Caenorhabditis elegans MES-3 is a highly divergent ortholog of the canonical PRC2 component SUZ12

Berend Snel[1,*], Sander van den Heuvel[2,*], Michael F. Seidl[1,*,#]

[1]Theoretical Biology & Bioinformatics, Department of Biology, Faculty of Science, University, Utrecht, Padualaan 8, 3584 CH Utrecht, the Netherlands.

[2]Developmental Biology, Department of Biology, Faculty of Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands

[#]Corresponding author: Michael F. Seidl, Theoretical Biology & Bioinformatics, Department of Biology, Faculty of Science, University, Utrecht, Padualaan 8, 3584 CH Utrecht, the Netherlands, m.f.seidl@uu.nl

[*]These authors contributed equally

Abstract

Polycomb Repressive Complex 2 (PRC2) catalyzes the mono-, di, and trimethylation of histone protein H3 on lysine 27 (H3K27). Trimethylation of H3K27 is strongly associated with transcriptionally silent chromatin and plays an important role in the regulation of cell identity and developmental gene expression[1]. The functional core of PRC2 is highly conserved in animals and consists of four subunits[1] (**Fig. 1a**). Notably, one of these subunits, SUZ12, has not been identified in the genetic model *Caenorhabditis elegans*, whereas *C. elegans* PRC2 contains the lineage-specific protein MES-3[2,3] (**Fig. 1a**). Here, we demonstrate that MES-3 is in fact a highly divergent ortholog of SUZ12. Unbiased sensitive sequence similarity searches uncovered consistent but insignificant reciprocal best matches between MES-3 and SUZ12, suggesting that these proteins could share a common evolutionary history. We substantiate this hypothesis by directly comparing the predicted structures of SUZ12 and MES-3, which revealed shared protein folds and residues of key domains. Thus, in agreement with the observations in previous genetic and biochemical studies[2,3], we here provide evidence that *C. elegans*, like other animals, contains a diverged yet evolutionary conserved core PRC2.

In animals, the functional core PRC2 is composed of the H2K27 methyltransferase EZH2/1, the H3K27me3 binding protein EED, SUZ12, and RBBP4/7[1] (**Fig. 1b)**. SUZ12 interacts with all members of the core PRC2 to form two distinct lobes[4,5]. The N-terminal region of SUZ12 contains five motifs and domains: the zinc finger binding (ZnB), WD-domain binding 1 (WDB1), C2 domain, zinc finger (Zn), and WD-domain binding 2 (WDB2) (**Fig. 1b**). This region of SUZ12 together with RBBP4/7 forms the targeting lobe that serves as a platform for co-factor binding[4,5]. The C-terminal region of SUZ12 contains a VEFS domain (**Fig. 1b**), which associates with EZH2/1 and EED to form the catalytic lobe of PRC2[4,5]. Thus, SUZ12 is critical for the assembly, integrity, and function of PRC2[4,5], in agreement with the conservation of SUZ12 as a core PRC2 component in animals (**Fig. 1a**). PRC2 in *C. elegans* contains MES-2 (EZH2/1) and MES-6 (EED), but a SUZ12 ortholog was previously not discovered. By contrast, *C. elegans* PRC2 includes MES-3[3], which appeared to lack obvious motifs or sequence similarity to SUZ12 and consequently has been considered a *C. elegans* specific subunit[2]. PRC2 in *C. elegans* is involved in germ line development and gene silencing, and MES-2, MES-3, and MES-6 are required for PRC2 activity[6,7]. Consequently, PRC2 in *C. elegans* and in animals are considered functional analogues, despite a seemingly divergent subunit composition[6]. In-depth sequence comparisons have recently turned up surprising homologies, which prompted us to investigate whether MES-3 could be a highly diverged homolog of SUZ12 instead of a *C. elegans* specific invention.

To identify MES-3 homologs in animals, we used unbiased sensitive profile-vs-profile searches to query the predicted proteome of human with MES-3 and query the worm proteome with SUZ12. Surprisingly, we recovered a consistent but insignificant bidirectional match between SUZ12 and MES-3 (16% identity; **Fig. 1c**) that is located at approximately the same regions in both proteins and covers 223 amino acids in MES-3. This region in SUZ12 spans part of the ZnB motif, the complete WDB1 motif, and most of the C2 domain (**Fig. 1b, c**). Notably, the conserved RBBP4/7 binding site of SUZ12[8] is also present in MES-3 (pos. 108-113; FLxRx[VL]) as well as a conserved glycine (pos. 299) (**Fig. 1c**); a missense mutation of this glycine in *Drosophila* leads to a partial loss-of-function phenotype[9,10]. Therefore, we conclude that the N-terminal region of SUZ12 and MES-3 shares extended sequence

similarity including residues previously shown to be critical for function, suggesting that these two proteins are homologs. However, the profile-to-profile searches did not detect similarity between the C-terminal sequence of MES-3 and the SUZ12 domain that mediates E(Z)H2 and EED interaction[4,5] (**Fig. 1b**).

