

Electronic Thesis and Dissertation Repository

---

10-19-2018 3:30 PM

## A comprehensive catalog of post-translational modifications of *Drosophila melanogaster* HOX protein, Sex combs reduced

Anirban Banerjee, *The University of Western Ontario*

Supervisor: Percival-Smith, Anthony, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biology

© Anirban Banerjee 2018

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Biochemistry Commons](#)

---

### Recommended Citation

Banerjee, Anirban, "A comprehensive catalog of post-translational modifications of *Drosophila melanogaster* HOX protein, Sex combs reduced" (2018). *Electronic Thesis and Dissertation Repository*. 5758.

<https://ir.lib.uwo.ca/etd/5758>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

## Abstract

During formation of the anterior-posterior axis, Homeotic selector (HOX) proteins determine the identity of *Drosophila* body segments. HOX proteins are transcription factors that regulate gene expression during development. Besides a highly conserved DNA-binding homeodomain (HD), HOX proteins also contain functionally important, evolutionarily conserved small motifs. These short motifs found in HOX proteins may be Short Linear Motifs (SLiMs). SLiMs are proposed to be sites of phosphorylation and this may regulate the activity of HOX proteins. The primary aim of this work was to develop a comprehensive catalogue of the sites of phosphorylation and other post-translational modifications (PTMs) for Fushi tarazu (FTZ) and 8 HOX proteins extracted from developing *Drosophila melanogaster* embryos. *Drosophila* were transformed with constructs that express FTZ or HOX proteins fused to a triple tag (TT) from a heat-shock promoter. The HOXTT proteins are biologically active during embryogenesis. Triple tagged Sex combs reduced (SCRTT) protein was extracted from developing embryos and purified using Ni-NTA beads under denaturing conditions. Multiple sites of PTMs were identified in purified SCRTT by tandem mass spectrometry (MS/MS). The identified PTMs include phosphorylation at S185, S201 and T324, acetylation at K218, K434 and K439, formylation at K218, K309, K325, K369, K434 and K439, methylation at S19, S166, K168 and T364, carboxylation at W307, K309 and E323 and hydroxylation at P22, P107, D108, D111, P269 and P306. In testing the hypothesis that HOX SLiMs are preferential sites of phosphorylation, I found that the distribution of phosphorylatable residues, S, T and Y was biased to SLiMs, but there was no support for the hypothesis that SLiMs are preferentially phosphorylated.

## Keywords

HOX, transcription factors, development, SLiMs, PTMs, phosphorylation

## Co-Authorship Statement

I performed all the experiments mentioned in the ‘Materials and Methods’ and ‘Results’ section of the thesis with a few exceptions: Dr. Anthony Percival-Smith made the fusion constructs, *pbTT*, *DfdTT*, *ScrTT*, *AntpTT* and *Abd-B(r)TT*. Dr. Dan Bath made the *ftzTT* construct. Dr. Anthony Percival-Smith microinjected fly embryos to establish stable germ line transformants expressing the triple tagged constructs. Dr. Lovesha Sivanantharajah provided the plasmid with *ScrTT* inserted behind a T7 promoter. Dr. Anthony Percival-Smith independently mapped post-translational modifications, including phosphorylation in SCRTT and his data has been used in the SLiM analysis and in the ‘Discussion’ section of the thesis. Drs. Anthony Percival-Smith and David Litchfield proofread the thesis.

## Dedication

I am dedicating this thesis to my parents, Mr. Amitava Banerjee (father) and Mrs. Rina Banerjee (mother). They have made many personal sacrifices for my betterment and today, I am a proud son. I thank them for everything they have done for me.

## Acknowledgements

I want to take this opportunity to express my gratitude to my supervisor, Dr. Anthony Percival-Smith and my advisors, Drs. David Litchfield and Susanne Kohalmi for their continuous guidance and support during my PhD program.

I offer special thanks to the graduate students of the Percival-Smith laboratory, Samantha Koot, Laura Garofalo, Alaa Briek, Sheng Cheng and Suhayr Siddiqui Solangi. A special thanks to all my undergraduate volunteer students, especially, Mr. Stuart Cameron, who helped me in my project.

A special thanks to Ms. Paula Pittock from the Biological Mass Spectrometry Laboratory, Ms. Victoria Clarke and Ms. Kristina Jurcic from the Functional Proteomics Facility at The University of Western Ontario for their help with processing my protein samples and giving me valuable advice regarding my data.

I want to thank some more people without whom I would not have been able to complete my thesis: Mr. Gurjit Randhawa for helping me with MATLAB programming to process my SLiM data and create a figure, Dr. Ben Rubin for his valuable advice on statistical analysis of my SLiM data, Dr. Robert Cumming and his lab members for allowing me to image my Western Blots, Dr. Adam Rabalski for his valuable advice on phosphopeptide enrichment, Mr. Mohamed Gatie of the Kelly lab for his help and suggestions on my work from time to time, and Mr. Aria Khalili, Mr. Ian Craig and Mr. Mitch Zimmer for guiding me with Adobe Photoshop.

A special thank you to a close friend of mine, Dr. Mohammad Hossain for his support and guidance as a friend and mentor.

I would like to thank a very special person in my life, my beloved wife, Mrs. Pixie Banerjee for her continuous support and trust in me. I have no words to thank her enough. I express my gratitude to my father-in-law, Mr. Asish Banerjee and my mother-in-law,

Mrs. Shelly Banerjee for their blessings and unconditional love without which it would have been hard for me to come this far.

Lastly, I want to thank my parents for everything. My life is dedicated to them and therefore, everything that I have achieved, or I will achieve is for them.

# Table of Contents

Abstract.....	i
Co-Authorship Statement.....	ii
Dedication.....	iii
Acknowledgments.....	iv
Table of Contents.....	vi
List of Tables.....	ix
List of Figures.....	x
List of Appendices.....	xii
List of Abbreviations.....	xiii
1 Introduction.....	1
1.1 Role of phosphorylation in the regulation of development.....	1
1.2 <i>Drosophila melanogaster</i> as a Model Organism.....	3
1.3 Genetic Regulation of <i>Drosophila</i> Embryogenesis.....	7
1.4 The <i>Homeotic Selector (Hox)</i> Genes.....	10
1.5 The Pair-rule gene, <i>fushi tarazu</i> and its Gene Product.....	15
1.6 Structure and function of HOX proteins.....	15
1.7 Phosphorylation of Homeodomain-containing Proteins.....	19
1.8 Mass Spectrometry (MS).....	23
1.9 MS/MS as a tool to map exact sites of PTMs.....	26
1.10 Results of MS/MS analysis of FTZ and HOX proteins to date.....	28
1.11 Objectives.....	28
2 Materials and Methods.....	29
2.1 <i>Drosophila</i> husbandry.....	29
2.2 Generation of <i>ftz</i> and <i>Hox</i> fusion constructs.....	29

2.3	Generation of transgenic fly lines .....	35
2.4	Heat-shock induction of HOXTT expression .....	35
2.5	GAL4 ectopic expression.....	37
2.6	First instar larval cuticle preparation .....	38
2.7	Affinity purification of HOXTT protein from embryos .....	39
2.8	Expression of SCRTT in <i>E. coli</i> .....	40
2.9	Affinity purification of SCRTT from <i>E. coli</i> .....	40
2.10	Protein separation and staining on a gel.....	41
2.11	Western Blot Analysis .....	41
2.12	Total protein quantification.....	42
2.13	Phosphopeptide enrichment and C18 desalting of non-phosphopeptides.....	42
2.14	Sample preparation for MS/MS .....	43
2.15	Mass spectrometry data analysis.....	44
2.16	Bioinformatic analysis of proteomic data .....	45
2.17	Statistical Analysis.....	45
3	Results .....	46
3.1	Establishing transgenic <i>Drosophila</i> that express HOXTT proteins from a heat-shock promoter.....	46
3.2	Expression of HOXTT proteins from the heat-shock promoter .....	48
3.3	HOXTT proteins are biologically active.....	50
3.4	Levels of HOXTT and FTZTT protein expression using the heat-shock ectopic expression system and GAL4-UAS system .....	56
3.5	Determining the amount of protein required for MS/MS analysis .....	58
3.6	Estimating whether an extract of heat-shocked embryos expressed enough SCRTT protein for MS analysis .....	62
3.7	Analytical workflow for affinity purification, digestion and mapping of PTMs in embryonically expressed SCRTT .....	64



3.8 Embryonic SCRTT is post-translationally modified .....	70
3.8.1 Identification of novel phosphosites in SCRTT.....	74
3.8.2 Identification of acetylated peptides of SCRTT .....	78
3.8.3 Identification of formylated peptides of SCRTT .....	81
3.8.4 Identification of methylated peptides of SCRTT.....	86
3.8.5 Identification of carboxylated peptides of SCRTT.....	89
3.8.6 Identification of hydroxylated peptides of SCRTT .....	93
3.9 Short Linear Motifs (SLiMs) in HOX and FTZ proteins.....	100
3.10 Distribution of S, T and Y is biased towards SLiMs .....	104
3.11 Mapping of HOX phosphosites in SLiMs vs. non-SLiMs.....	106
4 Discussion .....	110
4.1 PTMs of embryonic SCRTT.....	110
4.1.1 Phosphorylation of SCRTT.....	116
4.1.2 PTMs found in the structure of the SCR-EXD-DNA complex .....	117
4.1.3 PTMs in potentially disordered regions of SCR.....	118
4.1.4 PTMs in regions of SCR which are not a functional or a conserved domain/motif or a SLiM .....	120
4.1.5 Competition of acetylation and formylation observed in SCRTT .....	120
4.1.6 Reversible PTMs identified in SCRTT.....	121
4.2 Affinity purification of SCRTT .....	122
4.3 Problems with detecting phosphorylation by MS/MS .....	122
4.4 SLiM analysis and their significance .....	124
4.5 Future directions .....	125
References.....	128
Appendices.....	160
Curriculum Vitae .....	163

## List of Tables

Table 1. Primers used for PCR-based construction of <i>HoxTT</i> fusion.....	34
Table 2. Heat-shock timepoints for HOXTT expression for first instar larval cuticles....	36
Table 3. <i>UAS</i> fly lines used for GAL4 ectopic expression .....	37
Table 4. Predicted and observed relative molecular masses ( $M_r$ ) of HOXTT proteins ....	50
Table 5. Quantification of SCRTT protein signals .....	64
Table 6. Purification table for SCRTT.....	67
Table 7. MS/MS analysis – Coverage and $AD_{orc}$ for SCRTT .....	69
Table 8. Post-translational modifications of SCRTT identified by MS/MS.....	70
Table 9. Post-translational modifications common to bacterial and embryonic SCRTT	100
Table 10. SLiMs in <i>Drosophila melanogaster</i> HOX and FTZ proteins.....	101
Table 11. S, T and Y distribution in SLiMs vs. non-SLiMs and their significance .....	105
Table 12. Mapping of HOX phosphosites in SLiMs vs. non-SLiMs.....	107
Table 13. Phosphosite distribution in SLiMs vs. non-SLiMs and their significance .....	109

## List of Figures

Figure 1. Hierarchical expression of segmentation genes during embryogenesis .....	9
Figure 2. Homeotic complex in <i>D. melanogaster</i> and the expression of <i>Hox</i> genes in the developing embryo.....	11
Figure 3. <i>Hox</i> cluster in Bilaterans .....	14
Figure 4. Schematic of SCR protein .....	17
Figure 5. Schematic of Tandem Mass Spectrometry (MS/MS).....	26
Figure 6. Mapping exact site of phosphorylation by tandem mass spectrometry (MS/MS) – an example .....	28
Figure 7. Schematic of <i>HoxTT</i> fusion construct.....	31
Figure 8. Schematic of PCR-based construction of <i>HoxTT</i> fusion.....	32
Figure 9. Selecting the transgenic fly line expressing maximum HOXTT protein .....	48
Figure 10. Expression of HOXTT proteins from the heat-shock promoter .....	49
Figure 11. First instar larval cuticular phenotypes of embryonic ectopic expression of HOX and HOXTT proteins.....	55
Figure 12. Comparison of the expression of triple tagged proteins from heat-shock and <i>UAS</i> promoters .....	58
Figure 13. Bacterially expressed SCRTT and sample preparation for MS/MS.....	60
Figure 14. Identification of phosphopeptides of a commercially purchased, pure phosphoprotein, $\alpha$ -casein .....	62
Figure 15. Comparison of bacterially expressed SCRTT and homogenized heat-shocked SCRTT embryos .....	63

Figure 16. Overall approach of affinity purification to PTM mapping in SCRTT.....	66
Figure 17. Comparison of <i>Drosophila</i> SCRTT vs. bacterial SCRTT to estimate the amount of protein to be analyzed by MS/MS .....	68
Figure 18. Identification of novel phosphorylation sites, Serine 185, Serine 201 and Threonine 324 in SCRTT.....	78
Figure 19. Acetylation of Lysine 218, 434 and 439 residues of SCRTT.....	81
Figure 20. Formylation of Lysine 218, 309, 325, 369, 434 and 439 residues of SCRTT	86
Figure 21. Methylation of Serine 19, Serine 166, Lysine 168 and Threonine 364 residues of SCRTT .....	89
Figure 22. Carboxylation of Tryptophan 307, Lysine 309 and Glutamic acid 323 residues of SCRTT .....	92
Figure 23. Hydroxylation of Proline 22, 107, 269, 306, Aspartic Acid 108 and 111 residues of SCRTT.....	98
Figure 24. A summary map of post-translational modifications in SCRTT identified by MS/MS.....	99
Figure 25. Schematic of HOX and FTZ proteins showing total and conserved SLiMs .	104
Figure 26. A summary map of post-translational modifications in SCRTT identified by MS/MS.....	112
Figure 27. 3-D crystal structure of SCR-EXD-DNA complex.....	115

## List of Appendices

Appendix A. Permission to use Figure 1 .....	160
Appendix B. Permission to use Figure 2 .....	161
Appendix C. Permission to use Figure 5 .....	162

## List of Abbreviations

A1	first abdominal segment
aa	amino acid
<i>abd-A</i>	<i>abdominal-A</i> (gene)
ABD-A	Abdominal-A (protein)
ABD-ATT	Triple tagged Abdominal-A (protein)
<i>Abd-B</i>	<i>Abdominal-B</i> (gene)
ABD-B	Abdominal-B (protein)
ABD-B(m)	Abdominal-B morphogenetic isoform (protein)
ABD-B(m)TT	Triple tagged Abdominal-B morphogenetic isoform (protein)
ABD-B(r)	Abdominal-B regulatory isoform (protein)
ABD-B(r)TT	Triple tagged Abdominal-B regulatory isoform (protein)
Ac	acetylation
AD	Average depth
AD <sub>orc</sub>	Average depth of regions covered
AEL	after egg laying
ANT-C	Antennapedia complex
<i>Antp</i>	<i>Antennapedia</i> (gene)
ANTP	Antennapedia (protein)
ANTPTT	Triple tagged Antennapedia (protein)
A-P	anterior-posterior
BAC	Bacterial Artificial Chromosome
<i>bcd</i>	<i>bicoid</i> (gene)
BCD	Bicoid (protein)
<i>BicD</i>	<i>Bicaudal D</i> (gene)
BIP2	Bric-à-brac interacting protein 2
BOSS	Bride-of-sevenless
BSA	Bovine serum albumin
BX-C	Bithorax complex

CAD	Collisionally activated dissociation
Car	carboxylation
Cas9	CRISPR-associated nuclease 9
CDK	cyclin-dependent kinase
CDS	coding sequence
CID	Collision-Induced Dissociation
CK1	casein kinase 1
CKII	casein kinase II
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
C-terminal	Carboxy terminal
CTD	C-terminal domain
1-D	1-dimensional
2-D	2-dimensional
3-D	3-dimensional
Da	Dalton
<i>Dfd</i>	<i>Deformed</i> (gene)
DFD	Deformed (protein)
DFDTT	Triple tagged Deformed (protein)
<i>dpp</i>	<i>decapentaplegic</i> (gene)
DRK	Downstream of receptor kinase
DSB	double-stranded breaks
DTT	1,4-dithiothreitol
EDTA	ethylenediaminetetraaceticacid
ELM	Eukaryotic linear motif
<i>en</i>	<i>engrailed</i> (gene)
EN	Engrailed (protein)
ESI	Electrospray Ionization
ETS	E26 transformation-specific or E-twenty-six
<i>eve</i>	<i>even skipped</i> (gene)
EVE	Even skipped (protein)
EXD	Extradenticle (protein)

FDR	false discovery rate
<i>fkf</i>	<i>fork head</i> (gene)
FLP	FLP recombinase
Fo	formylation
FRT	FLP recombinase target site
FT	flow-through
<i>ftz</i>	<i>fushi tarazu</i> (gene)
FTZ	Fushi tarazu (protein)
FTZ-F1	Fushi tarazu factor 1 (protein)
FTZTT	Triple tagged Fushi tarazu (protein)
GEF	Guanine exchange factor
<i>gsb</i>	<i>gooseberry</i> (gene)
GSK3	glycogen synthase kinase 3
HD	Homeodomain
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	high fidelity
HOM-C	Homeotic complex
<i>Hox</i>	<i>Homeotic selector</i> (gene)
HOX	Homeotic selector (protein)
HOXTT	Triple tagged Homeotic selector (protein)
HRP	horseradish peroxidase
<i>hsp</i>	heat-shock promoter
<i>hsp70</i>	<i>heat shock protein 70</i> (gene)
Hyd	hydroxylation
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
kDa	kilodalton
<i>Kr</i>	<i>Krüppel</i> (gene)
Lab	Labial
<i>lab</i>	<i>labial</i> (gene)
LAB	Labial (protein)
LABTT	Triple tagged Labial (protein)



LB	Luria-Bertani
LC	Liquid Chromatography
MALDI	Matrix-Assisted Laser Desorption/Ionization
MAPK	Mitogen-activated protein kinase
MAPKK	Mitogen-activated protein kinase kinase
MAPKKK	Mitogen-activated protein kinase kinase kinase
Mb	Mega base pairs
Md	Mandibular
Me	methylation
M <sub>r</sub>	Relative molecular mass
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MS/MS/MS	3-Stage Mass Spectrometry
Mx	Maxillary
<i>m/z</i>	mass/charge
NE	nuclear extract
Ni-NTA	Nickel-nitrilotriacetic acid
NP-40	Nonidet P-40
N-terminal	amino terminal
OD	Optical density
<i>odd</i>	<i>odd-skipped</i> (gene)
PAGE	Polyacrylamide gel electrophoresis
<i>pb</i>	<i>proboscipedia</i> (gene)
PB	Proboscipedia (protein)
PBTT	Triple tagged Proboscipedia (protein)
Pc-G	Polycomb group
PCR	Polymerase chain reaction
PIKK	Phosphatidylinositol 3-kinase-related kinases
PKA	Protein kinase A
Plk1	Polo-like kinase 1
PMSF	phenylmethanesulfonylfluoride

PNT	Pointed
<i>PP2A,B'</i>	<i>Protein phosphatase 2A</i> (gene)
PP2A,B'	Protein phosphatase 2A (protein)
pS	phosphoserine
pT	phosphothreonine
PTM	Post-translational modification
<i>p</i> -value	probability value
PVDF	Polyvinylidene difluoride
pY	phosphotyrosine
R7	rhabdomere 7
rpm	revolutions per minute
RT	retention time
RTK	Receptor tyrosine kinase
<i>run</i>	<i>runt</i> (gene)
<i>Scr</i>	<i>Sex combs reduced</i> (gene)
SCR	Sex combs reduced (protein)
SCRTT	Triple tagged Sex combs reduced (protein)
SDB-RPS	styrenedivinylbenzene-reversed phase sulphonated
SDS	Sodium dodecyl sulphate
SEV	Sevenless
SH2	Src homology 2
SH3	Src homology 3
SLiM	Short Linear Motif
SOS	Son-of-sevenless
T1	first thoracic segment
TALEN	Transcription Activator-Like Effector Nuclease
<i>Taq</i>	<i>Thermus aquaticus</i>
TEV	Tobacco Etch Virus
TF	transcription factor
TGFβ	Transforming Growth Factor β
TOF	Time-of-Flight

Trx-G	Trithorax group
TT	triple tag
<i>UAS</i>	Upstream Activation Sequence
<i>Ubx</i>	<i>Ultrabithorax</i> (gene)
UBX	Ultrabithorax (protein)
UBXTT	Triple tagged Ultrabithorax (protein)
w/v	weight/volume
v:v	volume:volume
ZFN	Zinc-Finger Nuclease

# 1 Introduction

## 1.1 Role of phosphorylation in the regulation of development

The central question in the study of development is how a multicellular organism forms from a single-celled fertilized egg. Development is the regulation of the vital processes of cell division, cell fate determination and cell differentiation. The developmental fate of genetically identical cells is first determined leading to differential gene expression and cellular differentiation. The coordination of these processes of the determination and differentiation of specific cell types leads to the synchronized formation of tissues and organs. Transcriptional regulation is a mechanism for the control of gene expression. Transcription factors (TFs) control expression of genes by binding to *cis*-acting regulatory DNA sequences recruiting RNA polymerase to the promoter (Bram & Kornberg, 1985; Giniger *et al.*, 1985; Brent & Ptashne, 1985; Keegan *et al.*, 1986; Ptashne, 1986; Ptashne, 1988). Many TFs have been identified that play central roles in the control of differential gene expression and development. Homeotic Selector (HOX) proteins are one such set of TFs.

Covalent post-translational modifications (PTMs) can change the behavior of proteins including TFs by: regulating the activity state, influencing translocation from one subcellular compartment to another, and controlling interactions with other proteins (Cohen, 2000; Mann & Jensen, 2003). Phosphorylation is one of the most important PTMs (Li *et al.*, 2013; Sacco *et al.*, 2012). It is often a regulatory component of 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). Protein kinases in eukaryotes catalyze the addition of phosphate group(s) (PO<sub>4</sub>) to the polar R group of amino acid residues of proteins, primarily, serine (S), threonine (T) and tyrosine (Y), and sometimes histidine is also phosphorylated (Manning *et al.*, 2002). The addition of a charged phosphate group changes a non-charged residue in a protein to a charged one, which can change the three-dimensional (3-D) conformation of the protein

(Barford & Johnson, 1989; Barford *et al.*, 1991; Sprang *et al.*, 1988; Groban *et al.*, 2006). This mechanism is reversible as phosphatases remove phosphate groups from the amino acids of proteins. Many TFs, enzymes and receptors are activated/deactivated by phosphorylation/dephosphorylation events mediated by kinases/phosphatases.

Phosphorylation does regulate the activity of TFs, and can influence a developmental process. One example of this regulation occurs during *Drosophila* eye development in the determination of photoreceptor cell (rhabdomere 7) (R7) fate. The *Drosophila* adult compound eye is made up of approximately 800 units known as ommatidia. Each ommatidia consists of 20 highly ordered cells of which 8 are photoreceptor (rhabdomere) cells, R1 to R8, 4 are non-neuronal cone cells, and 8 are accessory cells. The ommatidial cells have unique morphologies and positions. The outer R1-R6 photoreceptors have large rhabdomeres, which are an array of microvilli along the retinal cell margin that contain the light-sensitive Rhodopsins. The outer R1-R6 photoreceptors surround the inner R7 and R8 photoreceptors with smaller rhabdomeres. This ordered cellular unit develops during the third instar larval stage by the ordered assembly and differentiation of a set of undifferentiated pluripotent cells of the eye imaginal disc (reviewed in Wolff and Ready, 1993). Undifferentiated cells receive specific signals from neighboring cells determining their cellular fate. The R8 photoreceptor cell is the first to differentiate followed by R2/R5, R3/R4, R1/R6, and R7 neuronal cells. The R8 cells induce the R7 precursor cells to adopt a R7 cell fate via the Sevenless receptor tyrosine kinase (RTK) pathway (reviewed in Wassarman, *et al.*, 1995). The R8 cells express a ligand called Bride-of-sevenless (BOSS) which binds the transmembrane RTK, Sevenless (SEV) expressed in the R7 precursor cells. This activates the autophosphorylation of tyrosine residues of the intracellular domain of SEV. Inside the R7 precursor cells, the adaptor protein, Downstream of receptor kinase (DRK) binds to phosphorylated SEV via its Src homology (SH) 2 domain and also to Son-of-sevenless (SOS) with its SH3 domain (Simon *et al.*, 1993; Olivier, *et al.*, 1993). SOS, a guanine nucleotide exchange factor (GEF) promotes the exchange of GDP for GTP in RAS1, activating this G-protein (Simon *et al.*, 1991; Rogge *et al.*, 1991). The activated RAS1 turns on a phosphorylation cascade that starts with the RAS1-induced activation of the mitogen-activated protein kinase kinase kinase (MAPKKK), RAF serine/threonine kinase (Moodie & Wolfman,

1994). RAF phosphorylates the mitogen-activated protein kinase kinase (MAPKK), DSOR1 tyrosine/threonine kinase (Tsuda *et al.*, 1993; Lu *et al.*, 1994), which subsequently phosphorylates the mitogen-activated protein kinase (MAPK), Rolled serine/threonine kinase (Biggs & Zipursky, 1992; Biggs *et al.*, 1994). Rolled phosphorylates two members of the E26 transformation-specific or E-twenty-six (ETS) family of TFs, Pointed (PNT) and Yan, thereby regulating their function (O'Neill *et al.*, 1994; Brunner *et al.*, 1994). In the absence of a R8 BOSS signal, Yan outcompetes the weak transcriptional activator, PNT, for binding sites, and represses the expression of R7 target genes (O'Neill *et al.*, 1994; Rebay & Rubin, 1995; Xu *et al.*, 2000), and cone cell fate is determined. Upon activation of SEV/RTK signaling, Rolled phosphorylates Yan, thereby promoting nuclear export and subsequent degradation and inactivation of Yan (O'Neill *et al.*, 1994; Rebay & Rubin, 1995). Rolled also phosphorylates the PNTP2 isoform of Pointed. Phosphorylated PNTP2 acts as a strong transcriptional activator that turns on target gene expression required for determination of R7 fate repressed by Yan (O'Neill *et al.*, 1994; Brunner *et al.*, 1994).

Like PNT and Yan, HOX proteins are developmentally important TFs involved in establishing segmental identity in all bilaterally symmetrical animals. HOX TFs are encoded by a set of homeobox genes called *Hox* genes. The 180 bp homeobox encodes the highly conserved, 60 amino acid homeodomain (HD) which is the DNA-binding domain of HOX proteins (McGinnis *et al.*, 1984). HOX proteins are phosphoproteins and mapping the sites of phosphorylation is an important first step for determining whether phosphorylation plays an important role in the determination of segmental identity.

