Characterization of the Polycomb group member super sex sombs (sxc)

GAMBETTA, Maria Cristina

Abstract
Polycomb group (PcG) proteins are conserved transcriptional repressors that cooperate to stably maintain the silenced state of several developmental regulator genes, in cells in which these genes should not be expressed. Thus-far characterized PcG proteins have the ability to interact with and/or modify chromatin. We have found that the previously uncharacterized Drosophila PcG gene super sex combs (sxc) encodes Ogt, the highly conserved glycosyltransferase that catalyzes the addition of N-acetylglucosamine (GlcNAc) sugar residues to proteins in the nucleus and cytosol. GlcNAc-modified proteins can be detected on Drosophila chromatin at sites of PcG protein binding. Furthermore, the PcG protein Polyhomeotic (Ph) is glycosylated by Sxc/Ogt in vivo. sxc/Ogt-null mutants lack nuclear GlcNAcylation, and fail to maintain Polycomb transcriptional repression even though PcG protein complexes are still bound to the chromatin of their target genes. Polycomb repression appears to be a critical function of Sxc/Ogt in Drosophila and may be mediated by the glycosylation of Polyhomeotic.

Reference

URN : urn:nbn:ch:unige-130963
DOI : 10.13097/archive-ouverte/unige:13096
Characterization of the Polycomb Group Member Super Sex Combs (Sxc)

THÈSE

en cotutelle
présentée à la Faculté des Sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention biologie

par
Maria Cristina GAMBETTA
de
New York, Etats-Unis

Thèse n° 4244

GENÈVE
Décembre 2010
Doctorat ès sciences  
mention biologie

Thèse en cotutelle avec le Laboratoire Européen de biologie moléculaire – Heidelberg, Allemagne

Thèse de Madame Maria Cristina GAMBITTA

intitulée :

" Characterization of the Polycomb Group Member 
Super Sex Combs (Sxc)"

La Faculté des sciences, sur le préavis de Messieurs D. PAULI, docteur et co-directeur de thèse (Département de zoologie et biologie animale), J. MÜLLER, docteur et co-directeur de thèse (Laboratoire Européen de biologie moléculaire – Heidelberg, Allemagne), F. KARCH, professeur associé (Département de zoologie et biologie animale), U. SCHIBLER, professeur ordinaire (Département de biologie moléculaire), et Madame E. FURLONG, docteure (Laboratoire Européen de biologie moléculaire – Heidelberg, Allemagne), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 24 août 2010

Thèse - 4244 -

Le Doyen, Jean-Marc TRISCONÉ

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "informations relatives aux thèses de doctorat à l'Université de Genève".
# Table of Contents

**Abstract** ...................................................................................................................... 1  
Résumé (français) ................................................................................................................ 1  
Abstract (English) ............................................................................................................... 3  

**Introduction** ................................................................................................................... 5  
PcG and trxG genes are required to maintain spatial Hox gene expression patterns........ 5  
The PcG gene *super sex combs (sxc)* ............................................................................. 8  
PcG protein complexes in *Drosophila* ............................................................................. 10  
  
  PRC1 (Polycomb repressor complex 1) ............................................................................. 10  
  PRC2 (Polycomb repressor complex 2) ............................................................................ 11  
  PhoRC (Pho repressive complex) .................................................................................... 12  
  PR-DUB (Polycomb repressive deubiquitinase) ............................................................... 13  
  dRAF (dRING associated factors) .................................................................................. 13  
PcG target genes ................................................................................................................ 15  
Molecular mechanisms of transcriptional repression by PcG proteins............................. 16  
  
  Recruitment of PcG proteins to their targets .................................................................. 16  
  
  PcG proteins interfere with transcriptional elongation by RNA polymerase II ............ 17  
  
  H2A ubiquitination by PRC1 and deubiquitination by PR-DUB ...................................... 19  
  Chromatin compaction by PRC1 .................................................................................... 20  
  H3K27 trimethylation by PRC2 ..................................................................................... 20  
  H3K36me2 demethylation by dRAF .............................................................................. 22  
Antagonization of Polycomb repression by trxG proteins .................................................. 22  
  
  Trx (Trithorax) ............................................................................................................... 23  
  Ash1 (Absent, small or homeotic discs 1) ........................................................................ 24  
  Ash2 (Absent, small or homeotic discs 2) ...................................................................... 24  
  UTX (Ubiquitously transcribed tetratricopeptide repeat, X chromosome) ...................... 25  
  Kis (Kismet) ................................................................................................................... 25  

**Aim** .................................................................................................................................. 27  

**Results** ............................................................................................................................ 28  
  
  sxc mutants ..................................................................................................................... 28  
  sxc function is required for the repression of several PcG target genes ......................... 28
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloning of the <strong>sxc</strong> gene</td>
<td>30</td>
</tr>
<tr>
<td><strong>sxc</strong> is <em>O-GlcNAc</em> transferase (<strong>Ogt</strong>)</td>
<td>31</td>
</tr>
<tr>
<td><strong>sxc/oog</strong> mutants lack Ogt activity</td>
<td>33</td>
</tr>
<tr>
<td>O-GlcNAc is detected at PREs of several PcG target genes</td>
<td>34</td>
</tr>
<tr>
<td>O-GlcNAc is present on <em>Ubx</em> PREs in both the OFF and ON states</td>
<td>35</td>
</tr>
<tr>
<td>Genome-wide distribution of O-GlcNAc on <em>Drosophila</em> larval chromatin</td>
<td>36</td>
</tr>
<tr>
<td><strong>Sxc/Ogt</strong> glycosylates Polyhomeotic (<strong>Ph</strong>)</td>
<td>39</td>
</tr>
<tr>
<td><strong>Ph</strong> recruitment to PREs is impaired in <strong>sxc/oog</strong> mutants</td>
<td>42</td>
</tr>
<tr>
<td><strong>Sxc/Ogt</strong> is dispensable for PRC2’s histone methyltransferase activity</td>
<td>44</td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td>46</td>
</tr>
<tr>
<td>Parallel identification of <strong>sxc/Ogt</strong> by Sinclair et al, 2009</td>
<td>47</td>
</tr>
<tr>
<td>Is <strong>Ph</strong> the major biologically relevant Sxc/Ogt substrate?</td>
<td>47</td>
</tr>
<tr>
<td>Other candidates for biologically relevant Sxc/Ogt substrates</td>
<td>49</td>
</tr>
<tr>
<td>Speculation on the possible roles of GlcNAcylation of <strong>Ph</strong> in Polycomb repression</td>
<td>51</td>
</tr>
<tr>
<td>Evolutionary conservation of the role of Ogt in Polycomb repression</td>
<td>53</td>
</tr>
<tr>
<td>Contribution of this work to understanding the function of O-GlcNAc</td>
<td>55</td>
</tr>
<tr>
<td>Link between Polycomb silencing and diet</td>
<td>56</td>
</tr>
<tr>
<td><strong>Conclusion</strong></td>
<td>57</td>
</tr>
<tr>
<td><strong>Materials and Methods</strong></td>
<td>58</td>
</tr>
<tr>
<td>Fly husbandry</td>
<td>58</td>
</tr>
<tr>
<td>Complementation tests in view of mapping <strong>sxc</strong></td>
<td>58</td>
</tr>
<tr>
<td>Sequencing <strong>sxc</strong> alleles</td>
<td>59</td>
</tr>
<tr>
<td>Immunostaining of larval tissues</td>
<td>60</td>
</tr>
<tr>
<td>Anti-Ogt Western blot on larval extracts</td>
<td>60</td>
</tr>
<tr>
<td>Chromatin preparation from larval imaginal discs</td>
<td>61</td>
</tr>
<tr>
<td>ChIP assays</td>
<td>62</td>
</tr>
<tr>
<td>qPCR to determine binding at chromosomal regions</td>
<td>63</td>
</tr>
<tr>
<td>ChIP-on-chip experiments</td>
<td>64</td>
</tr>
<tr>
<td>Preparation of soluble embryonic nuclear extracts</td>
<td>65</td>
</tr>
<tr>
<td>WGA-agarose pull-downs</td>
<td>66</td>
</tr>
<tr>
<td>List of antibodies used</td>
<td>68</td>
</tr>
<tr>
<td><strong>References</strong></td>
<td>69</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Appendix</td>
<td>79</td>
</tr>
<tr>
<td>Supplementary figures</td>
<td>79</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>90</td>
</tr>
<tr>
<td>Publication</td>
<td>91</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>95</td>
</tr>
<tr>
<td>Remerciements (français)</td>
<td>95</td>
</tr>
<tr>
<td>Acknowledgements (English)</td>
<td>97</td>
</tr>
</tbody>
</table>
Résumé (français)

Les protéines du groupe Polycomb (PcG) sont des répresseurs transcriptionnels requis pour le maintien de la répression de gènes cibles, dans les cellules dans lesquelles ces gènes ne devraient pas être exprimés. Les protéines du PcG ont originellement été identifiées chez la drosophile, mais subséquemment il a été montré qu'elles sont évolutionnairement conservées chez les animaux et les plantes. Chez les animaux, les gènes soumis à la régulation par les protéines du PcG comprennent de nombreux gènes régulateurs du développement, aussi hautement conservés, et qui sont impliqués dans une pléthore de décisions de différentiation cellulaire. De ce fait, les facteurs du PcG sont critiques pour conserver une identité cellulaire correcte, en maintenant des programmes développementaux alternatifs sous silence. Ainsi, les protéines du PcG jouent un rôle essentiel dans la structuration du corps, et sont en outre requis, chez les cellules embryonnaires souches de mammifères, pour le maintien de la pluripotence. Leur implication étendue dans le développement des animaux a motivé une intense recherche sur les protéines du PcG dans ces dernières années. De plus, la répression Polycomb de gènes régulateurs du développement est étudiée en tant que paradigme de la mémoire épigénétique des états transcriptionnels, grâce à laquelle la répression de gènes cibles est perpétuée à travers de nombreuses générations cellulaires, en l'absence des facteurs qui ont originellement établi l'état réprimé. Des études moléculaires réalisées chez la drosophile au cours de la dernière décennie ont révélé que les protéines du PcG s'assemblent en des complexes multi-protéiques qui colocalisent au niveau de sites discrets sur la chromatine des gènes qu'ils régulent; ces sites sont appelés Eléments de Réponse Polycomb (PREs). La majorité des protéines du PcG identifiées à ce jour ont des activités enzymatiques de modification de la chromatine, ou la capacité de se lier spécifiquement à des modifications post-traductionnelles d'histones. Ceci a indiqué que les protéines du PcG agissent au niveau de la chromatine de gènes cibles afin de réprimer la transcription.

La plupart des gènes du PcG de la drosophile ont été identifiés grâce à leurs phénotypes mutants caractéristiques, qui suggèrent que de nombreuses erreurs dans les décisions d'identité cellulaire ont été commises. Ceci est aussi le cas du gène super sex combs (sxc), qui fut isolé dans une expérience de mutagenèse aléatoire il y a plus de 25 ans (Ingham, 1984). Au cours de mon travail de thèse, j'ai montré que les mutants sxc ne maintiennent pas la répression de plusieurs gènes cibles de manière efficace, ce qui permet de confirmer que sxc est un gène classique du PcG. J'ai
caractérise le gène sxc moléculairement et ai trouvé qu'il encode la O-GlcNAc transférase (Ogt), une enzyme hautement conservée chez les animaux et les plantes. Sxc/Ogt est la seule glycosyltransférase qui réside dans le noyau et le cytosol, où elle catalyse le transfert de monosaccharides N-acétylglucosamine (GlcNAc) aux sérines ou thréonines d'une grande variété de protéines substrats, un processus appelé "O-GlcNAcylation". J'ai montré que les mutants sxc sont dépourvus d'activité Ogt, ce qui suggère que la glycosylation de un ou plusieurs substrats est nécessaire pour une répression Polycomb efficace chez la drosophile. En examinant la distribution de la modification GlcNAc sur la chromatine larvaire à l'échelle du génome, nous avons trouvé que les PREs sont enrichis en cette modification. J'ai pu montrer que la présence de GlcNAc ne corrèle pas avec l'état transcriptionnel d'un des gènes cibles les mieux étudiés, ce qui suggère que la présence de la modification GlcNAc ne déclenche pas directement la répression génique. La colocalization de la modification GlcNAc avec les protéines du PcG sur la chromatine m'a poussé à tester si les protéines du PcG sont elles-mêmes glycosylées par Sxc/Ogt. L'analyse de plusieurs protéines du PcG a identifié Polyhoméotique (Ph) comme un substrat de Sxc/Ogt. En parallèle aux efforts pour identifier des substrats de Sxc/Ogt à travers lesquels Sxc/Ogt pourrait peut-être contribuer à la répression Polycomb, j'ai aussi examiné l'état de la chromatine de gènes cibles chez les mutants sxc. La liaison de Ph sur la chromatine de gènes cibles était réduite par moins qu'un facteur de deux dans les mutants sxc, tandis que la liaison à la chromatine et l'activité enzymatique de deux autres facteurs du PcG analysés étaient, dans une large mesure, inaffectées. En conclusion, nos résultats suggèrent qu'en l'absence de Sxc/Ogt et de O-GlcNAcylation, les complexes Polycomb se lient sur la chromatine de leurs gènes cibles, mais leur capacité de réprimer la transcription est compromise.

En résumé, cette thèse présente l'identification de la O-GlcNAcylation de protéines comme une étape critique de la répression transcriptionnelle par le système Polycomb, et suggère que la répression Polycomb est le processus principal qui est perturbé en l'absence de Sxc/Ogt chez la drosophile. Nos résultats soulèvent la possibilité que Sxc/Ogt pourrait contribuer à la répression Polycomb, au moins en partie, à travers la glycosylation de la protéine du PcG Ph, par des mécanismes qui restent encore à être élucidés.
Abstract (English)

Polycomb group (PcG) proteins are transcriptional repressors required for the stable maintenance of the silenced state of their target genes, in cells in which these genes should not be expressed. PcG proteins were originally identified in the model organism Drosophila melanogaster, but were found to be evolutionarily conserved in both animals and plants. In animals, genes silenced by PcG proteins include many developmental regulator genes that are also highly conserved, and are important for a plethora of cell-fate decisions. Consequently, PcG factors play a critical role in maintaining correct cell identities by shutting off alternative developmental programs in cells. In addition to their role as essential factors for the control of body patterning, PcG proteins were also found to be critical for maintaining the pluripotency of mammalian embryonic stem cells. Because of their widespread role in development, PcG proteins have become subject to intense research over the past years. PcG regulation of developmental regulator genes also represents a prime example for studying epigenetic inheritance of transcriptional states, whereby repression of target genes is inherited through multiple cell divisions in the absence of the factors that initially established the repressed state. Molecular studies in Drosophila over the past decade revealed that PcG proteins assemble into multiprotein complexes that co-bind at discrete sites on the chromatin of the genes that they regulate; these sites are called Polycomb Response Elements (PREs). Most PcG proteins identified to date have the capacity to enzymatically modify chromatin and/or interact with particular histone post-translational modifications. This has led to the view that the PcG system acts at the level of the chromatin of target genes to repress transcription.

Most Drosophila PcG genes have been identified through characteristic mutant phenotypes that suggested that multiple cell fate decisions go awry. This is also the case of the super sex combs (sxc) gene that was identified in a random mutagenesis screen over 25 years ago (Ingham, 1984). During the course of my PhD, I showed that sxc mutants fail to efficiently maintain the repression of many common PcG target genes, thus confirming that sxc is a classical PcG gene. I molecularly characterized the sxc gene and found that it encodes O-GlcNAc transferase (Ogt), a highly conserved enzyme in animals and plants. Sxc/Ogt is the only known glycosyltransferase that resides in the nucleus and in the cytosol, where it catalyzes the transfer of single N-acetylglucosamine (GlcNAc) sugar residues onto serines or threonines of a wide variety of substrate proteins, a process referred to as “O-GlcNAcylation”. I showed that sxc mutants lack Ogt activity, suggesting that glycosylation of one
or more substrates is necessary for efficient Polycomb repression in Drosophila. By investigating the distribution of the GlcNAc modification on the chromatin of larvae on a genome-wide scale, we found that it is highly enriched at PREs. I could also show that the presence of GlcNAc does not correlate with the transcriptional status of one of the best-studied PcG target genes, suggesting that the modification itself does not directly trigger gene repression. The colocalization of the GlcNAc modification and PcG proteins in chromatin then prompted me to test whether PcG proteins themselves might be glycosylated by Sxc/Ogt. Analyses of several PcG proteins identified Polyhomeotic (Ph) as an Sxc/Ogt substrate. In parallel to the efforts to identify Sxc/Ogt substrate candidates that might be biologically relevant for Polycomb repression, I also characterized the chromatin status of PcG target genes in sxc mutants. Binding of Ph to the chromatin of its targets was reduced by less than twofold in the chromatin of sxc mutants, whereas the binding and chromatin-modifying activity of two other PcG factors analysed were largely unaffected. In conclusion, these studies suggest that in the absence of Sxc/Ogt and protein O-GlcNAcylation, PcG protein complexes bind to their target genes but are compromised in repressing transcription.

In summary, this thesis reports the identification of protein O-GlcNAcylation as a critical step in transcriptional repression by the PcG system, and suggests that Polycomb repression is the main process that is disrupted in the absence of Sxc/Ogt in Drosophila. Our findings raise the possibility that Sxc/Ogt might contribute to Polycomb repression at least partially through glycosylation of the PcG member Ph, through yet to be identified mechanisms.
Introduction

During the development of multicellular organisms, cells resulting from the divisions of a totipotent embryonic cell gradually differentiate into functionally diverse cell-types. Different cell types contain the same genetic information but are distinguished by which part of this information is used, i.e. which genes are transcribed and which are not. Such cell-type specific gene expression patterns are often established early in development by transiently present factors, and are then maintained through epigenetic mechanisms for many cell divisions throughout the lifetime of the organism, after the disappearance of the initial cues that established the specified state. Understanding the molecular basis of the epigenetic mechanisms associated with such a cellular memory is valuable, as it might possibly be applied to medical areas such as regeneration and stem cell research, where it could provide a means to revert terminally differentiated cells back, maybe even to the point of totipotency; and cancer treatment, as cancer represents a failure in the maintenance of the differentiated state.

The Polycomb and trithorax groups of genes (abbreviated PcG and trxG, respectively) are part of this cellular memory system, acting on a subset of genes in the genome. PcG proteins are necessary for the maintenance of the repressed state of these genes in cells where they should not be expressed, whereas the trxG proteins antagonize the action of PcG proteins to maintain the same genes active in cells where they have to be expressed.

PcG and trxG genes are required to maintain spatial Hox gene expression patterns

In the early Drosophila embryo, maternal protein products set up the positional information along the egg axes, and this pattern sets the stage for position-specific transcriptional activation of zygotic segmentation genes, which divide the embryo into 14 segments. Among these segmentation factors, gap and pair-rule proteins are the transcription factors that initially establish transcriptional off and on states of homeotic selector (Hox) genes, which become expressed in a highly conserved pattern of body segments. Both pair-rule and gap products are no longer present after 4 hours of development. In contrast, Hox gene expression patterns have to be maintained throughout the development and life of the fly. Hox genes are so-called “master regulator” genes that specify the
identity of each segment in the embryo, larva and adult. Gain or loss of expression of Hox genes in incorrect segments results in homeotic transformations, in which body segments are transformed towards the identity of segments that are normally present in a different region of the body.

PcG genes were first identified in *Drosophila melanogaster* based on mutant phenotypes characterized by homeotic transformations in embryos and adults that suggested that several Hox genes were being expressed in segments where they should be silent (Lewis, 1978; Struhl, 1981). Analyses of the Hox gene transcripts and proteins in PcG mutant embryos revealed that the initial pattern of Hox gene expression is correctly established in PcG mutant embryos, but that Hox genes subsequently become ectopically expressed in many more segments (Struhl and Akam, 1985; Simon et al, 1992). This suggested that PcG genes are required for the maintenance of repression of Hox genes in segments in which they should not be expressed, but not for the initial establishment of repression.

Table 1 lists the 20 *Drosophila* PcG factors discovered to date. Mutations in all of these PcG genes result in the derepression of Hox genes. However, some PcG genes are duplicated in the genome and function in a partially redundant fashion, and the mutant phenotype only becomes overt in animals lacking both paralogs. Human homologs have been identified in the cases of all *Drosophila* PcG genes.

In a similar manner, trxG genes were initially discovered because mutations in these genes cause the opposite phenotype, i.e. their products keep Hox genes active in those segments where they should be expressed (Ingham, 1981). trxG genes are thus necessary to maintain the active state of Hox genes in their correct segments. Since trxG genes function is antagonistic to that of PcG genes, most trxG genes were subsequently identified in genetic screens searching for mutations that could suppress the homeotic transformation phenotypes of PcG mutants (Kennison and Russell, 1987; Kennison and Tamkun, 1988; Shearn, 1989). In these screens, several proteins playing a broader role in transcriptional activation were isolated. This contrasts with classical trxG proteins like Trx or Ash1, that do not seem to function as simple activators for the expression of target genes, but are primarily required to counteract silencing by PcG proteins (Klymenko and Müller, 2004).

This suggested that PcG and trxG proteins are the actors of a “cellular memory” system, whose role is to maintain the correct pattern of Hox gene expression after the disappearance of the initiators of this pattern of expression (i.e. the early segmentation genes). PcG and trxG genes are ubiquitously

---

1 The names of the majority of identified PcG genes refers to a frequent mutant phenotype consisting of the presence of ectopic sex combs in males, which can be explained by the ectopic expression of the Hox gene *Sex combs reduced* (*Scr*) in the second and third pair of legs in addition to its normal expression in the first pair of legs, where it directs the development of sex combs.
expressed throughout development, but in each segment, PcG and trxG factors maintain the segment-specific expression patterns of Hox genes. PcG proteins perpetuate the repression of genes that had been previously repressed by the gap proteins, whereas trxG proteins maintain the activation of those Hox genes which were expressed early on.

