University of Glasgow | College of Medical, Veterinary & Life Sciences

Head of College Scholars List Scheme

Summer Studentship

Report Form

This report should be completed by the student with his/her project supervisor. It should summarise the work undertaken during the project and, once completed, should be sent by email to: jill.morrison@glasgow.ac.uk within four weeks of the end of the studentship.

1. Student

   Surname: Martak Forename: Tomas

   E-mail address: 2266686M@student.gla.ac.uk

2. Supervisor

   Surname: Smith Forename: Brian

   E-mail address: Brian.Smith@glasgow.ac.uk

3. Research Project Report

   3.1 Project Title (maximum 20 words):

   Structure of the CLF-C5 domain and its interactions with the EMF2-VEFS domain.

   3.2 Project Lay Summary (copied from application):

   Plants use epigenetic memory, that is, changes to their DNA that do not change its sequence, to regulate their development. For example, vernalisation is the process by which plants know when winter is past and that they should germinate, or should flower. These memory processes and our understanding of how to manipulate them have a significant impact on agriculture.

   Part of the underlying machinery involves proteins that can modify DNA or other proteins that package it. In this project, we will study the 3-D structures of two key proteins, CURLY LEAF and EMBRYONIC FLOWER 2, to understand how they function.

3.3 Start Date: 21.5.2018 Finish Date: 29.6.2018

3.4 Original project aims and objectives (100 words max):

The original aim of this research was to understand how CLF-ARATH recognises EMF2-VEFS through its C5 region. Towards this aim, the two objectives of the summer project were 1) to express and purify CLF-C5 and characterize its structure by biophysical means including NMR spectroscopy, and 2) to test whether a CLF-EMF2 complex can be assembled by co-expression in E. coli.

3.5 Methodology: Summarise and include reference to training received in research methods

etc. (250 words max):

3.5.1 DNA extraction and quantification

Expression vectors were extracted from competent E. coli with the *Wizard*® Plus SV *Minipreps* DNA Purification System using the corresponding protocol, and DNA concentration was measured using a NanoDrop™ Spectrophotometer.

3.5.2 Protein visualisation, sequence analysis and design

The human PRC2 was examined in PyMOL v. 2.1.1. informed by the domains identified in UniProt and by the SMART (Simple Modular Architecture Research Tool). Sequence alignments of EZH2 and the two Arabidopsis homologues were conducted in Jalview using CLUSTAL. Finally, CLF-C5 and SWN-C5 constructs were developed in silico using CLC Genomics Workbench 11.

3.5.3 PCR and agarose gel electrophoresis

Custom forward and reverse primers were designed to isolate the SWN-C5 and EMF2 sequences. Primer annealing temperatures were calculated using Thermo Fisher and NEB online calculators. The PCR was run according to NEB protocol using *Taq* DNA polymerase. DNA products were separated and visualized by agarose gel electrophoresis was conducted using Invitrogen *UltraPure™ Agarose, TAE buffer, GelGreen®, and NEB 6X Purple Gel Loading Dye and NEB 1kb DNA ladders.*

3.5.4 Vector linearization and subsequent purification from gel

pET30a-EMF2-VEFS was linearised with SalI (NEB) following the manufacturer's protocol. The band corresponding to the linearised vector was extracted from the agarose gel following electrophoresis and purified with a Machery-Nagel NucleoSpin® kit. Elution in NE buffer was repeated an additional 2 times to maximise plasmid recovery.

3.5.5 Cloning and transformation

CLF-C5 and SWN-C5 inserts were cloned into the linearised vector using both Trans pEASY ® and NEBuilder® HiFi Assembly cloning kits following their respective protocols and the supplied competent E. coli strains.

3.6 Results: Summarise key findings (300 words max). Please include any relevant tables or images as an appendix to this report:

3.6.1 Sequencing CLF-C5 and EMF2-VEFS vectors

After extracting and sequencing the CLF-C5 vector, it was found to be a mislabelled SWN-C5-containing bait from a previous Y2H assay without an appropriate promoter for large-scale expression. The identity of the EMF2-VEFS vector was found correct after sequencing.

