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

Anirban Banerjee*, Anthony Percival-Smith

Department of Biology, The University of Western Ontario, 1151 Richmond Street, London, Ontario N6A 5B7, Canada

*Corresponding author

Email address: abanerj7@uwo.ca (Anirban Banerjee)

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 (SCRTT) 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.

Keywords

Homeotic selector, Sex combs reduced, transcription factors, post-translational modification, short linear motifs, Tandem Mass Spectrometry

Abbreviations

A-P, anterior-posterior; HOX, homeotic selector; SCR, Sex combs reduced; TT, triple tag; PB, Proboscipedia; HD, homeodomain; PTM, post-translational modification; TF, transcription factor; HAT, histone acetyltransferase; HDAC, histone deacetylase; MS, mass spectrometry; SLiM, short linear motif; AEL, after egg laying; AD, average depth; AD_{orc}, average depth of regions covered; CTD, C-terminal domain; ESI, Electrospray ionization; FDR, false discovery rate; LC, liquid chromatography

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 (reviewed in Akam, 1998; Lewis, 1978). 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) (Gehring *et al.*, 1994). 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 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 (reviewed in Primon *et al.*, 2019; Draime *et al.*, 2018; Sivanantharajah & Percival-Smith, 2015). 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* (Lewis *et al.*, 1980; Struhl, 1982; Panzer *et al.*, 1992; Percival-Smith *et al.*, 1997; Percival-Smith *et al.*, 2013). Besides the highly conserved HD, which is essential for function, there are other small conserved motifs in SCR that are also required for function (Figure 1). 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 (Sivanantharajah & Percival-Smith, 2009; Percival-Smith *et al.*, 2013; Sivanantharajah & Percival-Smith, 2014). The non-uniform effect of *Hox* mutant alleles across tissues, termed differential pleiotropy, suggest that the conserved domains and motifs contribute to HOX activity in a small, additive and tissue-specific manner (Hittinger *et al.*, 2005; Prince *et al.*, 2008; Sivanantharajah & Percival-Smith, 2009; Sivanantharajah & Percival-Smith, 2014; Merabet *et al.*, 2011; Percival-Smith *et al.*, 2013; Sivanantharajah & Percival-Smith, 2015). The short, conserved motifs found in SCR and other HOX proteins may also be Short Linear Motifs (SLiMs) (Sivanantharajah & Percival-Smith, 2015). 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 (Neduva *et al.*, 2005; reviewed in Neduva & Russell, 2006; Davey *et al.*, 2012; Van Roey *et al.*, 2014). SLiMs function as sites of protein-protein interaction, PTMs and cell compartment targeting signals (Puntervoll *et al.*, 2003; Neduva *et al.*, 2005; Davey *et al.*, 2012; Iakoucheva *et al.*, 2004; Khan & Lewis, 2005; Gould *et al.*, 2010; Dinkel *et al.*, 2016).

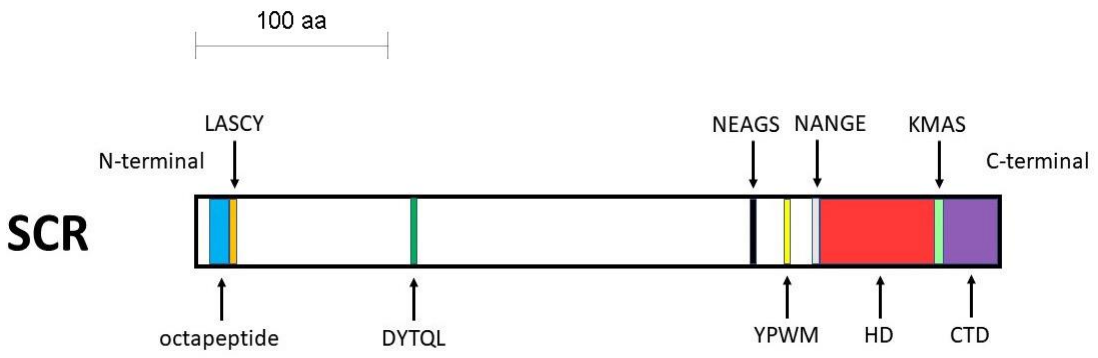

Figure 1. Schematic of SCR protein. The block diagram is drawn to scale. The functional regions of SCR are color-coded. The octapeptide motif is labeled in blue, LASCY motif in orange, DYTQL motif in dark green, NEAGS motif in black, YPWM motif in yellow, NANGE motif in grey, HD in red, KMAS motif in light green and CTD in purple.

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 (Mylin *et al.*, 1989; Hunter & Karin, 1992; Ardito *et al.*, 2017). HOX are phosphoproteins (Berry & Gehring, 2000; Jaffe *et al.*, 1997; Gavis & Hogness, 1991; Stultz *et al.*, 2006; Krause *et al.*, 1988; Krause & Gehring, 1989; Bourbon *et al.*, 1995; Driever & Nüsslein-Volhard, 1989; Gay *et al.*, 1988; Ronchi *et al.*, 1993; Dong *et al.*, 1998; Janody *et al.*, 2000). 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* (Berry & Gehring, 2000). However, a null *PP2A-B'* allele does not affect SCR activity suggesting that dephosphorylation by PP2A-B' plays no role in regulating SCR activity (Moazzen *et al.*, 2009). 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 (Berry & Gehring, 2000; Jaffe *et al.*, 1997; Gavis & Hogness, 1991; Stultz *et al.*, 2006; Krause *et al.*, 1988; Krause & Gehring, 1989; Bourbon *et al.*, 1995; Driever & Nüsslein-Volhard, 1989; Gay *et al.*, 1988; Ronchi *et al.*, 1993; Dong *et al.*, 1998; Janody *et al.*, 2000). We have used Tandem Mass Spectrometry (MS/MS) on SCR extracted from developing embryos to provide data about a possible repertoire of PTMs (Johnson & Eyers, 2010).

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) (Tiefenbach *et al.*, 2010) was isolated from a pUAST construct (Percival-Smith *et al.*, 2013) and inserted into pET-3a and pCaSpeR (Studier & Moffatt, 1986; Studier *et al.*, 1990; Thummel & Pirrotta, 1992) 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 (Rubin and Spradling, 1982).

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 (Brand & Perrimon, 1993). For ectopic expression of SCRTT protein from the heat-shock promoter *ScrTT* fusion gene, a heat-shock was administered at 5 hours AEL at 37.5°C for 30 minutes. First instar cuticles were prepared as described (Wieschaus & Nüsslein-Volhard, 1986) 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 (Brand & Perrimon, 1993), 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 of a filter basket and 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 (Loughran & Walls, 2011; Haneskog, 2006). 3g of heat-shocked embryos was homogenized in 15ml of lysis buffer (15mM HEPES pH 7.6, 10mM KCl, 5mM $MgCl_2$, 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 NaH_2PO_4, 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 0-5ºC. 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 NaH_2PO_4, 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 (25µl) 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; Sambrook *et al.*, 1989) and was heated to 90°C for 10 minutes. The 50µl 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 Coomasie blue (Coomasie 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® SpotPicker™ 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.; Zhang *et al.*, 2012) 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 MS^2 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 (Humphrey *et al.*, 2015). The TiO_2 flow-through containing potential non-phosphopeptides were desalted using a C18-StageTip (Rappsilber *et al.*, 2003) prior to MS/MS analysis. The C18 StageTip was solvated thrice with 200µl of 80% acetonitrile, 0.2% formic acid and 19.8% water followed by centrifugation at 3000g for 2 mins at room temperature or until no liquid remained in the tip. The C18 StageTip was equilibrated thrice with 200µl 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 100µl 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 200µl 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 100µl 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 18µl 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 (Studier & Moffatt, 1986) and purified using Ni-NTA chromatography under denaturing conditions (Hochuli *et al.*, 1987; Hochuli *et al.*, 1988). Protein concentration was determined by the Bradford protein assay (Bradford, 1976).

