

RESEARCH ARTICLE

Post-translational modifications of *Drosophila melanogaster* HOX protein, Sex combs reduced

Anirban Banerjee¹*, Anthony Percival-Smith

Department of Biology, The University of Western Ontario, London, Ontario, Canada

* abanerj7@uwo.ca

OPEN ACCESS

Citation: Banerjee A, Percival-Smith A (2020) Post-translational modifications of *Drosophila melanogaster* HOX protein, Sex combs reduced. PLoS ONE 15(1): e0227642. <https://doi.org/10.1371/journal.pone.0227642>

Editor: Jinsong Zhang, Saint Louis University School of Medicine, UNITED STATES

Received: September 17, 2019

Accepted: December 23, 2019

Published: January 13, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0227642>

Copyright: © 2020 Banerjee, Percival-Smith. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by the Natural Sciences and Engineering Research Council of

Abstract

Homeotic selector (HOX) transcription factors (TFs) regulate gene expression that determines the identity of *Drosophila* segments along the anterior-posterior (A-P) axis. The current challenge with HOX proteins is understanding how they achieve their functional specificity while sharing a highly conserved homeodomain (HD) that recognize the same DNA binding sites. One mechanism proposed to regulate HOX activity is differential post-translational modification (PTM). As a first step in investigating this hypothesis, the sites of PTM on a Sex combs reduced protein fused to a triple tag (SCR^{TT}) extracted from developing embryos were identified by Tandem Mass Spectrometry (MS/MS). The PTMs identified include phosphorylation at S185, S201, T315, S316, T317 and T324, acetylation at K218, S223, S227, K309, K434 and K439, formylation at K218, K309, K325, K341, K369, K434 and K439, methylation at S19, S166, K168 and T364, carboxylation at D108, K298, W307, K309, E323, K325 and K369, and hydroxylation at P22, Y87, P107, D108, D111, P269, P306, R310, N321, K325, Y334, R366, P392 and Y398. Of the 44 modifications, 18 map to functionally important regions of SCR. Besides a highly conserved DNA-binding HD, HOX proteins also have functionally important, evolutionarily conserved small motifs, which may be Short Linear Motifs (SLiMs). SLiMs are proposed to be preferential sites of phosphorylation. Although 6 of 7 phosphosites map to regions of predicted SLiMs, we find no support for the hypothesis that the individual S, T and Y residues of predicted SLiMs are phosphorylated more frequently than S, T and Y residues outside of predicted SLiMs.

Introduction

The identity of body segments along the Anterior-Posterior (A-P) axis of Bilaterians is determined by a set of developmental control genes called *Homeotic selector* (*Hox*) genes [1,2]. These genes encode transcription factors (TFs) that regulate expression of target genes by binding to DNA-binding sites with a 60 amino acid DNA-binding homeodomain (HD) [3]. That HOX proteins determine distinct segmental identities and regulate distinct patterns of gene expression while recognizing similar DNA binding sites is a major paradox. Interactions with the cofactor Extradenticle (EXD) is one mechanism proposed to mediate the functional

Canada (NSERC) discovery grant No. RGPIN-2015-06356 to AP-S (URL: http://www.nserc-crsng.gc.ca/index_eng.asp). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

specificity of HOX proteins. In addition to HOX cofactor protein interactions, PTMs are also proposed to have a role in the regulation of functional specificity [4–6]. Here the PTMs of a HOX protein, Sex combs reduced (SCR) are mapped as a first step towards understanding this regulation of HOX protein activity.

SCR is a well-studied HOX protein which is essential for the formation of larval salivary glands, adult proboscis and adult prothoracic legs in *Drosophila melanogaster* [7–11]. Besides the highly conserved HD, which is essential for function, there are other small conserved motifs in SCR, the N-terminal octapeptide motif (MSSYQFVNS), LASCY motif, DYTQL motif, NEAGS motif, YPWM motif, NANGE motif, KMAS motif and the C-terminal domain (CTD), that are also important for function (Fig 1A) [6,11–16]. Amino acid substitutions in these small conserved motifs affect SCR activity both for proboscis and sex comb determination and for the suppression of ectopic proboscis formation [11,14,15]. The non-uniform effect of *Hox* mutant alleles across tissues, termed differential pleiotropy, suggest that the conserved

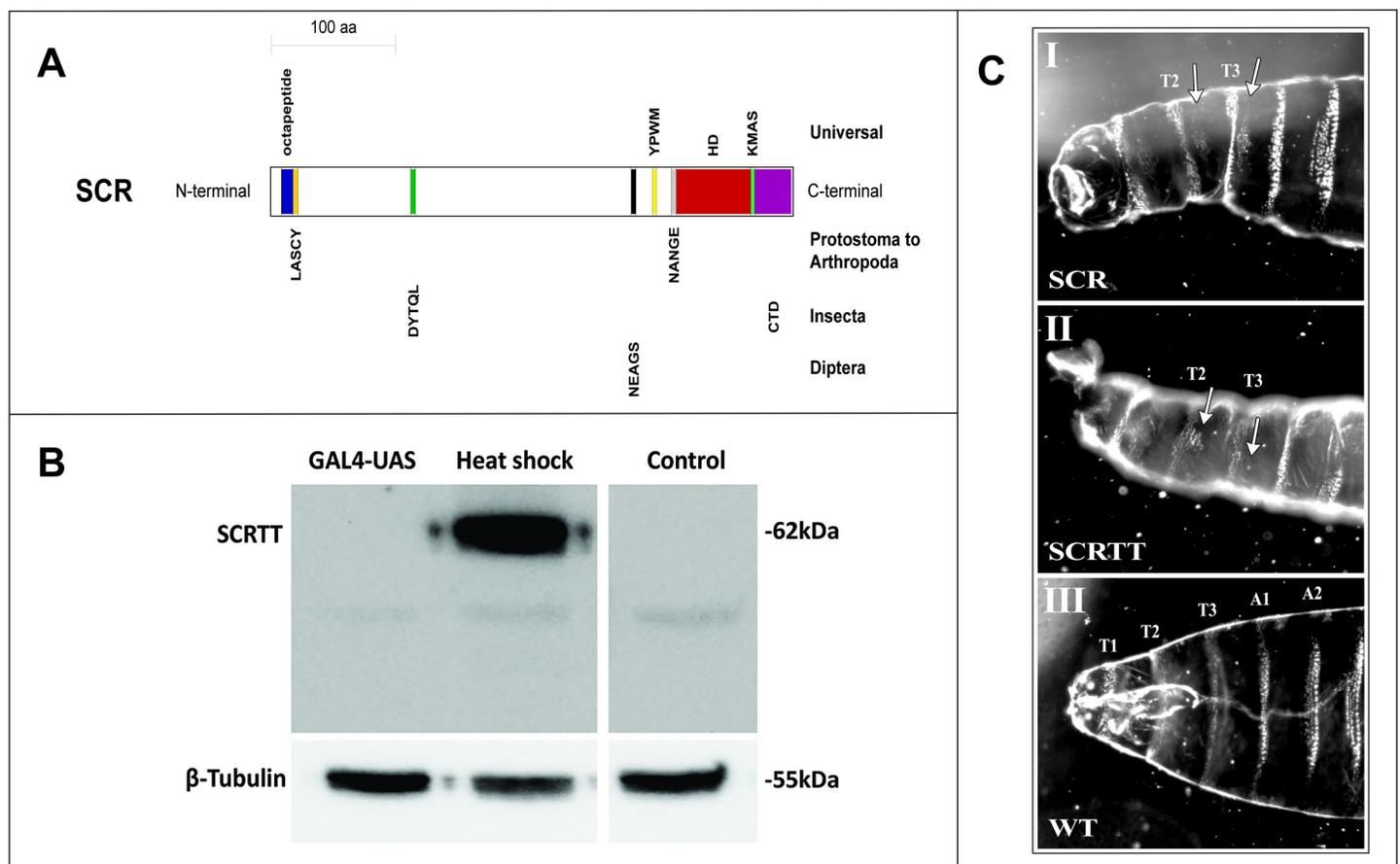


Fig 1. Sex combs reduced functional domains/motifs, expression and phenotype. (A) Schematic of SCR protein showing highly conserved, functional domains/motifs. The block diagram is drawn to scale and the domains/motifs are color-coded. The octapeptide motif is labeled in blue, LASCY motif in orange, DYTQL motif in dark green, NEAGS motif in black, YPWM motif in yellow, NANGE motif in grey, HD in dark red, KMAS motif in light green and CTD in purple. The taxonomic level of conservation of the domains/motifs is indicated on the right. (B) Comparison of the expression of SCR protein from heat-shock and UAS promoters. The method of expression is indicated above the lanes of the Western Blot of protein extracted from embryos. The position of SCR protein is indicated on the left, and the positions of molecular weight markers are indicated on the right. The level of β-tubulin expression is shown at the bottom. *y w* is the untransformed control. (C) First instar larval cuticular phenotypes of embryonic ectopic expression of SCR and SCR protein. The anterior half of the larva is shown. The untagged SCR protein expressed with the GAL4-UAS system using a ubiquitous *armadillo*-GAL4 is shown in panel (I); whereas, the SCR protein expressed from a heat-shock promoter is shown in panel (II). T1, T2 and T3 refer to first, second and third thoracic segments. A1 and A2 refer to first and second abdominal segments. (I) and (II) Ectopic expression of SCR and SCR protein, respectively (T2 and T3 beards marked with arrows). (III) Control wild-type (WT) first instar larval cuticle.

<https://doi.org/10.1371/journal.pone.0227642.g001>

domains and motifs contribute to HOX activity in a small, additive and tissue-specific manner [6,11,14,15,17–19]. The short, conserved motifs found in SCR and other HOX proteins may also be Short Linear Motifs (SLiMs) [6]. SLiMs or Eukaryotic Linear Motifs (ELMs) are short stretches of protein sequence (typically 3–10 amino acids long containing 2–3 specificity-determining residues) present in regions of intrinsic disorder [20–23]. SLiMs function as sites of protein-protein interaction, PTMs and cell compartment targeting signals [20–28]. Experimentally validated SLiMs have been curated into the ELM database and the database is used to predict SLiMs [28]. Although SLiMs have a low information content and are predicted to occur frequently in the proteome, the 22% of human disease mutations that occur in intrinsically disordered regions of human proteins map frequently to functionally important residues of validated SLiMs and also map frequently to predicted SLiMs suggesting that predicted SLiM sequences are correlated with function [29]. Viral proteins have SLiMs that mimic the SLiMs in host proteins which enables the virus to evade host immune functions [30,31]. Phosphorylation of the PDZ-binding motif (PBM), which is a SLiM, in the disordered C-terminal tail of ribosomal S6 kinase 1 (RSK1) protein is important for epidermal growth factor regulation [32]. The annotation of SLiMs in the ELM database suggest that a significant proportion are targets of phosphorylation.

Phosphorylation is a major regulatory mechanism used in cellular signaling pathways that often terminate in the regulation of TF activity and subsequent regulation of gene expression [33–35]. HOX are phosphoproteins [36–47]. PP2A-B' is proposed to activate SCR by dephosphorylating residues in the N-terminal arm of the SCR HD that when phosphorylated inhibit interaction of the SCR HD with DNA. Indeed, peptides with N-terminal arm sequence are phosphorylated by cAMP-dependent protein kinase A (PKA) and dephosphorylated by Serine-threonine protein phosphatase 2A (PP2A-B') *in vitro* [36]. However, a null PP2A-B' allele does not affect SCR activity suggesting that dephosphorylation by PP2A-B' plays no role in regulating SCR activity [48]. A drawback of the methodologies employed to study phosphorylation of SCR and other *Drosophila* HOX and HD-containing proteins so far is they provide no direct and definitive information on which amino acid residue is phosphorylated in the developing *Drosophila* embryo [36–47]. We have used Tandem Mass Spectrometry (MS/MS) on SCR extracted from developing embryos to provide data about a possible repertoire of PTMs [49].

Materials and methods

Drosophila husbandry

The fly stocks were maintained at 23°C and 60% humidity. The flies were grown in 20ml vials and ~300ml bottles containing corn meal food (1% (w/v) *Drosophila*-grade agar, 6% (w/v) sucrose, 10% (w/v) fine-ground cornmeal, 1.5% (w/v) yeast and 0.375% (w/v) 2-methyl hydroxybenzoate—an anti-fungal agent). To collect embryos, flies were allowed to lay eggs/embryos on apple juice plates (2.5% (w/v) *Drosophila*-grade agar, 6% (w/v) sucrose, 50% apple juice and 0.3% (w/v) 2-methyl hydroxybenzoate) smeared with a yeast paste (with 0.375% (w/v) 2-methyl hydroxybenzoate).