Protein structure is typically more conserved than primary sequence and better allows detection of diverged homologs. Since the protein structure of MES-3 is not yet experimentally resolved, we used deep-learning driven protein structure prediction of both MES-3 and SUZ12. The SUZ12 structure has six functional motifs and domains that were predicted with high precision as they resemble the experimentally determined structure (RMSD = 0.56-1.14; global TM-score = 0.70; global Dali Z-score = 14.8 **Fig. S1a-e**). Like SUZ12, the predicted MES-3 structure is partially disordered (**Fig. 1d; S1f-h**), but nevertheless has a globular N-terminal region mainly formed by β-sheets and a C-terminal region mainly formed by α-helices (**Fig. 1d, e**), and both regions were modelled with high confidence (**Fig. S1g**). Interestingly, the C2 domain of SUZ12 shares significant structural similarity with the N-terminal structural regions of MES-3 (**Fig. 1d, e; Fig. S1i**; RMSD = 1.607; TM-score = 0.60; Dali Z-score = 11.6), corroborating our profile-vs-profile results (**Fig. 1c**). The structural similarity (MES-3, pos. 150-365) extends beyond the region of shared sequence similarity identified above (MES-3, pos. 150-312), and thus encompasses the complete C2 domain (**Fig. 1d; Fig. S1i**). Nevertheless, we also observed some differences in the predicted structures such as the occurrence of an unmatched alpha helix in MES-3 (**Fig. 1e; Fig. S1i**) or the absence of amino acids in MES-3 known to be involved in the interaction between SUZ12 and RBBP4/7 (e.g., SUZ12: R196[5]).

Likewise, we observed structural similarity between the C-terminal domain of MES-3 and the VEFS domain in SUZ12 (**Fig. 1b, d, f**; RMSD = 3.676; TM-score = 0.55; Dali Z-score = 8.3). The MES-3 VEFS-like region is considerably shorter compared with SUZ12 and lacks amino acids that are thought to be involved in the stimulation of histone methyltransferase activity (SUZ12, pos. 580 to 612[10]) and specifically SUZ12 E610 and K611[10], which are invariant in plants, animals, and fungi (**Fig. S1j**). By contrast, several bulky or hydrophobic aromatic residues whose deletion impacts PRC2 assembly[9,10] are conserved, e.g., SUZ12 pos. F639, I647, L652, and F656 can be aligned to identical residues in

85 superposition of the SUZ12 and MES3-VEFS predicted structures (**Fig. S1j**). This suggests that even
86 though the overall sequence similarity is very low, the VEFS domain is overall well conserved in MES-
87 3.

88 Our sequence and structural similarity searches, however, were not able to detect the Zn domain in
89 MES-3 (**Fig. 1b)**, which is normally one of the easiest to identify domains. The absence of Zn is
90 unanticipated as Zn and ZnB in SUZ12 form an intramolecular contact that interact with the accessory
91 PRC2 subunit JARID2[5]. JARID2 contributes PRC2 targeting in embryonic stem cells[5], and even though
92 SUZ12 Zn is not required for methyltransferase activity *in vitro*, Zn is required for PRC2 nucleosome
93 binding *in vivo*, likely by mediating SUZ12-JARID2 interactions[5,10]. Similarly, we were not able to
94 detect WDB2 (**Fig. 1b**), which together with WDB1 in SUZ12 is closely associated with RBBP4/7 to
95 form the targeting lobe[5]. Thus, while the contact surface of SUZ12 with the core subunits EZH2/1 and
96 EED seems highly conserved in MES-3, MES-3 seems to lack specific elements of the areas interacting
97 with RBBP4/7 and accessory subunits that characterize the two subcomplexes in mammals.

