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

Anirban Banerjee*, Anthony Percival-Smith

Department of Biology, Western University, 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* body segments along the anterior-posterior (A-P) axis. Besides a highly conserved DNA-binding homeodomain (HD), HOX proteins also contain functionally important, evolutionarily conserved small motifs, which may be Short Linear Motifs (SLiMs). SLiMs are proposed to be sites of phosphorylation and this may regulate the activity of HOX proteins. Sex combs reduced protein fused to a triple tag (SCRTT) was extracted from developing embryos that express SCRTT from a heat-shock promoter fusion gene and purified using Ni-NTA beads under denaturing conditions. Multiple sites of post-translational modification (PTM) were identified in purified SCRTT by Tandem Mass Spectrometry (MS/MS). The identified PTMs 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. We found no support for the hypothesis that predicted SLiMs in SCR are preferential sites of phosphorylation.

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; CID, Collision-Induced Dissociation; 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 Bilaterans 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). The HOX protein, Sex combs reduced (SCR) 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, there are other small motifs in SCR required for function (Figure 1). Amino acid changes associated with *Scr* mutant alleles map to most of the conserved sequences of SCR and exhibit differential pleiotropy, which is the non-uniform effect of mutant alleles across tissues, suggesting that the conserved domains and motifs contribute to SCR 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). In addition, analysis of the induction of ectopic proboscises by Proboscipedia (PB) and SCR, changes of amino acid sequence of highly conserved domains and motifs identified roles for some of the conserved SCR motifs in suppression of proboscis identity (Percival-Smith *et al.*, 2013). The short, conserved motifs found in SCR 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 eukaryotic proteins that lack a native tertiary structure, called 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 including phosphorylation and also as 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).

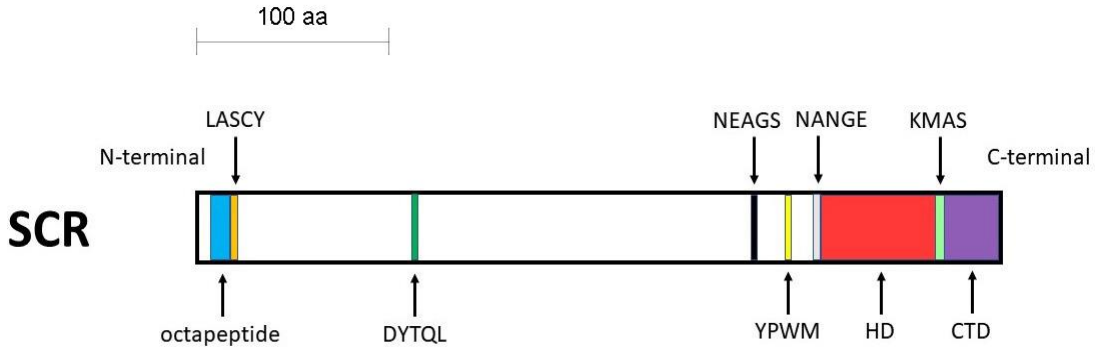

Figure 1. Schematic of SCR protein. The block diagram is drawn to scale. The functional regions of SCR are color-coded. The octapeptide motif is labeled in blue, LASCY motif in orange, DYTQL motif in dark green, NEAGS motif in black, YPWM motif in yellow, NANGE motif in grey, HD in red, KMAS motif in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

Often phosphorylation is a regulatory mechanism in cellular signaling pathways that terminate in the regulation of TF activity and subsequent transcriptional regulation (Mylin *et al.*, 1989; Hunter & Karin, 1992; Ardito *et al.*, 2017). HOX and many other HD-containing proteins including SCR 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). *In vitro* SCR peptides are phosphorylated by cAMP-dependent protein kinase A (PKA) and dephosphorylated by Serine-threonine protein phosphatase 2A (PP2A-B') (Berry & Gehring, 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. However, a null allele in *PP2A-B'* 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 is they provide no information on which amino acid residue of a protein 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). A method that can determine the amino acid phosphorylated in a protein extracted from developing embryos is Tandem Mass Spectrometry (MS/MS) (Johnson & Eyers, 2010). We have cataloged the PTMs of SCR extracted from developing embryos.

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* $w^{67c23.2}$ via *P*-element-mediated transformation (Rubin and Spradling, 1982).

Ectopic expression of SCR and SCRTT for examination of first instar cuticles

To ectopically express the untagged SCR using the GAL4-*UAS* system (Brand & Perrimon, 1993), adult 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. Ectopic expression of SCRTT protein from the heat-shock promoter *ScrTT* fusion gene was induced by a heat-shock administered at 5 hours AEL at 37.5°C for 30 minutes. The embryos were allowed to develop to the first instar larval stage and first instar cuticles were prepared as described (Wieschaus & Nüsslein-Volhard, 1986). The larval cuticles were imaged with darkfield optics on a Leica® Leitz™ DMRBE microscope.

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

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. 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.

Protein separation

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) was added to the embryos post-heat-shock and the embryos were crushed using a homogenizer. The samples were heated to approx. 90°C for 10 minutes prior to loading onto a 0.75mm thick, 11% separating (373mM Tris-HCl pH 8.8, 0.1% SDS) and 5% stacking (124mM Tris-HCl pH 6.8, 0.1% SDS) SDS-Polyacrylamide gel for protein separation.

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 three rinses and three 15 minute 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 three rinses and three 15 minute 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.