## 1.2 *Drosophila melanogaster* as a Model Organism

The experimental advantages of *Drosophila* have assisted in the development of *Drosophila* into a sophisticated model organism for the study of development, neurobiology and behavior. For studies in developmental genetics, *Drosophila* has a well-characterized life cycle, genome, genetics and the ability to reintroduce DNA into the genome. *Drosophila* is also a good model for some biochemical studies.

The life cycle of *D. melanogaster* is short, approximately 10-12 days from egg laying to eclosion of an adult at 25°C and the eggs develop externally. The life-cycle of *D. melanogaster* consists of 4 distinct stages: embryo, larva, pupa and imago. After internal fertilization, eggs are laid externally in large numbers (average 700-1000 per female) allowing large-scale sample collection (Ashburner, 1989). Although development is a continuous cycle, the fusion of the paternal and maternal nuclei creating the zygotic nucleus is often used to mark the beginning of development. During the subsequent syncytial blastoderm stage which lasts for 2h after egg laying (AEL), the zygotic nucleus and embryonic nuclei undergo rapid and synchronous nuclear divisions to establish the multi-nucleate, single-celled syncytial blastoderm. Prior to the tenth nuclear division, a few nuclei of the syncytial blastoderm migrate to the posterior pole of the embryo where cell membranes assemble around them to form germ-line pole cells. The rest of the nuclei, destined to become the soma, finish the thirteen synchronous divisions before the cell membrane starts developing around the nuclei marking the formation of the monolayer cellular blastoderm. The cellular blastoderm stage lasts from 2h10m to 2h50m AEL. Then from 2h50m to 3h AEL, the cellular blastoderm undergoes gastrulation which gives rise to an embryo with the three germ layers: ectoderm, mesoderm and endoderm. During gastrulation, germ band extension (3h10m to 7h20m AEL) begins when the posterior part of the embryo folds dorsally over the embryo, thus establishing a U-shaped A-P axis and then, germ band retraction occurs (7h20m to 9h20m AEL). During germ band extension and retraction, organogenesis is also occurring. The head involutes and the cuticle is secreted. The free-living first instar larva hatches approximately 24h AEL marking the end of embryogenesis (Campos-Ortega & Hartenstein, 1985; Ashburner, 1989). After 48h and 72h AEL, the larva will molt and form second and third instar larva, respectively. At about 5 days AEL, the third instar larva becomes sessile and pupariation begins. During the pupal stages, metamorphosis occurs when much of the larval tissue undergoes histolysis and adult structures develop from cells of the imaginal discs. After approximately 10 days AEL at 25°C, the adult fly or imago ecloses from the pupal case. The cycle repeats once the adult reaches sexual maturity (Campos-Ortega & Hartenstein, 1985).

The *Drosophila melanogaster* genome is well characterized having been sequenced and extensively annotated (Adams *et al.*, 2000). The *Drosophila* nuclear genome is approximately 143 Mb in size consisting of 3 autosomes and 2 sex chromosomes containing a total of 17,682 genes. A free integrated database of *Drosophila* genomic and proteomic information called FlyBase is available for use (Gramates *et al.*, 2017).

Sophisticated genetic tools are available to study gene function in *Drosophila* (Greenspan, 2004). Ectopic mis-expression of a gene can be achieved with either the binary GAL4-*UAS* system, driving the expression of a gene from a heat-shock promoter or the FLP-FRT mediated ectopic expression system (D'Avino & Thummel, 1999; Brand & Perrimon, 1993; Struhl, 1985; Struhl & Basler, 1993). Post-transcriptional inhibition of gene expression can be achieved with the RNA interference technique that can act as a reversible switch turning on or off the activity of a gene (Kavi *et al.*, 2008). Gene editing techniques initiated by protein-induced, double-stranded DNA breaks (DSBs) like Transcription Activator-Like Effector Nucleases (TALENs) and Zinc-Finger Nucleases (ZFNs) (Bibikova *et al.*, 2002) and a relatively new riboprotein technique called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9 (CRISPR-associated nuclease 9) are being widely used to modify and manipulate the *Drosophila* genome (Cong *et al.*, 2013; Gratz *et al.*, 2013).

The GAL4-*UAS* ectopic expression system allows temporal and tissue-specific expression of a gene of interest. The yeast transcriptional activator GAL4 binds to the Upstream Activation Sequence (*UAS*) and drives expression of the gene(s) downstream of *UAS* in a temporal and tissue-specific manner. For this binary system, two independent fly lines are required: one with the promoter GAL4 fusion (driver GAL4) that expressed GAL4 in a specific spatial and temporal pattern and the other with the *UAS* sequence fused to the downstream gene of interest. The gene of interest downstream of *UAS* is not expressed in the absence of the GAL4 driver. In the progeny of a cross of the two fly lines, GAL4 binds to *UAS*, thus allowing the specific spatial and temporal expression of the gene of interest (D'Avino & Thummel, 1999; Brand & Perrimon, 1993).



Another system used for ectopically expressing a gene of interest is the fusion of a *Drosophila heat shock protein 70 (hsp70)* gene promoter. Temporal control of gene expression can be achieved by when a heat-shock is administered (Struhl, 1985; D'Avino & Thummel, 1999).

Many transgenesis tools exist to reintroduce DNA into the fly germ line (Venken & Bellen, 2005; Venken & Bellen, 2007). The common fly transgenesis tools include the untargeted *P*-element-mediated transformation (Rubin & Spradling, 1982; Ryder & Russell, 2003), site-specific transgenesis using the bacteriophage  $\Phi$ C31 integrase (Groth *et al.*, 2004) and recombineering using bacterial artificial chromosomes (BACs) (Venken *et al.*, 2006). Moreover, different readily available mutant and transgenic lines can be purchased for experiments from different stock centres like Bloomington *Drosophila* Stock Center at Indiana University, USA.

Presently, *P*-element mediated transformation is the major method for reintroduction of genetic material into the *Drosophila* genome (Rubin & Spradling, 1982; Ryder & Russell, 2003). A naturally-occurring transposable element, the *P*-element is used in this technique. The *P*-element has two components: 1) a transposase gene, and 2) inverted repeats flanking the transposase gene. The transposase enzyme binds to the inverted repeats to mobilize the piece of DNA containing the transposase gene flanked by the repeats. For transgenesis, a helper plasmid carrying the transposase gene with inactive inverted repeats and a vector plasmid carrying the gene of interest along with an eye color marker gene flanked by inverted repeats are required. Each of the two plasmids contain a bacterial origin of replication and an antibiotic selectable marker thereby allowing the use of molecular cloning strategies. Flies at the syncytial blastoderm stage are injected with the two plasmids. Upon microinjection, the transposase enzyme catalyzes the excision of the DNA flanked by the two inverted repeats and randomly inserts the excised DNA fragment into the genome of the germ-line cells. Germ-line transformant flies are detected by screening for the eye color marker phenotype.

Lastly, modern methods of creating protein fusions have improved biochemical analysis in *Drosophila* (Kanca *et al.*, 2017). Protein fusions are expressed from chimeric genes of

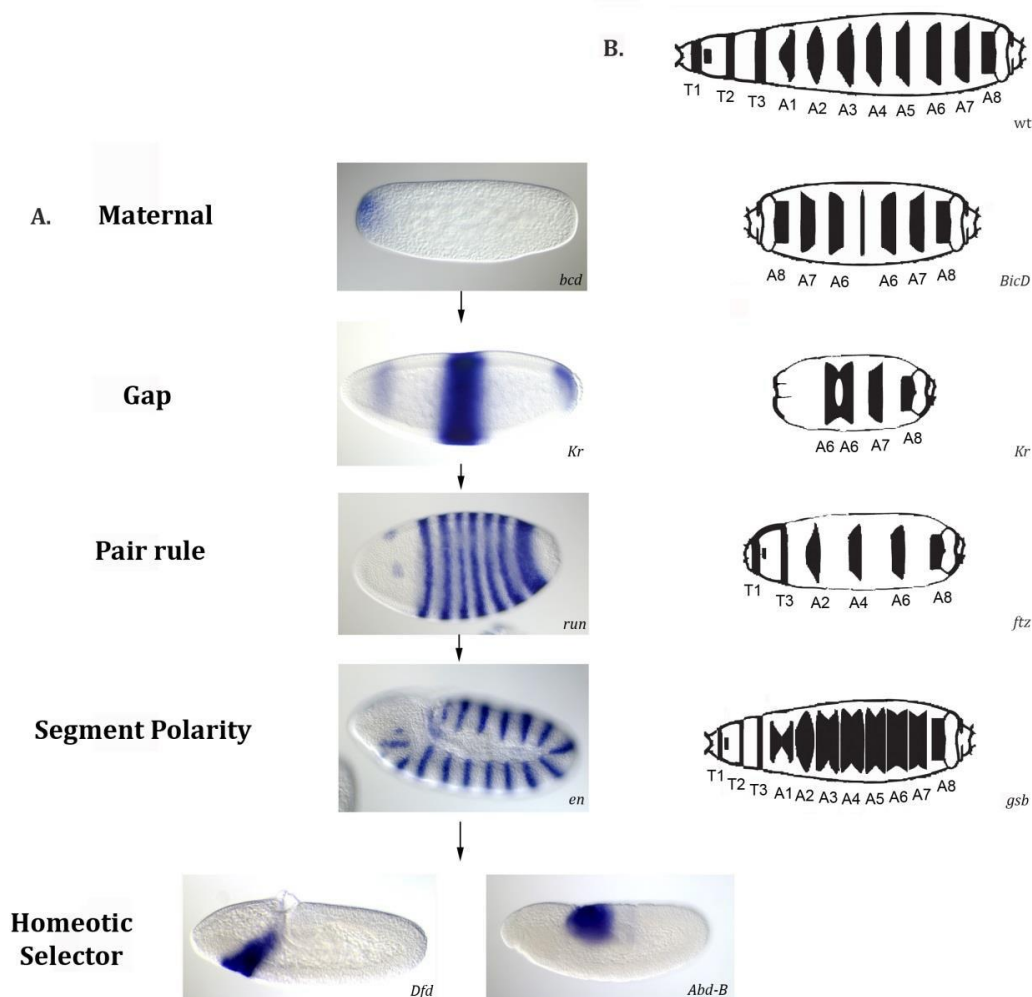
the fusion of the gene in frame with sequences that encode a protein tag. This tag may serve as an epitope to which monoclonal antibodies bind and allow detection of the protein fused to the tag. The detection of epitope-tagged proteins with monoclonal antibodies is achieved in techniques like Western Blot, Immunohistochemistry, and Immunofluorescence microscopy. Affinity purification tags are also used to purify proteins from a complex protein mixture with an aim to either unravel the interactome or map PTMs using Mass Spectrometry (MS) (reviewed in Aebersold & Mann, 2003; Guerrero & Kleiner, 2005; Han *et al.*, 2008).

### 1.3 Genetic Regulation of *Drosophila* Embryogenesis

The body plan of *Drosophila melanogaster* is segmented along the Anterior (A)-Posterior (P) axis and the process of segmentation is regulated by hierarchical interactions between sets of genes (Figure 1) (reviewed in Akam, 1987; St Johnston & Nüsslein-Volhard, 1992). *Drosophila* has two offset metamereric registers: the early 14 parasegmental register and the later morphological register of 15 segments. The 15 segments are composed of: 3 head (Mandibular (Md), Maxillary (Mx) and Labial (Lab)), 3 thoracic (T1-T3) and 9 abdominal segments (A1-A9) where Md is the anterior-most and A9 is the posterior-most segment (Martinez-Arias & Lawrence, 1985). Segments are distinct developmental units that form perpendicular to the A-P axis. Many genes are required for proper A-P patterning of the embryo. These are grouped into five different hierarchical classes based on their role in patterning: Maternal effect coordinate, Gap, Pair-rule, Segment polarity and Homeotic selector (*Hox*) genes. The phenotypes of the mutations in segmentation genes suggested their role in segmentation (Nüsslein-Volhard & Wieschaus, 1980; Jürgens *et al.*, 1984; Nüsslein-Volhard *et al.*, 1984; Wieschaus *et al.*, 1984; Nüsslein-Volhard, 1979). The first class are the maternal effect coordinate genes required for establishing the A-P axis (Nüsslein-Volhard *et al.*, 1987; Nüsslein-Volhard, 1991). Coordinate genes are transcribed from the maternal genome during oogenesis and the mRNA transcripts are deposited in the egg. These stored mRNA are translated after fertilization and before the activation of transcription of the zygotic genome. The protein of the mRNA of the *bicoid* gene determines the anterior end (head and thoracic segments)

of the developing embryo (Driever & Nüsslein-Volhard, 1988; Nüsslein-Volhard *et al.*, 1987), and the proteins of the mRNA of the *nanos* and *caudal* genes determine the posterior end (abdominal segments) of the developing embryo (Lehmann & Nüsslein-Volhard, 1991). The proteins encoded by the mRNA of the coordinate genes transcriptionally and post-transcriptionally regulate the spatial expression of the second class of segmentation genes, the gap genes (Driever & Nüsslein-Volhard, 1989). The gap genes are amongst the first genes expressed from the zygotic genome and they divide the embryo into distinct domains along the A-P axis. *hunchback*, *Krüppel*, *giant*, *knirps*, *tailless* and *huckebein* are examples of gap genes, and each regulates the formation of a distinct set of contiguous segments (Nüsslein-Volhard & Wieschaus, 1980; Jürgens *et al.*, 1984; Wieschaus *et al.*, 1984). Loss of the function of a Gap protein results in an embryo that develops with a loss of a contiguous set of segments. Gap proteins are transcription factors that control the spatial expression of primary pair-rule genes like *even skipped* (*eve*) and *hairy* (Jäckle *et al.*, 1992). The transcription factors encoded by primary pair-rule genes regulate the expression of secondary pair-rule genes, *fushi tarazu* (*ftz*) and *odd-skipped* (*odd*) (Carroll & Vavra, 1989; Vavra & Carroll, 1989). The primary and secondary pair-rule genes determine the number of segments that form (Frasch & Levine, 1987). Loss-of-function mutations in *ftz* result in an embryo with only odd-numbered parasegments (Jürgens *et al.*, 1984); whereas, loss-of-function mutations in *eve* results in an embryo with only even numbered parasegments (Nüsslein-Volhard & Wieschaus, 1980; Nüsslein-Volhard *et al.*, 1984). The pair-rule genes encode transcription factors that control the expression of a fourth set of genes called the segment polarity genes, which are required to pattern each segment. The gene products of segment polarity genes, *engrailed*, *hedgehog* and *wingless*, are responsible for establishing A-P polarity within each segment (reviewed in Akam, 1987). Mutations in these genes result in segmentation defects due to problems with polarity that can result in polarity reversal within each segment (Nüsslein-Volhard & Wieschaus, 1980; Jürgens *et al.*, 1984; Nüsslein-Volhard *et al.*, 1984).

The above four classes of genes lay out the segmented body plan of the embryo. While the fifth class, the *Homeotic selector* (*Hox*) genes determine the individual structure or identity of each segment (reviewed in Akam, 1998).



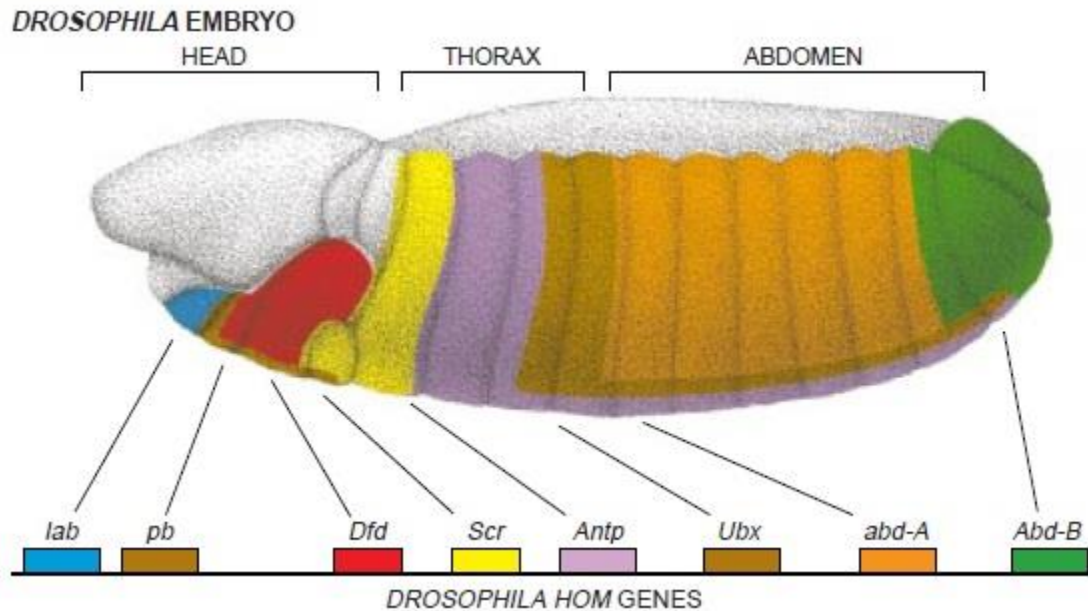
**Figure 1. Hierarchical expression of segmentation genes during embryogenesis.** A. *In situ* hybridization showing the mRNA localization of an example gene for each level of the segmentation hierarchy are shown (images from Tomancak *et al.*, 2007; Tomancak *et al.*, 2002). Localization of the mRNA of maternal effect gene, *bicoid* (*bcd*), gap gene, *Krüppel* (*Kr*), pair-rule gene, *runt* (*run*), segment polarity gene, *engrailed* (*en*), *Hox* genes, *Deformed* (*Dfd*) and *Abdominal-B* (*Abd-B*) are shown. The anterior end of the embryo is on the left. B. Diagrammatic representation of first instar larval cuticular phenotypes of wild-type and mutants for each class of segmentation genes. The phenotypes for loss-of-function mutation in the maternal effect gene, *Bicaudal D* (*BicD*), gap gene, *Kr*, pair-rule gene, *fushi tarazu* (*ftz*) and segment polarity gene, *gooseberry* (*gsb*) are shown here (Nüsslein-Volhard & Wieschaus, 1980). This figure has been adapted from Sivanantharajah, 2013 with permission (Appendix A).

## 1.4 The *Homeotic Selector (Hox)* Genes

*Homeotic selector (Hox)* genes are developmental control genes involved in establishing the identity of body segments along the A-P axis of Bilaterans (reviewed in Akam, 1998; Lewis, 1978). In 1894, Bateson coined the term ‘homeosis’ which refer to the transformation of one body part into the likeness of another (Bateson, 1894). An example of an insect homeosis is the observation of natural variants with a leg in place of an antenna. In 1915, Calvin Bridges isolated a spontaneous mutation in *Drosophila*, dubbed *bithorax*, with a genetically heritable homeotic transformation in which wing tissue replaces part of the haltere (the posterior flight appendage in flies). *Hox* genes were identified by either finding spontaneous mutant flies with homeotic transformations, or screening flies for body plan changes that have been exposed to a mutagen (Jürgens *et al.*, 1984; reviewed in Akam, 1987). Homeotic transformations are caused by loss, or ectopic expression, of *Hox* genes during the process of development.

The Homeotic Complex (HOM-C) in *Drosophila* (Akam, 1989) located on chromosome 3R consists of 8 *Hox* genes (Lewis, 1978) which are divided into two gene clusters, the Antennapedia Complex (ANT-C) and the Bithorax Complex (BX-C) (Figure 2). The Antennapedia complex consists of 5 genes that are involved in determining head and thoracic segmental identity: *labial (lab)*, *proboscipedia (pb)*, *Deformed (Dfd)*, *Sex combs reduced (Scr)*, and *Antennapedia (Antp)* (Kaufman *et al.*, 1980; Scott *et al.*, 1983). The Bithorax complex consists of 3 genes that determine the abdominal identity of posterior segments: *Ultrabithorax (Ubx)*, *abdominal-A (abd-A)* and *Abdominal-B (Abd-B)* (Lewis, 1978; Sánchez-Herrero *et al.*, 1985). Although both the ANT-C and the BX-C are located on chromosome 3, 10Mb of DNA separates them in *Drosophila*; however, in other insects, all the *Hox* genes are found in a single HOM-C (Beeman, 1987; Ferrier & Akam, 1996; Devenport *et al.*, 2000; Powers *et al.*, 2000). The order of the *Hox* genes on the chromosome corresponds to the rostral (head) to caudal (rear) order of the segments in which they are expressed (Lewis, 1978; Harding *et al.*, 1985). For instance, *lab*, the first *Hox* gene of the HOM-C, is expressed in the anterior-most segment, and *Abd-B*, the terminal *Hox* gene of the complex, is expressed in the posterior-most segment. This pattern of expression of the *Hox* genes is known as ‘spatial collinearity’ (Lewis, 1985;

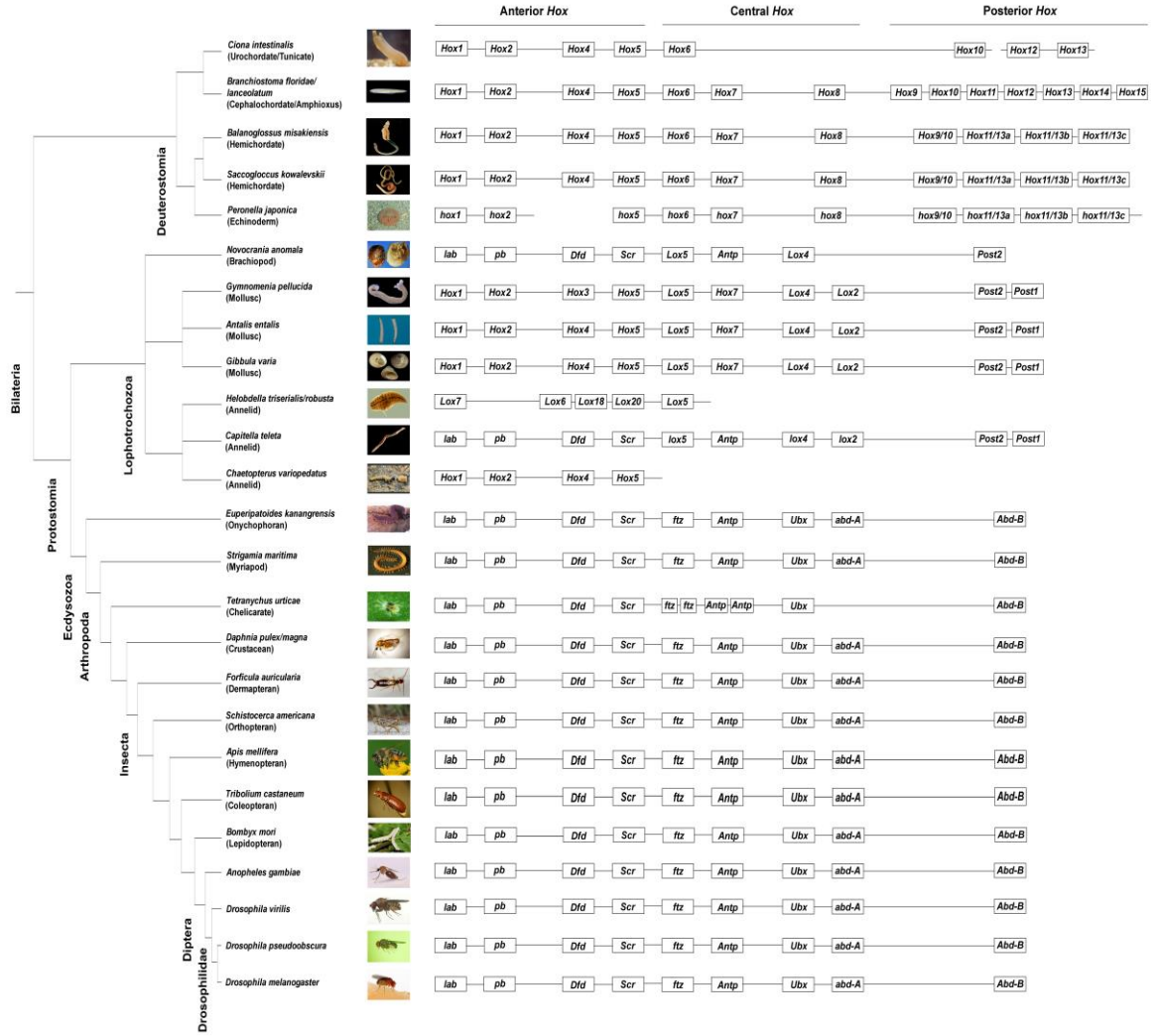
Duboule, 2007; reviewed in Gaunt, 2015).



**Figure 2. Homeotic complex in *D. melanogaster* and the expression of *Hox* genes in the developing embryo.** Eight *Hox* genes, *lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abd-A* and *Abd-B* constitute the *D. melanogaster* homeotic complex. Each gene and the region in the developing embryo where the gene is expressed is shown with a particular color. This image has been adapted from McGinnis & Kuziora, 1994 and has been used with permission from the creator of the image, Tomo Narashima (Appendix B).

*Hox* genes are one of the most important set of toolkit genes. Toolkit genes are a small set of genes in the genome that function as determinants of the body plan of an organism (Carroll *et al.*, 2005). Toolkit genes share a common set of characteristics. Most toolkit genes encode transcription factors or signaling molecules. The spatial and temporal expression of the toolkit genes correlates with the regions of the animal where the genes

function. Toolkit genes are highly conserved across animal species. The *Hox* genes provide an example of extensive conservation. Primitive *Hox* clusters are found in the diploblastic Cnidarian lineage, but not in Ctenophores, Placozoa or Porifera lineages (Murtha *et al.*, 1991; Schierwater *et al.*, 1991; Schummer *et al.*, 1992; Finnerty & Martindale, 1999). A gene cluster called *ParaHox* is thought to be an evolutionary sister of the *Hox* gene cluster (Brooke *et al.*, 1998) and exists in Cnidarians and Bilaterans (Finnerty & Martindale, 1999). Both *Hox* and *ParaHox* gene clusters are thought to have diverged from a *ProtoHox* cluster which probably existed in the Bilateran-Cnidarian ancestor (reviewed in Ferrier & Holland, 2001). *Hox* gene clusters are highly conserved across Bilaterans, from protostomes to deuterostomes (Figure 3). The pattern of *Hox* gene expression along the A-P axis is conserved among many animal groups (Negre *et al.*, 2005). The spatial collinear expression of *Hox* genes observed in *Drosophila* (Duboule, 2007) is also observed in many animal phyla (Barnett & Thomas, 2013; Hughes & Kaufman, 2002; Fröbius *et al.*, 2008) including deuterostomes like echinoderms and cephalochordates (Monteiro & Ferrier, 2006; reviewed in Gaunt, 2015). The requirement of *Hox* genes in determining segmental identity is conserved among Bilaterans. Loss-of-function mutations in vertebrate *Hox* genes result in homeotic transformations, which suggests the conservation of requirement of *Hox* genes. Both viable *Hoxa4* and *Hoxa5* (orthologs of *Drosophila Scr*) homozygous mutant mice show homeotic transformation where the 7<sup>th</sup> cervical vertebra is transformed to thoracic vertebrate identity (Horan *et al.*, 1994; Jeannotte *et al.*, 1993). The functions of the protein encoded by *Hox* genes are highly conserved in animals. Ectopic expression of mice and human *Hox* orthologs in *Drosophila* induce the same segmental transformations as their *Drosophila* counterparts (Zhao *et al.*, 1993; Malicki *et al.*, 1990; McGinnis *et al.*, 1990). Thus, the *Hox* genes across different animal phyla are conserved at the levels of structure, requirement and function.





