RICE UNIVERSITY

Identification of a novel PRC2 recruiter in mammalian cells

by

Yufeng Gou

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Master of Science

APPROVED, THESIS COMMITTEE

Jianpeng Ma, Chair
Professor, Department of Bioengineering

Michael Diehl
Associate Professor, Department of Bioengineering

Yizhi Tao
Associate Professor, Department of Biochemistry and Cell Biology

Qinghua Wang
Assistant Professor, Department of Biochemistry and Molecular biology BCM

HOUSTON, TEXAS
December 2012
ABSTRACT

Identification of a novel PRC2 recruiter in mammalian cells

by

Yufeng Gou

Polycomb repressive complex (PRC) 2 functions to repress thousands of target genes, and they are responsible for stem cell differentiation and carcinogenesis. However, how PRC2 are recruited to specific regions of their target genes remains elusive. In Drosophila, nine sequence-specific transcription factors including Zeste have been shown to act as PRC2 recruiters, but little is known about their homologues in mammalian cells, as a straightforward homology search failed to work in most cases. Aided by three-dimensional structure prediction and the use of the genomes of intermediate bridging species, we have identified a new protein, Zeste Homologue in humans (ZH), as the human homologue of Drosophila Zeste. Gel shift assays indicated that ZH binds to Zeste recognition sequence via its N-terminal DNA binding domain. ZH physically interacted with the components of PRC2 in GST-pull down assays. Chip-seq and Chip-qPCR experiments showed the co-localization of ZH and PRC2 complex. Together, these findings revealed the critical function of ZH in recruiting PRC2 complexes to their target genes.
Acknowledgments

Firstly, I want to thank my advisor, Dr. Jianpeng Ma and Dr. Qinghua Wang who support me from the very beginning I entered the lab, especially when my projects had difficulties. They are not only my best mentors but also are my good friends in my scientific career.

I also want to thank all my committee members who contribute lots in my projects direction and thesis refinement. They always show me their critic thinking in science which taught me lots.

I want to thank all the lab mates in Dr. Ma and Dr. Wang’s labs. All of them are my good friends and help me lots in my work. Without their kindly help, I cannot make good progress on my projects.

Finally, I want to thank my parents and my family. Without their support, I won’t choose scientist as my career.
# Contents

Acknowledgments ......................................................................................................................... iii

Contents ........................................................................................................................................ iv

List of Figures ................................................................................................................................. v

List of Tables .................................................................................................................................. vi

Nomenclature ................................................................................................................................. vii

Introduction ..................................................................................................................................... 1

Material and Methods .................................................................................................................. 4

2.1. Cloning, expression and purification of proteins ................................................................. 4

2.2. Kit, reagent and antibodies .................................................................................................. 5

2.3. GST pull down assay ............................................................................................................. 5

2.4. Gel Shift Assays .................................................................................................................... 5

2.5. Chromatin Immunoprecipitation-sequencing .................................................................... 6

Results and Discussion ................................................................................................................ 7

3.1. Computational prediction of Drosophila Zeste Homologue (ZH) in human genome .......... 7

3.2. ZH can interact with core subunits of PRC2 complex ...................................................... 10

3.3. ZH can bind to a specific DNA element, the Zeste binding site in vitro ......................... 12

3.4. Most of the ZH binding sites can occupied by PRC2 ..................................................... 14

3.5. Conclusion and Discussion .................................................................................................. 16

References ..................................................................................................................................... 18
List of Figures

Figure 3.1 – Alignment of ZH and Zeste proteins................................................................. 8

Figure 3.2 – ZH binds PRC2 complex by directly interacting with EED with its N terminal domain....................................................................................................................... 11

Figure 3.3 – ZH binds to a specific DNA element, a zeste binding site....................... 13

Figure 3.4 – Chip-qpcr results of anti-ZH and anti-EZH2 in human A549 cells 16

Figure 3.5 – The final model of ZH as a PRC2 recruiter in mammalian cells .... 17
List of Tables

Table 3.4 - ZH binding fragments and its overlapping with H3K27me3 identified by chip-seq ........................................................................................................................................................................ 15
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRC2</td>
<td>Polycomb Repressive Complex 2</td>
</tr>
<tr>
<td>ZH</td>
<td>Zeste Homologue</td>
</tr>
<tr>
<td>CHIP</td>
<td>Chromatin Immuno-Precipitation</td>
</tr>
</tbody>
</table>
Proper regulation of gene expression is very important for cell fate decision, environmental response and phenomena inheritance. In multicellular organisms, the gene regulation is essential to trigger cell differentiation and maintain cell morphogenesis. Different gene expression profiles in cells create different cell types with different functions to make sure the organism works systematically. Any misregulation of gene expression leads to dysfunction of cells even the death of the organism.

