COVER SHEET

TITLE: Impact of flanking residue modifications on binding specificity of H3K9 methylation reader; Dppa3

AUTHOR'S NAME: ____________________

MAJOR: ____________________

DEPARTMENT: ____________________

MENTOR: ____________________

DEPARTMENT: ____________________

MENTOR(2): ____________________

DEPARTMENT(2): ____________________

YEAR: ____________________

(The following statement must be included if you want your paper included in the library's electronic repository.)

The author hereby grants to University of Wisconsin-Madison the permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

ABSTRACT

Impact of flanking residue modifications on binding specificity of H3K9 methylation reader; Dppa3

Somatic cells can be converted to an embryonic stem cell-like state by transcription factor-mediated reprogramming. These induced pluripotent stem cells (iPSCs), like ESCs, have the ability to divide indefinitely and differentiate into any tissue under the correct stimuli. This property makes them valuable for regenerative therapy. Chromatin modification is thought to play a role in maintaining the plasticity of pluripotency. Notably, histone modifications that can be recognized by specific reader proteins are crucial to chromatin regulation. Recent work has found that modifications to flanking residues of key histone modification sites can play a regulatory role in the binding of reader proteins. Furthermore, proteins that recognize H3K9 methylation have been implicated as having significant impacts on late stage reprogramming. We are therefore interested in assessing the impact of flanking residues on the binding of Dppa3, which has been implicated as important for reprogramming, to the H3K9 methylation site. To this end we are purifying recombinant versions of this protein and determining the impact of flanking residue modifications to its binding specificity using an *in vitro* histone peptide array

Michael Zaiken/Biochemistry
Author Name/Major

Rupa Sridharan/Cell and Regenerative Biology
Mentor Name/Department

Author Signature

Mentor Signature

5.6.15

Date

Title: Impact of flanking residue modifications on binding specificity of H3K9 methylation reader; Dppa3

Abstract:

Somatic cells can be converted to an embryonic stem cell (ESC)-like state by transcription factor-mediated reprogramming. These induced pluripotent stem cells (iPSCs), like ESCs, have the ability to divide indefinitely and differentiate into any tissue under the correct stimuli. This property makes them valuable for regenerative therapy. Chromatin modification is known to play a role in maintaining the plasticity of pluripotency. Histone modifications are recognized by specific "reader" proteins for further downstream function. The recognition and binding by the reader protein to a specific site can be influenced by modifications on flanking amino acid residues. High levels of Histone H3 lysine 9 methylation (H3K9me) is a barrier to the late stage of reprogramming. We are therefore interested in assessing the impact of flanking residues on the binding of Dppa3, a protein implicated as important for reprogramming, to the H3K9 methylation site. To this end we are purifying recombinant versions of Dppa3 to determine the impact of flanking residue modifications to their binding specificity using an *in vitro* histone peptide array.

Introduction:

With the advent of stem cell culturing techniques, regenerative medicine has become an exciting new avenue for therapy. Of particular importance was the discovery that adult somatic cells can be converted to an embryonic stem cell (ESC) like state through the overexpression of four transcription factors; Oct4, Sox2, Klf4, and cMyc. These induced pluripotent stem cells (iPSCs) are an extremely valuable tool for regenerative therapy because they can be derived from

the patient's own cells which removes the likelihood of transplant rejection and avoids ethical problems sometimes ascribed to ESCs.[1] Despite this significant potential, there are major hurdles for translating iPSCs into clinical practice.

The primary roadblock to translating iPSCs to clinical use is the low efficiency of transcription factor mediated reprogramming. Of the original somatic cell culture only between 0.5% and 5% of cells will successfully reprogram to the pluripotent state. This is further exasperated by the fact that during the reprogramming process the cells in culture exist as a heterogeneous mixture. Different cells in the culture will be at different stages of reprogramming at any given time during the process. This can make it difficult to assess which cells have completed the process and which have stalled. Furthermore it is not currently possible to predict which cells in the original somatic culture will successfully complete the conversion. All of this is compounded by how slow the reprogramming process is; it takes approximately 21 days to successfully derive iPSCs from fibroblasts.[2]

To overcome heterogeneity of the reprogramming mixture, we can isolate pre-iPSC and propagate them clonally. These are an intermediate in the reprogramming process characterized by ESC like cell morphologies and certain cell surface markers (SSEA-1, Ecad), but they do not yet have endogenous activation of pluripotency associated genes (Nanog, Oct4, Sox2, Klf4) and are not themselves pluripotent (Figure 1A).