**Figure 3. *Hox* cluster in Bilaterans.** The composition of the *Hox* clusters of insects (Akam, 1989), *Drosophila melanogaster*, *Drosophila pseudoobscura*, *Drosophila virilis*, *Anopheles gambiae*, *Bombyx mori*, *Tribolium castaneum*, *Apis mellifera*, *Schistocerca americana*, *Forficula auricularia*, non-insect arthropods, *Daphnia* sp. (Kim *et al.*, 2018), *Tetranychus urticae* (Barnett & Thomas, 2013; Pace *et al.*, 2016), *Strigamia maritima* (Hughes & Kaufman, 2002), non-arthropod ecdysozoan, *Euperipatoides kanangrensis* (Janssen *et al.*, 2014), lophotrochozoan annelids, *Chaetopterus variopedatus* (Irvine & Martindale, 2000), *Capitella teleta* (Fröblius *et al.*, 2008), *Helobdella* sp. (Kourakis *et al.*, 1997; Kourakis & Martindale, 2001), molluscs, *Gibbula varia* (Samadi & Steiner, 2010), *Antalis entails* (De Oliveira *et al.*, 2016), *Gymnomenia pellucida* (De Oliveira *et al.*, 2016), brachiopod, *Novocrania anomala* (Schiemann *et al.*, 2017), deuterostomes, *Peronella japonica* (Tsuchimoto & Yamaguchi, 2014), *Saccoglossus kowalevskii* (Freeman *et al.*, 2012), *Balanoglossus misakiensis*, *Branchiostoma* sp. (Pascual-Anaya *et al.*, 2012) and *Ciona intestinalis* (Ikuta & Saiga, 2005) are shown. Each of the *Hox* homologs are aligned illustrating that gene duplication and gene loss has occurred in some species. The anterior, central (Hueber *et al.*, 2013) and posterior *Hox* genes have been indicated at the top. The phylogenetic tree on the left is not drawn to scale and it merely depicts the evolutionary relationship among the organisms and not evolutionary time of divergence (Seetharam & Stuart, 2013; Ishiwata *et al.*, 2011; Meusemann *et al.*, 2010; Swalla & Smith, 2008).

## 1.5 The Pair-rule gene, *fushi tarazu* and its Gene Product

Fushi tarazu (FTZ) is a transcription factor that is encoded by the pair-rule gene, *fushi tarazu* (*ftz*) in *D. melanogaster*. *ftz* gene expression begins at the late syncytial blastoderm stage and continues in the cellular blastoderm (Karr & Kornberg, 1989). After gastrulation, FTZ is expressed in the central nervous system where it is required for the morphogenesis of neuronal cells (Doe *et al.*, 1988; Hiromi & Gehring, 1987). Like HOX proteins, FTZ has a conserved DNA-binding domain called the Homeodomain (HD) which is encoded by a 180 bp DNA sequence known as the homeobox (McGinnis *et al.*, 1984; Laughon & Scott, 1984). *ftz* is located in the ANT-C *Hox* cluster on Chromosome 3R. The FTZ ancestor was a HOX protein but, with evolutionary time, lost its homeotic function due to the loss of the YPWM motif, and gained a role in segmentation with addition of the LXXLL motif. In the Coleopteran insect, *Tribolium castaneum* (red flour beetle), FTZ has retained its homeotic function and is capable of generating homeotic transformations in larvae and adults (Löhr *et al.*, 2001; Löhr & Pick, 2005).

FTZ is required for the formation of the even-numbered parasegments (Hafen *et al.*, 1984; Frasch & Levine, 1987; Lawrence & Johnston, 1989). Homozygosity for null mutations in *ftz* is lethal during embryogenesis (Wakimoto *et al.*, 1984). The larvae only develop denticle belts for the odd-numbered parasegments T1, T3, A2, A4, A6 and A8 (Jürgens *et al.*, 1984; Wakimoto & Kaufman, 1981; Weiner *et al.*, 1984). Ectopic, ubiquitous expression of *ftz* in the cellular blastoderm has also been found to be lethal as only the even-numbered parasegments T2, A1, A3, A5 and A7 develop in these embryos (Struhl, 1985; Ish-Horowicz *et al.*, 1989). The pair-rule protein, Even skipped (EVE) is required for the formation of the odd-numbered parasegments (Frasch & Levine, 1987).

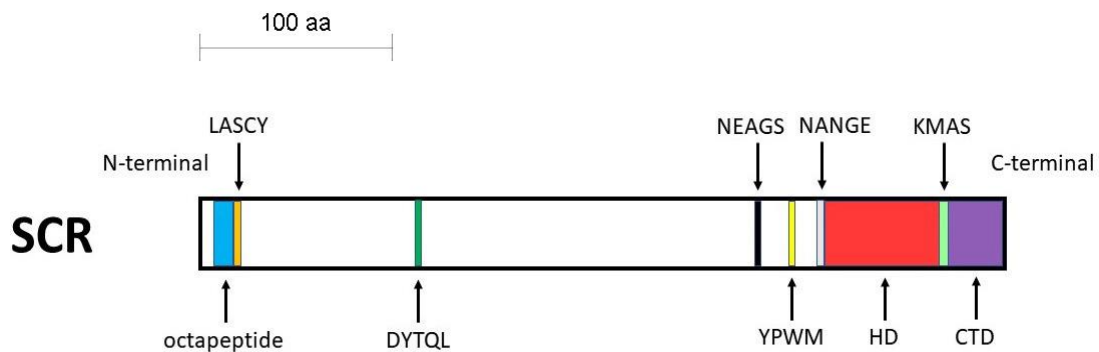
## 1.6 Structure and function of HOX proteins

HOX proteins determine segmental identity along the A-P axis of a developing embryo. HOX proteins are transcription factors that regulate expression of target genes by binding to DNA-binding sites with a 60 amino acid DNA-binding HD which is highly conserved

among the 8 HOX proteins in *Drosophila* (McGinnis *et al.*, 1984). HOX regulated, or downstream, genes generally code for either transcription factors or signaling molecules which play a major role in the formation of segment-specific structures (Hombria & Lovegrove, 2003; Pearson *et al.*, 2005). For instance, a HOX downstream gene, *decapentaplegic* (*dpp*), encoding a member of the Transforming Growth Factor  $\beta$  (TGF $\beta$ ) family of proteins, is positively regulated by UBX and negatively regulated by ABD-A for the formation of the second midgut constriction in *Drosophila* (Capovilla *et al.*, 1994). SCR controls the expression of the target salivary gland gene, *fork head* (*fh*), which encodes a TF (Andrew, 1998; Panzer *et al.*, 1992).

HOX TFs have a variety of domains and motifs that are important for function. The highly conserved, 60 amino acid DNA-binding HD (McGinnis *et al.*, 1984) is essential for HOX function. In addition, there are other small motifs in a HOX protein that are structurally and functionally important. Multiple sequence alignment of homologs of a HOX protein, SCR, identified motifs that are evolutionarily conserved at various taxonomic levels: Bilateran-specific, Deuterostome-specific, Protostome-specific, Lophotrochozoan-specific, Ecdysozoan-specific, Arthropod-specific, Insect-specific, Diptera-specific and *Drosophila*-specific (Sivanantharajah, 2013; Sivanantharajah & Percival-Smith, 2015). SCR, which is essential for the formation of larval salivary glands, adult proboscis and adult prothoracic legs (Lewis *et al.*, 1980; Struhl, 1982; Panzer *et al.*, 1992; Percival-Smith *et al.*, 1997; Percival-Smith *et al.*, 2013), contains 9 highly conserved functional regions – the N-terminal octapeptide motif (MSSYQFVNS), LASCY motif, DYTQL motif, NEAGS motif, YPWM motif, NANGE motif, HD, KMAS motif and the C-terminal domain (CTD) (Ryoo & Mann, 1999; Sivanantharajah, 2013; Tour *et al.*, 2005; Joshi *et al.*, 2007; Sivanantharajah & Percival-Smith, 2009; Percival-Smith *et al.*, 2013, Sivanantharajah & Percival-Smith, 2014; Sivanantharajah & Percival-Smith, 2015) (Figure 4). Sequencing of multiple *Scr* mutant alleles led to the identification of amino acid changes in most of the conserved sequences of SCR (Sivanantharajah, 2013; Sivanantharajah & Percival-Smith, 2009; Sivanantharajah & Percival-Smith, 2014). In addition, the mutations in highly conserved domains and motifs of HOX proteins, SCR, ANTP, UBX and ABD-A exhibit differential pleiotropy, which is

the non-uniform effect of mutant alleles across tissues, suggesting that the conserved domains and motifs in HOX proteins 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). SCR is proposed to have two different activities: SCR<sup>lab</sup> activity which determines the embryonic and adult labial identity and SCR<sup>T1</sup> activity which determines larval T1 and adult tarsus identity (Percival-Smith *et al.*, 1997; Percival-Smith *et al.*, 2013). Both SCR and PB are required for the determination of the adult labial structure, the proboscis (Percival-Smith *et al.*, 1997). PB is thought to mediate the switch from SCR<sup>T1</sup> to SCR<sup>lab</sup> activity during metamorphosis (Percival-Smith *et al.*, 2013). The LASCY, DYTQL and NANGE motifs and the CTD of SCR play a role in suppressing ectopic adult proboscis determination (Percival-Smith *et al.*, 2013). The short, conserved motifs found in HOX proteins may also be Short Linear Motifs (SLiMs).



**Figure 4. 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.

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). Many amino acid positions in SLiMs are degenerate, which means that there is flexibility in the amino acid allowed at those positions. SLiMs are required for cellular signaling and protein regulation. Although it was initially thought that large, ordered protein domains are solely responsible for determining specificity of a protein, many studies have shown that eukaryotic proteins use disordered regions with SLiMs to achieve functional specificity (Dyson & Wright, 2005; Tompa *et al.*, 2009; Gould *et al.*, 2010; Davey *et al.*, 2012). SLiMs function as sites of protein-protein interaction, PTMs and also as cell compartment targeting signals (Puntervoll *et al.*, 2003; Neduva *et al.*, 2005; Davey *et al.*, 2012).

SLiMs are conserved at various taxonomic levels. The short and degenerate nature of SLiMs make them evolutionarily plastic, i.e., they can appear and disappear during evolution (reviewed in Neduva & Russell, 2005). A single mutation in a SLiM can render the motif inactive or active which makes them to appear and disappear during evolution. Due to this flexibility, SLiMs arise independently in unrelated yet functionally similar proteins as a result of convergent evolution (Davey *et al.*, 2012).

SLiMs are sites of PTMs (Puntervoll *et al.*, 2003; Iakoucheva *et al.*, 2004; Khan & Lewis, 2005; Gould *et al.*, 2010; Dinkel *et al.*, 2016). Phosphorylation is one of the most important PTMs known for regulation of protein activity. One of the first studies showing phosphorylation in an unstructured region of a protein was reported in the year 2000 (Ojala *et al.*, 2000). Kaposi's sarcoma herpesvirus/human herpesvirus 8-encoded v-cyclin binds to the cellular cyclin-dependent kinase (CDK6) and this heterodimeric complex phosphorylates two serine residues (S70 and S87) of the cellular, anti-apoptotic/pro-growth protein, BCL-2, thus inactivating it and thereby, promoting apoptosis (Ojala *et al.*, 2000). The residues, S70 and S87 map to predicted SLiMs which are part of an unstructured loop of the cellular BCL-2 protein.

HOX proteins contain regions of intrinsic disorder, which are predicted to have SLiMs (Sivanantharajah & Percival-Smith, 2015; Dinkel *et al.*, 2016). In SCR, the highly conserved YPWM motif that confers HOX homeotic function (Löhr & Pick, 2005) is predicted to be a SLiM (Sivanantharajah & Percival-Smith, 2015). *Drosophila* HOX proteins, SCR and UBX interact with the HD-containing HOX cofactor, Extradenticle (EXD) via their YPWM motif, thus regulating target gene expression (Ryoo & Mann, 1999; Peifer & Wieschaus, 1990; Mann & Chan, 1996; Passner *et al.*, 1999; Merabet *et al.*, 2007; Joshi *et al.*, 2007). When the YPWM motif of the HOX protein interacts with EXD, the YPWM motif becomes ordered. The YPWM motif of ANTP also interacts with a TATA-binding protein associated factor, Bric-à-brac interacting protein 2 (BIP2) and this promotes the formation of wing tissues (Prince *et al.*, 2008). Another predicted SLiM in SCR, DISPK (Sivanantharajah & Percival-Smith, 2015) is phosphorylated at the serine residue (S216) (Zhai *et al.*, 2008; Figure 24), but the role of this motif in SCR activity is not known.

The aim of this study is to identify sites of phosphorylation in HOX and FTZ proteins. And because SLiMs are predicted sites of phosphorylation, it is important to determine whether the sites of protein phosphorylation are in SLiMs or not.

## 1.7 Phosphorylation of Homeodomain-containing Proteins

HOX and many other HD-containing proteins are phosphoproteins. Four different techniques have been used to determine whether *Drosophila* HOX and other HD-containing proteins are phosphorylated: 2-dimensional (2-D) gel analysis, yeast two-hybrid screen, *in vitro* kinase assay and genetic analysis.

2-D gel analyses (O'Farrell *et al.*, 1975; O'Farrell, 1977) determined that the HOX protein, UBX, HOX cofactor, EXD, pair-rule protein, FTZ and segment polarity protein, Engrailed (EN) are phosphorylated (Gavis & Hogness, 1991; Stultz *et al.*, 2006; Krause *et al.*, 1988; Krause & Gehring, 1989; Bourbon *et al.*, 1995). Phosphate addition to amino acids create a characteristic pattern of spots on a 2-D gel due to an increase in negative charge. UBX extracted from embryos as well as from cultured *Drosophila* cells are

phosphorylated at multiple sites (Gavis & Hogness, 1991). A Western Blot analysis showed six different spots with increasing negative charge and molecular weight which suggested phosphorylation of UBX isoforms (Gavis & Hogness, 1991). The HOX cofactor, EXD extracted from embryos was shown to be a phosphoprotein using 2-D gel analysis and corresponding Western Blots (Stultz *et al.*, 2006). It was also reported that nuclear-localized EXD is more extensively phosphorylated than the cytoplasmic EXD suggesting that EXD is involved in signal transduction upon phosphorylation by a protein kinase (Stultz *et al.*, 2006). Ectopically expressed FTZ extracted from developing embryos is post-translationally modified (Krause *et al.*, 1988) and the modifications were phosphorylation at as many as 16 sites (Krause & Gehring, 1989). Endogenously expressed FTZ extracted from developing embryo is also phosphorylated at multiple sites (Krause & Gehring, 1989). 2-D gel analysis of EN purified from *Drosophila* embryos (ectopic and endogenous) and embryo-derived cultured cells showed that EN is phosphorylated at a minimum of three sites. (Bourbon *et al.*, 1995). For all the above proteins, treatment with a phosphatase resulted in the disappearance of the spots on the 2-D gel, which suggested that the modifications are phosphorylation (Gavis & Hogness, 1991; Stultz *et al.*, 2006; Krause & Gehring, 1989; Bourbon *et al.*, 1995).

*In vivo* metabolic incorporation of radioactive orthophosphate,  $^{32}\text{PO}_4^-$  followed by 1-dimensional (1-D) gel and subsequent autoradiography and/or Western Blot also identified SCR, Bicoid (BCD) and EN as phosphoproteins (Berry & Gehring, 2000; Driever & Nüsslein-Volhard, 1989; Gay *et al.*, 1988). The N-terminal arm of the HD of SCR extracted from a mammalian cell-line is phosphorylated, and was dephosphorylated by treatment with a phosphatase (Berry & Gehring, 2000). Using *in vivo*  $^{32}\text{P}$  labeling, BCD expressed from a *Drosophila*-derived cell line was reported to be a phosphoprotein (Driever & Nüsslein-Volhard, 1989). This study was later extended by another group to show that embryonic BCD was phosphorylated by comparing the electrophoretic mobility of phosphatase-treated BCD and non-treated BCD on a 1-D gel (Ronchi *et al.*, 1993). EN expressed in the embryo and cultured cells were found to be phosphorylated (Gay *et al.*, 1988). Treatment with a phosphatase lowered the relative molecular mass ( $M_r$ ) of SCR and BCD on a 1-D gel which confirmed the modification as phosphorylation (Berry & Gehring, 2000; Driever & Nüsslein-Volhard, 1989; Ronchi *et al.*, 1993). *In vivo*

phosphorylation assays with  $^{32}\text{P}$  followed by 2-D gel and subsequent autoradiography and/or Western Blot identified UBX, FTZ, and EN as phosphoproteins (Gavis & Hogness, 1991; Krause & Gehring, 1989; Bourbon *et al.*, 1995).

The yeast two-hybrid screen was employed to identify HOX interacting proteins and indirectly suggest that ANTP and SCR are phosphoproteins (Jaffe *et al.*, 1997; Berry & Gehring, 2000). ANTP was found to interact with serine/threonine kinase casein kinase II (CKII) in the yeast two-hybrid assay, which suggested that ANTP might be phosphorylated (Jaffe *et al.*, 1997). In addition, genetic manipulation of predicted CKII phosphosites of ANTP suggested a role for these sites in regulating ANTP activity. Berry and Gehring showed that the SCR HD interacts with the Serine-threonine protein phosphatase 2A (PP2A,B'). PP2A,B' was proposed to positively regulate SCR function by dephosphorylating residues in the N-terminal arm of the SCR HD that when phosphorylated inhibited interaction of the SCR HD with DNA. However, subsequently, a null mutation in *PP2A,B'* was generated which did not affect SCR activity suggesting that dephosphorylation played no role in regulating SCR activity (Moazzen *et al.*, 2009).

To show that SCR, ANTP, UBX, FTZ, EN and BCD proteins are kinase substrates, *in vitro* kinase assays were performed (Berry & Gehring, 2000; Jaffe *et al.*, 1997; Gavis & Hogness, 1991; Dong *et al.*, 1998; Bourbon *et al.*, 1995; Janody *et al.*, 2000). The kinase incorporates radioactive phosphate,  $^{32}\text{PO}_4^-$  to its target amino acid residues, which can be detected by autoradiography. The N-terminal arm of SCR is phosphorylated by cAMP-dependent Protein Kinase A (PKA) (Berry & Gehring, 2000). *In vitro* kinase assay showed that ANTP is a substrate of CKII (Jaffe *et al.*, 1997). Bacterially-expressed UBX was found to be phosphorylated when mixed with a *Drosophila* cell extract (Gavis & Hogness, 1991). The T263 residue of the N-terminal region of FTZ HD is phosphorylated by *Drosophila* embryo extracts and cAMP-dependent PKA (Dong *et al.*, 1998). *In vitro* kinase assays showed that CKII or a CKII-related kinase phosphorylates EN at multiple amino acid residues between 392 to 409 (Bourbon *et al.*, 1995). *In vitro*, MAP kinase, a member of the Torso signaling pathway, phosphorylates the maternal effect protein, BCD at multiple S/T residues, S165, T188, T193, S195, T197, T200 and T353 which are predicted targets of MAP kinase activity (Janody *et al.*, 2000).



Genetic analysis that either mimics constitutive dephosphorylation by changing a phosphorylatable amino acid residue like S/T/Y to a non-phosphorylatable residue, alanine (A) or mimics constitutive phosphorylation by changing the phosphorylatable residue to a negatively charged, aspartic acid (D) or glutamic acid (E) residue was used to identify or functionally validate sites of phosphorylation in SCR, ANTP, FTZ, EN and BCD (Berry & Gehring, 2000; Jaffe *et al.*, 1997; Dong *et al.*, 1998; Bourbon *et al.*, 1995; Janody *et al.*, 2000). Site-directed alteration of both T and S residues at position 6 and 7 of the SCR HD, respectively to alanine significantly reduced PKA-mediated SCR phosphorylation *in vitro*, which suggested that the two sites were indeed targets of PKA-mediated phosphorylation, and that this non-phosphorylated form of SCR was active; whereas, the form of SCR mimicking constitutive phosphorylation was inactive (Berry & Gehring, 2000). CKII was unable to phosphorylate ANTP *in vitro* when four putative CKII targets were mutated to alanine suggesting that the four amino acid residues were indeed sites of CKII-mediated phosphorylation. Mutation of the four putative CKII targets in ANTP to alanine did not abolish the wild-type functions of ANTP *in vivo*, but additional functions, which are normally repressed for wild-type ANTP, were seen. However, genetic modification of the four putative CKII targets of ANTP to glutamic acid which mimics phosphorylation resulted in reduced ANTP function *in vivo* which suggested that dephosphorylation of ANTP was required for proper homeotic activity (Jaffe *et al.*, 1997). Mutating the T263 residue of the N-terminal region of the FTZ HD to alanine resulted in reduced <sup>32</sup>P labeling *in vitro* when FTZ was mixed with either embryo extracts or purified PKA. Mutating T263 to alanine resulted in the loss of FTZ activity *in vivo*; however, when phosphorylation of FTZ was mimicked by changing T263 to aspartic acid, FTZ activity was not affected suggesting that T263 of FTZ is required to be phosphorylated in the developing embryo for FTZ to function (Dong *et al.*, 1998). T263 phosphorylation plays a tissue-specific role: it is required for the formation of even-numbered parasegments but not for the development of central nervous system (Dong *et al.*, 1998; Dong & Krause, 1999). For bacterially-expressed EN, *in vitro* <sup>32</sup>P incorporation by CKII was reduced significantly when S394, S397, S401 and S402 were mutated to alanine and the DNA-binding capacity of mutated EN was significantly reduced (Bourbon *et al.*, 1995). Mutation of *in vitro* MAP kinase substrates of BCD, S165, T188,

T193, S195, T197, T200 and T353 to alanine abolished MAP-kinase-mediated phosphorylation of these residues in both cell culture and embryo. Phosphorylation of these 7 amino acid residues was required for the strengthening of BCD activity gradient by Torso RTK (Janody *et al.*, 2000).

The methodologies employed above do not directly determine which amino acid residue of a protein is phosphorylated in the developing *Drosophila* embryo. 2-D gel analysis suggests that a protein is phosphorylated, but it does not identify the amino acid that is phosphorylated. *In vivo* <sup>32</sup>P labeling of proteins also does not identify specific positions of phosphorylation. Yeast two-hybrid screens do not directly test whether a protein is truly phosphorylated or not; they only potentially suggest that the protein is phosphorylated because of an interaction with a kinase or a phosphatase. *In vitro* kinase assays show that a protein or peptide can be phosphorylated *in vitro* by either a factor present in the embryo extract or by a known kinase; however, the kinases used *in vitro* might not have the same specificity as *in vivo*, and therefore, the phosphorylation data obtained from *in vitro* kinase assays may be an artifact. Also, validation of the function of any phosphosite identified by *in vitro* kinase assays with genetic mimicry of constitutive phosphorylation/dephosphorylation may also be an artifact. Therefore, a direct way of determining which amino acid is phosphorylated on a HOX protein extracted from a developing embryo is needed. A method that does this is Tandem Mass Spectrometry (MS/MS).

## 1.8 Mass Spectrometry (MS)

Mass spectrometry (MS) is a popular proteomics technique used to identify and characterize organic and inorganic molecules (Whitelegge, 2009). Protein analysis and sequencing can be successfully carried out using MS. MS is also used as a tool to identify covalent modifications. The method of a single-step MS consists of 3 different steps: 1) ionization, i.e., generation of gaseous ions; 2) separation of the ions by mass to charge ratio ( $m/z$ ) in an analyzer; and 3) detection of the ions on a detector.

One of the commonly used ionization techniques is called Matrix-Assisted Laser Desorption/Ionization (MALDI) where the protein to be analyzed is first digested with a protease, generally, trypsin and is called the 'analyte' and this is mixed with a suitable matrix material and applied to a metal plate which is then exposed to a high energy laser that causes ablation and desorption of the analyte/matrix. Finally, the analyte molecules are ionized by being protonated or deprotonated in the hot plume of gases and can then be transported to a particular kind of mass spectrometer for separation of ions (reviewed in Singhal *et al.*, 2015).

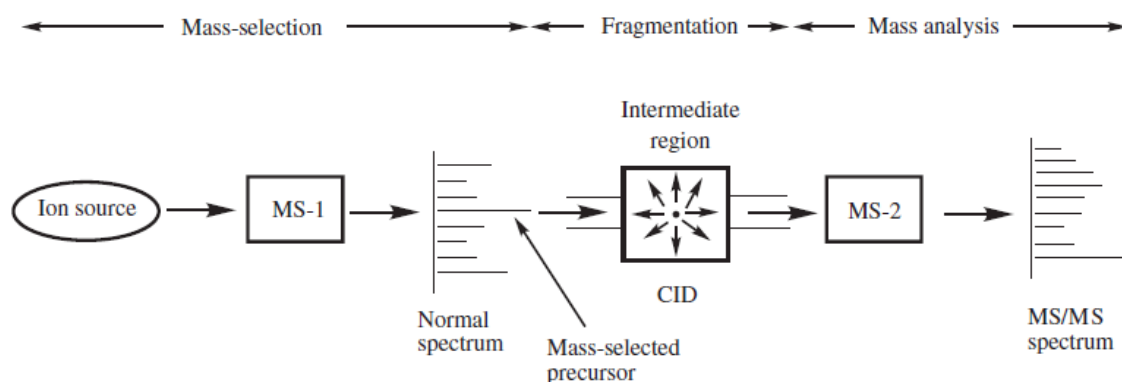
Another method of ionization is the Electrospray Ionization (ESI) technique, which is a soft technique and uses high voltage applied to a liquid to create aerosols (Fenn *et al.*, 1990). The analyte is dispersed by electrospray in a fine aerosol. The solvent for ESI are made by mixing water and volatile organic compounds like methanol, acetonitrile, etc. Compounds like acetic acid are added to the mixture to decrease droplet size by increasing conductivity and also acts as a source of protons. ESI is very useful for analysis of macromolecules as it overcomes the fragmentation of the macromolecules when ionized by techniques like MALDI. ESI also produces multiply charged ions unlike MALDI. The solution-phase information is also retained in the gas-phase in the ESI technique. Mass spectrometry using ESI is called Electrospray Ionization Mass Spectrometry (ESI-MS).