Gene regulation is a complicated process at multiple levels including transcription regulation, post-transcriptional modification, RNA transportation and translation, in which epigenetic regulation of gene expression plays very important roles. Given the same genetic profile for a multicellular organism, epigenetic regulation decides what type of cells to be differentiated into. Abnormal regulation
of genes at epigenetic level is the key factor for many diseases including developmental disorder and cancer.

Epigenetic regulation is the modification of genome without changing the DNA sequences, which mainly includes DNA modifications and histone modifications. Histone H3K27 and H3K4 are the two key sites of histone modifications, which can directly regulate the expression status of their target genes. The trimethylation of H3K27 can repress its target gene expression while the trimethylation of H3K4 can activate its target gene expression. In stem cells, most housekeeping genes have both H3K4 and H3K27 modifications, thus cooperation of those two markers is essential to control the genes’ expression.

Polycomb repressive complex 2 (PRC2) is a histone methyltransferase complex which can initially trimethylate H3K27. This complex has three core elements, EED, EZH2 and SUZ12, in which EZH2 is the key trimethylase. All of those elements are highly conserved from yeast to human.

Numerous genes in vivo are regulated by its trimethylation status of H3K27, which requires the regulation accurately and specifically. However the key enzyme complex of H3K27me3, PRC2, cannot directly bind to DNA. How PRC2 is recruited to its target genes specifically is still unclear till now. In Drosophila, four well studied transcription factors, Pho/Phol, Gaf, Zeste and Psq can recruit PRC2 to its target genes. Only two of them, Pho/Phol and Gaf, have their homologues identified in mammalian cells. Given that the gene coverage of PRC2 is much wider than that of
the identified transcription factors, it is necessary to study other possible transcription factors which can recruit PRC2 complex specifically to its target genes.

We identified a new protein named Zeste Homologue (ZH) which is a possible mammalian homologue of Zeste by computational prediction and alignments. Given that Zeste can recruit PRC2 in Drosophila, we hypothesize that ZH is a potential PRC2 recruiter in mammalian cells. To test this hypothesis, we validated the DNA binding ability of ZH and its interaction with PRC2 complex both in vivo and in vitro. Most DNA binding elements of ZH identified by Chip-seq assay co-localized with PRC2 complex in human cells. Based on the results of all these experiments, we conclude that ZH is a new PRC2 recruiter in mammalian cells.
Chapter 2

Material and Methods

2.1. Cloning, expression and purification of proteins

The full length ZH and its N-terminal and C-terminal domains were amplified from human cDNA library, and were cloned into pOPINE plasmid with GST tag. The proteins were expressed in Rocetta cells at 18°C for overnight with 1mM IPTG induction. Cells were lysed in lysis buffer (10mM HEPES 7.6, 300mM KCl, 10%Glycerol); Proteins were purified using GST beads and eluted in elution buffer (10mM HEPES 7.6, 300mM KCl, 10%Glycerol). For the proteins without GST tag, the GST was removed by PreScission Protease.

The EED protein was constructed into pGex6p-1 with GST tag on its N terminal. The protein were expressed at 16°C for overnight and purified by GST beads. PreScission Protease was used to remove the GST tag when necessary.
2.2. Kit, reagent and antibodies

Gel shift kit from Roche (Cat.03353591910) was used to perform the gel shift assay. Anti Ezh2 antibody is from BD Biosciences (Cat.612666); anti suz12 antibody is from Santa Cruz Biotechnology (SC-67105); anti EED antibody is from Millipore (Cat.05-1320); anti ZH antibody is from Abcam.

