mol). We found that their binding could be predicted for 9 out of the 20 sequences.

The reason why we fail to predict binding for the remaining 11 complexes is that we find with FoldX an important loss of interaction energy and also that large intraclashes are predicted that make the resulting structure not reliable for prediction.

These results suggests that more structures are needed in order to use protein design tools to predict binding to new sequences with a high success rate and some of these mutants are now being solved by the CNIO partner.

Using theI-CreI scaffold to design mutants for specific DNA sequences

Cellectis provided 3 different DNA sequences for which we tried to design the best I-CreI variants with the three structures available at that time: the wild type one, amel3-4 and ini3-4. Our protocol includes the following steps:

- Find in the template residues that could be grouped as they interact with neighboring bases. Six spots were defined (Figure 2.4), made of 2 or 3 residues, only in one of the monomers, as the binding sequences are palindromic.
- Mutate the interacting residues to Alanine
- Mutate the DNA seq in the crystal to the chosen DNA
- Do an exhaustive library for each of the 6 spots. That means 20^2 (400) or 20^3 (8,000) were generated
- Sorting the results for each spot according to its interaction energy ($\Delta\Delta G$), we selected the first candidate taking for each spot the best binder predicted
- That candidate was then optimized by choosing residues that could improve specificity, wherever it was appropriate.

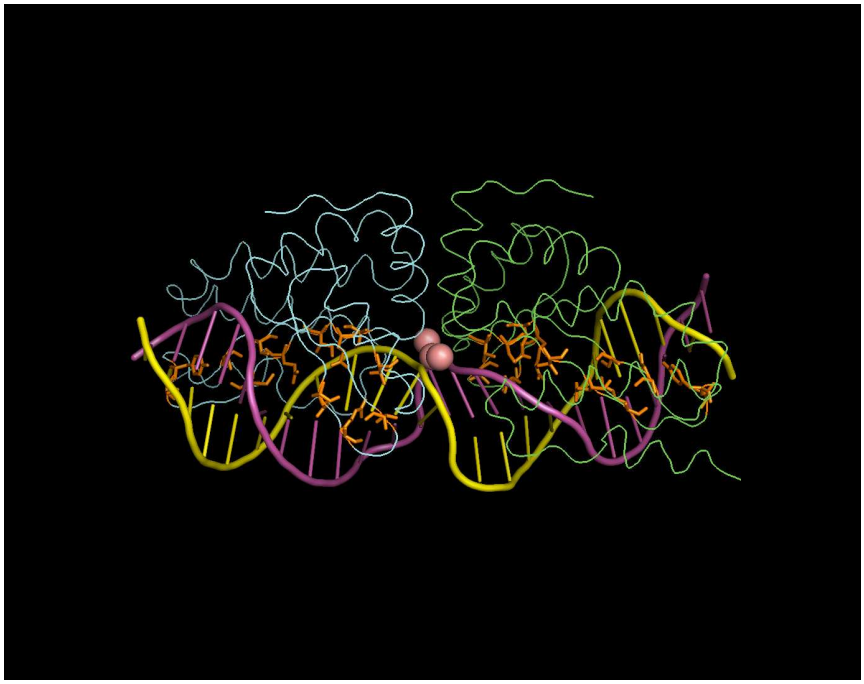

***Figure2.3**: Mutated residues in orange*

The designed I-CreI variants were sent to Cellectis and failed to cleave the DNA for which they were designed. Testing is underway to see if they bind as predicted. Also, Cellectis is performing random mutagenesis on them to see if they find a close variant that might cleave.

Publications

Redondo, P., Prieto, J., Muñoz, I.G., Alibés, A., Stricher, F., Serrano, L., Cabaniols, J-P., Daboussi, F., Arnould, S., Perez, C., Duchateau, P., Pâques, F., Blanco, F.J., Montoya, G. "Molecular basis of recognition and cleavage of the human XPC gene DNA by engineered homing endonuclease heterodimers" (2008) Nature, 456, 107-11

Marcaida, M.J., Prieto, J., Redondo, P., Nadra, A.D., Alibés, A., Serrano, L., Grizot, S., Duchateau, P., Pâques, F., Blanco, F.J., Montoya, G. "Crystal structure of I-DmoI in complex with its target DNA provides new insights into meganuclease engineering" (2008) Proc. Natl Acad. Sci USA, 105, 16888-16893

Alibés, A., Serrano, L., Nadra, A.D. "Structure-based DNA binding prediction and design" *In press*. Zinc Finger Proteins: Methods and Protocols. Humana Press. Eds. Joel Mackay and David Segal.

Alibés, A., Nadra, A.D., De Masi, F., Bulyk, M.L., Stricher, F., Serrano, L. "Structure-based prediction of Protein-DNA binding specificities" *Submitted*.

Del. no.	Deliverable name	Original delivery dates (**)	Revised delivery dates (**)	Status (***)	Nature
D2.1	New FoldX version	Month 12		**Finished**	**RP**
D2.2	First list of mutants with Predicted specificity	Month 12	**Month 18**	**Finished**	**RP**
D2.3	First In vitro validation	Month 24		**Finished**	**RP**
D2.4	Second list of mutants with Predicted specificity	Month 24		**Finished**	**RP**
D2.5	Second In vitro validation	Month 36		**Finished**	**RP**
D2.6	First structurally-constrained library	Month 12	**Month 18**	**Finished**	**RP**
D2.7	Second structurally-constrained library	Month 24		**Finished**	**RP**

Status of the milestones for the period and forecast for the other milestones of the WP.

Milest. no.	Description	Original target date (**)	Revised target date (**)	Status (***)
M2.1	List of designed meganucleases	Month 12	**Month 18**	**Finished**
M2.2	Experimental analysis	Month 24		**Finished**

Workpackage 3: screening of meganuclease libraries

Lead Partner: **CELLECTIS (Fréderic Pâques)**
People involved: Jean Charles Epinat, Anne-Sophie Petit, Marie-Ange Carrico, Virginie Gerbault, Diane Leclerre, Catherine Mikonio, and Noemie Pinard participated to the High throughput screening in yeast and CHO (CELLECTIS).
The protein engineering study was conducted by Philippe Duchateau, Sylvain Arnould, and Sylvestre Grizot (CELLECTIS)

Workpackage objectives and starting point of work at beginning of reporting period

The aim of workpackage 3 is: (i) to create novel meganucleases to enrich the collection of I-CreI derivatives in order (ii) to increase the rate of success of the combinatorial approach used to generate artificial meganucleases. Also, we will apply the same methodology to 3 other scaffolds. At the kick-

off meeting held on November 9th 2006, the partners decided to put the emphasis on scaffolds I-MsoI, I-SceI and DmoCre (Figure 3.1).

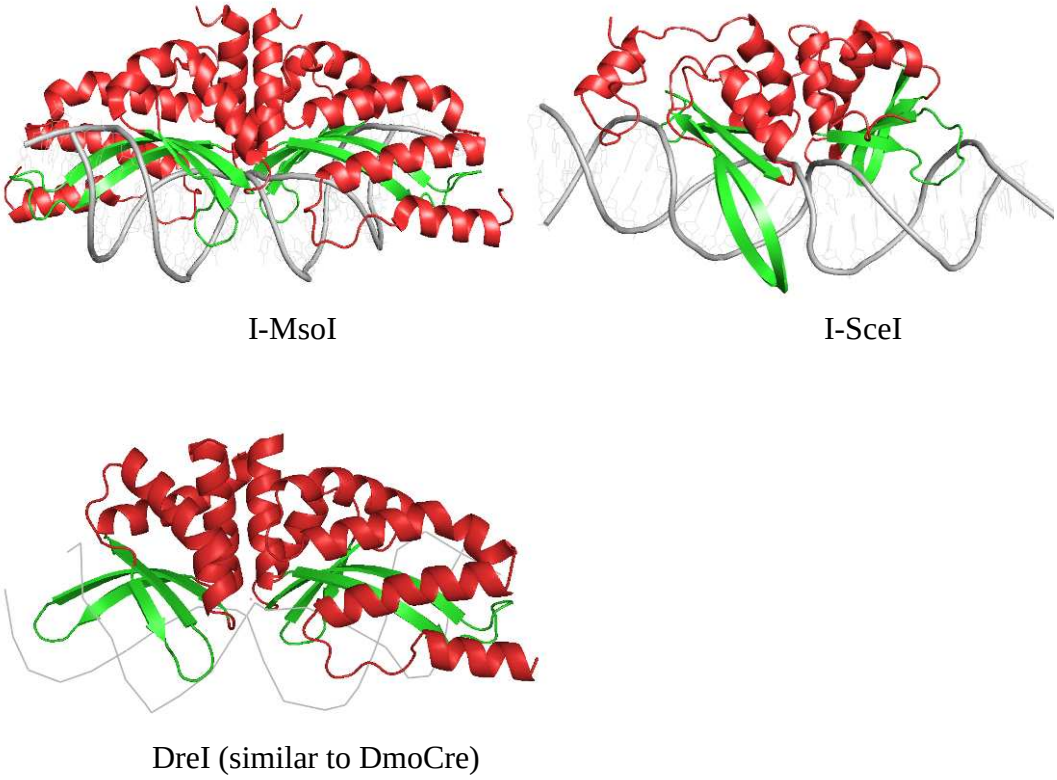

Figure 3.1 : 3D-structure of the 3 chosen meganucleases (green= DNA-binding domains)

Progress towards objectives

A/ I-MsoI

I-MsoI is an I-CreI isoschizomer, therefore the entire collection of I-CreI targets owned by Cellectis can be used. We have analysed the 3D-structure of the protein and we were able to define 2 sub-domains interacting with distinct sequences of the target (Figure 3.2).

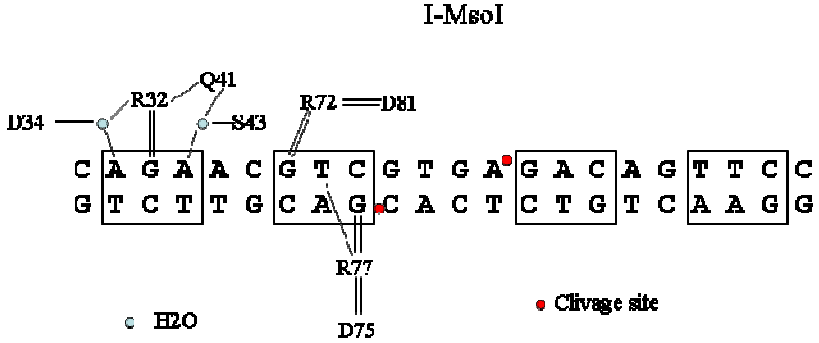

Figure 3.2: DNA targets and proteins sub-domains of meganucleases I-MsoI

The cleavage profile of I-MsoI was first determined on the 64 10NNN_P and 64 5NNN_P palindromic targets derived from the I-CreI C1221 target. The I-MsoI protein is able to cleave 20 out

of the 64 10NNN_P targets (Figure 3.3) while only one target (5GTC) is cleaved amount the 64 5NNN DNA sequences.

On one hand, a first set of libraries mutated at positions 32, 41, or 32, 43 were generated and screened in yeast against the 10NNN_P targets. 246 positive clones able to cleave at least one 10NNN_P target were isolated, resulting, after sequencing, in 96 unique meganucleases. Altogether the I-MsoI derivate proteins are able to cleave 37 out of the 64 10NNN_P targets (Figure 10A).

On other hand, two other libraries mutated at positions 75, 77 or 47, 72, 75 were generated and screened in yeast against the 5NNN_P targets. No new cutter could be identified with the library concerning the 5NNN targets. A close examination of the I-MsoI structure revealed the presence of 2 aspartates at positions 77 and 81 that would have to be mutated in order to decrease energetic tension caused by the replacement of the basic residues R72 and R75 in the library that satisfy the hydrogen-acceptor potential of the buried D77 and D81 in the I-MsoI structure.

.

A/ B/

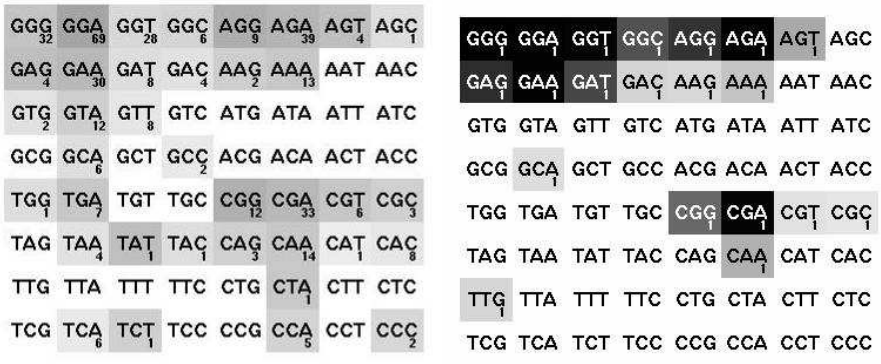

Figure 3.3: hitmap against the 64 10NNN_P targets. A, Mlib1 library; B, I-MsoI profiling. The number below each cleaved target is the number of I-MsoI mutants with different sequences cleaving this target. For each target, the grey level is proportional to the mean of cleavage

So far, the homing endonuclease I-MsoI is still not active in CHO-K1 cells. Therefore, in parallel to the precedent study and in order to generate a molecule with high activity in CHO cells, we have performed a random mutagenesis on the molecule and screen the library for endonuclease activity in vivo in CHO cells against the C1221 target. This first experiment allowed us to identify one mutant with weak cleavage activity.

In order to achieve high cleavage activity with this scaffold, a second run of random mutagenesis is now in progress. However, it did not yield to an I-MsoI mutant active in CHO cells. For that reason, the I-MsoI scaffold was definitively put on stand-by during the end of the present program.

B/I-SceI

We have analysed the 3D-structure of the monomeric protein I-SceI and we have defined 2 potentially independent sub-domains of the protein interacting with distinct sequences of the target (Figure 3.4). The respective 64 and 16 degenerate targets have been designed and cloned in the appropriate vector.

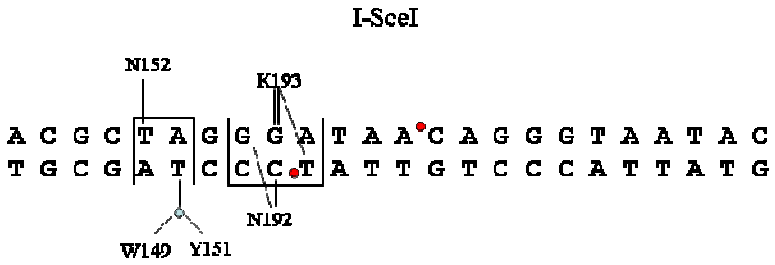

Figure 3.4: DNA targets and proteins sub-domains of meganucleases I-SceI

Two sets of degenerated targets concerning the positions 7NN and 4NNN were cloned in the appropriate vector for yeast screening.
A first library mutated at positions 150 and 152 did not give any new cutter when screened against the 16 degenerate 7NN I-SceI targets. As a consequence a new library, regarding positions 150, 151, and 152 has been constructed but not tested yet.

An I-SceI library mutated at positions 193 and 192 has been generated and has been screened in yeast against the 64 4NNN_P targets. No new cutters could be identified.

The disappointed results obtained with the I-SceI scaffold prompted us to lower the priority on this scaffold and put the emphasis on the DmoCre and I-CreI scaffolds and to begin a study on a brand new scaffold: the I-ChuI protein.

C/I-ChuI

I-ChuI is a brand new scaffold. The protein isolated from *C. humicola* is a single chain homing endonuclease with two LAGLIDADG elements like I-SceI that cleaves the following DNA sequence: 5'-GGTTTGGCACCTCGATGTCGGCTCA-3'. Before undertaking any protein engineering study, we would like to solve the structure of the protein in complex with its DNA target in order to determine the protein-DNA interaction map. However, when we first checked the protein for activity using a yeast screening assay, we found the protein inactive. We performed therefore an optimization step by random mutagenesis that allowed us to isolate five positive clones that all carry at least the Q142R mutation (Figure 3.5). As none of the positives displayed a strong cleavage activity, a second optimization step was done taking the I-ChuI Q142R mutant as template that yielded several mutants with a strong cleavage activity in yeast. Three of them were subcloned in a mammalian expression vector and probed for activity using an extrachromosomal SSA assay in CHO cells. Figure 3.5 shows that they are active even not reaching the I-SceI activity level. These three mutants have been transferred to Partner 1 for expression assays.