Affinity purification of SCRTT protein from embryos

To induce expression of SCRTT proteins from the heat-shock promoter *ScrTT* fusion gene for protein purification, embryos at 0-16 hours AEL were collected from the apple juice plates on nylon mesh screens in 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. Heat-shocked embryos were stored at -80°C. To purify SCRTT protein from embryo extracts, subcellular fractionation followed by metal affinity chromatography in denaturing conditions was employed (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. This 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 supernatant was discarded and the pellet (pellet 2) was resuspended using the 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 pellet was discarded and 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 to allow the 6X His tag to bind to the nickel beads. The mixture was run through, and the beads packed in, a column by gravity flow; the flow-through was reapplied to the column to ensure efficient retention of the tagged protein in the column. The final flow-through (Ni FT) was discarded. 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. The beads were stored at -80°C.

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.). Once the gel was destained, it was stored in 5% glacial acetic acid at 4°C. At the Functional Proteomics Facility, Western University, London, Ontario, Canada, approximately 10 spots were picked from a stained 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. Collision-induced dissociation (CID) was used to fragment peptide ions.

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 units, 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.

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 S6) was submitted to the ELM database (Dinkel *et al.*, 2016) to retrieve predicted short linear motif (SLiM) sequences. Since HOX transcription factors are known to interact with nuclear and cytoplasmic components (Merabet & Dard, 2014; Wiellette *et al.*, 1999), only the SLiMs for proteins localized to nucleus and cytoplasm were considered for the analysis. In addition, only SLiMs that are part of intrinsically disordered regions of the protein were considered for the analysis; therefore, 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 across different phyla, were retrieved from NCBI or ORCAE (only for *T. urticae*; Sterck *et al.*, 2012) database (accession numbers in Table S6) 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 S8). 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 detect and purify the SCR protein, the CDS of the *Scr* gene was fused in frame to CDS of 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). SCR isoform A (417 aa) (FlyBase ID FBpp0081163) was expressed from these constructs. Using the pI/Mw tool of ExPASy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics; Artimo *et al.*, 2012), the relative molecular mass (M_r) of SCRTT protein was predicted to be 49.8. However, on a Western Blot, the SCRTT protein expressed from both the *UAS* and heat-shock promoters ran with a higher M_r of 62 (Figure 2).

The two major systems for inducing ectopic expression of proteins in *Drosophila* are inducible promoters, such as the heat-shock promoter, and binary or two component systems, such as the GAL4-*UAS* system. The expression of SCRTT using the heat-shock promoter and GAL4-*UAS* system with the *armadillo*-GAL4 driver were compared (Figure 2). The heat-shock promoter resulted in higher levels of accumulation of SCRTT (Figure 2). The fold increase of expression of heat-shock relative to *UAS* was too great to be accurately quantified for SCRTT.

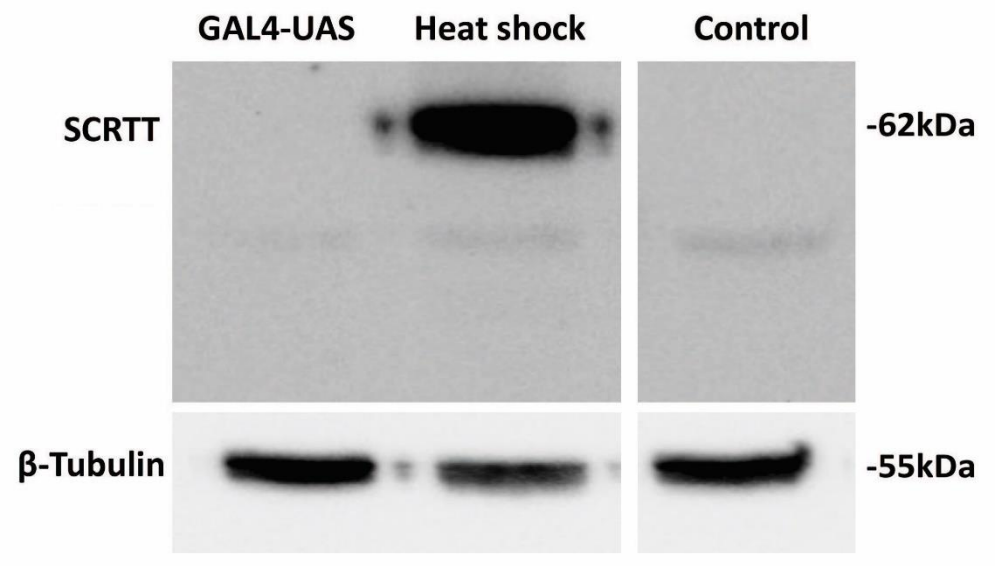

Figure 2. Comparison of the expression of SCRTT protein from heat-shock and *UAS* promoters. The protein extracted from embryos expressing SCRTT from either the heat-shock or *UAS* promoter was analyzed on a Western Blot. The method of expression is indicated above the lanes. The position of SCRTT is 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 was biologically active, it was ectopically expressed in all cells during embryogenesis with a heat-shock, and the first instar larval cuticular phenotype was compared with that of ectopic expression of the untagged SCR protein. The untagged SCR protein was expressed 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).