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Complex</th>
<th>Human homologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional sex combs</td>
<td>Asx</td>
<td>PR-DUB</td>
<td>ASXL1, ASXL2</td>
</tr>
<tr>
<td>Calypso</td>
<td>Crm</td>
<td>?</td>
<td>BAP1</td>
</tr>
<tr>
<td>Cramped</td>
<td>E(z)</td>
<td>PRC2</td>
<td>EZH1, EZH2</td>
</tr>
<tr>
<td>Extra sex combs</td>
<td>Esc</td>
<td>PRC2</td>
<td>EED</td>
</tr>
<tr>
<td>Extra sex combs-like</td>
<td>Escl</td>
<td>putative PRC2-like</td>
<td>Q9CUQ5</td>
</tr>
<tr>
<td>Multi sex combs</td>
<td>Mxc</td>
<td>?</td>
<td>YY1, YY2</td>
</tr>
<tr>
<td>Pleiohomeotic</td>
<td>Pho</td>
<td>PhoRC, Pho-INO80</td>
<td></td>
</tr>
<tr>
<td>Pleiohomeotic-like</td>
<td>Phol</td>
<td>putative PhoRC-like</td>
<td></td>
</tr>
<tr>
<td>Polycomb</td>
<td>Pcl</td>
<td>PRC1</td>
<td>NPCD, M33 (CBX2), CBX4,</td>
</tr>
<tr>
<td>Polycomblike</td>
<td>Pcl (substoichiometrically)</td>
<td>PHF19, MTF2 (M96)</td>
<td></td>
</tr>
<tr>
<td>Polyhomeotic-distal</td>
<td>Ph-d</td>
<td>PRC1</td>
<td>PHC1, PHC2, PHC3</td>
</tr>
<tr>
<td>Polyhomeotic-proximal</td>
<td>Ph-p</td>
<td>PRC1</td>
<td></td>
</tr>
<tr>
<td>Posterior sex combs</td>
<td>Psc</td>
<td>PRC1, dRAF</td>
<td>BMI1, MEL18</td>
</tr>
<tr>
<td>Scm-like with four MBTs</td>
<td>dSfmbt</td>
<td>PhoRC</td>
<td>L3MBTL2, MBTD1</td>
</tr>
<tr>
<td>Sex combs extra</td>
<td>Sce/Ring</td>
<td>PRC1, dRAF</td>
<td>RING1A or RING1, RING1B or</td>
</tr>
<tr>
<td>Sex comb on midleg</td>
<td>Scm</td>
<td>PRC1 (substoichiometrically)</td>
<td>SCMH1, SCML2</td>
</tr>
<tr>
<td>Super sex combs</td>
<td>Sxc</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Suppressor of zeste 2</td>
<td>Su(z)2</td>
<td>PRC1</td>
<td>BMI1, MEL18</td>
</tr>
<tr>
<td>Suppressor of zeste 12</td>
<td>Su(z)12</td>
<td>PRC2</td>
<td>SUZ12</td>
</tr>
</tbody>
</table>

**Table 1: Polycomb group proteins**

The following PcG proteins are partially-redundant paralogs resulting from gene duplication: Pho and Pho-like; Esc and Esc-like; Ph-d and Ph-p; Psc and Su(z)2. Human homologs of Drosophila PcG proteins have been identified, and in most cases their function in Polycomb repression is conserved (reviewed in Schwartz and Pirrotta, 2007).
The PcG gene **super sex combs (sxc)**

sxc mutants were originally isolated in a random mutagenesis screen for pharate adult lethals showing homeotic transformations of the body plan (Ingham, 1984). sxc mutants show transformations of the identities of body structures into those normally present on other body segments. For example, in sxc- pharate adults, antennae are partially transformed to legs, wings resemble a haltere, and the 2nd and 3rd pairs of legs show features specific to the 1st pair, evidenced by the presence of sex combs on all 6 legs in pharate adult males (figure i A) (Ingham, 1984; Ingham, 1985). These phenotypic transformations suggested that multiple Hox genes were being expressed outside of their correct segments, namely Antennapedia (Antp), Ultrabithorax (Ubx) and Sex combs reduced (Scr), in each of the cases just mentioned. The Ubx gene was indeed shown to be misexpressed in sxc mutant larval wing imaginal discs (Ingham, 1985), confirming that sxc is a bona fide Polycomb group gene.

sxc mutants which further lack maternal Sxc product were generated by pole cell transplantation (Ingham, 1984). These die much earlier, at the end of embryogenesis, as 1st instar pharate larvae. Although the initial pattern of Hox gene expression is correctly set up in these embryos, the denticle belt patterning of more anterior segments resembles that of the most posterior segment in older embryos (figure i B) (Ingham, 1984), hinting to widespread ectopic expression of the Hox gene Abdominal-B (Abd-B). This suggests that sxc function is required throughout fly development to maintain Hox gene repression in the correct cells.
A. \textit{sxc}\textsuperscript{-} pharate adults

\textit{sxc}\textsuperscript{mat+zyg-} pharate adult

Ingham, 1984 and 1985

B. \textit{sxc}\textsuperscript{-} pharate 1st instar larvae

\textit{wt} embryo \hspace{1cm} \textit{sxc}\textsuperscript{mat-zyg-} embryo

Ingham, 1984

Figure i: \textit{sxc} mutant phenotype

(A) Phenotype of \textit{sxc} null mutants derived from heterozygous parents, which develop until the pharate adult stage and die before eclosion from the pupal case. Pictures of the phenotypic transformations of antennae to legs, wings to halteres and of the 2nd and 3rd pairs of legs to the 1st pair, are taken from Ingham, 1984 and Ingham, 1985.

(B) Comparison of a wild-type 1st instar larva with an \textit{sxc} null mutant which further lacks maternal wildtype Sxc product, and dies at the end of embryogenesis. Small white arrowheads highlight the ventral denticle belt patterns of segments in \textit{sxc-} embryos, which resemble those of more posterior segments. The boundary between the last thoracic segment (T3) and first abdominal segment (A1) is indicated by a white arrow.

The aim of my thesis was to molecularly characterize the gene responsible for the \textit{sxc} mutant phenotype. In the rest of the introduction, I will describe the molecular functions that have been described for the greater part of identified PcG proteins, and discuss a current view of how PcG and trxG proteins molecularly regulate the transcriptional status of target genes.
PcG protein complexes in *Drosophila*

Molecular characterization of PcG proteins and their activities can help unravel the molecular mechanisms through which they stably repress their target genes. Almost all of the *Drosophila* PcG proteins characterized to date have been found to assemble in multiprotein complexes, such as Polycomb repressor complexes 1 and 2 (PRC1 and PRC2), Pho repressive complex (PhoRC), Polycomb repressive deubiquitinase complex (PR-DUB) or the dRing-associated factors complex (dRAF) (table 1). These complexes were originally purified from *Drosophila* embryos and can be reconstituted from individual recombinant proteins *in vitro*. Nevertheless, a much greater diversity of stable associations between PcG proteins (and other non-PcG proteins), notably in the case of PRC1 components, is believed to exist (Lagarou et al, 2008). It also possible that associations between PcG proteins are developmentally regulated, such that the complexes that exist in later developmental stages (larva, adult) differ from those found in embryos (e.g. Savla et al, 2008).

**PRC1 (Polycomb repressor complex 1)**

The PRC1 complex contains the core subunits Ph, Psc, Sce/Ring and Pc, and the substoichiometric component Scm (Shao et al, 1999; Saurin et al, 2001). All of these components are PcG proteins.

Pc contains a chromodomain which has been proposed to bind to trimethylated lysine 27 of histone H3 (H3K27) (Fischle et al, 2003).

In embryos, most of the soluble pool of Scm is not stably associated with PRC1 (Peterson et al, 2004). Interestingly, Scm shares a very similar and unique domain architecture with the other PcG protein dSfmbt, as well as with another protein called L(3)mbt (see figure ii). Specifically, these 3 proteins contain an N-terminal zinc finger motif, followed by 2-4 MBT (malignant brain tumor) repeats, and a C-terminal SAM (sterile alpha motif) domain. The Scm MBT repeat domain binds to monomethylated lysines of histone H3 peptides modified on lysine 4 (H3K4), 9 (H3K9), 27 (H3K27), 36 (H3K36) and histone H4 peptides modified on lysine 20 (H4K20), with a low affinity (Grimm et al, 2007). Surprisingly, complete deletion of the MBT-repeat domain still permits a mutant Scm protein to partially maintain PcG repression of target genes in a genetic rescue assay (Grimm et al, 2007).
The PRC1-members Scm and Ph and the PhoRC member dSfmbt all contain SAM domains at their C-terminus. The SAM domains of Scm and Ph mediate their homo- and hetero-dimeric interaction \textit{in vitro}; however, if this interaction also exists \textit{in vivo}, it is very weak (Peterson et al, 1997; Peterson et al, 2004; Klymenko et al, 2006; Grimm et al, 2009). Transgenic expression of Scm with mutations in its SAM domain that abolish interaction with Ph and/or Scm \textit{in vitro} do not rescue the lethality of Scm null mutant flies (Peterson et al, 2004). Tissue-specific overexpression in imaginal discs of the Scm SAM domain leads to PcG mutant phenotypes, suggesting that the SAM domain of Scm competes for interaction with the SAM domains of the other PcG proteins Ph or dSfmbt, but not of other non-PcG SAM-containing proteins (Peterson et al, 2004).

In addition to the SAM domains, the PRC1-members Ph and Scm and the PhoRC member dSfmbt share an FCS-type Cys$_2$-Cys$_2$ zinc finger (named after the signature FCS residues associated with the third cysteine), that are distinct from classical zinc fingers of known DNA-binding proteins. Scm and the PhoRC-member dSfmbt interact \textit{in vitro} through their N-termini, containing the zinc finger domains, thus independently of their SAM domains (Grimm et al, 2009).

**PRC2 (Polycomb repressor complex 2)**

PRC2 is composed of the PcG proteins Esc, E(z), Su(z)12, as well as Nurf55, a component of several chromatin-modifying complexes (Tie et al, 2001; Müller et al, 2002; Nekrasov et al, 2007). Only a small proportion of total nuclear Nurf55 is included in PRC2, and it is also a component of the CAF ("chromatin assembly factor") and NURF ("nucleosome remodeling factor") complexes, which are distinct from PRC2 (Tie et al, 2001). The PcG protein Pcl is a substoichiometric component of PRC2 (Nekrasov et al, 2007). In addition, the deacetylase Rpd3 might be transiently associated with PRC2 (Tie et al, 2001; Czermin et al, 2002).

The E(z) subunit contains a SET (Suvar3-9, Enhancer-of-zeste, Trithorax) catalytic domain, and is responsible for all methylation of H3K27 in flies (Ebert et al, 2004).
PhoRC (Pho repressive complex)

The PcG protein Pleiohomeotic (Pho) and its paralog Pho-like (Phol) are the only known PcG proteins capable of binding to specific DNA sequences similar to that of their human homolog, the transcription factor Yin Yang-1 (YY1) (Brown et al, 1998; Brown et al, 2003). Drosophila Pho and Phol and human YY1 all contain 4 highly conserved Gli-type zinc fingers (Brown et al, 1998; Brown et al, 2003). Pho and Phol act to a large extent redundantly, as removal of both functions causes more severe derepression of PcG target genes (Brown et al, 2003). The tandem affinity purification (TAP) of Pho from embryos revealed that it is present in two separable protein complexes, PhoRC and Pho-INO80 (Klymenko et al, 2006).

PhoRC contains the PcG protein dSfmbt in addition to Pho (Klymenko et al, 2006). Phol is also associated with dSfmbt in vivo (Klymenko et al, 2006). As previously mentioned, the dSfmbt protein contains a highly similar domain architecture to the PRC1 member Scm. The MBT domain of dSfmbt selectively binds to histone tail peptides mono- or di-methylated at a variety of lysine positions in vitro, with a ~1000-fold higher affinity than the Scm MBT domain: H4K20me1/2, H3K9me1/2, H3K4me1/2, H3K27me1/2 and H3K36me1/2 (Grimm et al, 2009). The biological relevance of this methyl-lysine binding capacity remains enigmatic, since a dSfmbt protein that is no longer able to bind to methyl-lysines, because of engineered mutations in the binding pocket located in MBT repeat 4, is nevertheless able to efficiently restore PcG repression of target genes when introduced in dSfmbt null mutants (Grimm et al, 2009). This is similar to what was observed in the case of mutations of the MBT domains of Scm. It is possible that the MBT domains of Scm and dSfmbt can partially functionally substitute for one another, since removal of dSfmbt in cells which express an Scm MBT point mutant protein causes a very strong enhancement of derepression of PcG target genes, which is not observed in either single dSfmbt or scm mutant cells (Grimm et al, 2007; Grimm et al, 2009). Moreover, dSfmbt and Scm proteins might directly interact in vivo, since a stable dSfmbt::Scm complex can be assembled from recombinant proteins (Grimm et al, 2009). The interacting regions of dSfmbt and Scm map to their N-terminal regions each containing a zinc finger motif (Grimm et al, 2009).
**PR-DUB (Polycomb repressive deubiquitinase)**

The polycomb group proteins Calypso and Asx form a stable complex *in vivo* and *in vitro*, named PR-DUB (Scheuermann et al, 2010). The PR-DUB complex is conserved, as it can be reconstituted from the recombinant human homologs BAP1 (BRCA1 associated protein 1) and ASXL1 (Scheuermann et al, 2010). PR-DUB is a major deubiquitinase for H2AK118ub1 in *Drosophila* (Scheuermann et al, 2010), and the catalytic activity of PR-DUB is essential for polycomb repression of target genes.

**dRAF (dRING associated factors)**

The dRAF complex was purified from *Drosophila* embryos by enriching for Sce/Ring-associated proteins after depletion of Pc-containing complexes such as PRC1 (Lagarou et al, 2008). The dRAF complex includes the PcG proteins Sce/Ring and Psc, but does not contain the PRC1 members Pc nor Ph (Lagarou et al, 2008). Sce/Ring and Psc are thus shared between PRC1 and dRAF. In addition, other proteins such as dKdm2 (an H3K36me2 demethylase), dRaf2, Mtor (a dynamic subunit of the nuclear pore complex) and Ulp1 (a SUMO peptidase also associated with the nuclear pore) also compose dRAF (Lagarou et al, 2008). It is at the moment unclear whether dKdm2 is a classical PcG protein, since no obvious homeotic phenotypes have been described in *dKdm2* mutants, yet *dKdm2* mutations have been shown to enhance the phenotype of PcG mutants and suppress the phenotype of certain trithorax mutants (Lagarou et al, 2008).
Figure ii: Schematic representations of PhoRC, PRC1, PRC2, dRAF or PR-DUB PcG protein members

Only one paralog (Pho, Ph-p, Psc, Esc) is shown in the case of PcG with paralogs; and the principal protein isoform is represented. Asx contains an atypical PhD Zn finger domain at its C-terminus (hatched box). Zn finger: zinc finger; MBT: malignant brain tumor; SAM: sterile alpha motif; HD: homology domain; RING: really interesting new gene; SANT: switching-defective protein 3 (Swi3), adaptor 2 (Ada2), nuclear receptor co-repressor (N-CoR), transcription factor (TFIIIB); SET: Suvar3-9, Enhancer-of-zeste, Trithorax; VEFS: VRN2-EMF2-EFS2-Su(z)12; WD40: tryptophan-aspartic acid 40; PHD Zn finger: plant homeodomain zinc finger; UCH: ubiquitin carboxy-terminal hydrolase.
PcG target genes

PcG proteins regulate transcription by associating with specific DNA cis-regulatory sequences called Polycomb response elements (PREs). PREs were first identified in Drosophila as cis-acting silencer sequences needed for Hox reporter gene repression mediated by PcG proteins (Müller and Bienz, 1991; Simon et al, 1993; Chan et al, 1994; Christen and Bienz, 1994). The silencing function of PREs is continuously required throughout development, and excision of a PRE from a silenced Hox reporter gene results in a rapid loss of repression (Busturia et al, 1997; Sengupta et al, 2004). PREs have subsequently been shown to be specifically bound by PcG and trxG proteins by chromatin immunoprecipitation (ChIP) analyses, a technique in which cells are cross-linked with formaldehyde in order to preserve chromatin structures and thus map protein-DNA interactions (Orlando et al, 1997).

Recently, genome-wide binding sites for several PcG proteins in Drosophila cells have been determined by DamID or ChIP-on-chip experiments, where enriched target sequences are hybridized to a genomic tiling array to map these fragments back to the genome (Nègre et al, 2006; Schwartz et al, 2006; Tolhuis et al, 2006; Oktaba et al, 2008; Schuettengruber et al, 2009; Schwartz et al, 2010; Scheuermann et al, 2010). In this way, PRC1, PRC2, PhoRC and PR-DUB members have been shown to colocalize at several PREs. Binding of PcG proteins at certain target genes seems to be developmentally regulated (Nègre et al, 2006; Oktaba et al, 2008). PcG binding sites have been shown to be depleted in nucleosomes. The PcG binding sites determined in these studies are putative PREs. In most cases, however, it remains yet to be shown whether the expression of the gene found in the close proximity of a candidate PRE is indeed regulated by PcG proteins or not. The majority of PhoRC binding sites are located within 1 kb from the nearest transcription start site (TSS) (Oktaba et al, 2008). The highly studied PREs of certain Hox genes, which are located dozens of kb away from their target gene promoters, are thus exceptional cases.

These genome-wide studies have revealed that PcG proteins appear to regulate a large number of genes in addition to the classically characterized Hox genes. PcG targets include several developmental regulators important for cell-fate decisions. A high proportion of PcG targets encode transcriptional regulators, as well as morphogens, receptors and signaling proteins that are involved in all of the main developmental pathways (reviewed in Schwartz and Pirrotta, 2007). Interestingly, comparison with the results of PcG protein binding profiles in mammalian embryonic stem cells (ESCs) have shown that several of these targets are conserved between flies and humans (Schwartz and Pirrotta, 2007).
PcG proteins also seem to limit the expression of cell-cycle proteins, and certain PcG mutant cells form tumors (Oktaba et al, 2008; Martinez et al, 2009; Classen et al, 2009), increasing the interest in PcG proteins for the cancer research field. Misregulation of several PcG proteins have indeed been observed in human tumours (reviewed in Bracken and Helin, 2009).

Interestingly, certain PcG genes themselves have been shown to contain PREs (Rastelli et al, 1993; Fauvarque et al, 1995), indicating that the expression of PcG proteins might be regulated by a negative-feedback loop, and that a balance in PcG protein levels might be necessary for proper regulation.

Trx was found to be present at nearly all PREs by ChIP-on-chip (Schwartz et al, 2010). Thus, PREs are also TREs (“trithorax response elements”).

Recently, evidence has been provided for the existence of mammalian cis-acting silencers capable of recruiting PcG proteins and repressing gene expression in a PcG/trxG-dependent manner (Sing et al, 2009; Woo et al, 2010). This provides evidence for the existence of mammalian PREs, and indicates conservation in the mechanisms that target PcG function in mammals and flies.

**Molecular mechanisms of transcriptional repression by PcG proteins**

PcG function is closely associated with the modulation of chromatin structure and covalent post-translational modifications of histones. PcG proteins are thought to negatively regulate transcription by RNA polymerase II at a late step of initiation.

**Recruitment of PcG proteins to their targets**

Surprisingly, PhoRC, PRC1, PRC2 and PR-DUB members have been found to be recruited to the Ubx PREs both in wing imaginal disc cells, where Ubx is repressed by the PcG system (UbxOFF), and in haltere and 3rd leg imaginal disc cells, where Ubx is actively transcribed (UbxON) (Papp and Müller, 2006; Scheuermann et al, 2010) (figure iii). Therefore, silencing or activation of Ubx is not determined at the stage of recruitment of PcG proteins. It is important to note, though, that studies comparing ON and OFF states of other Hox genes in Drosophila tissue culture cells have noted the absence of PcG proteins from PREs in the ON state (e.g. Beisel et al, 2007; Schwartz et al, 2010).
Since Pho and Phol are the only DNA-binding PcG proteins, and since PREs are devoid of nucleosomes, not allowing recruitment of other PcG proteins through interactions with chromatin, it is generally thought that binding of PhoRC to PREs could serve as a nucleation site for recruitment of other PcG complexes. In support of this hypothesis, mutation of Pho binding sequences in a Ubx PRE disrupts the PRE’s silencing activity (Fritsch et al, 1999) as well as binding of PhoRC and Ph to this PRE in vivo (Klymenko et al, 2006). Moreover, Pc (a PRC1 member) and E(z) and Pcl (PRC2 members) no longer bind to a Ubx PRE in imaginal discs of Pho phol double mutant larvae (Wang et al, 2004 b; Savla et al, 2008). However, it seems that Pho and Phol are not required for recruitment of PcG proteins to most other PREs, since immunostaining experiments of polytene chromosomes from salivary glands of pho phol double mutant larvae revealed that binding of all of the tested PcG proteins seemed unaffected at most sites (Brown et al, 2003).

Intriguingly, in contrast to other PRC1 and PRC2 members, recruitment of Scm to a Ubx PRE seems to be independent of Pho and Phol in Drosophila S2 cells (Wang et al, 2010 a). Moreover, Scm seems to have a recruiting role like Pho, because reduction of Scm protein levels also leads to a reduction in PRE-bound levels of the tested PRC2 and of the PRC1 components, although high levels of PRE-bound Pho are retained (Wang et al, 2010 a). This suggests that Scm recruitment to the PRE is independent of PhoRC, PRC1 and PRC2, perhaps through the association of Scm with a yet unidentified DNA-binding factor. DNA-bound PhoRC and Scm would in turn help to recruit other PcG complexes to PREs.

However, if interactions between PhoRC or Scm and members of other PcG complexes do exist, they are probably weak, since biochemical purification of PcG complexes have so far shown that these are distinct entities which do not share common components. Alternatively, the complexes might only interact once both tethered to chromatin, precluding finding their association in soluble embryonic nuclear extracts. Several DNA-binding factors are known to be present at PREs, and their role in the recruitment of PcG proteins remains to be investigated.

**PcG proteins interfere with transcriptional elongation by RNA polymerase II**

By comparing the UbxON and UbxOFF chromatin states it was revealed that the TATA-binding protein (TBP) and the transcription elongation factor Suppressor of Ty 5 (Spt5) are present both at the Ubx TSS, and at PREs, irrespective of the transcriptional state of Ubx (Papp and Müller, 2006) (figure
iii). Spt5 can be detected spread across the transcribed region in *Ubx*ON chromatin only, consistent with a model where Spt5 is associated with an elongating RNA polymerase II in this chromatin, but associated with a stalled RNA polymerase II at the *Ubx* TSS in the OFF state (Papp and Müller, 2006). These observations suggest that, in the Polycomb-repressed state, *Ubx* transcription is blocked at a late step of transcriptional initiation. Recent studies have shown that paused polymerases are present downstream of the promoters of many silent genes in *Drosophila*, suggesting that the regulation of early elongation is a relatively widespread phenomenon (Muse et al, 2007; Zeitlinger et al, 2007). PRC1 may block transcription via direct interactions with components of the basal transcription machinery, as evidenced by the recovery of transcription factor IID (TFIID) subunits in PRC1 purifications (Breiling et al, 2001; Saurin et al, 2001).

Eukaryotic transcription involves a highly coordinated cycle of events, any step of which could be subject to regulation by PcG and trxG proteins. Basal transcription factors and RNA Polymerase II are first recruited to the promoter. The C-terminal domain (CTD) of the largest subunit of RNA polymerase II is then phosphorylated on Ser 5, coinciding with promoter clearance. Transition from transcription initiation to elongation coincides with phosphorylation of Ser 2 by the P-TEFb complex. CTD phosphorylation modulates interactions between RNA polymerase II and other factors involved in transcription.

It is possible that PcG proteins interfere with transcriptional elongation by impeding the recruitment of the trxG chromatin remodeller Kismet (Kis). Indeed, Kis is constitutively recruited to PREs, but is only detected at the *Ubx* promoter in the ON state (Papp and Müller, 2006). Studies examining polytene chromosomes from salivary glands of *kis* mutant larvae have gained insight into Kis function. Kis seems to be associated with virtually all sites of transcriptionally active chromatin (Srinivasan et al, 2005). In *kis* mutant larvae, RNA polymerase II seems to be correctly recruited to promoters, and phosphorylated on serine 5 (Srinivasan et al, 2005). However, CTD phosphorylation on serine 2 is drastically reduced, despite normal recruitment of the Cyclin T subunit of P-TEFb (Srinivasan et al, 2005; Srinivasan et al, 2008). In addition, recruitment of the SPT6 and Chromodomain-helicase-DNA-binding protein 1 (CHD1) elongation factors is also lost (Srinivasan et al, 2005). These observations strongly suggest that Kis facilitates an early step in transcriptional elongation by RNA polymerase II.
H2A ubiquitination by dRAF and deubiquitination by PR-DUB

The *Drosophila* PcG proteins Sce/Ring and Psc dimerise through their N-terminal RING finger domains, which are a signature motif for ubiquitin E3 ligases, capable of catalyzing the transfer of a ubiquitin molecule onto a lysine residue of a substrate protein (Buchwald et al, 2006). Both human RING::BMI1 and *Drosophila* Sce/Ring::Psc complexes monoubiquitinate H2A on K119 or K118, respectively (Wang et al, 2004 a; Buchwald et al, 2006; Lagarou et al, 2008). Sce/Ring seems to function predominantly as an E3 ligase, whereas Psc is thought to be an important co-factor for boosting the activity of Sce/Ring (Buchwald et al, 2006; Lagarou et al, 2008). In *Drosophila*, Sce/Ring and Psc are present in both the PRC1 and dRAF complexes. dRAF, rather than PRC1, is responsible for H2A ubiquitination in *Drosophila* cells (Lagarou et al, 2008). This is due to the presence of the dRAF-specific subunit dKdm2, which is required to significantly boost the E3 ligase activity of the Sce/Ring::Psc dimer, to the point that knock-down of either dKdm2, Psc or Sce/Ring in tissue culture cells leads to comparable drastic reductions in H2Aub1 levels, that are not seen upon knock-down of the PRC1-specific members Ph or Pc (Lagarou et al, 2008). Similarly, RING1A and RING1B are the major H2A E3 ligases in human cells (de Napoles et al, 2004).