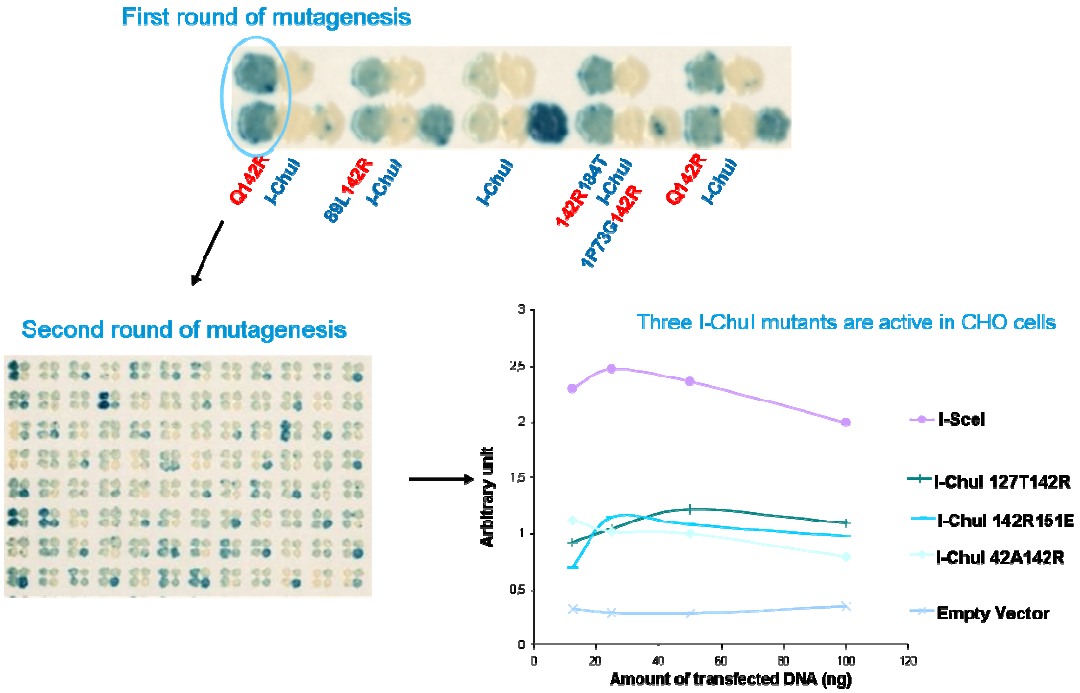

Figure 3.5 Scheme of the activity optimization process for the I-ChuI protein

D/DmoCre

DmoCre is a hybrid protein for which the N-terminal domain of I-DmoI has been fused to one I-CreI monomer. The resulting protein cleaves the hybrid D2-C1 hybrid target (figure 3.6).

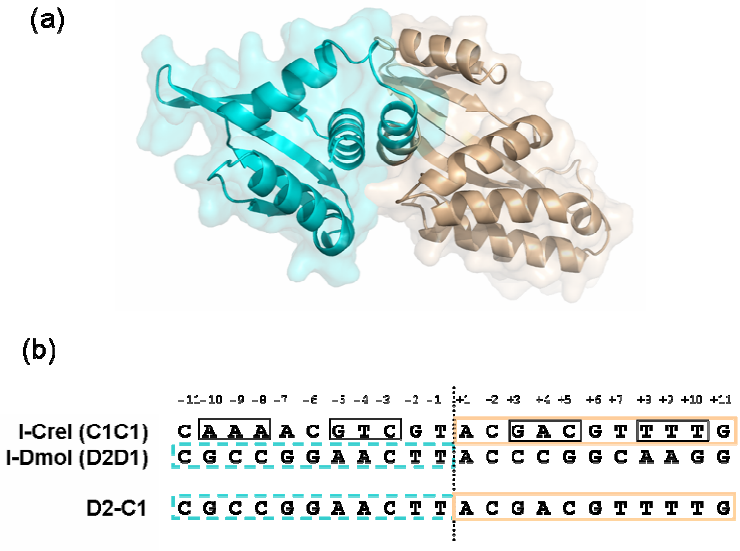

Figure 3.6: (a) Representation of DmoCre with the N-terminal domain of I-DmoI (cyan) is fused to one I-CreI monomer (brown). (b) The 22bp DNA targets for the I-CreI, I-DmoI and DmoCre proteins are depicted. The I-CreI 10NNN and 5NNN regions are boxed.

We analysed the I-DmoI part of the artificial DmoCre protein. Based on the published 3D-structure of E-DreI (similar to DmoCre), we have identified 3 sub-domains of the protein interacting with the DNA target (figure 3.7).

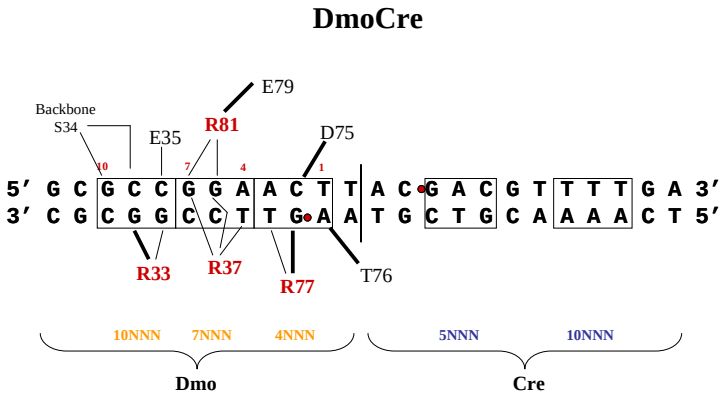

Figure 3.7: DNA targets and proteins sub-domains of meganucleases DmoCre

Recently, partner 1 was able to resolve the 3D structure of the homing endonuclease I-DmoI bound to its cognate DNA target. This structure revealed the protein domains interacting with the DNA. It appears that these domains of interaction were relatively conserved in the "Dmo part" of the Hybrid molecule DmoCre, and are in agreement with the 3 DNA-binding domains previously identified based on E-DreI structure (a similar hybrid molecule).

D1/*Generation of DmoCre mutant libraries*

The cleavage profile of DmoCre was determined on the 64 D(10NNN)-C1, D(7NNN)-C1 and D(4NNN)-C1 targets derived from the D2C1 target. The DmoCre protein is able to cleave 16 out of the 64 D(10NNN)-C1 targets, 10 D(7NNN)-C1 targets and only 6 D(4NNN)-C1 (Figure 3.8).

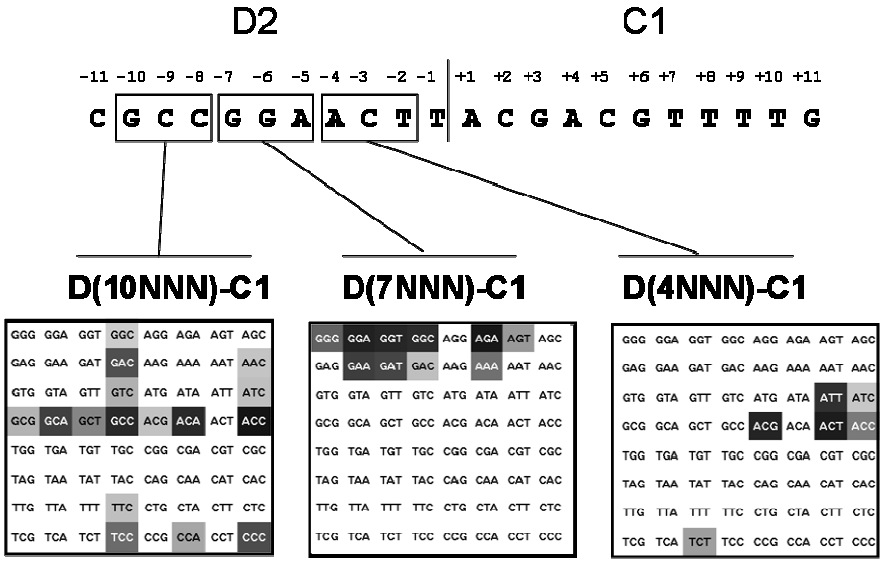

Figure 3.8: Cleavage profile of DmoCre against the three groups of degenerated targets

We tried to locally alter the DNA specificity of the I-DmoI moiety of the molecule.

In order to create a DmoCre library targeting the group of D(10NNN)-C1 targets, the residues Tyr29, Arg33 and Glu35 of the protein were randomized. 2232 clones, representing 28% of the library diversity, were screened against the 64 D(10NNN)-C1 targets. 387 active mutants with unique

sequences were thereby identified. Figure 3 a shows that the DmoCre protein is able to cleave 16 out of the 64 targets, five of which were very faintly cleaved, while the whole collection of newly produced DmoCre variants allowed cleavage in 50 out of 64 D(10NNN)-C1 targets (Figure 16). Moreover, with an average of 6 targets cleaved per variant, these engineered meganucleases appeared to retain a specificity equivalent, if not higher than, the DmoCre protein.

Using the same strategy, the residues Asp75, Thr76 and Arg77 of the protein were randomized in order to generate a DmoCre library targeting the DNA triplet at positions -4, -3 and -2 of D2-C1 (D(4NNN)-C1). The screening of the library (2232 clones) against the 64 D(4NNN)-C1 DNA targets yielded 570 active mutants with unique sequences. Six D(4NNN)-C1 targets were cleaved by the initial DmoCreV5 protein while 23 out of the 64 D(4NNN)-C1 DNA triplets were recognized by the new DmoCre variants (Figure 3.9). Furthermore these mutants appear highly specific since the average number of tolerated targets is 3 targets per protein. We noticed that the D(4ANN)-C1 sequences were preferentially targeted by the library (Figure 16). As an example of this engineering step, a mutant called S1 specific for the D(4AGT)-C1 target and carrying the D75R and R77V mutations was chosen and cloned into a mammalian expression vector for further characterizations. Analysis of the E-DreI and I-DmoI structures in complex with their DNA targets show that the residue Thr41 makes a Van der Waals contact with the methyl group of the complementary strand containing a thymidine at position -4. Therefore, we generated a new library randomized at residues Thr41, Asp75 and Arg77, while keeping the Thr76 fixed to limit the diversity of the library. Screening against the D(4NNN)-C1 targets allowed the isolation of 221 unique mutants which cleaved an average of 2 targets. In addition to the 35 D(4NNN)-C1 sequences recognized by at least one mutant, the randomization of the residue Thr41 allowed the cleavage of new D(4GNN)-C1 and D(4TNN)-C1 targets. Still, almost no cleavage for the D(4CNN)-C1 targets could be detected.

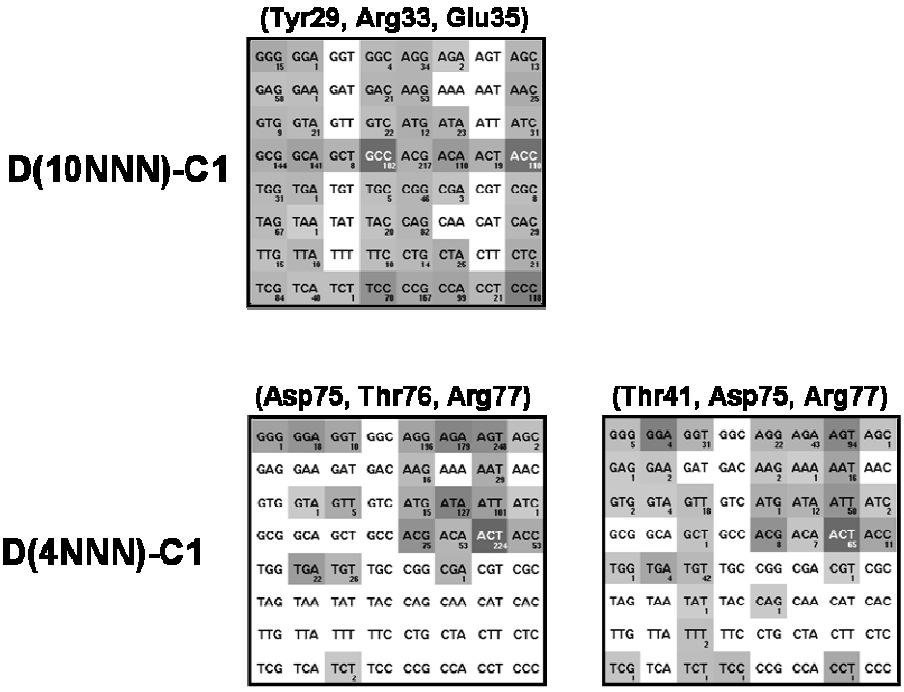

Figure 3.9: Hitmaps of three DmoCre mutant libraries that were screened in yeast

The analysis of the E-DreI and I-DmoI structures shows that Arg37 and Arg81 interact extensively with the nucleotides at positions -7, -6 and -5 of the D2-C1 target and appear resistant to engineering since mutations in any of the arginines led either to an inactive protein or to a protein with

much reduced activity (data not shown). However, the original DmoCre meganuclease is able to cleave 10 out of the 64 D(7NNN)-C1 DNA targets in our yeast assay (Figure 3.8)

D2/ *Combinatorial approach applied to Intra-domain: the Dmo part of DmoCre*.

Hundreds of DmoCre mutants recognizing the D(10NNN)-C1 or D(4NNN)-C1 targets have been obtained. We have previously described for the I-CreI protein, how to combine two mutation sets by a combinatorial approach to generate engineered I-CreI mutants. We then decided to apply this approach to the I-DmoI sub-domain of DmoCre. The two sets of mutations Tyr29, Arg33, Glu35 on one hand, and Asp75, Thr76, Arg77 on the other hand, constitute two distinct groups of mutations, which could therefore be combined into the same DmoCre coding sequence. Four D(10NNN)-C1 targets were selected (D(10CAG)-C1, D(10CCA)-C1, D(10CCG)-C1, D(10GCG)-C1), together with 24 corresponding DmoCre variants for each target. The same procedure was performed for three D(4NNN)-C1 targets where 3 sets of 24 DmoCre mutants recognizing respectively, D(4AGA)-C1, D(4AGT)-C1 and D(4ATA)-C1 DNA sequences, were chosen. Finally, the two groups of mutations were combined by *in vivo* cloning in yeast. Eight mutant libraries were created in such a manner and screened against their corresponding combined targets. None of the libraries enabled us to detect active proteins (suggesting a low intra-domain modularity within the I-DmoI moiety of the molecule. This result prompted us to adopt an alternative approach in which the two mutation sets would be introduced sequentially. Thus, a pool of eight DmoCre variants cleaving the D(4AGT)-C1 target was selected as a template to generate a mutant library randomized at residues Tyr29 and Arg33. The residue Glu35 was ignored in order to maintain the diversity of the library. The library was then screened against four combined targets (D(10GCG4AGT)-C1, D(10CCG4AGT)-C1 and D(10CAG4AGT)-C1, D(10CCA4AGT)-C1). No active protein cleaving the first three targets could be detected, but 13 unique proteins were able to cleave the D(10CCA4AGT)-C1 target, albeit with a low efficiency. Four mutants were further selected as templates in an error-prone PCR experiment in order to improve the cleavage activity. Although their endonuclease activities could be improved, these mutants did not reach the level of I-SceI activity in yeast, and no activity could be detected using our extrachromosomal assay in CHO-K1 cells (data not shown). Altogether, this data demonstrates the modularity of the I-DmoI domain of the DmoCre meganuclease, even though extensive engineering appears difficult.

D3/ *Combinatorial approach applied to Inter-domain: insertion of I-CreI mutants into the DmoCre scaffold*.

In addition to its monomeric architecture, the DmoCre scaffold offers the possibility to use I-CreI derived mutants that have already been obtained and characterized. We have already shown in previous reports that it was possible to introduce I-CreI mutants targeting the 10GTT-5CAG (Ca target) and 10TGG-5GAG (Cb target) DNA sequences within the DmoCre protein. Here we repeated this process using I-CreI mutants with different specificity and tested the activity of the resulting DmoCre variants against the adequate target using our yeast screening assay. As shown in table 3.1, this process of inserting I-CreI mutants into the DmoCre scaffold works with a good efficiency and success rate.

target's specificity of the I-CreI mutants (Cx)	Number and strength of the I-CreI mutants	Number of positive DmoCre mutants and maximal cleavage activity on D2-Cx target
5TTG	16 0.11 < AU < 0.30	16 AU_{max} = 0.56
5GAG	36 0.2< AU < 1.0	32 AU_{max} = 0.92
10CCA	14 0.15< AU < 0.70	12 AU_{max} = 0.92
10TGG-5GCT	26 0.61 < AU < 0.94	19 AU_{max} = 0.89
Ca: 10GTT-5CAG	33 0.4 < AU < 1.0	8 AU_{max} = 0.40; (0.91)*
Cb: 10TGG-5GAG	35 0.2 < AU < 0.74	6 AU_{max} = 0.48; (0.97)*

*: cleavage activity after protein optimization AU: arbitrary unit

Table 3.1: Cleavage activity of DmoCre variant with altered specificity within the I-CreI moiety.

We tried also to optimize the D2-Ca and D2-Cb cutters in order to increase their cleavage activities on their respective target. For each target, a corresponding pool of three active DmoCre mutants was used as a template in error-prone PCR, covering the entire length of the molecule. Two thousand two hundred and thirty two clones generated by *in vivo* cloning in yeast were then screened against the target of interest. For both D2-Ca and D2-Cb targets, highly active proteins could thus be obtained (Figure 17).

Two improved meganucleases, Va and Vb, specific to D2-Ca and D2-Cb respectively, were selected for further activity analyses in mammalian cells. The ORF's were further sub-cloned into a mammalian expression vector and their activity was determined in our extrachromosomal SSA assay in CHO-K1 cells. Figure 3.10 shows the activity of the original DmoCre meganuclease and the two selected DmoCre variants tested against all three targets (D2-C1, D2-Ca, and D2-Cb). Each meganuclease showed specific cleavage of its cognate target.

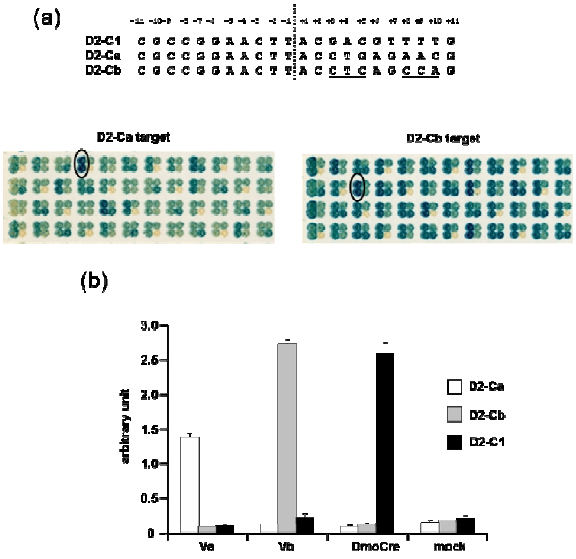

Figure 3.10: (a) Cleavage activity improvement against the D2-Ca and D2-Cb targets. The dots corresponding to the Va and Vb variants are circled. (b) Extrchromosomal SSA assay of the Va, Vb and DmoCre proteins against the D2-Ca, D2-Cb and D2-C1 targets.