Bioinformatic analysis of proteomic data

D. melanogaster SCR protein sequence (NCBI accession number in Table S8) was submitted to the ELM database (Dinkel *et al.*, 2016) to retrieve predicted nuclear and cytoplasmic short linear motif (SLiM) sequences as HOX transcription factors interact with nuclear and cytoplasmic components (Merabet & Dard, 2014; Wiellette *et al.*, 1999). In addition, SLiMs with any amino acids from known ordered regions of SCR were excluded from the analysis (Billeter *et al.*, 1990; Joshi *et al.*, 2007). 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*; Sterck *et al.*, 2012) database (accession numbers in Table S8) and a multiple sequence alignment was performed using the tools, MAFFT version 7 (Katoh *et al.*, 2017) and Clustal Omega (Sievers *et al.*, 2011). Each SLiM of SCR was manually checked for conservation across species (Figure S9). A SLiM was considered to be conserved only if they aligned perfectly in both MAFTT and Clustal Omega irrespective of the

maximum length of the SLiM. SLiMs less than five amino acids long were not considered as a conserved SLiM unless conserved beyond Diptera. The minimum length of a SLiM conserved beyond Diptera was 4 amino acids.

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 (Fisher, 1922).

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 3X FLAG, Strep II and 6X His tags (Tiefenbach *et al.*, 2010; Percival-Smith *et al.*, 2013) and cloned behind the *UAS* promoter of pUAST (Percival-Smith *et al.*, 2013), the heat-shock promoter (*hsp*) of pCaSpeR (Thummel & Pirrotta, 1992) and the T7 promoter of pET-3a (Studier & Moffatt, 1986; Studier *et al.*, 1990). Two major systems for ectopic expression of proteins in *Drosophila* are the heat-shock inducible promoter and the GAL4-*UAS* binary system. The expression of SCRTT using these two ectopic expression systems were compared (Figure 2). The heat-shock promoter resulted in higher levels of accumulation of SCRTT (Figure 2). 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 (Artimo *et al.*, 2012). However, on a Western Blot, the SCRTT protein expressed during embryogenesis ran with a higher M_r of 62 (Figure 2).

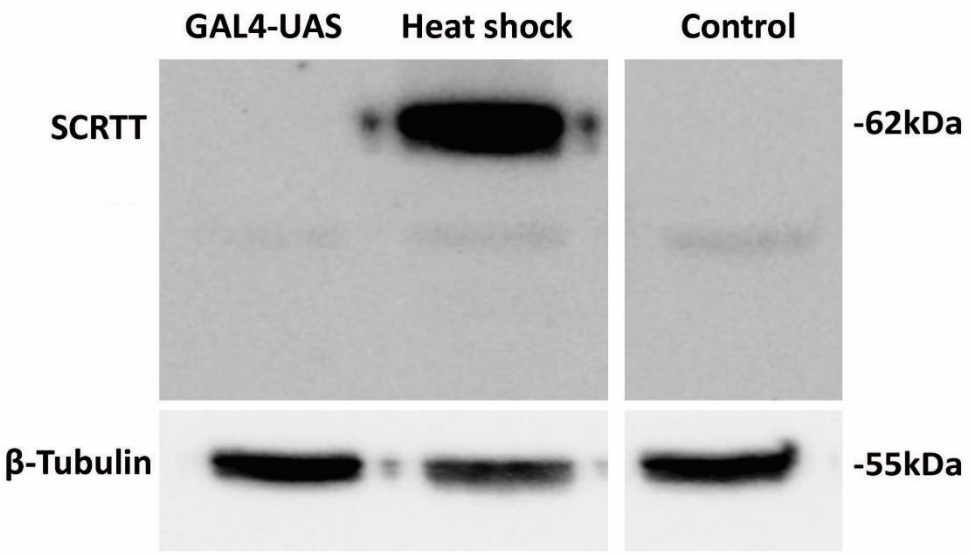

Figure 2. Comparison of the expression of SCRTT protein from heat-shock and *UAS* promoters. The method of expression is indicated above the lanes of the Western Blot of protein extracted from embryos. The position of SCRTT is indicated on the left, and the positions of molecular weight markers are indicated on the right. The level of β-tubulin expression is shown at the bottom. *y w* is the untransformed control.

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 (Brand & Perrimon, 1993). 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* (Figure 3A & B, Gibson *et al.*, 1990; Percival-Smith *et al.*, 2013; Zhao *et al.*, 1993).

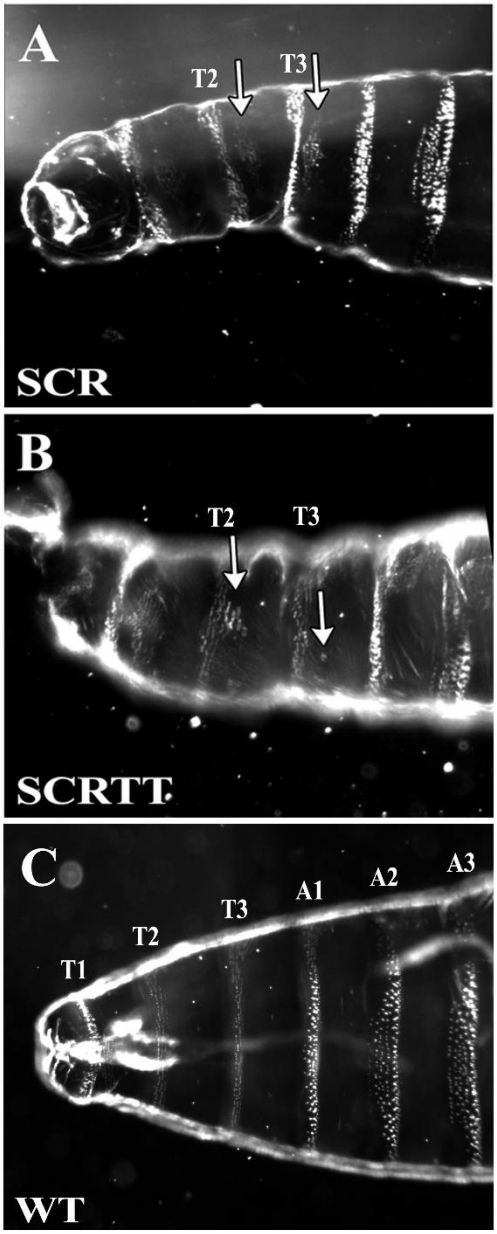

Figure 3. First instar larval cuticular phenotypes of embryonic ectopic expression of SCR and SCRTT proteins. The anterior half of the larva is shown. The untagged SCR protein expressed with the GAL4-*UAS* system using a ubiquitous *armadillo*-GAL4 is shown in panel A; whereas, the SCRTT protein expressed from a heat-shock promoter is shown in panel B. T1, T2 and T3 refer to first, second and third thoracic segments. A1, A2 and A3 refer to first, second and third abdominal segments. A & B. Ectopic expression of SCR and SCRTT, respectively (T2 and T3 beards marked with arrows). C. Control wild-type (WT) first instar larval cuticle.