In order to clear these hurdles and improve iPSC reprogramming efficiency much research has been done to characterize the regulatory changes that occur during the reprogramming process. We interested in investigating how different combinations of epigenetic changes, and exogenous factors can modify the reprogramming process. Of these changes, our

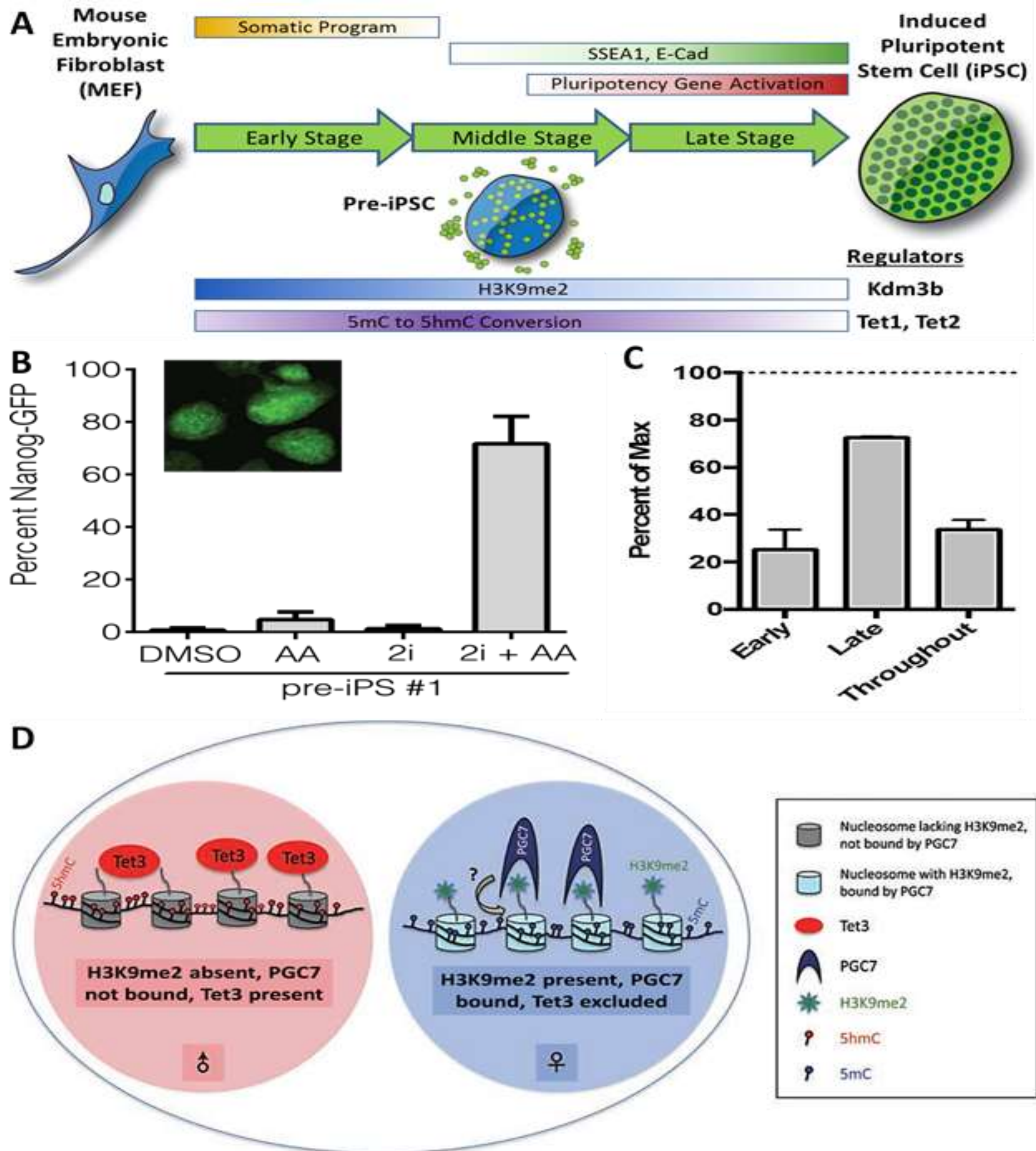

Figure 1: (**A**) Schematic diagram of transcription factor mediated stem cell reprogramming. Shows important changes in chromatin landscape and transcriptome for the process. (Image from Sridharan *et. al.* (**B**) Synergistic effect of AA and 2i on iPSC reprogramming as measured by a GFP reporter under control of pluripotency gene Nanog promoter.[4] **(C)** Impact of Kdm3b knockdown at different time points during pre-iPSC to iPSC reprogramming. Shows that efficiency loss is greatest during the early phase.[4] **(D)** Schematic of Dppa3's role during embryogenesis following fertilization. Shows Dppa3 localizes to H3Kpme2, blocking Tet access, preserving 5mC.[6]

lab is particularly interested in studying the role of histone 3 lysine 9 mono and di methylation (H3K9me), and 5'-methylation of the DNA (5mC). H3K9me2, has previously been implicated as a barrier to late stage reprogramming. In order to pass into the last stage of reprogramming, this mark needs to be largely lost from the genome.[3] Similarly, throughout the reprogramming process there is a global conversion of 5mC to 5'hydroxymethylation (5hmC), which is a precursor for dynamic removal of the methylation (Figure 1A).[2]

