Fidelity and Dynamics of DNA Methylation in Plants

Qin Yao[1]*(qyao@bnl.gov), Changjun Liu[1], John Shanklin[1], Chuan He[2] and **John Dunn[1]**
[1]Department of Biology, Brookhaven National Laboratory, Upton, NY 11793, [2]Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois, USA 60637

Project Goals:

The main goal of our program is to identify fundamental pathways and protein-DNA interactions maintaining inheritance of epigenetic states of plant DNA and chromatin during cell divisions. During semi-conservative DNA replication, hemi-methylated sites are generated where only cytosines in the parental strands remain modified. The correct methylation patterns in the daughter strands need to be reestablished to ensure proper epigenetic homeostasis, i.e. expression of essential genes and silencing of parasitic transposons. Recently we found that Arabidopsis genomic DNA contains 5-hydroxymethylcytosine (5hmC) as well as 5-methylcytosine (5mC)[(1)] which raises several questions: 1) where are these residues located; 2) do proteins that help maintain proper DNA methylation patterns in replicated DNA also maintain these patterns after replication across sites with 5hmC; and 3) do these two modifications play distinct epigenetic roles? To help address these questions, we are using a sensitive chemical labeling method to capture DNA fragments with 5hmC to discover its distribution pattern in the *Arabidopsis* genome and *in vitro* DNA binding studies to determine whether proteins needed for high-fidelity maintenance of cytosine methylation lose their intrinsic preference for hemi-methylated DNA when 5mC is replaced by 5hmC. Ultimately, an efficient and single-base resolution method for 5hmC identification is highly desirable. To this end we will integrate J binding protein, which has been shown to recognize single 5hmC residues[(2)] with next generation sequencing techniques to develop a new high-resolution 5hmC mapping tool which can be easily operated in a standard molecular lab. This project will also attempt to use genome-wide association analysis to enhance our understanding that any observed changes in DNA methylation status are non-random, gene-specific alterations that can result in gene activation or inactivation, especially for genes involved in lignocellulose biosynthesis and oil formation. All the newly generated epigenetic data and tools will be deposited into the BESC Knowledge Base (KBase).

In *Arabidopsis*, several different mechanisms are utilized to reestablish appropriate epigenetic methylation marks in the daughter strand cytosines. One of these mechanisms involves the five member VIM (Variant in Methylation) family of proteins which help reestablish daughter strand cytosine methylation at symmetric CG dinucleotides. Using recombinant proteins and modified double-stranded deoxyoligonucleotides, we observed that full-length VIM1 binds preferentially to hemimethylated DNA with a single modified 5mCG site; a result consistent with its known role in preserving DNA methylation in vivo following DNA replication. However, when 5hmC replaces one or both cytosine residues at a palindromic CpG site, VIM1 binds with approximately 15-fold lower affinity. VIM3, another member of the five member VIM family, has a single amino acid change, S317A, relative to VIM1, 2, 4 and 5 in the region known to be important for binding to methylated DNA. Changing this residue in VIM1 did not significantly

affect its DNA-protein binding properties. These results suggest that 5hmC may contribute to VIM-mediated passive loss of cytosine methylation in vivo during Arabidopsis DNA replication. Work is now in progress to extend these observations to other proteins involved in maintaining epigenetic fidelity to determine their ability to discriminate between substrates with 5mC vs. 5hmC using purified recombinant proteins as well as protein binding microarrays (PBMs)[3].

Previous studies of DNA hydroxymethylation suggest hydroxylation of 5mC may promote transcriptional de-repression by dissociation of mC-binding proteins and/or recruitment of effector proteins[4]. Our preliminary profiling of 5hmC residues in *Arabidopsis* reveals that 5hmC is enriched selectively in gene body regions rather than more distal regions and that there is a significantly different pattern around transcription start regions (TSS). This relative distribution is associated with gene expression levels. There is a highly negative correlation between enrichment in proximal regions to TSS's but no significant correlation between enrichment in the gene bodies and gene expression level. These results suggested that conversion of 5mC to 5hmC, possibly coupled to active or passive demethylation, might act as a switch to fine tune epigenetic homeostasis or perhaps have an active role in reprogramming the epigenome. Functional epigenomic studies will eventually address the mechanism from epigenotype to phenotype.

1. Yao, Q., Song, C.X., He, C., Dunn, J.J. Detection of 5-hydroxymethylcytosine in Arabidopsis and its potential impact on recognition of VIM1 (under review).
2. Robertson, A.B., Dahl, J.A., Vagbo, C.B., Tripathi, P., Krokan, H.E. and Klungland, A. (2011) A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res, 39, e55.
3. M.L. Bulyk, Analysis of sequence specificities of DNA-binding proteins with protein binding microarrays. (2006) Methods Enzymol 410, 279-299.
4. Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X. et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol, 29, 68-72.