The ions formed by MALDI or ESI are separated by  $m/z$  ratio. One of the simplest techniques to do this is called Time-of-Flight (TOF). In this method, the ions are accelerated by an electric field of known strength. The  $m/z$  ratio determines how fast the ions will travel towards the detector. All ions of the same charge have the same kinetic energy, and therefore, lighter ions travel faster than heavier ions. The time it takes for an ion to travel a known distance is measured. In recent times, more sensitive and accurate ion trap techniques are commonly used to sort the ions according to their  $m/z$  ratios and one such instrument, Orbitrap has been used in this study (reviewed in Nolting *et al.*, 2017). Orbitrap is an ion trap mass analyzer which consists of two outer and a central electrode. It functions as both an analyzer and a detector. Different ions oscillate at different frequencies around the central electrode and in between the outer electrodes

which results in their separation. Oscillation frequencies of the ions result in the acquisition of the mass spectra of the ions.

Multi-step MS is often performed to increase the power of protein analysis. The first part of a two-step MS is same as the single-step MS which can involve either MALDI-TOF MS or ESI-TOF-MS or ESI-Orbitrap-MS. In addition to this, the peptides from the first mass spectrometer (MS-1) are fragmented into smaller components or ions. The fragmented ions are run through another mass spectrometer (MS-2) for obtaining the second spectrum (Figure 5).

The most common technique of fragmentation is Collision-Induced Dissociation (CID), which involves the collision of the peptides with a gas molecule (Wells & McLuckey, 2005). Prediction of fragmentation products based on known covalent bond strength is done which gives  $m/z$  values both before and after fragmentation. These values can be used to determine the identity and sequence of an unknown peptide. Liquid Chromatography (LC) can be coupled with MS as it allows the introduction of the peptides individually into the mass spectrometer. LC increases the purity of the peptide to be ionized (Whitelegge, 2009). In this study, LC-ESI-Orbitrap-MS/MS was used to detect and characterize post-translational modifications of tagged HOX proteins.



**Figure 5. Schematic of Tandem Mass Spectrometry (MS/MS).** The sample/analyte is introduced into the first mass spectrometer (MS-1) and a spectrum is obtained. Mass-selected precursor ion from the MS<sup>1</sup> spectra are fragmented and introduced into the second mass spectrometer (MS-2) which gives the MS<sup>2</sup> spectrum of the ion fragments. © 2012 Sindona G, Taverna D. Published in [short citation] under CC BY 3.0 license. Available from: <http://dx.doi.org/10.5772/52142> (Appendix C)

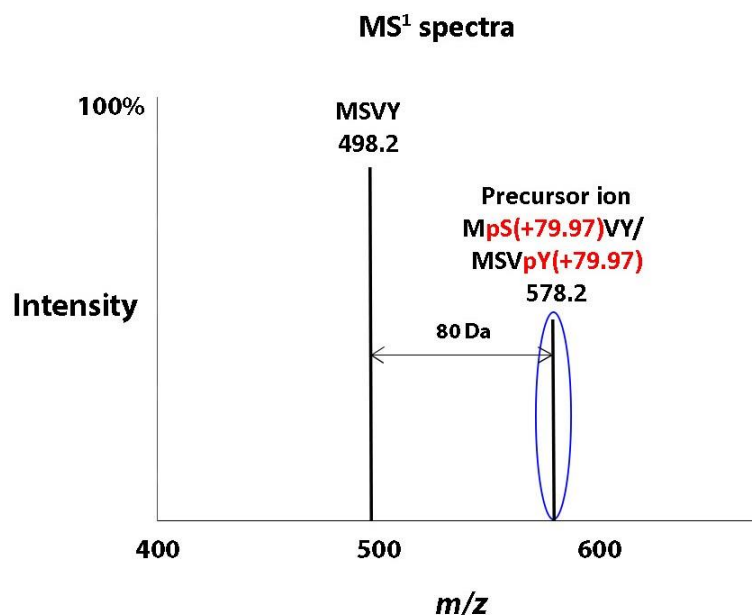
## 1.9 MS/MS as a tool to map exact sites of PTMs

MS/MS is a fundamental tool to identify PTMs of proteins. Protein phosphorylation is detected by an increase in amino acid residue mass of +80 Da, which is the addition of HPO<sub>3</sub> (Witze *et al.*, 2007). MS/MS is able to identify the exact amino acid residue bearing a covalent modification. Figure 6 is an example of how MS/MS is able to determine the exact site of phosphorylation in a peptide. For example, a mixture of three peptides (one unmodified and two phosphorylated) with the same primary amino acid sequence, MSVY, MpS(+79.97)VY and MSVpY(+79.97) were analyzed by a mass spectrometer. In MS<sup>1</sup> spectra, a peak was detected for the singly charged unmodified peptide which corresponded to its mass, 498.2 Da. A single peak was detected for both the singly charged phosphorylated peptides which corresponded to a mass of 578.2 Da (Figure 6A). The overall mass shift of the phosphorylated peptides from the unmodified peptide was 80 Da which suggested the attachment of a phosphate group to the peptide. However, a single-step MS is not able to determine which amino acid residue of the peptide is phosphorylated.

A two-step MS (MS/MS) is required to map the exact site of phosphorylation (Figure 6B). The precursor ion from MS-1 is fragmented at the peptide linkages, thereby resulting in the formation of b and y fragment ions. Assuming the precursor ion is MpS(+79.97)VY, the mass shift of b1 and b2 fragment ions and y2 and y3 fragment ions in the MS<sup>2</sup> spectra equaled the mass of a serine residue (87 Da) and a phosphate group (80 Da), thus assigning the phosphorylation at the serine residue. If the phosphorylation was on the tyrosine residue, the fragment ion profile in MS<sup>2</sup> spectra would have been

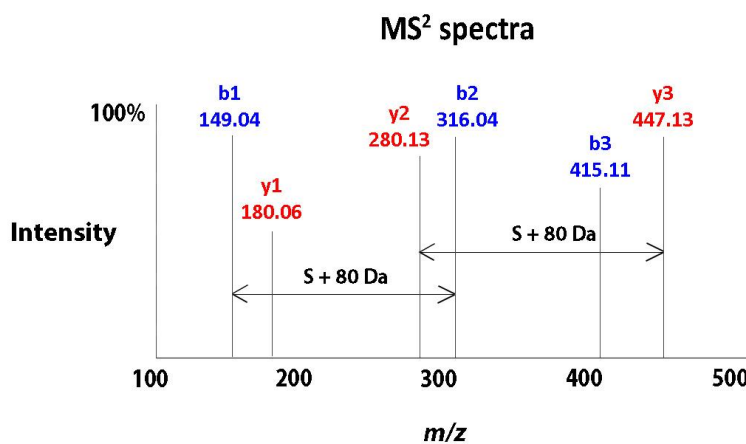
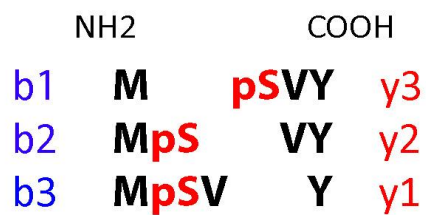
different. The objective of this thesis is to determine sites of PTMs in FTZ and HOX proteins using MS/MS.

A



B

Assume the precursor ion is MpS(+79.97)VY



**Figure 6. Mapping exact site of phosphorylation by tandem mass spectrometry (MS/MS) – an example.** A. MS<sup>1</sup> spectra of the peptide mixture is shown. The peptide sequence and  $m/z$  ratio are indicated at the top of the spectra. The precursor ion is indicated with a blue ellipse. B. MS<sup>2</sup> spectra of the fragmented precursor ion is shown. All the fragment ions and their  $m/z$  ratios are labelled in the spectra.

## 1.10 Results of MS/MS analysis of FTZ and HOX proteins to date

Dr. Dan Bath from our laboratory purified a tagged FTZ protein from *Drosophila* embryos under native conditions using anti-FLAG agarose beads and used MALDI-TOF MS to map six putative sites of phosphorylation: S107 or T108, T115 (within the FTZ-F1 binding site), T280 (within the HD) and Y399 or H402 (Bath, 2010). It was shown that phosphorylation of T115 abolishes FTZ activity (Bath, 2010).

Among the 8 HOX proteins in *Drosophila*, mass spectrometry data to determine exact phosphorylation sites exist for only UBX and SCR proteins. A *Drosophila* embryo phosphoproteome analysis showed that UBX is phosphorylated at T170, S174 and S177 and SCR is phosphorylated at S216 (Zhai *et al.*, 2008).

For mice, MS data showing exact sites of phosphorylation exists for HOXB6 (S214) and HOXD10 (S238 and S239) (Huttlin *et al.*, 2010). For human, MS data showing exact sites of phosphorylation exists for HOXB2 (S274), HOXB4 (S90), HOXB6 (S214), HOXB9 (T133), HOXC9 (S159), HOXC10 (T8 and S189) and HOXD9 (S146) (Olsen *et al.*, 2010; Matsuoka *et al.*, 2007; Zhou *et al.*, 2013).

## 1.11 Objectives

Although previous analysis of HOX protein has indicated that there may be a role for phosphorylation in the regulation of their activity, the lack of data on precisely which residues are phosphorylated is an important missing piece of information. The first

objective of this research study is to develop a comprehensive catalog of the sites of phosphorylation and other PTMs for FTZ and all 8 HOX proteins extracted from developing *Drosophila* embryos. The second objective is to test the hypothesis that SLiMs in FTZ and HOX proteins are preferential sites of phosphorylation.

## 2 Materials and Methods

### 2.1 *Drosophila* husbandry

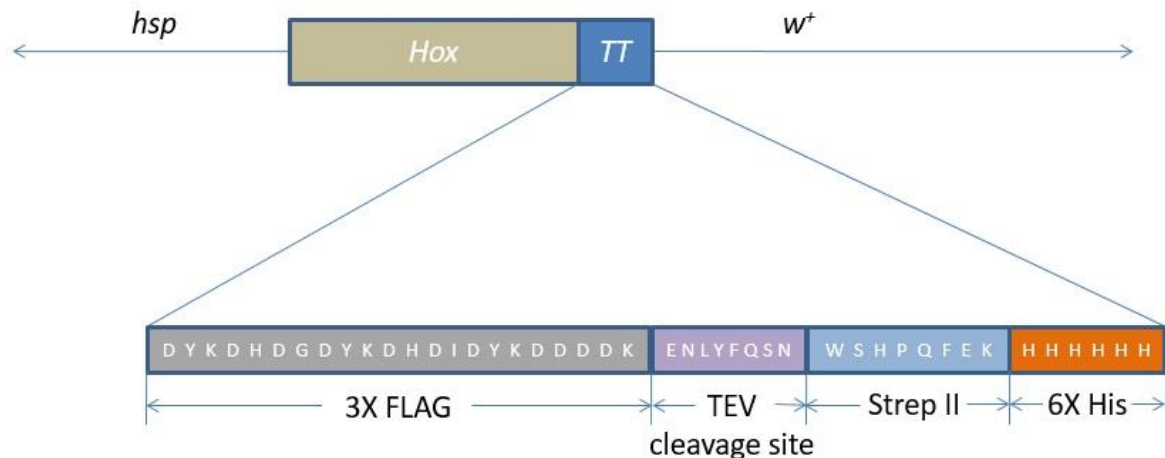
The fly stocks were maintained at 23°C and 60% humidity. The flies were grown in 20ml vials and ~300ml bottles containing corn meal food (1% (w/v) *Drosophila*-grade agar, 6% (w/v) sucrose, 10% (w/v) fine-ground cornmeal, 1.5% (w/v) yeast and 0.375% (w/v) 2-methyl hydroxybenzoate – an anti-fungal agent). 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) which promotes egg production.

### 2.2 Generation of *ftz* and *Hox* fusion constructs

Constructs expressing FTZ and HOX fused to the triple tag (TT) (containing 3X FLAG, Strep II and 6X His tags) (Figure 7) (Tiefenbach *et al.*, 2010) from the heat-shock promoter were constructed. Dr. Dan Bath assembled the fusion construct, *ftzTT* expressing FTZ protein fused to the TT from the heat-shock promoter (Bath, 2010). I assembled the constructs expressing HOX proteins fused to the triple tag from the heat-shock promoter for the *Hox* genes, *lab*, *abd-A* and *Abd-B(m)*. Genomic DNA was isolated from each of the three *UAS Hox* fly lines with genotypes, *y w; P{UAS-*lab*, *w*<sup>+</sup>}* (Bloomington Stock # 7300), *y w; P{UAS-*abd-A*, *w*<sup>+</sup>}* (Bloomington Stock # 912) and *y w; P{UAS-*Abd-Bm*, *w*<sup>+</sup>}* (Bloomington Stock # 913), respectively (obtained from Bloomington *Drosophila* Stock Center, Indiana University, Bloomington, Indiana). 5-40

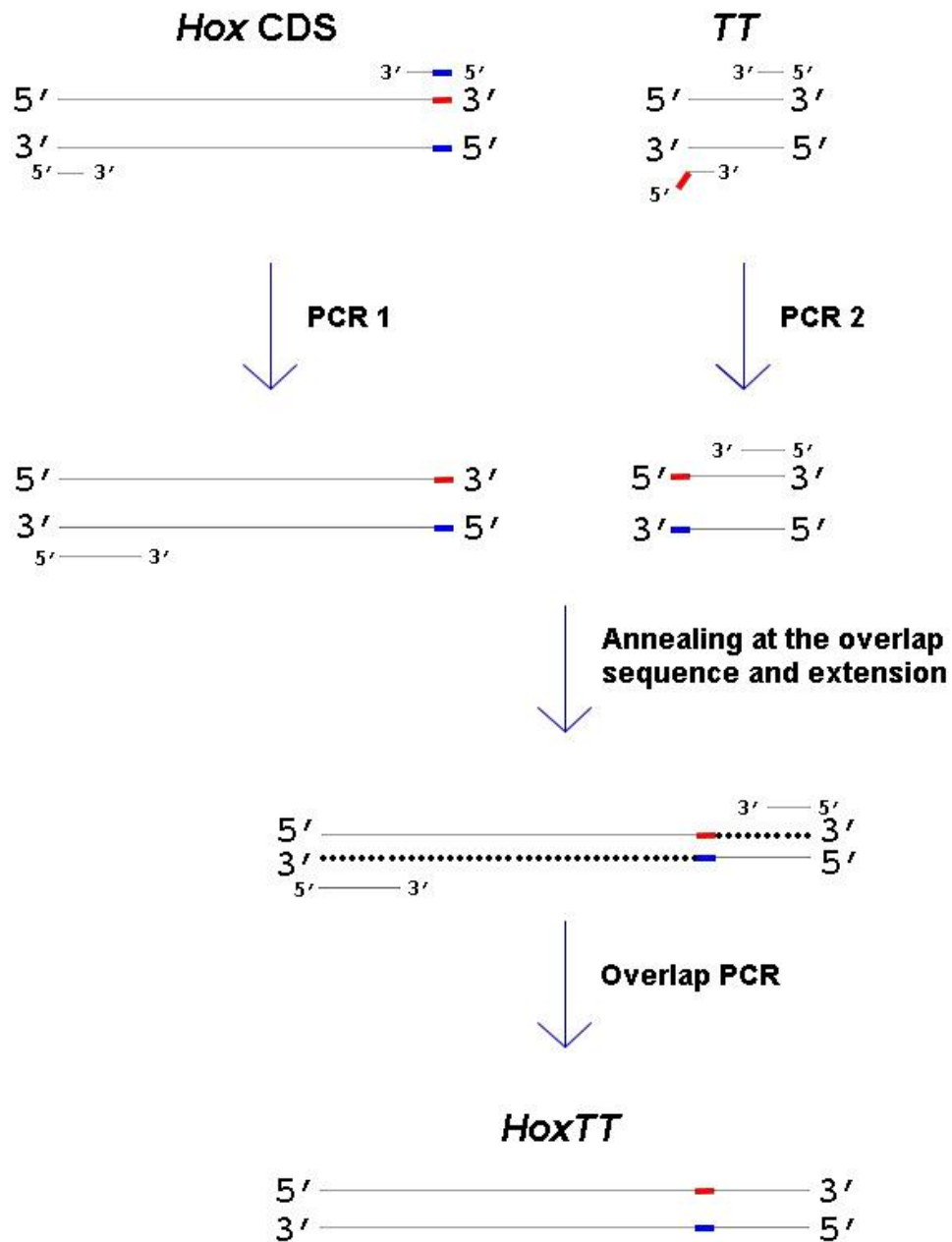


adult flies were homogenized in 300 $\mu$ l of freshly made isolation buffer (100mM Tris-HCl pH 9.5, 20mM EDTA, 1% SDS). Potassium acetate (8M) was added to the lysate for a final concentration of 1.12M followed by RNase A at a final concentration of 1 $\mu$ g/ml. After a brief vortex and 30 min of incubation on ice, the mixture was centrifuged at 14,000 rpm for 15 mins at 4°C. The supernatant was removed and mixed with 150 $\mu$ l of phenol and 150 $\mu$ l of chloroform and the mixture was centrifuged at 13,200 rpm for 7 mins at room temperature. The top aqueous layer was pipetted into a new tube, an equal volume of chloroform was added, mixed, and followed by centrifugation at 13,200 rpm for 4 mins. 600 $\mu$ l of anhydrous ethanol was added to the top aqueous layer and the mixture was incubated for 5 mins at room temperature. The precipitated DNA was collected by centrifugation at 13,200 rpm for 20 mins. The DNA pellet was washed with 400 $\mu$ l of 70% ethanol by gently pipetting in and then, pipetting out the ethanol without disturbing the pellet. The pellet was resuspended in 100 $\mu$ l MilliQ water, 250 $\mu$ l of anhydrous ethanol and 10 $\mu$ l of sodium acetate (3M, pH 5.2) for a final concentration of 83mM followed by a 2-3 min centrifugation at room temperature. The pellet was washed again with 400 $\mu$ l of 70% ethanol as before. After vacuum desiccation, the dry pellet was resuspended in 10mM Tris-HCl buffer at pH 8.



**Figure 7. Schematic of *HoxTT* fusion construct.** The *Hox* coding sequence represented by the grey box is fused in frame to a sequence encoding the triple tag (TT) represented by the blue box and this *HoxTT* fusion is inserted downstream of a heat-shock promoter (*hsp*).  $w^+$  is the *mini-white* marker which gives a red eye color phenotype. The arrowheads represent the inverted *P*-element repeats. The primary amino acid sequence of the C-terminal triple tag components, 3X FLAG, Strep II and 6X His are shown below. A TEV protease (Tobacco Etch Virus nuclear-inclusion-a endopeptidase) cleavage site is also part of the triple tag.

To create the triple tag fusions, three PCR reactions were performed. In the first PCR reaction, DNA extracted from *Drosophila* was used to amplify the *Hox* coding sequence (CDS). In the second PCR reaction, plasmid DNA carrying *UbxTT* (Percival-Smith, unpublished) was used to amplify the *TT*. The two DNA fragments were isolated and used in a third PCR to amplify a DNA fragment containing a fusion of the *Hox* coding sequence in frame to the *TT* (Shuldiner *et al.*, 1990; Shuldiner *et al.*, 1991; McPherson & Møller, 2000). This fusion occurs because the primers to the 3' end of the *Hox* sequence and 5' end of the *TT* have complementary sequence (Figure 8). The primers used for the three PCR reactions are tabulated in Table 1.



**Figure 8. Schematic of PCR-based construction of *HoxTT* fusion.** The 5' end of the forward primer for amplifying the *TT* has an additional *Hox* 3' sequence added which is indicated by the bent red box. The sequences indicated by the red and blue box are complementary.

All PCR steps were done using either Platinum *Taq* DNA Polymerase High Fidelity (Thermo Fisher Sci.) or Herculase II Fusion DNA Polymerase (Agilent Genomics). The PCR products were purified using EZ-10 Spin Column PCR Products Purification Kit (Bio Basic) and if needed, products were also extracted from a 1% (w/v) agarose gel and then, purified using a EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic) or QIAquick Gel Extraction Kit (Qiagen). To insert the fusion DNA fragments into the pCaSpeR vector (Thummel & Pirrotta, 1992), the restriction enzymes, *BglII* and *NotI* (New England Biolabs) were used to digest the *Hox* fusion DNA fragments at their 5' and 3' ends, respectively. For the *UbxTT* construct, which I inserted into the pCaSpeR vector, the 5' end was digested with *EcoRI*-HF and the 3' end with *XbaI* (New England Biolabs). The pCaSpeR vector was digested with the appropriate restriction enzyme. Ligation of the fusion DNA fragment with the digested vector used T4 DNA ligase (Thermo Fisher Sci.). The ligation mixture was transformed into Subcloning Efficiency DH5 $\alpha$  Competent *E. coli* Cells (Thermo Fisher Sci.). Transformed colonies were grown in culture overnight for plasmid isolation using a Miniprep Kit from Thermo Fisher Sci. The DNA sequence was confirmed at the DNA Sequencing Facility at the London Regional Genomics Centre, Robarts Research Institute, London, Ontario, Canada.

**Table 1. Primers used for PCR-based construction of *HoxTT* fusion**

<b>1) Amplification of <i>Hox</i> coding sequence</b>		
<b><i>Hox</i> gene amplified</b>	<b>Oligo name</b>	<b>Oligo sequence (5' to 3')</b>
<i>Lab</i>	Lab-1 (5')	CTTGCAAGATCTATGATGGACGTAAGCAGCATGTAC
	Lab-2 (3')	GTCTTTGTAGTCACTTTGCTTGCTCGTGGGTGC
<i>abd-A</i>	abdA-1 (5')	CTTGCAAGATCTATGTATCCGTACGTGTCTAAC
	abdA-2 (3')	GTCTTTGTAGTCGGAGTTGACTTTGCTGACCGC
<i>Abd-B(m)</i>	abdB fwd 2 (5')	CTTGCAAGATCTCGCCGCCGAGTTCGAGTGC
	abdB-2 (3')	GTCTTTGTAGTCCTGGTGCATCTTGGCGGCATG
<b>2) Adding a 3' <i>Hox</i> sequence to 5' end of <i>TT</i> and amplification</b>		
<b>Sequence amplified</b>	<b>Oligo name</b>	<b>Oligo sequence (5' to 3')</b>
<i>TT for lab</i>	Lab-3 (5')	AGCAAGCAAAGTGACTACAAAGACCATGACGG
	TTnot (3')	CTTACTGCGGCCGCCTAGTGATGGTGATGGTGATGC
<i>TT for abd-A</i>	abdA-3 (5')	AAAGTCAACTCCGACTACAAAGACCATGACG
	TTnot (3')	CTTACTGCGGCCGCCTAGTGATGGTGATGGTGATGC
<i>TT for Abd-B(m)</i>	abdB-3 (5')	AAGATGCACCAGGACTACAAAGACCATGACG
	TTnot (3')	CTTACTGCGGCCGCCTAGTGATGGTGATGGTGATGC

3) Overlap PCR to fuse <i>Hox</i> sequence with <i>TT</i> and amplification		
Construct	Oligo name	Oligo sequence (5' to 3')
<i>labTT</i>	Lab-1 (5')	CTTGCAAGATCTATGATGGACGTAAGCAGCATGTAC
	TTnot (3')	CTTACTGCGGCCGCCTAGTGATGGTGATGGTGATGC
<i>abd-ATT</i>	abdA-1 (5')	CTTGCAAGATCTATGTATCCGTACGTGTCTAAC
	TTnot (3')	CTTACTGCGGCCGCCTAGTGATGGTGATGGTGATGC
<i>Abd-B(m)TT</i>	abdB fwd 2 (5')	CTTGCAAGATCTCGCCGCCGCAGTTCGAGTGC
	TTnot (3')	CTTACTGCGGCCGCCTAGTGATGGTGATGGTGATGC

## 2.3 Generation of transgenic fly lines

The *P*-elements carried on all plasmid constructs were inserted into the genome of *Drosophila melanogaster* laboratory strain,  $y w^{67c23.2}$  via *P*-element-mediated transformation (Rubin and Spradling, 1982). pCaSpeR vector contains a *mini-white* ( $w^+$ ) marker such that germline insertion of the construct is detected by the red-eyed phenotype in the G1 generation. Flies homozygous for the  $P\{hsp-HoxTT, w^+\}$  insertion was established by standard crosses (Greenspan, 2004).

## 2.4 Heat-shock induction of HOXTT expression

To induce expression of HOXTT proteins from the heat-shock promoter for protein purification, the eggs/embryos at 4-8 hours AEL were collected from the apple juice plates on nylon mesh screens of a filter basket. The embryos were heat-shocked for 30 minutes at 37.5°C by immersion of the filter basket in a circulating water bath.

To observe the characteristic first instar larval cuticular phenotype induced by ectopic expression of HOXTT proteins from the heat-shock promoter, heat-shock was administered at different temperatures and at different timepoints specific for each *HoxTT* construct (Table 2). The nylon mesh containing embryos was carefully removed from the filter basket, folded, and inserted into a clean 1.7ml microfuge tube prior to the heat-shock. The tube was then immersed in a circulating water bath during the heat-shock treatment. After the heat-shock, the mesh with the embryos were kept on the apple juice plate for another 24 hours to allow development to first instar larval stage. For embryos requiring heat-shock at two different timepoints during development, the nylon mesh was kept inside the 1.7ml microfuge tube with its lid open for the period after the first heat-shock until the second.