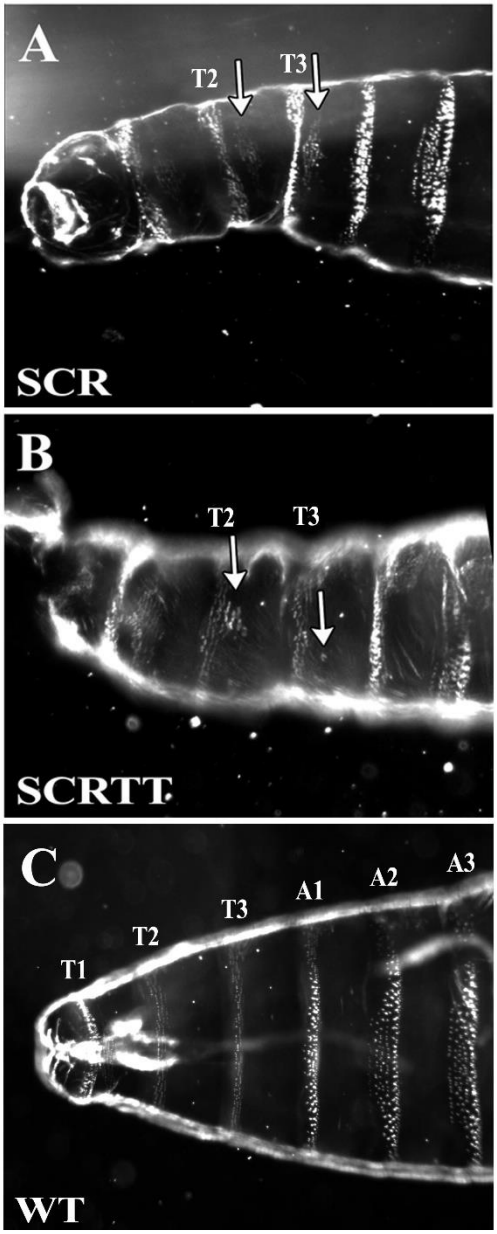

Figure 3. First instar larval cuticular phenotypes of embryonic ectopic expression of SCR and SCRTT proteins. The anterior half of the larva is shown. The untagged SCR protein expressed with the GAL4-*UAS* system using a ubiquitous *armadillo*-GAL4 is shown in panel A; whereas, the SCRTT protein expressed 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

The overall strategy of PTM detection by MS/MS analysis of SCRTT isolated from developing embryos (Figure 4A) required purification of SCRTT. 3g of heat-shocked embryos collected between 0 and 16h AEL were lysed. The nuclear fraction was collected and washed, and the proteins of the nuclear extract were denatured. Denatured SCRTT was purified from the nuclear extract by Ni-NTA affinity chromatography (Hochuli *et al.*, 1987; Hochuli *et al.*, 1988). The purification of SCRTT was monitored by Western Blot analysis showing 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, the amount in the band on the SDS gel that was isolated for enzyme digestion was estimated to be in the range of 10µg.