Combinations of small molecules have been found to be able to enhance the reprogramming efficiency of iPSC conversion. Work in our lab has identified that the addition of ascorbic acid (AA) as well as a MEK and GSK3 pathway inhibitor (2i) can increase the efficiency of pre-iPSC to iPSC reprogramming to approximately 70% (Figure 1B).[4] Furthermore our lab identified that for the efficiency to be this high the cells must be treated with AA for approximately 48 hours at the beginning of the reprogramming process. From this work, we found that if an H3K9me2 specific histone demethylase, Kdm3b, was knocked down during the 48 hour period where ascorbic acid was critical to elevate reprogramming efficiency, then this elevated efficiency was lost (Figure 1C).[4, 5] Also during this same 48 hour time frame there was a global rise in 5hmC in the culture.[4] This temporal correlation of events suggests that there may be some connection between H3K9me2 and 5mC to 5hmC conversion, and that it may be mediated by AA.

These two epigenetic marks, 5mC to 5hmC conversion and H3K9me2, are known to be linked during embryogenesis via the activity of a protein called Dppa3/PGC7/Stella. Following fertilization Dppa3 relocalizes from the cytoplasm into the two pronuclei, where it selectively binds at loci that have H3K9me2.[6,7] At these loci Dppa3 is hypothesized to exclude the Tet enzymes from acting, thus preserving 5mC (Figure 1D).[6] A recent study has also found that

reprogrammed Dppa3 knockout MEFs arrest at the pre-iPSC stage. This study further found that adding exogenous Dppa3 into the arrested knockout pre-iPSCs does not rescue the phenotype; the cells remain arrested at the pre-iPSC stage. Interestingly, the addition of AA to these arrested cells does rescue the phenotype, and the cells are able to complete reprogramming.[8]

These studies further suggest that there may be a connection between H3K9me2, 5mC to 5hmC conversion, reprogramming efficiency, and AA. We hypothesize that what links these phenomena is Dppa3. In order to begin assessing this hypothesize we are performing early characterization studies on Dppa3. We are most interested in assessing two questions. First does Dppa3 bind to H3K9me2 on its own and if so do flanking residue modifications effect its binding specificity? Second what other proteins associate with Dppa3 in ESCs? Here we present preliminary work in assessing these questions.

***In Vitro* Dppa3 Tag Screening:**

To assess whether flanking residue modifications impact the binding specificity of Dppa3 to H3K9me2 we will use a histone tail peptide binding array containing different permutations of H3K9 methylation, and different flanking residue post translational modifications. Use of the array requires successful purification of Dppa3. To this end we are purifying recombinant tagged Dppa3 from *E. coli*. Multiple different tags were screened for successful expression of the tagged construct. These included a Glutathione-S-transferase (GST) tag, as well as both a C (CHT) and N (NHT) terminus HaloTag (Figure 2A)[9]. The HaloTag consists of a 6x HisTag for nickel column purification, and a HaloTag, which is a luciferase. The luciferase would be used to assay binding location on the array. Each recombinant protein was placed under control of an IPTG inducible promoter and transformed into two different strains of *E. coli*; BL21-DE3, and BL21-