**Table 2. Heat-shock timepoints for HOXTT expression for first instar larval cuticles**

HOXTT	Fly genotype	Heat-shock temperature (°C)	Heat-shock period	Heat-shock timepoint (AEL)
LABTT	<i>y w; P{hsp-labTT, w<sup>+</sup>}</i>	37.5	30 mins	2
PBTT	<i>y w; P{hsp-pbTT, w<sup>+</sup>}</i>	36.5	30 mins	4h
DFDTT	<i>y w; P{hsp-DfdTT, w<sup>+</sup>}</i>	36.5	1 hr	3.5h
SCRTT	<i>y w; P{hsp-ScrTT, w<sup>+</sup>}</i>	37.5	30 mins	5h
ANTPTT	<i>y w; P{hsp-AntpTT, w<sup>+</sup>}</i>	36.5	30 mins	6h
UBXTT	<i>y w; P{hsp-UbxTT, w<sup>+</sup>}</i>	35.5	1 hr	5h

ABD-ATT	$y w; P\{hsp-abd-ATT, w^+\}$	36.5	30 mins	6 & 8h (2X heat-shock)
ABD-B(m)TT	$y w; P\{hsp-Abd-B(m)TT, w^+\}$	36.5	30 mins	4 & 6h (2X heat-shock)
ABD-B(r)TT	$y w; P\{hsp-Abd-B(r)TT, w^+\}$	37	2 hrs & 1 hr*	5 & 8h (2X heat-shock)

\* The first heat-shock was administered at for 2 hours at 5h AEL and the second for 1 hour at 8h AEL.

## 2.5 GAL4 ectopic expression

To ectopically express the untagged or triple tagged *Hox* genes using the GAL4-*UAS* system (Brand & Perrimon, 1993), adult virgin female flies of the *UAS Hox* line with genotype,  $y w; P\{UAS-Hox, w^+\}$  or *UAS HoxTT* line with genotype,  $y w; P\{UAS-HoxTT, w^+\}$ , respectively were crossed with the GAL4 driver males with the genotype,  $y w; P\{Armadillo-Gal4, w^+\}$ . The fly lines used are tabulated in Table 3.

**Table 3. *UAS* fly lines used for GAL4 ectopic expression**

Untagged/triple tagged	Protein expressed	Fly genotype	Bloomington Stock # or Reference
	LAB	$y w; P\{UAS-lab, w^+\}$	7300
	PB	$y w; P\{UAS-pb, w^+\}$	7298



Untagged	DFD	$y w; P\{UAS-Dfd, w^+\}$	7299
	SCR	$y w; P\{UAS-Scr, w^+\}$	7302
	ANTP	$y w; P\{UAS-Antp, w^+\}$	7301
	UBX	$y w; P\{UAS-Ubx, w^+\}$	911
	ABD-B(m)	$y w; P\{UAS-Abd-Bm, w^+\}$	913
Triple tagged	FTZTT	$y w; P\{UAS-ftzTT, w^+\}$	Percival-Smith, unpublished
	SCR TT	$y w; P\{UAS-ScrTT, w^+\}$	Percival-Smith <i>et al.</i> , 2013
	UBX TT	$y w; P\{UAS-UbxTT, w^+\}$	Percival-Smith, unpublished

## 2.6 First instar larval cuticle preparation

First instar larval cuticles were prepared to examine segmental morphology. Larvae were allowed to develop until 24-28 hours AEL before dechoriation with 3% sodium hypochlorite solution. The dechorionated larvae were rinsed with deionized water and devittelinized with 1:1 v:v Heptane:Methanol. After two washes with methanol, the larvae were transferred to a clean glass slide and were allowed to dry. Then, the larvae were mounted in a 1:1 mixture of Hoyer's mountant and lactic acid (Wieschaus & Nüsslein-Volhard, 1986). The larval cuticles were observed with darkfield optics on a Leica® Leitz™ DMRBE microscope.

## 2.7 Affinity purification of HOXTT protein from embryos

To purify HOXTT protein from embryo extracts, subcellular fractionation followed by metal affinity chromatography was employed. Embryos for a transgenic HOXTT fly line were collected and heat-shocked at 37.5°C for 30 minutes at 0-8 or 0-14h AEL over many days and were stored at -80°C. 3g of the pooled embryos were homogenized in 15ml of lysis buffer (15mM HEPES pH 7.6, 10mM KCl, 5mM MgCl<sub>2</sub>, 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 top layer of the pellet (pellet 1) excluding the dark colored debris 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<sub>2</sub>PO<sub>4</sub>, pH 7.5, 300mM NaCl, 20mM imidazole, 1% NP-40, with protease inhibitors: 0.2mM phenylmethanesulfonylfluoride (PMSF), 1.3mM benzamidine and 0.3mM Aprotinin) and was centrifuged at 12,000 rpm for 10 minutes in a Sorval SS-34 rotor. All preceding steps were performed at 0-5°C. The nuclear pellet was discarded and the nuclear extract (NE) was mixed with solid urea, which was added to give 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 same 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 effective 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<sub>2</sub>PO<sub>4</sub>, 300mM NaCl. The beads were stored at -80°C. During all purification steps, aliquots were taken for subsequent analysis.

## 2.8 Expression of SCRTT in *E. coli*

A plasmid expressing SCRTT (Sivanantharajah, 2013) was transformed into chemically competent BL21(DE3) cells (Studier & Moffatt, 1986) and a transformant colony was inoculated in 3ml of Luria-Bertani (LB) growth medium (1L: 10g Tryptone, 5g Yeast extract, 5g NaCl and 10g D-glucose) containing 100µg/ml ampicillin (selectable marker for the plasmid). This culture was allowed to grow overnight at 37°C to stationary phase. The overnight 3ml cell culture was added to 250ml of fresh LB containing 100µg/ml ampicillin and shaken at 37°C until OD<sub>600</sub>=0.8 was reached. 1mM IPTG was added to the culture and was shaken at 30°C for approximately 5 hours to induce expression of SCRTT. The 250ml of induced cell culture was centrifuged in a plastic bottle at 12,000 rpm or lower for 10 minutes. The supernatant was discarded and the pellet was stored at -80°C.

## 2.9 Affinity purification of SCRTT from *E. coli*

The frozen *E. coli* cell pellet was resuspended in lysis buffer (8M urea, 50mM NaH<sub>2</sub>PO<sub>4</sub>, 300mM NaCl, 1% Triton X-100, and 3mM imidazole; pH 7-8) and was sonicated six times. Following sonication, lysed cells were centrifuged at 10,000 rpm for 10 minutes in a 30ml Corex tube in a Sorval SS-34 rotor. The pellet was discarded and 300µl of Ni-NTA sepharose beads that had been equilibrated in lysis buffer were added and shaken for 15 minutes to ensure maximum binding of 6X His to Ni beads. After shaking, 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 effective retention of the tagged protein in

the column. The beads were first washed with the lysis buffer with urea twice and then with the buffer (50mM NaH<sub>2</sub>PO<sub>4</sub>, 300mM NaCl, 0.5% Triton X-100, and 10mM imidazole; pH 7-8). The tagged protein was eluted with the elution buffer (50mM NaH<sub>2</sub>PO<sub>4</sub>, 300mM NaCl, 0.1% Triton X-100, and 750mM imidazole; pH 7-8). Eluate fractions (f) 2, 3 and 4 were pooled and were stored at -80°C. Aliquots were taken at all purification steps and all steps were performed at room temperature.

## 2.10 Protein separation and staining on a gel

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. Then, the protein gel was either stained with GelCode™ Blue Stain Reagent (ThermoFisher Sci.) or with 1 mg/ml Coomassie blue (Coomassie Brilliant Blue™ R-250 from ThermoFisher Sci.). In case of GelCode™ Blue Stain Reagent, distilled water was used to fix and destain the gel. For Coomassie Blue R250 staining, the gel was first fixed with 50% methanol and 10% glacial acetic acid for 5-10 minutes at RT prior to staining and then, was destained with 45% methanol and 10% glacial acetic acid until the protein bands were clearly visible. The protein gel was stored in 5% glacial acetic acid at 4°C. Digital images of the gel were taken using a ChemiDoc™ Imaging System (Bio-Rad).

## 2.11 Western Blot Analysis

After size separation using an SDS-polyacrylamide gel, the proteins 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). For the detection of triple-tagged proteins, anti-FLAG M2 monoclonal antibody (Sigma) 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 to remove unbound or non-specifically bound primary antibodies, the blot was probed with a horseradish peroxidase-conjugated antibody, goat anti-mouse HRP (ThermoFisher Sci.) at a dilution of 3,000-fold in Blotto for one hour at room temperature. After three rinses and three 15 minute washes with PBT, the HRP was detected using SuperSignal™ West Femto Maximum Sensitivity Chemiluminescent Substrate (ThermoFisher Sci.). Digital images of the blot were taken using a ChemiDoc™ Imaging System (Bio-Rad). For some experiments, equal loading needed to be ensured, and therefore, the membrane which had been probed for FLAG epitopes 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- $\beta$ -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.

## 2.12 Total protein quantification

Bio-Rad Protein Assay Dye Reagent Concentrate was used to estimate the total protein concentration in samples by the standard Bradford dye-binding method (Bradford, 1976). Bio-Rad's Standard Procedure for Microtiter Plates was used to perform the assay. Bovine Serum Albumin (BSA) was used as the standard.

## 2.13 Phosphopeptide enrichment and C18 desalting of non-phosphopeptides

Phosphopeptide enrichment of trypsinized  $\alpha$ -casein or SCRTT used the EasyPhos protocol which employs TiO<sub>2</sub> beads (Humphrey *et al.*, 2015). For the SCRTT sample, the TiO<sub>2</sub> 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 20% 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<sub>2</sub> flow-through was reduced to 100µl and 0.2% formic acid was added. 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 98% 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 20% 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.

## 2.14 Sample preparation for MS/MS

The Ni-NTA bead slurry (25µl) from the protein purification was mixed with equal volume of 2xSDS dye and was heated to approx. 90°C for 10 minutes. 50µl of the 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 either stained with GelCode™ Blue Stain Reagent (ThermoFisher Sci.) or with 1 mg/ml Coomassie blue (Coomassie Brilliant Blue™ R-250 from ThermoFisher Sci.). Once the gel was destained, it was stored in 5% glacial acetic acid at 4°C. Approximately 10 spots were picked from the desired band of the stained-gel using an Ettan® SpotPicker™ and were submitted for in-gel digestion with a protease (either trypsin, chymotrypsin or thermolysin) and subsequent lyophilization of the peptides at the Functional Proteomics Facility, Western University, London, ON. The lyophilized peptides of HOXTT proteins were submitted at the Biological Mass Spectrometry Laboratory, Western University, London, ON for protein identification and characterization of post-translational modifications using nano

LC-ESI-Orbitrap-MS/MS technique. The instrument, Thermo Scientific Orbitrap Elite mass spectrometer was used for this study. Collision-induced dissociation (CID), also known as collisionally activated dissociation (CAD) was used to fragment peptide ions. During low energy fragmentation, the C-N bond of the peptide linkage is susceptible to cleavage, and b ions are fragments that appear to extend from the N terminus and y ions are fragments that appear to extend from the C-terminus (reviewed in Steen & Mann, 2004).

## 2.15 Mass spectrometry data analysis

LC-ESI-Orbitrap-MS/MS data was analyzed by the Biological Mass Spectrometry Laboratory, Western University, London, ON. 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. The false discovery rate (FDR) was set at 1%. Coverage is given in the analysis as percent coverage. A quantitative measure of depth, which I refer to as Average Depth (AD), is to add up all the lengths of chemically distinct peptides identified and divide by the length of the protein. Since, the protease used often generate fragments too large or too small for analysis, maximum coverage may never reach 100%. To address this problem, Average Depth of regions covered ( $AD_{orc}$ ) is calculated, which is Average Depth divided by proportion of the protein covered. 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 higher peptide score and a lower P-value gives more confidence to the peptide match. Minimal ion intensity, which is the relative intensities of position-determining fragment ions in a  $MS^2$  spectrum was set to 5%. 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.

## 2.16 Bioinformatic analysis of proteomic data

*D. melanogaster* FTZ and HOX protein sequences (NCBI accession numbers in Table S23) were submitted in the ELM database (Dinkel *et al.*, 2016) to retrieve short linear motif (SLiM) sequences predicted for that protein. Since HOX transcription factors are known to interact with nuclear and cytoplasmic components (Merabet & Dard, 2014; Wielleite *et al.*, 1999), only the SLiMs for proteins localized to nucleus and cytoplasm were considered for the analysis. SLiMs that are part of intrinsically disordered regions of the protein were considered for the analysis. If a SLiM was comprised of amino acids from both ordered and disordered regions of a protein, it was excluded from the analysis. Amino acids 7 to 59 of the 60 amino acid homeodomain of all HOX proteins are ordered/structured (Billeter *et al.*, 1990). For FTZ, amino acids 8 to 53 of the 60 amino acid homeodomain is ordered/structured (Qian *et al.*, 1994). To determine whether a SLiM was conserved, HOX protein sequences of various protostome and deuterostome species across different phyla, were retrieved from NCBI or ORCAE (only for *T. urticae*) database (accession numbers in Table S23) 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 a HOX protein (except for ABD-B(r)) was manually checked for conservation across species (Figures S4-12). A SLiM was considered to be conserved only if they aligned perfectly in both MAFFT and Clustal Omega. SLiMs less than five amino acids long were not considered as a conserved SLiM unless conserved beyond Diptera.

## 2.17 Statistical Analysis

To investigate whether the distribution of serine (S), threonine (T) and tyrosine (Y) in FTZ and HOX proteins was biased to SLiMs vs. non-SLiMs and whether the phosphate distribution in HOX SLiMs vs. non-SLiMs were significant, Fisher's Exact Test was employed (Fisher, 1922).



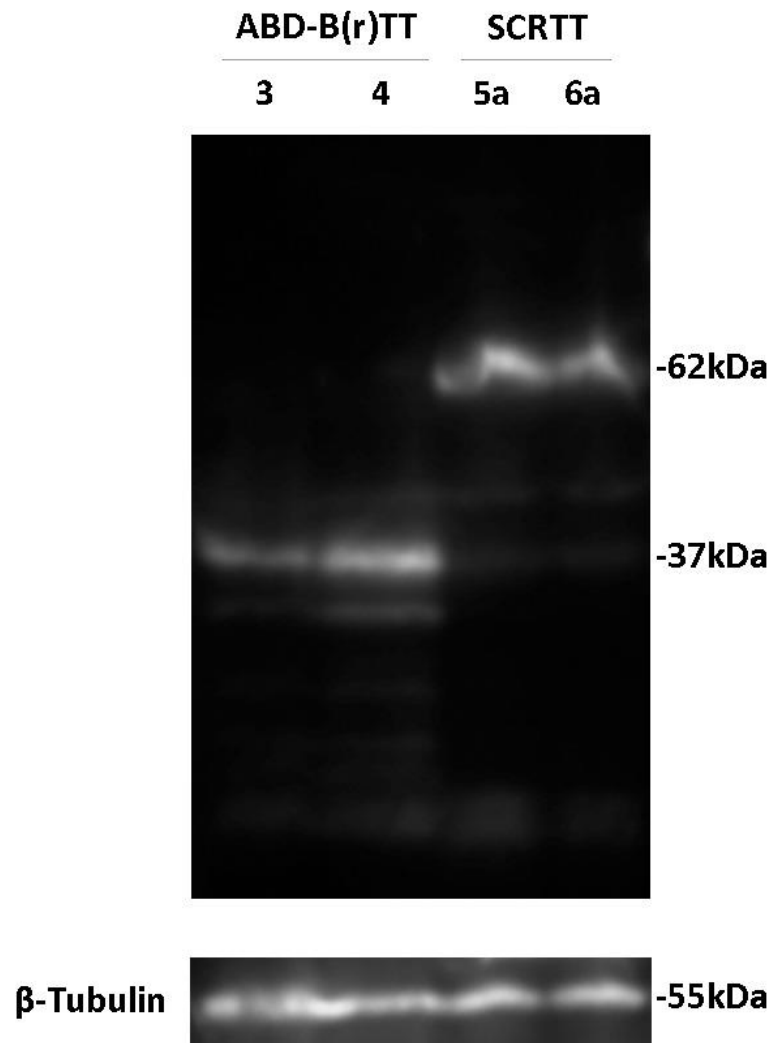
## 3 Results

### 3.1 Establishing transgenic *Drosophila* that express HOXTT proteins from a heat-shock promoter

To detect and purify HOX and FTZ proteins, the coding sequence (CDS) of these genes were fused in frame to CDS of the triple tag (TT) encoding 3X FLAG, Strep II and 6X His tags (Tiefenbach *et al.*, 2010; Figure 7). The FLAG and His tags of the C-terminal TT allow for easy detection and purification of HOXTT proteins. These gene fusions were cloned behind the heat-shock promoter (*hsp*) of the pCaSpeR vector (Thummel & Pirrotta, 1992). I cloned 4 *HoxTT* constructs, *labTT*, *UbxTT*, *abd-ATT* and *Abd-B(m)TT* for the expression of LABTT, UBXTT, ABD-ATT and ABD-B(m)TT proteins, respectively. Dr. Dan Bath made the *ftzTT* construct for the expression of FTZTT (Bath, 2010). Dr. Anthony Percival-Smith made the constructs, *pbTT*, *DfdTT*, *ScrTT*, *AntpTT* and *Abd-B(r)TT* for expression of PBTT, DFDTT, SCR TT, ANTP TT and ABD-B(r)TT, respectively (Percival-Smith *et al.*, 2013 for constructs expressing SCR TT and PBTT; the rest Percival-Smith, unpublished data). The protein isoforms expressed from the constructs are FTZ isoform A (410 aa) (FlyBase ID FBpp0081139), LAB isoform A (629 aa) (FlyBase ID FBpp0081194), PB isoform B (777 aa) (FlyBase ID FBpp0088334), DFD isoform A (586 aa) (FlyBase ID FBpp0081138), SCR isoform A (417 aa) (FlyBase ID FBpp0081163), ANTP isoform I (378 aa) (FlyBase ID FBpp0089245), UBX isoform A (389 aa) (FlyBase ID FBpp0082793), ABD-A isoform A (330 aa) (FlyBase ID FBpp0082828), ABD-B isoforms B or M (long morphogenetic isoform ‘m’ – 493 aa) (FlyBase ID FBpp0082826) and A or R (short regulatory isoform ‘r’ – 270 aa) (FlyBase ID FBpp0082824) (Casanova *et al.*, 1986). The ten constructs were introduced into the genome of *y1Df(1)67c23.2* (*y w*) flies with P-element mediated transformation by Dr. Anthony Percival-Smith. Multiple independent germ line transformants were established for each HOXTT construct. I screened the HOXTT transformant lines for maximum heat-shock dependent HOXTT expression. For each of the nine transformant lines, I selected the fly lines expressing the most HOXTT protein by comparing the amount of HOXTT protein expressed from independent transformant lines for a particular construct in a

Western Blot. Figure 9 is the comparison of two independent transformant lines of ABD-B(r)TT and SCR TT.

ABD-B(r)TT transformant line 4 expressed 1.8-fold more tagged protein than ABD-B(r)TT line 3. SCR TT transformant line 5a expressed 1.2-fold more tagged protein than SCR TT line 6a. I chose ABD-B(r)TT line 4 and SCR TT line 5a for further analysis. Dr. Dan Bath screened transformant lines for maximum expression of FTZTT (Bath, 2010). In summary, a transgenic fly line expressing each HOXTT protein was identified for analysis. The selected fly lines were expanded, and developing embryos were collected in a large scale and heat-shocked to induce expression of HOXTT proteins for affinity purification.

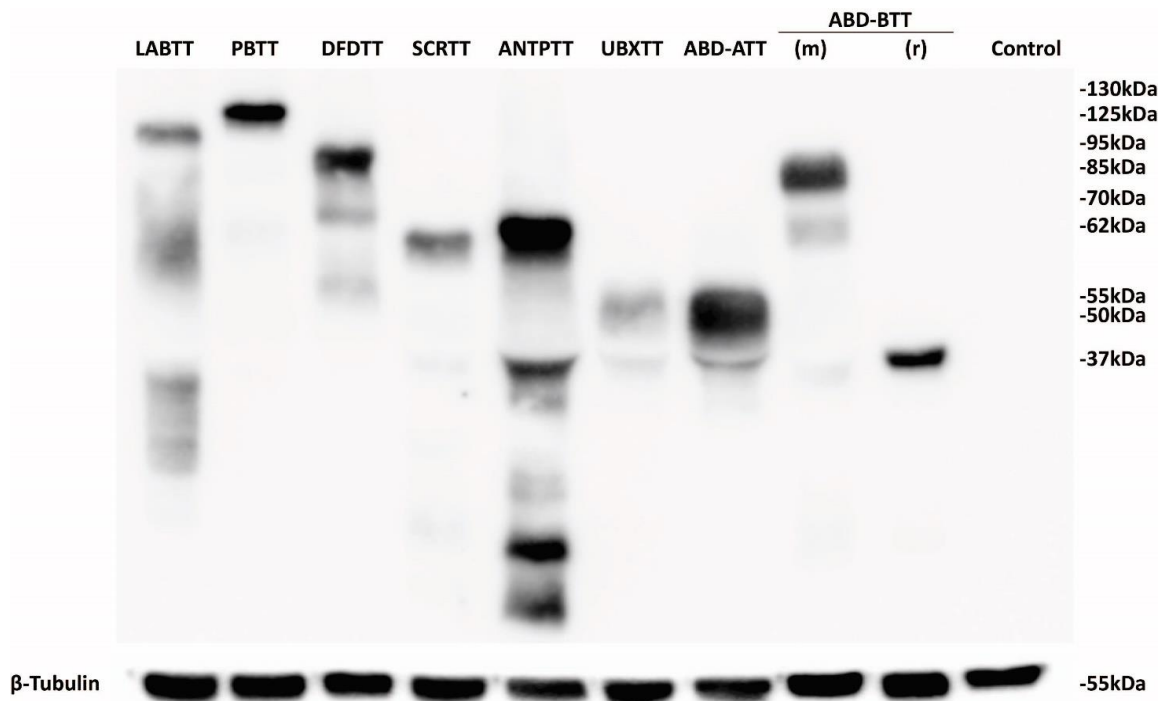


**Figure 9. Selecting the transgenic fly line expressing maximum HOXTT protein.**

Equal amounts of heat-shocked embryos from each independent transformant fly line expressing ABD-B(r)TT and SCR TT proteins were homogenized and loaded onto a 1-D SDS-polyacrylamide gel. HOXTT proteins were detected by Western Blot technique using anti-FLAG antibody. The transformant line number and the protein they are expressing is indicated on top of the gel. The position of protein molecular weight markers is indicated on the right. Along the bottom is the  $\beta$ -tubulin loading control detected using anti- $\beta$ -tubulin antibody.

### 3.2 Expression of HOXTT proteins from the heat-shock promoter

Using the pI/Mw tool of ExPASy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics; Artimo *et al.*, 2012), the relative molecular masses ( $M_r$ ) of HOXTT proteins were predicted (Table 4). Dr. Dan Bath determined the predicted relative molecular mass of FTZTT to be 55 and using a Western Blot analysis, a higher relative molecular mass of FTZTT of 70 was observed (Bath, 2010). To test whether the identified transformant fly lines expressed the HOXTT proteins of the predicted size, developing embryos were heat-shocked, homogenized and loaded onto a 1-D SDS polyacrylamide gel for size separation and subsequent detection of the tagged protein using Western Blot analysis. HOXTT proteins, with the exception of ABD-B(r)TT, ran at a higher relative molecular mass than predicted as observed for FTZTT (Table 4). The differences in signal intensities indicate that the HOXTT proteins were not expressed at the same level. In addition, LABTT, DFDTT, ANTPTT, UBXTT, ABD-ATT and ABD-B(m)TT undergo degradation resulting in multiple signals at different molecular weight positions for each protein (Figure 10).



**Figure 10. Expression of HOXTT proteins from the heat-shock promoter.** Western Blot was used to analyze protein extracted from heat-shocked embryos of nine HOXTT expressing fly lines. Anti-FLAG antibody was used to detect the HOXTT proteins. The HOXTT expressed is indicated above the lane. The control is protein extracted from *y w* embryos. The position of protein molecular weight markers is indicated on the right. Along the bottom is the  $\beta$ -tubulin loading control detected using anti- $\beta$ -tubulin antibody.