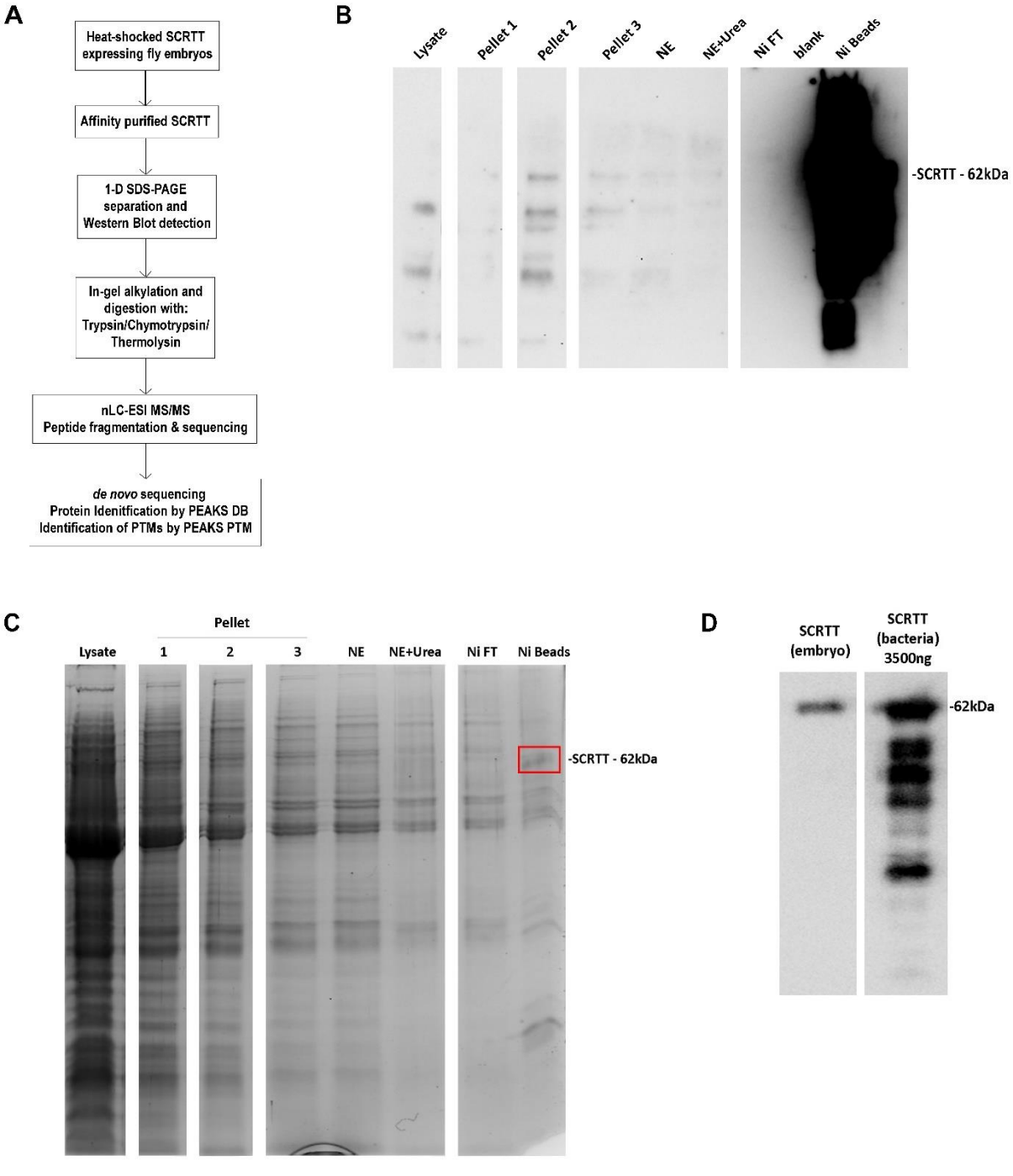

Figure 4. Overall approach for PTM mapping of SCRTT. A. Analytical workflow for affinity purification, digestion and sequence mapping of PTMs in SCRTT expressed from developing embryos. B. Overexposed Western Blot showing SCRTT at 62 kDa in the Ni beads fraction. Purification fraction is indicated on 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.

PeptideCutter (ExPASy) was used to predict peptide fragment size suitable for analysis by LC-MS/MS, with the conclusion that multiple enzymes would be needed to cover the primary sequence of SCRTT. The enzymes chosen for protein digestion were trypsin, chymotrypsin and thermolysin. 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. The coverage and AD_{orc} for each sample analyzed, and the combined coverage and AD_{orc} of various combinations of sample were determined (Table 1). 96% of the primary sequence of SCRTT was covered in the analysis.

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}.

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 may be the result of a biochemical process and not potential by-products of sample preparation were interrogated and characterized in more detail (Table S1). These 44 modifications are phosphorylation, acetylation, formylation, methylation, carboxylation and hydroxylation.

Evidence for the PTMs of SCRTT

Table S1 summarizes all the modified peptides of SCRTT identified and indicates which peptide's MS^2 spectra is shown as supporting evidence (Figures S2-7). For all modifications interrogated, seven 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 the intensity of the informative b and/or y ions were greater than 5%; fourth whether multiple MS^2 spectra were identified for peptides with a particular modification; fifth whether overlapping or differently modified peptides were identified with the same modification; sixth whether a modification was identified in multiple independent samples used for MS/MS analysis (Table 1) and seventh whether the difference of the modified peptide score and the cut-off score was greater than 8. Two additional diagnostic criteria were assessed for phosphorylation: whether b and/or y ions with a neutral loss of phosphoric acid (98 Da) in the MS^2 spectra were present; and second whether the phosphorylated peptide eluted earlier (lower retention time, RT) from the Liquid Chromatography (LC) column than the unphosphorylated but otherwise chemically same peptide (Beausoleil *et al.*, 2004; Kim *et al.*, 2007). The supporting evidence summarized in Table 2 is given with each example MS^2 spectra of a modified peptide (Figures S2-7).

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 distinct peptides[b]	No. of samples observed[c]	Peptide score - cut-off score[d]	b/y ions with a neutral loss of phosphoric acid (98 Da)	RT of phosphopeptide < RT of unphosphorylated peptide
Phosphorylation	S2A	S185	b	b	1	1	1	0.44	b, y	Yes
	S2B	S201	y	y	1	1	1	3.46	b, y	Yes
	S2C	T315	b, y	b	1	1	1	3.9	b, y	No
	S2D	S316	b, y	b, y	2	1	1	3.78	b, y	No
	S2E	T317	b, y	b, y	3	2	1	6.72	b, y	No
	S2F	T324	y	y	9	3	2	**9.43**	b, y	No
Acetylation	S3A	K218	y	y	3	1	2	**18.87**		
	S3B	S223	b, y	y	2	1	1	**11.54**		
	S3C	S227	b, y	-	4	2	1	**14.32**		
	S3D	K309	b	-	1	1	1	3		
	S3E	K434	b, y	b, y	1	1	1	3.2		
	S3F	K439	b, y	b, y	2	1	1	3.15		
Formylation	S4A	K218	y	y	35	1	4	**26.47**		
	S4B	K309	b	-	4	2	3	**12.52**		
	S4C	K325	b, y	y	21	2	3	**17.74**		
	S4D	K341	b, y	b, y	2	1	1	7.21		
	S4E	K369	b, y	b, y	8	1	3	6.03		
	S4F	K434	b, y	b, y	2	1	1	6.07		
	S4G	K439	b, y	b, y	1	1	1	1.63		
Methylation	S5A	S19	b, y	b, y	6	2	2	**8.23**		
	S5B	S166	b, y	b, y	12	6	2	2.79		
	S5C	K168	b, y	b, y	3	2	2	2.39		
	S5D	T364	b, y	y	6	2	2	**13.91**		
Carboxylation	S6A	D108	b, y	b, y	6	2	1	**15.71**		
	S6B	K298	y	-	5	1	1	3.43		

	S6C	W307	b, y	y	16	4	3	**8.09**
	S6D	K309	b	-	14	3	3	**13.22**
	S6E	E323	b, y	b, y	20	5	5	**26.93**
	S6F	K325	b, y	b	8	5	2	**10.76**
	S6G	K369	b, y	b, y	1	1	1	2.37
Hydroxylation	S7A	P22	b, y	b, y	1	1	1	2.62
	S7B	Y87	b, y	b, y	1	1	1	3.29
	S7C	P107	b, y	b, y	15	5	7	7.82
	S7D	D108	b, y	b, y	20	7	7	**10.82**
	S7E	D111	b, y	b, y	2	2	2	6.05
	S7F	P269	y	y	6	5	3	**16.26**
	S7G	P306	b, y	b, y	4	2	2	**19.6**
	S7H	R310	y	-	1	1	1	7.64
	S7I	N321	b	b	1	1	1	3.66
	S7J	K325	y	-	1	1	1	**8.22**
	S7K	Y334	b, y	-	2	1	1	7.06
	S7L	R366	y	-	11	2	1	**13.6**
	S7M	P392	b, y	b, y	1	1	1	2.29
	S7N	Y398	y	-	1	1	1	2.07

[a] Total number of MS^2 spectra reporting the modification from the analysis of the 8 samples (Table 1)
[b] Total number of peptides with either overlapping primary sequence or a distinct set of modifications
[c] Total number of samples the modification was observed in (Table 1)
[d] 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 units, which corresponds to a lower P-value.