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

For affinity purification of SCRTT, 3g of heat-shocked embryos containing the *hspScrTT* fusion gene were collected between 0 and 16h AEL and 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 (Hochuli *et al.*, 1987; Hochuli *et al.*, 1988). The purification of SCRTT was monitored by Western Blot analysis and shows concentration of SCRTT on the Ni-NTA beads (Figure 4B). 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 (Figure 4C). To determine whether this purification provided the amount of SCRTT required for MS/MS, a sample of the Ni beads containing purified SCRTT was run alongside a sample of 3500ng of SCRTT purified from bacteria (Figure 4D). The signal for the SCRTT purified from *Drosophila* embryos was 2.4-fold less suggesting the band contained about 1500ng of SCRTT. For the MS/MS analysis, approximately 10µg of SCRTT extracted from *Drosophila* embryos was analyzed per sample.

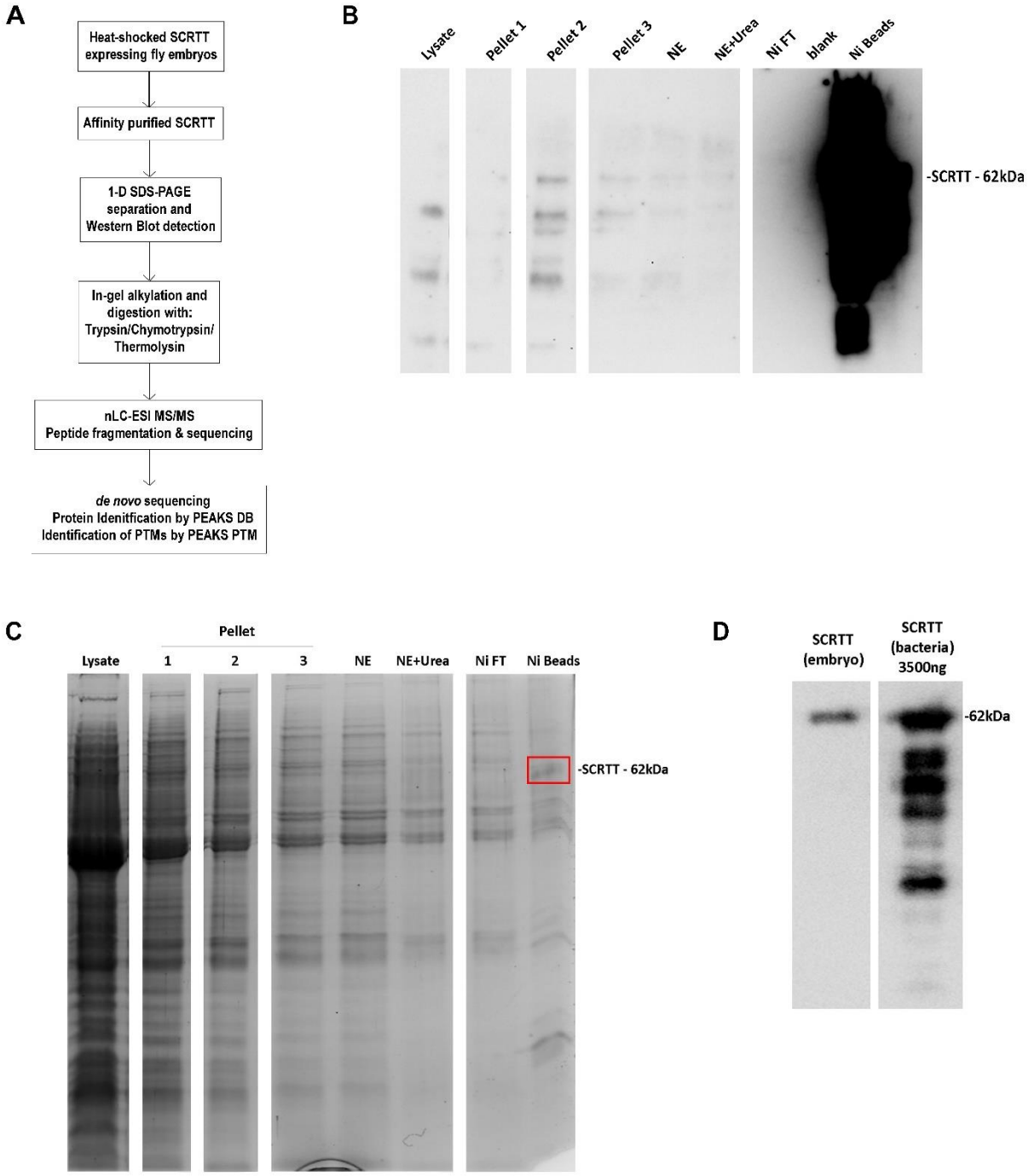
A
Heat-shocked SCRTT expressing fly embryos
Affinity purified SCRTT
1-D SDS-PAGE separation and Western Blot detection
In-gel alkylation and digestion with: Trypsin/Chymotrypsin/ Thermolysin
nLC-ESI MS/MS Peptide fragmentation & sequencing
de novo sequencing
Protein Idenitfication by PEAKS DB
Identification of PTMs by PEAKS PTM
B
Lysate
Pellet 1
Pellet 2
Pellet 3
NE
NE+Urea
Ni FT
blank
Ni Beads
-SCRTT - 62kDa
C
Pellet
Lysate
1
2
3
NE
NE+Urea
Ni FT
Ni Beads
-SCRTT - 62kDa
D
SCRTT (embryo)
SCRTT (bacteria) 3500ng
-62kDa

Figure 4. Overall approach for PTM mapping of SCRTT. A. Analytical workflow for affinity purification, digestion and sequence mapping of PTMs in SCRTT expressed from developing embryos. B. Overexposed Western Blot showing SCRTT at 62 kDa in the Ni beads fraction. Purification fraction is indicated on top of the lane. C. Coomasie-stained 1-D SDS-polyacrylamide gel of affinity purified SCRTT. Purification fraction is indicated on top of the lane. A band of SCRTT at 62 kDa (marked with a red box) was observed in the Ni beads fraction. D. Comparison of *Drosophila* SCRTT vs. bacterial SCRTT to estimate the amount of protein to be analyzed by MS/MS. The Western Blot shows signals for SCRTT at 62 kDa indicated on the right. The source of SCRTT is at the top of the lane. NE Nuclear Extract, Ni Nickel and Ni FT Nickel Flow-Through. For each of the panels B, C and D, all lanes shown are from the same gel and the white space between lanes indicate the splicing out of irrelevant lanes.

Eight samples were analyzed with MS/MS of which four were digested with trypsin, two with chymotrypsin and two with thermolysin. Figure S1 shows the distribution of chemically distinct peptides over the primary sequence of SCRTT for each sample. Trypsin, chymotrypsin and thermolysin were chosen for digestion of SCRTT based on the analysis of predicted peptide generated by these enzymes (Artimo *et al.*, 2012). 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 SCRTT.

Table 1. MS/MS analysis – Coverage and AD_{orc} for SCRTT

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*
Chymotrypsin	5	80	8.7
	6	76	8.3
	5+6	82	12.51*
Thermolysin	7	67	7.22
	8	60	5.92
	7+8	71	9.48*
Trypsin+Chymotrypsin+Thermolysin	-	96	43.29*

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

Evidence for the PTMs of SCRTT

Each spectrum obtained by LC-MS/MS was interrogated by PEAKS DB search followed by identification of PTMs using PEAKS PTM algorithm. Embryonic SCRTT is post-translationally modified (Figure S1). 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 Figure S1). These 44 modifications are phosphorylation, acetylation, formylation, methylation, carboxylation or hydroxylation (Figures S3-8). 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 criteria was assessed: whether b and/or y ions with a neutral loss of phosphoric acid (98 Da) were present in the MS^2 spectra.

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 MS^2 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 Figure S1A but was identified in the analysis of tryptic peptides (A2).