**Table 4. Predicted and observed relative molecular masses ( $M_r$ ) of HOXTT proteins**

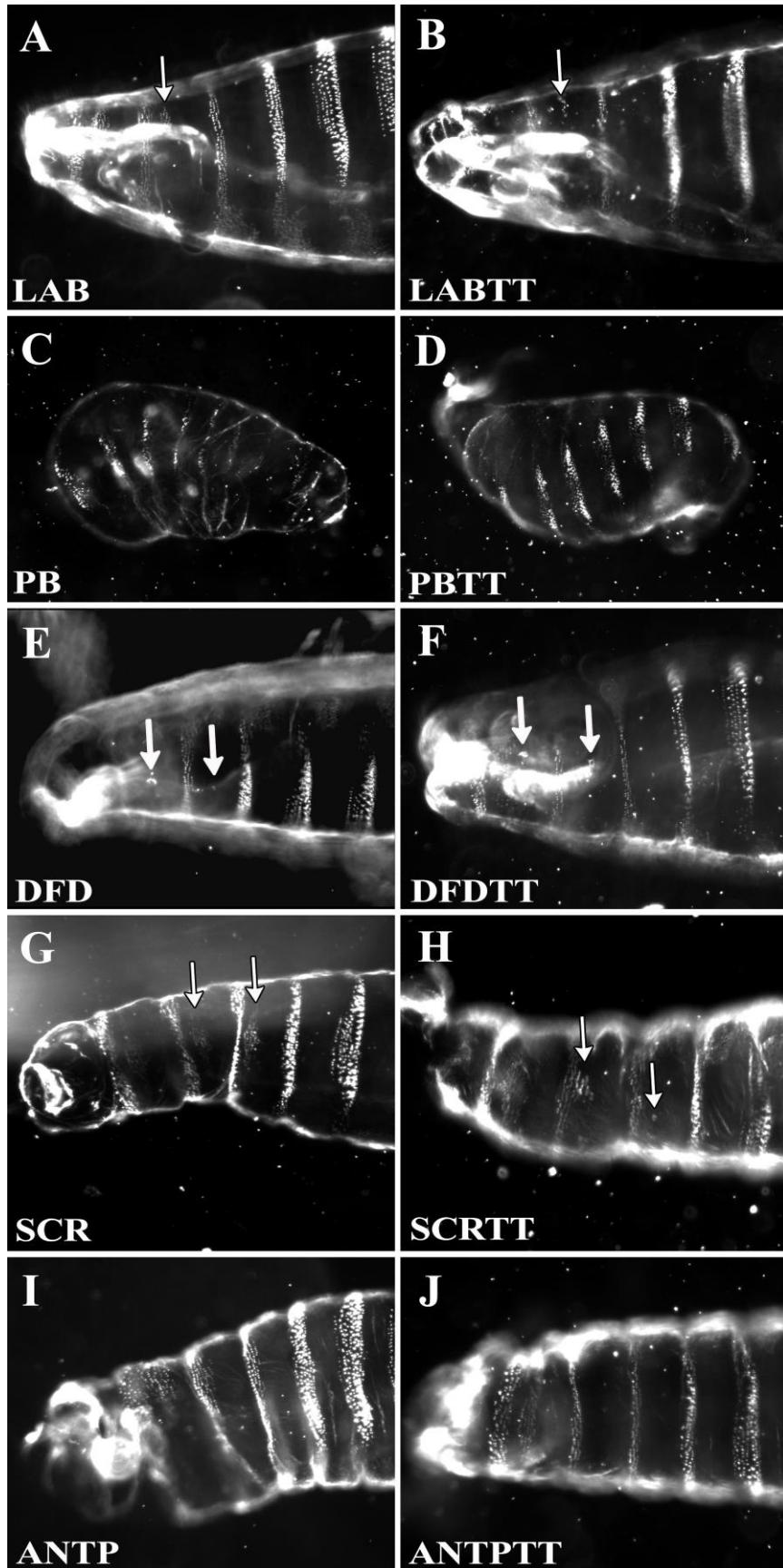
Tagged Proteins	Predicted $M_r$	Observed $M_r$
LABTT	73.1	125
PBTT	88.7	130
DFD TT	69	95
SCR TT	49.8	62
ANTPTT	48.3	62
UBXTT	45.6	55
ABD-ATT	41.8	52
ABD-B(m)TT	60.7	85
ABD-B(r)TT	35.9	37

### 3.3 HOXTT proteins are biologically active

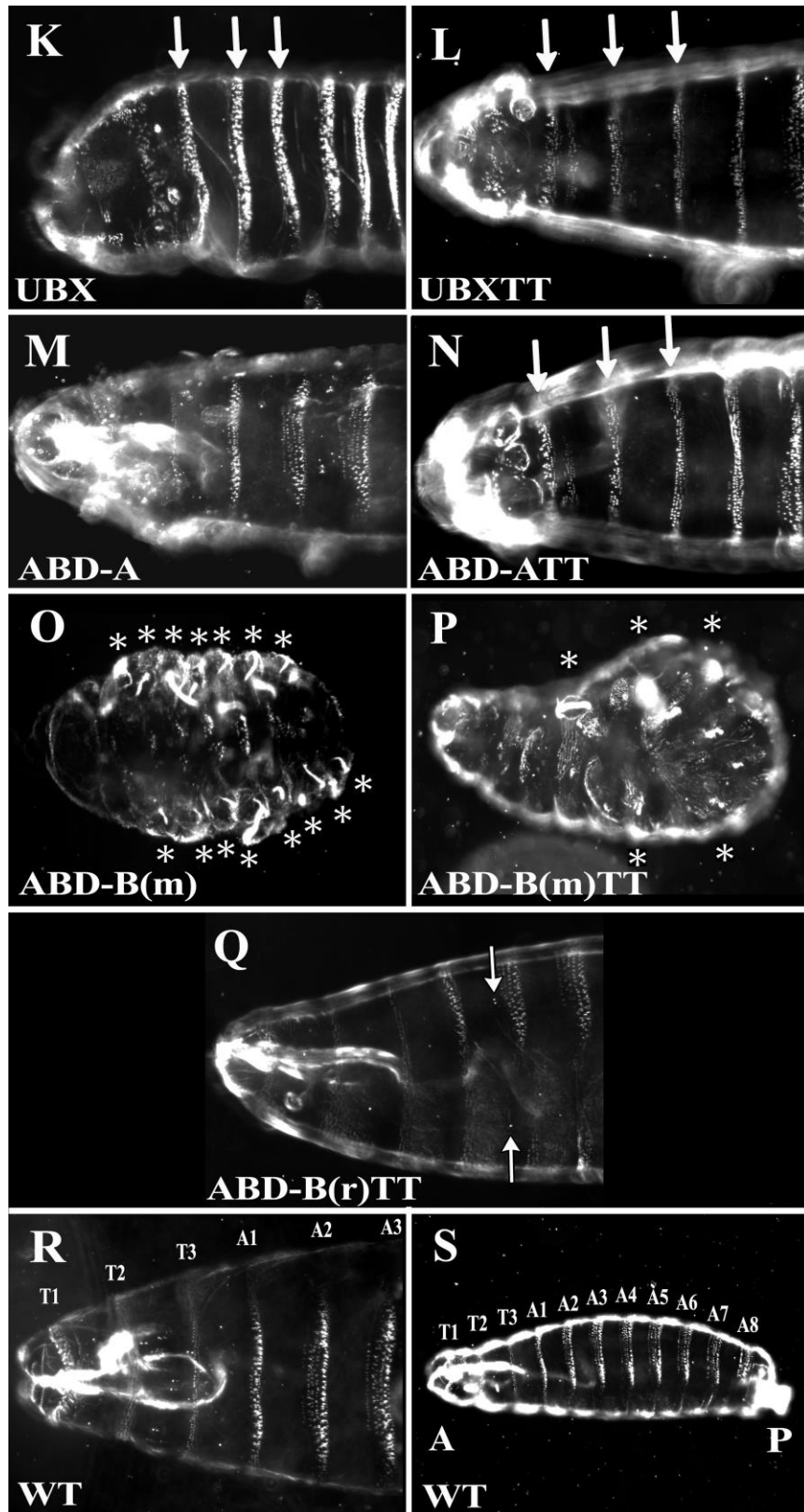
To determine whether the HOXTT proteins are biologically active, I ectopically expressed the HOXTT proteins in all cells during embryogenesis and assessed whether the first instar larval cuticle exhibited the same phenotype as ectopic expression of the untagged HOX protein. The untagged HOX proteins were expressed using the GAL4-UAS system (Brand & Perrimon, 1993). HOX proteins expressed from UAS promoters were expressed in all cells of the embryo with the *armadillo*-GAL4 driver (Figure 11A,

C, E, G, I, K, M & O). HOXTT proteins were expressed in all cells by the administration of a heat-shock during embryogenesis (Figure 11B, D, F, H, J, L, N, P & Q). Ectopic expression of HOX proteins from the *UAS* promoter gave all phenotypes reported in the literature with the exception of LAB, ABD-A and ABD-B(r) (Gehring *et al.*, 2009; Li *et al.*, 1999; Gibson *et al.*, 1990; Percival-Smith *et al.*, 2013; Grenier & Carroll, 2000; Lovegrove *et al.*, 2006). Ectopic expression of LAB from the *UAS* promoter gave a phenotype not previously reported in the literature of ectopic thoracic 1 (T1) beards in T2 (Figure 11A) and this new phenotype was also observed when LABTT was ectopically expressed from the heat-shock promoter (Figure 11B). Ectopic expression of ABD-ATT from the heat-shock promoter did result in transformation of the thoracic denticle belts toward abdominal identity as reported in the literature (Sánchez-Herrero *et al.*, 1994; Figure 11N); whereas, the control ABD-A protein expressed from the *UAS* promoter did not give the reported phenotype (Figure 11M). For ABD-B(r), the expected ectopic expression phenotype of ectopic ventral pits with the tagged version was observed (Kuziora, 1993; Figure 11Q). The untagged version of ABD-B(r) expressed from a *UAS* promoter is not available for comparison. Ectopic expression of all other HOXTT proteins resulted in the phenotype reported in the literature (Percival-Smith *et al.*, 1997; Kuziora & McGinnis, 1988; Zhao *et al.*, 1993; Schneuwly *et al.*, 1997; Gibson & Gehring, 1988; González-Reyes & Morata, 1991; Sánchez-Herrero *et al.*, 1994; Lamka *et al.*, 1992; Kuziora, 1993). Both PB and PBTT induced defects of germ band retraction and head involution (Figure 11C & D, Gehring *et al.*, 2009; Percival-Smith *et al.*, 1997). Both DFD and DFDTT induced ectopic cirri (maxillary structures) in T1 and T2 segments (Figure 11E & F, Li *et al.*, 1999; Kuziora & McGinnis, 1988). Both SCR and SCR TT expression induced ectopic T1 beards in T2 and T3 (Figure 11G & H, Gibson *et al.*, 1990; Percival-Smith *et al.*, 2013; Zhao *et al.*, 1993). Ectopic expression of ANTP and ANTP TT induced head involution defects (Figure 11I & J, Gehring *et al.*, 2009; Schneuwly *et al.*, 1997; Gibson & Gehring, 1988). Both UBX and UBXTT expression induced transformation of thoracic segments to abdominal fate (Figure 11K & L, Grenier & Carroll, 2000; González-Reyes & Morata, 1991); however, the ectopic expression phenotype is stronger for the untagged UBX (Figure 11K). For UBXTT, the setae of the denticle belts of thoracic segments were transformed to abdominal fate, i.e., the size of

the thoracic setae increased to match the size of the abdominal setae (Figure 11L). Both ABD-B(m) and ABD-B(m)TT expression induced formation of ectopic filzkörper (structures that normally develop on A8) in thoracic and abdominal segments (Figure 11O & P, Lovegrove *et al.*, 2006; Kuziora, 1993). These results indicate that the triple tag does not interfere with the biological activity of HOX proteins *in vivo*.







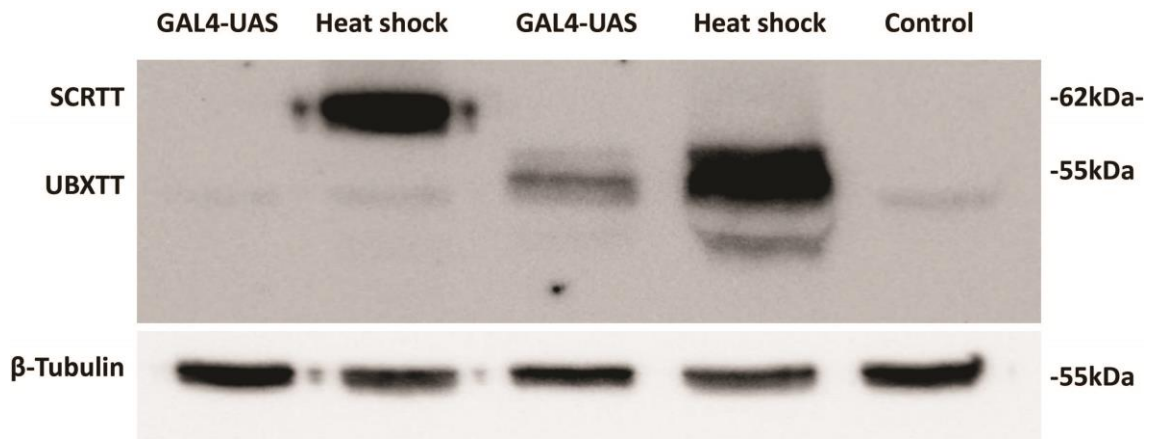
**Figure 11. First instar larval cuticular phenotypes of embryonic ectopic expression**

**of HOX and HOXTT proteins.** The untagged HOX proteins expressed with the GAL4-*UAS* system using a ubiquitous *armadillo*-GAL4 are shown in panels A, C, E, G, I, K, M and O; whereas, the HOXTT proteins expressed from a heat-shock promoter are shown in panels B, D, F, H, J, L, N, P and Q. T1, T2 and T3 refer to first, second and third thoracic segments. A1, A2, A3,...A8 refer to first, second, third,...eighth abdominal segments. A & B. Ectopic expression of LAB and LABTT, respectively (T2 beard marked with an arrow). C & D. Ectopic expression of PB and PBTT, respectively. E & F. Ectopic expression of DFD and DFDTT, respectively (cirri in T1 and T2 marked with arrows). G & H. Ectopic expression of SCR and SCR TT, respectively (T2 and T3 beards marked with arrows). I & J. Ectopic expression of ANTP and ANTP TT, respectively. K & L. Ectopic expression of UBX and UBX TT, respectively (abdominal transformations marked with arrows). M & N. Ectopic expression of ABD-A and ABD-ATT, respectively (abdominal transformations marked with arrows). O & P. Ectopic expression of ABD-B(m) and ABD-B(m)TT, respectively (ectopic filzkörper marked with asterisks, \*). Q. Ectopic expression of ABD-B(r)TT (ectopic ventral pits marked with arrows). R & S. Control wild-type first instar larval cuticles; anterior (A) and full anterior (A) to posterior (P), respectively.

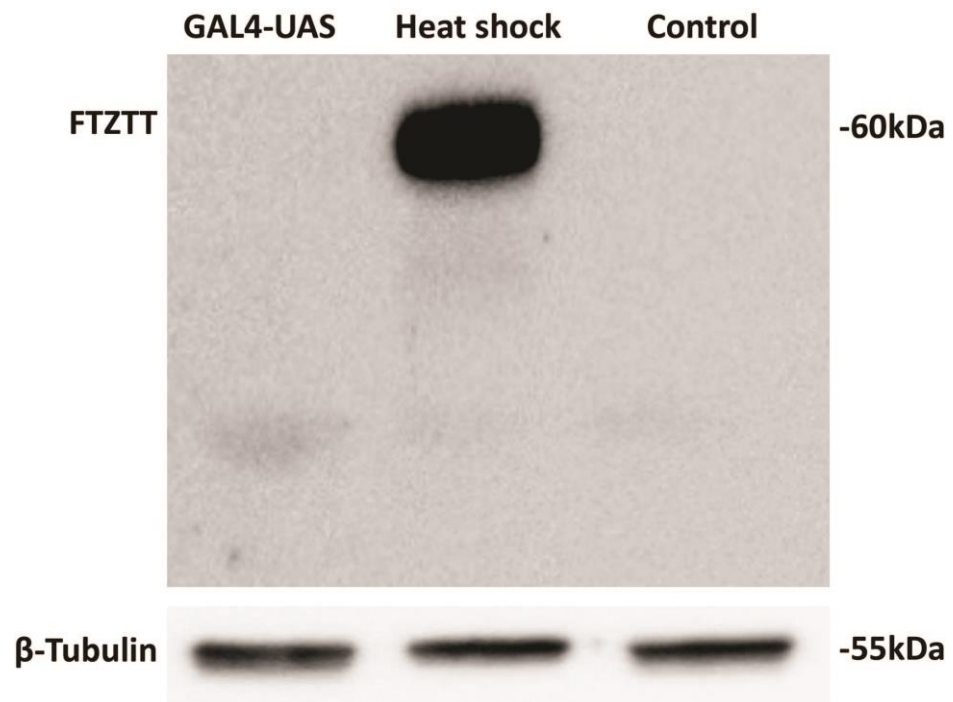
### 3.4 Levels of HOXTT and FTZTT protein expression using the heat-shock ectopic expression system and GAL4-*UAS* system

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 first consideration in the purification of any protein is finding a source with the highest initial concentration. I have compared the expression of HOXTT and FTZTT using the heat-shock promoter and GAL4-*UAS* system. FTZTT, SCR<sub>1</sub>TT and UBXTT were expressed during embryogenesis either from a heat-shock promoter or from a *UAS* promoter driven with the *armadillo*-GAL4 driver (Figure 12A & B). In all cases, the heat-shock promoter resulted in higher levels of accumulation of FTZTT, SCR<sub>1</sub>TT and UBXTT (Figure 12A & B). The fold increase of expression of heat-shock relative to *UAS* was too great to be accurately quantified for FTZTT and SCR<sub>1</sub>TT, but was 6.5-fold for UBXTT.

A.



B.

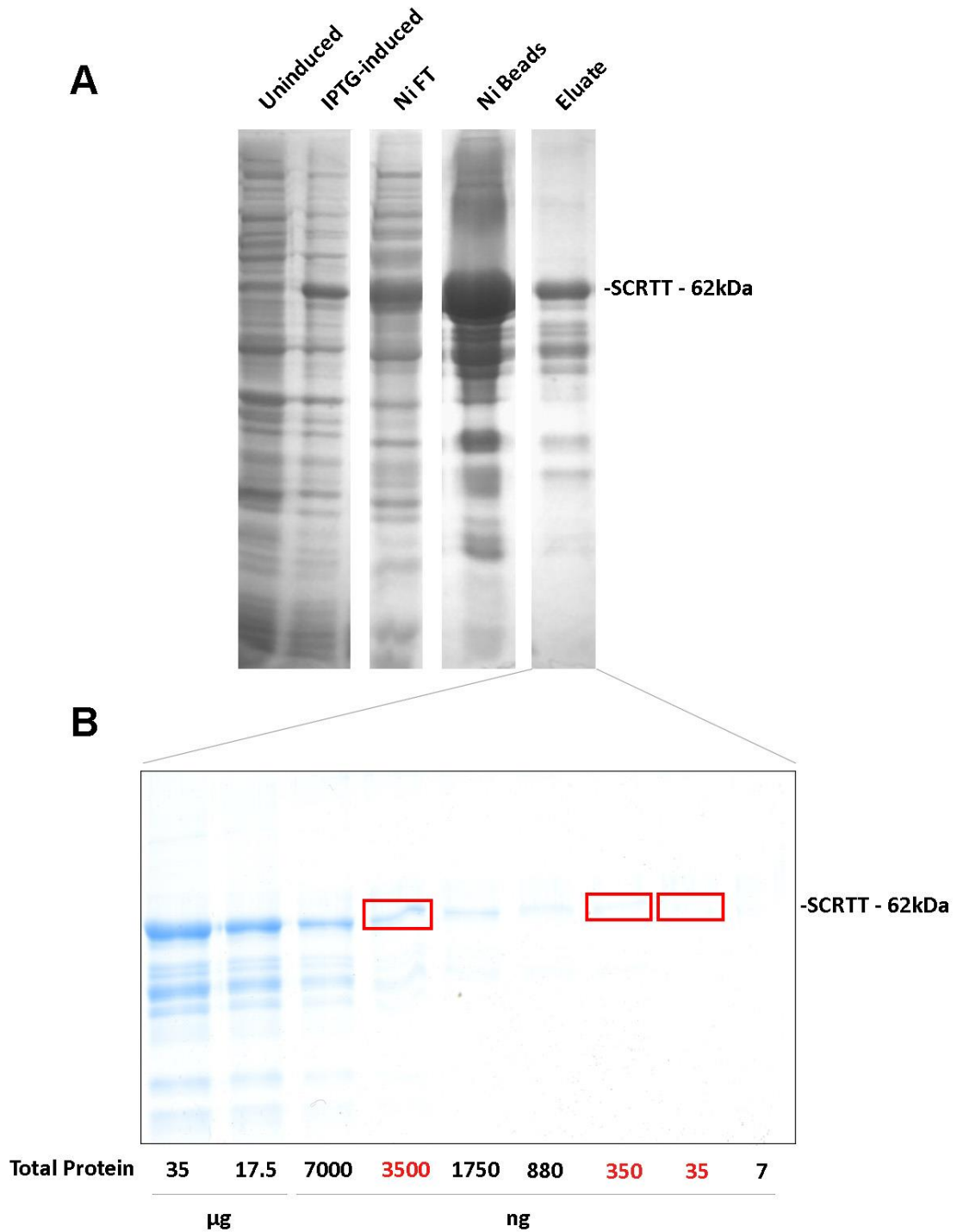


**Figure 12. Comparison of the expression of triple tagged proteins from heat-shock and *UAS* promoters.** Proteins extracted from embryos expressing FTZTT, SCR $TT$  and UBXTT from either the heat-shock or *UAS* promoter were analyzed on a Western Blot. Anti-FLAG antibody was used to detect the triple tagged proteins. A. SCR $TT$  and UBXTT. The protein and method of expression are indicated above the lanes. The position of SCR $TT$  and UBXTT are indicated on the left, and the positions of molecular weight markers are indicated on the right. The level of  $\beta$ -tubulin expression is shown at the bottom. *y w* is the untransformed control. B. FTZTT. The protein and method of expression are indicated above the lanes. The position of FTZTT is indicated on the left, and the positions of molecular weight markers are indicated on the right. The level of  $\beta$ -tubulin expression is shown at the bottom. *y w* is the untransformed control. Anti- $\beta$ -tubulin antibody was used to detect  $\beta$ -tubulin protein.

### 3.5 Determining the amount of protein required for MS/MS analysis

To determine the amount of protein required for a MS/MS analysis to detect post-translational modifications, I purified SCR $TT$  from bacteria and purchased pure bovine  $\alpha$ -casein, a known heavily phosphorylated protein (Larsen *et al.*, 2005). Using a plasmid provided by Dr. Lovesha Sivanantharajah with *ScrTT* inserted behind a T7 promoter, I purified SCR $TT$  from *E. coli* using Ni-NTA chromatography (Hochuli *et al.*, 1987; Hochuli *et al.*, 1988) (Figure 13A). Different amounts of protein were run on a SDS-polyacrylamide gel and lanes containing 3500ng, 350ng and 35ng of protein were MS/MS analyzed (Figure 13B). A summary of the peptides and their position on the primary sequence of SCR $TT$  for the three samples analyzed is shown in Figure S1. The percent coverage of 3500ng of SCR $TT$  was the greatest at 65% with 350ng having 52% coverage and 35ng having 20% coverage. Another quantitative measure is Average Depth of regions covered ( $AD_{orc}$ ) (defined as AD (the total length of chemically distinct peptides identified divided by the length of the protein) divided by proportion of coverage). The  $AD_{orc}$  for 3500ng of SCR $TT$  is the highest at 25.95 followed by 350ng at

7.10 and 35ng at 2.25. The bacterially expressed SCRTT protein is extensively modified but none of the modifications include phosphorylation.

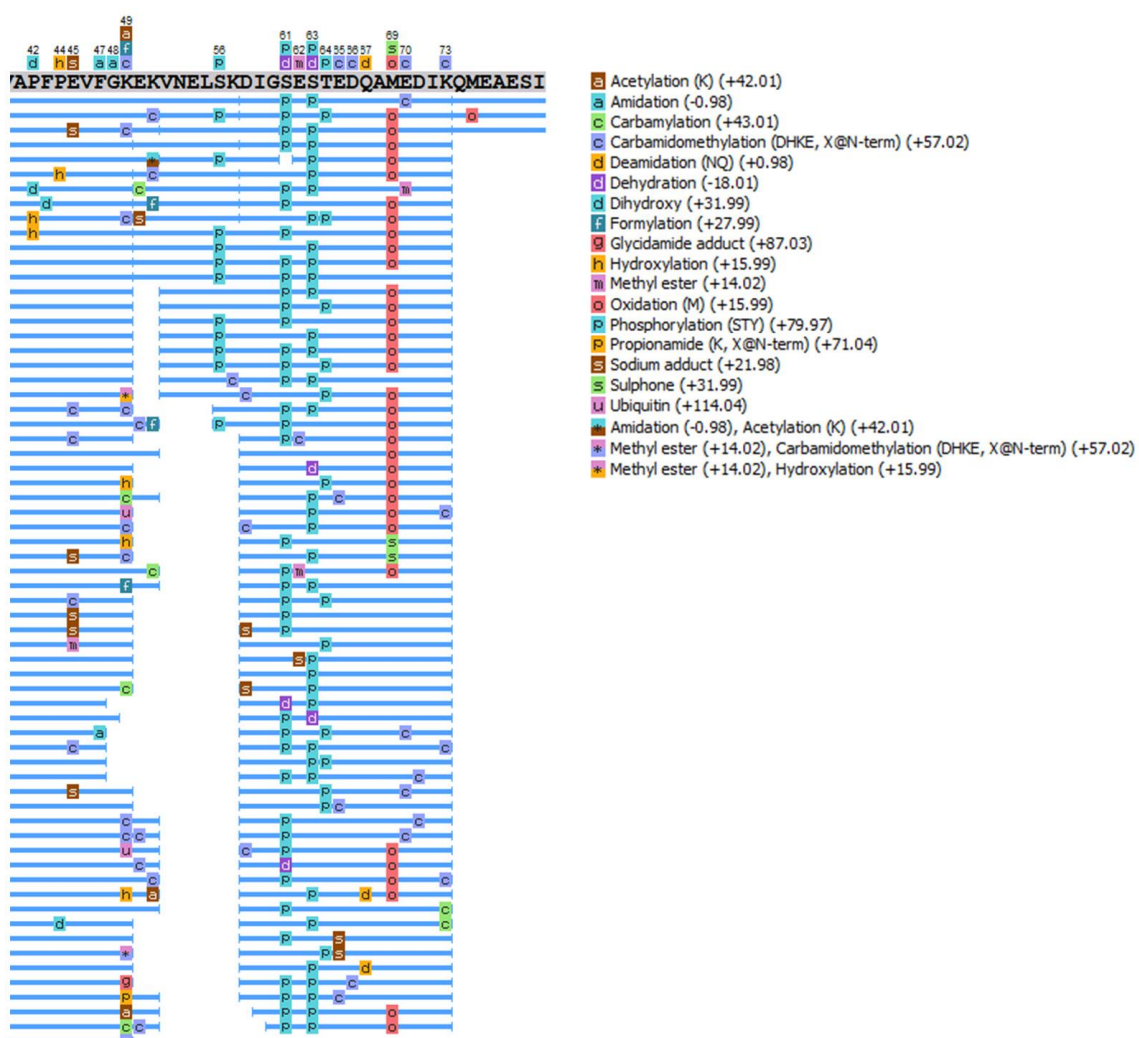


**Figure 13. Bacterially expressed SCRTT and sample preparation for MS/MS.**

SCRTT was expressed from an IPTG-inducible promoter and was affinity purified using a Ni-NTA column. Samples of fractions collected during purification were run on a SDS-polyacrylamide gel. A. A Coomassie-stained SDS-polyacrylamide gel showing total proteins in samples of fractions collected during purification. The fraction is indicated above the lane. The position of SCRTT protein along with its molecular weight is indicated on the right. All lanes shown are from the same gel and the white space between lanes indicate the splicing out of irrelevant lanes. B. A stained SDS gel shows eluate dilutions and three dilutions with total protein concentrations, 3500ng, 350ng and 35ng were used for MS/MS analysis. The total protein concentration for each dilution is indicated below the gel. The SCRTT protein bands corresponding to the above three dilutions are marked with red boxes. Ni stands for Nickel and Ni FT stands for Nickel Flow-Through.