D4/ *Combination of mutation sets present in both I-DmoI and I-CreI sub-domains*

To confirm the modular architecture of the DmoCre molecule, we assembled mutations affecting the specificity of both the I-DmoI and I-CreI moieties. The improved meganucleases Va and Vb, specific to D2-Ca and D2-Cb were selected and used as templates to generate I-DmoI libraries randomized at residues Tyr29, Arg33, Glu35, or, Asp75, Thr76, Arg77. The four resulting libraries were then screened in yeast against 18 randomly chosen hybrid targets. Table 3.2 summarizes the results, confirming that the modularity between the I-DmoI and I-CreI sub-domains as active DmoCre variants could be obtained against 13 out of the 18 tested targets.

Library template (target specificity)	**randomized positions**	**Target tested**	**Maximal cleavage activity observed in yeast (AU value)**
Vb (D2-Cb)	75, 76, 77	D(4AGT)-Cb	0.91
		D(4ATA)-Cb	0.57
		D(4AGA)-Cb	0.66
		D(4AGG)-Cb	0.50
	29, 33, 35	D(10CTC)-Cb	0.19
		D(10TAG)-Cb	0
		D(10GCG)-Cb	0.64
		D(10ACG)-Cb	0.64
		D(10CCA)-Cb	0.41
Va (D2-Ca)	75, 76, 77	D(4AGT)-Ca	0.87
		D(4ATA)-Ca	0
		D(4AGA)-Ca	0.34
		D(4AGG)-Ca	0.37
	29, 33, 35	D(10CTC)-Ca	0
		D(10TAG)-Ca	0
		D(10GCG)-Ca	0.43
		D(10ACG)-Ca	0
		D(10CCA)-Ca	0.2

Table 3.2: Cleavage activity in yeast of DmoCre variants with altered specificity within both I-DmoI and I-CreI moieties.

DmoCre variants, able to cleave the D(4AGT)-Cb [D(4AGT)-C(10TGG5GAG)] and D4AGT-Ca [D(4AGT)-C(10GTT5CAG)] targets, were further selected for *in vivo* activity characterization. For each target, a pool of three mutants, M1 to M3 and P1 to P3, showing good activity on their respective targets in yeast, were selected and their activities further improved. Upon random mutagenesis, three new optimized variants, OM1 to OM3, and OP1 to OP3, displaying high cleavage activities against D4(AGT)-Ca or D(4AGT)-Cb, respectively, were chosen and cloned into a mammalian expression vector in order to perform an extrachromosomal recombination assay in CHO-K1 cells. Figure 3.11 shows that all the DmoCre variants tested were active in CHO-K1 cells.However, the engineering step that shifts the protein specificity from the D2-Ca or D2-Cb target to D(4AGT)-Ca or D(4AGT)-Cb, respectively, also reduced, to a certain extent, the mutant cleavage activity which could be partially recovered after our optimization step. Importantly, the newly obtained DmoCre mutants are specific for their targets as they display no cleavage activity towards the D2-Ca or D2-Cb targets. Analysis of mutant sequences confirmed that activity improvements

resulted mainly in the addition of mutations in residues located in the I-CreI moiety of the DmoCre molecule or in the region linking the two sub-domains (residues 93 to 108).

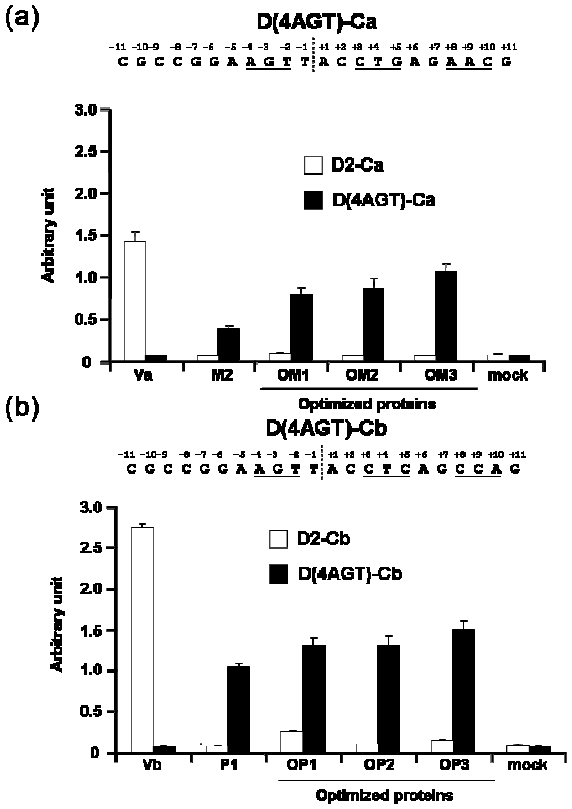

Figure 3.11: Extrachromosmal SSA assay of different mutants against the D2-CA and D(4AGT)-Ca targets (a), and the D2-Cb and D(4AGT)-Cb targets (b).

Finally, a set of DmoCre variants representative of each engineering step (Va, Vb, S1, OM3 and OP3) (Figure 3.12) were tested in a dose-response study (from 0.5 to 25 ng of meganuclease coding vectors) using an extrachromosomal assay in CHO-K1. We compared the cleavage activities of several engineered DmoCre variants specific to D2-Cb, D2Ca, D(4AGT)-C1, D(4AGT)-Ca and D(4AGT)-Cb respectively, as well as the parental hybrid DmoCre molecule and the natural meganuclease I-SceI against their cognate targets. Figure 19 shows that all of the mutants are active in mammalian cells, however, the DmoCre variants show less activity, compared to the initial DmoCre protein.

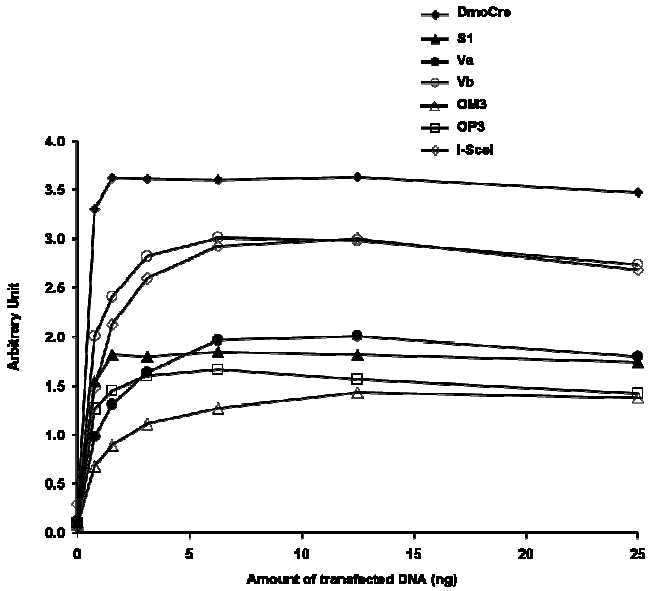

Figure 3.12: Extrachromosomal SSA assay in CHO cells of DmoCre, I-SceI and different engineered DmoCre based meganucleases.

D5/ *The engineered DmoCre mutants display a minimal toxicity*

The specificity/toxicity ratio is a crucial parameter in DSB-induced recombination technology. To evaluate the potential toxicity of engineered meganucleases we have developed cell survival assay in CHO-K1 cells. Briefly, cells are co-transfected with increasing amounts of the meganculease expression vector and a fixed amount of GFP coding vector. Six days after transfection, the percentage of cells still expressing GFP is measured and compared to the same percentage for cells transfected with an empty vector. Thus, the five representative engineered meganucleases described above (Va, Vb, S1, OM3 and OP3), along with the I-SceI and DmoCre meganucleases, were evaluated for their potential toxicity. Two meganucleases derived from I-CreI were added as control. The V2(G19S) / V3 (G19S) heterodimer is an inactive meganuclease while MegaX is a meganuclease with relaxed specificity. Figure 3.13 shows that the five engineered DmoCre mutants, as well as the initial DmoCre protein, do not display any visible toxicity and present the same profile as I-SceI. As expected, the inactive meganuclease shows a flat pattern, while expression of the non-specific MegaX induces signinficant toxicity.

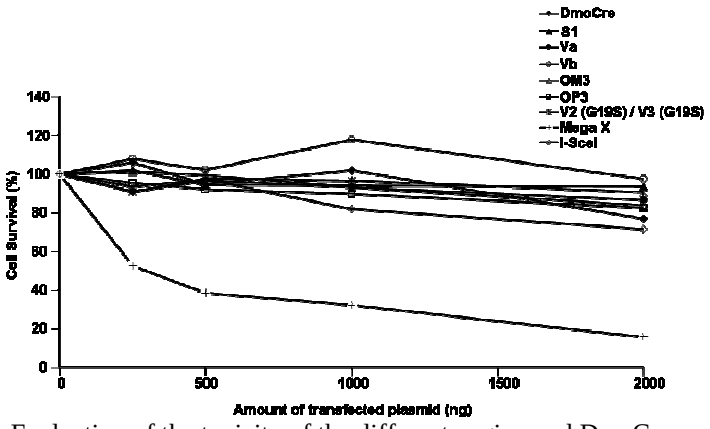

Figure 3.13: Evaluation of the toxicity of the different engineered DmoCre meganucleases

Toxicity might be related to the specificity profile of the endonuclease, which may generate off-site cleavage events. The characterization of the engineered DmoCre mutants was completed by monitoring the γ-H2AX foci content in CHO-K1 cells, expressing the endonucleases, since the γ-H2AX focus formation is one the first responses of the cell to DNA DSBs. CHO-K1 cells were transfected with 1 µg of the expression vector for the different meganucleases carrying a C-terminal HA epitope. At this high dose, no toxicity was revealed by the cell survival assay. Expression of DmoCre, as well as the five representative engineered DmoCre mutants, induces on average, approximately two to three γ-H2AX foci per transfected cells, similar to the background level (expression of the inactive V2 (G19S) / V3 (G19S) meganuclease) and to the number of foci induced by I-SceI (Figure 3.14).In contrast, the expression of the non-specific MegaX that displayed toxicity in the cell survival assay induces an average number of 12 γ-H2AX foci per transfected cell. All together, this data indicates that the modification of the protein specificity has not been achieved at the expense of toxicity.

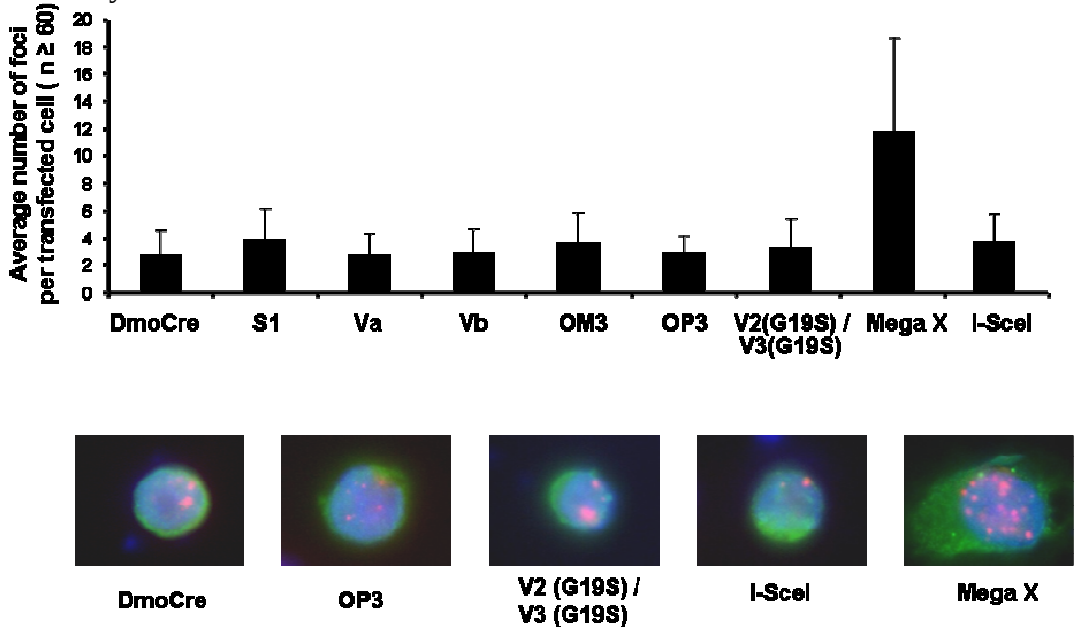

Figure 3.14: The potential genotoxicity of the engineered DmoCre meganucleases was monitored by visualizing the formation of γ-H2AX foci at DNA double-strand break. Green, HA labeling; red, γ-H2AX foci; and blue, DAPI staining

E/ I-CreI

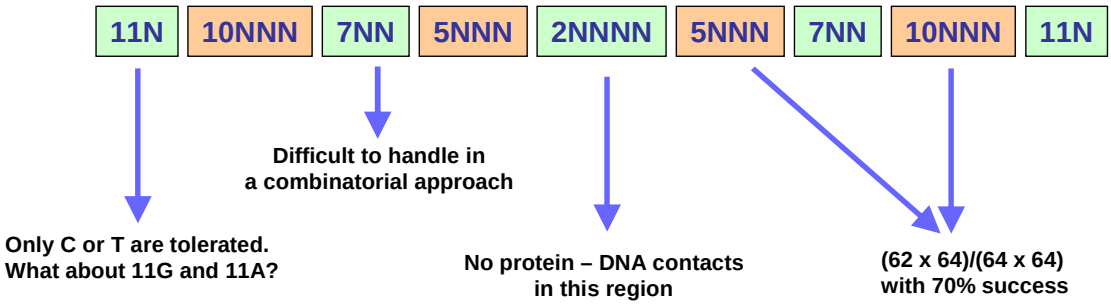

Figure 3.15: constraints and limitations of the I-CreI target

The combinatorial process using I-CreI protein with locally altered specificity involves the combination of 2 half of the I-CreI meganuclease binding respectively the 10NNN and the 5NNN nucleotide of the C1221 DNA target. Even though this approach has proven to be fast and very effective to produce meganucleases with tailored specificity, it has some constrains and limitations illustrated in Figure 3.15.

The database of locally altered I-CreI proteins contains protein mutants based on wild type I-CreI (D75) and I-CreI D75N. However the combination of D75N protein binding the 5NNN targets, with a D75 wild type protein, binding the 10NNN target, involves the introduction of the mutation D75N; and therefore can lead to a change in the expected specificity. Overall, this results in a decrease of the success rate of our combinatorial assembly process.

We have then produced I-CreI libraries mutated at position 44, 68, and 70 without the substitution D75N. The library was screened against the 64 5NNN_P I-CreI palindromic targets. 153 clones were identified as positives in the D75-based library called Ulib2D75. These clones cut collectively 39 targets out of the 64.

These cutters have been integrated into the list of modules available for the combinatorial process and have been used up to now into 26 combinatorial libraries. However, since many of these libraries are still in process, we have not been able to draw conclusions about the efficiency of these mutants compared to the mutants from the other older libraries.

Another direction was to try to extend the I-CreI target space to targets not limited to C or T at positions -11/+11. In the WT I-CreI, Serine at position 32 interacts with positions -11/+11 through a water molecule and might thus be responsible to the no recognition of A or G at positions -11/+11. Therefore we screened libraries resulted from randomization at 2 positions including position 32. These libraries were screened against 11GNNN_P and 11ANNN_P target collections. Also, new libraries were designed in which positions 32, 30, 33 and either 38 or 40 were randomized, with a limited subset at some positions to avoid too much complexity. Overall, 304 unique sequences were found cutting 11GNNN_P or 11ANNN_P targets, with 54 out of the 64 11GNNN_P targets cut and 47 out of the 64 for the 11ANNN_P targets. These new modules were integrated into the target search which now can lead to hits with A or G at position -11/+11. This lead to a 90% increase of the hit frequency along genomes. The real success rate of these new 11A/G targets remains to be monitored as they will be used in the next months.

In previous works, we found that it was difficult to produce locally altered I-CreI proteins able to cleave the 5ANN_P and 5NNA_P targets with a high efficiency. A close analysis of the 3D structure of the I-CreI meganuclease shows that the amino acid at position 24 could be responsible for the constraint on the 5ANN_P targets. Moreover the analysis of the I-CreI mutants available in our database, revealed that the highly active mutants against at least one of the 5ANN_P targets appears to have a Valine at position 24 while the wild type amino acid is a isoleucine. To verify this hypothesis, we have chosen a potential target carrying a 5ANN sequence and produced the 2 collections of meganucleases differing only by the amino acid at position 24. The comparison of the 2 collections shows that Valine at position 24 increases the number of cutters available against the 5ANN targets: 2 libraries targeting a palindromic target T**TTC**CT**ATC**GTACGATAGGAAA (“10TTC5ATC_P”). 35 mutants (with or without the mutation V24) targeting 5ATC were combined with 37 mutants targeting 10TTC. 2232 clones were tested in each cases. Only 2 weak cutters could be identified when the mutation V24 was not present although 39 clones positives appears positives when containing V24.