Generation of *ScrTT* fusion constructs

A *NotI* DNA fragment encoding SCR fused to the triple tag (TT) (containing 3X FLAG, Strep II and 6X His tags) [50] was isolated from a pUAST construct [11] and inserted into pET-3a and pCaSpeR [51–53] using standard molecular cloning techniques. The pCaSpeR *ScrTT* construct was used to transform the *Drosophila melanogaster* strain, $y^1 w^{67c23.2}$ via *P*-element-mediated transformation [54].

Ectopic expression of SCR and SCRTT and preparation of first instar cuticles

For ectopic expression of SCR, virgin female flies of the genotype, $y w; P\{UAS-Scr, w^+\}$ (Bloomington stock # 7302) were crossed with the GAL4 driver males of the genotype, $y w; P\{Armadillo-Gal4, w^+\}$ (Bloomington stock # 1560) and progeny were collected [55]. For ectopic expression of SCRTT protein from the heat-shock promoter *ScrTT* fusion gene, a heat-shock was administered at 5 hours AEL at 36.5°C for 30 minutes. First instar cuticles were prepared as described [56] and imaged with darkfield optics on a Leica[®] Leitz™ DMRBE microscope.

Ectopic expression of SCRTT from the heat-shock and UAS promoters

To ectopically express SCRTT using the GAL4-UAS system [55], adult virgin female flies of the genotype, $y w; P\{UAS-ScrTT, w^+\}$ were crossed with the GAL4 driver males of the genotype, $y w; P\{Armadillo-Gal4, w^+\}$ (Bloomington stock # 1560) and the progeny expressed SCRTT ubiquitously. To induce expression of SCRTT from the heat-shock promoter, *D. melanogaster* embryos at 0–16 hours AEL were collected from the apple juice plates on nylon mesh screens with an opening size of 0.3mm stretched across the bottom of a cylindrical basket 3cm in diameter and 5cm high. The embryos were heat-shocked for 30 minutes at 37.5°C by immersion of the filter basket in a circulating water bath.

Affinity purification of SCRTT protein from embryos

SCRTT protein was purified from embryo extracts using subcellular fractionation followed by metal affinity chromatography in denaturing conditions [57,58]. 3g of heat-shocked embryos was homogenized in 15ml of lysis buffer (15mM HEPES pH 7.6, 10mM KCl, 5mM MgCl₂, 2mM EDTA, 350mM sucrose, 0.032% 2-mercaptoethanol, with protease inhibitors: 0.2mM phenylmethanesulfonylfluoride (PMSF), 1.3mM benzamidine and 0.3mM Aprotinin) using a 40ml Dounce Homogenizer. The lysate was centrifuged in a Corex tube at 10,000 rpm for 15 minutes in a Sorval SS-34 rotor, and the supernatant was discarded. The white top layer of the pellet (pellet 1), leaving the dark colored debris behind, was carefully resuspended in a resuspension buffer 1 (15mM HEPES pH 7.6, 10mM KCl, 0.1mM EDTA, 350mM sucrose, 0.006% 2-mercaptoethanol, with protease inhibitors: 0.2mM phenylmethanesulfonylfluoride (PMSF), 1.3mM benzamidine and 0.3mM Aprotinin) and was centrifuged at 10,000 rpm for 15 minutes in a Sorval SS-34 rotor. The pellet (pellet 2) was resuspended in resuspension buffer 2 (15mM HEPES pH 7.6, 10mM KCl, 350mM sucrose, with protease inhibitors: 0.2mM phenylmethanesulfonylfluoride (PMSF), 1.3mM benzamidine and 0.3mM Aprotinin) and was centrifuged at 10,000 rpm for 15 minutes in a Sorval SS-34 rotor. The pellet (pellet 3) was resuspended in a nuclear lysis buffer (50mM NaH₂PO₄, pH 7.5, 300mM NaCl, 20mM imidazole, 1% NP-40, with protease inhibitors: 0.2mM phenylmethanesulfonylfluoride (PMSF), 1.3mM benzamidine and 0.3mM Aprotinin) and was centrifuged at 12,000 rpm for 10 minutes in a Sorval SS-34 rotor. All preceding steps were performed at 0–5°C. The nuclear extract (NE) was mixed with solid urea to a final concentration of 8M, and the mixture was gently rocked at room temperature until the urea dissolved. The denatured nuclear extract (NE+Urea) was mixed with 250μl of Ni-NTA sepharose beads (IBA Lifesciences) that had been equilibrated with the denaturing nuclear lysis buffer, and gently rocked for 15 minutes at room temperature. The beads were packed in a column by gravity flow and the flow-through was reapplied to the column. The beads in the column were washed twice with the denaturing nuclear lysis buffer and then, washed twice with a buffer containing 50mM NaH₂PO₄, 300mM NaCl pH 7.5. The beads were stored at -80°C.

Western blot analysis

The proteins in an SDS-polyacrylamide gel were transferred onto an Immobilon[®]-P PVDF transfer membrane (Millipore Sigma) by electroblotting at 250mA for two hours in ice cold transfer buffer (25mM Tris, 192mM glycine and 10% methanol). The blots were blocked at room temperature for one hour in Blotto (PBT: 10% PBS and 0.1% Tween-20, and 3% skim milk). Anti-FLAG M2 monoclonal antibody (Sigma-Aldrich) at a dilution of 60,000-fold in Blotto was incubated with the blot for one hour at room temperature. After washes with PBT, horseradish peroxidase (HRP)-conjugated goat anti-mouse antibody (ThermoFisher Sci.) at a dilution of 3,000-fold in Blotto was added and incubated for one hour at room temperature. After washes with PBT, HRP was detected using SuperSignal[™] West Femto Maximum Sensitivity Chemiluminescent Substrate (ThermoFisher Sci.). Digital images were recorded using a ChemiDoc[™] Imaging System (Bio-Rad). For some experiments, the membrane was stripped for one hour at room temperature using Restore[™] Western Blot Stripping Buffer (ThermoFisher Sci.) to remove the anti-FLAG antibody and was blocked at room temperature for one hour in Blotto followed by incubation with anti- β -tubulin monoclonal antibody (E7 concentrated from Developmental Studies Hybridoma Bank, University of Iowa, Iowa City, IA) at a dilution of 1,500-fold in PBT.

Sample preparation for MS/MS

The Ni-NTA bead slurry (25 μ l) from the protein purification was mixed with equal volume of 2xSDS buffer (100mM Tris-HCl pH 6.8, 200mM 1,4-dithiothreitol (DTT), 4% SDS, 20% glycerol, 1% 2-mercaptoethanol, ~1 mg/ml bromophenol blue) [59] and was heated to 90°C for 10 minutes. The 50 μ l sample was loaded onto a 1.5mm thick SDS-Polyacrylamide gel (11% separating and 5% stacking gel) for size separation of proteins. The gel was stained with 1 mg/ml Coomassie blue (Coomassie Brilliant Blue[™] R-250 from ThermoFisher Sci.). The destained gel was stored in 5% glacial acetic acid at 4°C. At the Functional Proteomics Facility, Western University, London, Ontario, Canada, 10 spots were picked from the gel using an Ettan[®] Spot-Picker[™] and were in-gel digested with either trypsin (Promega), chymotrypsin (Sigma-Aldrich) or thermolysin (Promega), and the peptides subsequently lyophilized. The peptides were analyzed with a Thermo Scientific Orbitrap Elite mass spectrometer, which uses the nano LC-ESI-Orbitrap-MS/MS technique, at the Biological Mass Spectrometry Laboratory, Western University, London, Ontario, Canada for protein identification and characterization of post-translational modifications.

Mass spectrometry data analysis

LC-ESI-Orbitrap-MS/MS data was analyzed at the Biological Mass Spectrometry Laboratory, Western University, London, Ontario, Canada. PEAKS[™] DB software versions 7, 7.5 or 8 (Bioinformatics Solutions Inc.) [60] were used to perform *de novo* sequencing and subsequent database search. PEAKS[™] PTM was used to identify post-translational modifications. PEAKS[™] DB uses a peptide score, which measures the quality of the peptide-spectrum match and separates the true and false identifications. Peptide score is given as $-10\log_{10}P$, where P refers to P-value. A high peptide score and a low P-value are associated with the confidence of the peptide match. The false discovery rate (FDR) was set at 1% which establishes a peptide cut-off score. A peptide must meet the cut-off score in order to be identified by PEAKS[™] DB. For our analysis, a modified peptide was associated with higher confidence if the peptide score was higher than the cut-off score by 8, which corresponds to a lower P-value. Minimal ion intensity, which is the relative intensities of position-determining fragment ions in a MS² spectrum was set to 5%. Coverage is given in the analysis as percent coverage. Average Depth (AD) is the

addition of the lengths of all chemically distinct peptides identified divided by the length of the protein. Since, the proteases used often generate fragments too large or too small for analysis, the Average Depth of regions covered (AD_{orc}) is calculated, which is Average Depth divided by proportion of the protein covered. To distinguish between the biologically relevant PTMs and artefactual modifications that might have arisen due to chemical handling, a manual investigation of the modifications was performed.

Phosphopeptide enrichment and C18 desalting of non-phosphopeptides

Phosphopeptide enrichment of trypsinized α -casein or SCRTT used the EasyPhos protocol employing TiO_2 beads [61]. The TiO_2 flow-through containing potential non-phosphopeptides were desalted using a C18-StageTip [62] prior to MS/MS analysis. The C18 StageTip was solvated thrice with 200 μ l of 80% acetonitrile, 0.2% formic acid and 19.8% water followed by centrifugation at 3,000g for 2 mins at room temperature or until no liquid remained in the tip. The C18 StageTip was equilibrated thrice with 200 μ l of 2% acetonitrile and 0.2% formic acid followed by centrifugation at 3,000g for 2 mins at room temperature. The TiO_2 flow-through was reduced in a SpeedVac to 100 μ l and adjusted to have a concentration of 0.2% formic acid. The sample was loaded onto the C18 StageTip and was centrifuged at 500g at room temperature until no liquid remained in the tip. The tip was then washed thrice with 200 μ l of aqueous buffer (2% acetonitrile, 0.2% formic acid and 97.8% water) followed by centrifugation at 500g at room temperature until no liquid remained in the tip. The peptides were eluted with 100 μ l of elution buffer (80% acetonitrile, 0.2% formic acid and 19.8% water) followed by centrifugation at 500g at room temperature until no liquid remained in the tip. The eluate was concentrated under vacuum using a SpeedVac to a volume of approximately 18 μ l and formic acid was added at a final concentration of 0.25–0.5%. The sample was analyzed by LC-MS/MS.

Purification of SCRTT from bacteria

SCRTT expression was induced in *E. coli* BL21(DE3) transformed with pBS *ScrTT* using standard methods [51] and purified using Ni-NTA chromatography under denaturing conditions [63,64]. Protein concentration was determined by the Bradford protein assay [65].

Bioinformatic analysis of proteomic data

D. melanogaster SCR protein sequence (NCBI accession number in S6 Table) was submitted to the ELM database [28] to retrieve predicted nuclear and cytoplasmic short linear motif (SLiM) sequences as HOX transcription factors interact with nuclear and cytoplasmic components [66,67]. SLiMs with any amino acids from known ordered regions of SCR were excluded from the analysis [16,68]. To determine whether a SLiM was conserved, SCR orthologous protein sequences of various protostome and deuterostome species belonging to different phyla, were retrieved from NCBI or ORCAE (only for *T. urticae*) [69] database (accession numbers in S6 Table) and a multiple sequence alignment was performed using the tools, MAFFT version 7 [70] and Clustal Omega [71]. Each predicted SLiM of SCR was manually checked for conservation across species (S10 Fig). A SLiM was considered to be conserved only if the entire predicted SLiM retrieved from the ELM database aligned perfectly in both MAFFT and Clustal Omega. If parts of a predicted SLiM and not the entire SLiM were conserved across species, that SLiM was not considered as conserved in our analysis. If a SLiM was less than four amino acids long and was conserved, it was not included in the list of conserved SLiMs (S5 Table and Fig 2). SLiMs less than five amino acids long were not considered as a conserved SLiM unless conserved beyond Diptera.

Statistical analysis

To determine the significance of the biased distribution of serine (S), threonine (T) and tyrosine (Y) in HOX proteins to SLiMs vs. non-SLiMs and the biased phosphate distribution in SCR SLiMs vs. non-SLiMs, Fisher's Exact Test was employed [72].