I also analyzed the MS/MS spectra of a known heavily phosphorylated protein  $\alpha$ -casein (Larsen *et al.*, 2005). 5 $\mu$ g of  $\alpha$ -casein was run on a SDS-polyacrylamide gel and the sample was analyzed by MS/MS. A summary of the peptides and their position on the primary sequence of  $\alpha$ -casein is shown in Figure S2. The coverage was 92% and the AD<sub>orc</sub> was 63.77. Of the 839  $\alpha$ -casein peptides identified, 158 (18.8%) were phosphopeptides (Table S22). Of the 64 peptides that cover the region 61-70 of  $\alpha$ -casein, only 3 peptides were not phosphorylated (Figure 14). The analysis identified 3 phosphosites in this region of the protein represented in multiple independent peptides. I used TiO<sub>2</sub> beads (Humphrey *et al.*, 2015) followed by MS/MS analysis to test whether phosphopeptides could be preferentially enriched. The coverage was 28% and AD<sub>orc</sub> was 21.86. Of the 91  $\alpha$ -casein peptides identified post-TiO<sub>2</sub> enrichment, 84 (92.3%) were phosphopeptides (Table S22) indicating an enrichment for phosphopeptides. Of the 158 chemically distinct phosphopeptides identified without a TiO<sub>2</sub>-enrichment step, 54 were identified post-TiO<sub>2</sub> treatment. This indicates a decrease in the number of chemically distinct phosphopeptides identified post-TiO<sub>2</sub> treatment relative to the initial sample

suggesting that not all phosphorylated peptides were enriched for. However, 30 unique phosphopeptides were identified post-TiO<sub>2</sub> enrichment, which were not initially detected (Table S22). Phosphopeptide enrichment was efficient at removing unphosphorylated peptides from the sample, but does not guarantee necessarily that every phosphorylated peptide will be identified post-enrichment. I conclude that MS/MS analysis of between 350ng and 3500ng of protein should be adequate to identify PTMs.

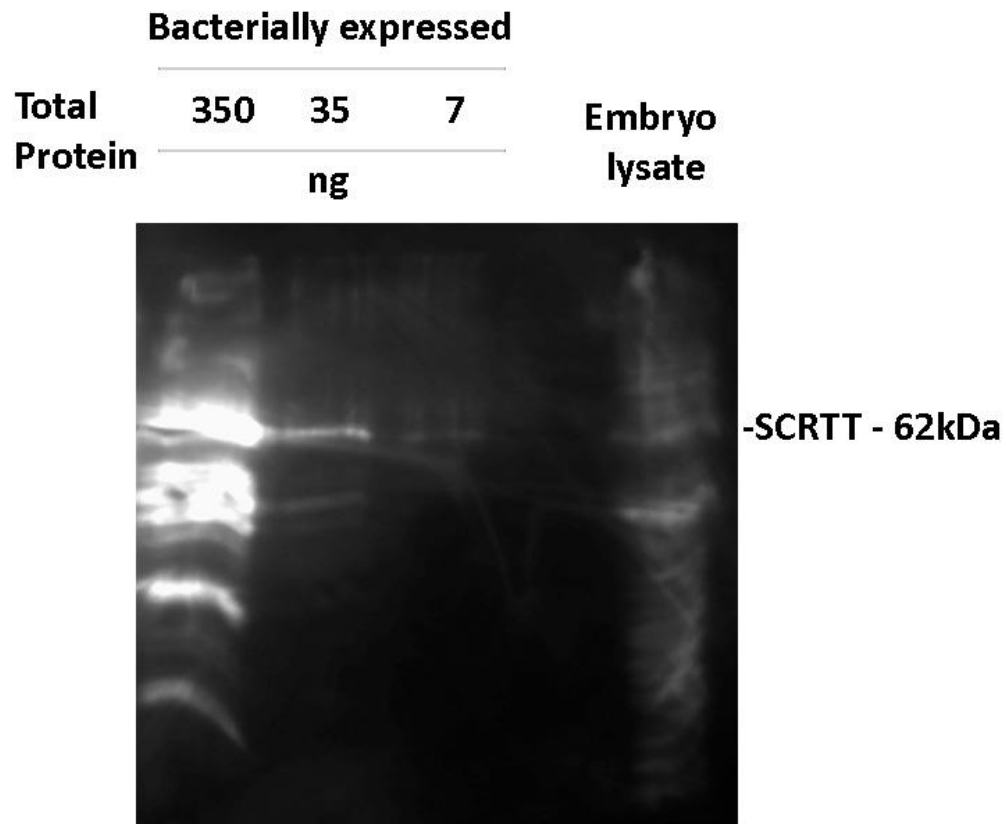




**Figure 14. Identification of phosphopeptides of a commercially purchased, pure phosphoprotein,  $\alpha$ -casein.** The figure shows a region of 41-80 of  $\alpha$ -casein (full protein in Figure S2) and each blue line underneath the primary protein sequence represents a chemically distinct peptide 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.

### 3.6 Estimating whether an extract of heat-shocked embryos expressed enough SCRTT protein for MS analysis

In order to establish whether 3 grams of heat-shocked embryos expressed enough HOXTT protein for MS analysis, a Western Blot of protein extracted from heat-shocked embryos expressing SCRTT and known amounts of bacterially expressed SCRTT was performed (Figure 15). The signals in lanes with 350ng, 35ng and 7ng of bacterial SCRTT were quantified. On the same Western Blot, 3.5 $\mu$ l of homogenized heat-shocked embryos expressing SCRTT were loaded for comparison with the bacterially expressed SCRTT. Table 5 shows the signal intensity values obtained by densitometry. The signal intensity to amount of protein loaded is not linear; however, the extract is estimated to have about 10ng of protein (Table 5). During a standard purification, 3g of embryos are homogenized in 15ml of the lysis buffer for a total volume is 18ml. If 3.5 $\mu$ l of embryo lysate has 10ng of protein, 18ml will have 51 $\mu$ g of protein, which if all purified and concentrated would provide 10 samples of 5000ng for MS analysis.



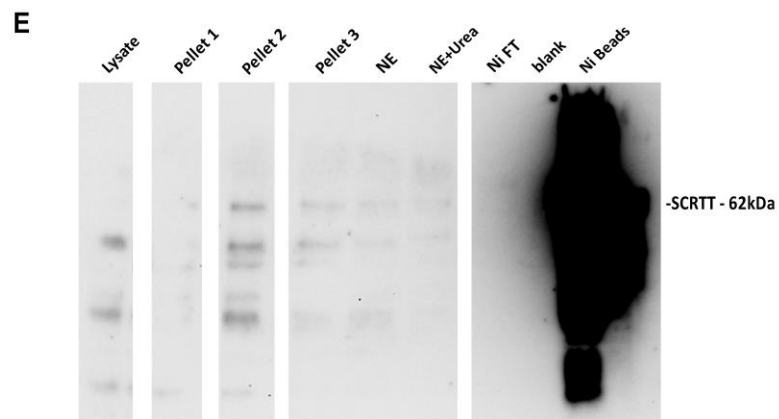
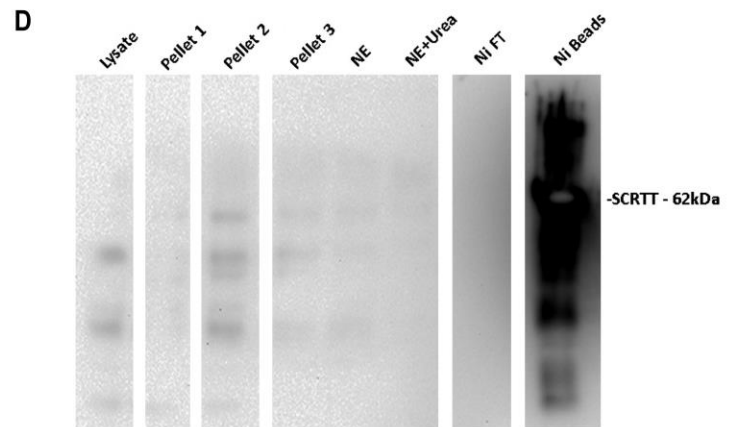
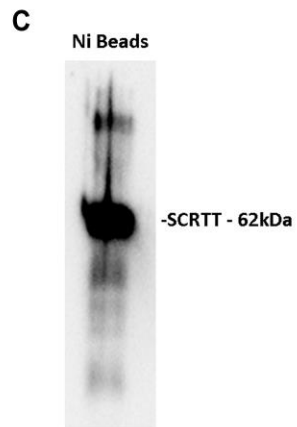
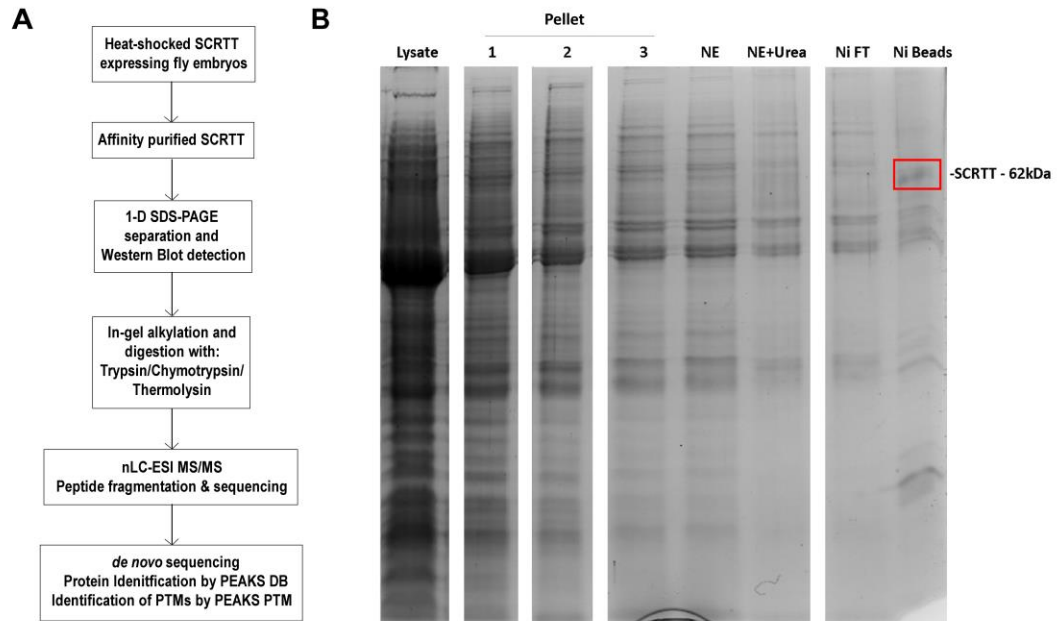
**Figure 15. Comparison of bacterially expressed SCRTT and homogenized heat-shocked SCRTT embryos.** Western Blot shows signal for bacterially expressed SCRTT for eluate dilutions with total protein amounts of 350ng, 35ng and 7ng (as shown in Figure 12B) and homogenized heat-shocked SCRTT embryos. The signals for 350ng, 35ng and 7ng were quantified. The rightmost lane is the SCRTT embryo lysate. Anti-FLAG antibody was used to detect SCRTT. The total protein concentration loaded on a gel is indicated above the lane. The position of SCRTT and its molecular weight is indicated on the right of the gel.

**Table 5. Quantification of SCRTT protein signals**

<b>Amount of protein loaded or sample</b>	<b>SCRTT signal intensity from Western Blot</b>
350ng	51,478
35ng	13,574
7ng	4,522
SCRTT from embryos	6,907

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

The overall strategy (Figure 16A) required optimization of the affinity purification procedure employed to enrich SCRTT from developing embryos for PTM detection by MS/MS analysis. 3g of heat-shocked embryos between 0 and 14h AEL were lysed. The nuclei fraction was collected, 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 and the amount of SCRTT in the various fractions were estimated by quantitation of the intensity of the signal, and the amount of total protein was monitored (Figure 16D & Table 6). During nuclear fractionation, the yield seemed to increase 50-fold which is most likely an artifact of the Western Blot analysis (Figure 16D & E; Table 6). Repeated attempts have shown that HOXTT proteins are not efficiently eluted from Ni-NTA beads; therefore, the final fold purification could not be estimated. 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 16B).

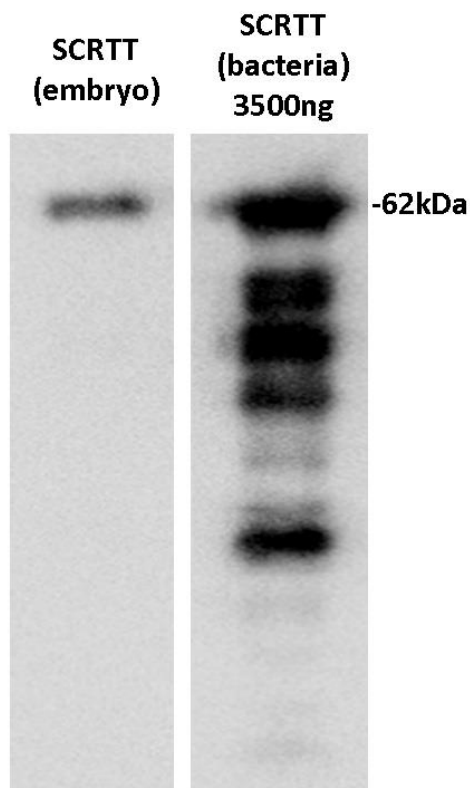


**Figure 16. Overall approach of affinity purification to PTM mapping in SCR TT.** A. Analytical workflow for affinity purification, digestion and sequence mapping of PTMs in SCR TT expressed from developing embryos. B. Coomassie-stained 1-D SDS-polyacrylamide gel of affinity purified SCR TT. Purification fraction is indicated on top of the lane. A band of SCR TT at 62 kDa (marked with a red box) was seen in the Ni beads fraction. C. Autoexposed Western Blot shows the position of SCR TT at 62 kDa in the Ni beads fraction. D. Optimally exposed Western Blot showing SCR TT at 62kDa in the Ni beads fraction. This Western Blot was used for the estimation of SCR TT signal intensities needed to construct the purification table (Table 6). E. Overexposed Western Blot showing SCR TT at 62 kDa in the Ni beads fraction. For panels C, D and E, anti-FLAG antibody was used to detect SCR TT. Purification fraction is indicated on top of the lane. NE stands for Nuclear Extract, Ni stands for Nickel and Ni FT stands for Nickel Flow-Through. For each of the panels B, D and E, all lanes shown are from the same gel and the white space between lanes indicate the splicing out of irrelevant lanes.

**Table 6. Purification table for SCRTT**

Fraction	Total volume (ml)	Total protein (mg)	Total Activity (units) x 10 <sup>7</sup>	Specific activity (Units/mg) x 10 <sup>5</sup>	Yield (%)	Fold Purification	Fold concentration
Lysate	18	130	1.7	1.3	100	1	1
Pellet 1	10	35	4.3	12	250	9.1	4.4
Pellet 2	10	23	78	340	4300	253	81
Pellet 3	10	14	22	160	1300	117	23
NE	10	16	18	110	1000	85	19
NE+Urea	14	17	23	140	1300	103	17
Ni Beads	0.25	-	87	-	5000	-	3600

To determine whether this purification provided the amount of SCRTT required for MS/MS, I ran a sample of the Ni beads containing purified SCRTT alongside a sample of 3500ng of bacterial SCRTT (Figure 17). 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, I estimate that the amount in the band on the SDS gel that I isolated for enzyme digestion was in the range of 10 $\mu$ g.



**Figure 17. 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. Anti-FLAG antibody was used to detect SCRTT. The source of SCRTT is mentioned on top of the lane. Both lanes shown are from the same gel and the white space between the 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 most of the primary sequence of SCRTT. The enzymes chosen for protein digestion were trypsin, chymotrypsin and thermolysin. A total of 6 samples digested with trypsin (3), chymotrypsin (1) and thermolysin (2) were analyzed with MS/MS. Figure S3 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, combined coverage and  $AD_{orc}$

for three samples digested with trypsin and two samples digested with thermolysin, and total coverage and total AD<sub>orc</sub> for all three enzymes were determined (Table 7). With all three enzymes combined, 96% of the primary sequence of SCRTT was covered.

**Table 7. MS/MS analysis – Coverage and AD<sub>orc</sub> for SCRTT**

Enzyme	Sample	% Coverage	AD <sub>orc</sub>
Trypsin	1	47	4.81
	2	66	10.79
	3	62	10.39
	1+2+3	72	14.44*
Chymotrypsin	1	80	8.7
Thermolysin	1	67	7.22
	2	60	5.92
	1+2	71	9.48*
Trypsin+Chymotrypsin+Thermolysin	-	96	25.09*

\* Duplicate peptides were removed for calculation of AD<sub>orc</sub>.



### 3.8 Embryonic SCRTT is post-translationally modified

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 and the modifications that may be the result of a biochemical process were followed up in more detail (Table 8 & Figure 24).

**Table 8. Post-translational modifications of SCRTT identified by MS/MS**

Location	Peptide	n*	Protease	Site modified	Modification
17-24	S.LAS(+14.02)C(+57.02) YYPQQ.M	1	Thermolysin	S19	Methylation
17-38	S.LASC(+57.02)YP(+15.99) QQMNPQQNHPGA GNSS.A	1	Thermolysin	P22	Hydroxylation
86-119	N.LYPNTPQAHYANQA AYGGQGNPD(+15.99) MVDYTQLQPQR.L	1	Trypsin	D108	Hydroxylation
95-119	H.YANQAAYGGQGNP(+15.99) DMVDYTQLQPQR.L	3	Trypsin	P107	Hydroxylation
95-119	H.YANQAAYGGQGNPD(+15.99) MVDYTQLQPQR.L	3	Trypsin	D108	Hydroxylation

98-119	N.QAAYGGQGNP(+15.99)DMVDYDTQLQPQR.L	1	Trypsin	P107	Hydroxylation
99-119	Q.AAYGGQGNPD(+15.99)MVDYDTQLQPQR.L	1	Trypsin	D108	Hydroxylation
101-114	A.YGGQGNP(+15.99)DMVDYDTQ.L	1	Thermolysin	P107	Hydroxylation
102-112	Y.GGQGNPD(+15.99)MVDY.T	1	Chymotrypsin	D108	Hydroxylation
102-114	Y.GGQGNPD(+15.99)MVDYDTQ.L	1	Thermolysin	D108	Hydroxylation
102-115	Y.GGQGNPD(+15.99)MVDYDTQL.Q	1	Chymotrypsin	D108	Hydroxylation
102-119	Y.GGQGNPDMVD(+15.99)YTQLQPQR.L	1	Trypsin	D111	Hydroxylation
102-120	Y.GGQGNP(+15.99)DMVDYDTQLQPQRL.L	1	Chymotrypsin	P107	Hydroxylation
156-169	Q.QQQQQQANISC(+57.02)K(+14.02)Y.A	1	Chymotrypsin	K168	Methylation
163-169	Q.ANIS(+14.02)C(+57.02)KY.A	1	Chymotrypsin	S166	Methylation
169-213	K.YANDPVTPGGSGGGVVS(+79.97)GSNNNN	1	Trypsin	S185	Phosphorylation

	NSANSNNNSQSLASP QDLSTR.D				
169-213	K.YANDPVTGGGG GVSGSNNSANSN(+.98)NNNS(+79.97)QSLASPQDLSTR.D	1	Trypsin	S201	Phosphorylation
214-230	R.DISPK(+27.99)LSPSS VVESVAR.S	3	Trypsin	K218	Formylation
214-230	R.DISPK(+42.01)LSPSS VVESVAR.S	2	Trypsin	K218	Acetylation
259-269	G.VSGGPGN(+.98)VNV P(+15.99).M	1	Thermolysin	P269	Hydroxylation
266-287	N.VNVP(+15.99)MHSP GGGDSSESDSGNE.A	1	Thermolysin	P269	Hydroxylation
266-294	N.VNVP(+15.99)MHSP GGGDSSESDSGNEAG SSQNS.G	1	Thermolysin	P269	Hydroxylation
300-309	K.NPPQIYPW(+43.99) MK.R	1	Trypsin	W307	Carboxylation
300-309	K.NPPQIYPW(+43.99) M(+15.99)K.R	1	Trypsin	W307	Carboxylation
300-309	K.NPPQIYPWMK(+43.99).R	1	Trypsin	K309	Carboxylation

300-309	K.NPPQIYPWM(+15.99)K(+43.99).R	1	Trypsin	K309	Carboxylation
300-310	K.NPPQIYP(+15.99)W MKR.V	1	Trypsin	P306	Hydroxylation
300-310	K.NPPQIYPW(+43.99) MKR.V	2	Trypsin	W307	Carboxylation
300-310	K.NPPQIYPWM(+15.99)K(+27.99)R.V	2	Trypsin	K309	Formylation
300-310	K.NPPQIYPWMK(+43.99)R.V	2	Trypsin	K309	Carboxylation
311-326	R.VHLGTSTVNANGE(+43.99)TKR.Q	2	Trypsin	E323	Carboxylation
311-326	R.VHLGTSTVNAN(+.98)GET(+79.97)KR.Q	1	Trypsin	T324	Phosphorylation
311-326	R.VHLGTSTVNANGETK(+27.99)R.Q	2	Trypsin	K325	Formylation
311-326	R.VHLGTSTVNAN(+.98)GETK(+27.99)R.Q	2	Trypsin	K325	Formylation
314-326	L.GTSTVNANGE(+43.99)TKR.Q	1	Chymotrypsin	E323	Carboxylation
355-366	R.IEIAHALC(+57.02)LT(+14.02)ER.Q	1	Trypsin	T364	Methylation

367-375	R.QIK(+27.99)IWFQNR .R	2	Trypsin	K369	Formylation
431-441	D.IDYK(+42.01)DDDDK EN.L	1	Thermolysin	K434	Acetylation
431-441	D.IDYK(+27.99)DDDDK EN.L	1	Thermolysin	K434	Formylation
431-441	D.IDYKDDDDK(+42.01) EN.L	1	Thermolysin	K439	Acetylation
431-441	D.IDYKDDDDK(+27.99) EN.L	1	Thermolysin	K439	Formylation

n\* – Number of independent MS/MS analysis in which the peptide was detected

Red – MS<sup>2</sup> spectra chosen for further confirmation of the modification.

### 3.8.1 Identification of novel phosphosites in SCRTT

MS/MS analysis of SCRTT identified three phosphosites: Serine 185, Serine 201 and Threonine 324 (Figure 18). Although phosphopeptide enrichment should increase the detection of phosphopeptides by MS, I found that no phosphopeptides were identified in a TiO<sub>2</sub>-enriched sample (Humphrey *et al.*, 2015). Also, no phosphopeptides were detected in the TiO<sub>2</sub> flow-through by MS/MS analysis.

The evidence for phosphorylation of S185 is an overall mass shift of the peptide, the difference in retention time (RT) on liquid chromatography, and identified b ions (during low energy fragmentation, the C-N bond of the peptide linkage is susceptible to cleavage, and b ions are fragments that appear to extend from the N terminus and y ions are fragments that appear to extend from the C-terminus) (Figure 18A). The mass shift between the quadruply charged phosphorylated and unphosphorylated peptides was 80

Da supporting phosphorylation of the peptide. Phosphorylated peptides have a lower RT, i.e., they elute earlier from the column than their corresponding unphosphorylated moiety due to the addition of a hydrophilic phosphate group which increases the overall hydrophilicity of the peptide (Beausoleil *et al.*, 2004; Kim *et al.*, 2007). The phosphorylated peptide,

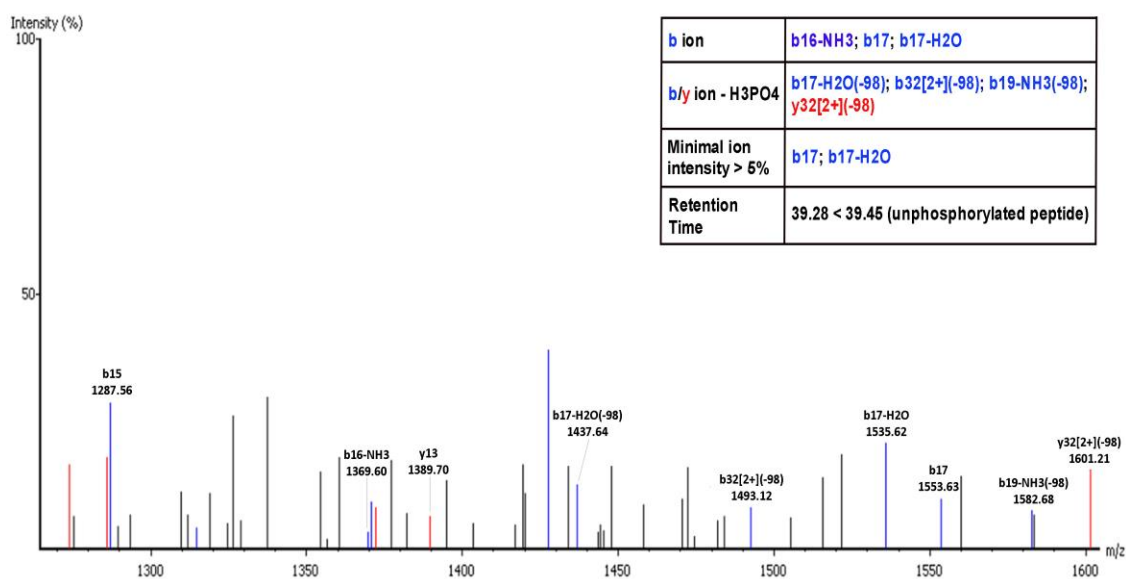
$^{169}\text{YANDPVTPGGSGGGVpS(+79.97)GSNNNNNSANSNNNNSQSLASPQDLSTR}$  ( $m/z$  1118.73<sup>4+</sup>) eluted at 39.28 min which was earlier than 39.45 min of the unphosphorylated peptide ( $m/z$  1098.74<sup>4+</sup> – MS<sup>2</sup> spectra not shown) (Figure 18A). For assigning which residue was phosphorylated, only b ions provide evidence of S185 phosphorylation (Figure 18A). The mass shift of ions, b17 and b16-NH<sub>3</sub> was 184 Da which equals the molecular mass of a serine residue (87 Da), NH<sub>3</sub> (17 Da) and the phosphate group (80 Da). The fragment ion b16 was detected at a minimal ion intensity of greater than 5%. Also, b and/or y ions with a neutral loss of phosphoric acid (with mass 98 Da) were identified which is a further indication that the fragmented peptide had a phosphate group attached to it.