Phosphorylation of SCRTT is substoichiometric

The percentage of phosphorylated peptides is low (Table 2) indicating that phosphorylation of SCRTT 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 (Larsen *et al.*, 2005) was analyzed. 12 phosphosites were detected in α-casein (Table S2; Figure S2). Multiple 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 (Figure 5). This suggests that phosphorylation is stable during MS/MS analysis.

TiO_2 beads were used to enrich for phosphopeptides from trypsinized α-casein and SCRTT samples (Humphrey *et al.*, 2015). For MS/MS analysis of α-casein, 92.3% of all peptides (84 out of 91) identified post-TiO_2 treatment were phosphorylated whereas without TiO_2 treatment, only 18.8% of all peptides (158 of 839) were phosphorylated showing that TiO_2 enriches for phosphopeptides (Table S3; Figure S2). However, only 54 phosphopeptides were common for both treatments. 104 phosphopeptides previously identified without TiO_2 treatment could not be identified post-TiO_2 treatment which indicates loss of phosphopeptides (Table S3). Although TiO_2 enriches for phosphopeptides, the yield is low and no phosphopeptides were detected upon TiO_2 enrichment of trypsinized SCRTT.

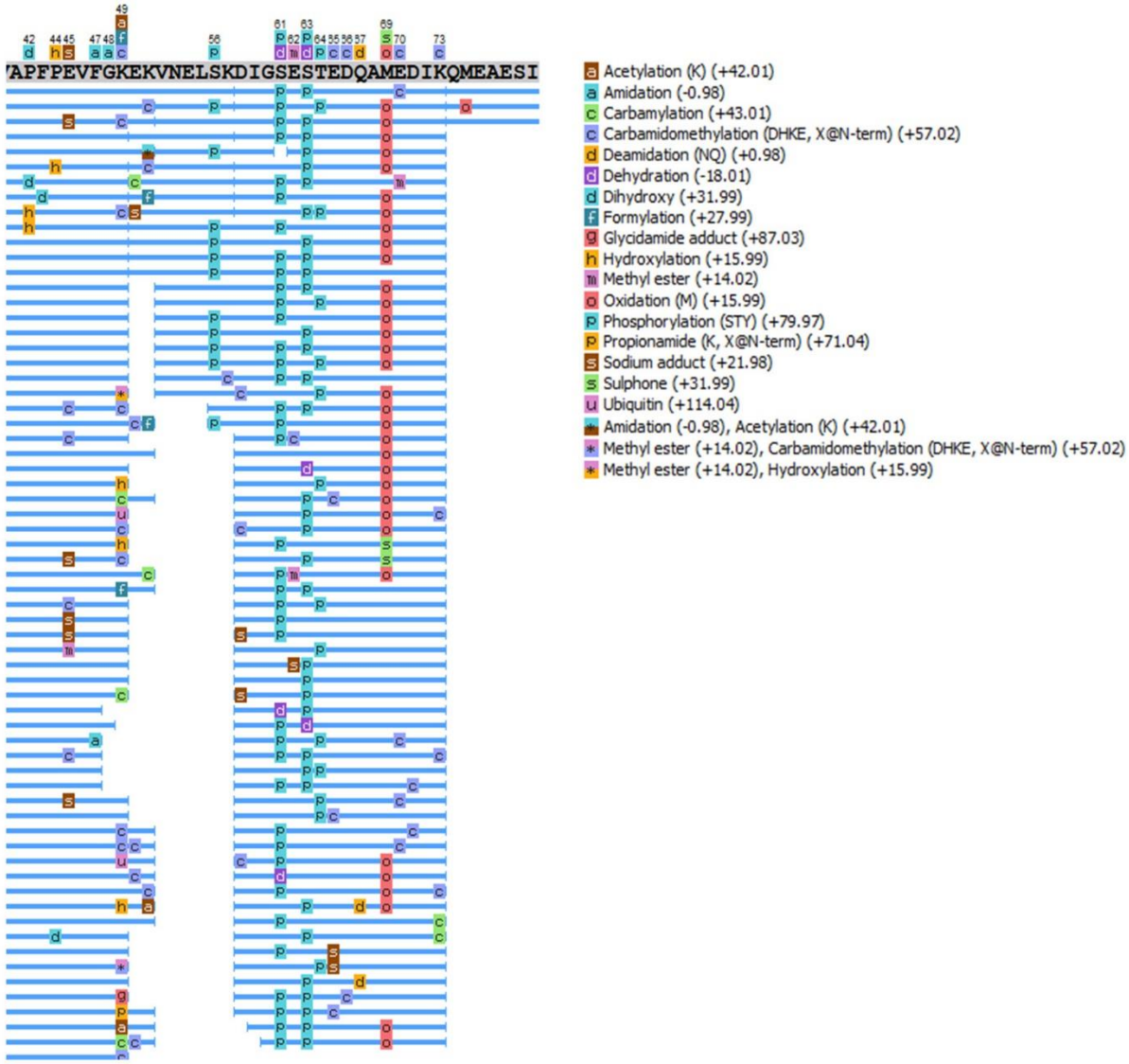

Figure 5. Identification of phosphopeptides of a commercially purchased, pure phosphoprotein, α-casein. The figure shows a region of 41-80 of α-casein (full protein in Figure S2A) and each blue line underneath the primary protein sequence represents a chemically distinct peptide identified by MS/MS analysis. 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.

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

Outside the HD, HOX proteins are highly disordered proteins (reviewed in Merabet & Dard, 2014). SLiMs present in disordered protein regions are proposed to be preferential sites of phosphorylation (Diella *et al.*, 2008; Dinkel *et al.*, 2012; Sivanantharajah & Percival-Smith 2015). To test this with SCRTT, the Eukaryotic Linear Motif (ELM) resource (Dinkel *et al.*, 2016) was screened for predicted SLiMs. Some SLiMs also correspond with regions of SCR conservation and are referred to as 'conserved SLiMs'. In SCR, 66% of the primary sequence was SLiM sequence and 35% were conserved SLiM sequence (Table 3; Figure 6). 6 out of 7 phosphosites were in a SLiM region and 1 of 7 phosphosites was in a conserved SLiM (Figure 6). 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 51 conserved SLiMs of SCR were phosphorylated (Tables S6 & S7) indicating that about 10% of predicted SLiMs 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).

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	239	66	51	129	35

Frequency of phosphorylation in SCR SLiMs vs. non-SLiMs					
Percentage of S, T and Y phosphorylated				*p*-value	
Total SLiMs		Conserved SLiMs			
SLiMs	Non-SLiMs	SLiMs	Non-SLiMs	Total SLiMs	Conserved SLiMs
10	17	3	20	0.5	0.04

Biased distribution of S, T and Y to SCR SLiMs					
Percentage of amino acid that are S, T and Y				*p*-value	
Total SLiMs		Conserved SLiMs			
SLiMs	Non-SLiMs	SLiMs	Non-SLiMs	Total SLiMs	Conserved SLiMs
25	5	28	13	3.4×10^{-7}	5.7×10^{-4}