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

SCR like many HOX proteins is disordered outside the HD (reviewed in Merabet & Dard, 2014). To test whether SLiMs found in disordered regions of SCR were preferential sites of phosphorylation, 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 5). 6 out of 7 phosphosites were in a SLiM region and 1 of 7 phosphosites was in a conserved SLiM (Figure 5). However, the 6:1 enrichment for phosphorylation of SLiMs was not significant, and the 1:6 phosphorylation of conserved SLiMs was significant indicating that phosphosites outside conserved SLiMs of SCR were preferential sites of phosphorylation (Table 3). The hypothesis that SLiMs are preferential sites of SCR phosphorylation is not supported by this data. The reason why 6 of the 7 phosphosites in SLiMs was not significant is due to the bias of S, T and Y residues to SLiMs. The percentage of S, T and Y was significantly higher in the SLiMs and conserved SLiMs than in non-SLiM portions of SCR (Table 3). In addition, 9 out of 81 SLiMs and 5 out of 51 conserved SLiMs of SCR were phosphorylated (Tables S4 & S5). This may indicate that about 10% of SLiMs are *bona fide* sites of SCR phosphorylation suggesting that a minority of identified SCR SLiMs were phosphorylated which also does not support the hypothesis that SLiMs are preferential sites of phosphorylation.

Table 3. SLiM analysis of SCR

<table>
<tr><th colspan="7">Percentage of SCR protein that are SLiMs</th></tr>
<tr><td>Protein (disordered) size (aa)</td><td>No. of SLiMs</td><td>SLiM size (aa)</td><td>% SLiM</td><td>No. of conserved SLiMs</td><td>Conserved SLiM size (aa)</td><td>% Conserved SLiM</td></tr>
<tr><td>364</td><td>81</td><td>239</td><td>66</td><td>51</td><td>129</td><td>35</td></tr>
<tr><th colspan="7">Distribution of phosphates in SCR SLiMs vs. non-SLiMs</th></tr>
<tr><td colspan="4">Percentage of S, T and Y phosphorylated</td><td colspan="3" rowspan="2">p-value</td></tr>
<tr><td colspan="2">Total SLiMs</td><td colspan="2">Conserved SLiMs</td></tr>
<tr><td>SLiMs</td><td>Non-SLiMs</td><td>SLiMs</td><td>Non-SLiMs</td><td>Total SLiMs</td><td colspan="2">Conserved SLiMs</td></tr>
<tr><td>10</td><td>17</td><td>3</td><td>20</td><td>0.504</td><td colspan="2">0.041</td></tr>
<tr><th colspan="7">Distribution of S, T and Y is biased to SCR SLiMs</th></tr>
<tr><td colspan="4">Percentage of amino acid that are S, T and Y</td><td colspan="3" rowspan="2">p-value</td></tr>
<tr><td colspan="2">Total SLiMs</td><td colspan="2">Conserved SLiMs</td></tr>
<tr><td>SLiMs</td><td>Non-SLiMs</td><td>SLiMs</td><td>Non- SLiMs</td><td>Total SLiMs</td><td colspan="2">Conserved SLiMs</td></tr>
<tr><td>25</td><td>5</td><td>28</td><td>13</td><td>3.38×10^{-7}</td><td colspan="2">5.7×10^{-4}</td></tr>
</table>

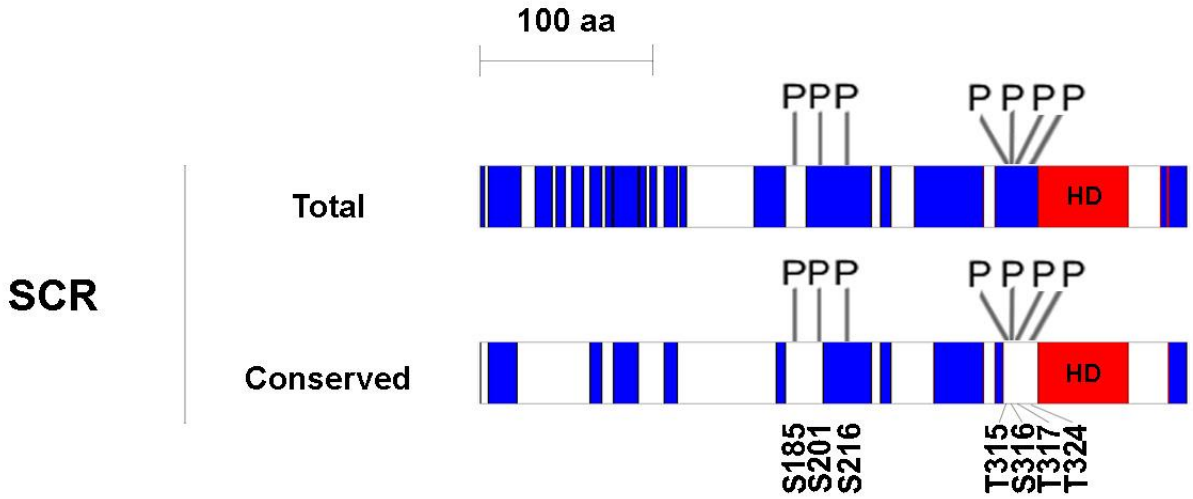

Figure 5. Schematic of SCR showing total and conserved SLiMs. The block diagrams are drawn to scale. The ordered region of the homeodomain is labeled by red and SLiMs are labeled blue. The SLiM data used to construct the figure are in the supplement (Tables S4 & S5). The phosphosites in SCR are marked to scale.