3.6.2 *In silico* PRC2 structural analysis

Examination of the 3-D structures of human PRC2 in tandem with analysis of sequence alignments of its domains with their *Arabidopsis* counterparts yielded unexpected results. Based on the common occurrence of 2 zinc fingers, the MCSS motif of the human EZH2 was found to be a homologue of the C5 motif previously described in plant EZ homologues. Additionally, it was found that the C5 motif needed to be co-expressed with the second SANT domain from the same protein and the VEFS box of a SuZ homologue to fold properly due to the extensive hydrophobic interactions between the two protein subunits and the Zn^{2+} binding site shared between the C5 and SANT domain (**Fig. 1**).

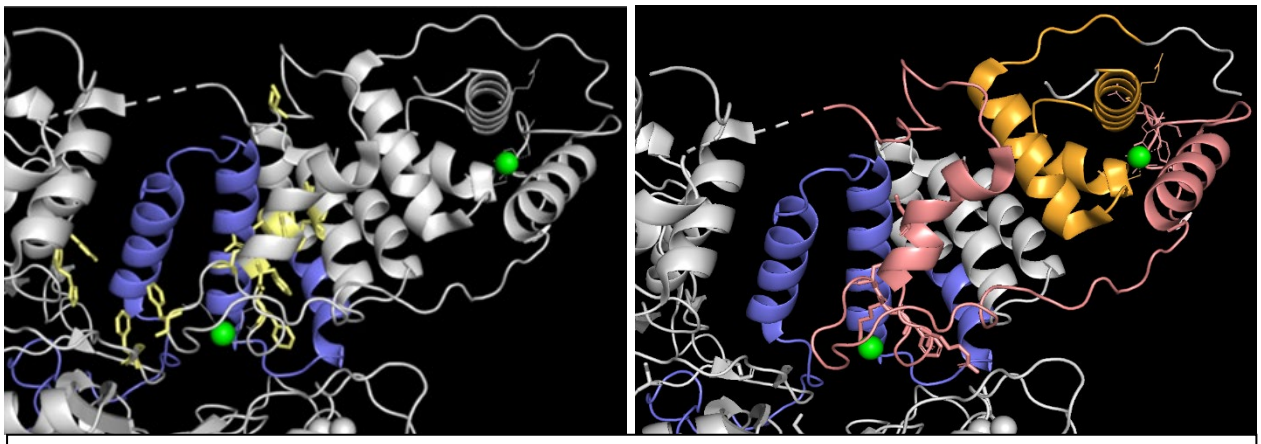

Figure 1. Motifs in the human PRC2 structure. The structural elements corresponding to the conserved sequence motifs are identified in different colours. The Suz12 VEFS box is shown in blue; zinc ions are shown in green; amino acid residues within 5Å of the SuZ12 VEFS box are shown in yellow; the EZH2 MCSS motif is shown in pink, and the EZH2 SANT 2 domain is shown in orange. The interaction between a part of the MCSS motif and the VEFS box is needed for the proper folding of both of the domains as is the Zn2+ binding site shared with the SANT 2 domain.

3.6.3 *In silico* CLF-C5-SANT2 and SWN-C5-SANT2 insert design

SWN-C5 and CLF-C5 constructs were designed to introduce sequences coding for the C5 and SANT2 domains of each protein joined by truncated versions of the intervening sequences. These intervening sequences are predicted to be sufficiently long to allow the C5 and SANT2 domains to fold and adopt the same relative orientation as seen in the PRC2 complex. A further 15 amino acid unstructured linker region was included so that when the inserts were cloned after the EMF2 VEFS box in the existing pET30a vector, the expressed polypeptide would produce a linked VEFS-C5-SANT2 protein. The inserts were also designed to contain two SalI sites seven nucleotides apart in this linker so that cleavage by SalI and re-

ligation with the loss of the intervening sequence would introduce a stop codon allowing expression of separate EZ and SuZ polypeptides from a polycistronic messenger RNA.