2.3. GST pull down assay

The purified GST-tag proteins were bound to the GST beads first and then incubated with cell lysates for overnight. Beads were collected and washed by washing buffer (20mM Hepes 7.6, 300mM NaCl and 10%Glycerol). The final proteins were eluted by elution buffer (20mM Tris-HCl 8.0, 10%SDS) followed by running SDS-PAGE and doing Western Blots.

2.4. Gel Shift Assays

The gel shift assay was performed following the protocol listed in the kit. Briefly, the DNA element was synthesized and labeled by DIG, and then incubated with purified proteins at 4C for at least 30mins. Native running buffer was used to help loading the mixture into native DNA PAGE gel. After running the PAGE, the gel was used to transfer its contents onto membrane which is used to detect the DIG signal by anti-DIG antibody.
2.5. Chromatin Immunoprecipitation-sequencing

The chromatin Immuno-precipitation experiments follow the protocol of reference of Nature Protocols with minors’ modification at the shearing DNA step. Human A549 cells and mouse P19 cells were cross-linked for 15min and then treated by lysis buffers to broken the cells. The chromatin DNA was sheared by sonication for 40 cycles to get the proper DNA elements, which was incubated with antibody-conjugated magnetic beads for overnight at 4C. The antibody-DNA complex was washed and eluted followed by reverse-crosslinking and RNA digestion. The final DNA was purified by phenol:chloroform extraction and EtOH precipitation.

The chip-seq data was processed on GALAXY which uses SICER to annotate the peaks with cutoff of p value $10^{-5}$, and false discovery rate 0.001. Annotated peaks were further analyzed in Genome Brower by comparing with exist histone modification peaks.
Chapter 3

Results and Discussion

3.1. Computational prediction of Drosophila Zeste Homologue (ZH) in human genome

Accurate recruitment of PRC2 complex to its target genes is essential for cell functions. But how PRC2 is recruited in mammalian cells is unknown. In Drosophila, four well studied transcription factors including Zeste have been identified as PRC2 recruiters. However in mammalian cells, only YY1 which is the homologue of Drosophila Pho protein has been identified as PRC2 recruiter. Here, we use primary sequence alignment and predicted 3D structure alignment to search Drosophila Zeste homologue in mammalian cells.

Protein Zeste has two separate domains, DNA binding domain ranging from amino acid 47-138 and Coiled-coil domain from amino acid 500-575. Both of those
domains are conserved across *Drosophila* genre. Because of the requirement of DNA binding ability for PRC2 recruiters, the DNA binding domain of Zeste was used as the Query in BLAST to search its similar proteins in human protein database. The highly similar protein named ZH was identified in this process.

Primary sequence alignment shown in Figure 3.1 (A) of DNA binding domain of Zeste and N terminal domain of ZH yields 22 identical residues, 28 conserved residues, and 11 semi-conserved residues. The identity percentage is 24% and the similarity percentage is 66%. This indicates the highly similarity between DNA binding domain of Zeste and N terminal domain of ZH at their primary sequence level.

![Fig3-1](image-url)

**Fig3-1.** A. Sequence alignment of ZH N-terminal domain with Zeste DNA binding domain; B. 3D structure prediction alignment of Zeste and ZH DNA binding domains; C. Schematic of ZH fragments.
Furthermore, 3D structures of those two proteins were predicted by I-TASSER. The best model of Zeste was predicted with the highest C-score = -4.29 and its PDB file was used to present the 3D structure of Zeste DNA binding domain. The best model of ZH was predicted with the highest C-score = -1.16, which was used to represent the 3D structure of ZH N terminal domain. Those two structure models were aligned by Pymol and the final result was shown in Figure3.1 (B). Both the DNA binding domain of Zeste and the N terminal domain of ZH are helix-loop-helix. The superimposition of those two proteins indicates a highly similarity at their 3D structure level. The alignment of predicted 3D structures between the two further confirms that the N terminal domain of ZH is similar to the DNA binding domain of Zeste.

The secondary structure of ZH was analyzed by online service of Network Protein Sequence Analysis. The first 140 amino acids at N terminal form helix-loop-helix which is the same as its predicted 3D structure, and the last 90 amino acids at C terminal is a typical coiled-coil domain which is similar to the C terminal domain of Zeste. The schematic domain of ZH is shown in Figure3.1 (C) giving two conserved domains: DNA binding domain and the coiled-coil domain, which is similar to the domain distribution of Drosophila Zeste.