5NNA_P targets are particularly difficult to reach because of an unfavourable steric clash between the thymidine on the other strand and threonine at position 46. To abolish the steric clash we performed mutagenesis on the residue at position 46 of the I-Cre protein. However, Thr46 interacts with the amino acid at position 73 therefore attempts to change position 46 were successful only when

coupled to a change at position 73. Finally we selected two I-CreI derivates mutated at positions 46 and 73 (A46M73 or G46M73).

Three libraries based on either A46M73 or G46M73 scaffolds were generated by mutated positions 44, 68, 70 or 44, 68, 75 and 77 and were screened against the 5NNA_P targets. 97 unique mutants were identified (16 and 81 respectively). These new modules were added in the standard target search on sequences. The outcome of these modules will be examined in a few months.

The figure 3.15 shows how the I-CreI target was subdivided into different regions. In orange are depicted the regions for which we alter the specificity of I-CreI and in green the regions for which we rely on the natural I-CreI degeneracy. Using our current combinatorial process to engineer meganucleases, we estimate that we can target one DNA sequence every 300 bp in the human genome As the central 2NNNN region is very difficult to tackle because there is no DNA-protein contact in this region, to expand our meganucleases repertoire, we tried to find I-CreI mutants with an altered specificity toward the 7NN region. For that purpose, we built a mutant library of 2232 clones in yeast by randomizing residues at positions 26, 28 and 42 and screened it against the 16 palindromic 7NN targets. The screening yielded 250 mutants with unique sequences that cleave overall the 16 targets.

In a second step, like we usually do for the 10NNN and 5NNN mutants, we tried to combine these 7NN cutters with mutants specific for the 5NNN region. Taking mutants specific respectively for the 7TA, 7TT and 7GA targets (three targets that are not cleaved by I-CreI) and mutants cleaving respectively the 5TAC, 5CTC targets, we built three different combinatorial mutant libraries and screened them respectively against the 7TATAC, 7TTCTC and 7GACTC combined targets. Figure 3.16 shows that the screening yielded strong mutants against the three combined targets.

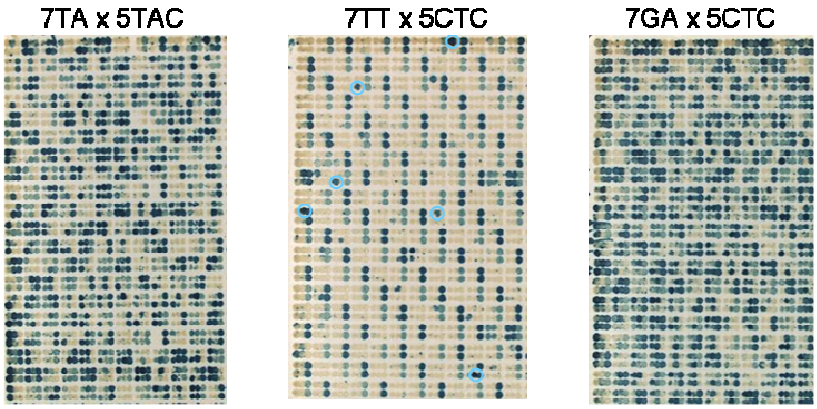

Figure 3.16: Primary screening of the three 7NNx5NNN combinatorial mutant libraries against respectively the 7TATAC, 7TTCTC and 7GACTC targets.

The last step of the meganuclease engineering process that consists in the introduction of the mutations responsible for the 10NNN specificity in a combined 7NNx5NNN I-CreI mutant is currently under evaluation.

Another important aspect of the meganuclease engineering is the level of activity of the engineered meganuclease at the end of the process. Dose – response studies with our extrachromosomal SSA assay have shown that the wild-type I-CreI protein is a better cutter than the N75 mutant. We would therefore try to keep constant the D75 during the engineering process. Analysis of the I-CreI structure shows that D75 interacts with R70, which itself interacts with the nucleaotide at position ±3. Keeping D75 constant implies therefore to look for R70-D75 I-CreI mutants that cleave targets with a cytosine at position 3. For that purpose, we built a mutant libray of

2232 clones in yeast randomized at position 44,68 and 77 and screened it against the 16 5NNC targets. The screening yielded 577 clones with a unique sequence, the 16 targets being cleaved.

However, if we limit the target search to DNA sequences that have a cytosine at position ±3, the hit frequency will increase by a 16 fold factor. This impact can be compensated by introducing the change in the specificity of the 7NN region. In fact, by combining the two aspects that have been developed above (7NN specificity and R70-D75 meganucleases), we built a mutant library of 4464 clones in yeast by randomizing the six residues 26, 28, 42, 44, 68 and 77 and screened it against the 256 7$(N)_4$C targets. This vast screening allows to spare the 7NNx5NNN combinatorial step. The primary screening yielded 428 mutants that cleave overall 150 7$(N)_4$C targets. Mutations responsible for the 10NNN specificity will now be introduced in these mutants.

F/ Improving the I-CreI scaffold

When using the I-CreI scaffold to generate new meganucleases with tailored specificity, we usually co-express 2 proteins in order to obtain heterodimerization of the monomers. However, we can not avoid the formation of the 2 respective homodimers. The presence in the cell of these 2 homodimers decreases the overall specificity of the custom made meganuclease and could sometime reveal to be toxic for cells. Therefore avoiding the homodimers formation could be a tremendous advantage. 2 approaches can be considered: we can fuse the 2 monomers to form a single chain molecule, and we have shown previously that it was possible without affecting the endonuclease activity of the protein; or we can mutate the protein-protein interface of the meganuclease to inhibit the formation of homodimers and to obtain an obligatory heterodimer. These 2 approaches are currently under investigation in collaboration with CRG.

<u>The obligate heterodimers</u>.
In previous work, we have produced a meganuclease RAG1 able to cleave a DNA sequence in the human RAG1 gene. The meganuclease RAG1 was used as a model.

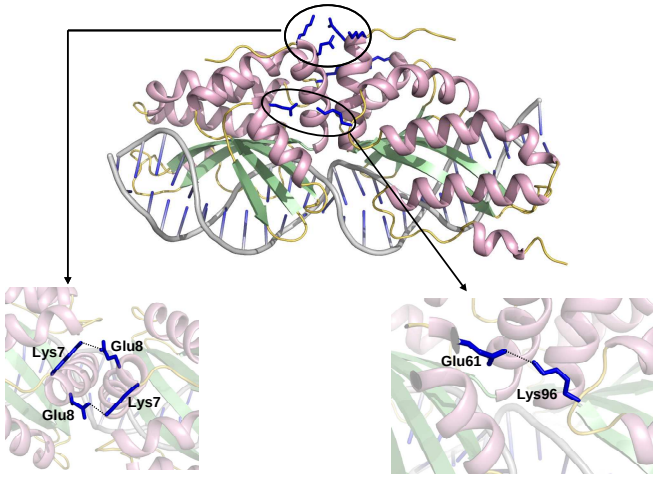

Figure 3.17: protein-protein interactions considered for the obligate heterodimer

The analysis of the 3D structure of I-CreI allowed us to identify 2 workable protein-protein interactions (Figure 3.17). These interactions concern the interaction between Lys7 and Glu8, and Glu61 and Lys96. The 4 possible double mutants (7-61, 8-61, 7-96, and 8-96) were produced. Their activity was tested in our extrachromosomic assay against the RAG1.10 target.

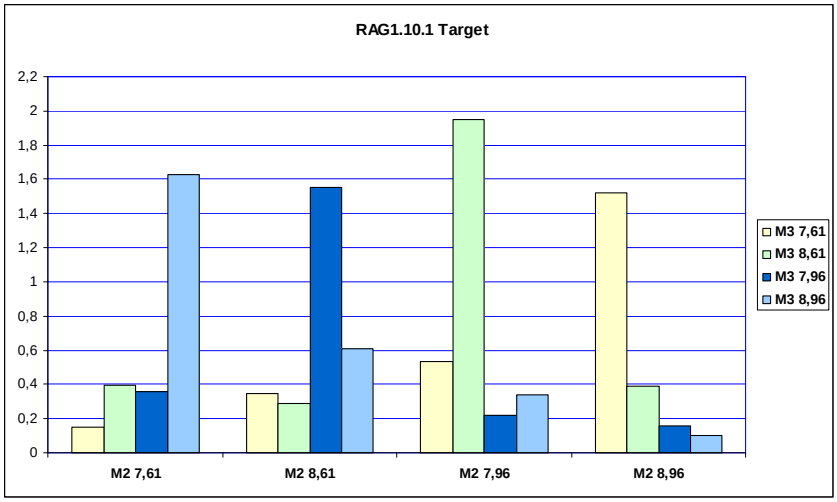

Figure 3.18: extrachromosomic assay of the obligate heterodimers against RAG1.10 DNA target

As illustrated in figure 3.18, the obligate heterodimer appears to be active only when the compensatory mutations are present in the 2 monomers. Furthermore, the activity of the respective homodimers is entirely abolished.

Single chain Molecule:

The meganuclease cleaving the RAG1.10 DNA target is once again our model. 9 differents linkers' peptides were used in order to fuse 2 monomers. Their activities were tested in yeast and in our mammalian extrachromosomic assay in CHO cells. The best candidate was chosen and its activity improved by introduction of the G19S mutation (previously shown to increase cleavage activity).

This final molecule appears to be highly active, and no homodimeric activity could be detected in extrachromosomic assy in CHO (Figure 3.19)

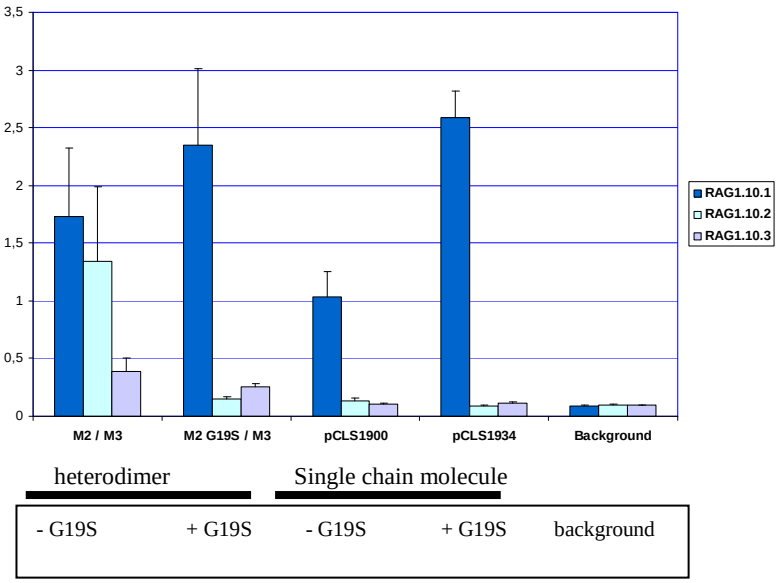

Figure 3.19: extrachromosomic assay in CHO cells. Cleavage activity was tested against the human target Rag1. 10.1 and its 2 palindromic derivates sequences

We also evaluated the potential toxicity of these new RAG meganucleases using the cell survival assay and the measurement of γH2AX foci that have been described above for the DmoCre protein.

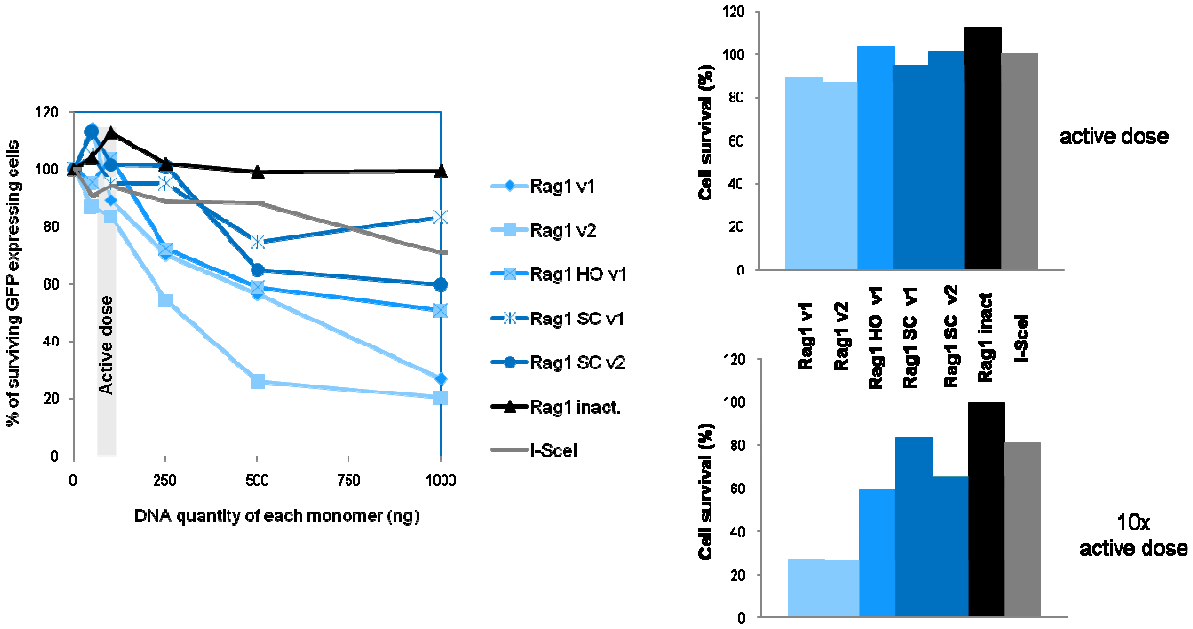

Figure 3.20: Toxicity of the different RAG meganucleases evaluated in a cell survival assay

Figure 3.20 shows the toxicity profile of the different RAG meganucleases in CHO-K1 cells. At the active dose (corresponding to the maximal activity in the gene targeting experiment), none of the meganucleases presents a significant toxicity, while at ten times the active dose the obligate heterodimer and the single chain, molecules are much less toxic than the RAG heterodimer. This characteristic was confirmed by measuring the content of γH2AX foci in cells transfected by the meganuclease expression vector (Figure 3.21)

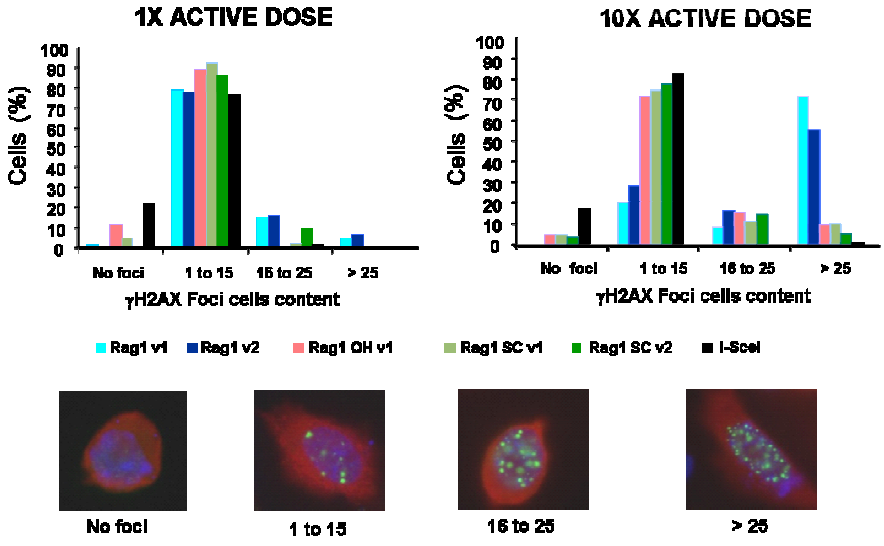

Figure 3.21: The potential genotoxicity of the different RAG meganucleases was monitored by visualizing the formation of γ-H2AX foci at DNA double-strand break. Red, HA labeling; green, γ-H2AX foci; and blue, DAPI staining

Again, at ten times the active dose, the obligate heterodimer and single chain molecule appear less toxic than the initial heterodimer.

As the single chain molecule facilitates the transfection issues by alleviating the need to co-transfect two monomers and presents a low toxicity pattern, it has become the scaffold of choice for undertaking genome engineering studies.

Deviations from the project workprogram

The original molecule DmoCre was active at 37°C in yeast and its activity in mammalian cells was limited. In order to generate a molecule with high activity in CHO cells, we have performed a random mutagenesis on the molecule and screened the library for highly active DmoCre. One variant was chosen for its activity in yeast and CHO cells.

Status of Deliverables and Milestones

Status of the deliverables for the period and forecast for the other deliverables of the WP.

Del. no.	Deliverable name	Original delivery date (**)	Revised delivery date (**)	Status (***)	Nature
D3.1	First set of libraries	Month 12		**finished**	**RP**
D3.2	Second set of libraries	Month 28		**finished**	**RP**
D3.3	1st set of novel non combinatorial meganucleases	Month 16		**finished**	**RP**
D3.4	2nd set of novel non combinatorial meganucleases	Month 32		**finished**	**RP**
D3.5	1st set of novel combinatorial meganucleases	Month 20		**finished**	**RP**
D3.6	2nd set of novel combinatorial meganucleases	Month 36		**finished**	**RP**
D3.7	MEGATOOLBASE available	Month 36		**finished**	**R**

Status of the milestones for the period and forecast for the other milestones of the WP.

Milest. no.	Description	Original target date (**)	Revised target date (**)	Status (***)
M3.1	Success of libraries designed	Month16		**Finished**
M3.2	Success of combinatorial	Month20		**Finished**
M3.3	Approach validated by MEGATOOLBASE	Month36		**finished**

Workpackage 4: Characterization and standartization of newly isolated meganucleases

Lead Partner: Fermentas (Arvydas Lubys)
People involved: Arvydas Lubys
Julius Gagilas
Jolanta Vitkutė
Nomeda Lauciūnienė
Lelija Trinkūnaitė
Arūnas Leipus
.