Discussion

PTMs of SCRTT

Bottom-up MS/MS analysis of SCRTT purified from developing *D. melanogaster* embryos identified many amino acid residues that were covalently modified (Figure 6). The analysis of PTMs disregarded modifications that could be due to sample preparation; however, some of these disregarded modifications may result from a biological process, and for 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 the conservation of protein domains and sequence motifs, genetic analysis of SCR function, analysis of the distribution of SLiMs and X-ray crystallography 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.

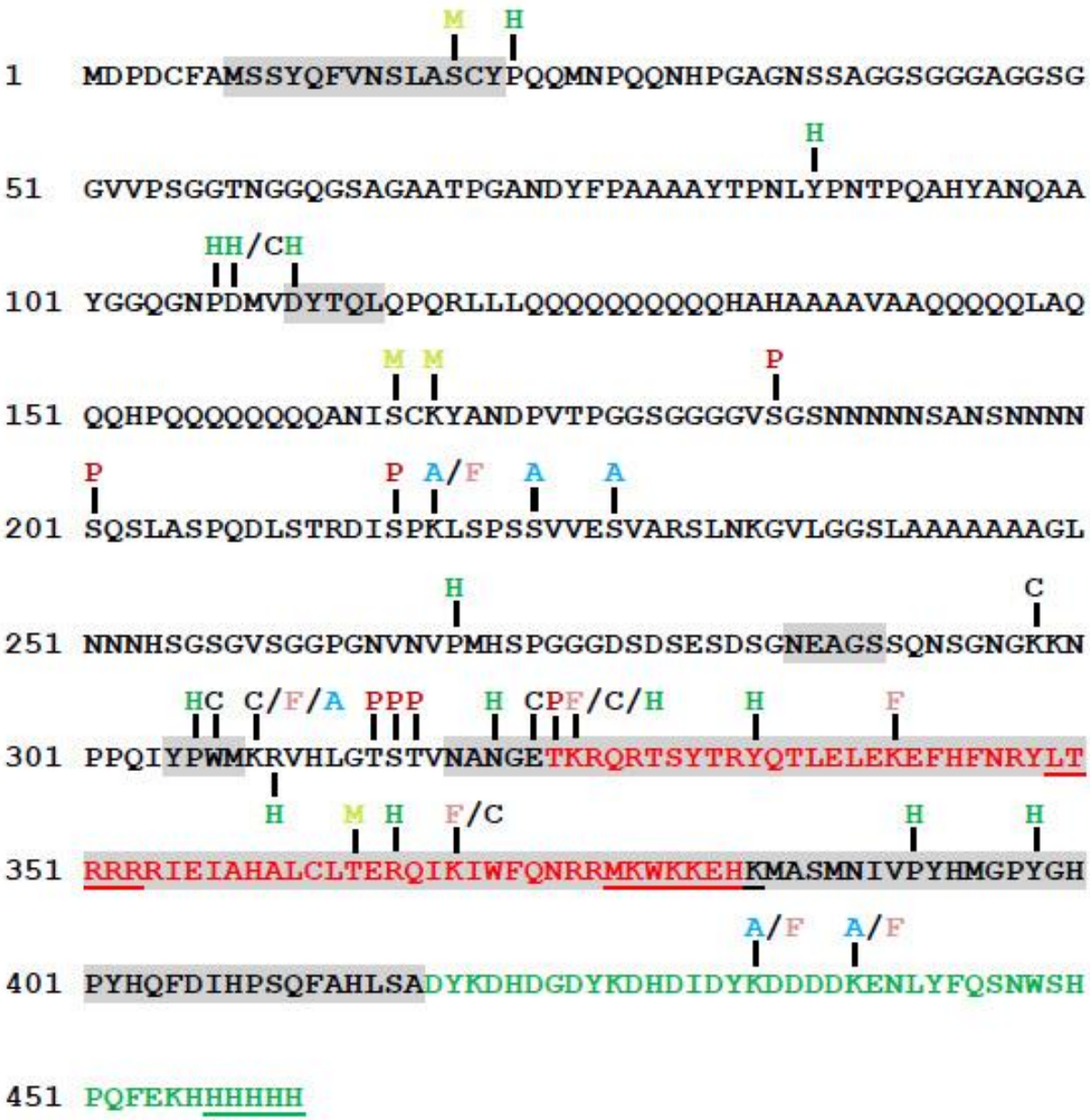

Figure 6. A summary map of post-translational modifications in SCRTT identified by MS/MS. Phosphorylation sites are indicated by "**P**", acetylation by "**A**", formylation by "**F**", methylation by "**M**", carboxylation by "**C**" and hydroxylation by "**H**". The homeodomain is highlighted in red and the triple tag is highlighted in green. The functional regions of SCR are shaded in grey. The amino acids underlined were not detected by MS/MS. Phosphorylation at S216 was identified by Zhai *et al.*, 2008.