All those computational prediction results shown above indicate that ZH is highly similar to Zeste at primary sequences, 3D structure alignments and domain distribution, indicating that ZH is a possible human homologue of Drosophila Zeste.
Given that Zeste can recruit PRC2 complex in *Drosophila*, we hypothesized that ZH is a PRC2 recruiter in mammalian cells.

### 3.2. ZH can interact with core subunits of PRC2 complex

In order to test the hypothesis that ZH is a PRC2 recruiter in mammalian cells, GST pull down assays were performed to test the interaction between ZH and the core subunits of PRC2: EZH2, SUZ12 and EED.

The full-length, N-terminal and C-terminal domain proteins of ZH were expressed in *E. coli* and purified by GST beads. 30µg purified proteins were incubated with 25µl GST beads each for overnight. The beads-conjugated proteins were then incubated with cell lysates for overnight. Final beads were eluted by SDS running buffer after washing, which were used for western blottings.

The western blotting results of the GST pull down assay with different ZH fragments by anti-EZH2, anti-SUZ12 and anti-EED were shown in figure 3.2 (A). Compared to the GST protein alone, both the full-length and the N-terminal domain of ZH, not the C-terminal domain of ZH can pull down all the three core subunits of PRC2. To confirm the interaction between ZH and PRC2 complex, GST tagged EED was purified and performed the reverse GST pull down assay to detect the accumulation of ZH. Figure 3.2 (B) shows that the EED protein can also pull down ZH, which combined together with the results in Figure 3.2 (A) indicates that ZH can interact with PRC2 by its N terminal domain.
Figure 3.2 – ZH binds PRC2 complex by directly interacting with EED with its N terminal domain. A. GST pull down assay of full-length, N terminal and C terminal domain of ZH with core subunits of PRC2; B. Reverse GST pulldown assay of EED with ZH; C. GST pull down assay of full-length ZH with PRC2 in knocking down EZH2 cells; D. Gel filtration of EED and N terminal domain of ZH complex; E. SDS-PAGE of figure3.2-D.

In order to identify the direct binding proteins in PRC2 with ZH, Ezh2 was knocked down in P19 cells which was used to perform the GST pull down assay by Full-length ZH. The results were shown in Figure3.2 (C). Compared to the wild type P19 cell lysate, knocking down EZH2, the accumulation of EZH2 and SUZ12 was decreased, but the accumulation of EED had no change. This indicates that EED is possible the directly recognized protein by ZH in PRC2.
Purified EED and purified N terminal domain of ZH were incubated at 4C for overnight, and then were loaded on the Superose S200 column of FPLC. The elution fraction was shown in Figure3.2 (D) and the corresponding SDS-PAGE results were shown in Figure3.2 (E). The Gel filtration results clearly show that the N terminal domain of ZH can form complex with EED directly in vitro. All the results of Figure 3.2 indicate that ZH can interact with PRC2 complex by directly binding to EED with its N terminal domain.

3.3. ZH can bind to a specific DNA element, the Zeste binding site in vitro

Another aspect to test the function of PRC2 recruitment of ZH is to test the DNA binding ability of ZH. Because of the high similarity between N terminal domain of ZH and the DNA binding domain of Zeste shown in Figure3.3 (A), the Zeste binding consensus sequence was used as the template for ZH binding.

A 22bp double-strand DNA element harboring one Zeste binding site was synthesized and labeled by DIG then was incubated with purified ZH at Room temperature for 30min followed by performing the gel shift assay. Final signal was detected by anti-DIG antibody. When the ZH concentration was increased shown in Figure3.3 (B), the signal of the shift bands was increased too. This means more ZH-DNA complex formed. This result indicates that ZH can bind to a specific DNA element harboring Zeste binding site.
Figure 3.3 – ZH binds to a specific DNA element, a zeste binding site. A. 3D structure prediction alignment of ZESTE and ZH DNA binding domains; B. Gel shift assay of fulllength ZH with Dna element harboring Zeste binding site; C. Schematic of ZH fragments for gel filtration; D. Gel shift assay of different fragments of ZH proteins with Zeste binding site; E. Binding affinity of ZH proteins with Zeste binding site.
Semi-Quantification of the DNA binding efficiency of ZH based on the gel-shift assay was shown in Figure 3.3(B). The shift band with highest ZH concentration was identified as 100% binding efficiency, compared to which the other shift bands with different ZH concentrations were then quantified. The quantified K_d value of full-length ZH-DNA binding was 1.5uM.