Workpackage objectives and starting point of work at beginning of reporting period

The workpackage objectives are: a) identification of candidates for protein isolation; b) purification of selected meganucleases; c) investigation of enzymatic properties of purified enzymes; d) investigation of meganucleases stability in purified preparations.

Progress towards objectives

The goal of the workpackage 4 is to purify selected meganucleases using conventional protein purification approaches and to investigate their enzymatic properties as well as stabilities during prolonged storage. At first meganucleases Rosav3-Rosav4, Rosav3, Rosav4, I-CreI and I-CreI N75 have been purified near to homogeneity and their samples in different storage buffers were kept for stability testing. Later several plasmids coding for new variants of meganucleases were received from Cellectis. Four RAG1-specific mutants: single chain M3-M2G19S (encoded by pCLS 2050), single chain M3--M2+G19S (pCLS 2216), heterodimer M2G19S/M3 (pCLS 2080), heterodimer M2+G19S/M3- (pCLS 1947) and four GS-specific mutants: heterodimer GS3/GS4 (pCLS 2148), heterodimer GS4/GS3 (pCLS 2149), mutant GS3 (pCLS 2150), mutant GS4 (pCLS 2151) were selected as candidates for purification and investigation of enzymatic properties.

Activity of Rosav3 and Rosav4 enzymes was investigated using either linearized (Figure4.1, A) or supercoiled (Figure 4.1, B) DNA substrates and applying increasing amounts of purified proteins.

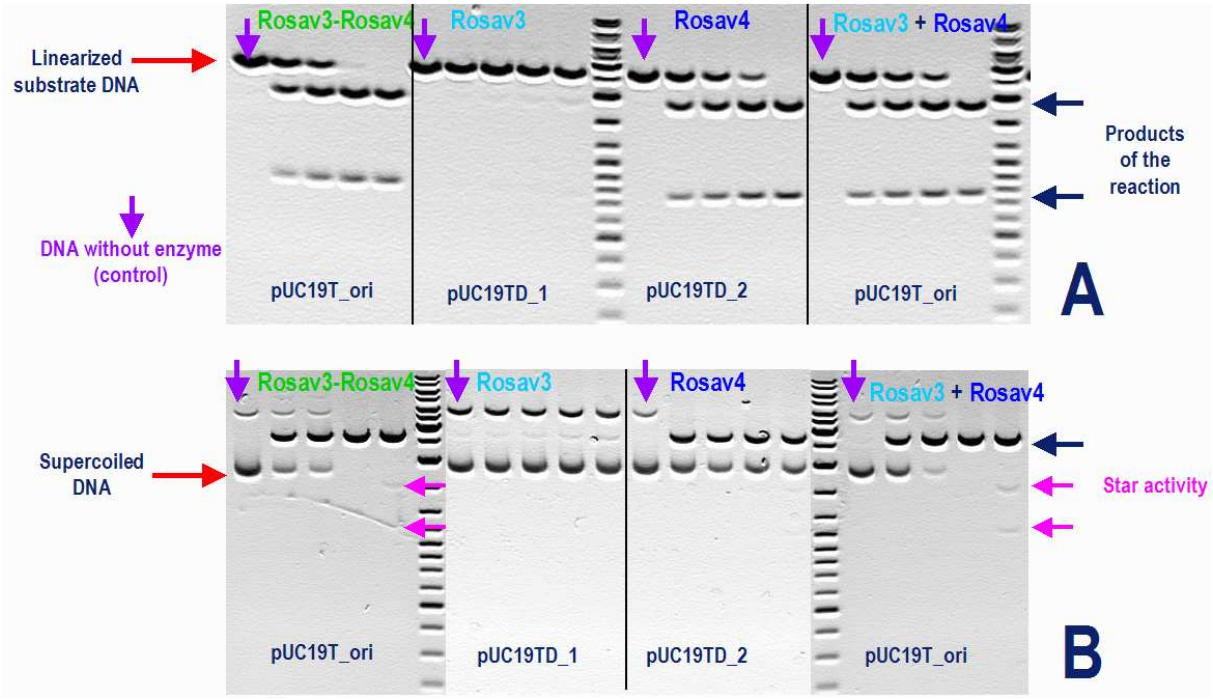

Figure 4.1. Activity comparison of heterodimeric meganuclease Rosav3-Rosav4, homodimeric meganucleases Rosav3 and Rosav4 and equimolar mixture of Rosav3 and Rosav4 (Rosav3 + Rosav4) using linearized (A) or supercoiled (B) DNA substrates. Abbreviations of plasmids used for the assay are indicated below the electrophoregram. Increasing amounts of enzymes were used. Reactions were incubated for 1 hour at 37°C in buffer: 33mM Tris-acetate (pH 7.9 at 37°C) 10mM Mg-acetate, 66mM K-acetate, 0.1mg/ml BSA.

Figure 4.1 shows that the activity of Rosav3 is very low on both substrates. In contrast, Rosav4 cleaved linearized DNA completely (at highest enzyme's concentration), while cleavage of supercoiled DNA substrate was partial. Unexpectedly, the equimolar mixture of Rosav3 and Rosav4 (Rosav3 + Rosav4) digested DNA of plasmid pUCT19_ori which carries the recognition target of heterodimeric meganuclease Rosav3-Rosav4 nearly as efficiently as the latter one. In addition, the mixture showed star activity on supercoiled DNA substrate like heterodimeric meganuclease.
Nonspecific cleavage was evaluated by incubation of meganucleases for 16 hours at 37°C with plasmid DNA which contains no recognition sequences (pGPS3 and pUC19) and with pUC19 derivatives possessing one recognition target (Figure 4.2) .

Figure 4.2 shows that:

1. Rosav3 cleaves DNA in case only if it contains the recognition sequence (pUC19TD_1 DNA). However, the cleavage of substrate DNA was not complete even after 16 h of incubation.
2. Rosav4 cleaves its own substrate (pUC19TD_2) with the same efficiency as the substrate of Rosav3-Rosav4 pUC19T_ori. It should be noted that two DNA fragments of similar size are generated in both cases, whereas only pUC19TD_2 has the unique target for Rosav4.
3. Heterodimeric meganuclease Rosav3-Rosav4 shows very high star activity on all DNA substrates tested. Similar level of star activity and nearly identical pattern of generated DNA fragments is observed using the equimolar mixture of Rosav3 and Rosav4 enzymes. These

data (and also data shown in Figure 4.1) suggest that during reaction heterodimeric enzyme may be formed by dimerization of monomers of Rosav3 and Rosav4 enzymes.

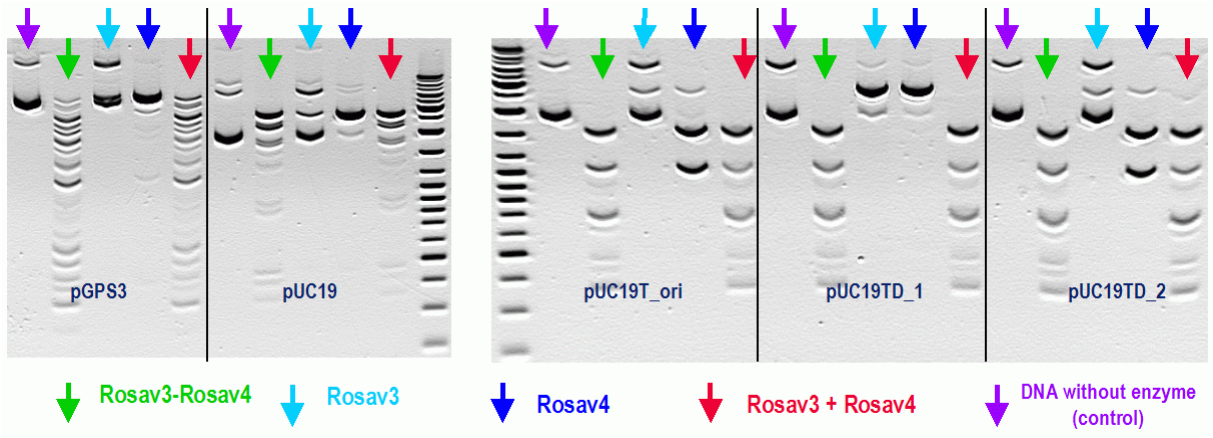

Figure 4. 2. Nonspecific activity of meganucleases on different DNA substrates.

Heterodimers GS3/GS4 and GS4/GS3 were purified near homogeneity using successive chromatography steps on IMAC Sepharose™ HP, *Strep*•Tactin® SP, Phosphocellulose P11 and ANX sepharose columns, mutants GS3 and GS4 - using chromatography on IMAC Sepharose™ HP and Phosphocellulose P11. Single chain meganucleases M3-M2G19S and M3--M2+G19S were partially purified from inclusion bodies - dissolving in urea followed by fast dilution for renaturation was used. We observed that partially purified preparations of these meganucleases completely lost their activities after one week storage at +4oC in buffer without DTT, but not in the presence of DTT. Necessity of DTT as stabilizing agent excluded the possibility to purify single chain meganucleases using their His6 affinity tags. Therefore, five ion exchange chromatography steps were used for purification of aforementioned enzymes to nearly homogeneity. Similarly, the activity of obligatory heterodimer M2+G19S/M3- was also lost during purification in case if DTT wasn't used, and this again excluded the possibility to use affinity chromatography on IMAC Sepharose™ HP. In this case four chromatography steps were required to purify to near homogeneity, but very low yield was observed – (0.01 mg per g of biomass). The yield of other purified meganucleases was in the range between 0.5 and 1.5 mg per g of biomass. Purity of all enzymes exceeded 97% according to SDS-PAGE data. Samples of each enzyme were kept in several different storage buffers in order to investigate their long term stability.

To improve the specificity of RAG1-specific mutants, Cellectis have developed two protein engineering strategies: the obligatory heterodimer design and the single chain molecule. Comparison of activities of obligatory heterodimeric meganuclease M2+G19S/M3-, heterodimer M2G19S/M3 with activities of corresponding single chain meganucleases M3--M2+G19S and M3-M2G19S, all cleaving the same target, are shown in figure 4.3.

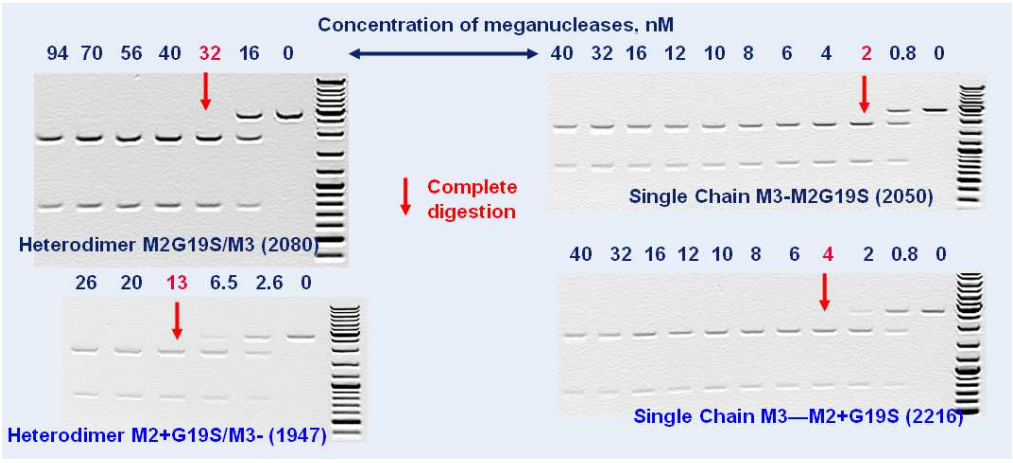

Figure 4.3. Comparison of activities of obligatory heterodimeric meganuclease M2+G19S/M3-, heterodimer M2G19S/M3 with activities of corresponding single chain meganucleases M3--M2+G19S and M3-M2G19S. Different concentrations of tested meganucleases were used for 1 h incubation at 37°C in buffer: 10mM TRIS-HCl (pH 8.0, 25°C), 10mM $MgCl_2$, 50mM NaCl. Target (PdmI linearized substrate DNA) concentration was 2nM in 25 µl final volume.

Figure 4.3 shows that considerably higher concentrations of heterodimeric meganucleases were needed for complete digestion of linearized substrate DNA compared to corresponding single chain meganucleases.

We found that activities of RAG1-specific meganucleases were considerably higher in case when DTT was added either to the reaction mixture or to the storage buffer. For instance, single chain M3-M2G19S meganuclease prepared from soluble fraction completely lost the activity after 1 month of storage, but addition of DTT partially restored the activity of this meganuclease (see figure 4.4). This observation suggests that inactivation of these enzymes during storage may be reversible.

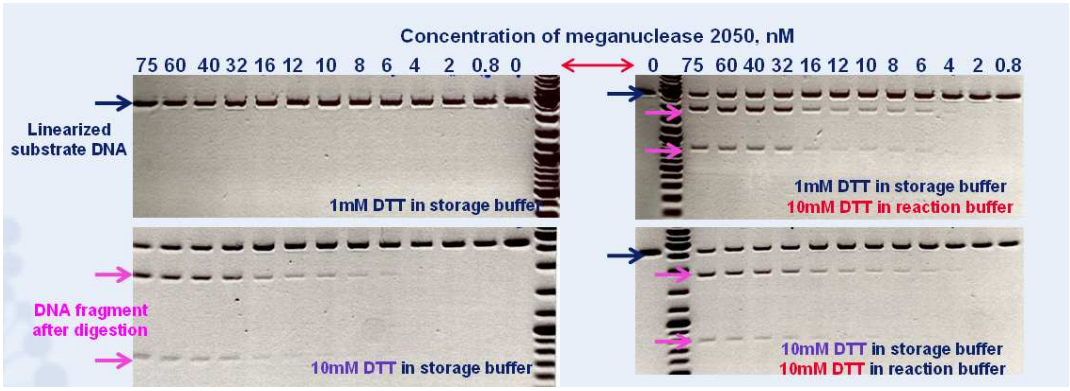

Figure 4.4. Reactivation of inactive single chain heterodimer M3-M2G19S by DTT. Different concentrations of meganuclease were used for 1 h incubation at 37°C in buffer: 10mM TRIS-HCl (pH 8.0, 25°C), 10mM $MgCl_2$, 50mM NaCl. Target (PdmI linearized substrate DNA) concentration was 2nM in 25 µl final volume.

Of note, the activity of remaining meganucleases is not influenced by DTT.

In order to compare enzymatic properties of meganucleases, their concentrations (nM) sufficient for complete cleavage of 2nM of substrate DNA during either 1h or 16h reaction were

determined. We found that an excess of some tested enzymes digested DNA at noncognate targets, i.e. showed so called "star" activity. The comparison of specific and nonspecific activities of homodimeric, heterodimeric and single chain meganucleases is presented in table 4.1.

Meganuclease	conc. nM, 1h	conc. nM, 16 h	Remarks
I-CreI	4	4	DNA cleavage was impaired at ten-fold or higher excess of enzyme. No star activity (up to 100 nM conc. of meganuclease/16h)
I-CreIN75	4	2	No star activity (up to 110 nM conc. of meganuclease/16h)
Rosav3-Rosav4	2	0.8	Star activity at 12 nM conc. of meganuclease (1h)
Mutant GS4 (2151)	4	1	No star activity (up to 100 nM conc. of meganuclease/16h)
Mutant GS3 (2150)	2	4	Star activity at 4 nM conc. of meganuclease (1h)
Heterodimer GS3/GS4 (2148)	4	2	Star activity at 32nM (1h) or 6 nM (16h) conc. of meganuclease
Heterodimer M2+G19S/M3- (1947)	13	6.5	No star activity (up to 26 nM conc. of meganuclease/16h)
Single Chain M3- -M2+G19S (2216)	4	2	No star activity (up to 160 nM conc. of meganuclease/16h)
Heterodimer M2G19S/M3 (2080)	32	16	No star activity (up to 94 nM conc. of meganuclease/16h)
Single Chain M3-M2G19S (2050)	2	2	No star activity (up to 120 nM conc. of meganuclease/16h)

Table 4.1. Concentration of meganucleases sufficient for complete cleavage of substrate DNA (2nM) after incubating either 1 or 16h at 37oC in buffer: 10mM TRIS-HCl (pH 8.0, 25°C), 10mM MgCl2, 50mM NaCl.

Mutant GS4 and single chain meganucleases M3-M2G19S and M3--M2+G19S showed higher specific activity and no star activity compared to other tested meganucleases.

The stabilities of meganucleases during storage at -20^{o}C in different storage buffers (with 50 % glycerol) were determined. The results of testing the stability of purified meganucleases in selected buffers are presented in table 4.2.

Meganuclease	Storage buffer	Stability, months	Remarks
I-CreI	A	9	
I-CreIN75	A	13	
Heterodimer Rosav3-Rosav4	A	19	
Rosav4	A	17	
Heterodimer GS3/GS4 (2148)	C	16	
Mutant GS4 (2151)	B	10	
Heterodimer M2G19S/M3 (2080)	C	3	DTT up to 10 mM was added
Single Chain M3-M2G19S (2050)	B	4	DTT up to 10 mM was added
Single Chain M3- -M2+G19S (2216)	B	2	DTT up to 10 mM was added After 5 months 50% activity was lost
Heterodimer M2+G19S/M3- (1947)	B	-	DTT up to 5 mM was added was not stable after 1 month

Storage buffer A
25mM Tris-HCl pH 8,0
20mM NaCl
0,2 % TX-100
1mM DTT
0,1mM EDTA
50 % glycerol

Storage buffer B
25mM Tris-HCl pH 8,0
100mM NaCl
0,2 % TX-100
1mM DTT
0,1mM EDTA
50 % glycerol

Storage buffer C
25mM Tris-HCl pH 8,0
200mM NaCl
0,2 % TX-100
1mM DTT
0,1mM EDTA
50 % glycerol

Table 4.2. Stability of tested meganucleases.