Phosphorylation of SCRTT

The analysis identified a clustered set of phosphorylations on the amino acid residues, T315, S316, T317 and T324 flanking the NANGE motif (Figure 5). The NANGE motif is important for the suppression of ectopic proboscises which may suggest that the phosphates have a role in regulating the transformation of SCR from having T1 determining activity to having labial determining activity (Percival-Smith *et al.*, 2013). This cluster of phosphosites map to a disordered region of SCR and hence, are not observed 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 are known to interact 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 none of the amino acid residues of SCR directly involved in minor groove interaction were found to be modified, the residues, T315, S316 and T317 are part of the linker that interacts with the minor groove and T324 is the first amino acid residue of the HD were phosphorylated. 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 the minor groove interaction.

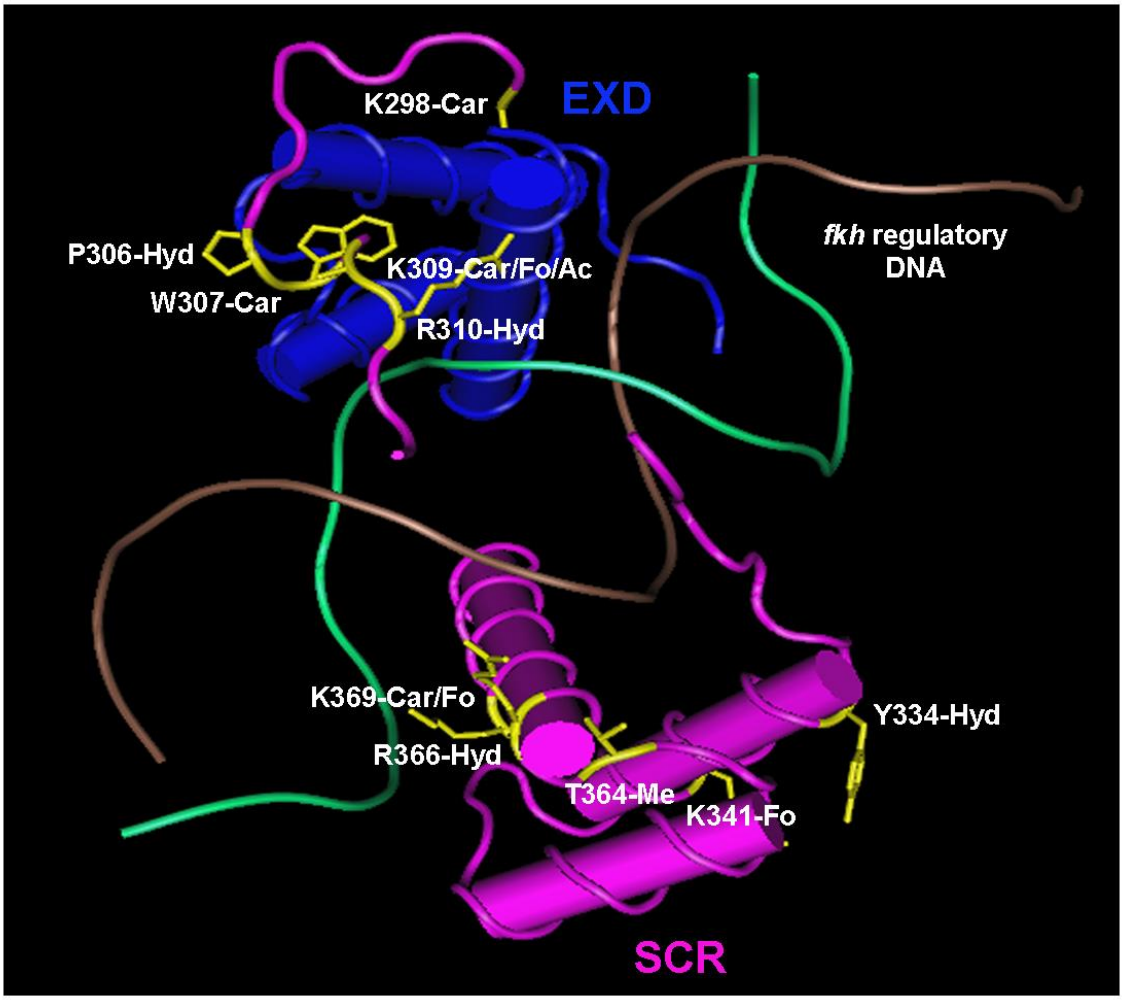

Figure 7. 3-D crystal structure of SCR-EXD-DNA complex. SCR is shown in pink and EXD in blue. The two strands of *fkh* regulatory DNA is shown in brown and green. The modified amino acids of SCR along with their side chains are shown in yellow. Ac – acetylation, Car – carboxylation, Fo – formylation, 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 a model for regulation of SCR activity proposed by Berry & Gehring, 2000, the 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*. This negative regulation of SCR activity is proposed to be reversed by phosphatase PP2A-B'. However, these residues were ruled out as targets of the phosphatase PP2A-B' because loss of PP2A-B' activity had no effect on SCR activity (Moazzen *et al.*, 2009), and here we have not detected phosphorylation of T329 and S330 which may suggest that their phosphorylation by cAMP-dependent protein kinase A is an *in vitro* artifact.

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

The crystal structure of SCR-EXD bound to *fkh* regulatory DNA encompasses the evolutionarily conserved functional motifs/domains, YPWM, NANGE and HD (Figure 7). 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 is known to interact 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 finding agrees with the finding that lysine residues of core histone proteins were N^{ϵ}-formylated and these lysine residues were also sites of frequent acetylation (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, thereby increasing transcription of genes (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, respectively acetylate and deacetylate non-histone proteins, including TFs, which may suggest a mechanism for SCR acetylation/deacetylation (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.