The correlation between the loss of H2A monoubiquitination and the derepression of PcG target genes that occur upon loss of Ring function, have led to the hypothesis that H2A ubiquitination is necessary for target gene silencing. However, this is still unclear since the *Sce*/Ring mutations analysed to date probably also lead to destabilization of Sce/Ring-containing protein complexes.

The *Drosophila* PcG protein Calypso and its human homolog BAP1 are ubiquitin carboxy-terminal hydrolases, a type of cystein proteases capable of cleaving the isopeptide bond between a ubiquitin residue and the lysine residue of a substrate protein. Recombinant Calypso/Asx and BAP1/ASXL1 complexes specifically deubiquinate nucleosomes containing H2Aub1, but not H2Bub1, in vitro (Scheuermann et al, 2010). Asx is a necessary cofactor for Calypso’s deubiquitinase activity (Scheuermann et al, 2010). Flies lacking Asx also show reduced Calypso protein levels, and show a 10-fold increase in H2Aub1 levels (Scheuermann et al, 2010). This suggests that PR-DUB is a major deubiquitinase specific for H2A in flies, and possibly also in human cells. The deubiquitinase activity of PR-DUB is required for Polycomb repression, since Calypso protein with a mutation in the catalytic cysteine is unable to rescue Hox gene derepression in calypso flies (Scheuermann et al, 2010).

The fact that the loss of H2A ubiquitination by Sce/Ring, and that the accumulation of H2Aub1 levels upon loss of PR-DUB activity, both lead to the derepression of overlapping PcG target genes...
seems paradoxal. The idea that Sce/Ring and PR-DUB function synergistically comes from the fact that the Sce/Ring mutant phenotype is aggravated (versus suppressed) by additional loss of PR-DUB (Scheuermann et al, 2010). The distribution of ubiquitinated H2AK118 on Drosophila chromatin is currently unclear. It is possible that the presence of H2AK118 at certain chromosomal regions, as well as its absence from other chromosomal regions, might be necessary for Polycomb repression of targets. Alternatively, it is possible that H2A ubiquitination and deubiquitination may have to occur in a temporally regulated cycle to maintain repression.

Chromatin compaction by PRC1

In nucleosome remodeling assays in vitro, PRC1 inhibits chromatin remodeling by SWI-SNF complexes (Shao et al., 1999; Francis et al., 2001; Saurin et al., 2001; Francis et al, 2004), providing a molecular explanation for the genetic interaction between Polycomb mutants and brahma trxG mutants. This might be a consequence of compaction of nucleosome arrays that is observed by electron microscopy upon addition of recombinant PRC1 (Francis et al, 2004).

In undifferentiated mouse ESCs, Ring1B is necessary to maintain Hox loci chromatin compaction and transcriptional repression (Eskeland et al, 2010). Intriguingly, an H2A ubiquitination-defective RING1B protein causes recondensation of Hox genes and rescues Hox gene derepression in RING1B\textsuperscript{-/-} ESCs, despite its incapacity to restore wild-type levels of H2AK119ub (which are nevertheless high, due to the activity of RING1A) (Eskeland et al, 2010). This suggests that the chromatin compaction activity of Ring-containing complexes is dissociable from H2A ubiquitination activity.

H3K27 trimethylation by PRC2

Recombinant tetrameric PRC2 (E(z), Esc, Su(z)12, Nurf55) is capable of methylating histone H3 on lysine K27 in vitro (Müller et al, 2002). The methyltransferase activity of PRC2 is conferred by the SET domain of the E(z) subunit; however, E(z) alone is by far less active than the PRC2 complex, suggesting that the other PRC2 members are necessary to boost catalytic activity and/or efficiently interact with the H3 substrate (Müller et al, 2002). The WD40 domain of Esc specifically binds to
histone tails carrying trimethyl-lysine residues associated with repressive chromatin marks, and this leads to the allosteric activation of the methyltransferase activity of PRC2 (Margueron et al, 2009). Su(z)12 and Nurf55 bind to nucleosomes, and might aid E(z) to efficiently interact with the H3 substrate in chromatin (Nekrasov et al, 2005).

Many lines of evidence consolidate the view that trimethylation of H3K27 by PRC2 is a crucial step in Polycomb repression. E(z) is responsible for the generation of all mono-, di- and trimethylation of H3K27 in Drosophila (Ebert et al, 2004; Papp and Müller, 2006; Schwartz et al, 2010), and H3K27me3 levels are drastically reduced in esc and Su(z)12 mutant cells as well (Ketel et al, 2005; Papp and Müller, 2006; Nekrasov et al, 2007). Specifically, point mutations in the E(z) SET domain or in the trimethyl-lysine binding pocket of Esc compromise Polycomb repression (Müller et al, 2002; Margueron et al, 2009).

Chromatin immunoprecipitation experiments in Drosophila have shown that although PRC2 binds in a localized manner at PREs of target genes, H3K27me3 is distributed across large Polycomb-repressed chromatin domains (e.g. Papp and Müller, 2006; Schwartz et al, 2010). Comparison of the chromatin of the Hox gene Ubx in the Polycomb-repressed (OFF) and active (ON) states in larvae revealed that the whole upstream control region of Ubx is constitutively trimethylated at H3K27 (Papp and Müller, 2006). However, presence or absence of H3K27me3 in the Ubx promoter and coding region correlates tightly with the gene being repressed or active, respectively (Papp and Müller, 2006) (figure iii). In contrast, H3K27me1/2 has been reported to be present genome-wide on more than 50% of total H3 in Drosophila (Ebert et al, 2004; Nekrasov et al, 2007). H3K27 trimethylation is thus a distinctive mark of PcG-repressed chromatin.

A fraction of endogenous PRC2 complexes are associated with the PcG protein Polycomblike (Pcl) (Nekrasov et al, 2007). In the absence of Pcl, PRC2 is still able to methylate histones, but does so in a non-targeted manner across chromosomes (Nekrasov et al, 2007). Pcl is needed to efficiently recruit PRC2 to PREs of target genes in order to generate local high levels of H3K27me3 (Nekrasov et al, 2007; Savla et al, 2008).

What could the molecular mechanism, through which H3K27me3 contributes to Polycomb-repression, be? A possible explanation for silencing of promoters by PREs at a distance would be that PRE-bound Pc interacts with H3K27me3 nucleosomes generated by PRC2, thus creating chromatin loops that bring PRE-bound factors in the proximity of the promoter and coding gene regions (Müller and Kassis, 2006; Schwartz and Pirrotta, 2007). These interactions might allow methylation of other residues such as H3K9 and H4K20, by yet unidentified histone methyltransferases (HMTases) possibly
present at PREs; as it was observed that in UbxOFF chromatin, H3K9me3 and H4K20me3 are spread across the promoter and transcribed region similarly to H3K27me3 (Papp and Müller, 2006) (figure iii). The molecular mechanisms by which trimethylation of these histones contributes to repression is not yet understood. PRE looping might also allow PcG proteins bound at PREs to establish specific contacts with promoter-bound components of the transcription machinery in order to silence their targets.

In a small number of exceptional cases, H3K27me3 can also be found on genes in a so-called “balanced” state, characterized by simultaneous binding of PcG factors and trxG proteins, in Drosophila tissue culture cells (Schwartz et al, 2010). Importantly, these genes are transcriptionally active, indicating that the presence of H3K27me3 is not directly responsible for inhibition of gene transcription (Schwartz et al, 2010).

**H3K36me2 demethylation by dRAF**

The dRAF complex member dKdm2 contains a JmjC demethylase domain, and seems to be a major H3K36me2 demethylase in Drosophila cells, testified by the important increase in H3K36me2 levels observed upon dKdm2 knock-down (Lagarou et al, 2008). In vivo, all H2Aub-modified nucleosomes are also demethylated on H3K36me2 (Lagarou et al, 2008). Since H3K36me2 is an active mark which might be mediated by the trxG protein Ash1 (Tanaka et al, 2007), it is proposed that dRAF contributes to silencing of Polycomb group target genes by coupling the removal of the active H3K36me2 mark to the addition of the H2Aub (Lagarou et al, 2008). Interestingly, homeotic transformations caused by reduced Hox gene expression in ash1 heterozygotes are suppressed in ash1 dkdkm2 double heterozygotes, suggesting that reduced dRAF activity alleviates the requirement for Ash1 (Lagarou et al, 2008).

**Antagonization of Polycomb repression by trxG proteins**

Repression by PcG proteins is counteracted by trxG regulators that also act through the modification of chromatin. In contrast to PcG complexes, however, which act in concert at their dedicated elements, many trxG members are involved in more general transcription processes. For
example, certain trxG proteins are assembled in complexes with ATP-dependent chromatin remodeling activity, such as the Brahma (BRM) and ISWI-containing complexes. The function of these complexes in gene regulation is most likely to facilitate DNA-accessibility for transcriptional regulators by promoting alterations of the nucleosome structure.

Another set of trxG factors display specific histone modifying activities that are more selectively required to counteract PcG silencing by modifying epigenetic marks. For example, the Trithorax (Trx), absent, small and homeotic discs 1 and 2 (ash1 and ash2), and ubiquitously transcribed tetratricopeptide repeat, X chromosome (Utx) loci are classic trxG genes, whose mutants show homeotic transformations consistent with loss of activity of multiple Hox genes (Ingham and Whittle, 1980; Capdevila and García-Bellido, 1981; LaJeunesse and Shearn, 1995; Herz et al, 2010).

I briefly describe the proposed anti-repressor functions of some classic trxG factors.

**Trx (Trithorax)**

Trx counteracts default silencing by PcG proteins: it is needed for Hox gene expression in wild-type *Drosophila*, but is dispensable for their transcription in PcG mutants (Klymenko and Müller, 2004). Trx is a highly conserved protein with homologues in yeast (Set1) and in humans (MLL). In yeast, Set1 is a member of the protein complex COMPASS (COMplex Proteins Associated with Set1; Miller et al, 2001) and MLL is part of the MLL complex in humans (human COMPASS; Yokoyama et al, 2004). Trx was reported to be present in a complex called TAC1 (Trithorax acetylation complex 1) that contains Trx, CREB-binding protein (CBP) and SET domain binding protein (Sbf1) (Petruk et al, 2001). The TAC1 complex has 2 histone-modifying enzymatic activities, conferred by the acetyltransferase CBP, and the methyltransferase activity of the SET domain of Trx. To date, Trx complexes related to MLL/COMPASS complexes have not been described yet, but it is likely that Trx also exists in such a complex in *Drosophila*. Trx has been reported to methylate H3K4 (Smith et al, 2004), although *in vivo* evidence suggests that it is not required for H3K4 trimethylation in *Drosophila* (Srinivasan et al, 2008; Schwartz et al, 2010).

Both *Drosophila* Trx and its human homolog Mll1 are cleaved into an N-terminal and a C-terminal fragment, by an endogenous protease (Hsieh et al, 2003). The N-terminal fragment of Trx contains several PHD (plant homeodomain) zinc finger domains, whereas the C-terminal portion contains the catalytic SET domain.
**Ash1 (Absent, small, or homeotic discs 1)**

Like Trx, Ash1 function is necessary to counteract default silencing by PcG proteins, but it is not a coactivator necessary for expression of the target gene (Klymenko and Müller, 2004). The Ash1 protein contains a SET domain, and has been shown to dimethylate H3K36 (Tanaka et al, 2007), although another HMTase is responsible for the bulk H3K36me2 in *Drosophila* (Bell et al, 2007).

The genome-wide binding profile of Ash1 was determined in *Drosophila* tissue culture cells (Schwartz et al, 2010). Ash1 and Trx N-ter co-bind interdependently over large domains which can extend up to 77.5 kb in length, and which are acetylated at H3K27 (Schwartz et al, 2010). Genes bound by Ash1 and Trx N-ter are actively transcribed (Schwartz et al, 2010).

In *ash1* mutants, *Ubx* becomes repressed in tissues in which it should normally be expressed. Analysis of ectopically silenced *Ubx* chromatin in *ash1* mutants revealed that the promoter and coding regions become extensively trimethylated at H3K27 and H3K9 (Papp and Müller, 2006). The fact that *Ubx* expression is restored in *ash1* mutant cells that also lack E(z) function (Klymenko and Müller, 2004) provides strong evidence that Ash1 is required in the ON state to prevent PRC2 and other HMTases from trimethylating the promoter and coding region at H3K27 and H3K9. Ash1 might dimethylate H3K36 on nucleosomes across these regions, and might function as an insulator element to delimit a domain where the repressive heterochromatin marks can not spread. It is also possible that erasure of the H3K27me3, H3K9me3 and H4K20me3 marks might be a downstream consequence of transcriptional activation brought about by Ash1::Trx N-ter.

**Ash2 (Absent, small, or homeotic discs 2)**

Like *trx* and *ash1*, *ash2* is a classic trxG gene. Ash2 protein has no domain involved in HMTase activity; nevertheless, loss of *ash2* function in wing imaginal discs results in a dramatic reduction of H3K4me3 (Beltran et al, 2007). In yeast and mammals, Ash2 is a COMPASS/MLL complex subunit. Thus, it was hypothesized that Ash2 function may be required to facilitate the interaction between nucleosomes and the enzyme responsible for H3K4 trimethylation.
**UTX (Ubiquitously transcribed tetratricopeptide repeat, X chromosome)**

UTX is necessary for maintaining proper expression of Hox genes in human cells (Lee et al., 2008), and *dUTX* mutant flies display homeotic transformations suggestive of loss-of-function of the HOX gene Scr (Herz et al., 2010), thus classifying it as a trxG protein. UTX contains a Jumonji C (JmjC) domain, which is a signature for histone demethylases. Indeed, recombinant UTX specifically demethylates H3K27me2/3 (Lee et al., 2007; Smith et al, 2008). Importantly, *dUTX* null mutant flies show a slight increase in global H3K27me3 levels (but not H3K27me2 nor H3K27me1 levels) (Herz et al., 2010). The relatively mild phenotype and increase in H3K27me3 levels observed upon loss of UTX might indicate that UTX is not the only H3K27me3 demethylase in cells. Nevertheless, UTX might reverse the silencing established by PcG proteins at certain targets by erasing the H3K27me3 marks across the body of Polycomb-repressed genes. H3K27me1 is indeed an active mark found in the coding regions of transcribed genes in human cells (Barski et al, 2007).

**Kis (Kismet)**

Zygotic *kis* mutants display homeotic transformations suggestive of loss-of-function of certain HOX genes, thus classifying it as a classical trxG gene (Daubresse et al, 1999). The long protein isoform of Kis contains an ATPase domain through which it is thought to catalyze ATP-dependent alterations in chromatin structure.

Interestingly, both Trx and Ash1 binding to chromosomes is dramatically reduced in the absence of Kis, although these proteins are still produced in *kis* mutants (Srinivasan et al, 2008). This raises the possibility that Kis may promote early elongation and counteract PcG repression by recruiting the Ash1 and Trx HMTases to chromatin (see “PcG proteins interfere with transcriptional elongation by RNA polymerase II”). It is not known whether Kis recruits Ash1 and Trx through its chromatin remodeling activity.
**Ubx OFF**

H3-K27me3
H3-K9me3
H4-K20me3

**Ubx ON**

H3-K27me3
H3-K9me3
H4-K20me3

*Figure iii: PcG and trxG protein binding and histone lysine methylation at the Ultrabithorax Hox gene*

Schematic representation of the binding of proteins and methylation marks in the Polycomb-repressed (OFF) and active (ON) states of the Ubx gene. Trx, PhoRC, PRC1, PRC2 (Papp and Müller, 2006) and PR-DUB (Scheuermann et al, 2010) are constitutively bound at the bxd and bx PREs. Trx, TBP and Spt5 are constitutively bound at the transcription start site (Papp and Müller, 2006). The twofold reduction in the level of PRC1 and PRC2 binding at the internal bx PRE in the ON state, and the reduction in the level of TBP and Spt5 binding at the TSS in the OFF state, are indicated by smaller symbols (Papp and Müller, 2006). The trxG proteins Kis and Ash1 are only bound in the ON state, at the TSS or 1 kb downstream of the TSS, respectively (Papp and Müller). The localization of the N-terminal portion of Trx was not analysed in this study, but it is probably associated with Ash1 (Schwartz et al, 2010). In the OFF state, H3K27, H3K9 and H4K20 trimethylation spreads across the entire Ubx gene, whereas in the ON state it is largely absent in the promoter and coding region, where H4K20me1 and H3K4me3 are present instead (Papp and Müller, 2006). In addition, H3K27ac is usually also co-extensive with Ash1::Trx N-ter domains (Schwartz et al, 2010). In the OFF state, dsfmbt tethered to PREs by Pho may interact with hypomethylated H3K27, H3K9 and H4K20 tails in flanking chromatin (Grimm et al, 2009), and thereby bring them into the vicinity of PRE-anchored HMTases that will hypermethylate them to the trimethylated state. Pc may interact with H3K27me3 through its chromodomain (Fischle et al, 2003) and thereby allow PRE-bound factors to act upon the Ubx TSS or transcribed region. This figure was taken from Papp and Müller, 2006.
Aim

Only a minority of the thus far genetically characterized PcG genes still await molecular characterization, and this included the super sex combs gene (Ingham et al, 1984). sxc mutants display Polycomb phenotypes shared by many other PcG mutants, and as such it was classified as a classical PcG gene (Ingham, 1984; Ingham, 1985). In a continuing effort to identify all players important for Polycomb repression, I pursued the work previously initiated by Cornelia Fritsch and Bernadett Papp in our lab to identify the sxc gene. Once we would discover the sxc gene, the goal would be to characterize the biochemical activity of this PcG protein, in view of understanding the molecular mechanisms through which it contributes to Polycomb repression, and how it cooperates with the previously described PcG protein complexes.
Results

sxc mutants

In addition to the originally isolated sxc<sup>1</sup>, sxc<sup>2</sup>, sxc<sup>3</sup>, sxc<sup>4</sup>, sxc<sup>5</sup> alleles (Ingham, 1984), we found that the l(2)NC130 and Df(2R)NC32 alleles, that were isolated in an independent chemical (ethyl methanesulfonate, or EMS) mutagenesis screen for lethality in trans with a deficiency located in the proximal region of chromosome 2R (Myster et al, 2004), failed to complement the lethality and sxc mutant phenotype in combination with the original sxc alleles. These alleles were therefore renamed sxc<sup>6</sup> and sxc<sup>7</sup>, respectively.

With the exception of sxc<sup>2</sup>, all sxc alleles in transheterozygous or hemizygous combinations die as pharate adults showing a comparably severe sxc mutant phenotype (see Introduction), suggesting that they are all null alleles. Homozygotes often die prior to reaching the pharate adult stage, probably as a consequence of homozygosity for additional mutations induced by the EMS mutagenesis. sxc<sup>2</sup> is a hypomorphic allele, since transheterozygous combinations with other sxc alleles develop into fertile adults showing wing abnormalities that suggest partial wing-to-haltere transformations (Ingham, 1984).

sxc function is required for the repression of several PcG target genes

As previously mentioned in the introduction, the phenotypes of sxc mutant embryos and pharate adults suggested that multiple Hox genes were being expressed outside of their correct segments; and this was indeed proven in the case of the Ubx gene (Ingham, 1984; 1985). In order to investigate how widespread the requirement of sxc function for repression of PcG targets is, I tested sxc mutants for expression of additional well-established PcG target genes outside of their normal expression domains. I examined third instar mutant larvae, in which maternally contributed protein or mRNA should have dropped to low levels. Imaginal discs from wild-type larvae and larvae transheterozygous for two null sxc mutations (sxc<sup>1</sup>/sxc<sup>7</sup>) were immunostained with antibodies against the protein products of the Hox genes Ubx, Abd-B and Scr, as well as of engrailed (en), Distal-less (Dll) and teashirt (tsh), which have been shown to be derepressed in other PcG mutants (Oktaba et al,
sxc mutant discs show consistent derepression of each of these targets in invariant patterns, and the degree of derepression varies in the case of one target gene to another (figure 1). Hence, sxc function is required to maintain the repression of multiple Polycomb target genes in larvae.

It is important to note that apart from these homeotic transformations, pharate adults show no other obvious developmental defects.

**Figure 1:** sxc mutants show derepression of several PcG targets

Wing (W), haltere (H) and 3rd leg (3L) imaginal discs from wild-type and sxc/sxc7 (sxc-) larvae were immunostained with antibodies against the protein products of the indicated Polycomb targets (magenta), and co-stained with Hoechst to visualize DNA (blue). In wild-type animals, Ubx and Abd-B are not expressed in the wing disc, Scr is not expressed in the 3rd leg disc except in the peripodial membrane cells (marked by an asterisk), En expression is limited to the posterior compartment, Dll expression is confined to the wing pouch, and Tsh expression is confined to the notum and hinge regions of the wing. sxc larvae misexpress these genes in invariant patterns indicated by arrowheads. Ubx and Scr are misexpressed at high levels in the wing pouch and throughout the 3rd leg disc, respectively. Misexpression of Abd-B in the center and of tsh in two small crescents of the wing pouch, as well as of Dll in two dots in the wing hinge, is at lower levels. En misexpression in certain cells of the anterior compartment is difficult to detect. sxc mutants lose Dll expression in the central wing pouch (asterisk), likely because of secondary repression of Dll by ectopic Ubx and/or Abd-B present in this region (i.e. as recently demonstrated in another PcG mutant: Oktaba et al, 2008).
Cloning of the \textit{sxc} gene

Despite the early discovery of \textit{sxc} mutants, the gene in which the \textit{sxc} mutations had occurred remained molecularly uncharacterized for 25 years. The \textit{sxc} gene had been mapped to the centromeric heterochromatin-euchromatin junction on the right arm of chromosome 2 (2R), after the observation that \textit{sxc} mutants failed to complement deficiencies in this region (Ingham, 1984). In accordance with this, the \textit{sxc} \textsuperscript{2} allele was cytologically characterized as a translocation between chromosomes 2 and 3 (T(2;3)sxc\textsuperscript{2}), with a breakpoint occurring in the proximal 2R region (Ingham, 1984). Owing to the high content of repetitive sequences, this portion of the \textit{Drosophila} genome sequence was only assembled in 2004, hindering the previous drawing of a complete map of \textit{sxc} candidate genes present in this region.

In order to find the \textit{sxc} gene, deficiencies (Df) near the 2R centromere were tested for their ability to complement the \textit{sxc}\textsuperscript{-} phenotype. In parallel, these X-ray induced chromosomal deficiencies were fine-mapped by complementation tests with available lethal mutations located between the \textit{RpL38} and \textit{l(2)09851} genes. \textit{sxc} alleles complemented deficiencies removing regions proximal to \textit{Nipped-A} (Df(2R)M41A10) and distal to \textit{CG42345} (Df(2R)nap11), but did not complement deficiencies missing regions in between (for example, Df(2R)nap8), thus indicating that the \textit{sxc} gene is located in the 160 kb genomic interval between \textit{Nipped-A} and \textit{CG42345} (figure S1).

Within this interval, I sequenced portions of the open reading frames (ORFs) of \textit{CG2682}, \textit{CG10392}, \textit{CG10465}, \textit{CG10395}, \textit{CG10417} and \textit{CG42345} in genomic DNA amplified from \textit{sxc}\textsuperscript{1-7} heterozygotes. Only \textit{CG10392} showed allele-specific changes in \textit{sxc}\textsuperscript{1}, \textit{sxc}\textsuperscript{4}, \textit{sxc}\textsuperscript{5}, \textit{sxc}\textsuperscript{6} and \textit{sxc}\textsuperscript{7} (figure 2 A). \textit{sxc}\textsuperscript{1}, \textit{sxc}\textsuperscript{4}, \textit{sxc}\textsuperscript{5} and \textit{sxc}\textsuperscript{7} contain base pair changes that change an amino-acid codon into a premature termination codon (STOP) or into a codon for a different amino-acid (figure 2 A). \textit{sxc}\textsuperscript{6} also contains a point mutation, which destroys the 3’ splice site (splice acceptor) of intron 4 (figure 2 A). This probably results in an aberrant messenger RNA that contains a premature STOP codon, which is expected to lead to degradation of this transcript by non-sense mediated decay, or to the production of truncated protein. I found no lesions in the DNA from the T(2;3)sxc\textsuperscript{2} allele, nor from \textit{sxc}\textsuperscript{3}. It is likely that the \textit{sxc}\textsuperscript{2} and \textit{sxc}\textsuperscript{3} mutations occur in gene regulatory regions or in portions of the \textit{CG10392} ORF which I did not sequence.