Homodimeric meganucleases Rosav4, I-CreI, I-CreIN75, GS4, heterodimers GS3/GS4, M2G19S/M3, Rosav3-Rosav4 and single chain meganuclease M3-M2G19S were stable in purified preparations in selected storage buffers during storage at -20^{o}C.

Deviations from the project workprogram

No corrective actions during this period

Status of Deliverables and Milestones

Table 1: Deliverables List

List all deliverables, giving date of submission and any proposed revision to plans.

Del. No.	Deliverable name	Original delivery date	Revised delivery date	Status	Nature
D4.1	Meganucleases for purification experiments identified	Month 20		finished	R
D4.2	Meganucleases purified	Month 20		finished	RP
D4.2	Properties of meganucleases determined	Month 36		finished	R
D4.3	Stability of meganucleases determined	Month 36		finished	R

Table 2: Milestones List

List all milestones, giving date of achievement and any proposed revision to plans.

Milestone no.	Milestone name	Workpackage no.	Date due	Actual/Forecast delivery date	status
M4.1	Decision regarding the first mutant to be purified	WP4	Month 16	Month 13	finished
M4.2	Decision regarding the second mutant to be purified	WP4	Month 20	Month 20	finished

Workpackage 5: In vivo testing of specific DSBR induced recombination in mouse cell

Lead Partner: **CELLECTIS (Fréderic Pâques)**
People involved: P Duchateau, A Gouble, C Jacqmarcq, V Guyot, A, Duclert, JM Filhol, E Gamelas Magalhaes, S Leduc, J Mikolajczak, S Thomas (CELLECTIS)

Workpackage objectives and starting point of work at beginning of reporting period

The aim of this workpackage is to validate the Megatools database. For this purpose, we useduse the novel proteins issue from the Megatools program to produce custom-made meganucleases specially designed to engineer rodent genome. 4 loci would have to be identified.

During this reporting period, we have performed targets search and identified potential targets on the mouse loci ROSA26 and glutamate synthetase, as well as on the CHO locus HPRT. Heterodimeric meganucleases were engineered. The initial cleavage activity of the meganucleases targeting was improved and tested for cleavage in mammalian cells, using a previously described assay based on transient transfection in CHO cells.

In parallel, using a meganuclease targeting the human gene RAG1 as a model, we developed several tools and methods for monitoring gene targeting efficiency. Eventhough meganucleases are known to be highly specific; monitoring potential toxicity is a major issue for some applications. We then developed assays to quantify any potential adverse effects of the meganuclease into the cells.

Progress towards objectives

During this period a major effort has been devoted to develop tools to monitor gene targeting events and potential toxicity of the meganucleases expressed in to cells.
We use our well characterized meganuclease engineered to target the human gene Rag as a model.

1/ We developed rapid assays to monitor the activity of the meganuclease at the endogenous site.

Targeted insertion:

First, in order to test a high number of transfected cells, we developed a robust PCR-screen strategy allowing identifying cells where recombination events took place. We showed that meganucleases can induce high frequency of targeted insertion making selection strategies dispensable (Figure 5.1).

In human cells, meganuclease can induce recombination events at frequencies up to 3 to 6%.

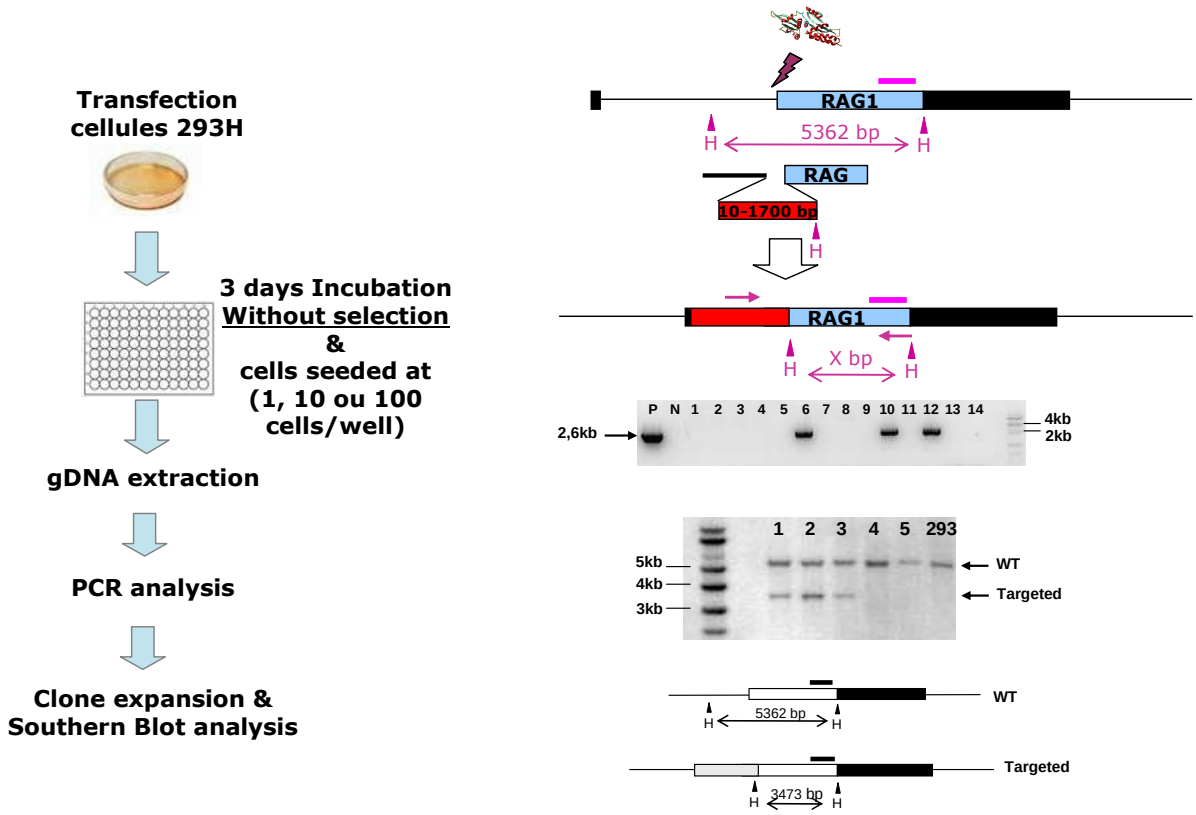

Figure 5.1: Targeted integration at the endogenous *RAG1* locus, driven by a single-chain meganuclease. Experimental outline and diagram of the gene targeting strategy used at the endogenous *RAG1* locus. The *RAG1* target sequence is located just upstream of exon 2 coding for the Rag1 protein. Exon 2 is boxed, with the open reading frame in white. Human 293 cells transfected with meganuclease expression plasmids and a repair substrate were cultured without selection for 72 hours and were then dispensed into 96-well plates at a concentration of 1, 10 or 100 cells/well. Cleavage of the native *RAG1* gene by the meganuclease yields a substrate for homologous recombination, which may use the repair matrix containing 1.7 kb of exogenous DNA flanked by homology arms as a repair matrix. Targeted integration events can be detected by PCR amplification, using a primer specific for the integrated exogenous DNA sequence and a primer present in the *RAG1* gene but outside the region of homology used in the donor matrix (arrows). Example of PCR screens for targeted integration events. PCR analysis of 14 different samples (1-14), each derived from 10 cells transfected with meganuclease expression plasmid and repair matrix. P indicates positive control (a recombinant plasmid), and N is the negative control (no DNA). Analysis of thegenomic locus was then perfomed bySouthern blot analysis: Genomic DNA preparations were digested with *Hin*dIII and Southern blotting was performed with a fragment of the *RAG1* gene lying outside the right homology arm. The locus maps indicate the restriction pattern of the wild-type locus (5.3 kb) and the targeted locus (3.4 kb). The probe is indicated by a solid black box. Five clones (1-5) samples derived from single transfected cells are analyzed, together with DNA from non transfected cells (293).In three samples, one of the alleles has been targeted. H, *Hin*dIII site.

Induced-mutagenesis:

We also developed a rapid strategy to test the meganuclease activity at the endogenous site. 2 DNA repair pathways can be involved after DNA DSB, the most predominant being the non-homologous end joining. Usually, NHEJ leads to perfect junction of the 2 broken ends of DNA. However, minor events as insertions or deletions of DNA fragments can be observed at the DSB. Thus we developed a rapid sequencing strategy allowing the detection of such events. In human cells, the transfection of the meganuclease targeting the RAG1 locus could induce up to 6% of mutagenesis at the endogenous site.

2/ Toxicity is a major issue in Double-strand break-induced gene targeting as it could be related to specificity and off-site cleavages. We then developed several assays in order to monitor potential toxicity of expressed meganucleases into cells.

We first developed a survival assay in CHO-K1 cells and were able to show that, at the active dose (minimal dose to induce maximal activity into the cells), none of the meganucleases tested affected the survival rate of the transfected cells, showing the high degree of specificity of these proteins. At higher dose, the single-chain design appears to improve the safety of the meganucleases as they can compare to I-SceI, the gold standard in the field. (figure 5.2).

Same conclusions could be obtained with the quantification of the γH2AX foci formation, either by classical quantification of foci number within the nucleus or by quantification by FACS of the γH2AX expression

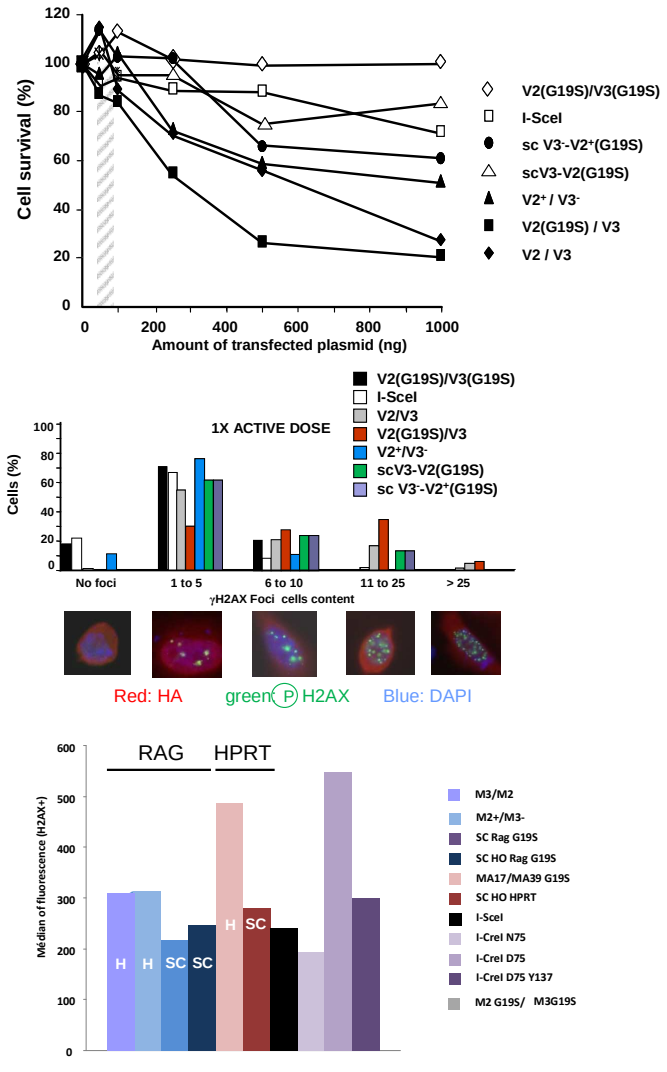

Figure 5.2: evaluation of the potential toxicity of the meganucleases:
3 different assays were developed. Top: cell survival assay. Various amounts of meganuclease expression vector and a constant amount of plasmid encoding GFP were used to cotransfect CHO-K1 cells. Cell survival is expressed as the percentage of cells expressing GFP six days after transfection. Dashed box, meganucleases active dose . The totally inactive V2(G19S)/V3(G19S) heterodimer is shown as a control for non toxicity. Middle: DNA damage was also visualized by the formation of H2AX foci at DNA double-strand breaks, and nuclear foci content was detremined. Representative images of cells treated with 10 times the active dose of meganuclease. Bottom: cells were transfected with various meganucleases and γH2AX expression was monitor by FACS.

Finally, we were able to induce gene targeting events into cells after profection of purified meganuclease. The use of purified protein would increase the safety of this technology as random integration of the plasmid expressing meganuclease is avoided (Figure5.3). 99% of the clones obtained after profection of the I-CreI protein fused to a cell penetrating peptide, showed perfect integration events.

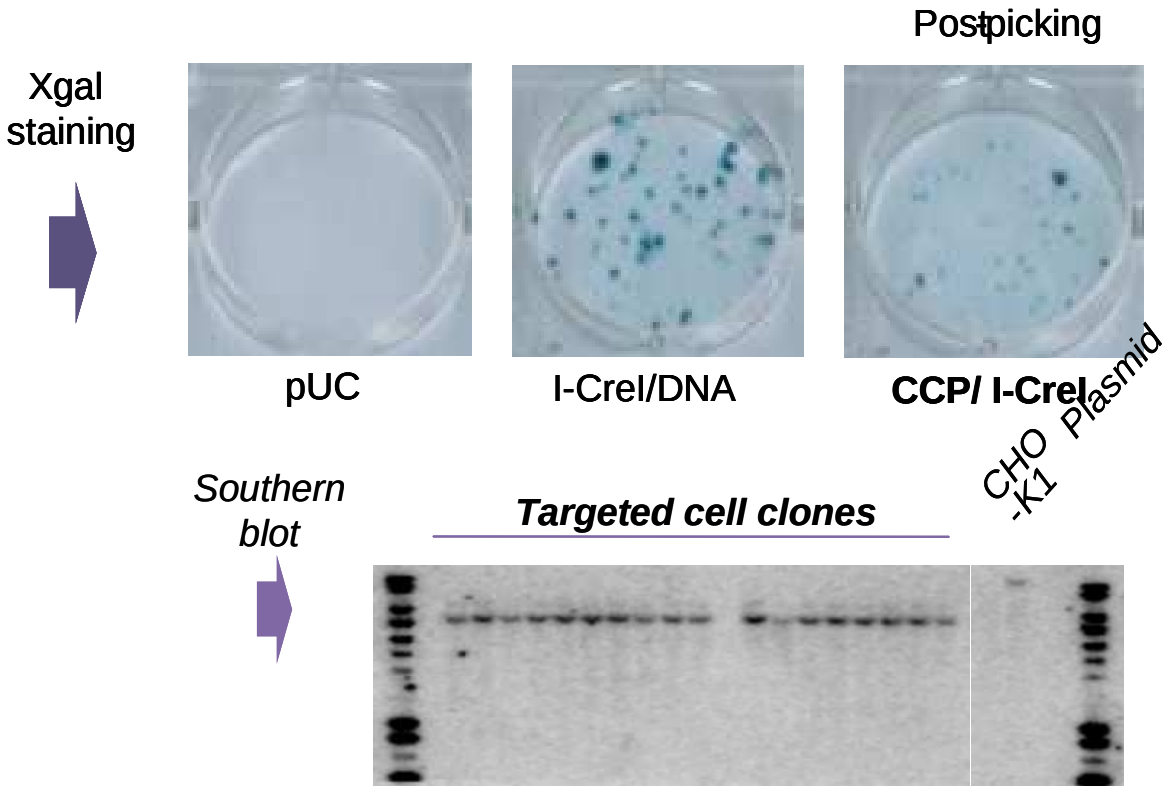

Figure5.3 : Profection of meganuclease induces recombination events in a chromosomal reportrer system. CHO-K1 cells carrying one chromosomal copy of the LACZ gene interrupted by a I-CreI site were either transfected with a I-CreI expressing plasmid (I-CreI/DNA), or profected with I-CreI protein fused to a cell penetrated peptide (CCP/I-CreI). Cells were stained with X-Gal 3 days after the experiment. Clones obtained after profection were picked and analyzed by Sourthern blotting.

In vivo testing of meganucleases targeting rodent genes:

1/ mouse ROSA26 gene

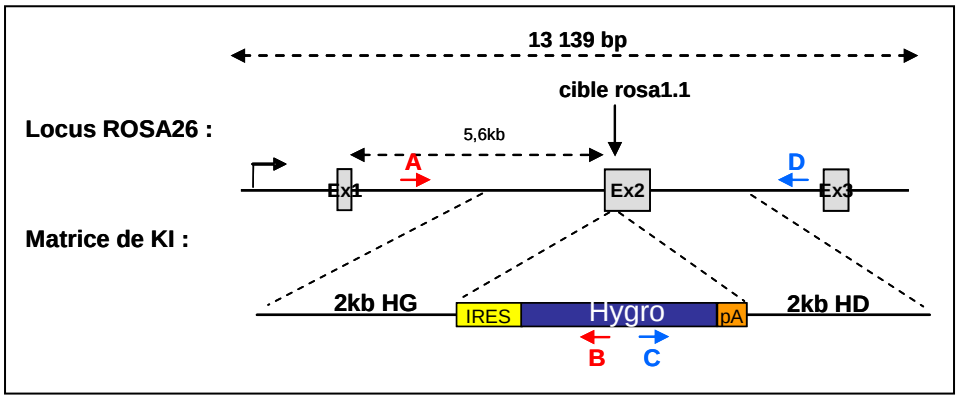

Figure 5.4: Rosa 26 gene organisation or genome engineering strategy..
The engineered meganuclease recognizes and cleaves a DNA sequence located in exon2 (fig 5.1). The DNA matrix used during the meganuclease induced Recombination experiments contains an expression cassette of the hygromycine resistance (hygro^{R}) gene preceded by a ribosome Binding site sequence (IRES).