Results

Expression of SCRTT protein

To map PTMs of SCR expressed during embryogenesis requires an initial concentrated source of protein that can be affinity purified. The CDS (expressing SCR isoform A (417 aa); FlyBase ID FBpp0081163) of *Scr* mRNA was fused in frame to the triple tag (TT) encoding C-terminal 3X FLAG, Strep II and 6X His tags [11,50] and cloned downstream of the *UAS* promoter of pUAST [11], the heat-shock promoter (*hsp*) of pCaSpeR [53] and the T7 promoter of pET-3a [51,52]. Two major systems for ectopic expression of proteins in *Drosophila* are the heat-shock inducible promoter and the GAL4-*UAS* binary system, using the strong, ubiquitous *armadillo*-GAL4 driver. The expression of SCRTT using these two ectopic expression systems were compared (Fig 1B). The heat-shock promoter resulted in higher levels of accumulation of SCRTT (Fig 1B). The fold increase of SCRTT expression of heat-shock relative to *UAS* was too great to be accurately quantified. The relative molecular mass (M_r) of SCRTT protein is calculated to be 49.8 [73]. However, on a Western Blot, the SCRTT protein expressed during embryogenesis ran with a higher M_r of 62 (Fig 1B).

SCRTT protein is biologically active

To assess whether SCRTT expressed from the heat-shock promoter was biologically active, the first instar larval cuticular phenotype of heat-shocked embryos expressing SCRTT was compared with the cuticles that result from expression of untagged SCR protein in all cells of the embryo using the GAL4-*UAS* system [55]. Both SCR and SCRTT ectopic expression induced ectopic T1 beards in T2 and T3, indicating that the triple tag does not interfere with the biological activity of SCR *in vivo* (Fig 1C) [11,74,75].

Analytical workflow for affinity purification, digestion and mapping of PTMs in embryonically expressed SCRTT

For affinity purification of SCRTT protein, 3g of *D. melanogaster* embryos, collected at 0–16 hours AEL and heat-shocked for 30 minutes at 37.5°C, were lysed. The nuclei were collected and washed, and the proteins of the nuclear extract were denatured and SCRTT was affinity purified by Ni-NTA chromatography [63,64]. The purification of SCRTT was monitored by Western Blot analysis and shows concentration of SCRTT on the Ni-NTA beads (Fig 3B). An SDS gel stained for total protein identified a band of the correct M_r for SCRTT from protein extracted from the Ni-NTA beads (Fig 3D). To determine whether this purification provided the amount of SCRTT required for MS/MS, a Western Blot analysis was performed where a sample of the Ni beads containing purified embryonic SCRTT was run alongside a sample of 3500ng of SCRTT purified from bacteria, an amount which upon MS/MS analysis resulted in efficient detection of SCRTT peptides (unpublished data) (Fig 3E). The signal for the sample of the SCRTT purified from *Drosophila* embryos was 2.4-fold less than the bacterial signal suggesting the band contained about 1500ng of SCRTT. Knowing the concentration of SCRTT purified and the amount required for efficient detection of peptides, approximately 10µg of SCRTT purified from embryos was digested and processed for each sample analyzed by MS/MS.

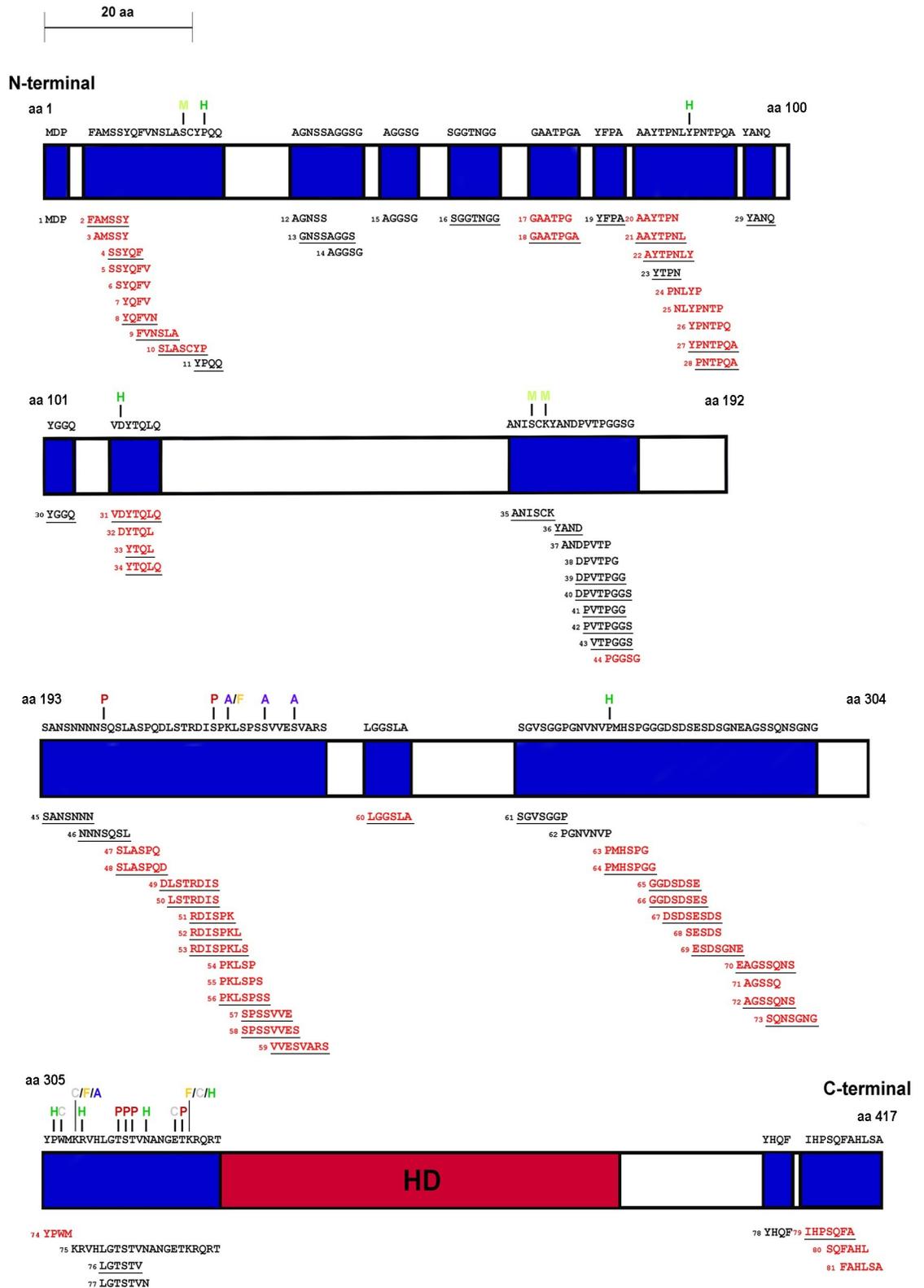


Fig 2. Schematic of SCR showing total and conserved SLiMs. The schematic is drawn to scale. The ordered region of the homeodomain is labeled by dark red and SLiMs are labeled blue. SLiMs that are conserved are red. The numbers beside SLiMs correspond to the SLiM data tables in the supplement (S4 and S5 Tables). The PTMs which map to SCR SLiMs are above the

modified residue and color-coded. Phosphorylation sites are indicated by P in dark red, acetylation by A in blue, formylation by F in orange, methylation by M in light green, carboxylation by C in grey and hydroxylation by H in green. SLiMs underlined are candidate phosphorylation sites.

<https://doi.org/10.1371/journal.pone.0227642.g002>

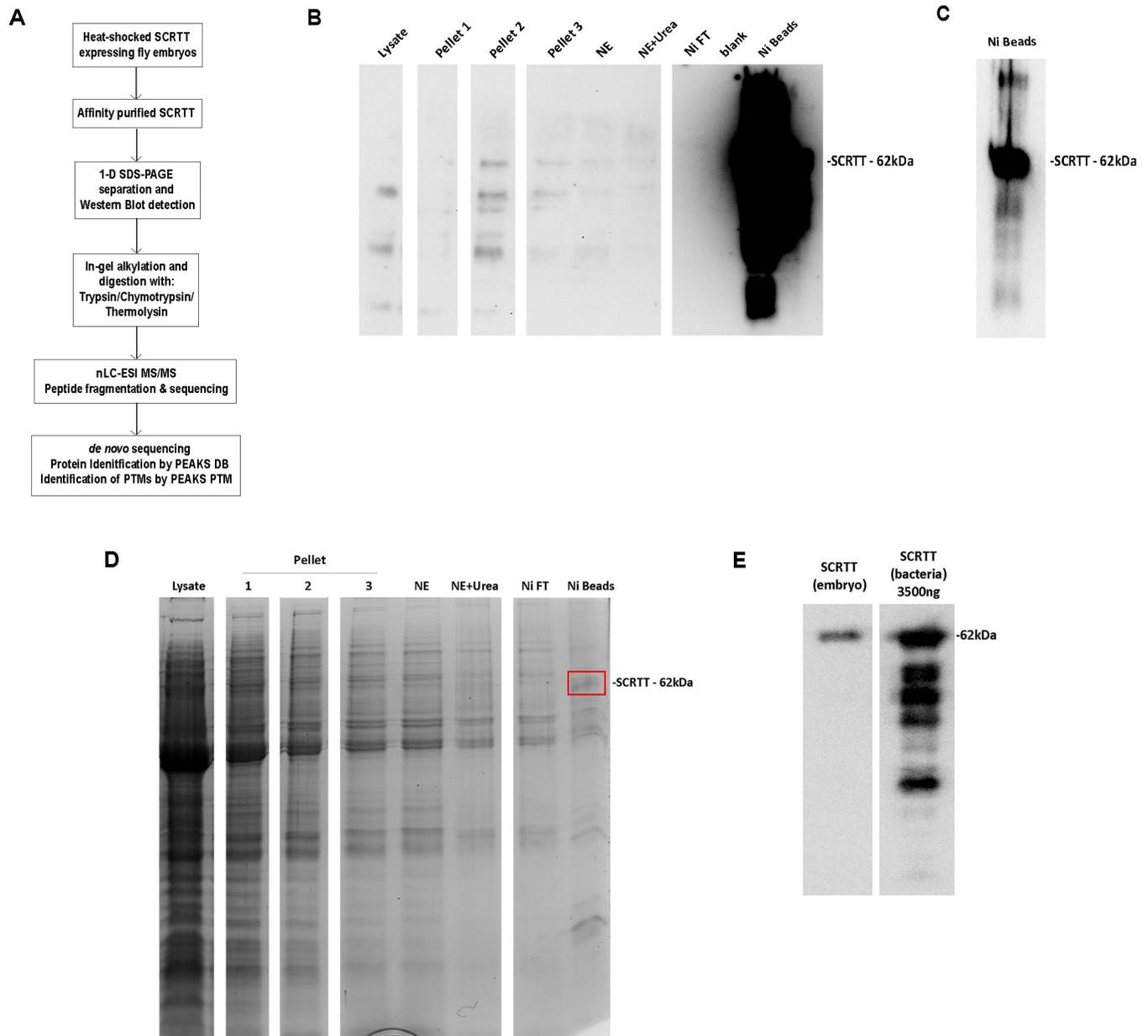


Fig 3. Overall approach for PTM mapping of SCR TT. (A) Analytical workflow for affinity purification, digestion and sequence mapping of PTMs in SCR TT expressed from developing embryos. (B) Overexposed Western Blot showing SCR TT at 62 kDa in the Ni beads fraction. Purification fraction is indicated at top of the lane. (C) Autoexposed Western Blot showing SCR TT at 62 kDa in the Ni beads fraction. (D) Coomassie-stained 1-D SDS-polyacrylamide gel of affinity purified SCR TT. Purification fraction is indicated at top of the lane. A band of SCR TT at 62 kDa (marked with a red box) was observed in the Ni beads fraction. (E) Comparison of *Drosophila* SCR TT vs. bacterial SCR TT to estimate the amount of protein to be analyzed by MS/MS. The Western Blot shows signals for SCR TT at 62 kDa indicated on the right. The source of SCR TT is at the top of the lane. NE Nuclear Extract, Ni Nickel and Ni FT Nickel Flow-Through. For each of the panels (B), (D) and (E), all lanes shown are from the same gel and the white space between lanes indicate the splicing out of irrelevant lanes.

<https://doi.org/10.1371/journal.pone.0227642.g003>

Eight samples were analyzed with MS/MS of which four were digested with trypsin, two with chymotrypsin and two with thermolysin. [S1 Fig](#) shows the distribution of chemically distinct peptides over the primary sequence of SCR_{TT} for each sample. Trypsin, chymotrypsin and thermolysin were chosen for digestion of SCR_{TT} based on the analysis of predicted peptide generated by these enzymes [73]. The coverage and average depth of region covered (AD_{orc}) for each sample analyzed, and the combined coverage and AD_{orc} of various combinations of samples were determined ([Table 1](#)). The final coverage was 96% of the primary sequence of SCR_{TT}.