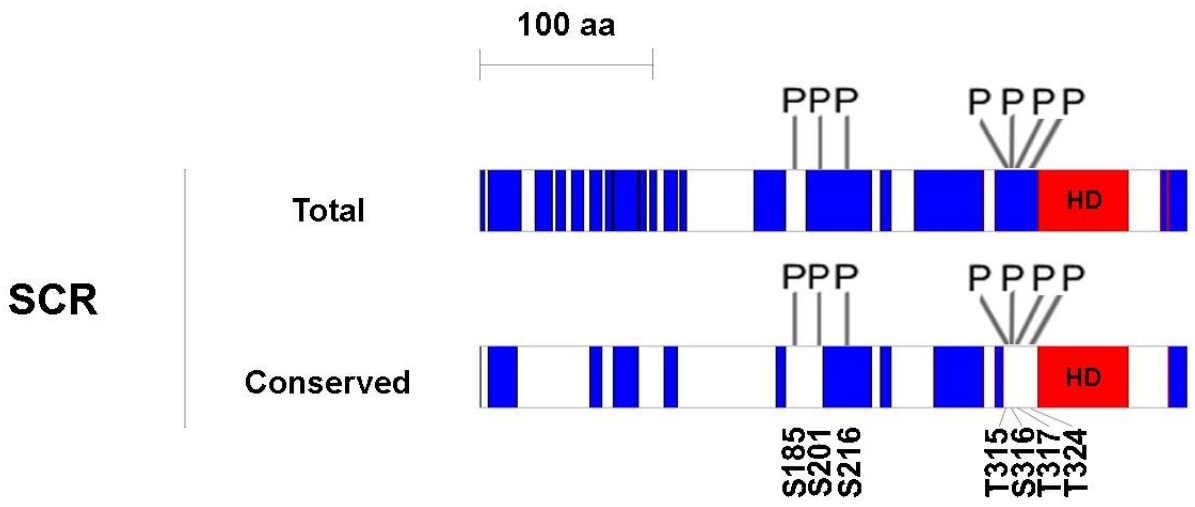

Figure 6. Schematic of SCR showing total and conserved SLiMs. The schematics are drawn to scale. The ordered region of the homeodomain is labeled by red and SLiMs are labeled blue. The SLiM data used to construct the figure are in the supplement (Tables S6 & S7). The phosphosites in SCR are marked to scale.

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 (Figure 7). 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 (O'Connell *et al.*, 2015). 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 (Figure 1; Sivanantharajah & Percival-Smith, 2009; Percival-Smith *et al.*, 2013; Sivanantharajah & Percival-Smith, 2014; Sivanantharajah, 2013), all are post-translationally modified with the exception of the octapeptide and KMAS motifs.

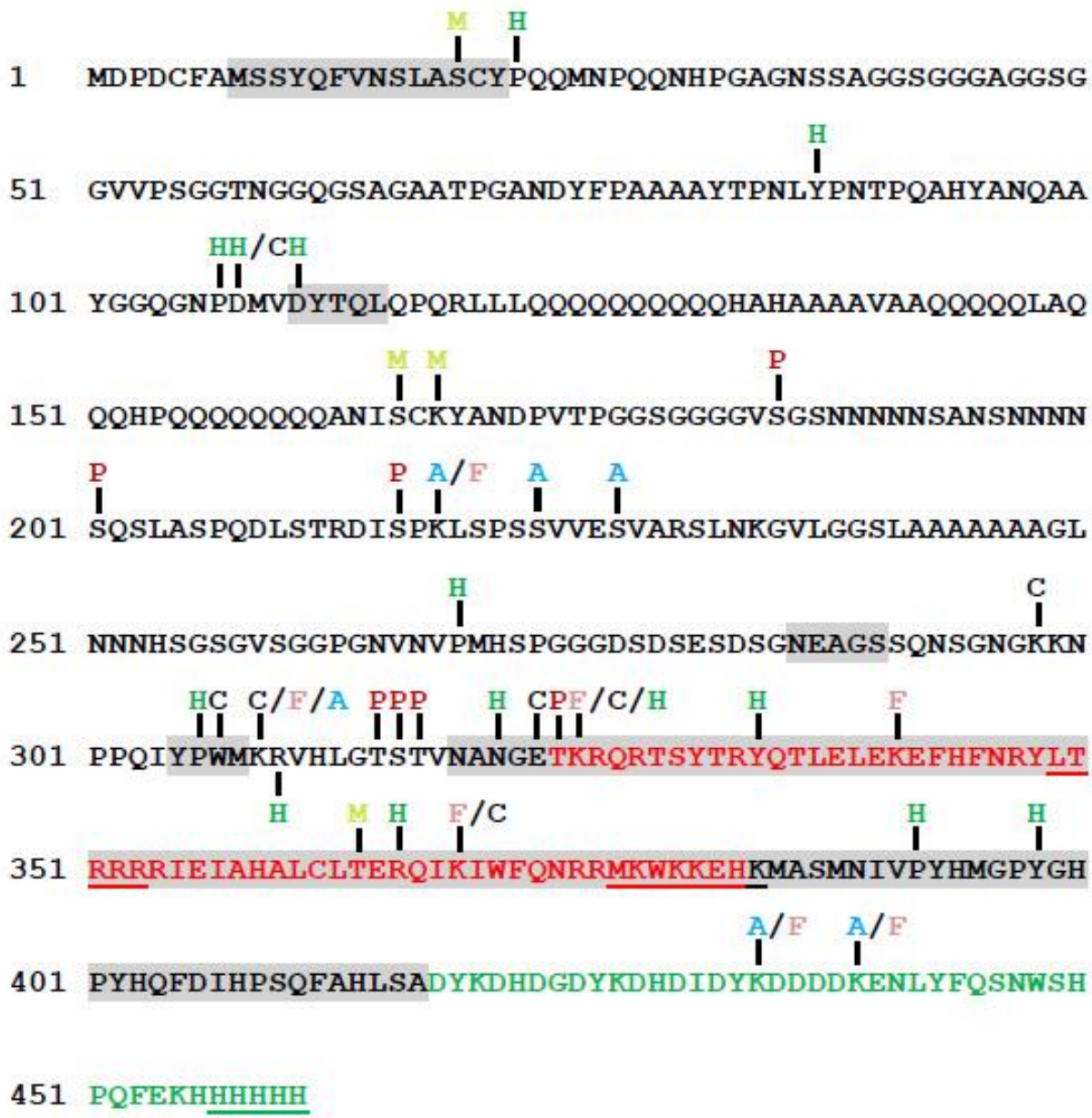

Figure 7. A summary map of post-translational modifications in SCRTT identified by MS/MS. Phosphorylation sites are indicated by "**P**", acetylation by "**A**", formylation by "**F**", methylation by "**M**", carboxylation by "**C**" and hydroxylation by "H". The homeodomain is highlighted in red and the triple tag is highlighted in green. The functional regions of SCR are shaded in grey (Figure 1). The amino acids underlined were not detected by MS/MS. Phosphorylation at S216 was identified by Zhai *et al.*, 2008.

Phosphorylation of SCRTT

A clustered set of phosphorylations on the amino acid residues, T315, S316, T317 and T324 flank the NANGE motif (Figure 7). 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 (Percival-Smith *et al.*, 2013). The phosphosites map to a region of SCR not ordered in the SCR-EXD complex binding to *fkh* DNA (Figure 7). 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 (Joshi *et al.*, 2007; Rohs *et al.*, 2009a; Rohs *et al.*, 2009b; Rohs *et al.*, 2010; Abe *et al.*, 2015). 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 might interfere with transient minor groove interactions.