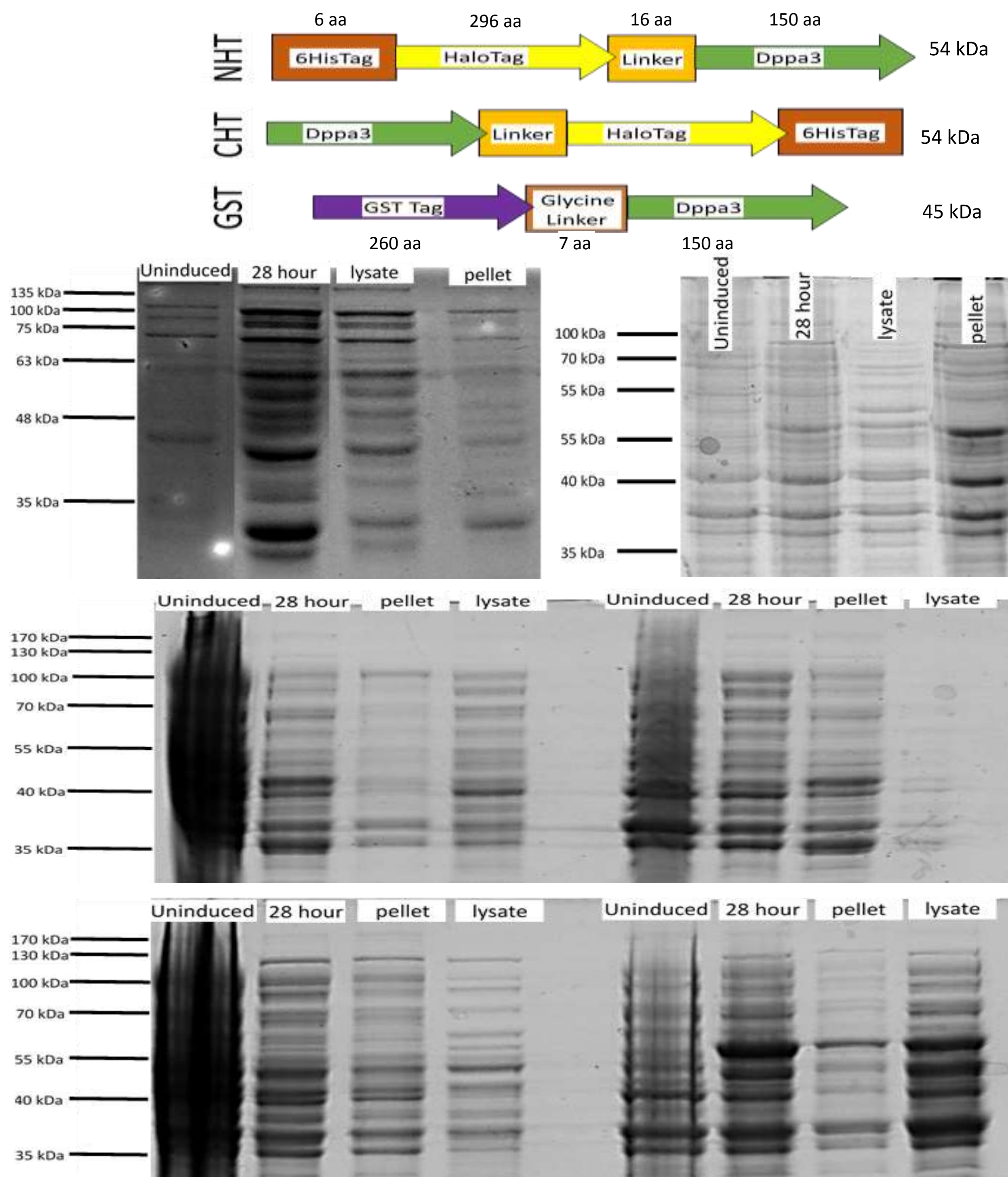

Figure 2: **(A)** Schematic diagram of each of three recombinant Dppa3 protein constructs. Images not representative of each domains size. Molecular weights and domain lengths are shown. **(B)** Induction analysis of NHT Dppa3 in BL21-DE3 *E.coli*. **(C)** Induction analysis of NHT Dppa3 in BL21-DE3 CodonPlus *E. Coli*. **(D)** Induction Analysis of GST Dppa3 in both the BL21-DE3 *E. Coli* (left) and the BL21-DE3 CodonPlus *E. Coli* (right). **(E)** Induction Analysis of CHT Dppa3 in both the BL21-DE3 *E. Coli* (left) and the BL21-DE3 CodonPlus *E. Coli* (right).

DE3 CodonPlus. The CodonPlus strain contained genes for mammalian tRNA synthases for assisting translation of the protein. Cultures were induced at 25C for 28 hours. We found that the GST tag did not express well in either strain of *E. coli* tested (Figure 2B). The NHT Dppa3 expressed at detectable levels in the BL21-DE3 strain and expressed well in the CodonPlus strain. Unfortunately, in the CodonPlus strain the NHT Dppa3 remained caught in the insoluble fraction following cell lysis (Figure 2C). Finally the CHT did not express in the standard BL21-DE3 strain, but expressed well in the CodonPlus strain. Furthermore the CHT Dppa3 remained in the soluble fragment following cell lysis (Figure 2D).

N-terminus HaloHis Tagged Dppa3 Purification:

Based on the induction results, we chose to proceed with the purification from the BL21-DE3 strain. Induction of the NHT Dppa3 was first confirmed via western blot with an anti-Dppa3 polyclonal antibody (Figure 3A). The NHT Dppa3 was then purified from the soluble fraction of the cell lysate using a Ni-agarose column (Figure 3B). Coomassie staining showed that the final product of the purification contained several background bands of both greater and lesser molecular weight than the target. Furthermore a Bradford assay showed that there was only 0.8mg/mL of total protein present in the final purified product. Following this several alternate induction techniques were used to try and improve the initial induction efficiency. The cells were induced at 18C and 30C for 4, 20, and 28 hours. However none of these parameters were able to improve the induction efficiency of the NHT Dppa3 (data not shown).