Evidence for the PTMs of SCR_{TT}

Each spectrum obtained by LC-MS/MS was interrogated by PEAKS DB search followed by identification of PTMs using PEAKS PTM algorithm. Embryonic SCR_{TT} is post-translationally modified ([S1 Fig](#)). The modifications that were the result of a biochemical process and not a potential by-product of sample preparation were interrogated and characterized in more detail ([Table 2](#) and [S1 Fig](#)). These 44 modifications are phosphorylation, acetylation, formylation, methylation, carboxylation or hydroxylation ([S4–S9 Figs](#)). For all modifications interrogated, six diagnostic criteria were assessed: first that the peptide had a mass shift indicative of the modification; second whether in the MS^2 spectra there were b and/or y ions that supported modification of a specific amino acid residue; third whether other MS^2 spectra were identified for peptides with a particular modification; fourth whether overlapping or differently modified peptides were identified with the same modification; fifth whether a modification was identified in multiple independent samples used for MS/MS analysis ([Table 2](#)) and sixth whether the difference of the modified peptide score and the cut-off score was greater than 8. For phosphorylation, an additional diagnostic criterion was assessed: whether b and/or y ions with a neutral loss of phosphoric acid (98 Da) were present in the MS^2 spectra.

Phosphorylation of SCR_{TT} is substoichiometric

The percentage of phosphorylated peptides is low ([Table 2](#)) indicating that phosphorylation of SCR_{TT} may be substoichiometric. To determine whether the lack of phosphate detection was due to the instrument used for the analysis, a heavily phosphorylated protein, α -casein [76] was analyzed. 12 phosphosites were detected in α -casein ([S2 Table and S2 Fig](#)). Multiple

Table 1. MS/MS analysis—Coverage and AD_{orc} for SCR_{TT}.

Enzyme	Sample	% Coverage	AD_{orc}
Trypsin	1	47	4.81
	2	66	10.79
	3	62	10.39
	4	76	27.29
	1+2+3+4	79	31.1 ^a
Chymotrypsin	5	80	8.7
	6	76	8.3
	5+6	82	12.51 ^a
Thermolysin	7	67	7.22
	8	60	5.92
	7+8	71	9.48 ^a
Trypsin+Chymotrypsin+Thermolysin	-	96	43.29 ^a

^a Duplicate peptides were removed for calculation of AD_{orc} .

<https://doi.org/10.1371/journal.pone.0227642.t001>

Table 2. Summary of the evidence for PTMs in SCRTT.

Modification	Figure	Residue	b/y ions	Minimal ion intensity > 5%	Total no. of spectra ^a	No. of peptides modified	Percentage of peptides modified (Total no. of peptides)	No. of samples observed ^b	Peptide score —cut-off score ^c	b/y ions with a neutral loss of phosphoric acid (98 Da)
Phosphorylation	S3A	S185 ^d	b	b	1	1	0.66 (151)	1	0.44	b, y
	S3B	S201	y	y	1	1	0.58 (171)	1	3.46	b, y
	S3C	T315	b, y	b	1	1	0.49 (206)	1	3.9	b, y
	S3D	S316	b, y	b, y	2	1	0.49 (206)	1	3.78	b, y
	S3E	T317	b, y	b, y	3	2	0.97 (206)	1	6.72	b, y
	S3F	T324	y	y	9	4	2.31 (173)	2	9.43	b, y
Acetylation	S4A	K218	y	y	3	2	4.35 (46)	2	18.87	
	S4B	S223	b, y	y	2	1	0.92 (109)	1	11.54	
	S4C	S227	b, y	-	4	2	1.85 (108)	1	14.32	
	S4D	K309	b	-	1	1	0.93 (108)	1	3	
	S4E	K434	b, y	b, y	1	1	2 (50)	1	3.2	
	S4F	K439	b, y	b, y	2	1	1.67 (60)	1	3.15	
Formylation	S5A	K218	y	y	35	4	8.7 (46)	4	26.47	
	S5B	K309	b	-	4	4	3.7 (108)	3	12.52	
	S5C	K325	b, y	y	21	6	3.51 (171)	3	17.74	
	S5D	K341	b, y	b, y	2	2	2.99 (67)	2	7.21	
	S5E	K369	b, y	b, y	8	3	9.68 (31)	3	6.03	
	S5F	K434	b, y	b, y	2	1	2 (50)	1	6.07	
	S5G	K439	b, y	b, y	1	1	1.67 (60)	1	1.63	
Methylation	S6A	S19	b, y	b, y	6	2	9.09 (22)	2	8.23	
	S6B	S166	b, y	b, y	12	7	11.48 (61)	2	2.79	
	S6C	K168	b, y	b, y	3	2	3.33 (60)	2	2.39	
	S6D	T364	b, y	y	6	3	4.62 (65)	2	13.91	
Carboxylation	S7A	D108	b, y	b, y	6	2	1.12 (179)	1	15.71	
	S7B	K298	y	-	5	2	8.33 (24)	2	3.43	
	S7C	W307	b, y	y	16	6	4.69 (128)	3	8.09	
	S7D	K309	b	-	14	6	5.56 (108)	3	13.22	
	S7E	E323	b, y	b, y	20	9	4.89 (184)	5	26.93	
	S7F	K325	b, y	b	8	5	2.92 (171)	2	10.76	
	S7G	K369	b, y	b, y	1	1	3.23 (31)	1	2.37	
Hydroxylation	S8A	P22	b, y	b, y	1	1	4.76 (21)	1	2.62	
	S8B	Y87	b, y	b, y	1	1	2.33 (43)	1	3.29	
	S8C	P107	b, y	b, y	15	8	4.47 (179)	7	7.82	
	S8D	D108	b, y	b, y	20	11	6.15 (179)	7	10.82	
	S8E	D111	b, y	b, y	2	3	1.73 (173)	3	6.05	
	S8F	P269	y	y	6	5	12.5 (40)	3	16.26	
	S8G	P306	b, y	b, y	4	4	3.13 (128)	2	19.6	
	S8H	R310	y	-	1	1	2.94 (34)	1	7.64	
	S8I	N321	b	b	1	1	0.51 (196)	1	3.66	
	S8J	K325	y	-	1	1	0.58 (171)	1	8.22	
	S8K	Y334	b, y	-	2	1	1.69 (59)	1	7.06	
	S8L	R366	y	-	11	6	9.38 (64)	2	13.6	
	S8M	P392	b, y	b, y	1	1	9.09 (11)	1	2.29	
S8N	Y398	y	-	1	1	10 (10)	1	2.07		

^a Total number of MS² spectra reporting the modification as identified by PEAKS and subsequently filtered for those with evidence from the analysis of the 8 samples (Table 1)

^b Total number of samples the modification was observed in (Table 1)

^c The peptide score differences greater than 8 are in bold. A peptide was associated with higher confidence if the peptide score was higher than the cut-off score by 8, which corresponds to a lower P-value.

^d S185 phosphosite is not labelled in S1A Fig but was identified in the analysis of tryptic peptides (A2).

<https://doi.org/10.1371/journal.pone.0227642.t002>

phosphosites map between 61–70 of α -casein, and of the 64 peptides identified for this region, 61 were phosphorylated and 24 of these peptides were phosphorylated at two amino acid residues, which is a high percentage of modified peptides indicating stoichiometric levels of phosphorylation of α -casein (S3 Fig). This suggests that phosphorylation is stable during MS/MS analysis.

TiO₂ beads were used to enrich for phosphopeptides from trypsinized α -casein and SCRTT samples [61]. For MS/MS analysis of α -casein, 92.3% of all peptides (84 out of 91) identified post-TiO₂ treatment were phosphorylated whereas without TiO₂ treatment, only 18.8% of all peptides (158 of 839) were phosphorylated showing that TiO₂ enriches for phosphopeptides (S3 Table and S2 Fig). However, only 54 phosphopeptides were common for both treatments. 104 phosphopeptides previously identified without TiO₂ treatment could not be identified post-TiO₂ treatment which indicates loss of phosphopeptides (S3 Table). Although TiO₂ enriches for phosphopeptides, the yield is low and no phosphopeptides were detected upon TiO₂ enrichment of trypsinized SCRTT.

Are Short Linear Motifs (SLiMs) in SCR favored sites of phosphorylation?

Outside the HD, HOX proteins are highly disordered proteins [66]. SLiMs present in disordered protein regions are proposed to be preferential sites of phosphorylation [6,77,78]. To test this with SCRTT, the Eukaryotic Linear Motif (ELM) resource [28] was screened for predicted SLiMs. Of the 81 SLiMs identified, 52 (64%) are annotated as potential phosphosites. If 100% of the amino acid sequence of a predicted SLiM was conserved across species, the SLiM was considered ‘conserved’. Of the 53 conserved SLiMs identified, 34 (64%) are annotated as potential phosphosites. In SCR, 63% of the primary sequence was SLiM sequence and 35% were conserved SLiM sequence (Table 3 and Fig 2). 6 out of 7 phosphosites were in a SLiM region and 1 of 7 phosphosites was in a conserved SLiM (Fig 2). There is an enrichment of phosphosites to SLiMs. Although this may suggest preferential phosphorylation of SLiMs, we also analyzed whether the amino acid residues that accept phosphates are more frequently phosphorylated in predicted SLiMs than outside of SLiMs; an additional expectation for SLiMs being preferential sites of phosphorylation. There was no significant increase in the frequency of phosphorylation of residues in SLiM versus non-SLiM regions, and a significant decrease in the frequency of phosphorylation of residues in conserved SLiMs (Table 3). In addition, 9 out of 81 SLiMs and 5 out of 53 conserved SLiMs of SCR were phosphorylated (Fig

Table 3. SLiM analysis of SCR.

Percentage of SCR protein that are SLiMs						
Protein (disordered) size (aa)	No. of SLiMs	SLiM size (aa)	% SLiM	No. of conserved SLiMs	Conserved SLiM size (aa)	% Conserved SLiM
364	81	228	63	53	129	35
Frequency of phosphorylation in SCR SLiMs vs. non-SLiMs						
Percentage of S, T and Y phosphorylated				<i>p</i> -value		
Total SLiMs		Conserved SLiMs				
SLiMs	Non-SLiMs	SLiMs	Non-SLiMs	Total SLiMs		Conserved SLiMs
10	14	3	20	0.6		0.04
Biased distribution of S, T and Y to SCR SLiMs						
Percentage of amino acid that are S, T and Y				<i>p</i> -value		
Total SLiMs		Conserved SLiMs				
SLiMs	Non-SLiMs	SLiMs	Non-SLiMs	Total SLiMs		Conserved SLiMs
26	5	28	13	2.0 x 10 ⁻⁷		5.7 x 10 ⁻⁴

<https://doi.org/10.1371/journal.pone.0227642.t003>

2 and S4 and S5 Tables) indicating that about 10% of predicted SLiMs are *bona fide* sites of SCR phosphorylation, and only 15% of SLiMs annotated as potential phosphosites are *bona fide* sites of SCR phosphorylation. This suggests that a minority of SCR predicted SLiMs were phosphorylated which also does not support the hypothesis that SLiMs are preferential sites of phosphorylation. The reason that 6 of the 7 phosphosites are in predicted SLiMs is due to the bias of S, T and Y residues to SLiMs; the percentage of S, T and Y is significantly higher in the SLiMs and conserved SLiMs than in non-SLiM portions of SCR (Table 3).

Discussion

PTMs of SCRTT

PTMs are one mechanism proposed for the functional specificity of HOX proteins, and mapping PTMs of HOX proteins is a first step in testing this proposal. Bottom-up MS/MS analysis of SCRTT purified from developing *D. melanogaster* embryos identified many amino acid residues that were covalently modified (Fig 4). The final analysis of PTMs did not include modifications that could be due to sample preparation; however, some of these uninterrogated modifications may be a result of a biological process that regulates SCR activity. For example, deamidation of N321, which is the second N of the NANGE motif, may have a role in DNA binding [79]. The potentially biologically relevant PTMs will be discussed in relation to conserved protein domains and sequence motifs, SCR function, distribution within predicted SLiMs and the structure of an SCR-EXD-DNA complex. Of the conserved domains/motifs of SCR important for SCR activity (Fig 1A) [6,11–16], all are post-translationally modified with the exception of the octapeptide and KMAS motifs.