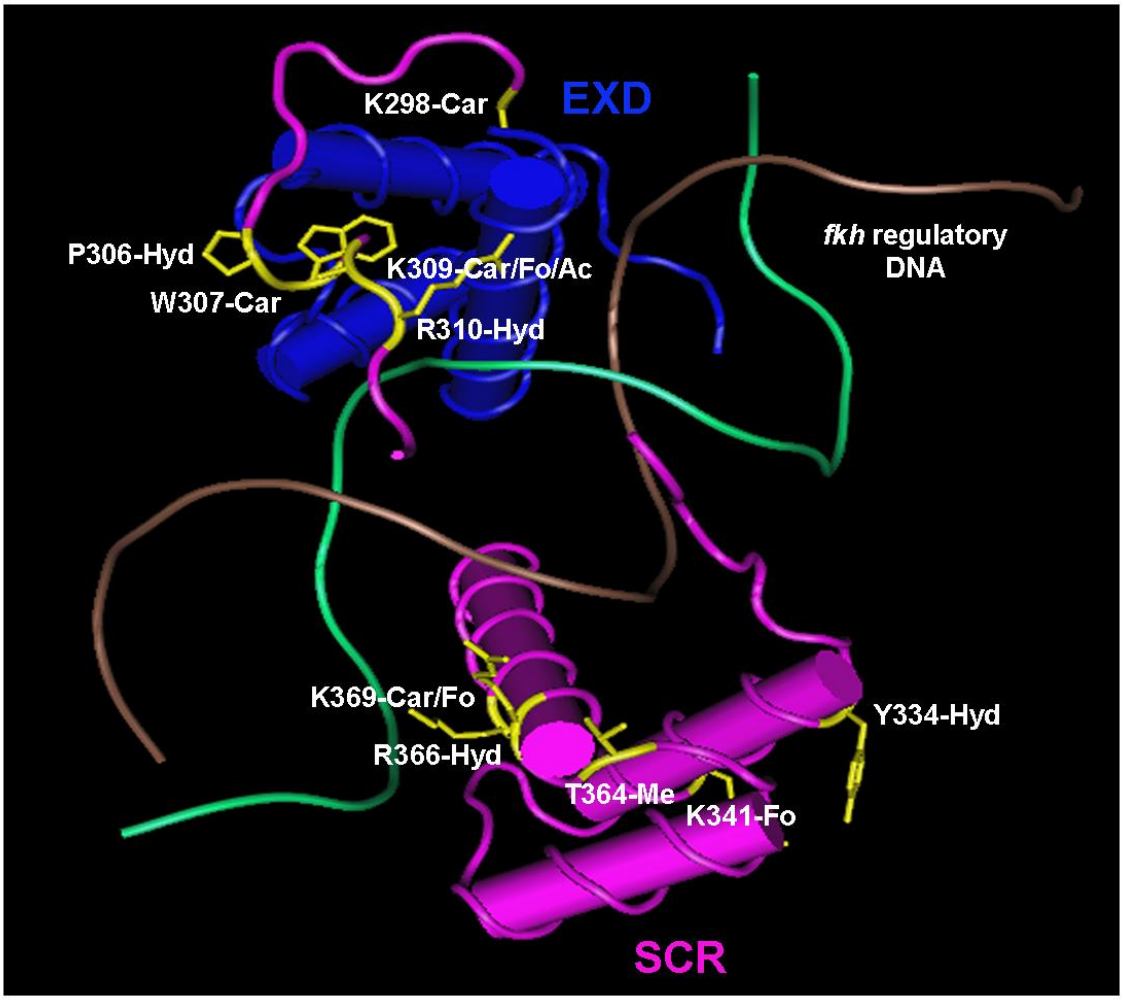

Figure 8. The structure of SCR-EXD-DNA complex determined by crystallography. SCR is shown in pink and EXD in blue. The two strands of *fkh* regulatory DNA is shown in brown and green. The modified amino acids of SCR along with their side chains are shown in yellow. Ac – acetylation, Car – carboxylation, Fo – formylation, Hyd – hydroxylation and Me – methylation. The structure coordinates with accession code 2R5Z (*fkh*250) were retrieved from RCSB Protein Data Bank (Joshi *et al.*, 2007). Cn3D 4.3.1 (NCBI) was used to annotate the 3-D structure.

In the model for regulation of SCR activity proposed by Berry & Gehring, 2000, 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 (Moazzen *et al.*, 2009). 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.

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 (Figure 8). The Bilateran-specific YPWM motif of SCR and UBX binds to a hydrophobic pocket on the surface of EXD HD (Passner *et al.*, 1999; Joshi *et al.*, 2007). The MS/MS analysis identified hydroxylation at P306 and carboxylation at W307 of 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 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 (Ryoo & Mann, 1999).

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 (Joshi *et al.*, 2007; Rohs *et al.*, 2009b). 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 (Joshi *et al.*, 2007). Carboxylation at K309 and hydroxylation at R310 adds negative charges to this region of the protein which might be a mechanism of inhibition of 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 (Otting *et al.*, 1990; Gehring *et al.*, 1994; Joshi *et al.*, 2007; Religa *et al.*, 2007). 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 (Jiang *et al.*, 2007).

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 (Wiśniewski *et al.*, 2008).

In histones, lysine acetylation by histone acetyltransferases (HAT) and lysine deacetylation by histone deacetylases (HDAC), respectively are involved in chromatin remodeling and gene expression (Allfrey *et al.*, 1964; reviewed in Roth *et al.*, 2001). Acetylation has also been reported to modify the activity of TFs, thereby regulating the ability of the TF to bind DNA (Gu & Roeder, 1997; reviewed in Kouzarides, 2000; Bannister & Miska, 2000). Besides histones, HATs and HDACs acetylate and deacetylate non-histone proteins respectively, including TFs

which may be a mechanism regulating SCR activity (reviewed in Park *et al.*, 2015; Wang *et al.*, 2011; Glozak *et al.*, 2005; Choudhary *et al.*, 2014).

Formylation of lysine residues is widespread in histones and other nuclear proteins and arise as a secondary modification due to oxidative DNA damage (Jiang *et al.*, 2007). 5'-oxidation of DNA deoxyribose results in the formation of a highly reactive 3-formylphosphate residue which outcompetes the acetylation mechanism and formylates the side-chain amino group of lysine (Jiang *et al.*, 2007). 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 (Sivanantharajah & Percival-Smith, 2015). 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 into SLiMs. In addition, a minority of predicted SLiMs were phosphosites. This lack of phosphorylation may reflect that either the ELM database overpredicts SLiMs, or many SLiMs are targets of other PTMs (Puntervoll *et al.*, 2003; Iakoucheva *et al.*, 2004; Khan & Lewis, 2005; Gould *et al.*, 2010; Dinkel *et al.*, 2016) like methylation, hydroxylation, carboxylation, acetylation and formylation. Examples of other PTMs in SLiM regions of SCR include a methylation at S19 and a hydroxylation at P22 of the *Drosophila*-specific SLiM, SLASCYP and methylation at S166 and K168 of the SLiM, ANISCK (Table S6).

Potential limitations

In order to detect PTMs by MS/MS analysis, a high concentration of the analyte protein is required which requires 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 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 (Krause *et al.*, 1988; Krause & Gehring, 1989; Gavis & Hogness, 1991) 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 (Steen *et al.*, 2006). In addition, multiply phosphorylated peptides were detected upon MS/MS analysis of a commercially purchased, pure, heavily phosphorylated, bovine α-casein protein (Larsen *et al.*, 2005) 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 might 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 (Hu *et al.*, 2019; Zhai *et al.*, 2008).

Conclusion

This study identified sites of phosphorylation and other PTMs in a tagged HOX protein, SCRTT, extracted from developing *Drosophila melanogaster* embryos. These modifications map to functionally important regions of SCR. In testing the hypothesis that HOX 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.

Acknowledgements

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 *ScrTT* construct, Ben Rubin for his suggestions on the statistical analysis, Gurjit Randhawa and Stuart Cameron for assistance with the project. We extend our gratitude to René Rezsohazy for valuable suggestions on the manuscript. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant to AP-S.

References

Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H. J., … Mann, R. S. (2015). Deconvolving the recognition of DNA shape from sequence. *Cell*, *161*(2), 307-318.

Akam, M. (1998). *Hox* genes, homeosis and the evolution of segmental identity: no need for hopeless monsters. *The International Journal of Developmental Biology*, *42*(3), 445-451.

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(5), 786-794.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., & Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine*, *40*(2), 271-280.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., … Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(Web Server issue), W597-W603.