\textit{CG10392} encodes O-linked N-acetylglucosamine (O-GlcNAc) transferase (Ogt), the highly conserved glycosyltransferase present in the nucleus and cytosol of animal and plant cells. Ogt catalyzes the transfer of single N-acetylglucosamine (GlcNAc) residues from the sugar-nucleotide
Results

Donor UDP-GlcNAc, to serines or threonines of a broad range of nuclear or cytosolic substrates. The Ogt protein contains 13 N-terminal tandem tetratricopeptide repeats (TPRs), and a C-terminal glycosyltransferase domain, composed of two lobes that fold together in the native structure (figures 2A and S2).

The molecular lesions found in 5 independent sxc alleles thus identified sxc as the gene encoding Ogt, and thus we renamed the gene sxc/Ogt.

**sxc is O-GlcNAc transferase (Ogt)**

I next tested whether the different sxc alleles produce Ogt protein, by probing sxc- extracts with an antibody raised against the N-terminus of human Ogt (named H-300), taking advantage of the fact that Drosophila and human Ogt proteins are highly similar (77% sequence identity) (figure S2 A). Protein extracts from imaginal discs of wild-type larvae and of sxc mutants carrying the sxc<sup>1</sup> allele in trans with another sxc allele were prepared, and comparable protein amounts of each extract were loaded, as evidenced by the content of α-tubulin (figure 2 B). Anti-human Ogt recognized a band of the predicted size of ~119 kDa in wild-type extracts (figure 2 B). Low levels of a truncated protein of ~80 kDa (asterix) was present in all sxc transheterozygotes, but was absent from wild-type extracts, and emanates from sxc<sup>1</sup> (figure 2 B and figure S3). In a subsequent Western, sxc<sup>7</sup> was also found to produce trace amounts of a ~50 kDa truncated protein (figure S3). The viable translocation allele sxc<sup>2</sup> produces apparently normal Ogt protein at seemingly wild-type levels, consistent with the fact that we did not find a molecular lesion in the sequenced portion of the Ogt gene in this allele. The translocation in sxc<sup>2</sup> might occur in a regulatory region outside of the Ogt coding sequence, and thereby possibly perturb the spatio-temporal expression of Ogt protein, leading to a mild sxc- phenotype. No Ogt protein produced from sxc<sup>3</sup> (for which we also did not find a mutation) or sxc<sup>6</sup> was detected by the H-300 antibody, suggesting that these alleles might be protein-null (figure 2 B). Alternatively, these alleles might encode an Ogt protein with N-terminal aberrations that abrogate recognition by H-300.

Interestingly, Ogt<sup>N948I</sup>, the protein with a point mutation in the catalytic domain encoded by sxc<sup>4</sup>, and Ogt<sup>Δ1031-1059</sup>, the C-terminally truncated protein encoded by sxc<sup>5</sup>, were both expressed as stable polypeptides (figure 2 B). The Asp948, mutated in sxc<sup>4</sup>, is conserved in bacterial (*Xanthomonas campestris*), fly, worm (*Caenorhabditis elegans*), mouse and human Ogt. It is located on the same loop as a conserved histidine, which makes contacts with a phosphate group of UDP-GlcNAc and is
Results

essential for Ogt catalytic activity (Martinez-Fleites et al, 2008; Clarke et al, 2008) (figure S2 C). Asp948 points away from UDP-GlcNAc, and hydrogen-bonds with residues of other parts of the catalytic domain. These interactions are predicted to correctly orient His950 to enable it to interact with UDP-GlcNAc. Substitution of Asp948 by isoleucine disrupts hydrogen bonding and is therefore predicted to result in incorrect positioning of the essential His950 residue, thereby abolishing Ogt’s catalytic activity.

OgtΔ1031-1059, encoded by sxc5, lacks the C-terminal helix which loops from the second half of the catalytic domain (CDII) back to the first (CDI) and is therefore required for cohesion of both halves. This is essential as both CDI and CDII contribute residues important for binding the UDP-GlcNAc donor and the target serine/threonine of the substrate protein at their interface, as well as for catalyzing the glycosyltransferase reaction (Martinez-Fleites et al, 2008) (figure S2 B).

In conclusion, the sxc4 and sxc5 mutations produce large amounts of (nearly) full-length but catalytically inactive Ogt, however the phenotype of these mutants is as severe as that of seemingly protein-null alleles, such as sxc3 of sxc6. This suggests that the glycosyltransferase activity of Sxc/Ogt, and by extension the GlcNAcylation of one or more substrates, is essential for Polycomb repression in Drosophila.

Figure 2: Molecular characterization of the Drosophila PcG gene sxc/Ogt

(A) Structure of the Ogt protein and lesions in sxc mutant alleles. The Drosophila Ogt protein (1'059 aa) is schematically drawn, with the 13 TPRs at the N-terminus and the C-terminal bipartite catalytic domain (CDI and CDII in grey and black,
respectively). Base changes in the corresponding positions of the Ogt coding sequence that were found in the DNA of sxc mutants are shown on the top, as well as the accompanying changes in the amino acid codon, or the disruption of the splice site in the case of sxc6.

(B) Detection of Ogt protein in larval extracts. Whole-cell protein extracts from wild-type and mutant larvae carrying sxc\textsuperscript{1} in trans with the other sxc alleles, were probed with antibodies against human Ogt (top), and α-tubulin (bottom) as a control of comparable loading amounts. Anti-Ogt recognizes a band of predicted size in wild-type extracts. Anti-Ogt recognizes aberrant Ogt proteins produced from certain sxc alleles, and does not recognize any protein produced from sxc\textsuperscript{3} nor sxc\textsuperscript{6}, raising the possibility that these alleles are protein-null. The asterisk indicates low levels of truncated protein produced by sxc\textsuperscript{3}. sxc\textsuperscript{7} was also found to produce trace amounts of a severely truncated Ogt protein in a subsequent Western (see figure S3). Seemingly wild-type levels of Ogt protein are produced from alleles sxc\textsuperscript{2}, sxc\textsuperscript{4} and sxc\textsuperscript{5} (note that in these cases, Ogt protein is produced by only one chromosome, as opposed to the wild-type where 2 copies of Ogt-producing alleles are present). Other faint bands are probably due to cross-reactivity of anti-Ogt with non-specific proteins. Positions of the 80 kDa and 58 kDa marker proteins are indicated.

**sxc/ogt mutants lack Ogt activity**

Among the hundreds of cellular proteins that have been reported to bear the O-GlcNAc modification, the most abundant seem to be the nuclear pore components at the nuclear envelope (Holt and Hart, 1986; reviewed in Hart et al, 2007). I verified that sxc/ogt mutants lose GlcNAcylation of nucleoporins, as was previously shown to occur in Drosophila S2 cultured cells upon Ogt RNAi (Jinek et al, 2004). The GlcNAcylated nuclear pore proteins can be readily visualized by staining with fluorescently-labeled wheat germ agglutinin (WGA), a lectin with high affinity for terminal GlcNAc residues. Indeed, staining of Drosophila salivary gland cells from wild-type 3\textsuperscript{rd} instar larvae with AlexaFluor488-WGA resulted in a strong nuclear-rim staining that overlapped with the detection of nucleoporins (figure 3). In addition, a strong cytoplasmic staining presumably resulted from labeling of terminal GlcNAc residues on highly branched sugars elaborated on proteins within the endoplasmic reticulum and Golgi organelles by other GlcNAc transferases unrelated to Sxc/Ogt (figure 3). In contrast, nuclear rim staining was absent in salivary gland cells from sxc\textsuperscript{1}/sxc\textsuperscript{7} mutant larvae, although nucleoporins were still intact. The truncated Sxc/Ogt proteins produced in these mutants thus lack glycosyltransferase activity, as predicted.
Results

Figure 3: \textit{sxc\textregistered ogt} mutants show loss of Ogt activity, as testified by the loss of GlcNAc on nucleoporins

Salivary glands from wild-type (\textit{wt}) and \textit{sxc\textregistered 1/sxc\textregistered 7} (\textit{sxc\textsuperscript{-}}) larvae were stained to visualize DNA (with Hoechst), nucleoporins (mAb414) and terminal GlcNAc residues (WGA-AlexaFluor488). mAb414 is a monoclonal antibody that recognizes the FxFG repeats present on multiple nucleoporins, such as Nup62 which is a well characterized Ogt substrate. Nucleoporins were only labeled by WGA in \textit{wt} (full arrowheads) but not \textit{sxc\textsuperscript{-}} cells (empty arrowheads).

It is unclear what effect the loss of glycosylation of nuclear pore components has. There is no evidence that either nucleoporin stability nor nuclear transport are affected upon loss of GlcNAc (for example, in \textit{Caenorhabditis elegans ogt} mutants: Hanover et al, 2005). Although we can not rule out that glycosylation of nuclear pore proteins by Sxc/Ogt somehow contributes to silencing of PcG target genes, we sought out to test whether Sxc/Ogt also modifies chromatin-bound transcription factors, which are more direct players in Polycomb repression.

O-GlcNAc is detected at PREs of several PcG target genes

Previous evidence for GlcNAcylation of chromatin-bound factors in fly cells comes from the observations that WGA binds to \textit{Drosophila} polytene chromosomes (Kelly and Hart, 1989). I therefore tested whether I could detect the GlcNAc modification on chromatin by anti-GlcNAc chromatin immunoprecipitation (ChIP).

For this, formaldehyde-crosslinked and sheared chromatin from wild-type and \textit{sxc\textregistered 1/sxc\textregistered 7} 3\textsuperscript{rd} instar larval imaginal discs was immunoprecipitated with a monoclonal antibody recognizing GlcNAc linked to serines or threonines of proteins (HGAC85). Recoveries of selected PRE and non-PRE regions of the Polycomb target genes \textit{Ubx, Abd-B, Scr, en, Dll, tsh} and \textit{pannier (pnr)}, as well as of
euchromatic and heterochromatic control regions that are not regulated by PcG proteins, were measured by quantitative PCR (qPCR). Interestingly, only PRE regions were enriched after anti-GlcNAc ChIP from wild-type chromatin (figure 4). Importantly, no enrichment was obtained from *sxc^l/sxc^7* mutant chromatin, confirming that the GlcNAc signals specifically emanate from proteins glycosylated by Sxc/Ogt.

The presence of GlcNAc at PcG protein binding sites suggested that a PcG protein itself might be a substrate of Sxc/Ogt. I note that ChIP experiments using the H-300 anti-Ogt antibody failed to produce any signals detectable at any of the *Ubx* regions, thus it remains unknown whether Sxc/Ogt is itself bound to chromatin.

**O-GlcNAc is present on Ubx PREs in both the OFF and ON states**

I asked whether the presence of the O-GlcNAc modification correlates with gene repression, or whether it is also present on a target gene when this gene is not being repressed by the Polycomb system (i.e. in a different tissue). *Ubx* is repressed by PcG proteins in all cells of the wing (*Ubx*OFF), but is expressed throughout the haltere and 3rd leg discs (*Ubx*ON). Therefore, I compared the GlcNAc profile at the *Ubx* gene in chromatins prepared from separate wing- (*Ubx*OFF) and haltere/3rd leg- (*Ubx*ON) imaginal discs.

O-GlcNAc signals were present at *Ubx* PREs in both wild-type wing and haltere/3rd leg disc chromatins, indicating that O-GlcNAc modified proteins are constitutively bound both when *Ubx* is repressed and active (figure 4). However, there are more GlcNAc-modified proteins bound at the internal *Ubx* PRE (bx PRE) in wing (*Ubx*OFF) chromatin compared to haltere/3rd leg disc chromatin (*Ubx*ON). Intriguingly, this has also been shown to be the case of PcG proteins that are members of the PRC1, PRC2 and PR-DUB complexes (Papp and Müller, 2006; Scheuermann et al, 2010).
**Figure 4**: Proteins modified by Sxc/Ogt are constitutively bound at the Ubx PREs

Chromatin was prepared from either wing or haltere/3rd leg imaginal discs, from wild-type (wt) and sxc1/sxc7 (sxc-) larvae. The recovery of the regions indicated by bars drawn above the schematized genes (numbers indicate the distance, in kilobases, from the transcription start site of the respective gene) after anti-GlcNAc immunoprecipitation, was measured by qPCR and is expressed as the percentage of input chromatin recovered. Mean values from experiments with 3 independent chromatins are presented, with error bars indicating mean +/- standard deviation. All tested PRE regions of the Ubx, Abd-B, Scr, en, Dll, tsh and pnr Polycomb targets were significantly enriched from wt chromatin. No notable signals were detected at other genic regions tested, nor at euchromatic or heterochromatic control regions that are not under Polycomb regulation. No enrichment was obtained from sxc- chromatin, confirming that the anti-GlcNAc antibody specifically recognizes GlcNAc residues added by Sxc/Ogt. High GlcNAc signals at the Ubx PREs in both wing (UbxOFF) and haltere/3rd leg (UbxON) wt chromatins were obtained, with a ~2.5-fold higher signal at the bx PRE in wing compared to haltere/3rd leg chromatins. Therefore, the GlcNAc profile across Ubx mirrors that of PRC1 or PRC2 members.

**Genome-wide distribution of O-GlcNAc on Drosophila larval chromatin**

In order to determine the genome-wide distribution of GlcNAc-modified proteins in Drosophila larval chromatin, Katarzyna Oktaba within the lab performed anti-GlcNAc ChIP-on-chip. Two independent chromatin preparations from wild-type wing imaginal discs were either subjected to immunoprecipitations with the anti-O-GlcNAc antibody (HGAC85) or affinity purifications with agarose beads covalently linked to WGA. The precipitated material was hybridized to high-density whole-genome tiling arrays and analysed with TileMap, using a stringent cutoff (Ji and Wong, 2005).
genomic regions significantly enriched by both the anti-O-GlcNAc antibody and the WGA lectin were analysed.

To allow the estimation of the overlap of GlcNAc sites with PcG protein binding sites, Kasia Oktaba also determined the genome-wide binding profile of the PRC1 member Polyhomeotic (Ph) in larval chromatin samples prepared from mixed wing, haltere and 3rd leg discs. Kasia compared the GlcNAc and Ph ChIP-on-chip results to the PhoRC binding profiles determined in wing, haltere and 3rd leg disc chromatin, that she had generated previously (Oktaba et al, 2008).

We found that, in accordance with the qPCR results, but at this much larger resolution and scale, GlcNAc signals were found to be localized at discrete regions that overlap to a great extent with sites bound by PcG proteins such as Ph, Pho and dSfmbt. An example of the visualization of the ChIP-on-chip signals in the chromosomal region termed the Antennapedia cluster, comprising the Hox genes labial (lab), proboscipedia (pb), Deformed (Dfd), Scr and Antennapedia (Antp), is shown in figure 5 B.

Stringent bioinformatic analysis led to the definition of high-confidence sets of 338 PhoRC binding sites and 1'681 Ph bound sites in larval chromatin (see Materials and Methods for the defining of each set of high-confidence binding sites). On a genome-wide scale, 235 of the 338 (i.e. 70%) high-confidence PhoRC binding sites overlapped with a Ph binding site. Ph and/or PhoRC high-confidence sites were included in only 490 of the 1'138 (i.e. 43%) total 10% top ranking GlcNAc sites; however Ph and/or PhoRC were present on nearly all (111/114) of the 1% highest-scoring GlcNAc sites.

We note that high thresholds were used to generate the high-confidence sets of PhoRC, Ph and GlcNAc binding sites, in order to minimize the inclusion of false-positive regions (i.e. unbound regions that are nevertheless considered as significant). The unusually high quality of the Ph ChIP-on-chip data allowed the definition of a large number of high-confidence Ph binding sites. It must be borne in mind that many true binding sites are not included in the high-confidence PhoRC and Ph sets of bound regions. Therefore, the overlaps shown in the Venn diagrams in figure 5 A likely underestimate the true overlap, which is difficult to ascertain due to the technical limitations of ChIP-on-chip experiments (affinity of the antibody for the fixed antigen, problems with antigen accessibility during the ChIP-ing procedure, biased chromatin fragmentation, skewing of relative enrichments of immunoprecipitated regions after the amplification steps prior to hybridization to microarrays, and so on). Indeed, analysis of the raw ChIP-on-chip data by eye showed us examples of sites which might only be bound by PcG proteins but are devoid of GlcNAc; however, intriguingly, we have not seen convincing examples of GlcNAc sites that are clearly unoccupied by a PcG protein.
The significant overlap between GlcNAc sites and PREs on a genome-wide scale led me to test whether one or several PcG proteins are directly GlcNAc-modified by Sxc/Ogt.

Figure 5: Overlap between GlcNAc and PcG binding sites in larval chromatin (work of Katarzyna Oktaba)

A. Venn diagrams showing the overlap between the high-confidence set of 338 PhoRC-bound sites, the 1'681 Ph-bound sites and either the 1'138 10% top-ranked GlcNAc sites (left) or the 114 1% top-ranked GlcNAc sites (right). Nearly all (111/114) of the best scoring GlcNAc sites overlap with a Ph or PhoRC binding site.

B. Screen-shot of the ChIP profiles of Pho (blue), dSfmbt (green), Ph (red) and GlcNAc (pink) across the Antennapedia complex. The Hox genes lab, pb, Dfd, Scr and Antp are represented with exons (black boxes) and introns (lines) of the alternative transcripts, above (plus strand) or below (minus strand) the genomic map (genomic coordinates from release 5). Hybridization intensities for oligonucleotide probes (resolution of 35 bp) are indicated in the y axis. Regions scored as high-confidence binding sites are marked by boxes underneath the corresponding peaks (in the case of O-GlcNAc, regions belonging to the 10% highest scoring are indicated). The microarray data have been deposited in the ArrayExpress database under the accession number E-TABM-697.
Sxc/Ogt glycosylates Polyhomeotic (Ph)

In order to test whether any PcG proteins are themselves GlcNAc modified, I used agarose beads with covalently linked WGA molecules to affinity-purify GlcNAc-modified proteins from soluble nuclear extracts of wild-type embryos. The purified material from a same experiment was then probed with antibodies against the PhoRC subunits Pho and dSfmbt, the PRC1 subunits Ph, Pc, Ring and Scm, the PRC2 subunits E(z), Su(z)12 and Nurf55, as well as the large subunit of RNA polymerase II (Rpb1).

GlcNAc affinity purification of these extracts led to notable recoveries of the core PRC1 members Ph, Pc and Ring, but not of the PRC1 substoichiometric component Scm, nor of the other PRC2 or PhoRC members tested (figure 6). From this experiment, it is not possible to conclude which amongst the Ph, Pc and Ring proteins are directly modified, as unglycosylated proteins can copurify by interacting with a GlcNAc-modified protein. I note that the large subunit of RNA polymerase II, which was reported in mammals to bear GlcNAc residues on several sites of the C-terminal domain (CTD) (Kelly et al, 1993), was not purified in this experiment.

![Figure 6: The core PRC1 complex is recovered after GlcNAc affinity purification of embryonic nuclear protein extracts](image)

Soluble nuclear extracts from wild-type embryos were subjected to affinity-purification with WGA-agarose under native conditions, and probed with antibodies against the indicated PRC1, PhoRC and PRC2 subunits, as well as against the large RNA Polymerase II subunit Rpb1. (I) 1% of input extract, (E) 10% of affinity-purified material released from the resin by GlcNAc elution. PRC1 components Ph, Pc and Ring are strongly enriched in WGA affinity-purified material; no enrichment of the other tested proteins is detectable. We estimate that ~5% of total core PRC1 was recovered after WGA-purification, if we consider that the signals observed in lanes E are half as intense as those observed in lane I.
In order to determine whether Ph, Pc or Ring bear the GlcNAc modification, and whether this modification is mediated by Sxc/Ogt, extracts from both wild-type and sxc¹/sxc⁷ larvae were prepared. Half of the wild-type extract was denatured prior to affinity purification, in order to disrupt protein secondary and tertiary structures, and thereby limit protein-protein interactions. Extracts containing comparable total amounts of proteins were then submitted to WGA pull-down. sxc¹/sxc⁷ starting material notably differed from the wild-type starting material in that it lacked full-length Sxc/Ogt (figure 7: compare lanes 5 to 1 and 3), as previously observed in figure 2; and, interestingly, sxc¹/sxc⁷ extracts contained higher amounts of Ph. Accumulation of Ph in sxc¹/sxc⁷ cells might reflect a failure to downregulate Ph expression by the PcG system through PREs present in both Ph-proximal and Ph-distal genes (Oktaba et al, 2008). After WGA affinity-purification, only Ph, but none of the other proteins tested, was strongly enriched from native wild-type extracts (figure 7: lane 4). Interestingly, Ph was also efficiently recovered after WGA pull-down from wild-type denatured extracts, suggesting that Ph is directly GlcNAc-modified (figure 7: lane 2). The Ph5 antibody that I used for Western blotting probably recognizes both Ph-p and Ph-d protein paralogs (see figure S10). The fact that two bands are clearly observed in the eluted material raises the possibility that both Ph paralogs are GlcNAc-modified. The traces of dSfmbt recovered after purification from wild-type extracts only under non-denaturing conditions (compare lanes 4 and 2 of figure 7) likely copurify through association with GlcNAc-modified Ph. Importantly, no proteins were purified from sxc¹/sxc⁷ extracts (figure 7: lane 6), confirming that glycosylation of Ph is mediated by Sxc/Ogt. A subsequent repetition of this experiment with independently prepared extracts and testing with additional antibodies also led to enrichment of Ph, but this approach again did not provide evidence for GlcNAcylation of the other PRC1 members Psc, Ring, Pc and Scm; the PRC2 members E(z), Su(z)12, Nurf, Esc and Pcl; the PhoRC members Pho and dSfmbt; the PR-DUB members Asx and Calypso; the C-terminal portion of Trx; the transcription elongation factor Spt5; and the large subunit of RNA polymerase II (figure S4).
**Figure 7 : Polyhomeotic (Ph) is GlcNAc-modified by Sxc/Ogt**

Extracts of wild-type (wt) and sxc\(^{-}/sxc\(^{-}\) (sxc\(^{-}\)) larvae were subjected to a pull-down with WGA-agarose under native or denaturing (denat.) conditions. 0.5% of the total input material (lanes I) and 10% of the total affinity purified material (lanes E) were electrophoresed and probed with antibodies against the indicated proteins. Only Ph was recovered after WGA pull-down from wild-type denatured extracts, and it was not purified from sxc\(^{-}\) extracts, indicating that Sxc/Ogt glycosylates Ph. We may estimate that over 10% of total cellular Ph is glycosylated, if we consider that the intensity of the Ph signal observed in lane 2 is more than twice as intense as that in lane 1. A greater proportion of Ph was purified from denatured extracts than from native extracts (compare lanes 2 and 4), likely thanks to the inactivation of lysosomal deglycosylases released upon cell disruption. Traces of dSfmbt and Ogt (and also Pc and Scm – seen on longer exposures) are visible in eluted material from wild-type extracts only in native conditions (compare lanes 2 and 4), hinting that they might have copurified through physical association with GlcNAcylated Ph. I do not consider the faint band recognized by the Pho antibody in the eluted material from wild-type denatured extracts as evidence for GlcNAcylated Pho protein, since this band is also visible with an equal intensity in material eluted from sxc\(^{-}/sxc\(^{-}\) extracts, which lack Sxc/Ogt activity (compare lanes 2 and 6). As the Pho band in lane 4 is stronger than in lanes 2 and 6, it is possible that trace amounts of Pho also copurify from wild-type native extracts through association with GlcNAcylated Pho. Finally, note that the Ring and dSfmbt proteins appear to migrate more rapidly in the sxc input (lane 5), but this is an artifact possibly arising from irregularities in the acrylamide gel (as evidenced by the migration of the molecular weight ladder on either side of the gel – not shown).
Other PRC1 core components did not copurify with GlcNAc-modified Ph in these WGA pull-down assays from larval extracts, although Ring and Pc were readily observed in purifications from embryonic nuclear extracts (figure 6). It is possible that differences in extract preparation (recovery of soluble proteins by osmosis of embryo nuclei versus sonication of whole larval cells) might affect the stability of protein complexes to different degrees. Alternatively, Ph might associate with distinct protein partners in larvae, since PRC1 has so far only been purified from Drosophila embryos.