Phosphorylation of SCRTT

A clustered set of phosphorylations on the amino acid residues, T315, S316, T317 and T324 flank the NANGE motif (Fig 4). The NANGE motif is important for the suppression of ectopic proboscis formation suggesting that phosphorylation may regulate the transition of SCR activity between either determining T1 identity or determining labial identity [11]. The phosphosites map to a region of SCR not ordered in the SCR-EXD complex bound to *fkh* DNA (Fig 5). A few amino acid residues of the linker region (15 residues between YPWM motif and the HD) and the N-terminal of the HD (residues 3–9 of 60) of SCR interact stably with the minor groove of *fkh* DNA [16,81–84]. Although the amino acid residues of SCR involved in stable minor groove interactions were not found to be modified, the residues, T315, S316 and T317 that are part of the linker between the YPWM motif and HD, and T324 which is the first amino acid residue of the HD, may interact with DNA transiently. Phosphorylation adds negative charge to amino acid residues, and therefore, phosphorylation of the linker region and N-terminal arm of HD of SCR may interfere with transient minor groove interactions.

In the model for regulation of SCR activity proposed by Berry & Gehring [36], phosphorylation of the 6th and 7th amino acid residues of the HD, T329 and S330 inhibits SCR DNA binding and activity. cAMP-dependent protein kinase A phosphorylates these residues *in vitro*, and this negative regulation of SCR activity is proposed to be reversed by phosphatase PP2A-B' that removes the phosphates on these residues *in vitro*. However, loss of PP2A-B' activity had no effect on SCR activity suggesting that PP2A-B' is not involved in regulation of SCR activity [48]. Further, we have not detected phosphorylation of T329 and S330 suggesting that their phosphorylation by cAMP-dependent protein kinase A may also be an *in vitro* artifact.

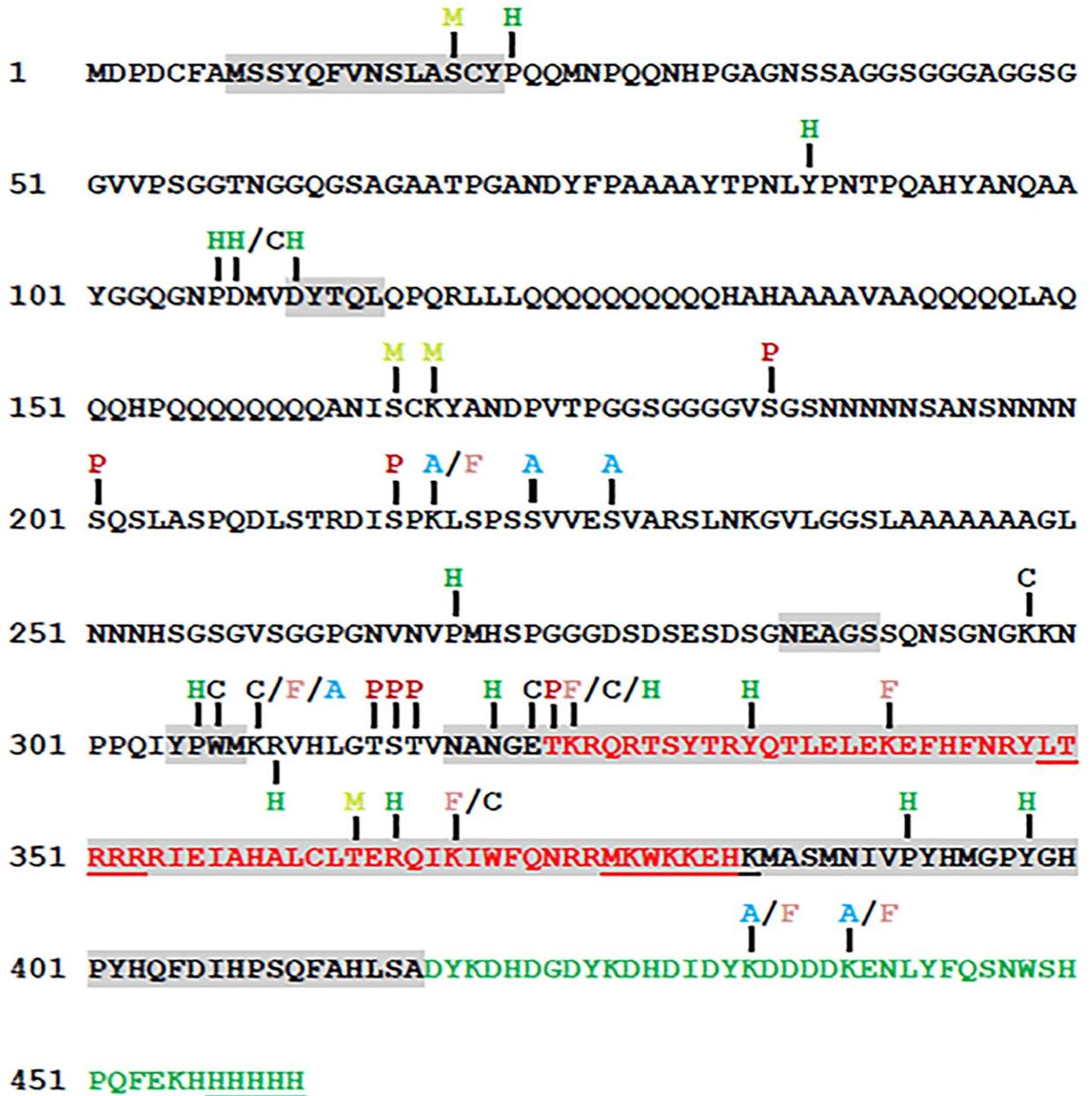


Fig 4. A summary map of post-translational modifications in SCR-EXD identified by MS/MS. Phosphorylation sites are indicated by P in dark red, acetylation by A in light blue, formylation by F in pink, methylation by M in yellowish green, carboxylation by C in black and hydroxylation by H in green. The homeodomain is highlighted in red and the triple tag is highlighted in green. The functional regions of SCR are shaded in grey (Fig 1A). The amino acids underlined were not detected by MS/MS. Phosphorylation at S216 was identified by a bulk proteomic analysis [80].

<https://doi.org/10.1371/journal.pone.0227642.g004>

PTMs of SCR residues found in the structure of the SCR-EXD-DNA complex

The structure of SCR-EXD bound to *fkh* regulatory DNA encompasses the evolutionarily conserved functional motifs/domains, YPWM, NANGE and HD (Fig 5). The Bilateran-specific YPWM motif of SCR and UBX binds to a hydrophobic pocket on the surface of EXD HD [16,85]. The MS/MS analysis identified hydroxylation at P306 and carboxylation at W307 of

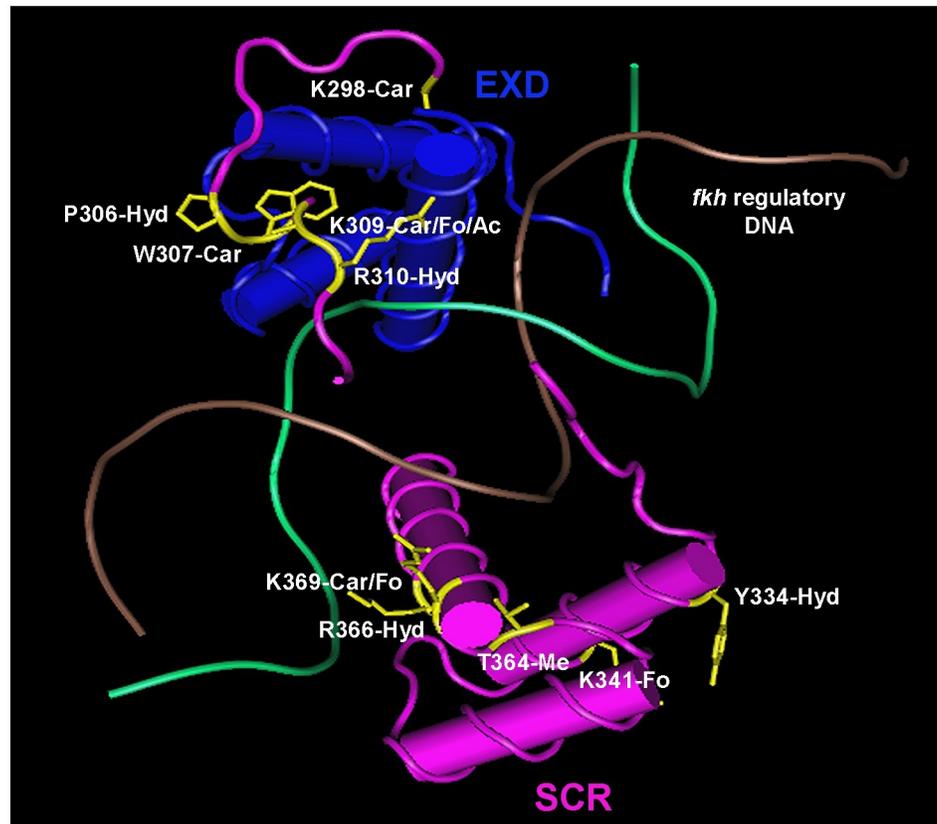


Fig 5. The structure of SCR-EXD-DNA complex determined by crystallography. SCR is shown in pink and EXD in blue. The two strands of *fkh* regulatory DNA is shown in brown and green. The modified amino acids of SCR along with their side chains are shown in yellow. Ac-acetylation, Car-carboxylation, Fo-formylation, Hyd-hydroxylation and Me-methylation. The structure coordinates with accession code 2R5Z (*fkh250*) were retrieved from RCSB Protein Data Bank [16]. Cn3D 4.3.1 (NCBI) was used to annotate the 3-D structure.

<https://doi.org/10.1371/journal.pone.0227642.g005>

the YPWM motif of SCR. These modifications render the YPWM motif hydrophilic and may interfere with the binding of YPWM to the hydrophobic pocket of the EXD HD. This might be a mechanism of regulation of SCR activity as SCR-EXD interaction is essential for activating the target, *fkh* gene which is required for salivary gland development [12].

The linker region of SCR interacts with the minor groove of *fkh* DNA. Narrowing of the DNA minor groove increases the negative electrostatic potential of the groove and proteins exploit this charged state of the groove by inserting a positively charged amino acid residue, thereby, making the interaction more stable [16,83]. Although K309 and R310 residues of SCR do not directly interact with the minor groove of DNA, they render the region of the protein positively charged which aids the neighboring H312 residue in making a strong contact with the DNA minor groove [16]. Carboxylation at K309 and hydroxylation at R310 adds negative charges to this region of the protein which may be a mechanism of inhibition of the SCR-DNA interaction, thereby, regulating the functional specificity of SCR.

The highly conserved HD is a compact self-folding protein domain which interacts with the major and minor groove of DNA [3,16,86,87]. All PTMs of the SCR HD are on the solvent exposed surface. These solvent exposed residues do not interact directly with DNA, and if they have a role in regulation of SCR activity, it is unlikely due to alterations in DNA binding. The formylation of K341 and K369 may arise as a secondary modification from oxidative DNA damage when the HD is bound to DNA [88].

Competition of acetylation and formylation observed in SCR

MS/MS analysis of SCRTT identified 4 lysine residues: K218 in the DISPK SLiM, K309 in the linker region, K434 and K439 in the triple tag that are acetylated in some peptides and formylated in others. This formylation/acetylation is found for lysine residues of core histone proteins [89].

In histones, lysine acetylation by histone acetyltransferases (HAT) and lysine deacetylation by histone deacetylases (HDAC), respectively are involved in chromatin remodeling and gene expression [90,91]. Acetylation has also been reported to modify the activity of TFs, thereby regulating the ability of the TF to bind DNA [92–94]. Besides histones, HATs and HDACs acetylate and deacetylate non-histone proteins respectively, including TFs which may be a mechanism regulating SCR activity [95–98].

Formylation of lysine residues is widespread in histones and other nuclear proteins and arise as a secondary modification due to oxidative DNA damage [88]. 5'-oxidation of DNA deoxyribose results in the formation of a highly reactive 3-formylphosphate residue which out-competes the acetylation mechanism and formylates the side-chain amino group of lysine [88]. Therefore, amino acid residues of a protein that are acetylated are also found to be formylated in many cases.

SLiM analysis

Out of 7 sites of SCR phosphorylation found, 6 were in SLiMs suggesting that predicted SLiMs are preferential sites of phosphorylation [6]. However, there is not a significant increase in the frequency of phosphorylation of S, T and Y residues in predicted SLiMs relative to the same residues outside the SLiMs. The main reason for 6 of 7 sites of phosphorylation mapping to SLiMs is because the phosphorylatable residues, S, T and Y are concentrated in predicted SLiMs. The ELM database search yielded 81 predicted SLiMs in *D. melanogaster* SCR that span 63% of the primary sequence of the disordered region of SCR (Table 3). Filtering the predicted SLiMs for evolutionary conservation did reduce the number of SLiMs to 53. As SLiMs have a low information content [20], a database search of SCR sequence may identify false positives, which may explain why a minority of predicted SLiMs were phosphosites. However, this lack of phosphorylation may also suggest that many SLiMs are targets of other PTMs [24–28] like methylation, hydroxylation, carboxylation, acetylation and formylation. Examples of other PTMs in predicted SLiMs of SCR include a methylation at S19 and a hydroxylation at P22 of the *Drosophila*-specific SLiM, SLAS-CYP and methylation at S166 and K168 of the SLiM, ANISCK (S4 Table and Fig 2).