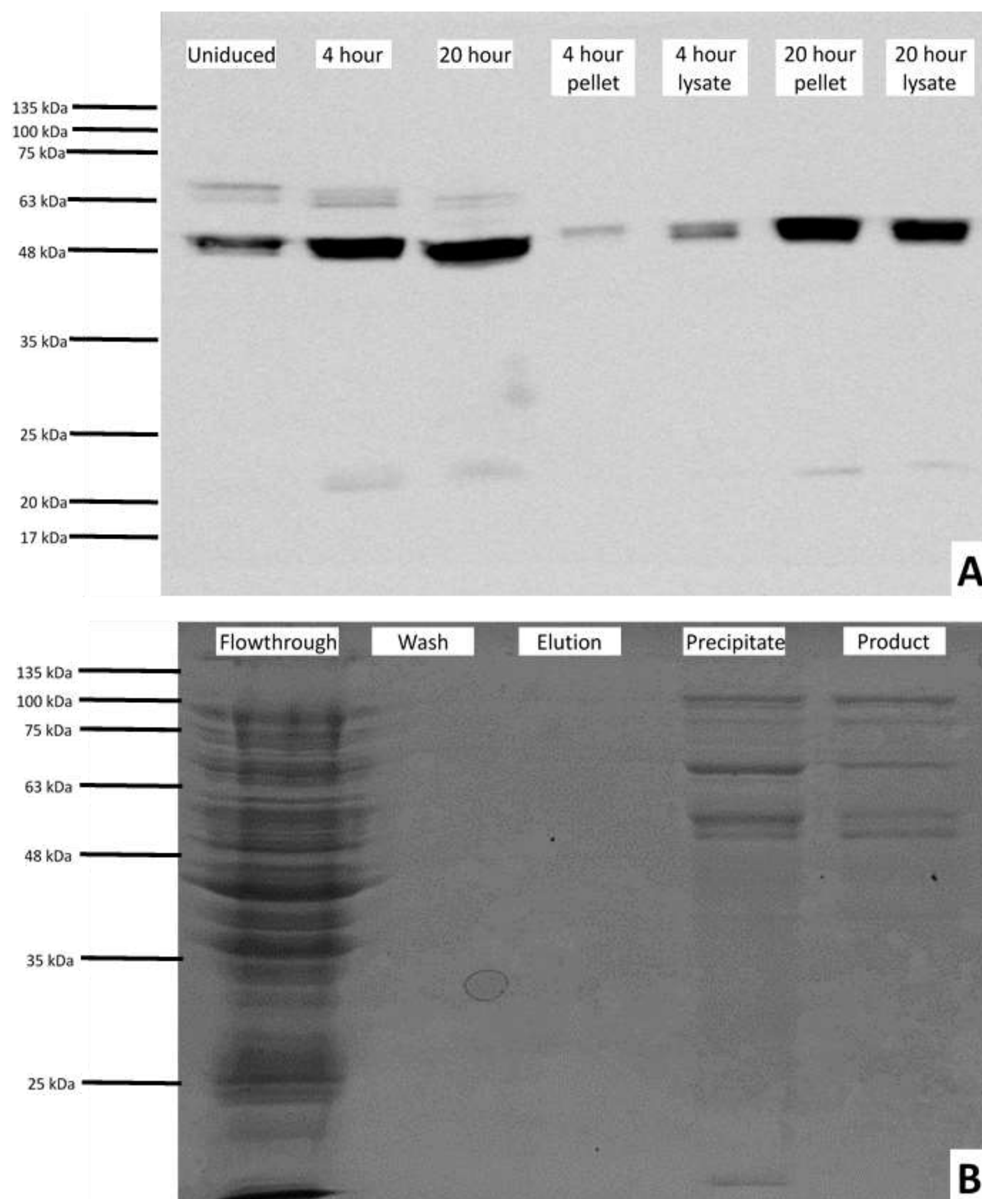

Figure 3: **(A)** Western blot with an anti-Dppa3 primary antibody. Shows samples of the NHT Dppa3 induction from BL21-DE3 *E. Coli*. Shows uninduced cells as well as samples from four and twenty hours of induction. Pellet and lysate refer to the insoluble and soluble components after sonication. **(B)** Coomassie stain of samples taken during purification of NHT Dppa3 from the soluble fragment.

