The genome in each cell of a complex organism is the same (with rare exceptions, for example in lymphocytes in mammals). The genome consists of four DNA "letters", which string up to encode the genomic information. Although all cells contain the same instruction set, different cell types have different developmental histories, morphologies and functions. The differences come about because cells express the same genome differently. This is because the expression of genes is controlled not only by the DNA sequence, but also by epigenomic marks, which reside either on the DNA itself (as cytosine methylation or hydroxymethylation) or on histones, on which DNA is wrapped.

In my studies, I will focus on the control of gene expression in plants, and more specifically in the model plant Arabidopsis. This plant is a member of the mustard family that is widely used in research, because it is a convenient model: the plants are small, and have a relatively short generation time between four and six weeks, which makes it possible to study the consequences of genetic alterations over several generations. Moreover, the genome is compact and large collections of genetically altered *Arabidopsis* plants have already been created, which facilitate many genetic studies. *Arabidopsis* is itself not an economically important plant, it is representative for a large group of plants (technically the *Brassic*a and *Sinapis* plants), which do have great economic and commercial value. They play a major role in feeding the world population, and they may also be useful for the production of biofuels.

One of the best know epigenetics markers is DNA methylation, a covalent modification of DNA which alters and typically represses gene expression. Thanks to elaborate machinery that operates together with the DNA replication machinery, some DNA methylation is heritable. I am interested in a form of DNA methylation that is not directly heritable as cells divide, but is instead controlled by small RNAs, which guide the methylation machinery. At the heart of this machinery is a protein called DRM2 (domains rearranged methyltransferase 2), which consists of two key parts: one called the catalytic domain does the "writing" of methylation, the other part is likely acting as a guide (akin to the hand that holds the "pen"). We know that this part of the protein is important, because it is conserved in different plants, and also because biochemical experiments have shown that its loss disrupts the "writing" process. Using different genetic, biochemical and biophysics techniques I will clarify the role of the non-catalytic part of this enzyme.

Much is already known about the catalytic part of DRM2. I will focus on the targeting part, which consists of three ubiquitin-associated (UBA) domains. Prior experiments by others have already shown that these are essential for the function of DRM2. Once the UBA domains are ablated, the protein stops to work. I want to understand how the UBA domains control DRM2 activity. Are they involved in targeting DRM2? Is there a division of labor between different UBA domains? Is each responsible for a subset of the genomic DRM2 targets? Or are different UBA domains perhaps controlling the function of the DRM2 catalytic domains at different stages of the development of the plant? Either way, what are the binding partners of the UBA domains? Do they directly sense epigenomic marks in chromatin (for example on histones), or do they interact with proteins to do this? And once the interaction partners are known, how do the interactions work in molecular detail?

In order to deduce the role of individual UBA domains, I will do genetic experiments. Fortunately, otherwise isogenic plants that either have or lack the DRM2 gene are already available. In the DRM2 deficient plant I will introduce the wild-type gene as a positive control, the gene fragment for the catalytic domain as a negative control, and DRM2 gene versions lacking the coding regions for isolated UBA domains or UBA domain combinations. I will then monitor DRM2 dependent methylation (technically called CHH methylation according to the sequence context) in the transgenic plants using a technique known as bisulfite sequencing. Bisulfite sequencing relies on a chemical trick (deamination of cytosine but not methyl cytosine under bisulfite conditions) to deduce the methylation status of DNA.

In order to understand how UBA domains target DRM2 to different genomic loci, I will attempt to identify the binding partners of isolated UBA domains or of combinations of the UBA domains (if the genetic experiments suggest that the UBA domains act in concert). I will use tagged versions of the UBA domains as "bait" to "fish" for interaction partners, and then characterize isolated interaction partners by mass spectrometry. In order to make the circumstances of this experiment as physiologically as possible, tagged UBA domains will be overexpressed in the plants under the control of the genome regions that control the expression of the DRM2 gene in the wild-type plant.

Finally, I will structurally characterize UBA domains or their combinations. Expression in plants will not provide me with enough material for these experiments. Therefore I will overproduce the domains in bacteria (*Escherichia coli*) that have been genetically modified to be used for such purposes. I will then try to coax protein molecules *in vitro* into regular arrays called crystals, which can be analyzed with X-ray light. The resulting data should provide me with atomic level detail on how the UBA domains are built and how they interact with their interaction partners.