Bannister, A. J., & Miska, E. A. (2000). Regulation of gene expression by transcription factor acetylation. *Cellular and Molecular Life Sciences*, *57*(8-9), 1184-1192.

Berry, M., & Gehring, W. (2000). Phosphorylation status of the SCR homeodomain determines its functional activity: essential role for protein phosphatase 2A,B'. *The EMBO Journal*, *19*(12), 2946–2957.

Billeter, M., Qian, Y., Otting, G., Müller, M, Gehring, W. J., & Wüthrich, K. (1990). Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by ^{1}H nuclear magnetic resonance spetroscopy. *Journal of Molecular Biology*, *214*(1),183-197.

Bourbon, H. M., Martin-Blanco, E., Rosen, D., & Kornberg, T. B. (1995). Phosphorylation of the *Drosophila* engrailed protein at a site outside its homeodomain enhances DNA binding. *Journal of Biological Chemistry*, *270*(19), 11130–11139.

Bradford, M. M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248-254.

Brand, A. H., & Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, *118*(2), 401–415.

Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., & Mann, M. (2014). The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature Reviews Molecular Cell Biology*, *15*(8), 536-550.

Davey, N. E., Van Roey, K., Weatheritt, R. J., Toedt, G., Uyar, B., Altenberg, B., … Gibson, T. J. (2012). Attributes of short linear motifs. *Molecular BioSystems, 8*(1), 268-281.

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., … Gibson, T. J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Frontiers in Bioscience, 13*, 6580-603.

Dinkel, H., Michael, S., Weatheritt, R. J., Davey, N. E., Van Roey, K., Altenberg, B., … Gibson, T. J. (2012). ELM--the database of eukaryotic linear motifs. *Nucleic Acids Research, 40*(Database issue), D242-51.

Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., … Gibson, T. J. (2016). ELM 2016--data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Research, 44*(D1), D294-D300.

Dong, J., Hung, L. H., Strome, R., & Krause, H. M. (1998). A phosphorylation site in the Ftz homeodomain is required for activity. *The EMBO Journal, 17*(8), 2308–2318.

Draime, A., Bridoux, L., Graba, Y., & Rezsohazy, R. (2018). Post-translational modifications of HOX proteins, an underestimated issue. *The International Journal of Developmental Biology, 62*(11-12), 733-744.

Driever, W., & Nüsslein-Volhard, C. (1989). The bicoid protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature, 337*(6203), 138–143.

Fisher, R. A. (1922). On the interpretation of χ2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society, 85*, 87-94.

Gavis, E. R., & Hogness, D. S. (1991). Phosphorylation, expression and function of the *Ultrabithorax* protein family in *Drosophila melanogaster*. *Development, 112*(4), 1077–1093.

Gay, N. J., Poole, S. J., & Kornberg, T. B. (1988). The *Drosophila engrailed* protein is phosphorylated by a serine-specific protein kinase. *Nucleic Acids Research, 16*(14A), 6637–6647.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., … Wüthrich, K. (1994). Homeodomain-DNA recognition. *Cell, 78*(2), 211-223.

Gibson, G., Schier, A., LeMotte, P., & Gehring, W. J. (1990). The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. *Cell, 62*(6), 1087-1103.

Glozak, M. A., Sengupta, N., Zhang, X., & Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. *Gene, 363*, 15-23.

Gould, C. M., Diella, F., Via, A., Puntervoll, P., Gemünd, C., Chabanis-Davidson, S., … Gibson, T. J. (2010). ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Research*, *38*(Database issue), D167-D180.

Gu, W., & Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, *90*(4), 595-606.

Haneskog, L. (2006). Purification of Histidine-Tagged Proteins under Denaturing Conditions Using IMAC. *Cold Spring Harbor Protocols*, *2006*(1).

Hittinger, C. T., Stern, D. L., Carroll, S. B. (2005). Pleiotropic functions of a conserved insect-specific Hox peptide motif. *Development*, *132*(23), 5261-5270.

Hochuli, E., Döbeli, H., & Schacher, A. (1987). New metal chelate adsorbent for proteins and peptides containing neighbouring histidine residues. *Journal of Chromatography*, *411*, 177–184.

Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., & Stüber, D. (1988). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Nature Biotechnology*, *6*, 1321–1325.

Hu, Y., Sopko, R., Chung, V., Foos, M., Studer, R. A., Landry, S. D., … Perrimon, N. (2019). iProteinDB: An Integrative Database of *Drosophila* Post-translational Modifications. *G3: Genes Genomes Genetics*, *9*(1), 1-11.

Humphrey, S. J., Azimifar, S. B., & Mann, M. (2015). High throughput phosphoproteomics reveals *in vivo* insulin signaling dynamics. *Nature Biotechnology*, *33*(9), 990-995.

Hunter, T., & Karin, M. (1992). The regulation of transcription by phosphorylation. *Cell*, *70*(3), 375-387.

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, *32*(3), 1037-1049.

Jaffe, L., Ryoo, H. D., & Mann, R. S. (1997). A role for phosphorylation by casein kinase II in modulating Antennapedia activity in *Drosophila*. *Genes & Development*, *11*(10), 1327–1340.

Janody, F., Sturny, R., Catala, F., Desplan, C., & Dostatni, N. (2000). Phosphorylation of Bicoid on MAP-kinase sites: contribution to its interaction with the torso pathway. *Development*, *127*(2), 279-289.

Jiang, T., Zhou, X., Taghizadeh, K., Dong, M., & Dedon, P. C. (2007). N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(1), 60-65.

Johnson, H., & Eyers, C. E. (2010). Analysis of Post-translational Modifications by LC-MS/MS. *Methods in Molecular Biology*, *658*, 93-108.

Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., … Mann, R. S. (2007). Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. *Cell*, *131*(3), 530–543.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, bbx108.

Khan, A. N., & Lewis, P. N. (2005). Unstructured conformations are a substrate requirement for the Sir2 family of NAD-dependent protein deacetylases. *The Journal of Biological Chemistry*, *280*(43), 36073-36078.

Kouzarides, T. (2000). Acetylation: a regulatory modification to rival phosphorylation? *The EMBO Journal*, *19*(6), 1176-1179.

Krause, H. M., & Gehring, W. J. (1989). Stage-specific phosphorylation of the *fushi tarazu* protein during *Drosophila* development. *The EMBO Journal*, *8*(4), 1197–1204.

Krause, H. M., Klemenz, R., & Gehring, W. J. (1988). Expression, modification, and localization of the *fushi tarazu* protein in *Drosophila* embryos. *Genes & Development*, *2*(8), 1021-1036.

Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., & Jørgensen, T. J. D. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & Cellular Proteomics*, *4*(7), 873–886.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature*, *276*(5688), 565-570.

Lewis, R. A., Wakimoto, B. T., Denell, R. E., & Kaufman, T. C. (1980). Genetic analysis of the Antennapedia gene complex (ANT-C) and adjacent chromosomal regions of *Drosophila melanogaster*. II. Polytene chromosome segments 84A-84B1,2. *Genetics*, *95*(2), 383-397.

Loughran, S. T., & Walls, D. (2011). Purification of Poly-Histidine-Tagged Proteins. *Methods in Molecular Biology*, *681*, 311-335.

Merabet, S., & Dard, A. (2014). Tracking context-specific transcription factors regulating Hox activity. *Developmental Dynamics*, *243*(1), 16-23.

Merabet, S., Litim-Mecheri, I., Karlsson, D., Dixit, R., Saadaoui, M., Monier, B., … Graba, Y. (2011). Insights into Hox protein function from a large scale combinatorial analysis of protein domains. *PLoS Genetics*, *7*(10), e1002302.