Finally, as in the case of embryonic nuclear extracts, WGA-purification from larval extracts failed to provide evidence for GlcNAcylation of the large subunit of RNA polymerase II. The fact that Rpb1 was also not recovered under denaturing conditions argues against the possibility that Rpb1 bears GlcNAc residues that are unaccessible to binding to WGA because buried inside the protein structure.

In conclusion, the finding that Sxc/Ogt directly modifies the PRC1-member Ph tempts us to speculate that glycosylation of Ph by Sxc/Ogt might be an essential step in Polycomb repression. However, we note that Drosophila cells probably contain several glycosylated nuclear proteins in addition to nucleoporins and Ph, as WGA purifications from embryonic nuclear extracts led to enrichment of other potential yet-unidentified proteins bearing O-GlcNAc (figure S5). Therefore, it remains possible that O-GlcNAcylation of other Sxc/Ogt substrates than Ph may contribute to Polycomb function.

**Ph recruitment to PREs is impaired in sxc/ogt mutants**

I next tested whether binding of PcG complexes to chromatin is impaired in the absence of glycosylation mediated by Sxc/Ogt. Wild-type or sxc1/sxc7 wing chromatins were immunoprecipitated with antibodies against the DNA-binding PhoRC member Pho, the PRC1 subunit Ph and the PRC2 subunit E(z). Binding of Pho and E(z) at the analysed PREs was largely unaffected in sxc/ogt mutants (figure 8). Ph was also bound at high levels at these PREs, yet signals were 1.5 to 2-fold reduced at nearly all PREs examined (figure 8).
Figure 8: Ph levels are ~twofold reduced at many PREs in sxc/ogt mutants

Comparison of ChIP profiles of Ph (PhoRC), E(z) (PRC2) and Ph (PRC1), in wild-type (blue bars) and sxc1/sxc7 (orange bars) chromatins. Results are presented as in figure 4, except that signals obtained from wild-type and sxc mutant chromatins at each region are shown side-by-side. Asterisks indicate regions that were not measured by qPCR. Pho and E(z) are bound at comparable levels in both wild-type and sxc1/sxc7 chromatin at the majority of the tested PREs, but Ph levels are 1.5 to 2-fold reduced at most PREs in sxc1/sxc7 chromatin.

It is not known whether such a decrease of PRE-bound Ph molecules is sufficient to compromise Polycomb repression. This reduction might represent a generalized lower chromosomal occupancy of Ph in all mutant wing disc cells, or a more severe loss of chromatin-bound Ph in only a subset of wing disc cells (for example, in those which appreciably misexpress PcG target genes). Ph is reduced at PREs of genes which do not show obvious misexpression, such as in the case of en (figure 1) and pnr - for which I was unable to detect misexpression of a pnr-LacZ reporter transgene introduced into an sxc1/sxc7 genetic background (figure S6). This suggests that such a reduction of PRE-bound Ph levels might not be sufficient to lose Polycomb silencing. Moreover, sxc1/sxc7 mutant wing disc cells ectopically express Scr (figure S7) and Tsh (figure 1) proteins, despite Ph [as well as Pho and E(z)] being bound at wild-type levels at the Scr and tsh PREs in sxc1/sxc7 mutant wing disc chromatin. This hints that the reduced binding of Ph to PREs is most likely not the only cause of gene
derepression in sxc/ogt mutants. Nevertheless, it is possible that binding of other PcG proteins not tested here is more severely affected in sxc/ogt mutants.

**Sxc/Ogt is dispensable for PRC2’s histone methyltransferase activity**

Finally, I asked whether generation of the repressive histone mark H3K27me3 by PRC2 across target genes was compromised in sxc/ogt mutants, by performing ChIP with an antibody raised against the H3K27me3 modification. I first verified that histone H3 levels across target genes were comparable between wild-type and sxc\textsuperscript{1}/sxc\textsuperscript{7} wing disc chromatins, by anti-H3 ChIP. H3 levels are indeed highly similar, and dip at nucleosome-free regions such as PREs and transcription start sites (TSSs) occupied by RNA polymerase II (figure 9). Lower nucleosome-occupancy at the Ubx and Scr TSSs in sxc\textsuperscript{1}/sxc\textsuperscript{7} chromatin compared to wild-type can be explained by the ectopic transcription of these genes in a significant proportion of sxc\textsuperscript{1}/sxc\textsuperscript{7} wing disc cells (figure 1 and figure S7). Histone levels at the Ubx TSS has also been shown to be lower in haltere/3rd leg disc chromatins (UbxON) than in wing disc chromatins (UbxOFF) (Papp and Müller, 2006). As Abd-B is also misexpressed in a substantial amount of wing disc cells, I expected to also see a reduction of histone H3 levels at the Abd-B TSS in sxc\textsuperscript{1}/sxc\textsuperscript{7} chromatin. Abd-B is a complex gene with 5 distinct promoters which lead to 5 alternative transcripts. Primers were designed to amplify the first TSS region of Abd-B. The very high H3 occupancy that we observe at this region suggests that this promoter is not used in wing disc cells to produce Abd-B transcripts.

H3K27me3 levels across target genes are at wild-type levels in sxc\textsuperscript{1}/sxc\textsuperscript{7} mutant chromatin, except in the transcribed regions of Ubx, Scr and tsh (figure 9). Since H3K27me3 levels are comparable to wild-type outside of the regions through which RNA polymerase II passes to transcribe genes (for example, in the upstream region of Ubx), we believe that this reduction of H3K27me3 levels in sxc\textsuperscript{1}/sxc\textsuperscript{7} chromatin is not due to impaired enzymatic activity of PRC2, but is rather a secondary consequence of the derepression of these genes in a significant portion of cells within the wing disc. For example, passage of RNA polymerase II may cause turnover of the methylated histones and their replacement with unmodified ones, which are not remethylated by PRC2 quickly enough to restore H3K27me3 levels. Reduction of H3K27me3 within the coding regions of the other target genes en, Dll and pnr might not be detectable because derepression occurs in a smaller proportion of wing disc cells (figures 1 and S6). If no transcription emanates from the Abd-B TSS which we examined here, then
the PRE region located 2 kb downstream should also not be transcribed. Moreover, the iab-7 PRE tested at position +72 kb is situated after the Abd-B transcription unit. This would explain why I also do not detect reduced H3K27me3 levels across the analysed regions of Abd-B.

In conclusion, these results suggest that Sxc/Ogt function is dispensable for recruitment of PRC2 and its ability to trimethylate H3K27 on target gene chromatin.

![Figure 9: sxc/ogt cells derepress target genes despite seemingly intact PRC2 HMTase activity](image)

ChIP signals in wild-type (blue) and sxc1/sxc7 (orange) wing discs showing levels of H3K27me3, and of total H3 as a reference for nucleosome occupancy. H3K27me3 is present at high levels across target genes in sxc1/sxc7 mutant chromatin, suggesting that PRC2 HMTase activity is intact. Lower H3K27me3 levels in sxc1/sxc7 mutant wing discs are only observed in the transcribed regions of genes that are presumably derepressed in a significant portion of wing disc cells. Thus Sxc/Ogt function is dispensable for the histone methyltransferase activity of PRC2.
I have shown that the classical PcG gene *super sex combs* encodes the highly conserved O-GlcNAc transferase, that catalyzes the addition of single GlcNAc sugar residues to serines or threonines of a wide variety of proteins located in the nucleus and cytosol of animal and plant cells. Specifically, I identified molecular lesions in the DNA of several independent sxc alleles, which lead to the production of aberrant or no Ogt protein (figure 2), and showed that sxc mutants lack GlcNAcylation of nucleoporins located at the nuclear periphery (figure 3) and of chromatin-bound proteins (figure 4). Particularly, the identification of genetically null sxc mutations that lead to the production of high levels of (nearly) full-length Ogt enzyme, but that are predicted to disrupt its catalytic activity, point to a requirement for the GlcNAc transferase activity of Sxc/Ogt (and by extension, for the glycosylation of one or more Sxc/Ogt substrates) for Polycomb repression. *sxc/ogt* function is necessary for stable silencing of several well-characterized PcG target genes (figure 1). Intriguingly, the fact that *sxc/ogt* mutants do not show other obvious phenotypes apart from those arising from misexpression of Hox genes, suggests that Polycomb repression is the main process that critically depends on O-GlcNAcylation in *Drosophila*. Genome-wide analysis of the distribution of GlcNAc on wild-type larval chromatin revealed that it is highly enriched at PREs (figure 5). In accordance with this observation, I identified the PRC1-member Ph as an Sxc/Ogt substrate, but did not detect GlcNAcylation of other PcG proteins, nor of the large subunit of RNA polymerase II - which has been reported to be O-GlcNAc-modified in mammalian cells (Kelly et al, 1993) (figures 6 and 7). I also tried to gain insight into the function of O-GlcNAcylation at the molecular level. O-GlcNAc is present on the chromatin of the highly studied PcG target gene *Ubx* in both the ON and OFF states, paralleling the constitutive binding of PcG proteins (Papp and Müller, 2006; Scheuermann et al, 2010), and indicating that O-GlcNAc does not function as a switch that triggers gene silencing (figure 4). O-GlcNAc does not seem to be required for the recruitment of the DNA-binding PcG protein Pho, nor of the PRC2 complex, which generates high levels of the repressive H3K27me3 mark across the body of Polycomb-repressed genes in *sxc/ogt* mutant chromatin (figures 8 and 9). In contrast, O-GlcNAc seems to be required for wild-type levels of Ph binding at many PREs (figure 8). Nevertheless, my results suggest that this reduction in Ph chromosomal occupancy can not in itself explain the loss of Polycomb repression observed in *sxc/ogt* mutants. Further experiments are required to test whether Sxc/Ogt contributes to Polycomb silencing through GlcNAcylation of Ph, and/or through other yet unidentified substrates; as well as to unravel the
molecular function of the O-GlcNAc post-translational modification on biologically relevant Sxc/Ogt substrates.

Parallel identification of sxc/Ogt by Sinclair et al, 2009

The identification of the sxc gene was reported in parallel to us by the labs of Barry Honda and Hugh Brock (Sinclair et al, 2009). An additional sxc null allele, sxc<sup>2637</sup>, was described, containing a transposable P-element insertion in exon 1 and which does not produce any detectable Ogt protein. In vitro tests of GlcNAc transferase activity towards recombinant Nup62 (a well characterized Ogt substrate) in protein extracts prepared from transheterozygous or hemizygous sxc mutants, showed that the sxc<sup>2637</sup>, sxc<sup>i</sup>, sxc<sup>3</sup>, and importantly sxc<sup>4</sup> (encoding Ogt<sup>N984I</sup>) mutations all largely abrogate Ogt activity. Unequivocal proof that sxc is indeed Ogt was provided by the observation that expression of wild-type Ogt from a transgene in an sxc<sup>3</sup>/sxc<sup>2637</sup> mutant background was able to significantly rescue lethality and give rise to flies that are fertile and mostly phenotypically normal. An exciting observation was that human Ogt also very efficiently rescued sxc<sup>3</sup>/sxc<sup>2637</sup> mutants. This indicates that human Ogt is able to fulfill all functions of Drosophila Ogt, consistent with the high similarity between the fly and human Ogt protein sequences (figure S2).

Sinclair and coworkers also detected several sites of GlcNAc modification on polytene chromosomes from larval salivary glands in wild-type animals. Moreover, GlcNAc sites colocalized with binding sites of the Polycomb group proteins Ph and Pc. Finally, although GlcNAc was not detectable on the chromatin of sxc/ogt mutants, chromatin-bound Pc was still present at high levels. In conclusion, Sinclair and coworkers present independent proof for many of our observations.

Is Ph the major biologically relevant Sxc/Ogt substrate?

Intriguingly, a plethora of cytosolic and nuclear proteins have been described to be O-GlcNAcylated in mammals (reviewed in Hart et al, 2007), and it is probable that Sxc/Ogt glycosylates a large number of proteins in fly cells as well. The absence of a prevalent consensus sequence amongst Ogt substrates does not allow their bioinformatic prediction. A large-scale mass spectrometry study on proteins affinity purified through their GlcNAc residues, from soluble cytosolic extracts of Drosophila S2
cells, led to the isolation of 51 proteins, involved in a series of cellular processes including translation, cytoskeletal organisation, proteasomal protein degradation, metabolism and intracellular signaling (Sprung et al, 2005). (It is important to note, however, that some of these proteins might have copurified unspecifically.) Moreover, there are several unidentified O-GlcNAc-modified proteins in Drosophila embryo nuclei (figure S5; Kelley and Hart, 1989). Visualisation of the electrophoretic mobility of these proteins suggests that Ph is not the major O-GlcNAcylated protein present in the nucleus (figure S5; Kelley and Hart, 1989). This raises the risk that the role of Sxc/Ogt in Polycomb repression might be indirect, by affecting the function of other factors, which in turn has repercussions on the activity of PcG proteins.

For example, it is possible that loss of GlcNAcylation of nucleoporins in sxc/ogt mutants might compromise Polycomb repression by affecting the nuclear import of a factor important for Polycomb repression, such as a PcG protein itself. Despite the fact that nuclear membranes contain the highest density of O-GlcNAc modified proteins (Holt and Hart, 1986), the role of glycosylation of nucleoporins remains elusive. As in flies, Caenorhabditis elegans ogt mutant cells lose WGA staining at the nuclear rim, demonstrating that nuclear pore complexes lack O-GlcNAc (Hanover et al, 2005). Nevertheless, glycosylation of nucleoporins in the worm appears to be dispensable for the stability of nucleoporins, and for the nuclear import of several analysed transcription factors (Hanover et al, 2005). It is important to note, though, that Ogt function seems to be exceptionally dispensable in Nematodes, since unlike the case of most other organisms, C. elegans ogt mutants are viable and fertile and mostly phenotypically normal (Hanover et al, 2005). I have also shown that nucleoporin stability seems to be largely unaffected by loss of GlcNAcylation in Drosophila salivary gland cells (figure 3). Moreover, nuclear import of analysed PcG proteins seems to be unaffected, as evidenced by high levels of Pho and E(z) binding at PREs in sxc/ogt mutant chromatin (figure 8). In addition, preliminary stainings of salivary gland cells from wild-type and sxc/ogt mutant larvae suggested that Ph protein was not accumulating in the cytoplasm of sxc/ogt mutant cells, and thus GlcNAcylation does not seem to be necessary for proper import of Ph into the nucleus (figure S8).

The highly specific PcG phenotype of sxc/ogt mutants and their relatively late lethality (reminder: sxc/ogt mutant embryos lacking maternal wild-type Sxc/Ogt contribution survive until the end of embryogenesis) suggests that Sxc/Ogt directly participates in Polycomb repression, and that other cellular processes are largely refractory to loss of protein O-GlcNAcylation in Drosophila. It is tempting to speculate that the function of Sxc/Ogt in gene silencing may be to GlcNAcylate Ph.
The next important step in this work is thus to confirm whether glycosylation of Ph directly contributes to Polycomb repression or not. The best way to do so would be to test the ability of an engineered Ph protein, which is no longer capable of being GlcNAc-modified by Sxc/Ogt (because mutated in residues previously determined to bear the GlcNAc modification), to rescue the mutant phenotype of ph null flies. If glycosylation of Ph is indeed required for efficient Polycomb repression, then this should result in a PcG phenotype. Furthermore, if Ph is the only Sxc/Ogt substrate important for Drosophila development, we would expect Polycomb repression in the resulting ph mutants to be defective to a similar degree as in sxc/ogt mutants. In other words, loss of glycosylation on all Sxc/Ogt substrates in the nucleus and cytosol (occurring in sxc/ogt mutants) would cause the same developmental defect as loss of glycosylation on only Ph. It is important to recall that the Ph gene is tandemly duplicated in flies, giving rise to two closely related and partially redundant paralogs (Ph-p and Ph-d). Therefore, O-GlcNAc sites would have to be determined on both paralogs, and both Ph-p and Ph-d proteins would have to be engineered if they are both GlcNAcylated, and expressed in a ph-p, ph-d double mutant to test for rescue.

It could also be insightful to test whether Ph is the major GlcNAcylated protein bound to Drosophila chromatin. This could be done by performing anti-GlcNAc ChIP using chromatin prepared from ph mutants, in order to test whether GlcNAc is still detectable on PREs in the absence of Ph protein. The limit of this experiment arises from the fact that secondary reductions of chromatin-bound proteins, such as other PcG proteins, may occur upon loss of Ph. Therefore, diminished GlcNAc levels might not be solely attributable to the reduction of chromatin-bound Ph. I performed initial attempts to analyse chromatin from larval imaginal discs in which both Ph-p and Ph-d levels were effectively knocked-down by tissue-specific RNAi. However, comparison of histone H3 occupancy at several chromosomal regions suggested that the chromatin structure is severely affected in cells with reduced Ph levels, precluding controlled comparisons with wild-type chromatin. The severity of the aberrations in chromatin structure may be linked to the tumorous nature of Ph knock-down cells, which undergo uncontrolled rounds of cell division.

Other candidates for biologically relevant Sxc/Ogt substrates

Although Ph is an attractive candidate, studies in mammals hint that glycosylation of other nuclear factors, if conserved in the fly, might also contribute to Polycomb repression. In mammals, O-
GlcNAc seems to be involved in much wider transcriptional regulation, given the long list of mammalian transcription factors that have been identified as being modified with O-GlcNAc (reviewed in Hart et al, 2007). It would be interesting to test whether certain of these transcription factors, that could be foreseen to influence Polycomb-mediated silencing, are also Sxc/Ogt substrates in *Drosophila*. We have done this in the cases of the large subunit of RNA polymerase II and of the fly YY1 homolog, Pho.

Early studies reported that the C-terminal domain (CTD) of the large subunit of RNA polymerase II, purified from calf thymus, is itself GlcNAc-modified on the serine/threonine at position 4 of the heptamer repeat sequence that composes the CTD (Kelly et al, 1993) (see figure S9 for an alignment of fly and human CTD sequences). The observation that the glycosylated and phosphorylated forms of the CTD are distinct suggested that glycosylation and phosphorylation of the CTD are mutually exclusive (Kelly et al, 1993). *In vitro* studies further suggested that O-GlcNAc and phosphate modification of the CTD are mutually exclusive at the level of the enzymes responsible for their addition, since recombinant OGT and TFIIH (the basal transcription factor responsible for phosphorylation of the CTD at serine 5) were incapable of modifying a CTD fragment that had previously been phosphorylated or O-GlcNAcylated, respectively (Comer and Hart, 2001). As phosphorylation of the CTD is implicated in promoter clearance and elongation, it was proposed that glycosylation of RNA polymerase II might repress transcription. We therefore hypothesized that Sxc/Ogt might be recruited to PcG target genes, and silence genes in the OFF state by glycosylating RNA polymerase II present at the TSS. However, our GlcNAc ChIP experiments argue against the existence of glycosylated RNA polymerase II on *Drosophila* chromatin. Indeed, O-GlcNAc signals at the TSSs of the analysed PcG target genes were not above background (figure 4). Moreover, we were not able to detect glycosylation of the large subunit of RNA polymerase II in our WGA pull-down experiments from extracts prepared from embryonic nuclei nor from larval cells (figures 6 and 7). It is possible that GlcNAc residues on the CTD are inaccessible to binding to the WGA affinity resin because occluded within the native protein structure. However, the fact that we did not detect RNA polymerase II enrichment even under denaturing conditions argues against this possibility (figure 7). Further evidence that glycosylation of the CTD of RNA polymerase II might not be an evolutionarily conserved mechanism of gene repression comes from the observation that O-GlcNAc is also present on chromatin in *C. elegans*, but peaks in average 100 bp upstream of gene TSSs (Love et al, 2010).

YY1, the mammalian Pho homolog, has also been reported to be O-GlcNAc modified in human cells (Hiromura et al, 2003). The O-GlcNAcylated form of transcription factor YY1 interacts less efficiently with the retinoblastoma protein (Rb) (Hiromura et al, 2003). Glycosylated YY1 shows
increased binding to DNA, although glycosylation does not increase YY1’s affinity for DNA, but rather presumably because of its release from the Rb protein (Hiromura et al, 2003). However, we also did not detect Pho enrichment after WGA-agarose pull-down from embryo nuclear extracts nor from larval extracts. It is nonetheless possible that Pho is glycosylated, but not at as many sites as Ph, or that GlcNAc residues on Pho are not clustered in a way that increases the affinity of binding by WGA. Indeed, it is important to bear in mind that our WGA-agarose pull-down experiments may have failed to efficiently enrich all O-GlcNAcylated proteins.

Speculation on the possible roles of GlcNAcylation of Ph in Polycomb repression

In relation to the large number of described O-GlcNAc-modified proteins, convincing evidence that O-GlcNAcylatation directly influences a modified protein’s function has only been obtained in a comparatively small number of cases. The proposed functional consequences of O-GlcNAc addition are diverse, and include protein stabilization (Yang et al, 2006), nuclear translocation (Dentin et al, 2008), modulation of interaction partners (Hiromura et al, 2003; Fujiki et al, 2009), and direct stimulation of enzymatic activity in the case of an O-GlcNAc-modified enzyme (Fujiki et al, 2009). Thus, O-GlcNAcylation of different proteins entails different molecular consequences. Moreover, it has been suggested that O-GlcNAcylation of different sites on a same protein can lead to different functional consequences (Yang et al, 2006). It has been shown in many cases that glycosylation exerts such effects by antagonizing the phosphorylation of the same or other residues within the same protein (reviewed in Hart et al, 2007).

If glycosylation of Ph is indeed important for the maintenance of transcriptional repression of PcG target genes, what might the molecular consequences of GlcNAc modification of Ph be? GlcNAcylation of Ph does not seem to be essential for Ph protein stability, since Ph was able to accumulate in sxc/ogt mutants to levels that were higher than in the wild-type (figure 7). Despite this increase, the levels of Ph at analysed target genes were 1.5-2 fold reduced on sxc/ogt mutant chromatin, which indicates that GlcNAc-modification of Ph might be necessary for efficient recruitment or stabilization of Ph on chromatin (figure 8). As previously mentioned, GlcNAcylation does not seem to be necessary for proper import of Ph into the nucleus (figure S8). Reduced recruitment of Ph to chromatin might arise from the destabilization of Ph-containing complexes, which might contain proteins necessary for chromosomal targeting. Nevertheless, Ph was present at wild-type levels at certain target genes which
were nonetheless derepressed in a significant portion of sxc/ogt mutant wing disc cells, for example in the case of Scr (figures 8 and S7). Therefore, I speculate that derepression of PcG target genes is not solely explained by a reduction in chromosomal occupancy by Ph.

It is possible that GlcNAcylation of Ph might boost the capacity of chromatin-bound Ph to exert its activity as a repressor. However, Ph should still maintain partial repressor activity in the absence of GlcNAcylation, since the ph mutant phenotype is by far more severe than that of sxc/ogt mutants (Dura et al, 1987; Oktaba et al, 2008). It is possible that only the most sensitive Ph targets become misregulated in the absence of GlcNAcylation, or Sxc/Ogt might only be required for regulation of a subset of Ph target genes. Although the requirement of Ph for Polycomb repression is undoubtedly one of the most critical amongst all other PcG proteins in Drosophila (i.e. the phenotype of ph null mutants is among the most severe PcG phenotypes), the molecular mechanisms through which Ph contributes to repression remain unclear.