98 We conclude that MES-3, even though diverged, structurally resembles SUZ12 in two large regions
99 that are involved in mediating EZH2/1, EED, and RBBP4/7 binding, and it is therefore conceivable
100 that, similarly to SUZ12, MES-3 is critical in assembling and maintaining a functional PRC2[4,5]. The
101 here uncovered sequence and structural similarities as well as the peculiar complementary phylogenetic
102 profiles strongly suggest that MES-3 and SUZ12 are in fact orthologs, albeit that MES-3 has undergone
103 rapid sequence divergence and loss of crucial amino acid motifs as well as the Zn domain. However,
104 further *C. elegans* specific evolution of the PRC2 assembly and architecture is likely to also play a role.
105 For instance, it has previously been noted that MES-2 lacks a region that in EZH2/1 mediates SUZ12
106 binding[6], and thus it can be anticipated that MES-2 evolved an additional compensatory mechanism to
107 partake in PRC2 formation and function. The here described similarities and differences between
108 SUZ12 and MES-3 should facilitate further experiments to elucidate the specific mechanisms by which
109 MES-3 acts in PRC2 in *C. elegans*. Our work joins a rapidly growing set of *in silico* predictions of
110 previously undetected homologies made possible by unprecedented advances in deep-learning driven
111 structure prediction.

ACKNOWLEDGEMENTS

We would like to thank Danny Hancock for the constructing the phylogenetic profiles of PRC2 core members.

AUTHORS CONTRIBUTION

B.S., S.v.d.H, and M.F.S. conceived the study, performed the experiments, analyzed the data, and drafted the manuscript.