The analysis of conserved SLiMs suggests that they are underphosphorylated relative to the remaining SCR sequence. The SCR predicted SLiM, KRVHLGTSTVNANGETKRQRT is phosphorylated at four sites (Figs 3 and 4 and S4 Table), but this SLiM is not considered conserved in our analysis (as the entire SLiM is not conserved 100% in the *Drosophila* species chosen for our analysis, *D. pseudoobscura* and *D. virilis*) although this SLiM is conserved 100% in many other *Drosophila* species like *D. sechellia*, *D. simulans*, *D. elegans*, etc. This SLiM contains the Arthropod-specific NANGE motif which is important for suppression of ectopic proboscis formation [11]. Moreover, this SLiM maps to the linker region which interacts with the minor groove of *fkh* DNA [16,81–84]. Hence, the phosphosites may function in regulating SCR activity. If this SLiM is considered 'conserved', 5 out of 7 phosphosites would map to conserved SLiMs; however, the frequency of phosphorylation of conserved SLiMs is not greater than the frequency of phosphorylation in non-conserved SLiM regions (p -value = 0.7).

Potential limitations

In order to detect PTMs by MS/MS analysis, a high concentration of the analyte protein is required which requires ectopic overexpression as endogenous protein levels are low. This

study identified PTMs on SCRTT protein expressed in all embryonic cells after administration of a heat-shock. One of the potential limitations of the ectopic approach is that the PTMs identified may not occur in cells where *Scr* is normally expressed. Another limitation of detecting PTMs on overexpressed proteins is that the heat-shock administered may affect the rate of protein modification or the amount of overexpressed protein is too high for the modification enzymes to modify completely. However, phosphoproteins expressed from heat-shock promoters are still heavily phosphorylated [38,40,41] ruling out an effect of the heat-shock on phosphorylation. An additional limitation is that PTMs are temporally and spatially regulated by the activity of modification enzymes, such as kinases and phosphatases for phosphorylation; therefore, SCR may only be post-translationally modified in a subset of developing embryonic cells resulting in detection of substoichiometric levels of modification. A final limitation is that although the total coverage of SCRTT is 96%, the depth of coverage in some regions of the protein is low and modifications in these under-represented regions may have been missed.

Although phosphorylation of proteins is common, detection of the phosphorylated amino acid residues is still a challenge. Three common explanations are used to address problems with phosphopeptide detection. Firstly, phosphopeptides are hydrophilic, and hence, they are lost during reversed-phase chromatography. Secondly, phosphopeptide ionization is selectively suppressed in the presence of unmodified peptides. Thirdly, the phosphopeptides have lower ionization or detection efficiency when compared to their unmodified moieties. There is no data to support the third argument [99]. In addition, multiply phosphorylated peptides were detected upon MS/MS analysis of a commercially purchased, pure, heavily phosphorylated, bovine α -casein protein [76] suggesting that the first two problems of phosphopeptide detection were not major issues. Therefore, the substoichiometric nature of SCRTT phosphorylation may not be a technical issue of detection but may arise during expression or purification of SCRTT. Phosphatases may remove phosphates during nuclear fractionation and nuclear lysis; however, this is unlikely during chromatography as it was performed in denaturing conditions. An attempt to enrich phosphopeptides of heavily phosphorylated α -casein using TiO_2 beads was successful although the yield of phosphopeptides was lower than with no enrichment. The TiO_2 -mediated phosphopeptide enrichment proved unsuccessful for SCRTT. The MS/MS analysis of SCR protein is most likely not exhaustive and the PTMs mapped for SCR may not be the complete set that can be detected by MS/MS analysis; however, the analysis of a purified SCR protein has detected more PTMs than the bulk proteomic analyses [80,100].

Finally, our study only identified PTMs and has not assessed the functional importance directly. Two problems are associated with a functional analysis of SCR PTMs: the substoichiometric levels of modification and differential pleiotropy of *Scr* mutant alleles. Because the substoichiometric levels of modification may suggest tissue specific modifications, this whole embryo analysis is unable to suggest the tissue to assay function. Also, differential pleiotropy suggests that HOX proteins outside the HD are composed of small domains/motifs that make small tissue-specific contributions to overall HOX activity [6]. Therefore, motifs like NANGE do not have strong phenotypes when mutant making clear-cut interpretation of a genetic analysis difficult [11].

Conclusion

This study identified sites of phosphorylation and other PTMs in a tagged HOX protein, SCRTT, extracted from developing *Drosophila melanogaster* embryos. 18 of 44 modifications map to functionally important regions of SCR. In testing the hypothesis that HOX predicted SLiMs are preferential sites of phosphorylation, we found that more phosphosites mapped to

predicted SLiMs but no support for the hypothesis that the S, T and Y residues of predicted SLiMs are more frequently phosphorylated.

Supporting information

S1 Fig. Coverage map of embryonic SCRTT. Each blue line underneath the primary protein sequence of SCRTT represents a chemically distinct peptide identified by MS/MS analysis. The modifications are indicated by letters or symbols on the blue lines. The C-terminal triple tag sequence is represented by the faded region. On the right is the legend for all modifications shown in the figure. Amino acid substitutions have been excluded from the figure. (A) Analysis of tryptic peptides. A1, A2, A3 and A4 are independent MS/MS analyses. (B) Analysis of chymotryptic peptides. A1 and A2 are independent MS/MS analyses. (C) Analysis of thermolytic peptides. A1 and A2 are independent MS/MS analyses. (PDF)

S2 Fig. Coverage map of commercially purchased bovine α -casein. Each blue line underneath the primary protein sequence of α -casein represents a chemically distinct tryptic peptide identified by MS/MS analysis. The modifications are indicated by letters or symbols on the blue lines. On the right is the legend for all modifications shown in the figure. Amino acid substitutions have been excluded from the figure. (A) Analysis of α -casein without TiO₂ treatment. (B) Analysis of α -casein with TiO₂ treatment. (PDF)

S3 Fig. Identification of phosphopeptides of a commercially purchased, pure phosphoprotein, α -casein. The figure shows a region of 41–80 of α -casein (full protein in [S2A Fig](#)) and each blue line underneath the primary protein sequence represents a chemically distinct peptide identified by MS/MS analysis. The peptides are heavily modified, and the modifications are indicated by letters or symbols on the blue lines. On the right is the legend for all modifications shown in the figure. Amino acid substitutions have been excluded from the figure. (PDF)

S4 Fig. Phosphorylation of Serine 185, Serine 201, Threonine 315, Serine 316, Threonine 317 and Threonine 324 residues of SCRTT. MS² spectra for the three phosphopeptides identified by LC-MS/MS are shown. (A) Phosphorylation of Serine 185. (B) Phosphorylation of Serine 201. The inset box shows fragment ions with m/z 1690 to 1790. (C) Phosphorylation of Threonine 315. (D) Phosphorylation of Serine 316. (E) Phosphorylation of Threonine 317. The inset box shows fragment ions with m/z 900 to 1400. (F) Phosphorylation of Threonine 324. The peptide sequence and m/z ratio are indicated on the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidences of phosphorylation. The relevant fragment ions and their m/z ratios supporting phosphorylation are labelled in the spectra. (PDF)

S5 Fig. Acetylation of Lysine 218, Serine 223, Serine 227, Lysine 309, Lysine 434 and Lysine 439 residues of SCRTT. MS² spectra of the peptide identified by LC-MS/MS is shown. (A) Acetylation of Lysine 218. (B) Acetylation of Serine 223. (C) Acetylation of Serine 227. (D) Acetylation of Lysine 309. The inset box shows fragment ions with m/z 1150 to 1350. (E) Acetylation of Lysine 434. (F) Acetylation of Lysine 439. The inset box shows fragment ions with m/z 390 to 580. The peptide sequence and m/z ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming acetylation. The relevant fragment ions and

their m/z ratios supporting acetylation are labelled in the spectra. (PDF)

S6 Fig. Formylation of Lysine 218, 309, 325, 341, 369, 434 and 439 residues of SCR TT. MS² spectra of the peptide identified by LC-MS/MS is shown. (A) Formylation of Lysine 218. The inset box shows fragment ions with m/z 300 to 700. (B) Formylation of Lysine 309. The inset box shows fragment ions with m/z 1100 to 1300. (C) Formylation of Lysine 325. The inset box shows fragment ions with m/z 700 to 1400. (D) Formylation of Lysine 341. (E) Formylation of Lysine 369. The inset box shows fragment ions with m/z 700 to 1100. (F) Formylation of Lysine 434. The inset box shows fragment ions with m/z 830 to 1010. (G) Formylation of Lysine 439. The inset box shows fragment ions with m/z 950 to 1200. The peptide sequence and m/z ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming formylation. The relevant fragment ions and their m/z ratios supporting formylation are labelled in the spectra. (PDF)

S7 Fig. Methylation of Serine 19, Serine 166, Lysine 168 and Threonine 364 residues of SCR TT. MS² spectra of the peptide identified by LC-MS/MS is shown. (A) Methylation of Lysine 19. (B) Methylation of Serine 166. (C) Methylation of Lysine 168. (D) Methylation of Threonine 364. The inset box shows fragment ions with m/z 1000 to 1150. The peptide sequence and m/z ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming methylation. The relevant fragment ions and their m/z ratios supporting methylation are labelled in the spectra. (PDF)

S8 Fig. Carboxylation of Aspartic acid 108, Lysine 298, Tryptophan 307, Lysine 309, Glutamic acid 323, Lysine 325 and Lysine 369 residues of SCR TT. MS² spectra of the peptide identified by LC-MS/MS is shown. (A) Carboxylation of Aspartic acid 108. (B) Carboxylation of Lysine 298. The inset box shows fragment ions with m/z 470 to 618. (C) Carboxylation of Tryptophan 307. The inset box shows fragment ions with m/z 780 to 1030. (D) Carboxylation of Lysine 309. The inset box shows fragment ions with m/z 990 to 1300. (E) Carboxylation of Glutamic acid 323. (F) Carboxylation of Lysine 325. (G) Carboxylation of Lysine 369. The inset box shows fragment ions with m/z 840 to 1020. The peptide sequence and m/z ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming carboxylation. The relevant fragment ions and their m/z ratios supporting carboxylation are labelled in the spectra. (PDF)

S9 Fig. Hydroxylation of Proline 22, Tyrosine 87, Proline 107, Aspartic Acid 108, Aspartic acid 111, Proline 269, Proline 306, Arginine 310, Asparagine 321, Lysine 325, Tyrosine 334, Arginine 366, Proline 392 and Tyrosine 398 residues of SCR TT. MS² spectra of the peptide identified by LC-MS/MS is shown. (A) Hydroxylation of Proline 22. The inset box shows fragment ions with m/z 835 to 920. (B) Hydroxylation of Tyrosine 87. (C) Hydroxylation of Proline 107. The inset box shows fragment ions with m/z 670 to 1000. (D) Hydroxylation of Aspartic acid 108. The inset box shows fragment ions with m/z 620 to 900. (E) Hydroxylation of Aspartic acid 111. The inset box shows fragment ions with m/z 420 to 600. (F) Hydroxylation of Proline 269. The inset box shows fragment ions with m/z 420 to 940. (G) Hydroxylation of Proline 306. (H) Hydroxylation of R310. (I) Hydroxylation of N321. (J) Hydroxylation of K325. (K) Hydroxylation of Y334. The inset box shows fragment ions with m/z 788 to 1040. (L) Hydroxylation of R366. (M) Hydroxylation of P392. The inset box shows

fragment ions with m/z 750 to 890. (N) Hydroxylation of Y398. The inset box shows fragment ions with m/z 945 to 1010. The peptide sequence and m/z ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming hydroxylation. The relevant fragment ions and their m/z ratios supporting hydroxylation are labelled in the spectra.

(PDF)

S10 Fig. Alignment of SLiMs of animal SCR homologs. SLiMs of *D. melanogaster* SCR which are conserved across various taxonomic groups were aligned using multiple sequence alignment tools, MAFFT v. 7 and Clustal Omega. The phylogenetic tree on the left is not drawn to scale and it merely depicts the relationship among the organisms and not evolutionary time of divergence. The block diagram below the aligned sequences shows the HD and four SLiMs conserved beyond Diptera.

(PDF)

S1 Raw Images. Original blot and gel images.