C-terminus HaloHis Tagged Dppa3 Purification:

Following the difficulties with purifying the NHT Dppa3, and based on the good induction results of the CodonPlus CHT Dppa3 strain, we chose to proceed with the purification of CHT Dppa3. We began by confirming that the observed band at 54kDa (Figure 2D) was in fact Dppa3. A western blot for Dppa3 showed that there were three bands that stained as Dppa3; one at ~54kD, one at ~50kDa, and one at ~37kDa (Figure 4B). We predicted that despite the high prevalence of the 37kDa band, that the lower two molecular weight products would be removed following purification. A comparable three bands appeared again in the final product following purification (Figure 4A). To determine what these bands were, three western blots were performed. The first was stained with an anti-Dppa3 polyclonal antibody. This showed that in the final product both the ~54kDa and the ~37kDa band were positive for the Dppa3 epitope (Figure 4B). The second western blot was stained with an anti-HaloTag polyclonal antibody. This showed that, in the purified product, only the 37kDa band was positive for the HaloTag epitope (Figure 4C). Finally we stained the third western blot with an anti-6xHisTag polyclonal antibody. Similar to the anti-HaloTag blot, this western showed that only the ~37kDa band was positive for the HisTag (Figure 4D).

Doxycycline Inducible Flag Tagged Dppa3 ESC line construction:

In order to address our second question, what proteins are associated with Dppa3, we prepared a flag tagged Dppa3 under the control of a doxycycline inducible promoter, and electoporated this construct into V6.5 ESCs. Following characterization, flag-Dppa3 will be salt extracted from the cells, treated with micrococcal nuclease, and immunoprecipitated. This product will be sent out for mass spectrometry analysis to identify proteins associated with Dppa3 in ESCs. To begin,

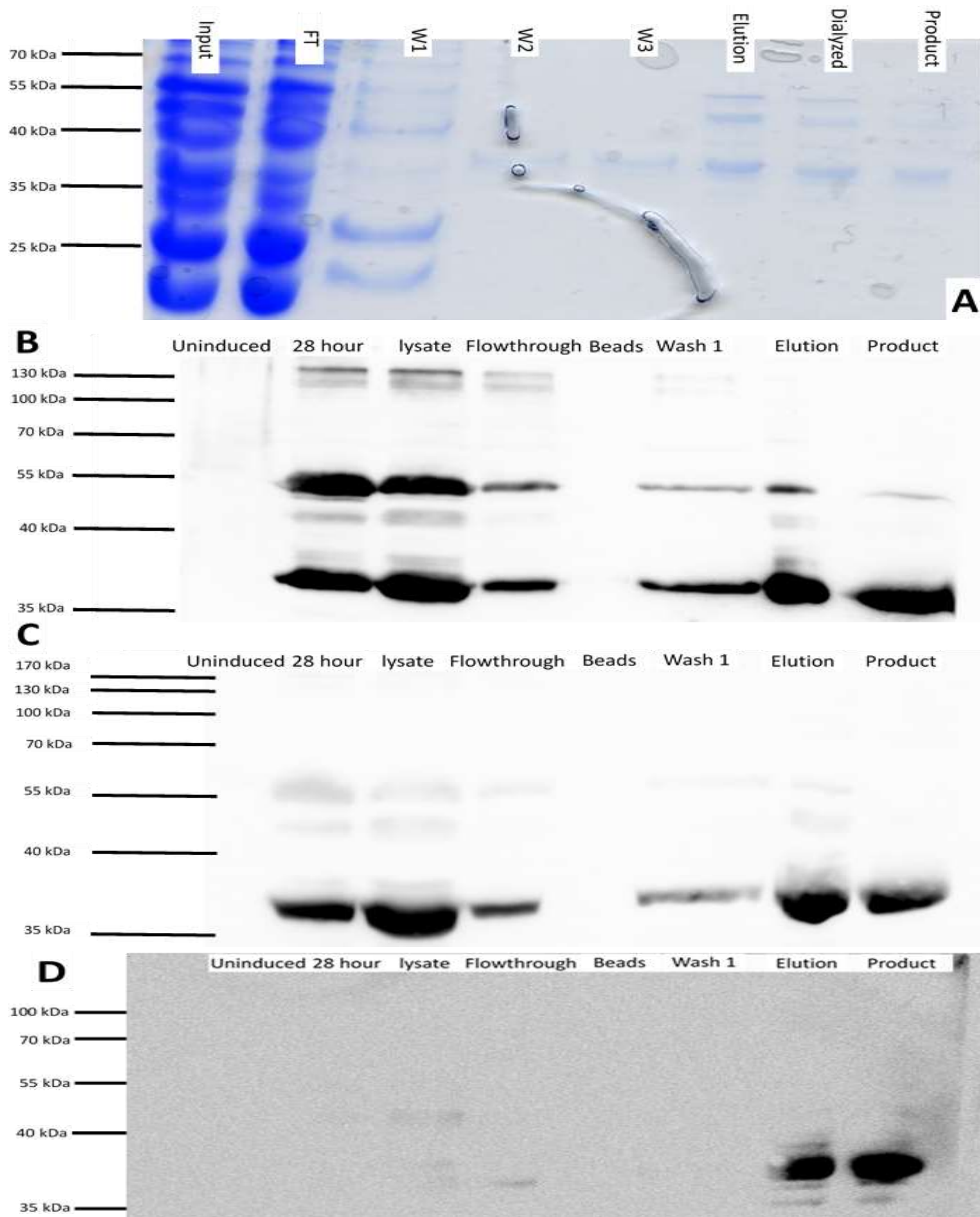

Figure 4: **(A)** Coomassie staining of purification samples. Final product bands at 37kDa, 50kDa, and 54kDa. **(B)** anti-Dppa3 western blot of purification samples. Final product bands at 37kDa and 54kDa. **(C)** anti-HaloTag western blot of purification samples. Final product at 37kDa. **(D)** anti-HisTag western blot of purification samples. Final product at 37kDa.