DECLARATION OF INTERESTS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

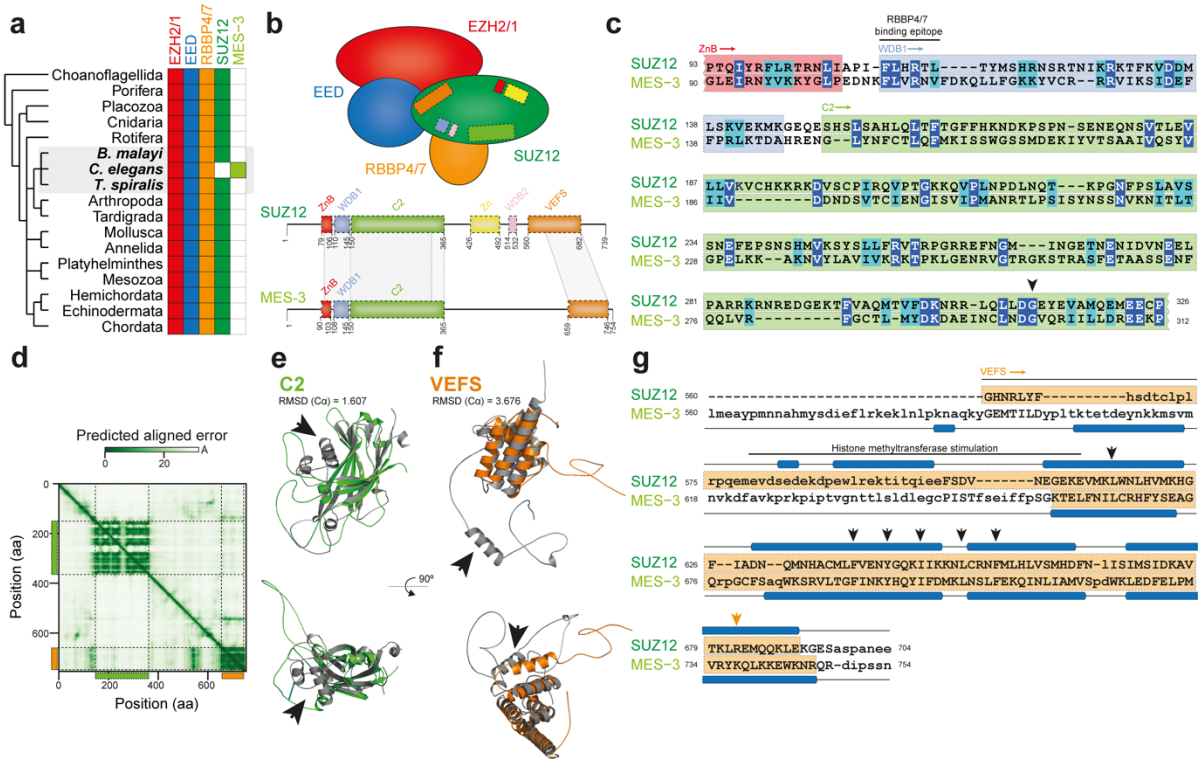

Figure 1 - MES-3 is a highly divergent ortholog of the canonical Polycomb Repressive Complex 2 component SUZ12. a. The Polycomb Repressive Complex 2 (PRC2) core components EZH2/1, EED, RBBP4/7, and SUZ12 are conserved in a broad range of metazoans; the presence of orthologs is indicated by filled boxes. Notably, based on sequence similarity searches, an ortholog of SUZ12 is absent in the nematode model species *Caenorhabditis elegans*, but present in other, closely related nematodes (*Brugia malayi* and *Trichinella spiralis*). *C. elegans* encodes the PRC2 core component MES-3[2,3] that lack obvious motifs or sequence similarity to SUZ12[2]. **b**. Schematic representation of the composition of the core PRC2. The zinc finger binding (ZnB; red), WD-domain binding 1 (WDB1; blue), C2 domain (green), zinc finger (Zn; yellow), WD-domain binding 2 (WDB2; pink), and VEFS (orange) motifs or domains involved in SUZ12 protein-interactions are shown in the schematic as well as along the protein sequence[4,5]. Schematic representation of the protein sequence of MES-3 is shown, and regions of here uncovered sequence (**c**) and structural (**e, f**) similarity are highlighted. **c**. Protein sequence alignment between the N-terminal region of SUZ12 and MES-3, as identified by sensitive profile-vs-profile sequence similarity searches, covers part of the zinc finger binding (ZnB; red), WD-domain binding 1 (WDB1; blue), and C2 domain (green). The conserved RBBP4/7 binding epitope as well as Gly299 are highlighted[8–10]. Identical amino acids are shown in blue and biochemically similar

amino acids are shown in turquoise. **d**. The predicted aligned error (in Å; based on model 2 ptm) of the MES-3 structure is shown as a heatmap and reveals two separated globular regions in the N- and C-terminus, the former overlaps with the profile-vs-profile match (**c**) and corresponds to the C2 domain of SUZ12 (**e; Fig. S1i;** RMSD = 1.607) while the latter overlaps with the regions that structurally resemble the VEFS domain (**f; Fig. S1j;** RMSD = 3.676). The black arrows (**e, f**) highlight regions that differ considerably between SUZ12 and MES-3 (**Fig. S1i, j**). **g.** Sequence-independent structure alignment of the VEFS regions of SUZ12 and MES-3 reveals significantly structural similarity (Dali Z-score = 8.3; TM-score = 0.55), especially along the alpha helices in the C-terminus; a region previously shown to stimulate histone methyltransferase activity in SUZ12[9] (pos. 580 to 612) is highlighted by a black bar, and individual amino acids important for PRC2 assembly[9] are shown by black arrows.

151 **REFERENCES**

152

153 1. Simon, J.A., and Kingston, R.E. (2009). Mechanisms of Polycomb gene silencing: Knowns and
154 unknowns. Nat Rev Mol Cell Bio *10*, 697–708.

155 2. Paulsen, J.E., Capowski, E.E., and Strome, S. (1995). Phenotypic and molecular analysis of *mes-3*,
156 a maternal-effect gene required for proliferation and viability of the germ line in *C. elegans*. Genetics
157 *141*, 1383–1398.

158 3. Xu, L., Fong, Y., and Strome, S. (2001). The *Caenorhabditis elegans* maternal-effect sterile
159 proteins, MES-2, MES-3, and MES-6, are associated in a complex in embryos. Proc National Acad
160 Sci *98*, 5061–5066.

161 4. Kasinath, V., Faini, M., Poepsel, S., Reif, D., Feng, X.A., Stjepanovic, G., Aebersold, R., and
162 Nogales, E. (2018). Structures of human PRC2 with its cofactors AEBP2 and JARID2. Science *359*,
163 940–944.

164 5. Chen, S., Jiao, L., Shubbar, M., Yang, X., and Liu, X. (2018). Unique structural platforms of Suz12
165 dictate distinct classes of PRC2 for chromatin binding. Mol Cell *69*, 840-852.e5.

166 6. Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit
167 contributions to histone methyltransferase activities of fly and worm Polycomb group complexes. Mol
168 Cell Biol *25*, 6857–6868.

169 7. Bender, L.B., Cao, R., Zhang, Y., and Strome, S. (2004). The MES-2/MES-3/MES-6 complex and
170 regulation of histone H3 methylation in *C. elegans*. Curr Biol *14*, 1639–1643.

171 8. Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess,
172 D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin
173 marks. Mol Cell *42*, 330–341.

174 9. Birve, A., Sengupta, A.K., Beuchle, D., Larsson, J., Kennison, J.A., Rasmuson-Lestander, A., and
175 Müller, J. (2001). Su(z)12 , a novel *Drosophila* Polycomb group gene that is conserved in vertebrates
176 and plants. Development *128*, 3371–3379.

177 10. Rai, A.N., Vargas, M.L., Wang, L., Andersen, E.F., Miller, E.L., and Simon, J.A. (2013).
178 Elements of the polycomb repressor SU(Z)12 needed for histone H3-K27 methylation, the interface
179 with E(Z), and *in vivo* function. Mol Cell Biol *33*, 4844–4856.

180