Several lines of evidence suggest that Ph might serve as a scaffolding protein. Ph contains 3 domains at its C-terminus which are conserved in Vertebrate orthologues: a homology domain (HD) of unknown function, an FCS-type zinc finger and a SAM domain (figures ii and S10). As mentioned in the introduction, the SAM domains of Scm and Ph mediate their interaction in vitro (Peterson et al, 1997; Peterson et al, 2004; Grimm et al, 2009). An FCS-type zinc finger is also present in Scm and dSfmbt, and possibly mediates their interaction in vitro (Grimm et al, 2009), raising the possibility that in Ph it may also function as a protein-protein interaction domain. It is therefore possible that Ph serves to bridge interactions between PcG or non-PcG proteins.

The bulkiness and amphipathic nature of GlcNAc moieties may alter Ph’s interactions with partner proteins. For example, it would be possible to test whether the assembly of PRC1 is dependent on GlcNAcylation of Ph, by treating purified or recombinant PRC1 with a recombinantly expressed O-GlcNAc hydrolase. Alternatively, Ph might be incorporated into other complexes in a GlcNAc-dependent manner. Indeed, there is evidence that Ph may be part of other multiprotein complexes than PRC1 (Lagarou et al, 2008). Although the PRC1 core complex members Pc and Ring copurified with Ph in the WGA-agarose purification from wild-type embryos (figure 6), other proteins such as dSfmbt seemed to copurify with Ph in such purifications from larvae (figure 7), suggesting that Ph might be part of different complexes at different stages of development. It could therefore be interesting to compare the protein interaction partners of Ph in the presence and absence of GlcNAcylation. GlcNAcylation of Ph does not, in any case, seem to be required for the protein stability of other PcG
proteins, since their protein levels are comparable in wild-type and sxc/ogt larval tissues (figures 7 and S4).

GlcNAc-dependent assembly of Ph-containing complexes might be a prerequisite for efficient recruitment of these proteins together to chromatin, and/or for the functionality of the other PcG proteins within the complex. It is therefore possible that the chromosomal occupancy of yet untested factors in addition to Ph itself might be compromised in sxc/ogt chromatin. We have shown that binding of the DNA-binding protein Pho and of the PRC2 catalytic subunit E(z) are largely unaffected in sxc/ogt mutant chromatin. Recruitment of the other PRC2 members, which are all necessary to generate high-levels of trimethylated H3K27 across PcG target genes (see Introduction), is probably also not affected in sxc/ogt mutants, since we have shown that H3K27me3 levels are largely unperturbed (figure 9). Thus, it would be interesting to test binding of other PRC1 members and of dSfmbt by additional ChIP experiments in wild-type and sxc/ogt chromatins. Reduced binding of PRC1 might affect its capacity to silence target genes by compacting chromatin, or by interfering with the general transcriptional machinery present at the TSS (see Introduction).

Evolutionary conservation of the role of Ogt in Polycomb repression

Interestingly, recent proteome-wide mass spectrometry studies in mammals, with the aim of discovering new O-GlcNAc-modified proteins, have identified O-GlcNAc sites on the Polyhomeotic-like 3 (PHC3) mouse and human Ph homologs (Chalkley et al, 2009; Wang et al, 2010b). The identified GlcNAcylated threonines are in the N-terminal portions of the PHC3 proteins, which are not well conserved in Drosophila Ph (see figure S10), thereby precluding the prediction of which residue(s) of Drosophila Ph might be modified. It is nevertheless possible that additional and possibly conserved GlcNAc sites on PHC3 exist which were not identified in these studies. This is because these studies did not reach saturation in the detection of all GlcNAcylated peptides present in the samples; and also because the sequence coverage of the PHC3 proteins by the peptides observable by mass spectrometry is not full. In addition, several cases of proteins which are O-GlcNAc-modified on several residues have been reported, for example within these same studies (Chalkley et al, 2009; Wang et al, 2010b). It has been shown in some cases that only a few O-GlcNAc sites on a protein are functional (Yang et al, 2006; Dentin et al, 2008); and it is possible that different GlcNAcylation sites on the same protein have different functions.
In any case, these studies indicate that the role of glycosylation of Ph by Ogt might be evolutionarily conserved between flies and mammals. Reinforcement of Polycomb repression might be the major role of Ogt in flies, whereas Ogt may have more pleiotropic effects in mammals by participating in several other parallel processes, which complicates the appreciation of the contribution of Ogt to the specific process of Polycomb repression (see next section).

It has not yet been investigated whether the role of Ogt in Polycomb repression might also be conserved in *C. elegans*. Phenotypic analysis of *ogt* worm mutants has thus far only revealed a role of Ogt in the negative regulation of insulin signaling (Hanover et al, 2005; Love et al, 2010). Nonetheless, this does not disprove a possible impairment of Polycomb repression in these animals, as misexpression of Hox genes in *C. elegans* leads to subtle cell lineage defects that are not obvious upon casual examination. Moreover, a recent O-GlcNAc ChIP-on-chip experiment similar to the one that we report here has shown that, as in flies, O-GlcNAc is detected at discrete sites on Nematode chromatin (Love et al, 2010). O-GlcNAc was found associated with 828 genes (Love et al, 2010). Unlike the RNA polymerase II binding profile, which peaks at the TSSs of many genes, O-GlcNAc peaks were centered 100 bp upstream of TSSs (Love et al, 2010). In accordance with what we have shown in *Drosophila*, there was no clear correlation between O-GlcNAc occupancy and the transcriptional activity of genes (Love et al, 2010). Importantly, O-GlcNAc was no longer detected on the chromatin of *ogt* animals (Love et al, 2010). Taken together, these observations suggest that O-GlcNAc, while not essential in the nematode, may nevertheless play a critical role in transcriptional regulation by PcG proteins.

Taken together, the studies of the biological roles of Ogt in other organisms point to a conserved implication of O-GlcNAc in transcriptional regulation, and do not rule out that a role of Ogt in Polycomb repression may be ancient and evolutionarily conserved. *Drosophila* appears to be a powerful model for dissecting the possible contribution of Ogt to Polycomb repression, for the principal reason that other major biological processes do not seem to be perturbed in *sxc/ogt* animals, as suggested by the highly specific Polycomb phenotype. Moreover, the late lethal stage of *sxc/ogt* mutants opens the door to many genetic and biochemical experiments in view of unraveling the molecular mechanisms through which glycosylation may contribute to gene silencing.
Contribution of this work to understanding the function of O-GlcNAc

The O-GlcNAc modification has been identified on a plethora of mammalian proteins, and has been shown to be highly dynamic and responsive to environmental cues (reviewed in Hart et al, 2007). Indeed, O-GlcNAc modification can be reversed by an antagonistic enzyme to Ogt, which is also present within the nucleus and the cytosol, called O-GlcNAc hydrolase (Oga), and O-GlcNAc has been shown to turn-over several times during the life of a modified protein. Given the wide variety of biological functions attributed to Ogt substrates, it has been supposed that O-GlcNAc plays a pleiotropic role in regulating several cellular processes as diverse as transcription, intracellular signaling and organization of the cytoskeleton (Hart et al, 2007). In mammalian cells, O-GlcNAc is believed to be implicated in several biological phenomena such as cell differentiation, the development of insulin resistance, lymphocyte activation and others (Hart et al, 2007). This has suscitated great interest in the potential functions of the O-GlcNAc post-translational modification, which shares many characteristics and seems to interplay with protein phosphorylation (Hart et al, 2007). Despite the discovery of O-GlcNAc a quarter of a century ago (Holt and Hart, 1986), its consequences for the functions of the proteins that bear it remain unknown in the majority of cases. The study of the function of O-GlcNAc is complicated by the fact that its presence is required for cell viability in Vertebrates, or results in early embryonic lethality (reviewed in Love et al, 2010). In addition, because it is a widespread modification, it is thought that perturbation of O-GlcNAc levels in cells leads to pleiotropic effects, complicating the evaluation of the importance of O-GlcNAc on a protein-by-protein basis.

In order to circumvent the limitations of manipulability of O-GlcNAc levels in Vertebrates, the biological requirements for Ogt have been investigated in other organisms. Ogt is a single copy gene in nearly all animals. The Nematode C. elegans seems to represent an opposite extreme, where Ogt seems to play a regulatory, but nonessential, role (Hanover et al, 2005). Through the work presented here, we have found that in Drosophila, Ogt seems to be primarily required for polycomb repression. The late lethality of sxc/ogt flies has allowed me to perform biochemical experiments that have supported the view that Sxc/Ogt is the sole glycosyltransferase that catalyzes O-GlcNAcylation in flies. Drosophila sxc/ogt mutants are thus an exciting new model to study the function of this abundant but mysterious post-translational modification. Further studies aimed at elucidating the functional consequences of glycosylation of Ph promise to bring an important contribution to the O-GlcNAc field, where evidences that O-GlcNAc modification at a specific site on a given protein alters its biological
properties are still much sought-after. Indeed, such demonstrations are necessary in order to confirm that O-GlcNAc is truly a key regulatory post-translational modification (Wells, 2007).

**Link between Polycomb silencing and diet**

Changes in intracellular O-GlcNAc levels are highly responsive to nutrient flux, leading to the idea that Ogt integrates information about the metabolic status of the cell through glycosylation of its nucleo-cytosolic substrates (reviewed in Love et al, 2010). The requirement of glycosylation for effective Polycomb repression would make this mechanism highly dependent on an energy-rich diet, and might explain the well known but poorly reported dependence between the degree of Polycomb silencing and diet (Schwartz and Pirrotta, 2009). This connection moreover raises exciting possibilities for the role of diet or calorie restriction on the wide range of processes that are affected by PcG mechanisms, including developmental processes, differentiation and tumor growth. The biological significance of such coupling remains to be investigated.
Conclusion

The work presented here suggests that glycosylation by Sxc/Ogt is a critical step in Polycomb repression, thus identifying an unprecedented role for the O-GlcNAc post-translational modification in the maintenance of heritable gene repression by the PcG system. We hypothesize that Sxc/Ogt might mediate its contribution to Polycomb repression through glycosylation of the PRC1-member Ph. Further investigation of the possible molecular consequences of O-GlcNAcylation of Ph may lead to significant contributions to both the Polycomb and O-GlcNAc research fields. On the one hand, it might provide further insight into possibly conserved molecular mechanisms of Polycomb repression, since glycosylation of Ph homologs is conserved in mammals. On the other hand, it would contribute much sought-after evidence to consolidate O-GlcNAc as a key regulatory post-translational modification, in the context of a developing organism.
Materials and Methods

Fly husbandry

Flies were grown on cornmeal-molasses-yeast medium supplemented with the mold inhibitor propionic acid. Crosses were performed at 25°C, and offspring was raised at 18°C, in order to slow down development and thereby allow the obtention of fat third instar larvae, from which protein extracts (for Western blot) or chromatin (for ChIP) were prepared. For complementation tests (see next paragraph), crosses were maintained at 25°C in order to accelerate the obtention of adult progeny, and to increase the penetrance of genetic adult markers such as Cy (causes curled wings).

Complementation tests in view of mapping sxc

Deficiencies and mutant alleles for genes within the [RpL38; l(2)09851] region on proximal 2R were obtained from the Bloomington fly stock center (Indiana, USA). Non-complementation of mutant stocks was testified in most cases by the absence of transheterozygous adult progeny (i.e. pre-adult lethality), or by the observation of a characteristic mutant phenotype in adult progeny (atrophied wings in the case of the apterous mutation). Progeny of each expected genotype were counted, by inspecting for the presence/absence of dominant markers conferred by balancer chromosomes, as well as for eye color. Indeed, nearly all of the mutant alleles ordered carried mutagenic insertions of transposable elements, which often contained a white transgene conferring darker eye color. Crosses were scored daily for at least one week after the beginning of the pupal eclosion, in order to wait for the appearance of progeny with developmental delay. In this way, I was able to confidently conclude as to the results of complementation tests. For complementation tests with sxc alleles, all deficiencies were crossed to at least sxc\(^2\), and in many cases to all 7 sxc alleles, and progeny were inspected for pharate adults, notably pharate adult males with extra sex combs. I was not able to reproduce results of complementation testing that had originally defined Df(2R)NC32 (sxc\(^2\)) as a chromosomal deletion.
(Myster et al, 2004); and subsequent identification of a point mutation in the coding sequence of Ogt confirmed that sxc\(^7\) allele was indeed not a deficiency.

**Sequencing sxc alleles**

Genomic DNA was prepared from \(\text{sxc}^{1-7}\) balanced heterozygotes. PCR oligos were designed to amplify exonic regions (with intron-exon junctions included) of candidate genes, in fragments of \(~\text{500 bp} - \text{3 kb}\) in length. Purified PCR products were sequenced by the EMBL Genomics Core Facility. Sequencing oligos were designed every \(~\text{600 bp}\), on both the plus and minus strands. In this way, nearly all DNA regions were sequenced at least twice independently, and “read” in both directions. Reads were aligned using Vector NTI software (Invitrogen). Chromatograms were scanned by eye for double peaks seen reproducibly in the independent sequence reads. Sequence differences between the mutagenized sxc chromosome and the balancer chromosome were called polymorphisms when this difference was shared by two or more independent sxc alleles. In this manner, we noticed that \(\text{sxc}^{1,2,6,7}\) on one hand, and \(\text{sxc}^{3,4,5}\) on the other hand, shared all polymorphisms. This can be explained by the fact that \(\text{sxc}^{3,4,5}\) were generated by mutagenizing a common \(\text{b, pr, cn, bw}\) chromosome (Ingham, 1984); whereas \(\text{sxc}^{1,2}\) and \(\text{sxc}^{6,7}\) were induced on a similar genetic background: on \(\text{cn, bw}\) (Ingham, 1984) or \(\text{cn, bw}\) (Myster et al, 2004) chromosomes, respectively. Only regions containing at least one polymorphism, testifying that the corresponding region had been amplified from both the sxc and balancer chromosomes, were considered as effectively sequenced. Ogt (CG10392) was the only gene in which we found allele-specific changes. Mutations suspected from the observation of double peaks were confirmed by cloning the PCR product amplified from sxc heterozygotes, and sequencing clones to confirm the presence of both variants at a given base position. The following oligo pairs were used for amplifying the sequenced portions of Ogt:

\[
\begin{align*}
\text{Ogt-3} : & \text{CAGTGTCTTTACTTAAGGGTGG} \quad \text{and Ogt-4} : \text{CTGGTATCTTTCATCGAAGC} \\
\text{Ogt-7} : & \text{AGCTAGTGGAAATGTGGACG} \quad \text{and Ogt-6} : \text{CTCTGCCGTGATCCTTAGG} \\
\text{Ogt-14a} : & \text{GCAAGGCATTTTGCACATGC} \quad \text{and Ogt-15a} : \text{GATTCCCATGTCCAAATCG} \\
\text{Ogt-10a} : & \text{CAAGCAACGTCAGAACC} \quad \text{and Ogt-11a} : \text{GGAAACGCTTCCCTCTCTACC}
\end{align*}
\]
**Immunostaining of larval tissues**

Inverted carcasses with imaginal discs and brains, or salivary glands, attached were dissected from wandering third instar larvae, and fixed for 20 minutes at room temperature, in 4% formaldehyde in PBS. Carcasses were then washed with PBT (0.1% Tween20 in PBS), blocked by washes with BBT (1% BSA, 0.1% Triton X-100 in PBS), and labeled with antibodies and/or Alexa488-WGA in a small volume of BBT (~50 uL) overnight at 4°C. After extensive washes with BBT, carcasses were incubated with fluorescently-labeled (Cy3) secondary antibodies and ~50 ng of Hoechst33342 DNA stain. After washes with BBT and then PBT, imaginal discs or salivary glands were mounted on microscope slides in Fluoromount-G. Confocal images were taken using a Leica SP2 confocal microscope.

**Anti-Ogt Western blot on larval extracts**

Protein extracts were prepared by boiling dissected imaginal discs and brains from 3rd instar larvae (for figure 2) or inverted cuticles of anterior quarters of larvae from which the the fat body and digestive tract had been removed (for figure S3), in 2xSDS sample buffer [120 mM Tris-Cl pH 6.8 @ 25C / 200 mM DTT / 4% (w/v) SDS / 0.02% bromphenol blue / 20% (v/v) glycerol]. Approximately 3 larvae were loaded per lane. Wild-type extracts were prepared from w¹¹¹⁸ larvae (which contain a mutation in the white gene required for red pigmentation of the eyes). sxc transheterozygotes were obtained by crossing sxc¹ / CyO GFP males to sxc²,⁴,⁵,⁶,⁷ / CyO GFP or sxc³ / SM5^TM6B virgin females (for figure 2) or sxc⁷ / CyO GFP virgin females to sxc¹,²,⁴,⁵,⁶ / CyO GFP males or sxc³ / SM5^TM6B males (for figure S3). Transheterozygote progeny were selected by the absence of the tubby larval marker and/or of GFP.

For the tubulin loading control presented in figure 2, extracts were diluted 20 times, in order to avoid saturation of the Western blot signal.

Proteins were transferred onto a nitrocellulose membrane by semi-dry blotting at 15V for 20 minutes [transfer buffer : 25 mM Tris pH 8.4-8.9 / 192 mM Glycin / 0.1% (w/v) SDS / 20% EtOH] (for figure 2) or wet protein transfer at 90V for 90 minutes [transfer buffer : 25 mM Tris pH 8.3 / 192 mM Glycine / 0.05% SDS / 20% methanol] (for figure S3). Membranes were then blocked in 5% (w/v) milk (for figure 2) or BSA (for figure S3). Primary antibodies were incubated overnight at 4°C, and
secondary antibodies for at least 2 hours at room temperature. Horseradish-peroxidase labeled secondary antibodies were used, and bands were visualised by chemiluminescence.

**Chromatin preparation from larval imaginal discs**

Wild-type chromatin was prepared from *w^{1118}* larvae (figures 4 and 8) or OregonR larvae (figure 9). *sxc^{i}/sxc^{7}* mutant third instar larvae were obtained from a large-scale cross between *sxc^{7} / CyO GFP* virgins and *sxc^{i} / CyO GFP* males. Eggs were laid on apple juice agar plates, and embryos were allowed to develop to 1st-2nd instar larvae, before non-GFP larvae were transferred into yeasted vials. Wandering 3rd instar larvae were picked for dissection after double-checking that they were indeed non-GFP.

The chromatin preparation protocol was modified from Orlando et al, 1997. Larvae were dissected in ice-cold PBS and, after removing gut and fatbody, inverted carcasses with attached discs were fixed for 20 minutes at room temperature by gently mixing in 1 mL crosslinking solution (1.8% formaldehyde / 50 mM Hepes pH 8 / 100 mM NaCl / 1 mM EDTA / 0.5 mM EGTA). The crosslinking solution was changed 4 times during fixation and, after final removal, crosslinking was stopped by washing for 10 minutes in 1 mL PBS / 0.01% Triton X-100 / 125 mM glycine with 5 buffer changes. Fixed carcasses were then washed for 10 minutes each in wash A (10 mM Hepes pH 7.6 / 10 mM EDTA / 0.5 mM EGTA / 0.25% Triton X-100), followed by wash B (10 mM Hepes pH 7.6 / 200 mM NaCl / 1 mM EDTA / 0.5 mM EGTA / 0.01% Triton X-100), changing the solution 5 times during the course of each wash. Carcasses were stored at 4°C in wash B for up to 5 days, until ~ 900 larvae (~300 larvae per each of the 3 independent chromatin preparations) were dissected.

For the ChIP experiments presented in figures 4, 8 and 9: wing imaginal discs were separated from haltere and 3rd leg imaginal discs in wash B. 3 pools of each kind of discs were generated from ~300 larvae each. Disc pools were sonicated in parallel, each in 2 mL sonication buffer (10 mM Hepes pH 7.6 / 1 mM EDTA / 0.5 mM EGTA / 1x “Complete” protease inhibitor cocktail – Roche) with a Bioruptor sonicator water bath (Diagenode, 10 x 30 seconds on / 30 seconds off cycles, high-energy settings). After sonication, N-lauroylsarcosine was added to 0.5% and the solution was incubated for 10 minutes at 4°C on a rotating wheel. Insoluble material was removed by centrifugation at 4°C for 10 minutes at 16’000 g, and soluble chromatin in the supernatant was dialysed using a membrane with a molecular weight cut-off of 3.5 kDa, at 4°C against dialysis buffer (10 mM Tris-HCl pH 8 / 1 mM EDTA /
0.5 mM EGTA / 4% glycerol) overnight, with two changes of dialysis buffer during the first few hours of dialysis. The dialysed chromatin was aliquoted, snap-frozen in liquid nitrogen and stored at -80°C.

For the ChIP-on-chip experiments presented in figure 5: wild-type chromatins were prepared from mixed wing, haltere and 3rd leg discs (Oktaba et al., 2008).

ChIP assays

The following procedure was modified from Orlando et al, 1997. For each chromatin preparation, 30 uL (corresponding to ~6 discs) were removed and the DNA content was quantified by qPCR after reversal of cross-links (treated from RNase step onwards as described below), to obtain an estimate of the DNA concentration in each batch of chromatin. This was used to determine what volume of each chromatin to use per immunoprecipitation (IP), in order to approximate a previously optimized amount of DNA corresponding to ~20-50 imaginal discs per IP. In addition, this DNA preparation was measured again in parallel to the immunoprecipitated DNA by qPCR, in order to calculate the IP efficiency (see next section).

IPs using the independent wild-type and sxc- chromatin preparations were performed in parallel. The appropriate volumes of each chromatin (~100 uL) were adjusted to 1 mL RIPA buffer conditions (10 mM Tris-HCl pH 8 / 140 mM NaCl / 1 mM EDTA / 1% Triton X-100 / 0.1% SDS / 0.1% sodium deoxycholate / 1x “Complete” protease inhibitor cocktail / 1 mM AEBSF). Chromatin was pre-cleared by adding 40 uL of 50% (v/v) Protein A Sepharose CL-4B (GE Healthcare) beads suspension, previously equilibrated in RIPA buffer for 1 hour at 4°C on a rotating wheel. The precleared chromatin was removed after centrifugation at 4°C for 2 minutes at 3’000 g, and incubated with specific antibodies at 4°C overnight. Antibody-chromatin complexes were captured by incubation with 40 uL of 50% (v/v) Protein A Sepharose beads, previously blocked with 1 mg / mL BSA, at 4°C for 3 hours. Beads were washed for 10 min at 4°C with 1 ml of the following buffers : 1 wash with RIPA buffer, followed by 4 washes with RIPA buffer containing 500 mM NaCl, then 1 wash with LiCl buffer (10 mM Tris-HCl pH 8 / 250 mM LiCl / 1 mM EDTA / 0.5 % NP-40 / 0.5 % sodium deoxycholate), then 2 washes with TE (10 mM Tris-HCl pH 8 / 1 mM EDTA). Beads were resuspended in 100 uL TE and incubated with 10 ug / mL RNase at 37°C for 30 minutes, then adjusted to 0.5% SDS / 0.5 mg/mL proteinase K and incubated at 37°C for at least 8 hours with shaking. Cross-links were reversed by incubating 6 hours at 65°C.
DNA was purified by phenol / chloroform extraction using Phase-Lock heavy gel tubes (QIAGEN), followed by ethanol precipitation. The DNA pellet was resuspended in 500 uL of 5 mM Tris-HCl pH 8, and stored at -20°C.

WGA chromatin precipitations were performed in the same way, except that 50 uL of 50% (v/v) succinylated WGA-agarose resin (Vector Labs, cat no. AL-1023S), previously washed 3 times with RIPA buffer, was added to each chromatin at 4°C overnight. Washes were then performed as for the antibody immunoprecipitates.

**qPCR to determine binding at chromosomal regions**

ChIP eluate was analysed using a real-time PCR instrument (Applied Biosystems 7500) using SYBR Green and standard settings. PCR was performed in duplicates using 10 uL (1/50th) of the immunoprecipitated samples and corresponding untreated samples (prepared from the 30 uL aliquot of chromatin). A standard curve was generated by measuring dilutions of a purified DNA standard on the same qPCR plate. This allowed the calculation of the amount of target sequence in immunoprecipitated chromatin relative to the amount of target sequence in input chromatin, and was expressed as the percentage of input DNA recovered after precipitation.