cells were treated with a titration of Dox concentrations. Whole cell extracts were analyzed via western blot and immunofluorescence, with an anti-Dppa3 and an anti-Flag primary antibody, to assess the correct dox concentration where the endogenous and flag tagged Dppa3's were expressed similarly. Western blot showed that this concentration was approximately 0.125μg/mL Dox (Figure 5A). Immunofluorescence confirmed that the culture was homogenous; all cells were expressing the flag tagged Dppa3 at comparable levels (Figure 5B). We also observed that at higher doxycycline concentrations the flag-Dppa3 began to relocalize into the cytoplasm from the nucleus. It is unclear whether this is an artifact of overexpression or a regulatory reaction to the higher Dppa3 concentration in the nucleus. From here, cells were treated with dox at 0.125μg/mL for 3 days. Following this, cells were lysed with a dounce homogenizer, and extracted at a gradient of salt concentrations. A western blot for Dppa3 was performed on each sample in order to assess whether Dppa3 was still present in the sample, or had been removed (Figure 5D). We found that at 0.42M NaCl, Dppa3 was still present in the pellet. This is the concentration typically used to extract transcription factors. Since this failed to remove the Dppa3 from the pellet, it suggests that Dppa3 is chromatin bound.

Discussion:

Given the difficulties that we have had successfully completing our purifications, both from the ESC line, and from *E. coli*, it is difficult to draw specific conclusions at this time. However there are several key points that can still be made. Our *in vitro* work has identified a product in the CHT Dppa3 expressing BL21-DE3 CodonPlus *E. coli* strain. This second product contains both the His and Halo tag epitopes, as well as the Dppa3 epitope. This is, at present, remarkably difficult to explain. In order for the band to stain positive as Dppa3, the Dppa3