(PDF)

S1 Table. Post-translational modifications of embryonic SCR^{TT} identified by LC-MS/MS.

(PDF)

S2 Table. Phosphosites in bovine α -casein identified by MS/MS.

(PDF)

S3 Table. Phosphopeptides of α -casein identified by MS/MS (TiO₂-enriched vs. non-TiO₂).

(PDF)

S4 Table. SLiMs in *D. melanogaster* SCR.

(PDF)

S5 Table. Conserved SLiMs in *D. melanogaster* SCR.

(PDF)

S6 Table. Accession numbers of animal SCR protein homolog sequences (retrieved from NCBI).

(PDF)

Acknowledgments

We thank Paula Pittock from the Biological Mass Spectrometry Laboratory, The University of Western Ontario, London, Ontario, Canada for analyzing the peptide samples by LC-ESI MS/MS and assisting us with the PTM characterization using the PEAKS™ DB software, Victoria Clarke and Kristina Jurcic from the Functional Proteomics Facility, The University of Western Ontario, London, Ontario, Canada for their assistance with the Ettan® SpotPicker™ instrument and in-gel digestion of our samples, Lovesha Sivanantharajah for providing us with the pBS *Scr^{TT}* construct, Ben Rubin for his suggestions on the statistical analysis, Stuart Cameron and Gurjit Randhawa for assistance with the project. We extend our gratitude to René Rezsöházy for valuable suggestions on the manuscript. We thank the two reviewers for their helpful suggestions.

Author Contributions

Conceptualization: Anirban Banerjee, Anthony Percival-Smith.

Data curation: Anirban Banerjee.

Formal analysis: Anirban Banerjee.

Funding acquisition: Anthony Percival-Smith.

Investigation: Anirban Banerjee, Anthony Percival-Smith.

Methodology: Anirban Banerjee, Anthony Percival-Smith.

Project administration: Anthony Percival-Smith.

Resources: Anirban Banerjee.

Software: Anirban Banerjee.

Supervision: Anthony Percival-Smith.

Validation: Anirban Banerjee.

Visualization: Anirban Banerjee.

Writing – original draft: Anirban Banerjee.

Writing – review & editing: Anirban Banerjee, Anthony Percival-Smith.

References

1. Akam M. Hox genes, homeosis and the evolution of segmental identity: no need for hopeless monsters. *Int J Dev Biol.* 1998; 42: 445–451. <https://doi.org/10.1387/ijdb.9654030> PMID: 9654030
2. Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature.* 1978; 276: 565–570. <https://doi.org/10.1038/276565a0> PMID: 103000
3. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, et al. Homeodomain-DNA recognition. *Cell.* 1994; 78: 211–223. [https://doi.org/10.1016/0092-8674\(94\)90292-5](https://doi.org/10.1016/0092-8674(94)90292-5) PMID: 8044836
4. Primon M, Hunter KD, Pandha HS, Morgan R. Kinase Regulation of HOX Transcription Factors. *Cancers.* 2019; 11: E508. <https://doi.org/10.3390/cancers11040508> PMID: 30974835
5. Draime A, Bridoux L, Graba Y, Rezsoszy R. Post-translational modifications of HOX proteins, an underestimated issue. *Int J Dev Biol.* 2018; 62: 733–744. <https://doi.org/10.1387/ijdb.180178rr> PMID: 30604843
6. Sivanantharajah L, Percival-Smith A. Differential pleiotropy and HOX functional organization. *Dev Biol.* 2015; 398: 1–10. <https://doi.org/10.1016/j.ydbio.2014.11.001> PMID: 25448696
7. Lewis RA, Wakimoto BT, Denell RE, Kaufman TC. Genetic analysis of the Antennapedia gene complex (ANT-C) and adjacent chromosomal regions of *Drosophila melanogaster*. II. Polytene chromosome segments 84A-84B1,2. *Genetics.* 1980; 95: 383–397. PMID: 17249042
8. Struhl G. Genes controlling segmental specification in the *Drosophila* thorax. *Proc Natl Acad Sci U S A.* 1982; 79: 7380–7384. <https://doi.org/10.1073/pnas.79.23.7380> PMID: 6961417
9. Panzer S, Weigel D, Beckendorf SK. Organogenesis in *Drosophila melanogaster*: embryonic salivary gland determination is controlled by homeotic and dorsoventral patterning genes. *Development.* 1992; 114: 49–57. PMID: 1349523
10. Percival-Smith A, Weber J, Gilfoyle E, Wilson P. Genetic characterization of the role of the two HOX proteins, Proboscipedia and Sex Combs Reduced, in determination of adult antennal, tarsal, maxillary palp and proboscis identities in *Drosophila melanogaster*. *Development.* 1997; 124: 5049–5062. PMID: 9362475
11. Percival-Smith A, Sivanantharajah L, Pelling JJH, Teft WA. Developmental competence and the induction of ectopic proboscises in *Drosophila melanogaster*. *Dev Genes Evol.* 2013; 223: 375–387. <https://doi.org/10.1007/s00427-013-0454-8> PMID: 24121940
12. Ryoo HD, Mann RS. The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev.* 1999; 13: 1704–1716. <https://doi.org/10.1101/gad.13.13.1704> PMID: 10398683
13. Tour E, Hittinger CT, McGinnis W. Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development.* 2005; 132: 5271–5281. <https://doi.org/10.1242/dev.02138> PMID: 16284118

14. Sivanantharajah L, Percival-Smith A. Analysis of the sequence and phenotype of *Drosophila* *Sex combs reduced* alleles reveals potential functions of conserved protein motifs of the Sex combs reduced protein. *Genetics*. 2009; 182: 191–203. <https://doi.org/10.1534/genetics.109.100438> PMID: 19293143
15. Sivanantharajah L, Percival-Smith A. Acquisition of a leucine zipper motif as a mechanism of antimorphy for an allele of the *Drosophila* *Hox* gene *Sex combs reduced*. *G3-Genes Genom Genet*. 2014; 4: 829–838. <https://doi.org/10.1534/g3.114.010769>
16. Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, Crickmore MA, et al. Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. *Cell*. 2007; 131: 530–543. <https://doi.org/10.1016/j.cell.2007.09.024> PMID: 17981120
17. Hittinger CT, Stern DL, Carroll SB. Pleiotropic functions of a conserved insect-specific Hox peptide motif. *Development*. 2005; 132: 5261–5270. <https://doi.org/10.1242/dev.02146> PMID: 16267091
18. Prince F, Katsuyama T, Oshima Y, Plaza S, Resendez-Perez D, Berry M, et al. The YPWM motif links Antennapedia to the basal transcriptional machinery. *Development*. 2008; 135: 1669–1679. <https://doi.org/10.1242/dev.018028> PMID: 18367556
19. Merabet S, Litim-Mecheri I, Karlsson D, Dixit R, Saadaoui M, Monier B, et al. Insights into Hox protein function from a large scale combinatorial analysis of protein domains. *PLoS Genet*. 2011; 7: e1002302. <https://doi.org/10.1371/journal.pgen.1002302> PMID: 22046139
20. Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, et al. Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol*. 2005; 3: e405. <https://doi.org/10.1371/journal.pbio.0030405> PMID: 16279839
21. Neduva V, Russell RB. Peptides mediating interaction networks: new leads at last. *Curr Opin Biotechnol*. 2006; 17: 465–471. <https://doi.org/10.1016/j.copbio.2006.08.002> PMID: 16962311
22. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. *Mol Biosyst*. 2012; 8: 268–281. <https://doi.org/10.1039/c1mb05231d> PMID: 21909575
23. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev*. 2014; 114: 6733–6778. <https://doi.org/10.1021/cr400585q> PMID: 24926813
24. Puntervoll P, Linding R, Gemünd C, Chabanis-Davidson S, Mattingsdal M, Cameron S, et al. ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res*. 2003; 31: 3625–3630. <https://doi.org/10.1093/nar/gkg545> PMID: 12824381
25. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res*. 2004; 32: 1037–1049. <https://doi.org/10.1093/nar/gkh253> PMID: 14960716
26. Khan AN, Lewis PN. Unstructured conformations are a substrate requirement for the Sir2 family of NAD-dependent protein deacetylases. *J Biol Chem*. 2005; 280: 36073–36078. <https://doi.org/10.1074/jbc.M508247200> PMID: 16131486
27. Gould CM, Diella F, Via A, Puntervoll P, Gemünd C, Chabanis-Davidson S, et al. ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res*. 2010; 38: D167–D180. <https://doi.org/10.1093/nar/gkp1016> PMID: 19920119
28. Gouw M, Michael S, Sámano-Sánchez H, Kumar M, Zeke A, Lang B, et al. The eukaryotic linear motif resource—2018 update. *Nucleic Acids Res*. 2018; 46: D428–D434. <https://doi.org/10.1093/nar/gkx1077> PMID: 29136216
29. Uyar B, Weatheritt RJ, Dinkel H, Davey NE, Gibson TJ. Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? *Mol Biosyst*. 2014; 10: 2626–2642. <https://doi.org/10.1039/c4mb00290c> PMID: 25057855
30. Hrabec P, O'Maille PE, Silberfarb A, Davis-Anderson K, Generous N, McMahon BH, et al. Resources to Discover and Use Short Linear Motifs in Viral Proteins. *Trends Biotechnol*. 2019; Article in press. <https://doi.org/10.1016/j.tibtech.2019.07.004>
31. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. *Trends Biochem Sci*. 2011; 36: 159–169. <https://doi.org/10.1016/j.tibs.2010.10.002> PMID: 21146412
32. Gógl G, Biri-Kovács B, Durbesson F, Jane P, Nomine Y, Kostmann C, et al. Rewiring of RSK-PDZ Interactome by Linear Motif Phosphorylation. *J Mol Biol*. 2019; 431: 1234–1249. <https://doi.org/10.1016/j.jmb.2019.01.038> PMID: 30726710
33. Mylin LM, Bhat JP, Hopper JE. Regulated phosphorylation and dephosphorylation of GAL4, a transcriptional activator. *Genes Dev*. 1989; 3: 1157–1165. <https://doi.org/10.1101/gad.3.8.1157> PMID: 2676720
34. Hunter T, Karin M. The regulation of transcription by phosphorylation. *Cell*. 1992; 70: 375–387. [https://doi.org/10.1016/0092-8674\(92\)90162-6](https://doi.org/10.1016/0092-8674(92)90162-6) PMID: 1643656