The following qPCR primer pairs were used. Distances (in kilobases) of the middle nucleotides in the amplified regions are given relative to the gene’s first transcription start site: *Ubx* for primers 1-10, *Abd-B* for 11-13, *Scr* for 14-15, *en* for 16-17, *Dll* for 18-19, *tsh* for 20-21, *pn* for 22-23, *dpr12* (CG34385) for 24 and *CG11665* for 25.

<table>
<thead>
<tr>
<th>Name</th>
<th>Forward primer (5’ to 3’)</th>
<th>Reverse primer (5’ to 3’)</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (a)</td>
<td>GCAGCATAAAACCGAAGGA</td>
<td>CGCCAAACATTGACAGGATAG</td>
<td>-30.7 kb</td>
</tr>
<tr>
<td>2 (a)</td>
<td>TAGCTTATCTGTATCTCGCTCTTA</td>
<td>CAGAACCAAAGTGCGACTGAG</td>
<td>-29.6 kb</td>
</tr>
<tr>
<td>3 (a)</td>
<td>AAGGCGGAAGAGAGACACAA</td>
<td>CGTTTTAAGTGCAGACTGAG</td>
<td>-29.4 kb</td>
</tr>
<tr>
<td>4 (a)</td>
<td>GCACGCTAAACCCCATAA</td>
<td>TCCACCTCCTCCTCCTCCTC</td>
<td>-28.9 kb</td>
</tr>
<tr>
<td>5 (a)</td>
<td>ATCGGTAGCTTGTCAGCA</td>
<td>GGCTACTTGACAGGCTGAG</td>
<td>-2.1 kb</td>
</tr>
<tr>
<td>6 (a)</td>
<td>TCCAATCTGCTGGGGAACGAATT</td>
<td>TTAGCCGAGACGTGAGCTTGGAG</td>
<td>+0.2 kb</td>
</tr>
<tr>
<td>7 (a)</td>
<td>ATCGGTAGCTTGTCAGCA</td>
<td>ATCCAAAGGAGCAAAGGAAC</td>
<td>+1 kb</td>
</tr>
<tr>
<td>8 (a)</td>
<td>GCCGTGGGGAACGTCTCAGAT</td>
<td>TCGTGGTCGTCCTCCTAAT</td>
<td>+26.9 kb</td>
</tr>
<tr>
<td>9 (a)</td>
<td>CCATAAAGAAATGCCACTTTCAG</td>
<td>CTCTCAGCTCCTACTGAT</td>
<td>+31.7 kb</td>
</tr>
</tbody>
</table>
### Materials and Methods

| 10 (a) | GTCTGGCCAAGGCAAATATT | CGAAAGGAGAGCGAATGTT | +34.6kb |
| 11 (d) | CACTTTCGAGCAAAGGCG | AACAGCGATTCACAGCAGAC | +0.1 kb |
| 12 (d) | CTGCTGGTACATTTGCACGG | TCTGTGTCTCTAATGGCTGCG | +2.1 kb |
| 13 (a) | GGAATACCGCACTGTCGTAGG | GCAGCCATCATGGATGTGAA | +72.3kb |
| 14 (d) | GAAGTGCGCCACGTTGTCGTT | TCCCTCTCTCGACACTCGTT | +0.2 kb |
| 15 (d) | ATGTAGGGCGGTAGCGCACTG | CCTGAGTTCATCAATCAAGG | +18.3kb |
| 16 (b) | GTTCACTCCCTCTCGAGTAG | GAAACGCAGATTGAAACGTC | -0.3 kb |
| 17 (b) | GCGGTGTGTAATTTTGTGGA | CGCCTTAAGGTGAGATTCAGTT | +3.1 kb |
| 18 (c) | CCTGCCACAAAGCGACATT | CCCTGCTAGAGCAGAAACT | -1 kb |
| 19 (b) | ATCCGAGGACACTTCAGCC | GGCGGTGTCAATATTTTGGT | +3.1 kb |
| 20 (b) | GCGTGATGTACCACCTTTTGGT | AAGCGCTTCTCTGAATGCTT | +5.9 kb |
| 21 (b) | AAGGATTTCATTCCTGACACC | AGGTCCCAAACGCACTACT | +19.4kb |
| 22 (b) | GAGCAGGGGTGTTGACAA | TCTTCTCTAGGGACTGCT | +3.9 kb |
| 23 (b) | AGTAGGCGTTGATGGACAGCAT | GTCTGTGCTGAATACGAGAG | +5.4 kb |
| 24 (a) | CCGAAGATGAGAGATGGAAAA | AAATGCGGGACATGGTAAG | -3.2 kb |
| 25 (a) | CAGGGATGAGGGATGTTTGG | TGCCCTGTGGTCTATCAACAC | +12.5kb |

(a) Papp and Müller, 2006  
(b) Oktaba et al, 2008  
(c) Nekrasov et al, 2007  
(d) Gambetta et al, 2009

### ChIP-on-chip experiments

This work was performed by Katarzyna Oktaba.

For the GlcNAc ChIP-on-chip, two independent chromatin preparations from wing imaginal discs from wild-type larvae were either subjected to immunoprecipitations with an anti-O-GlcNAc antibody (HGAC85) or affinity purifications with agarose beads covalently linked to succinylated WGA.

For Ph ChIP-on-chip, two independent chromatin preparations from mixed wing, haltere and 3rd leg discs from wild-type larvae were immunoprecipitated with two affinity purified antibodies raised against different portions (Ph86-520 and Ph766-984) of the *Drosophila Ph-p* protein.

Pho and dSfmbt ChIP-on-chip is described in Oktaba et al, 2008. Two independent chromatin preparations from mixed wing, haltere and 3rd leg discs from wild-type larvae were immunoprecipitated with two antibodies raised against distinct epitopes of the Pho protein (Pho324-520 and Pho2-382). In
parallel, three independent wing, haltere and 3rd leg disc chromatins were immunoprecipitated with an affinity purified antibody raised against \textit{dSfmbt}531-980.

(For the detailed procedure, please see Gambetta et al., 2009.) The precipitated material was amplified by linker ligation-mediated PCR, fragmented (ideally ~25-100 bp fragments), labeled and hybridized to high-density whole-genome tiling arrays. Affymetrix GeneChip \textit{Drosophila} Tiling 1.0R Arrays were used, which consist of 25-mer probes designed in euchromatic and some heterochromatic regions, with an average 10 bp spacing. Analysis of the hybridization intensities by TileMap software generated a list of scored candidate bound regions. Threshold scores, above which we consider that regions are bound with high-confidence, were determined as follows:

- In the case of Ph, Pho and dSfmbt, data from each of the independent IPs were analysed together. The cut-offs for Pho-bound and dSfmbt-bound regions was set so as to recover at least 70% of published PREs in each case. Of these regions, there were 338 that were bound simultaneously by Pho and dSfmbt, and thus constitute the high-confidence PhoRC binding sites.

- The cut-off for Ph-bound regions was set so that 70% of the previously identified 338 PhoRC-bound regions were recovered. This led to the identification of 1'681 regions bound by Ph in larval imaginal discs.

- Only genomic regions significantly enriched by both the anti-O-GlcNAc antibody and the WGA lectin were analysed. As the analysis of the GlcNAc sites was done without prior knowledge of positive regions, we analysed both the top 10%- and top 1%-scoring regions proposed by TileMap.

\textbf{Preparation of soluble embryonic nuclear extracts}

Embryonic nuclear extracts were prepared as reported in Klymenko et al, 2006. Steps were performed on ice or at 4°C. 0-12 hour old wild-type embryos were dechorionated, then taken up in NU1 buffer (15 mM HEPES pH 7.6 / 10 mM KCl / 5 mM MgCl\textsubscript{2} / 0.1 mM EDTA pH 7.9 / 0.5 mM EGTA pH 7.9 / 350 mM Sucrose / 2 mM DTT / 0.2 mM PMSF), at a ratio of 1 mL NU1 buffer per 1 gram of embryos. Embryos were homogenized with a glass dounce homogenizer. Lysate containing nuclei from 30 to 40 grams of embryos was filtered by gravity through a double layer or miracloth mounted on a funnel. After the lysate was mostly drained through the miracloth, the cloth was rinsed with 2-3 volumes of buffer NU1. Nuclei were pelleted by centrifugation in a precooled Superlite GSA rotor at 9'000 rpm for 15 minutes. After wiping off the lipid (upper) layer and discarding the supernatant
(intermediate liquid portion), the nuclei pellet was resuspended in 0.5 mL of low-salt buffer (15 mM HEPES / 20 mM KCl / 1.5 mM MgCl₂ / 0.2 mM EDTA pH 7.9 / 20% glycerol / 1 mM DTT / “Complete” protease inhibitor cocktail) per gram of embryos. Care was taken so that the yolk was not resuspended with the nuclei. Resuspended nuclei were transferred into a 50 mL falcon tube and lysed by addition of 0.5 mL of high-salt buffer (15 mM HEPES / 800 mM KCl / 1.5 mM MgCl₂ / 0.2 mM EDTA pH 7.9 / 20% glycerol / 1 mM DTT) per gram of embryos, for 20 minutes with mixing. After lysis was complete, soluble nuclear material was separated from insoluble chromatin and lipids by centrifugation in a pre-cooled SW40 rotor at 38’000 rpm for 1 hour in ultraclear tubes. After centrifugation, a thin lipid layer was removed by suction (upper layer), and the soluble nuclear extract was carefully taken using a glass pipet (intermediate liquid portion). Care was taken not to contaminate the sample with the lower debris portion. Soluble nuclear extract was then dialysed against dialysis buffer (15 mM HEPES / 200 mM KCl / 1.5 mM MgCl₂ / 0.2 mM EDTA pH 7.9 / 20% glycerol / 1 mM DTT) using a dialysis membrane with a molecular weight cut-off of 6-8’000 kDa. After dialysis, soluble nuclear extract was snap-frozen in liquid nitrogen, and stored at -80°C.

WGA-agarose pull-downs

For GlcNAc affinity purification from wild-type embryonic soluble nuclear extracts, 2 mg of extract were incubated in a final volume of 1 mL IP buffer (15 mM Hepes pH 7.9 / 200 mM KCl / 1.5 mM MgCl₂ / 0.2 mM EDTA pH 8 / 0.25% NP-40 / 20% glycerol / 0.3 mM DTT / 1x “Complete” protease inhibitor cocktail / 1 mM PMSF) with 100 µl of a 50% slurry of washed succinylated WGA-agarose resin (Vector Labs) for 12 hours at 4°C. Succinylated WGA has high affinity for GlcNAc, and a reduced affinity for sialic acids. Beads were washed with IP buffer containing 0.5 mM DTT and 0.4% NP-40, followed by a 1 hour incubation with 1 M GlcNAc (Galab) on ice to elute resin-bound proteins. Boiling of the WGA-agarose beads in SDS sample buffer after GlcNAc elution revealed that a considerable amount of proteins remained on the beads after GlcNAc elution. Note: It was not possible to denature the embryonic protein extracts with SDS because of the presence of potassium chloride, which causes SDS::protein precipitation.

For GlcNAc affinity purification from larvae, inverted larval carcasses of wild-type (w¹¹¹B in the experiment presented in figure 7, or OregonR flies in the experiment presented in figure S4) or sxc²/sxc⁷ mutant larvae were cleared of fat body and digestive tract to leave only brain and imaginal discs
Materials and Methods

Attached. Extracts were processed in parallel in the following steps. Carcasses were sonicated in IP buffer (20 mM Hepes pH 7.9 / 200 mM NaCl / 1.5 mM MgCl₂ / 1 mM EDTA pH 8 / 0.5 mM EGTA / 20% glycerol / 1x “Complete” protease inhibitor cocktail / 1 mM AEBSF) and soluble extract was isolated by centrifugation at 20,800 g for 10 minutes at 4°C to remove debris. Half of this native extract was directly used for purification, while the other half was first denatured by adjusting to 0.5% SDS and 5 mM DTT and boiling for 3 min. In the experiment in figure 7, denatured extracts were diluted 5x with IP buffer to a final concentration of 0.1% SDS and 1 mM DTT, and adjusted to ~0.4% NP-40; whereas native extracts were diluted 5x and adjusted to a final concentration of (0% SDS), 0.2 mM DTT and ~0.4% NP-40. In the experiment in figure S4, native and denatured extracts were diluted with IP buffer to a final common concentration of 0.1% SDS, 1 mM DTT, and ~0.4% NP-40. Diluted extracts (final volumes of ~1.5 mL) were incubated with 100 µl of a pre-washed 50% WGA-agarose slurry for 2 hours at 4°C. (Longer incubations resulted in a decrease in enrichment, probably due to digestion of GlcNAc residues by lysosomal glycosidases released upon sonication.) After washes with IP buffer containing 0.4% NP-40, beads were boiled in 100 µL of SDS sample buffer to elute resin-bound material. Approximately 0.5-1 larva was loaded for input (0.5% or 1% of total starting material), and 10 larvae were loaded for the elutions (10% of total eluted material).

Western blots were performed using the wet transfer method (see “Anti-Ogt Western blot on larval extracts”). 5% BSA was used for blocking the membranes in the experiments presented in figures 6 and 7; whereas 5% milk was used in experiment S4.
### List of antibodies used

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Immunogen</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abd-B</td>
<td></td>
<td>gift from S. Celniker; described in Celniker et al, 1990</td>
</tr>
<tr>
<td>Asx</td>
<td>aa 210-336</td>
<td>Scheuermann et al, 2010</td>
</tr>
<tr>
<td>Calypso</td>
<td>aa 1-471</td>
<td>Scheuermann et al, 2010</td>
</tr>
<tr>
<td>CTD110.6</td>
<td>synthetic YSPTS(O-GlcNAc)PSK glycopeptide</td>
<td>Covance (MMS-248R)</td>
</tr>
<tr>
<td>Dll</td>
<td></td>
<td>gift from S. Cohen; described in Vachon et al, 1992</td>
</tr>
<tr>
<td>dSfmbt(MBT)</td>
<td>aa 531-980 (dSfmbt MBT domain)</td>
<td>Klymenko et al, 2006</td>
</tr>
<tr>
<td>En</td>
<td></td>
<td>Developmental Studies Hybridoma Bank (clone 4D9)</td>
</tr>
<tr>
<td>Esc</td>
<td></td>
<td>gift from V. Pirrotta; unpublished</td>
</tr>
<tr>
<td>E(z)3TAF</td>
<td>full-length E(z)</td>
<td>M. Nekrasov; described in Gambetta et al, 2009</td>
</tr>
<tr>
<td>H3</td>
<td>synthetic peptide from C-terminus of Human Histone H3</td>
<td>Abcam (ab 1791)</td>
</tr>
<tr>
<td>H-300</td>
<td>aa 1-300 of human Ogt iso 3 (Swissprot entry O15294)</td>
<td>santa cruz biotechnology (sc-32921)</td>
</tr>
<tr>
<td>HGAC85</td>
<td>streptococcal group A carbohydrate (GAC)</td>
<td>Abcam (Ab2735)</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>synthetic peptide containing the sequence [ARme3KSAP]</td>
<td>Upstate (UBI#07-449)</td>
</tr>
<tr>
<td>mAb414</td>
<td>nuclear pore complex mixture (believed to recognize FXFG repeats of certain nucleoporins)</td>
<td>Covance (MMS-120R)</td>
</tr>
<tr>
<td>Nurf55</td>
<td>full-length Nurf55</td>
<td>gift from A. Nowak; described in Gambetta et al, 2009</td>
</tr>
<tr>
<td>PCCD2</td>
<td>aa 15-78 (Pc chromodomain)</td>
<td>Papp and Müller, 2006</td>
</tr>
<tr>
<td>Pcl(Tu)</td>
<td>Pol tudor domain</td>
<td>T. Klymenko; unpublished</td>
</tr>
<tr>
<td>Ph2</td>
<td>aa 766-984 of Ph-p</td>
<td>K. Oktaba; described in Oktaba et al, 2008</td>
</tr>
<tr>
<td>Ph5</td>
<td>aa 86-520 of Ph-p</td>
<td>K. Oktaba; described in Gambetta et al, 2009</td>
</tr>
<tr>
<td>Pho(ZF)</td>
<td>aa 324-520 (Pho zinc finger domain)</td>
<td>Papp and Müller, 2006</td>
</tr>
<tr>
<td>Psc7a</td>
<td>aa 122-519</td>
<td>K. Oktaba; unpublished</td>
</tr>
<tr>
<td>Ring</td>
<td>aa 1-274</td>
<td>gift from M. Vidal; described in Gorfinkiel et al, 2004</td>
</tr>
<tr>
<td>Scm(MBT)3BWJ</td>
<td>aa 174-435 (Scm MBT domain)</td>
<td>A. Gaytan and N. Ly-Hartig; described in Gambetta et al, 2009</td>
</tr>
<tr>
<td>Scr</td>
<td></td>
<td>gift from D. Brower; described in Glicksman and Brower, 1988</td>
</tr>
<tr>
<td>Spt5</td>
<td></td>
<td>gift from J. Lis; described in Andrulis et al, 2000</td>
</tr>
<tr>
<td>Su(z)12ΔS</td>
<td>aa 1-448</td>
<td>A. Sengupta; unpublished</td>
</tr>
<tr>
<td>Trx C-ter</td>
<td>aa 2’388-2’674</td>
<td>gift from R. Paro; described in Beisel et al, 2007</td>
</tr>
<tr>
<td>Tsh</td>
<td></td>
<td>gift from S. Cohen; described in Ng et al, 1996</td>
</tr>
<tr>
<td>α-tubulin</td>
<td></td>
<td>Sigma (T9026)</td>
</tr>
<tr>
<td>Ubx</td>
<td></td>
<td>gift from R. White; described in White and Wilcox, 1984</td>
</tr>
<tr>
<td>8WG16</td>
<td>RNA polymerase II purified from wheat germ extract</td>
<td>Covance (MMS-126R)</td>
</tr>
</tbody>
</table>

**Table 2**: List of antibodies used for immunostaining, Western blot or ChIP experiments.
References


Capdevila, M.O. and García-Bellido, A. (1981), 'Genes involved in the activation of the Bithorax Complex of Drosophila.', Wilhelm Roux’s Arch. 179, 125-143.


Margueron, R.; Justin, N.; Ohno, K.; Sharpe, M. L.; Son, J.; Drury, W. J.; Voigt, P.; Martin, S. R.; Taylor,


Appendix

Supplementary figures

Figure S1: Mapping of sxc by complementation tests with deficiencies in the proximal region of chromosome 2R

Genes located in the 2R heterochromatin-euchromatin junction are listed on the left, with RpL38 being the most proximal (i.e. closest to the centromere) and 2R(09851 being the most distal (i.e. closest to the telomere). "(...)" indicates that there are non-explicated genes in between. 5 different deficiencies were fine-mapped by complementation tests with the lethal mutations available in the region, highlighted in purple (the names of the mutant alleles are specified). Each deficiency chromosome is schematized with a grey circle indicating the centromere, present regions colored in blue, and the deleted interval marked by dashed lines. The interval separating a lethal mutation which complements a deficiency and the next lethal mutation which no longer complements the deficiency, is represented by an empty rectangle with question marks, because the breakpoint of the deficiency occurs at an unknown position within. The mapped deficiencies were then tested for their ability to complement sxc alleles: *+* / "-" indicates complementation / no complementation. We were thus able to delimit the interval of sxc candidate genes to [Nipped-A; CG42345].
Figure S2: Effects of sxc/Ogt mutations

(A) Alignment of the protein sequences of Drosophila Ogt (dmOgt) and human Ogt (hsOgt) isoform 3, using the Blosum 62 scoring matrix. Base differences observed in dmOgt are colored in green (conservative amino acid substitutions) or red (non-conservative amino acid substitutions). Both proteins contain 13 N-terminal TPRs, each composed of an A (yellow) and B (pink or orange) helix, and a bipartite catalytic domain at the C-terminus (CDI: catalytic domain I, in blue; CDII: catalytic domain II, in purple). TPR-like domains (TLRs), shown in green, fold in a similar way to TPRs although they do not share the TPR consensus sequence.
Much of the variation between human and Drosophila Ogt proteins occurs in the N-terminus and loop regions. The commercial H-300 antibody used in this work was raised against aa 1-300 of human Ogt iso 3 (shown in bold). Structural domains are colour-coded as in (B) and (C). Non-sense mutations found in sxc2, sxc4 and sxc7 alleles are boxed in red; the missense mutation in sxc4 is boxed in orange.

(B) Cartoon view of Ogt in complex with a UDP-GlcNAc analogue (Clarke et al, 2008). This structure is a composite of the crystal structure of bacterial (Xanthomonas campestris) Ogt (Clarke et al, 2008), with the TPRs elongated to 13 in total by superimposing the human Ogt TPR crystal structure (Jinek et al, 2004). The active site is positioned between the two halves of the catalytic domain. sxc4 and sxc5 alleles produce high levels of aberrant OgtN948I and OgtΔ1031-1059 proteins, respectively (figure 2). XcOgt N464, which is equivalent to dmOgt N948, resides within the catalytic domain and is pointed to. The C-terminal helix of XcOgt, analogous to the one missing in dmOgtΔ1031-1059, is marked by a red bracket, and is necessary for the cohesion of both halves of the catalytic domain.