To identify the minimum fragments of ZH responsible for the DNA binding, ZH protein was truncated into three parts in Figure 3.3 (C) and they were used for the gel shift assays, in which the F2 fragment had only the DNA binding domain of ZH. Figure3.3 (D) shows that the full-length, F1 and F2 proteins have shift bands in the gel shift assay indicating that the N terminal DNA binding domain of ZH is responsible for the DNA binding. Quantification of the binding efficiencies of different ZH fragments with DNA was shown in Figure3.3 (E). The F3 fragment has no shift band further confirmed that the N terminal domain of ZH has DNA binding ability.

3.4. Most of the ZH binding sites can be occupied by PRC2

The results shown above clearly indicate the PRC2 recruitment function of ZH in vitro. But the target genes regulated by ZH is unknown. To address this problem, Chromatin immunoprecipitation experiments by anti-ZH antibody followed by sequencing was performed to identify the target genes directly regulated by ZH both in human A549 cells and mouse P19 cells.
The chip-seq results were shown in Table 3.4 with significant P value less than 10^{-5} and the False Discovery Rate less than 0.001. There are 972 fragments in human and 829 fragments in mouse identified as the ZH directly binding DNA elements.

<table>
<thead>
<tr>
<th>Species</th>
<th># of ZH binding fragments</th>
<th># of fragments overlapping with H3K27me3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>972</td>
<td>787</td>
</tr>
<tr>
<td>Mouse</td>
<td>829</td>
<td>660</td>
</tr>
</tbody>
</table>

Table 3.4 – ZH binding fragments and its overlapping with H3K27me3 identified by chip-seq.

The identified ZH binding elements were further analyzed with the signal of H3K27me3 markers by Genome Brower database. 787 of the 972 fragments identified in human and 660 of the 829 fragments identified in mouse have the H3K27me3 markers. Given the fact that H3K27me3 is the product of PRC2 complex, nearly 80% of identified DNA sites directly recognized by ZH are overlapped with H3K27me3 marker, which means the co-occurrence of ZH and PRC2 complex on those DNA sites.

To validate the findings discussed above, 10 fragments from the overlapped sites were validated byChip-qPCR experiments. The anti-ZH and anti-EZH2 were
used to immunoprecipitate the chromatin-DNA complex with IgG as the negative control, and then the final DNA was extracted to perform the quantitative PCR assay to detect the accumulation of all the 10 fragments. Figure 3.4 shows the results and indicates that the entire 10 fragments can accumulated in both ZH and EZH2 precipitation samples. This result confirms the co-occurrence of ZH and PRC2 complex in vivo.

Figure 3.4 – Chip-qpcr results of anti-ZH and anti-EZH2 in human A549 cells.

3.5. Conclusion and Discussion

Because of the important function of PRC2 in cell differentiation and cancer development, how PRC2 is recruited to its target genes is essential. Here, we identified a new protein ZH in mammalian cells which is a possible homologue of Drosophila Zeste as a PRC2 recruiter. To test this function of ZH, the interaction between ZH and PRC2 as well as the DNA binding ability of ZH were investigated in
this thesis. GST pull down assay between ZH and the core subunits of PRC2 indicates the interaction of ZH and PRC2 complex. Gel shift assay in vitro showed that ZH can bind to a specific DNA element harboring Zeste binding site. The chip-seq experiment further identified that nearly 80% of the ZH binding sites can also be occupied by PRC2 complex. All those experiments indicate the model shown in Figure 3.5.

![Figure 3.5 - The final model of ZH as a PRC2 recruiter in mammalian cells.](image)

Since ZH is a new protein and nothing is known about it. Further work also needs to do to study its detail mechanism as a PRC2 recruiter as well as to study its molecular function during cell differentiation or cancer development.
References