35. Ardito F, Giuliani M, Perrone D, Troiano G, Lo Muzio L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med*. 2017; 40: 271–280. <https://doi.org/10.3892/ijmm.2017.3036> PMID: 28656226
36. Berry M, Gehring W. Phosphorylation status of the SCR homeodomain determines its functional activity: essential role for protein phosphatase 2A,B'. *EMBO J*. 2000; 19: 2946–2957. <https://doi.org/10.1093/emboj/19.12.2946> PMID: 10856239
37. Jaffe L, Ryoo HD, Mann RS. A role for phosphorylation by casein kinase II in modulating Antennapedia activity in *Drosophila*. *Genes Dev*. 1997; 11: 1327–1340. <https://doi.org/10.1101/gad.11.10.1327> PMID: 9171376
38. Gavis ER, Hogness, DS. Phosphorylation, expression and function of the *Ultrabithorax* protein family in *Drosophila melanogaster*. *Development*. 1991; 112: 1077–1093. PMID: 1682129
39. Stultz BG, Jackson DG, Mortin MA, Yang X, Beachy PA, Hursh DA. Transcriptional activation by *extradenticle* in the *Drosophila* visceral mesoderm. *Dev Biol*. 2006; 290: 482–494. <https://doi.org/10.1016/j.ydbio.2005.11.041> PMID: 16403493
40. Krause HM, Klemenz R, Gehring WJ. Expression, modification, and localization of the *fushi tarazu* protein in *Drosophila* embryos. *Genes Dev*. 1988; 2: 1021–1036. <https://doi.org/10.1101/gad.2.8.1021> PMID: 3049237
41. Krause HM, Gehring WJ. Stage-specific phosphorylation of the *fushi tarazu* protein during *Drosophila* development. *EMBO J*. 1989; 8: 1197–1204. PMID: 2743978
42. Bourbon HM, Martin-Blanco E, Rosen D, Kornberg TB. Phosphorylation of the *Drosophila* engrailed protein at a site outside its homeodomain enhances DNA binding. *J Biol Chem*. 1995; 270: 11130–11139. <https://doi.org/10.1074/jbc.270.19.11130> PMID: 7744743
43. Driever W, Nüsslein-Volhard C. The bicoid protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature*. 1989; 337: 138–143. <https://doi.org/10.1038/337138a0> PMID: 2911348
44. Gay NJ, Poole SJ, Kornberg TB. The *Drosophila* engrailed protein is phosphorylated by a serine-specific protein kinase. *Nucleic Acids Res*. 1988; 16: 6637–6647. <https://doi.org/10.1093/nar/16.14.6637> PMID: 2899884
45. Ronchi E, Treisman J, Dostatni N, Struhl G, Desplan C. Down-regulation of the *Drosophila* morphogen Bicoid by the torso-receptor mediated signal transduction cascade. *Cell*. 1993; 74: 347–355. [https://doi.org/10.1016/0092-8674\(93\)90425-p](https://doi.org/10.1016/0092-8674(93)90425-p) PMID: 8343961
46. Dong J, Hung LH, Strome R, Krause HM. A phosphorylation site in the Ftz homeodomain is required for activity. *EMBO J*. 1998; 17: 2308–2318. <https://doi.org/10.1093/emboj/17.8.2308> PMID: 9545243
47. Janody F, Sturny R, Catala F, Desplan C, Dostatni N. Phosphorylation of Bicoid on MAP-kinase sites: contribution to its interaction with the torso pathway. *Development*. 2000; 127: 279–289. PMID: 10603346
48. Moazzen H, Rosenfeld R, Percival-Smith A. Non-requirement of a regulatory subunit of Protein Phosphatase 2A, PP2A-B', for activation of Sex comb reduced activity in *Drosophila melanogaster*. *Mech Dev*. 2009; 126: 605–610. <https://doi.org/10.1016/j.mod.2009.06.1084> PMID: 19563886
49. Johnson H, Eyers CE. Analysis of Post-translational Modifications by LC-MS/MS. In: Cutillas P, Timms J. (eds) LC-MS/MS in Proteomics. *Methods Mol Biol. (Methods and Protocols)*. Humana Press, Totowa, NJ. 2010. vol. 658. pp. 93–108. https://doi.org/10.1007/978-1-60761-780-8_5
50. Tiefenbach J, Moll PR, Nelson MR, Hu C, Baev L, Kislinger T, et al. A Live Zebrafish-Based Screening System for Human Nuclear Receptor Ligand and Cofactor Discovery. *PLoS ONE*, 2010; 5: 1–12. <https://doi.org/10.1371/journal.pone.0009797>
51. Studier FW, Moffatt BA. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes. *J Mol Biol*. 1986; 189: 113–130. [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2) PMID: 3537305
52. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol*. 1990; 185: 60–89. [https://doi.org/10.1016/0076-6879\(90\)85008-c](https://doi.org/10.1016/0076-6879(90)85008-c) PMID: 2199796
53. Thummel C, Pirrotta V. New pCaSpeR P-element vectors. *D. I. S*. 1992; 71: 150.
54. Rubin GM, Spradling AC. Genetic transformation of *Drosophila* with transposable element vectors. *Science*. 1982; 218: 348–353. <https://doi.org/10.1126/science.6289436> PMID: 6289436
55. Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*. 1993; 118: 401–415. PMID: 8223268
56. Wieschaus E, Nüsslein-Volhard C. Looking at embryos. In: Roberts DB, editor. *Drosophila: A Practical Approach*. IRL, Oxford; 1986. pp. 199–228.

57. Loughran ST, Walls D. Purification of Poly-Histidine-Tagged Proteins. *Methods Mol Biol.* 2011; 681: 311–335. https://doi.org/10.1007/978-1-60761-913-0_17 PMID: 20978973
58. Haneskog L. Purification of Histidine-Tagged Proteins under Denaturing Conditions Using IMAC. *CSH Protoc.* 2006; 2006: pdb.prot4221. <https://doi.org/10.1101/pdb.prot4221>
59. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 1989.
60. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: *De Novo* Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol Cell Proteomics.* 2012; 11: M111.010587. <https://doi.org/10.1074/mcp.M111.010587> PMID: 22186715
61. Humphrey SJ, Azimifar SB, Mann M. High throughput phosphoproteomics reveals *in vivo* insulin signaling dynamics. *Nat Biotechnol.* 2015; 33: 990–995. <https://doi.org/10.1038/nbt.3327> PMID: 26280412
62. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for Matrix-Assisted Laser Desorption/Ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem.* 2003; 75: 663–670. <https://doi.org/10.1021/ac026117i> PMID: 12585499
63. Hochuli E, Döbeli H, Schacher A. New metal chelate adsorbent for proteins and peptides containing neighbouring histidine residues. *J Chromatogr.* 1987; 411: 177–184. [https://doi.org/10.1016/s0021-9673\(00\)93969-4](https://doi.org/10.1016/s0021-9673(00)93969-4) PMID: 3443622
64. Hochuli E, Bannwarth W, Döbeli H, Gentz R, Stüber D. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Nat Biotechnol.* 1988; 6: 1321–1325. <https://doi.org/10.1038/nbt1188-1321>
65. Bradford MM. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 1976; 72: 248–254. <https://doi.org/10.1006/abio.1976.9999> PMID: 942051
66. Merabet S, Dard A. Tracking context-specific transcription factors regulating Hox activity. *Dev Dyn.* 2014; 243: 16–23. <https://doi.org/10.1002/dvdy.24002> PMID: 23794379
67. Wiellette EL, Harding KW, Mace KA, Ronshaugen MR, Wang FY, McGinnis W. *spen* encodes and RNP motif protein that interacts with Hox pathways to repress the development of head-like sclerites in the *Drosophila* trunk. *Development.* 1999; 126: 5373–5385. PMID: 10556062
68. Billeter M, Qian Y, Otting G, Müller M, Gehring WJ, Wüthrich K. Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by ¹H nuclear magnetic resonance spectroscopy. *J Mol Biol.* 1990; 214: 183–197. [https://doi.org/10.1016/0022-2836\(90\)90155-f](https://doi.org/10.1016/0022-2836(90)90155-f) PMID: 2164583
69. Sterck L, Billiau K, Abeel T, Rouzé P, Van de Peer Y. ORCAE: online resource for community annotation of eukaryotes. *Nat Methods.* 2012; 9: 1041. <https://doi.org/10.1038/nmeth.2242> PMID: 23132114
70. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019; 20: 1160–1166 <https://doi.org/10.1093/bib/bbx108> PMID: 28968734
71. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011; 7: 539. <https://doi.org/10.1038/msb.2011.75> PMID: 21988835
72. Fisher RA. On the interpretation of χ^2 from contingency tables, and the calculation of P. *J Royal Stat Soc.* 1922; 85: 87–94. <https://doi.org/10.2307/2340521>
73. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 2012; 40: W597–W603. <https://doi.org/10.1093/nar/gks400> PMID: 22661580
74. Gibson G, Schier A, LeMotte P, Gehring WJ. The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. *Cell.* 1990; 62: 1087–1103. [https://doi.org/10.1016/0092-8674\(90\)90386-s](https://doi.org/10.1016/0092-8674(90)90386-s) PMID: 1976044
75. Zhao JJ, Lazzarini RA, Pick L. The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila Sex combs reduced* gene. *Genes Dev.* 1993; 7: 343–354. <https://doi.org/10.1101/gad.7.3.343> PMID: 8095481
76. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jørgensen TJD. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics.* 2005; 4: 873–886. <https://doi.org/10.1074/mcp.T500007-MCP200> PMID: 15858219
77. Diella F, Haslam N, Chica C, Budd A, Michael S, Brown NP, et al. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci.* 2008; 13: 6580–6603. <https://doi.org/10.2741/3175> PMID: 18508681

78. Dinkel H, Michael S, Weatheritt RJ, Davey NE, Van Roey K, Altenberg B, et al. ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res.* 2012; 40: D242–D251. <https://doi.org/10.1093/nar/gkr1064> PMID: 22110040
79. O'Connell NE, Lelli K, Mann RS, Palmer AG 3rd Asparagine deamidation reduces DNA-binding affinity of the *Drosophila melanogaster* Scr homeodomain. *FEBS Lett.* 2015; 589: 3237–3241. <https://doi.org/10.1016/j.febslet.2015.09.020> PMID: 26435141
80. Zhai B, Villén J, Beausoleil SA, Mintseris J, Gygi SP. Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J Proteome Res.* 2008; 7: 1675–1682. <https://doi.org/10.1021/pr700696a> PMID: 18327897
81. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. Origins of Specificity in Protein-DNA Recognition. *Annu Rev Biochem.* 2010; 79: 233–269. <https://doi.org/10.1146/annurev-biochem-060408-091030> PMID: 20334529
82. Rohs R, West SM, Liu P, Honig B. Nuance in the double-helix and its role in protein-DNA recognition. *Curr Opin Struct Biol.* 2009; 19: 171–177. <https://doi.org/10.1016/j.sbi.2009.03.002> PMID: 19362815
83. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. The role of DNA shape in protein-DNA recognition. *Nature.* 2009; 461: 1248–1253. <https://doi.org/10.1038/nature08473> PMID: 19865164
84. Abe N, Dror I, Yang L, Slattey M, Zhou T, Bussemaker HJ, et al. Deconvolving the recognition of DNA shape from sequence. *Cell.* 2015; 161: 307–318. <https://doi.org/10.1016/j.cell.2015.02.008> PMID: 25843630
85. Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature.* 1999; 397: 714–719. <https://doi.org/10.1038/17833> PMID: 10067897
86. Otting G, Qian YQ, Billeter M, Müller M, Affolter M, Gehring WJ, et al. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* 1990; 9: 3085–3092. <https://doi.org/10.1002/j.1460-2075.1990.tb07505.x> PMID: 1976507
87. Religa TL, Johnson CM, Vu DM, Brewer SH, Dyer RB, Fersht AR. The helix-turn-helix motif as an ultrafast independently folding domain: The pathway of folding of Engrailed homeodomain. *Proc Natl Acad Sci U S A.* 2007; 104: 9272–9277. <https://doi.org/10.1073/pnas.0703434104> PMID: 17517666
88. Jiang T, Zhou X, Taghizadeh K, Dong M, Dedon PC. N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc Natl Acad Sci U S A.* 2007; 104: 60–65. <https://doi.org/10.1073/pnas.0606775103> PMID: 17190813
89. Wiśniewski JR, Zougman A, Mann M. N^ε-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic Acids Res.* 2008; 36: 570–577. <https://doi.org/10.1093/nar/gkm1057> PMID: 18056081
90. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A.* 1964; 51: 786–794. <https://doi.org/10.1073/pnas.51.5.786> PMID: 14172992
91. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. *Ann Rev Biochem.* 2001; 70: 81–120. <https://doi.org/10.1146/annurev.biochem.70.1.81> PMID: 11395403
92. Gu W, Roeder RG. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell.* 1997; 90: 595–606. [https://doi.org/10.1016/s0092-8674\(00\)80521-8](https://doi.org/10.1016/s0092-8674(00)80521-8) PMID: 9288740
93. Kouzarides T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* 2000; 19: 1176–1179. <https://doi.org/10.1093/emboj/19.6.1176> PMID: 10716917
94. Bannister AJ, Miska EA. Regulation of gene expression by transcription factor acetylation. *Cell Mol Life Sci.* 2000; 57: 1184–1192. <https://doi.org/10.1007/pl00000758> PMID: 11028911
95. Park JM, Jo SH, Kim MY, Kim TH, Ahn YH. Role of transcription factor acetylation in the regulation of metabolic homeostasis. *Protein Cell.* 2015; 6: 804–813. <https://doi.org/10.1007/s13238-015-0204-y> PMID: 26334401
96. Wang C, Tian L, Popov VM, Pestell RG. Acetylation and nuclear receptor action. *J Steroid Biochem Mol Biol.* 2011; 123: 91–100. <https://doi.org/10.1016/j.jsbmb.2010.12.003> PMID: 21167281
97. Glozak MA, Sengupta N, Zhang X, Seto E. Acetylation and deacetylation of non-histone proteins. *Gene.* 2005; 363: 15–23. <https://doi.org/10.1016/j.gene.2005.09.010> PMID: 16289629
98. Choudhary C, Weinert BT, Nishida Y, Verdin E, Mann M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol.* 2014; 15: 536–550. <https://doi.org/10.1038/nrm3841> PMID: 25053359
99. Steen H, Jeganathirajah JA, Rush J, Morrice N, Kirschner MW. Phosphorylation analysis by mass spectrometry. Myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics.* 2006; 5: 172–181. <https://doi.org/10.1074/mcp.M500135-MCP200> PMID: 16204703

100. Hu Y, Sopko R, Chung V, Foos M, Studer RA, Landry SD, et al. iProteinDB: An Integrative Database of *Drosophila* Post-translational Modifications. *G3-Genes Genom Genet*. 2019; 9: 1–11. <https://doi.org/10.1534/g3.118.200637>