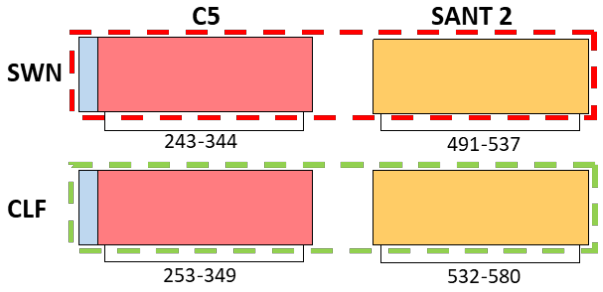

Figure 2. Schematic representation of the CLF-C5-SANT2 and SWN-C5-SANT2 constructs. Parts of domains included in the construct are identified with different colours. C5 domains are shown in pink, SANT 2 domains are shown in orange and the 15 amino acid linker region before the C5 motif is shown in blue. The numbers below the boxes show the parts of SWN and CLF used in the constructs. The unstructured regions between the C5 motif and SANT 2 domain were left out to improve folding efficiency.

3.6.4 Linearisation of pET30a-EMF2-VEFS with SalI followed by insertion of CLF-C5 and SWN-C5 and transformation

The pET30a-EMF2-VEFS plasmid was successfully extracted from an overnight bacterial culture and digested with SalI, yielding an appropriate band following agarose gel electrophoresis. The linearized plasmid was incubated in a DNA assembly reaction with the DNAString encoding either the CLF or the SWN domains and the products used to transform chemically competent *E.coli.* Colonies of the transformed cells were able to grow on selective medium. To check for the presence of the new inserts, plasmid DNA extracted from these colonies was used as template for PCR with T7 promoter and T7 terminator primers. The PCR products were resolved by agarose gel electrophoresis and visualized revealing bands that could plausibly be the size corresponding to the EMF2-CLF and EMF2-SWN coding sequences.

3.7 Discussion (500 words max):

3.7.1 The EZH2 MCSS as an analogue of *Arabidopsis* EZ C5 motifs

Although the catalytic activity of the PRC2 complex is mediated by the SET domain of its EZ subunit, interaction between the EZ and SuZ subunits is crucial for the adoption of their correct conformation and targeting of the complex's activity. The platform for this interaction is the SuZ VEFS box, which interacts extensively with the EZ MCSS motif, which we found to be a homologue of the C5 motifs of *Arabidopsis* PRC2 EZ homologues (CLF, SWN, and MEA). This could have some implications for the better understanding of the role of the individual subunits of the PRC2 complex in humans.

3.7.2 Cloning results and future research

The effectiveness of the cloning undertaken was debatable. The transformants obtained grew on selective media, but the parent plasmid would also permit growth if it re-circularized. Two of the colonies that showed additional sequences to have been inserted as judged by PCR with T7 promoter and T7 terminator primers were sequenced but yielded inconclusive results, and the length of the internship left no time or resources for further tests.

In future research it will be important to repeat the cloning process and to test a sufficient number of transformants. Successfully assembled vectors could then be extracted and used

to further study the interaction between the EMF2 VEFS box and the C5 motifs of CLF and SWN.

4. Reflection by the student on the experience and value of the studentship (300 words max):

Although we didn't manage to complete the project and had to scrap the first and main objective, i.e. the characterisation of the structure of CLF-C5 by NMR spectroscopy, it is precisely because of this that I ended up learning more that I possibly would've done. I improved my troubleshooting skills and worked toward being able to conduct experiments independently. I was able to gain experience in using a range of software and online tools including PyMOL, CLC Genomics Workbench, Jalview, SMART, and BLAST. Throughout the internship I also learned and improved on many molecular biology techniques, such as setting up bacterial cultures, large-scale expression, plasmid minipreps, casting and running gels, DNA purification from agarose gel, running a PCR, restriction digest, cloning, and bacterial transformation. This short internship has given me a good skill-set that I can use and improve upon in my future projects and internships.

I'm grateful to the HOCS List Scheme for giving me the opportunity to meet Dr Smith, and to Dr Smith for patiently overseeing my work and going out of his way to correct my blunders.

5. Dissemination: (note any presentations/publications submitted/planned from the work):

6. Signatures:

Supervisor	Date
	 25.7.2018