(C) Zoom into the position of N464 of XcOgt (equivalent to dmOgt N948). (Left panel) N464 (in pink) is located on the same loop as two other residues (a tyrosine and a histidine, in yellow) that are predicted to contact the substrate serine/threonine (not shown) or UDP-GlcNAc (in green), respectively (Clarke et al, 2008). (Right panel) N464 projects in the opposite direction and contacts other residues of CDII (in purple), with which it forms hydrogen bonds (dashed yellow lines). These interactions might serve to correctly position the Tyr463 and His466 components of the catalytic domain. Substitution of N464 by an isoleucine (as in dmOgtN948I encoded by sxc4) disrupts these ionic interactions, and thereby probably disrupts the integrity of the catalytic domain. The PDB file 2JLB (Clarke et al, 2008) was downloaded from www.pdb.org and processed with MacPyMOL software, with the help of Morgane Agez.
Figure S3: Independent anti-Ogt Western blot complementary to figure 2

Extracts from ~3 wild-type or sxc/ogt mutant larvae of the indicated genotypes were loaded in each lane, but extracts were not controlled for equal loading. Blots were probed with the H-300 anti-Ogt antibody as in figure 2. The position of full-length Ogt, produced from the wild-type, sxc² and sxc⁴ alleles, is indicated by a blue asterisk. The predicted molecular weights of the truncated proteins encoded by sxc⁷, sxc¹ and sxc⁵ are 62 kDa, 100 kDa and 115 kDa, respectively. The expected truncated Sxc/Ogt product is produced from sxc⁵ (marked by an orange asterisk). Trace amounts of severely truncated proteins are produced from the sxc⁷ (marked by green asterisks) and sxc¹ alleles (marked by a red asterisk), suggesting that these proteins are further degraded by intracellular proteases. Several cross-reacting bands are apparent, owing to the fact that the membrane was blocked in 5% BSA (not milk, as in figure 2) and to the fact that the exposure is relatively long. However these conditions are necessary in order to observe the truncated proteins produced from sxc² and sxc⁷.
Figure S4: Independent repetition of the WGA-agarose pull-down experiment from larval whole-cell extracts shown in figure 7

Wild-type native and denatured (denat.) extracts, as well as sxc\(^{-}\)/sxc\(^{-}\) native extracts were subjected to WGA-agarose affinity purification. 1% of the total starting material is loaded in "I" (input) lanes, and 10% of total eluted material is loaded in "E" (eluted) lanes. Note that the increase in Ph protein levels observed in sxc\(^{-}\)/sxc\(^{-}\) mutant extracts is not as prominent as seen in the experiment presented in figure 7; and, in general, loadings of each extract do not seem as equal. Purified material was tested for several PcG proteins, as well as for the C-terminal portion of Trx, transcription elongation factor.
Spt5, and the large subunit of RNA polymerase II (Rpb1). It is important to note that the Esc antibody used does not recognize recombinant Esc, and therefore we are unsure whether it specifically recognizes endogenous Esc. Ph was the sole protein to be enriched after purification from wild-type extracts, and was not purified from sxc1/sxc7 extracts. Ph enrichment is unfortunately much lower than in the experiment shown in figure 7, and might be due to a greater digestion of GlcNAc residues by lysosomal glycosidases released upon cell disruption in this experiment. Therefore, the possibility exists that certain of the tested proteins might in fact be GlcNAcylated by Sxc/Ogt in vivo, but might not be recovered under these experimental conditions. Note that, in contrast to what was seen in the experiment presented in figure 7, no traces of other proteins are detected in the purification from native wild-type extracts. This might be explained by the fact that the native extracts in this experiment contained 1 mM DTT (versus 0.2 mM DTT in the previous experiment) and 0.1% SDS (versus 0% SDS in the previous experiment) (see Materials and Methods for details), which might already disrupt protein-protein interactions.
GlcNAc-eluted material from the WGA-agarose pull-down experiment from wild-type embryonic nuclear extracts, presented in figure 6, was also probed with two independent antibodies against O-GlcNAc (HGAC85 and CTD110.6) and with the mAb414 antibody which recognizes the FxFG repeats present in several nucleoporins. I = 1% of input; E = 10% of purified material. Positions of marker proteins and their sizes are indicated in blue. Both anti-O-GlcNAc antibodies recognize a common set of bands in the purified material, suggesting that WGA-agarose pull-down effectively leads to the enrichment of several O-GlcNAc modified proteins present in embryo nuclear extracts. Importantly, a band corresponding to the size of Ph (which runs slightly above the 175 kDa marker) is not visible, raising the possibility that O-GlcNAcylated Ph might represent a minority of total O-GlcNAcylated proteins present in embryo nuclei. The mAb414 antibody recognizes a series of bands in the input material, of which 6 are recovered in the eluted material (the identity of these bands has not been investigated). Certain of the bands recognized by anti-O-GlcNAc run at the same size as the WGA-enriched nucleoporin bands (green asterisks), and might thus represent highly abundant O-GlcNAc modified nucleoporins (the lowest band might represent Drosophila Nup62, which has a predicted molecular weight of 41 kDa). Note that both anti-O-GlcNAc antibodies do not seem to recognize O-GlcNAcylated proteins in larval whole cell extracts, since an identical banding pattern is observed with both antibodies in wild-type and sxc^{+/sxc^{-}} extracts (data not shown).
Figure S6: sxc-/ogt- wing discs do not derepress pnr

The pnr-N21 transgene, which expresses β-galactosidase in the same pattern as Pnr because under the control of the pnr promoter, was introduced into an sxc1/sxc7 genetic background in order to test for misregulation of pnr expression in sxc/ogt mutants (in the absence of an anti-Pnr antibody). Wing imaginal discs from transgene-containing wild-type or sxc1/sxc7 third-instar larvae were immunostained with an antibody against β-galactosidase (red). Merged images with Hoechst, to visualize DNA, are also shown. As expected, the reporter transgene is expressed in the hinge region of wild-type wing discs. Observation of several mutant discs suggests that pnr is not misexpressed in sxc1/sxc7 mutants. pnr has nevertheless been shown to be misexpressed in wing discs of other PcG mutants such as ph and Psc-Su(z)2 mutants (Oktaba et al, 2008).

Figure S7: Derepression of Scr in sxc/ogt wing discs

Wing imaginal discs from wild-type or sxc1/sxc7 third-instar larvae were immunostained with an antibody against the Scr protein (red). Merged images with Hoechst, to visualize DNA, are also shown. Scr is normally not expressed in the wing disc (note that the images are overexposed and therefore appear red), but is misexpressed at low levels in specific regions of the pouch, hinge and notum of sxc1/sxc7 mutant wing discs.
Figure S8: Nuclear localization of Ph appears normal in sxc- /ogt- cells
Salivary glands from wild-type (wt) or sxc- /sxc7 (sxc-) 3rd instar larvae were immunostained with anti-Ph and costained with Hoechst to visualize DNA. Ph is nuclear in both wt and sxc1/sxc7 mutant cells. Note that by immuno-fluorescence in these salivary glands, I was unable to detect the increase in Ph protein levels that I observed by Western blotting protein extracts from sxc/sxc7 imaginal discs and brains (figure 8).

Figure S9: Conservation between human, cow and fly CTDs
Alignment of the C-terminal portions of the large subunit of RNA polymerase II from humans (HsRpb1, starting at amino acid 1'541), Bos taurus (BtRpb1) and Drosophila (DmRpb1), using the Blosum 62 scoring matrix. The heptapeptide repeats (52 in human and cow, and 32 in Drosophila) are colored. Conservatory amino acid substitutions are in green; non-conservative substitutions are in red. This shows that some heptapeptide repeats are conserved in the Drosophila CTD.
Figure S10: Alignment of Drosophila Ph and mammalian PHC3 orthologs

The long isoforms of Drosophila Ph-proximal (Ph-p) and Ph-distal (Ph-d) proteins are aligned with human and mouse PHC3 orthologs, using the Blosum 62 scoring matrix. The N-terminal portions can only poorly be aligned. Homologous regions are found in the C-terminus of the proteins, and include homology domain 1 (HD1) (pink), an FCS-type zinc finger (green) and a SAM domain (blue). Threonines found to be GlcNAcylated in recent proteome-wide mass spectrometry studies in human and mouse are in red and underlined. Regions of Drosophila Ph against which Ph5 (aa 86-520) and Ph2 (aa 766-984) antibodies were raised are in bold. Given the high similarity of Ph-p and Ph-d sequences in these regions, it is probable that these antibodies recognize both paralogs.
Abbreviations

aa = amino acid
Abd-B = Abdominal-B
AEBSF = 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
Antp = Antennapedia
Ash = Absent, small and homeotic discs
CDS = coding sequence
ChIP = chromatin immunoprecipitation
CTD = C-terminal domain of RNA polymerase II
DamID = DNA adenine methyltransferase identification
Df = deficiency
Dfd = Deformed
dl = Distal-less
dRAF = dRING associated factors
dSfmbt = Scm-related containing four mbt domains
DTT = Dithiothreitol
En = Engrailed
EMS = ethyl methanesulfonate
ESCs = embryonic stem cells
E(z) = Enhancer of zeste
FISH = fluorescence in situ hybridization
GlcNAc = N-acetylglucosamine
HD = homology domain
HMTrase = histone methyltransferase
IP = immunoprecipitation
kb = kilobases
kDa = kilo Daltons
Kis = Kismet
lab = labial
MBT = malignant brain tumor
me1 = monomethylation
me2 = dimethylation
me3 = trimethylation
MW = molecular weight
NURF = nucleosome remodeling factor
Oga = O-GlcNAc hydrolase
O-GlcNAc = O-linked N-acetylglucosamine
Ogt = O-GlcNAc transferase
ORF = open reading frame
pb = proboscipedia
PBS = phosphate buffered saline
Pc = Polycomb
PcG = Polycomb group
Pcl = Polycomb-like
Ph = Polyhomeotic
PHC3 = Polyhomeotic-like 3
PHD = plant homeodomain
Ph-d / Ph-p = Polyhomeotic-distal / proximal
Pho = Pleiohomeotic
PhoRC = Pho repressive complex
Pnr = Pannier
PMSF = phenylmethanesulfonylfluoride
PRC1 = Polycomb repressor complex 1
PRC2 = Polycomb repressor complex 2
PR-DUB = Polycomb repressive deubiquitinase
PRE = Polycomb Response Element
Psc = Posterior sex combs
qPCR = quantitative PCR
RYBP = Ring and YY1 Binding Protein
SAM = sterile alpha motif
SANT = Swi3, Ada2, N-CoR, TFIIIB
Sc = Sex combs extra
Scm = Sex combs on midleg
Scr = Sex combs reduced
SDS = sodium dodecylsulfate
SET = Suvar3-9, Enhacer-of-zeste, Trithorax
Spt5 = Suppressor of Ty 5
Su(z)2 = Suppressor of zeste 2
Su(z)12 = Suppressor of zeste 12
Sxc = Super sex combs
TLR = TPR-like
TPR = tetratricopeptide repeat
trxG = Trithorax group
Tsh = Teashirt
TSS = transcription start site
Ubx = Ultrabithorax
UCH = ubiquitin carboxy-terminal hydrolase
UDP = uridine diphosphate
UTX = Ubiquitously transcribed tetratricopeptide repeat, X chromosome
VEFS = VRN2-EMF2-FIS2-Su(z)12
WGA = wheat germ agglutinin
wt = wild-type
Zn = zinc
Essential Role of the Glycosyltransferase Sxc/Ogt in Polycrom Repression

Maria Cristina Gambetta, Katarzyna Oktava, Jürg Müller*

Polycrom group proteins are conserved transcriptional repressors that control animal and plant development. Here, we found that the Drosophila Polycrom group gene super sex combs (sxc) encodes Ogt, the highly conserved glycosyltransferase that catalyzes the addition of N-acetylglucosamine (GlcNAc) to proteins in animals and plants. Genome-wide profiling in Drosophila revealed that GlcNAc-modified proteins are highly enriched at Polycrom response elements. Among different Polycrom group proteins, Polyhomoecte is glycosylated by Sxc/Ogt in vivo. sxc/Ogt-null mutants lack DlK/Gnaicylation and failed to maintain Polycrom transcriptional repression even though Polycrom group protein complexes were bound at their target sites. Polycrom repression appears to be a critical function of Sxc/Ogt in Drosophila and may be mediated by the glycosylation of Polyhomoecte.

The Drosophila gene super sex combs (sxc)

was originally identified because mutations in this gene caused lethality at the pupal stage and hemeosomatic transformations of multiple body segments into segments normally present in more posterior body regions (1). This phenotype stage and hemeosomatic transformations of multiple body segments into segments normally present in more posterior body regions (1). This phenotype suggested the involvement of hemeosomatic (Hox) genes were expressed outside of their normal expression domains, and the Hox gene Ultrabithorax (Ubx) was indeed misexpressed in sxc mutant larvae (2,3). To explore the role of sxc in Page repression, we stained larvae that were transheterozygous for two different sxc-null mutations with antibodies against the protein products of the PcG target genes Ubx, Abd-B, and Abd-B (Sail-B), Sex comb reduced (Dcr), engrailed (en), Distal-less (Dll), and teashirt (tsh). Each of these genes was expressed outside of its normal expression domain (Fig. 1A). sxc is thus needed for repression of multiple PcG target genes in Drosophila larvae.
mutant larvae with an antibody to Ogt. No Ogt protein was detected in sxc/sxc, sxc/sxc, or sxc/sxc mutants, suggesting that these are protein-null mutations (Fig. 1B). OgtN948I, the mutant Ogt protein encoded by sxc, and OgtN948I proteins, containing lesions predicted to affect the fold of the glycosyltransferase domain (6, 7), therefore seem to be completely nonfunctional. An intact glycosyltransferase domain in Ogt is thus critical for Polycrome repression.

More than 100 different proteins have been reported to be the O-GlcNac modification in mammalian cells, including RNA polymerase II and several transcription factors and coregulators (4, 5). Labeling of Drosophila polytene chromosomes with wheat germ agglutinin (WGA), a lectin that binds with high affinity to GlcNAc residues (8, 9), suggested that GlcNAcylated proteins are also present on Drosophila chromosomes (10). We determined the genome-wide distribution of GlcNAc-modified proteins in the chromatin of imaginal disc cells from Drosophila larvae by performing chromatin immunoprecipitation (ChIP) assays with an antibody that recognizes the O-GlcNac modification (11) and with a WGA-affinity resin (12). We hybridized the precipitated material to high-density whole-genome tiling arrays and analyzed them using TileMap (12) to retrieve putative chromatin domains with high levels of GlcNAc modification.

The work reported here shows that GlcNAcylation by Sxc/Ogt plays an essential role in PcG repression of PcG target genes in flies, and worms, and the only known glycosyltransferase that adds GlcNAc moieties to nuclear and cytosolic proteins (4, 5). Ogt is essential in

The similarity between the GlcNAc and PcG protein ChIP profiles prompted us to test whether any of the PcG proteins themselves may carry the O-GlcNac modification. We used WGA-agarose to affinity-purify GlcNAc-modified proteins from wild-type or sxc mutant larval extracts and probed the purified material with antibodies against the PhoRC subunits Pho (S4) and E(z) (S5) and with antibodies against the PRC1 subunits Ph, Ring, and Sce; and the PRC2 subunits Su(z)12 and Nurf55. Ph, but none of the other proteins tested, was strongly enriched after WGA-affinity purification from wild-type but not sxc mutant larvae (Fig. 4). This enrichment of Ph was also observed under denaturing conditions (Fig. 4). This suggests that Ph itself is GlcNAcylated by Sxc/Ogt. PRC1 components Ring, Pce, or Sce did not copurify with GlcNAc-modified Ph in these WGA pull-down assays from larval extracts (Fig. 4), although Ring and Pce were readily detected together with Ph in such purifications from embryonic nuclear extracts (fig. S4). This finding could be explained by differential association of Ph with PRC1 components and/or different accessibility of GlcNAc moieties on Ph during embryonic and larval development.

WGA-purifications failed to provide evidence for GlcNAcylation of the large subunit of RNA poly

Fig. 2. O-GlcNAc–modified proteins are localized at PREs in Drosophila. (A) Venn diagrams showing the overlap of 1139 10% top-ranked O-GlcNAc sites (left) or 114 1% top-ranked O-GlcNAc sites (right) with 1689 Ph and 338 PhoRC-bound regions in larval imaginal discs (tables S1 and S2) (3). (B) ChIP profiles of O-GlcNAc, Ph, and PhoRC subunits Pho and dSfmbt at the Antennapedia complex in imaginal disc cells. Hybridization intensities for oligonucleotide probes are plotted as color bars above the genomic map (release 5, kilobase coordinates) of Drosophila melanogaster; significantly enriched regions are marked below plots (3). The HOX genes labial (lab), proboscipedia (pb), Deformed (Dfd), Scr, and Antennapedia (Antp) on the plus (above) and minus (below) strand are represented with exons (black boxes) and introns (thin black lines).

Fig. 3. O-GlcNAc modification and PcG protein binding in wild-type and sxc mutant chromatin. (A) ChIP analysis monitoring O-GlcNAc modification in wing (magenta) and haltere/third leg (pink) imaginal discs from wt and sxc7/sxc1 mutant (sxc–) larvae. Graphs show results from independent immunoprecipitation reactions with antibody to O-GlcNAc. ChIP signals, quantified by means of quantitative polymerase chain reaction, are presented as mean percentage of input chromatin precipitated at each region; error bars indicate ±SD (3). Locations of PREs (purple boxes) and other regions relative to transcription start sites are indicated in kilobases; euchromatic (eu.) and heterochromatic (het.) control regions were mapped outside of these genes. In wt larvae, GlcNAc ChIP signals at the −30 kb Ubx PRE were comparable in wing and haltere/third leg chromatin, but at the +32 kb PRE the signal in haltere/third-leg chromatin was two- to threefold lower than in wing chromatin, paralleling PRC1 and PRC2 binding (14). (B) ChIP signals in WT (blue bars) and sxc– wing discs (orange bars) representing Pho, E(z), and Ph binding.

www.sciencemag.org SCIENCE VOL 325 3 JULY 2009
mice, in which it is required for the viability of embryonic stem cells (17, 18), but dispensable for the normal development of Caenorhabditis elegans (19). In contrast, Drosophila mutants lacking Sxx/Ogt and O-GlcNAcylation display a specific chromatin phenotype. Polycistronic repression, sxx mutants show no obvious developmental defects, suggesting that O-GlcNAcylation is the main process that critically depends on O-GlcNAcylation in Drosophila. We provide evidence that Ph is GlcNAcylated. Although it remains to be determined whether Ph is indeed the relevant Sxx/Ogt substrate in PhO-GlcNAcylation repression, it is tempting to speculate that the function of Sxx/Ogt in gene silencing may be GlcNAcylated Ph. The sxx null phenotype is not as severe as that of other Ph mutants, notably that of ph (18, 19) and this study). Thus, if GlcNAcylation of Ph contributes to its function, Ph still retains partial repression activity in the absence of this modification. One possibility would be that GlcNAcylation of Ph is needed for efficient anchoring of Ph to PREs and/or for the capacity of PRE tethered Ph to maintain a repressed chromatin state at target genes.

Drosophila PhO-GlcNAc proteins are conserved in vertebrates, and the PhO-GlcNAcylation system represents a large set of orthologous developmental regulator genes both during Drosophila development and in mammalian embryonic stem cells (13, 22). Thus, GlcNAcylation of Polycomb homologs, or perhaps other PhO-GlcNAc proteins, may also be an evolutionarily ancient and essential function of O-GlcNAcylation in vertebrates.

References and Notes
3. Materials and methods are available as supporting material on Science Online.

Ligand-Gated Chloride Channels Are Receptors for Biogenic Amines in C. elegans
Niels Ringstad,* Namiko Abe,‡ H. Robert Horvitz‡

Biogenic amines such as serotonin and dopamine are intercellular signaling molecules that function widely as neurotransmitters and neuromodulators. We have identified in the nematode Caenorhabditis elegans three ligand-gated chloride channels that are receptors for biogenic amines: LGC-53 is a high-affinity dopamine receptor, LGC-55 is a high-affinity tyramine receptor, and LGC-40 is a low-affinity serotonin receptor that is also gated by choline and acetylcholine. Lgc-55 mutants are defective in a behavior that requires endogenous tyramine, which indicates that this ionotropic tyramine receptor functions in tyramine signaling in vivo. Our studies suggest that direct activation of membrane chloride conductances is a general mechanism of action for biogenic amines in the modulation of C. elegans behavior.

Biogenic amines function in diverse neuronal circuits as neurotransmitters and neuromodulators. For many psychiatric disorders, including major depression, schizophrenia, and bipolar affective disorder, target signaling pathways of such biogenic amines as serotonin, dopamine, and noradrenaline (1). Biogenic amine signaling pathways are also targets of drugs of abuse (2). Almost all known biogenic amine receptors are G protein-coupled receptors (GPCRs) that signal though the activation of heterotrimeric guanine nucleotide-binding proteins (G proteins), which activate second-messenger signaling pathways. However, there exists a second type of biogenic amine receptor.

Previously published online as Science Express on 28 May 2009.

1. Address correspondence to J. Müller; E-mail: jmueller@ibmol.bf.bf.de
2. To whom correspondence should be addressed. E-mail: horvitz@mit.edu
3. Supported by grants from the Deutsche Forschungsgemeinschaft. The microarray data have been deposited in the ArrayExpress database under the accession number E-TABM-769.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1169727/SOM
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References
22 December 2009; accepted 19 May 2009
Published online 28 May 2009; 10.1126/science.1169727
Include this information when citing this paper.

Biogenic amines function in diverse neuronal circuits as neurotransmitters and neuromodulators. For many psychiatric disorders, including major depression, schizophrenia, and bipolar affective disorder, target signaling pathways of such biogenic amines as serotonin, dopamine, and noradrenaline (1). Biogenic amine signaling pathways are also targets of drugs of abuse (2). Almost all known biogenic amine receptors are G protein-coupled receptors (GPCRs) that signal through the activation of heterotrimeric guanine nucleotide-binding proteins (G proteins), which activate second-messenger signaling pathways. However, there exists a second type of biogenic amine receptor:
Remerciements (français)

J'exprime ma gratitude sincère aux personnes suivantes :

À Jürg, pour m'avoir offert une position au sein de son laboratoire, ainsi que pour son suivi et soutien tout au long de mon doctorat.
À Daniel, pour son dévouement au suivi du progrès de mon travail, ainsi que pour ses commentaires qui en ont contribué à l'avancement.
Aux autres membres du comité de thèse, Eileen Furlong et François Spitz, pour leur soutien lors des réunions annuelles. Je suis aussi reconnaissante à Eileen pour ses efforts pour participer à ma défense à Genève.
À Pierre Spierer et Milanka Stojkovic, pour leur aide dans l'établissement d'une collaboration entre l'Université de Genève et le Laboratoire Européen de Biologie Moléculaire (EMBL).
À François Karch et Ueli Schibler, pour leur participation à ma défense.
Spécialement à Denis Basquin, pour son amitié, des discussions stimulantes et pour son aide précieuse à plusieurs étapes tout au long de ma thèse. Aussi, pour son dessin des complexes de protéines PcG de la figure S4.
À Bernadett Papp, pour ses efforts initiaux pour cloner sxc et ses conseils aux débuts du travail.
À Kasia Oktaba, pour avoir généré les profiles de ChIP-on-chip de GlcNAc et de Ph, ainsi que pour m'avoir initié aux techniques de la préparation de chromatine, de ChIP et de qPCR.
À Phil Ingham, pour son son input au commencement du projet.
À Kate Osborne, pour des discussions stimulantes et pour avoir confirmé les mutations sxc de manière indépendante, ainsi que pour ses efforts de contribution au début du projet.
À Barry Honda et Hugh Brock, qui ont publié l'identification de sxc en parallèle, pour des discussions utiles.
À Andrés Gaytán, pour m'avoir fourni la majeure partie de la chromatine larvaire wild-type que j'ai utilisée dans ce travail.
À Raquel Matos, pour son aide dans la préparation des extraits nucléaires d'embryons.
À Maxime Nekrasov, Agnieszka Nowak et Nga Ly-Hartig, pour avoir généré les anticorps contre E(z), Nurf55 et Scm, respectivement.
À Morgane Agez, pour son amitié et son aide dans l'utilisation du programme informatique MacPyMOL.
À Jos de Graaf et Vladimir Benes du Service de Génomique de l'EMBL, pour l'hybridation du matériel obtenu par immunoprécipitation de chromatine à des microarrays, ainsi que pour le soutien pour qPCR et séquençage d'ADN.

À tous ceux de l'EMBL, pour leur générosité, bonne humeur et enthousiasme qui m'ont fait sentir privilégiée d'en avoir pris part.

L'allocation de recherche a été financée par la Deutsche Forschungsgemeinschaft (DFG).

Je dédie ma thèse à ma famille et à mes amis.
Acknowledgements (English)

I express my sincere gratitude to the following people:

To Jürg, for having offered me a position in his laboratory, as well as for his mentoring and generous support during my PhD.

To Daniel, for his commitment to following the progress of my work and for his valuable feedback.

To my TAC members Eileen Furlong and François Spitz, for their support during my yearly TAC meetings. I am also grateful to Eileen for her efforts to participate in my thesis defense in Geneva.

To Pierre Spierer and Milanka Stojkovic, for their help in setting up the collaboration between the University of Geneva and the European Molecular Biology Laboratory (EMBL).

To François Karch and Ueli Schibler, for participating in my thesis defense.

Especially to Denis Basquin, for his friendship, helpful discussions and precious help at several steps throughout my PhD. Also, for the drawing of the PcG protein complexes in figure S4.

To Bernadett Papp, for initial efforts to map sxc and tutoring in the early steps of my work.

To Kasia Oktaba, for generating the GlcNAc and Ph ChIP-on-chip profiles, as well as for initiating me to the chromatin preparation, ChIP and qPCR techniques.

To Phil Ingham, for input during the early stages of the project.

To Kate Osborne, for helpful discussions and for having confirmed the sxc molecular lesions independently, as well as for her efforts to make contributions in the early stage of the project.

To Barry Honda and Hugh Brock, who published their results on sxc in parallel to ours, for helpful discussions.

To Andrés Gaytán, for having provided me with most of the wild-type larval chromatin that I used in this study.

To Raquel Matos, for help in preparing embryonic nuclear extracts.

To Maxime Nekrasov, Agnieszka Nowak and Nga Ly-Hartig, for raising E(z), Nurf55 and Scm antibodies, respectively.

To Morgane Agez, for her friendship and for help with utilization of MacPyMOL software.

To Jos deGraaf and Vladimir Benes from the EMBL Genomics Core Facility, for hybridization of chromatin immunoprecipitated material to arrays, as well as for qPCR and DNA sequencing support.
To everyone at EMBL, for their generosity, good humour and enthusiasm which made me feel privileged to be a part of it.

My PhD fellowship was funded by the Deutsche Forschungsgemeinschaft (DFG).

I dedicate my PhD thesis to my family and friends.