Moazzen, H., Rosenfeld, R., & Percival-Smith, A. (2009). Non-requirement of a regulatory subunit of Protein Phosphatase 2A, PP2A-B', for activation of Sex comb reduced activity in *Drosophila melanogaster*. *Mechanisms of Development*, *126*(8-9), 605–610.

Mylin, L. M., Bhat, J. P., & Hopper, J. E. (1989). Regulated phosphorylation and dephosphorylation of GAL4, a transcriptional activator. *Genes & Development*, *3*(8), 1157–1165.

Neduva, V., Linding, R., Su-Angrand, I., Stark, A., de Masi, F., Gibson, T. J., …Russell, R. B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biology*, *3*(12), e405.

Neduva, V., & Russell, R. B. (2006). Peptides mediating interaction networks: new leads at last. *Current Opinion in Biotechnology*, *17*(5), 465-471.

O'Connell, N. E., Lelli, K., Mann, R. S., & Palmer, A. G. 3rd (2015). Asparagine deamidation reduces DNA-binding affinity of the *Drosophila melanogaster* Scr homeodomain. *FEBS Letters*, *589*(21), 3237-3241.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J., & Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *The EMBO Journal*, *9*(10), 3085-3092.

Panzer, S., Weigel, D., & Beckendorf, S. K. (1992). Organogenesis in *Drosophila melanogaster*: embryonic salivary gland determination is controlled by homeotic and dorsoventral patterning genes. *Development*, *114*(1), 49–57.

Park, J. M., Jo, S. H., Kim, M. Y., Kim, T. H., & Ahn, Y. H. (2015). Role of transcription factor acetylation in the regulation of metabolic homeostasis. *Protein Cell*, *6*(11), 804-813.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., & Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature*, *397*(6721), 714-719.

Percival-Smith, A., Weber, J., Gilfoyle, E., & Wilson, P. (1997) 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 melanogsater*. *Development*, *124*(24), 5049-5062.

Percival-Smith, A., Sivanantharajah, L., Pelling, J. J. H., & Teft, W. A. (2013). Developmental competence and the induction of ectopic proboscises in *Drosophila melanogaster*. *Development Genes and Evolution*, *223*(6), 375-387.

Primon, M., Hunter, K. D., Pandha, H. S., & Morgan, R. (2019). Kinase Regulation of HOX Transcription Factors. *Cancers*, *11*(4), E508.

Prince, F., Katsuyama, T., Oshima, Y., Plaza, S., Resendez-Perez, D., Berry, M., … Gehring, W. J. (2008). The YPWM motif links Antennapedia to the basal transcriptional machinery. *Development*, *135*(9), 1669–1679.

Puntervoll, P., Linding, R., Gemünd, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S. … Gibson, T. J. (2003). ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Research*, *31*(13), 3625-3630.

Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and go extraction tips for Matrix-Assisted Laser Desorption or Ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical Chemistry*, *75*(3), 663-670.

Religa, T. L., Johnson, C. M., Vu, D. M., Brewer, S. H., Dyer, R. B., & Fersht, A. R. (2007). The helix-turn-helix motif as an ultrafast independently folding domain: The pathway of folding of Engrailed homeodomain. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(22), 9272-9277.

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of Specificity in Protein-DNA Recognition. *Annual Review of Biochemistry*, *79*, 233-269.

Rohs, R., West, S. M., Liu, P., & Honig, B. (2009a). Nuance in the double-helix and its role in protein-DNA recognition. *Current Opinion in Structural Biology*, *19*(2), 171-177.

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., & Honig, B. (2009b). The role of DNA shape in protein-DNA recognition. *Nature*, *461*(7268), 1248-1253.

Ronchi, E., Treisman, J., Dostatni, N., Struhl, G., & Desplan, C. (1993). Down-regulation of the Drosophila morphogen Bicoid by the torso-receptor mediated signal transduction cascade. *Cell*, *74*(2), 347-355.

Roth, S. Y., Denu, J. M., & Allis, C. D. (2001). Histone acetyltransferases. *Annual Review of Biochemistry*, *70*, 81-120.

Rubin, G. M., & Spradling, A. C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science*, *218*(4570), 348–353.

Ryoo, H. D., & Mann, R. S. (1999). The control of trunk Hox specificity and activity by Extradenticle. *Genes & Development*, *13*(13), 1704–1716.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539.

Sivanantharajah, L., & Percival-Smith, A. (2009). 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*, *182*(1), 191-203.

Sivanantharajah, L., & Percival-Smith, A. (2014). Acquisition of a leucine zipper motif as a mechanism of antimorphy for an allele of the *Drosophila Hox* gene *Sex combs reduced*. *G3: Genes Genomes Genetics*, *4*(5), 829-838.

Sivanantharajah, L., & Percival-Smith, A. (2015). Differential pleiotropy and HOX functional organization. *Developmental Biology*, *398*(1), 1-10.

Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N., & Kirschner, M. W. (2006). Phosphorylation analysis by mass spectrometry. Myths, facts, and the consequences for qualitative and quantitative measurements. *Molecular & Cellular Proteomics*, *5*(1), 172-181.

Sterck, L., Billiau, K., Abeel, T., Rouzé, P., & Van de Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. *Nature Methods*, *9*(11), 1041.

Struhl, G. (1982). Genes controlling segmental specification in the *Drosophila* thorax. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(23), 7380-7384.

Studier, F. W., & Moffatt, B. A. (1986). Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes. *Journal of Molecular Biology*, *189*(1), 113-130.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990). Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods in Enzymology*, *185*, 60-89.

Stultz, B. G., Jackson, D. G., Mortin, M. A, Yang, X., Beachy, P. A, & Hursh, D. A. (2006). Transcriptional activation by *extradenticle* in the *Drosophila* visceral mesoderm. *Developmental Biology*, *290*(2), 482-494.

Thummel, C., & Pirrotta, V. (1992). New pCaSpeR P-element vectors. *Drosophila Information Service*, *71*, 150.

Tiefenbach, J., Moll, P. R., Nelson, M. R., Hu, C., Baev, L., Kislinger, T., & Krause, H. M. (2010). A Live Zebrafish-Based Screening System for Human Nuclear Receptor Ligand and Cofactor Discovery. *PLoS ONE*, *5*, 1-12.

Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., … Davey, N. E. (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chemical Reviews*, *114*(13), 6733-6778.

Wang, C., Tian, L., Popov, V. M., & Pestell, R. G. (2011). Acetylation and nuclear receptor action. *Journal of Steroid Biochemistry and Molecular Biology*, *123*(3-5), 91-100.

Wiellette, E. L., Harding, K. W., Mace, K. A., Ronshaugen, M. R., Wang, F. Y., & McGinnis, W. (1999). spen encodes and RNP motif protein that interacts with Hox pathways to repress the development of head-like sclerites in the Drosophila trunk. *Development*, *126*(23), 5373-5385.

Wieschaus, E., & Nüsslein-Volhard, C. (1986). Looking at embryos, pp. 199-228 in *Drosophila*: A Practical Approach, edited by D. B. ROBERTS. IRL, Oxford.

Wiśniewski, J. R., Zougman, A., & Mann, M. (2008). 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 Research*, *36*(2), 570-577.

Zhai, B., Villén, J., Beausoleil, S. A., Mintseris, J., & Gygi, S. P. (2008). Phosphoproteome analysis of *Drosophila melanogaster* embryos. *Journal of Proteome Research*, *7*(4), 1675-1682.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D, … Ma, B. (2012). PEAKS DB: *De Novo* Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Molecular & Cellular Proteomics*, *11*(4), M111.010587, 1-8.

Zhao, J. J., Lazzarini, R. A., & Pick, L. (1993). The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila Sex combs reduced* gene. *Genes & Development*, *7*(3), 343–354.