Problems with detecting phosphorylation by MS/MS

Phosphorylation is a major modification involved in the regulation of various cellular processes (Li *et al.*, 2013; Sacco *et al.*, 2012; Mylin *et al.*, 1989; Hunter & Karin, 1992; Ardito *et al.*, 2017). Although phosphorylation of proteins is common, detection of the phosphorylated amino acid residues is still a challenge. Three common arguments are used to address the problem of phosphopeptide detection. Firstly, phosphopeptides are hydrophilic, and hence, they are lost during loading on reversed-phase columns like C18. 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 was no data to support the third argument (Steen *et al.*, 2006). However, with respect to the first two arguments, multiply phosphorylated peptides were detected upon MS/MS analysis of a commercially purchased, pure, bovine α-casein which is a known heavily phosphorylated protein (Larsen *et al.*, 2005) (data not shown). Of the 61 identified phosphorylated peptides that cover the region 61-70 of α-casein, 24 peptides were phosphorylated at two amino acid residues (data not shown). This suggests that phosphopeptides can be successfully detected if the protein is pure and concentrated. This also suggests that phosphorylation is stable during MS/MS analysis although neutral loss of phosphoric acid in case of S- and T-phosphorylated species is common. In conclusion, α-casein was purchased as a pure protein, and its phosphorylation was not substoichiometric as in the case of SCRTT which was enriched from a mixture of embryonic proteins. It is important to know why only a few peptides of SCRTT are phosphorylated. Phosphorylation is a reversible mechanism and phosphatases can readily dephosphorylate proteins during cell signaling. Not all proteins in a cell might be phosphorylated at a given time. Also, a particular protein might have different phosphorylation status in different cells. During nuclear isolation and affinity purification of a particular protein for MS/MS analysis, it is common for the protein to be phosphorylated substoichiometrically. This is the likely reason why not many phosphopeptides were detected upon MS/MS analysis of embryonic SCRTT. In addition, SCRTT phosphorylation might occur in a tissue-specific manner and therefore, not all SCRTT proteins are phosphorylated at one given time. SCRTT was ectopically expressed in the developing embryo with a heat-shock. The overexpressed protein might not be phosphorylated in the same manner as the endogenous protein. A likely reason is that the heat-shock treatment might interfere with the proper functioning of the PTM machinery *in vivo*.

A proteolytic digest of a substoichiometrically phosphorylated protein results in the formation of a huge number of unmodified peptides which dominates the survey mass spectrum. One way to overcome this problem is to enrich for phosphopeptides. I used the EasyPhos protocol which employed TiO_2 beads to enrich trypsinized α-casein and SCRTT for phosphopeptides (Humphrey *et al.*, 2015). This protocol uses the hydrophilic styrenedivinylbenzene-reversed phase sulphonated (SDB-RPS) StageTip which binds the hydrophilic phosphopeptides, thereby minimizing the loss of phosphopeptides as seen for hydrophobic reversed-phase columns, for

example, C18. Although there was a 92.3% phosphopeptide enrichment of α-casein post-TiO_2 treatment (84 were phosphopeptides out of 91 peptides detected), only 54 phosphopeptides were detected post-TiO_2 out of 158 phosphopeptides initially detected without TiO_2 enrichment which indicates loss of phosphopeptides (raw data not shown). No phosphopeptides were detected upon TiO_2 enrichment of trypsinized SCRTT. Even the MS/MS analysis of the TiO_2 flow-through did not identify any phosphopeptides which suggests that there was very few or no phosphopeptides in the SCRTT sample analyzed. It is likely that SCRTT is phosphorylated substoichiometrically and hence, a higher concentration of SCRTT protein is needed to detect phosphopeptides.

SLiM analysis

The hypothesis that the SLiMs of SCR are preferential sites of phosphorylation was not supported by our data (Sivanantharajah & Percival-Smith, 2015). Out of 7 sites of SCR phosphorylation found, 6 were in SLiMs. This is not statistically significant because the phosphorylatable residues, S, T and Y are concentrated in SLiMs. In addition, a minority of SLiMs predicted were phosphosites. This lack of phosphorylation may be a reflection that either the ELM database overpredicts the number of SLiMs, or 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, which are also found in SCR. 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.

Conclusion

This study identified sites of phosphorylation and other PTMs in a tagged HOX protein, SCRTT extracted from developing *Drosophila melanogaster* embryos. The phosphorylation map can be used to design genetic tests to investigate the role of phosphorylation at these sites. The role of other PTMs can also be investigated using the PTM map. Moreover, this protocol can be extended to other tagged HOX proteins for PTM characterization.

Acknowledgements

We thank Paula Pittock from the Biological Mass Spectrometry Laboratory, Western University, 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, Western University, 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 for his help with creating a figure and an undergraduate volunteer student, Stuart Cameron for helping with the project. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant 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.

Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., … Gygi, S. P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(33), 12130-12135.

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.

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.

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 $\chi 2$ 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.

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.

Kim, J., Petritis, K., Shen, Y., Camp, D. G. 2nd, Moore, R. J., & Smith, R. D. (2007). Phosphopeptide elution times in reversed-phase liquid chromatography. *Journal of Chromatography A*, *1172*(1), 9-18.

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.

Li, X., Wilmanns, M., Thornton, J., & Köhn, M. (2013). Elucidating human phosphatase-substrate networks. *Science Signaling*, *6*(275), rs10.

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.

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.

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.

Sacco, F., Perfetto, L., Castagnoli, L., & Cesareni, G. (2012). The human phosphatase interactome: An intricate family portrait. *FEBS Letters*, *586*(17), 2732-2739.

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.