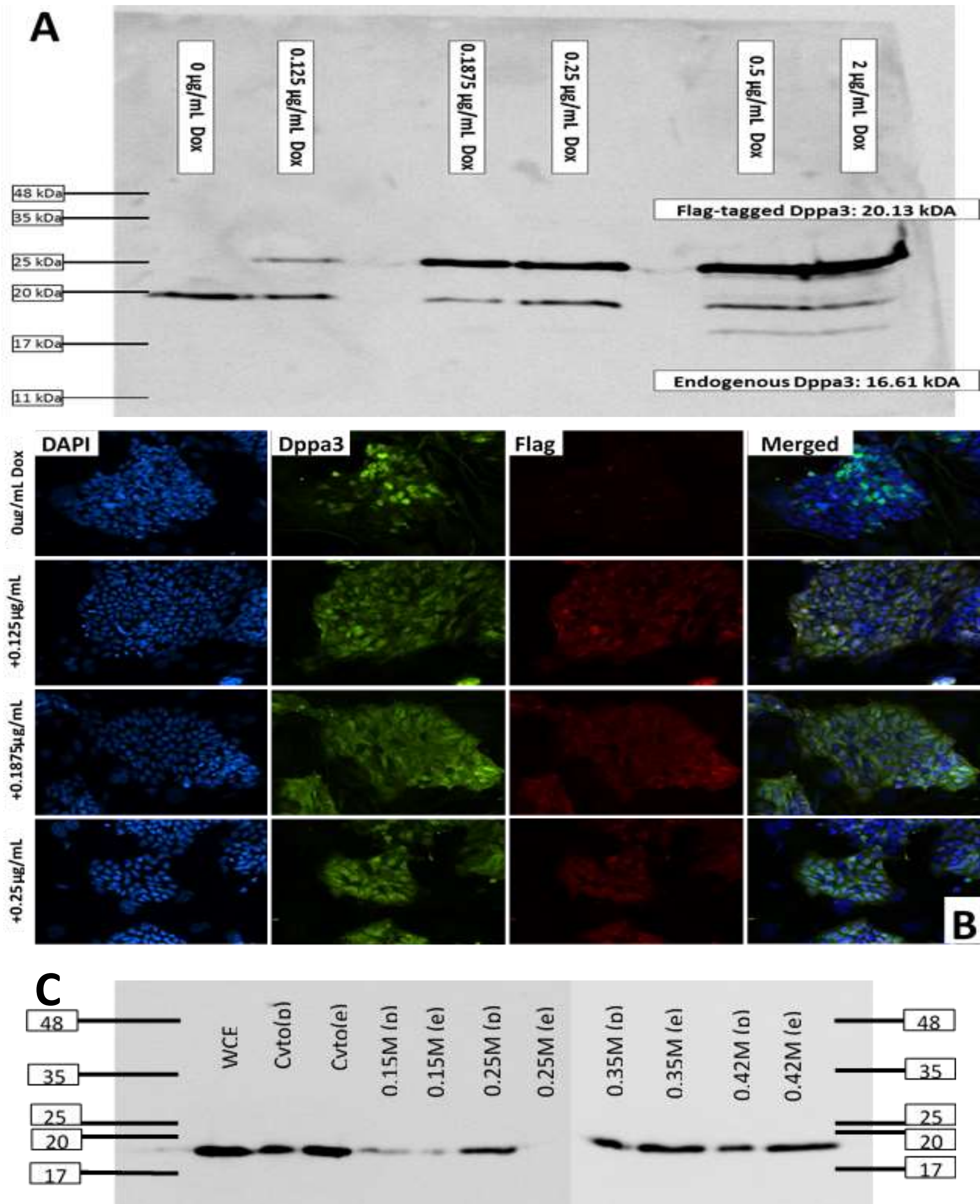

Figure 5: **(A)** anti-Dppa3 western blot of whole cell extracts taken from ESC line. Samples grown in progressively higher concentrations of doxycycline. Upper bands are flag tagged Dppa3, lower bands are endogenous. **(B)** Immunofluorescence stains of ESC line. Samples are grown in progressively higher concentrations of doxycycline. **(C)** anti-Dppa3 western blot, of extracts taken from ESC line. Extraction performed in steadily increasing concentrations of sodium chloride. Both pellet and extract are shown.

epitope must be present. This epitope is all the way at the N-terminus end of the protein. Meanwhile the HisTag epitope is all the way at the C-terminus end. Therefore in order to contain both epitopes the entire 54kDa protein should be present. Furthermore there is no way that a single internal protease site could result in a single band at 37kDa that contains both epitope. This makes identifying this product exceptionally difficult. At this time, our explanation of this product is that a second plasmid may have been transformed into the *E. coli* along with the CHT Dppa3 expressing plasmid. This second plasmid could be producing an alternate product that contains all three epitopes. To correct for this we are currently repeating the transformation of the strain, with a fresh preparation of the plasmid that we have confirmed to be free of contamination.

As for the purification of the flag-tagged Dppa3 from the ESC line, methods for effective purification of the product are currently being devised in our lab. When these preliminary studies are completed we will continue with this purification. Following identification of Dppa3 associated products in ESCs, we will use this cell line to construct a chimera mouse that contains both the flag-tagged Dppa3, as well as an inducible system for iPSC reprogramming. From these mice we will isolate embryonic fibroblasts (MEFs) and perform reprogramming experiments. At each time point of the reprogramming we plan to purify Dppa3 and use mass spectrometry to test for associated proteins. In addition to this we will use chromatin immunoprecipitation and sequencing (ChIP-seq) to identify where on the genome Dppa3 is bound at each time point of the reprogramming process.

Here we have presented preliminary work in the purification of Dppa3 from both ESCs and in a recombinant form from *E coli.* While we have difficulties with these purifications that have prevented the completion of functional assays, this data represents the first step of a much

larger project to characterize the role of Dppa3 in chromatin regulation during iPSC reprogramming.

References:

1. Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature *502*, 462–471.

2. Plath, K. & Lowry, W. E. (2011). Progress in understanding reprogramming to the induced pluripotent state. *Nat. Rev. Genet.* **12**, 253–265

3. Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nature Cell Biology *15*, 872–882.

4. Tran, K., Jackson, S., P. G. Olfus, Z., Zaidan, N. Z., Leng, N., Kendziorski, C., Roy, S., Sridharan, R. (2015). Collaborative rewiring of the pluripotency network by chromatin and signalling modulating pathways. *Nature Communications 6.*

5. Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2012). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nature Genetics *45*, 34–42.

6. Kang et al. 2013. PGC7, H3K9me2 and Tet3: regulators of DNA methylation in zygotes. Cell Research 23, 6-9.

7. Nakamura et al. 2012. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature 486, 415-420

8. Xu et al. 2015 Dppa3 expression is critical for generation of fully reprogrammed iPS cells and maintenance of *Dlk1-Dio3* imprinting. Nature Communications.

9. Promega. (2011). HaloTag® Technology: Driving mammalian protein expression forward. *Promega*. http://www.promega.com/products/pm/halotag-technology/halotag-technology/