During the previous reporting period, we showed that gene insertion events could be detected in occurred in about 1 out of 1000 transfected mouse L cells.

During the course of this project, we showed that I-CreI based single-chain meganuclease improved the safety of the meganuclease. Thus, a single chain version of the meganuclease targeting the ROSA26 gene was designed and new gene targeting experiments were performed in NIH3T3 cells. After co-transfection of the meganuclease and the DNA repair matrix, depicted in figure 5.4, genomic DNA issue from Hygromycin resistant clones were subject to southern blot analysis. 4/30 clones showed targeted integration. However, these events were accompanied by random insertion of the repair matrix.

New set of experiments were carried out without selection. No integration events could be detected. Chromatin conformation can impact the efficiency of the process as the recognition site can be poorly accessible by the meganuclease. The mouse ROSA26 is a valuable locus as the gene is ubiquitously expressed and therefore could represent a good target for transgenic or protein production studies. Thus, we are pursuing our effort to create new research tools for functional genomics in mouse, and decided to produce new meganucleases targeting different sites in the same locus. 4 sites have been identified and the meganucleases are now in the production process.

2/ mouse GS gene

During the last period, using the Cellectis plateform, we were able to produce heterodimeric proteins able to cleave a target located in exon 2 of the mouse GS gene. Activity of the engineered meganucleases has been optimized. Finally, highly active single chain meganucleases in mammalian cells could be produced.

This gene is of potential interest as in absence of glutamine, GS is essential for cell survival (fig 5.5). Thus, it can be used as selectable marker. It is notice that the DNA target sequence also present in the Chinese Hamster ovary cells genome. Thus a unique meganuclease has the potential application in 2 rodent's species.

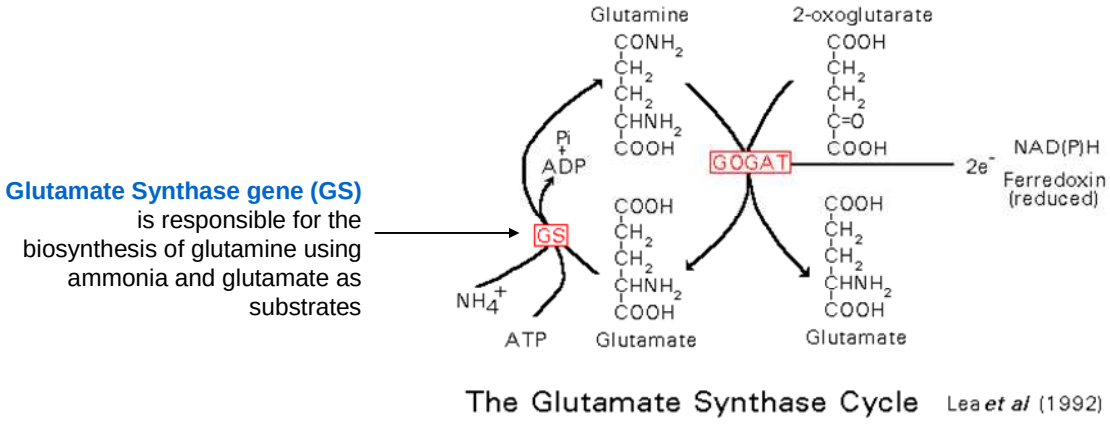

Figure 5.5: glutamate Synthase cycle. In the absence of glutamine, GS is essential for the survival of the cell. In cells that express little or no GS (example: mouse myeloma cells), the GS gene can be used as a selectable marker.

During the previous period, we were able to produce meganuclease presenting the high activity in mammalian cells. We have then designed active single-chain molecules carrying engineered protein-protein interface. As shown in figure 5.6, this new design improves the safety as activities of the homodimers, resulting of co-expression of both monomers, are abolished.

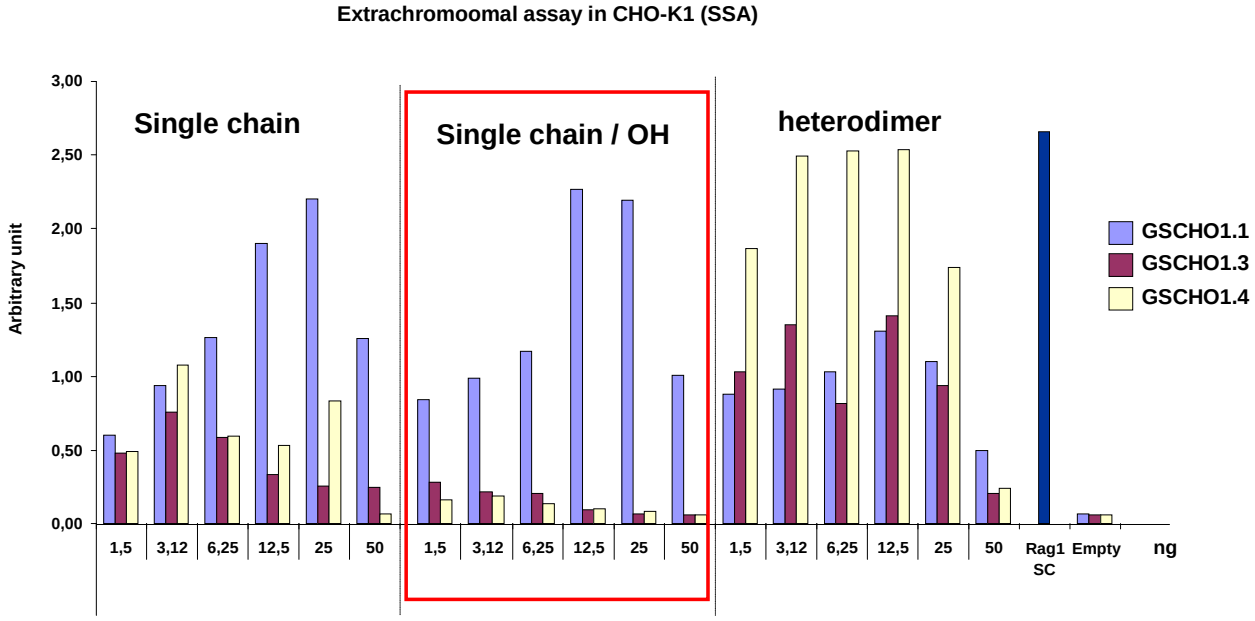

Figure 5.6 . Cleavage activity in mammalian cells of the meganucleases targeting GS gene. Activities of different meganuclease designs tested with our SSA extra-chromosomic assay in CHO-K1 cells. Meganucleases were tested against the non palindromic target found in the genome (GSCHO1.1), as well as against the 2 derived palindromic targets (GSCHO1.3, GSCHO1.4). Single chain meganuclease targeting the rag1 locus is used as a control (Rag1 sc). OH: mutation affecting the protein interface

Moreover, 2 single-chain meganucleases have been tested for their potential toxicity using cell survival assay (figure 5.7). In dose-response experiment, none of these meganucleases showed adverse effect.

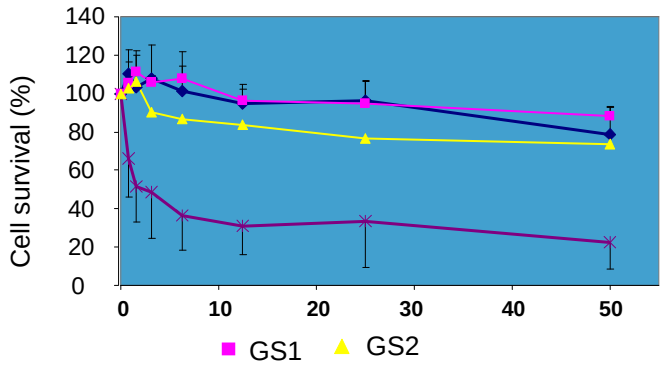

Figure 5.7: cell survival assay. 2 single-chain meganucleases targeting the GS locus were transfected in CHO-LK1 cells (yellow and red). Cell survival was quantified 6 days after transfection. Wild type I-CreI (purple) and I-SceI (dark blue) were used as control.

The best meganucleases are now tested in NIH3T3 cells. The frequency of mutagenesis induced by the meganuclease at the endogenous locus is evaluated.

3/ Chinese hamster HPRT gene

We have screened the Hamster's Hypoxanthine ribosyl transferase gene for potential target. This gene is of interest as it is the main actor in the purine biosynthesis in absence of active DHFR. Thus in CHO-K1 which is DHFR negative, the insertion of any DNA sequences in this locus would disrupt the HPRT-based metabolic pathway: the cell would become sensitive to 6-thioguanine (fig5.8). Therefore a meganuclease targeting the HPRT gene in CHO-K1 could represent a powerful tool.

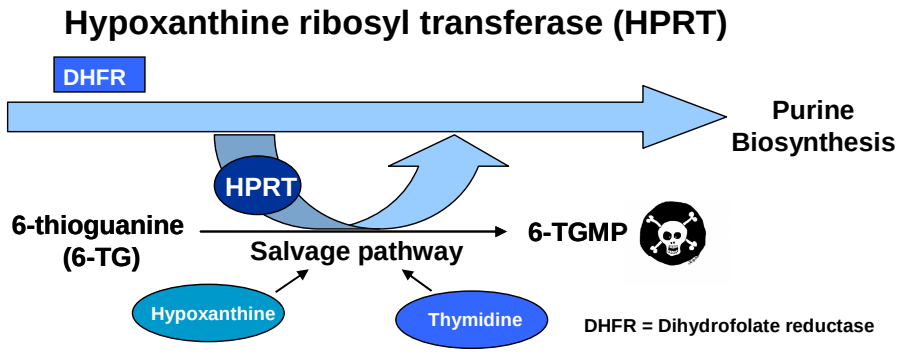

Figure 5.8: Involvement of the HPRT gene in the purine biosynthesis. In absence of DHFR (Dihydrofolate reductase), HPRT is the only metabolic pathway to produce Purine. The 6-thioguanine molecule can be used as a selective agent as its incorporation by HPRT will lead to cell death**.**

During the previous period, we have identified a potential target in exon 3 of the hamster HPRT gene (Hypoxanthine ribosyl transferase), and engineered heterodimeric proteins showing cleavage activity in yeast.

During this period, activity in mammalian cell has been improved, and active single-chain proteins have been produced and targeted gene insertion experiments have been carried out.
Figure 5.9 depicts the HPRT locus in CHO cells and the repair matrix used for the experiment.

Cells were co-transfected with plasmid expressing single-chain meganuclease and the repair matrix. Following double selection (hygromycin and 6-TG), cellular clones were screened for KI events. A frequency of KI events of 1.2 x 10-4 could be achieved in CHO-K1 cells.

A

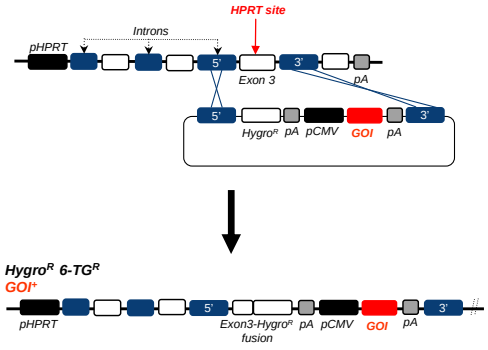

Figure 5.9: Frequency of HygroR/6-TG+ clones after HPRT-mediated targeting in CHO-K1 cells.
A; genomic organization of the HPRT locus and DNA repair Matrix design. GOI: gene of interest. **B**; Table summarizing the gene targeting experiments with heterodimeric (HeD) or single-chain meganucleases (SC).

B

Frequency of Hygro^R/6-TG$^+$ clones

KI matrix only	KI matrix + HeD G19S	KI matrix + SC G19S
7 x10^{-6}	5 x10^{-5}	1.2 x10^{-4}

Moreover, southern blot analysis showed that 99% of the clones contain only targeted insertion.
Finally, we showed that the expression of the inserted

gene over time was a high stable (figure 5.10).

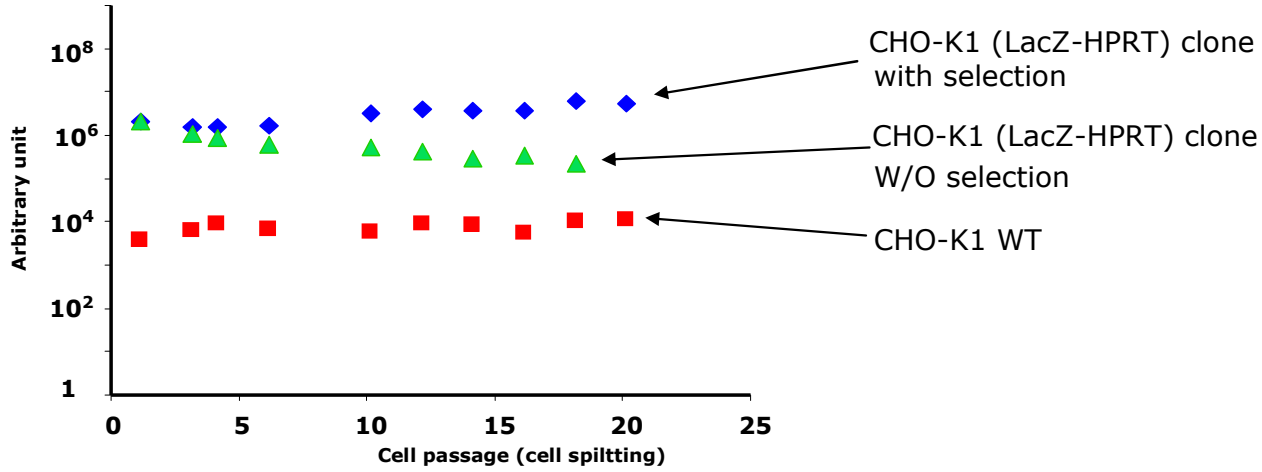

Figure 5.10: targeted insertion at the HPRT locus allows stable expression The LacZ gene wasinserted in the HPRT locus by targeted insertion and ernzymatic activities were quantified at each cell passage. Cells were cultured with or without selection pressure.

4/ New targets: During the course of the project, a new approach was implemented to produce meganucleases. Our proprietary Omegabase contains thousands protein resulting from our combinatorial process. These "half meganucleases can potentially, by co-expression within the cells, form active heterodimeric meganucleases. This approach would considerably decrease the time of delivery of our meganucleases. we search the mouse genome for sequences compatible with a combination of 2 "half meganucleases" already available in our database. Even though, the MEGATOOLS project is now completed, we are pursuing our effort to create new research tools for functional genomics in mouse. Targets were selected in diffrents mouse genes and meganucleases were directly tested in our mammalian cells assay (Figure 5.11). New meganucleases with cleavage activity similar to I-SceI are now available.

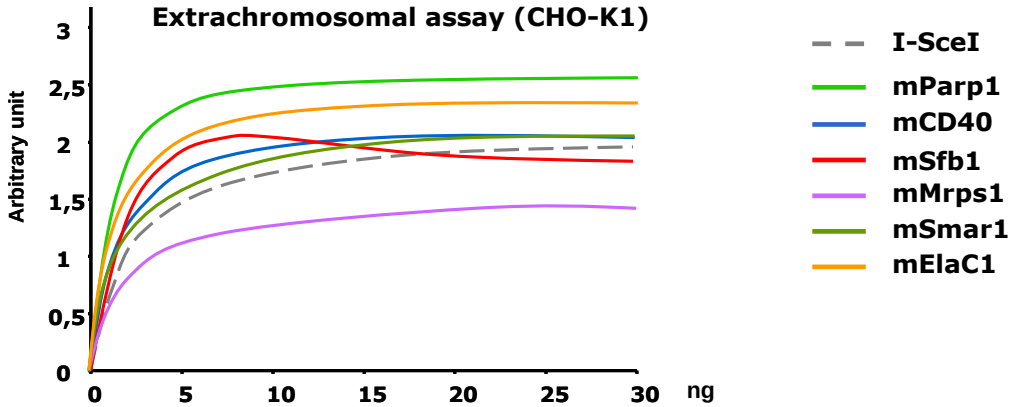

Figure 5.11: Cleavage activity in mammalian cells of the meganucleases targeting mouse genes. Activities of different meganucleases were tested inour SSA extra-chromosomic assay in CHO-K1 cells. genes targeted by the meganucleases are indicated.

Deviations from the project workprogram

The meganuclease targeting the ROSA26 locus showed disappointing result in gene insertion experiments, probably due to the chromatine structure that could make inaccessible the recognition site.

Status of Deliverables and Milestones

Del. no.	Deliverable name	Original delivery date (**)	Revised delivery date (**)	Status (***)	Nature
D5.1	Repair matrix for target 1 & 2	Month 20		**Finished**	**R**
D5.2	Repair matrix for target 3 & 4	Month 28		**Finished**	**R**
D5.3	KO of targets 1 & 2 in mouse cells	Month 28		**Finished**	**R**
D5.4	KO of targets 2 & 4 in mouse cells	Month 36		**partial**	**R**
D5.5	Establishment of standard procedure	Month 36		**Finished**	**R**

Status of the milestones for the period and forecast for the other milestones of the WP5.

Milest. no.	Description	Original target date (**)	Revised target date (**)	Status (***)
M5.1	First cloning of repair matrix	Month 20		**Finished**
M5.2	First KO	Month 28		**Finished**
M5.3	Full validation of the approach	Month 36		**Finished**

2. DISSEMINATION AND USE

A WWW site (http://www.cellectis.com/megatools/,) has been constructed at the coordinator's server for the dissemination of project results and exchange of information between partners.

Section1 - Exploitable knowledge and its Use

Overview table

Exploitable Knowledge (description)	**Exploitable product(s) or measure(s)**	**Sector(s)** of application	Timetable for commercial use	Patents or other IPR protection	**Owner** & Other Partner(s) involved
1.meganuclease targeting the glutamate synthetase mouse gene	*Plasmid expressing the meganuclease and its use thereof*	*1. Industrial inspection* *2.Research tools*	*2010*	*Patented (sept 2008)*	***CELLECTIS SA***
2.meganuclease targeting the Hypoxanthine ribosyltransferase *mouse & hamster genes*	*Plasmid expressing the meganuclease and its use thereof*	*1.Industrial inspection* *2.Research tools*	2009	*Patented (Nov 2006)* *Patented (oct 2008)*	***CELLECTIS SA***
3.meganuclease targeting the ROSA26*mouse gene*	*Plasmid expressing the meganuclease and its use thereof*	*1. Industrial inspection* *2.Research tools*	2011	*Patented (June 2007)*	***CELLECTIS SA***
4 engineering of the chimeric DmoCre meganuclease	*Meganuclease with locally altered specificity and Methods to combine them*	*1. Industrial inspection* *2.Research tools*	2010	*Patented (dec 2007)*	***CELLECTIS SA***
5. engineering of new domains of I-CreI meganuclease	*Meganuclease with locally altered specificity and Methods to combine them*	*1. Industrial inspection*	2010	*(patent is planned for 2010)*	***CELLECTIS SA***

Meganuclease targeting the glutamate synthetase mouse gene

This gene is of interest as GS is responsible for the biosynthesis of glutamine using ammonia and glutamate as substrates. Thus, In absence of glutamine I n the medium, GS is essential for cell survival and can therefore be used as selectable marker.

The ability to destroy GS activity in any mouse or Hamster cells is highly attractive in the protein production industry.

Interestingly, this target sequence is also present in the Chinese Hamster ovary cells genome.

Meganuclease targeting the Hypoxanthine ribosyltransferase *mouse & hamster genes*

This gene is of interest as it is the main actor in the purine biosynthesis in absence of active DHFR Thus in CHO-K1 which is DHFR negative, the insertion of any DNA sequences in this locus would disrupt the HPRT-based metabolic pathway: the cell would become sensitive to 6-thioguanine. Therefore a meganuclease targeting the HPRT gene in CHO-K1 represents a powerful tool for the production of cell lines expressing any gene of interest.

Meganuclease targeting the ROSA26*mouse gene*

This gene is of interest as ROSA26 gene is ubiquitously expressed in mice. The ability to insert gene expression cassette at the locus would tremendous help the production of transgeneis mice for the research community.

Unfortunately, the recognition site that has been selected in our first attempt appears to be not accessible, probably because of chromatin structure constrains. Thus, 4 new recognition sites have been selected and engineered meganucleases are now under production

Engineering of the chimeric DmoCre meganuclease

Homing endonucleases have become valuable tools for genome engineering. Their sequence recognition repertoires can be expanded by modifying their specificities or by creating chimeric proteins through domain swapping between two sub-domains of different homing endonucleases. However, these two approaches can be combined to create engineered meganucleases with new specificities.

The ability to engineered new scaffold has a tremendous potential to expand the number of DNA sequences targetable by meganucleases.

Engineering of new domains of I-CreI meganuclease.

Expanding the target space that can be reached by a particular meganuclease is under highly investigation around the world. The ability to engineered I-CreI based meganucleases with new locally altered specificity, and to be able to combine them will improved the production process of meganucleases. First, it will increase the number of DNA target that we can reach with this protein scaffold. Second, the overall specificity of the enginnered I-CreI beased variant will be improved as the safety of this technoçlogy.

Section2 – Dissemination of knowledge

The dissemination activities below include those falling in the first reporting period and those definitely planned.

Overview table

Planned /actual Dates	Type	Type of audience	Countries addressed	Size of audience	Partner responsible /involved
Jan 07	*Project web-site*	*General public*	**All**	*NA*	
Jul 07	*Conference*	*Research*	**all**	*500*	*CELLECTIS*
May 07	*Poster*	*Research*	**France**	*300*	*CELLECTIS*
May 07	*Poster*	*Research*	**all**	*2000*	*CELLECTIS*
July 07	*Poster*	*Research*	**all**	*500*	*CELLECTIS*
Dec 07	*Conference*	*Industry*	**Germany**	*100*	*CELLECTIS*
2007	*Publication*	*Research*	**All**	*NA*	*CNIO, CRG, CELLECTIS*
Apr 08	*Publication*	*Research*	**All**	*NA*	*CRG, CELLECTIS*
May 08		*research*	**all**	*2000*	*CELLECTIS, CNIO*
Sept 08	*Posters*	*research*	**Spain**	*2000*	*CNIO, CELLECTIS*
2008	*Publication*	*Research*	**All**	*NA*	*CRG, CELLECTIS*
2008	*Publication*	*Research*	**All**	*NA*	*CNIO, CRG, CELLECTIS*
2008	*Publication*	*Research*	**All**	*NA*	*CNIO, CELLECTIS*
2009	*Publication*	*Research*	**All**	*NA*	*CNIO, CELLECTIS*
2009	*Publication*	*Research*	**All**	*NA*	*CELLECTIS*
2009	*Publication pending*	*Research*	**All**	*NA*	*CRG*
2009	*Publication pending*	*Research*	**All**	*NA*	*CRG*
2009	*Publication*	*Research*	**All**	*NA*	*CNIO*
2009	*Publication pending*	*Research*	**All**	*NA*	*CNIO, CRG, CELLECTIS*
2009	*Conference*	*Research*	**Spain**	*300*	*CNIO*
2009	*Conference*	*Research*	**Europe**	*150*	*CNIO*
2009	*Conference*	*Research*	**All**	*600*	*CNIO*
2009	*Conference*	*Research*	**All**		*CRG*

The project has a dedicated website (http://www.cellectis.com/megatools) the majority of which is open to the public. A project fact sheet about Megatools can be found on the webpage.

Publications:

1) Crystallization and preliminary X-ray diffraction analysis on the homing endonuclease I-Dmo-I in complex with its DNA target.
P. Redondo, J. Prieto, E.Ramos, F. Blanco & G. Montoya (2007) Acta Cryst F Struct. Biol. 63(Pt 12):1017-20.
2) Generation and analysis of mesophilic variants of the thermostable I-DmoI homing endonuclease.
J. Prieto, J-C Epinat, P. Redondo, E. Ramos, D. Padró &, G. Montoya, F. Pâques and F. Blanco. J. Biol. Chem. (2008) Feb 15;283(7):4364-74.
3) Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences.
Fajardo-Sanchez E, Stricher F, Pâques F, Isalan M, Serrano L.
Nucleic Acids Res. 2008 Apr;36(7):2163-73.
4) Molecular basis of recognition and cleavage of the human Xeroderma pigmentosum group C gene by engineered homing endonuclease heterodimers.
P. Redondo , J. Prieto , I. Muñoz , A. Alibés , F. Strichter , L. Serrano, S. Arnould , C. Perez , J.P. Cabaniols , P. Duchateau, F. Paques , F. Blanco & G. Montoya. (2008) Nature Nov 6;456(7218):107-11
5) Crystal structure of I-DmoI in complex with its target DNA provides new insights into meganuclease engineering.
P. Redondo, M. Marcaida, J. Prieto, E. Ramos, S. Grizot, P. Duchateau, F. Paques, F. Blanco & G. Montoya (2008) Proc Natl Acad Sci U S A. 2008 Nov 4;105(44):16888-93.
6) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas S, Lemaire L, Montoya G, Blanco FJ, Pâques F, Duchateau P. Nucleic Acids Res. 2009 Sep;37(16):5405-19
7) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds
Sylvestre Grizot, Jean-Charles Epinat, Séverine Thomas, Sandra Rolland, Frédéric Pâques, Philippe Duchateau. Nucleic Acids Res. 2009 (accepted for publication)
8) Structure-based DNA binding prediction and design Alibés, A., Serrano, L., Nadra, A.D. Zinc Finger Proteins: Methods and Protocols. Humana Press. Eds. Joel Mackay and David Segal. *In press*.

9) Structure-based prediction of Protein-DNA binding specificities Alibés, A., Nadra, A.D., De Masi, F., Bulyk, M.L., Stricher, F., Serrano, L " *Submitted*.

10) Homing endonucleases: from basics to therapeutic applications (2009) Maria J. Marcaida, Ines G. Muñoz Francisco J. Blanco • Jesus Prieto Guillermo Montoya. Cellular and Molecular Life Sciences In press

11) Molecular basis of SCID DNA recognition by engineered heterodimers and single chain meganucleases. (2009) Ines G. Muñoz, S. Subramanian, J. Coloma, Nekane Merino, Maider Villate Frédéric Pâques, sylvestre Grizot, Philippe Duchateau, Francisco J. Blanco, Jesus Prieto, Guillermo Montoya. Submitted.

Meeting:

Meganucleases with tailored specificities for genome engineering purposes
Frédéric Pâques. Enzymes engineering meeting, July 2007, China

Induction de la recombinaison homologue par des méganucléases artificielles
Julie Smith, Jean Pierre Cabaniols, Laetitia Lemaire, Laurence Guianvarch, Sophie Leduc, Frédéric pâques et Philippe Duchateau. 7ème colloque des 3R, May 2007, France

Gene targeting induced by engineered endonucleases derived from the I-CreI meganucleases
Jean-Pierre Cabaniols, Sophie Leduc, Cécile Jacqmarcq, Christophe Perez, Sylvestre Grizot, Sylvain Arnould, Philippe Duchateau and Frédéric Pâques. 10th Annual meeting of American Society of Gene Therapy, May 2007, USA

Meganucleases with tailored cleavage specificities can induce efficient homologous gene targeting
Smith J, Cabinols J-P, Gouble A, Grizot S, Perez C, Duchateau P and Pâques F. Faseb SRC, genetic recombination of genome rearrangement, July 2007, USA

Use of engineered meganucleases to trigger efficient DSB-induced homologous gene targeting.
A. Gouble, J. Smith, J.P. Cabinols, S. Grizot, C. Perez, P. Duchateau and F. Pâques . DNA repair, damage signaling, and carcinogenesis. First German-French Meeting, September 2007, France

Improvement of meganuclease specificity for genome engineering
Sylvestre Grizot, Séverine Thomas, Jean-Pierre Cabaniols, Christophe Perez, Sophie Leduc, Guillermo Montoya, Francisco J. Blanco, Jesús Prieto, Frédéric Pâques and Philippe Duchateau. 11th Annual meeting of American Society of Gene Therapy, May 2008, USA

Generation and analysis of mesophilic variants of the thermostable I-DmoI homing endonuclease.
J. Prieto, J-C Epinat, P. Redondo, E. Ramos, D. Padró, G. Montoya, F. Pâques and F. J. Blanco. SEBBM (Spanish society for biochemistry and molecular biology). september 2008. Spain

Sructural biology of homing endonucleases and their possible therapeutic applications.
Guillermo Montoya SEBBM septemeber 2009 Oviedo

Sructural biology of homing endonucleases and their possible therapeutic applications
University of Santander, Spain, 2009

Sructural biology of homing endonucleases and their possible therapeutic applications
International gene theraphy congress 2009, Granada Spain.

THE DNA-PROTEOME: Recent advances towards establishing the protein-DNA interaction space. Barcelona, Spain. 20-22 APRIL, 2009

-Structure-based DNA binding prediction and design. Alejandro D. Nadra, Andreu Alibés, Francois Stricher, Luis Serrano.

-Structure-based prediction of protein-DNA binding specificities. Andreu Alibés, Alejandro D. Nadra, Francois Stricher, Luis Serrano.

Section 3 - Publishable results

Meganucleases can be engineered to change their recognition site sequence. This breakthrough in protein engineering alleviates the need to pre-engineer cell lines with a natural meganuclease recognition site. Engineered meganucleases achieve targeted integration on “wild type” cell lines.

The first ever engineered meganuclease for the hamster genome targets the natural HPRT (Hypoxanthine-guanine PhosphoRibosylTransferase) locus. HPRT is involved in purine metabolism and converts 6-thioguanine (6-TG) to its monophosphate form (6-TGMP). 6-TGMP interferes subsequently with de novo synthesis of purines.

CHO-K1 HPRT+ cells are naturally sensitive to 6’TG. Upon expression of the engineered meganuclease targeting HPRT following cell transfection, a DNA double-strand break is made which triggers repair through homologous recombination. The transfected integration matrix is used for repair and contains upstream and downstream homology regions of the HPRT locus, a hygromycin resistance gene, and promoter and terminator sequences of the Gene Of Interest (GOI) coding sequence. Upon repair, the GOI is integrated at the HPRT meganuclease recognition site, and the cells become hygromycin and 6-TG resistant (HPRT deficient).

- **targeted gene integration in your lab’s cell line**
- **guaranteed gene expression**
- **sustained protein production**
- **clone homogeneity**
- **fast & effortless**

cGPS® and cGPS®Custom are the unique ways to stably express a protein from your cell line!

Publications:

1) Crystallization and preliminary X-ray diffraction analysis on the homing endonuclease I-Dmo-I in complex with its DNA target.

P. Redondo, J. Prieto, E.Ramos, F. Blanco & G. Montoya (2007) Acta Cryst F Struct. Biol. 63(Pt 12):1017-20.

2) Generation and analysis of mesophilic variants of the thermostable I-DmoI homing endonuclease.
J. Prieto, J-C Epinat, P. Redondo, E. Ramos, D. Padró &, G. Montoya, F. Pâques and F. Blanco. J. Biol. Chem. (2008) Feb 15;283(7):4364-74.

3) Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences.
Fajardo-Sanchez E, Stricher F, Pâques F, Isalan M, Serrano L.
Nucleic Acids Res. 2008 Apr;36(7):2163-73.

4) Molecular basis of recognition and cleavage of the human Xeroderma pigmentosum group C gene by engineered homing endonuclease heterodimers.
P. Redondo , J. Prieto , I. Muñoz , A. Alibés , F. Strichter , L. Serrano, S. Arnould , C. Perez , J.P. Cabaniols , P. Duchateau, F. Paques , F. Blanco & G. Montoya. (2008) Nature Nov 6;456(7218):107-11

5) Crystal structure of I-DmoI in complex with its target DNA provides new insights into meganuclease engineering.
P. Redondo, M. Marcaida, J. Prieto, E. Ramos, S. Grizot, P. Duchateau, F. Paques, F. Blanco & G. Montoya (2008) Proc Natl Acad Sci U S A. 2008 Nov 4;105(44):16888-93.

6) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas S, Lemaire L, Montoya G, Blanco FJ, Pâques F, Duchateau P. Nucleic Acids Res. 2009 Sep;37(16):5405-19

7) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds
Sylvestre Grizot, Jean-Charles Epinat, Séverine Thomas, Sandra Rolland, Frédéric Pâques, Philippe Duchateau. Nucleic Acids Res. 2009 (accepted for publication)

8) Structure-based DNA binding prediction and design Alibés, A., Serrano, L., Nadra, A.D. Zinc Finger Proteins: Methods and Protocols. Humana Press. Eds. Joel Mackay and David Segal. *In press*.

9) Structure-based prediction of Protein-DNA binding specificities Alibés, A., Nadra, A.D., De Masi, F., Bulyk, M.L., Stricher, F., Serrano, L " *Submitted*.

10) Homing endonucleases: from basics to therapeutic applications (2009) Maria J. Marcaida, Ines G. Muñoz Francisco J. Blanco • Jesus Prieto Guillermo Montoya. Cellular and Molecular Life Sciences In press

11) Molecular basis of SCID DNA recognition by engineered heterodimers and single chain meganucleases. (2009) Ines G. Muñoz, S. Subramanian, J. Coloma, Nekane Merino, Maider Villate Frédéric Pâques, sylvestre Grizot, Philippe Duchateau, Francisco J. Blanco, Jesus Prieto, Guillermo Montoya. Submitted.

Patent:

8 sept 2008 - PCT/IB2008/003109 (non publiée à ce jour) "MEGANUCLEASE VARIANTS CLEAVING A DNA TARGET SEQUENCE FROM THE GLUTAMINE SYNTHETASE GENE AND USES THEREOF"

- HPRT 14 nov 2006 - PCT/IB2007/004281 (publiée sous le n° WO 2008059382) "MEGANUCLEASE VARIANTS CLEAVING A DNA TARGET SEQUENCE FROM A HPRT GENE AND USE THEREOF"

12 dec 2007 - PCT/IB2007004418 (publiée sous le n° WO WO2009074873) "IMPROVED CHIMERIC MEGANUCLEASE ENZYMES AND USES THEREOF"

6 juin 2007 - PCT/IB2007/002830 (publiée sous le n° WO2008152523) "MEGANUCLEASE VARIANTS CLEAVING A DNA TARGET SEQUENCE FROM THE MOUSE ROSA26 LOCUS AND USES THEREOF"