Evidence of phosphorylation of S201 is an overall mass shift, lower RT and identified y ions (Figure 18B). The mass shift between the phosphorylated and unphosphorylated peptides was 80 Da, the mass of a phosphate group. The phosphorylated peptide,  $^{169}\text{YANDPVTPGGSGGGVSGSNNNNNSANSN(+.98)NNNpS(+79.97)QSLASPQDLSTR}$  ( $m/z$  1118.98<sup>4+</sup>) eluted at 38.64 min which was earlier than 40.21 min of the unphosphorylated peptide ( $m/z$  1098.98<sup>4+</sup> – MS<sup>2</sup> spectra not shown) (Figure 18B). For assigning that S201 was phosphorylated, the mass shift of y13-NH<sub>3</sub> and y12 ions equaled 150 Da, which when added to the mass of NH<sub>3</sub> (17 Da) equaled the mass of a phosphorylated serine residue (167 Da) (Figure 18B). The fragment ions, y13-NH<sub>3</sub> and y12 were detected at a minimal ion intensity of greater than 5%. Also, b and/or y ions with a neutral loss of phosphoric acid (with mass 98 Da) were identified which is a further indication that the fragmented peptide had a phosphate group attached to it.

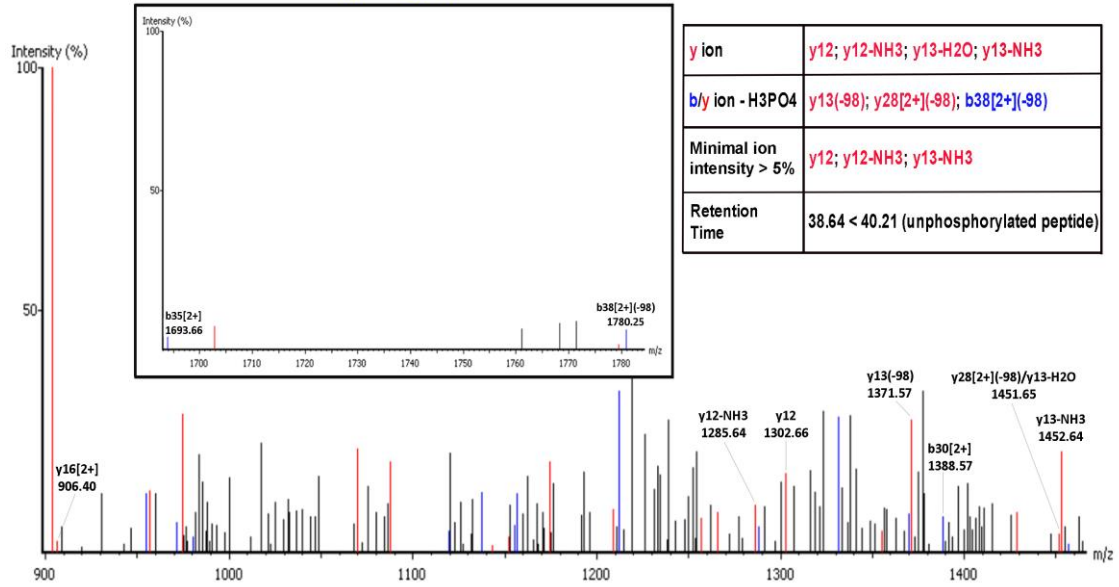
Evidence of phosphorylation of T324 is an overall mass shift and identified y ions (Figure 18C). The mass shift between the phosphorylated and unphosphorylated peptides was 80 Da, the mass of a phosphate group. For assigning that T324 was phosphorylated

in the peptide,  $^{311}\text{VHLGTSTVNAN}(+.98)\text{GEP}(+.79.97)\text{KR}$  ( $m/z$  441.96<sup>4+</sup>), the mass shift of y2 and y3 equals the mass of a phosphorylated threonine (181 Da) (Figure 18C). The fragment ion, y2 was detected at a minimal ion intensity of greater than 5%. Also, b and/or y ions with a neutral loss of phosphoric acid (with mass 98 Da) were identified which is a further indication that the fragmented peptide had a phosphate group attached to it.

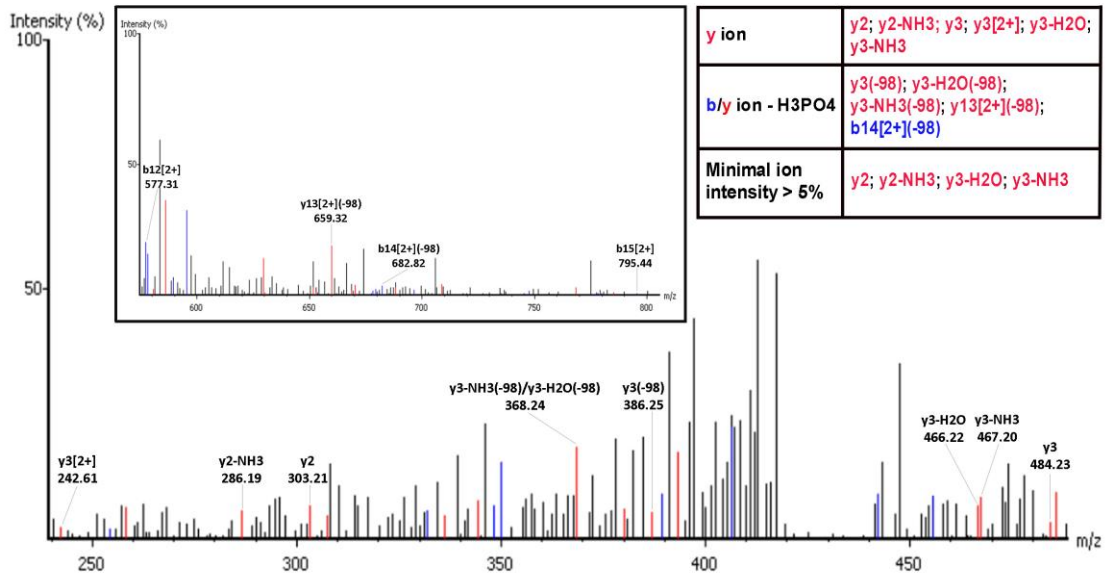
### A MS<sup>2</sup> $m/z$ 1118.73



## B MS<sup>2</sup> m/z 1118.98



## C MS<sup>2</sup> m/z 441.96





**Figure 18. Identification of novel phosphorylation sites, Serine 185, Serine 201 and Threonine 324 in SCRTT.** MS<sup>2</sup> spectra for the three phosphopeptides identified by LC-MS/MS are shown. A. Phosphorylation of Serine 185. B. Phosphorylation of Serine 201. The inset box shows fragment ions with  $m/z$  1690 to 1790. C. Phosphorylation of Threonine 324. The inset box shows fragment ions with  $m/z$  550 to 800. The peptide sequence and  $m/z$  ratio are indicated on the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidences of phosphorylation. The relevant fragment ions and their  $m/z$  ratios supporting phosphorylation are labelled in the spectra.

### 3.8.2 Identification of acetylated peptides of SCRTT

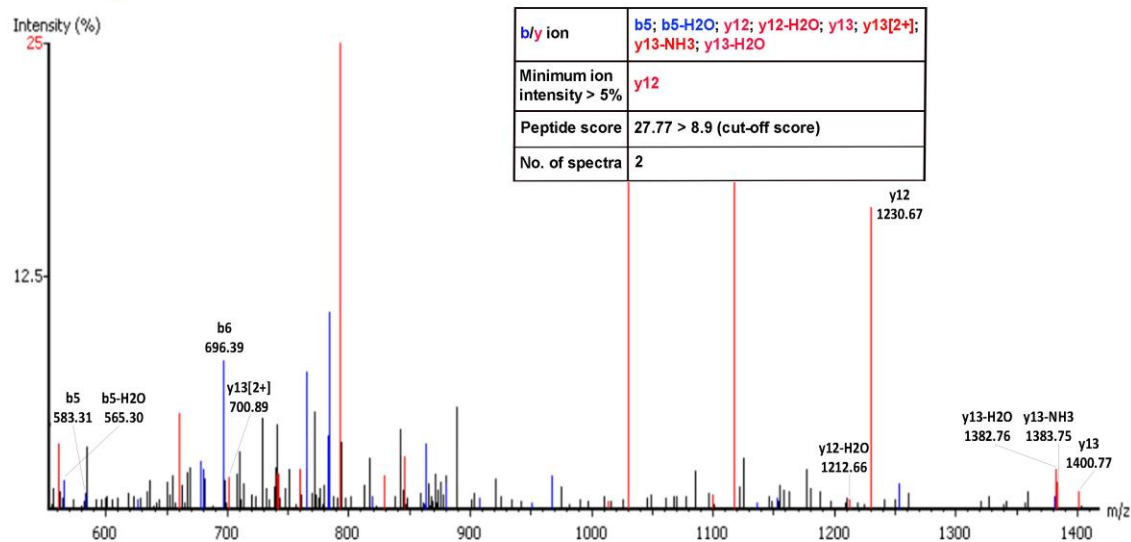
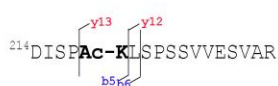
Three sites of acetylation were identified by LC-MS/MS of SCRTT: Lysine 218, Lysine 434 and Lysine 439 (Figure 19).

The evidence for acetylation of K218 is an overall mass shift and y ions. The overall mass shift between the acetylated peptide, <sup>214</sup>DISPAc-K(+42.01)LSPSSVVESVAR ( $m/z$  906.99<sup>2+</sup>) and the non-acetylated peptide equaled the mass of an acetyl group suggesting the attachment of an acetyl group to the peptide. For assigning that K218 is acetylated, the mass shift of the fragment ions, y12 and y13 was 170 Da, which is the mass of an acetylated lysine residue (Figure 19A). The y12 ion was detected at a minimal ion intensity of above 5%. The peptide score was 3-fold higher than the cut-off score and 2 independent spectra identified this acetylated peptide.

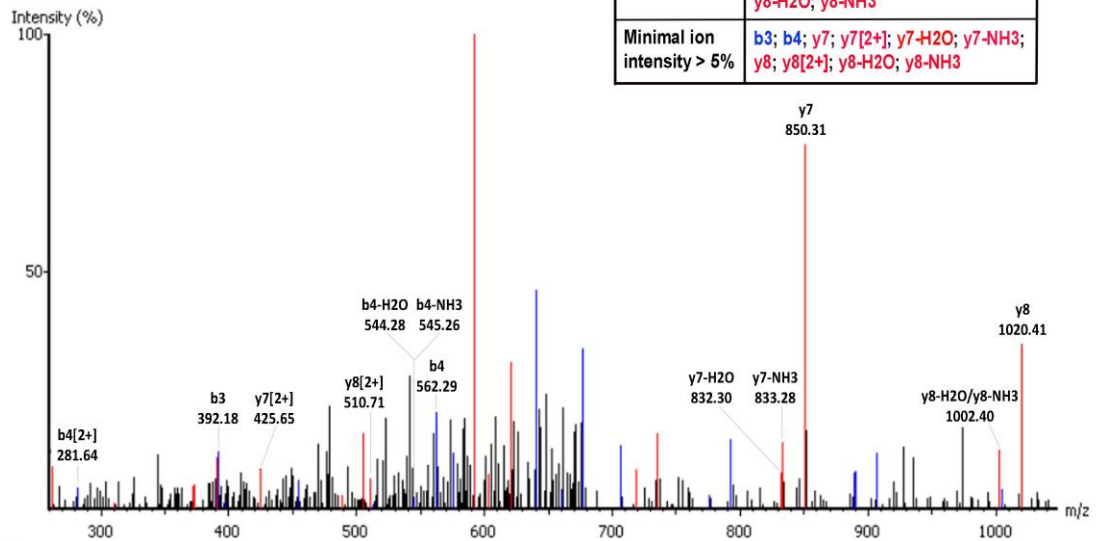
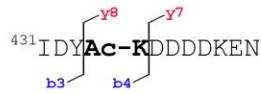
MS/MS analysis of a thermolysin digest of SCRTT yielded two acetylated peptides, <sup>431</sup>IDYAc-K(+42.01)DDDDKEN ( $m/z$  706.3<sup>2+</sup>) and <sup>431</sup>IDYKDDDAAc-K(+42.01)EN ( $m/z$  706.3<sup>2+</sup>) that have the same primary amino acid sequence and are part of the triple tag (Figure 19B & C). The overall mass shift between the acetylated and non-acetylated peptides was 42 Da, which supports acetylation of the peptides. For mapping acetylation to K434 in the <sup>431</sup>IDYAc-KDDDDKEN peptide, the mass shift of b3 and b4 fragment

ions and y7 and y8 fragment ions was 170 Da, which is the mass of an acetylated lysine supporting the acetylation of K434 (Figure 19B). Also, for mapping acetylation to K439 in the  $^{431}\text{IDYKDDDAc-KEN}$  peptide, the mass shift of b8 and b9 fragment ions and y2 and y3[2+] fragment ions equaled that of an acetylated lysine supporting K439 acetylation (Figure 19C). All 4 b and y fragment ions that support K434 acetylation were detected at a minimal ion intensity of greater than 5%. The b8 and b9 fragment ions that support K439 acetylation were detected at a minimal ion intensity of greater than 5%. For the peptide,  $^{431}\text{IDYKDDDAc-KEN}$ , 2 independent spectra were obtained.

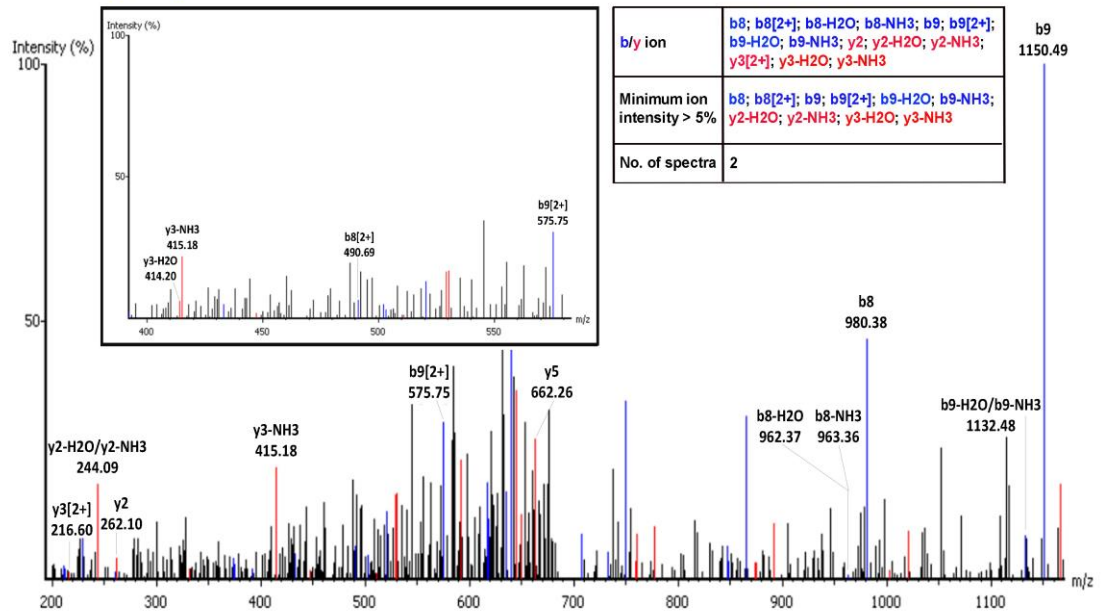
### A MS<sup>2</sup> m/z 906.99



### B MS<sup>2</sup> m/z 706.3



### C MS<sup>2</sup> m/z 706.3



**Figure 19. Acetylation of Lysine 218, 434 and 439 residues of SCRTT.** MS<sup>2</sup> spectra of the peptide identified by LC-MS/MS is shown. A. Acetylation of Lysine 218. B. Acetylation of Lysine 434. C. Acetylation of Lysine 439. The inset box shows fragment ions with  $m/z$  390 to 580. The peptide sequence and  $m/z$  ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming acetylation. The relevant fragment ions and their  $m/z$  ratios supporting acetylation are labelled in the spectra.

### 3.8.3 Identification of formylated peptides of SCRTT

Six sites of formylation were identified by LC-MS/MS of SCRTT: Lysine 218, Lysine 309, Lysine 325, Lysine 369, Lysine 434 and Lysine 439 (Figure 20).

The evidence of formylation of K218 is an overall mass shift and y ions. The overall mass shift between the formylated and non-formylated peptides equaled the mass of a formyl group, which supports formylation of the peptide. For assigning that K218 is formylated in the peptide,

<sup>214</sup>DISPFo-K(+27.99)LSPSSVVESVAR ( $m/z$  899.98<sup>2+</sup>), the mass shift of the fragment ions y12 and y13 was 156 Da which is the mass of a formylated lysine (Figure 20A). The y12 ion was detected at a minimal ion intensity of above 5%. The peptide score was 4-fold higher than the cut-off score and 10 independent spectra identified this formylated peptide.

The evidence of formylation of K309 is an overall mass shift and b ions. The overall mass shift between the formylated and non-formylated peptides equaled the mass of a formyl group. For assigning that K309 is formylated in the peptide,

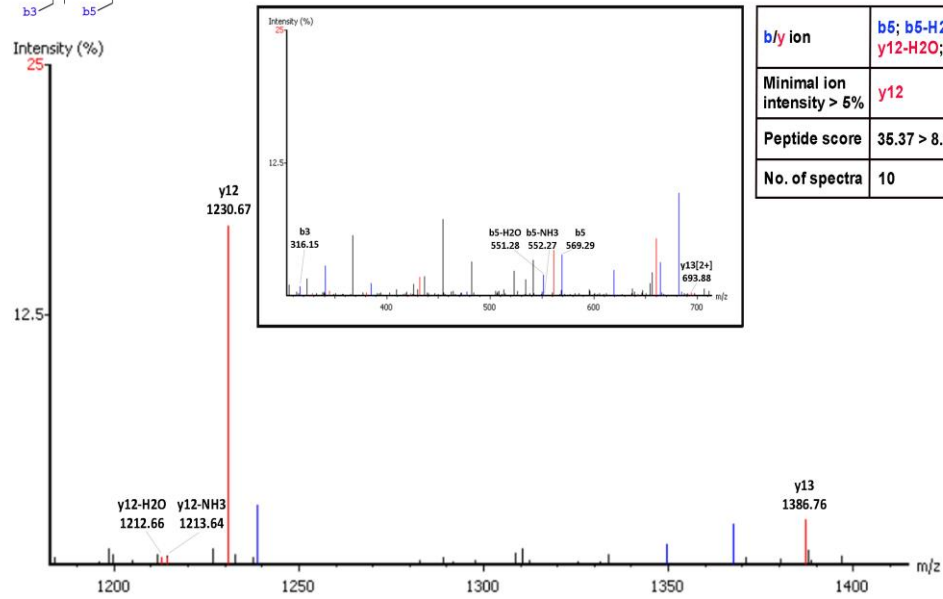
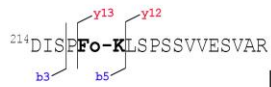
<sup>300</sup>NPPQIYPWM(+15.99)Fo-K(+27.99)R ( $m/z$  737.37<sup>2+</sup>), the mass shift of b9 and b10 ions was 156 Da which is the mass of a formylated lysine (Figure 20B). The peptide score was 2.4-fold higher than the cut-off score and 2 independent spectra were obtained for the formylated peptide.

The evidence of formylation of K325 is an overall mass shift and b and y ions. Two chemically distinct peptides formylated at K325 were identified by MS/MS of a tryptic digest of SCRTT (Table 8). The peptide with the best spectra of the two,  $^{311}\text{VHLGTSTVNANGETFo-K}(+27.99)\text{R}$  ( $m/z$  571.3<sup>3+</sup>) shifted by a mass of 28 Da when compared to its non-formylated moiety. For assigning that K325 is formylated in the peptide, the mass shift of y1 and y2 fragment ions and b14 and b15[2+] fragment ions equaled that of a formylated lysine residue (Figure 20C). Both y1 and y2 ions were detected at a minimal ion intensity greater than 5%. The peptide score was 2.6-fold higher than the cut-off score and 3 independent spectra were obtained for the formylated peptide.

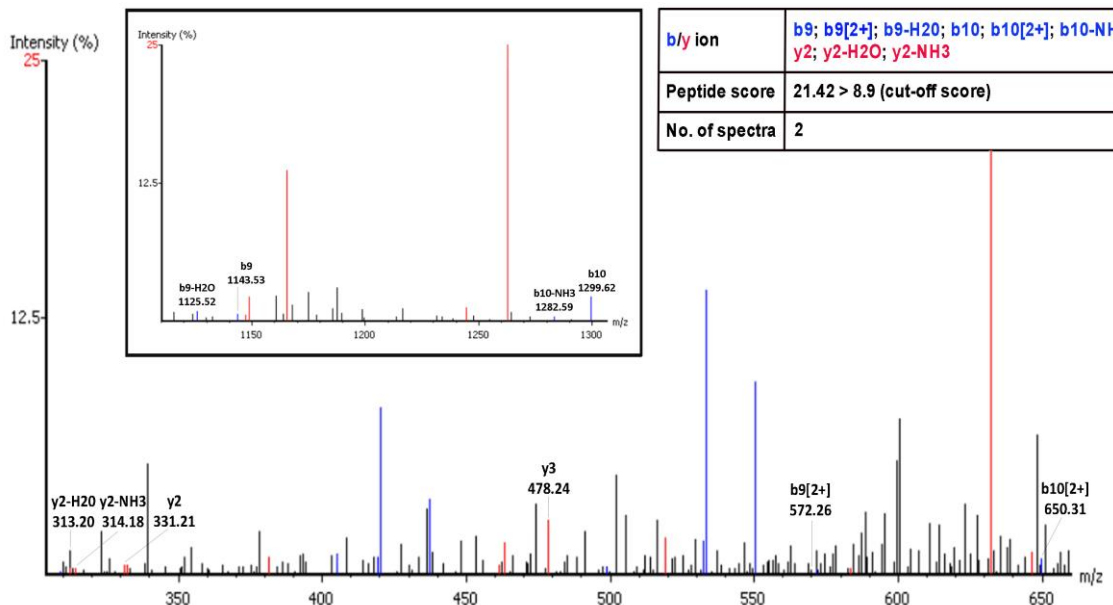
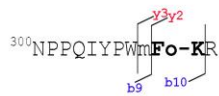
The evidence for formylation of K369 is identified b and y ions. For assigning that K369 is formylated in the peptide,  $^{367}\text{QIFo-K}(+27.99)\text{IWFQNR}$  ( $m/z$  630.85<sup>2+</sup>), the mass shift of b2 and b3 fragment ions and y6 and y7 fragment ions equaled the mass of a formylated lysine (Figure 20D). All 4 fragment ions supporting K369 formylation were detected at a minimal ion intensity greater than 5% and 3 independent spectra were identified for this formylated peptide.

MS/MS analysis of a thermolysin digest of SCRTT yielded two formylated peptides,  $^{431}\text{IDYFo-K}(+27.99)\text{DDDDKEN}$  ( $m/z$  699.29<sup>2+</sup>) and  $^{431}\text{IDYKDDDFo-K}(+27.99)\text{EN}$  ( $m/z$  699.29<sup>2+</sup>) of the same primary amino acid sequence and that are part of the triple tag (Figure 20E & F). The overall mass shift between the formylated and non-formylated peptides was 28 Da, which supports formylation of the peptides. For assigning that K434 is formylated in the peptide,  $^{431}\text{IDYFo-KDDDDKEN}$ , the mass shift of b3 and b4 fragment ions and y7 and y8 fragment ions equaled the mass of a formylated lysine (Figure 20E). For assigning that K439 is formylated in the peptide,  $^{431}\text{IDYKDDDAc-KEN}$ , the mass shift of b8 and b9 fragment ions equaled that of a formylated lysine (Figure 20F). All b and y ions that support K434 and K439 formylation were detected at a minimal ion intensity of greater than 5%. For the peptide,  $^{431}\text{IDYFo-KDDDDKEN}$ , 2 independent spectra were obtained.

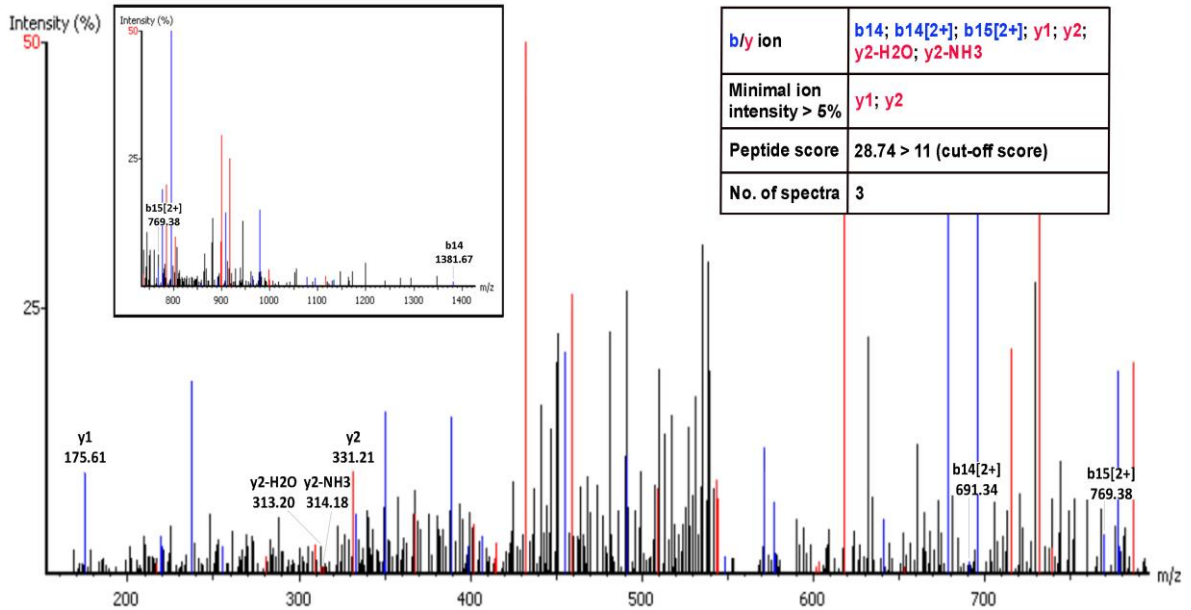
### A MS<sup>2</sup> *m/z* 899.98



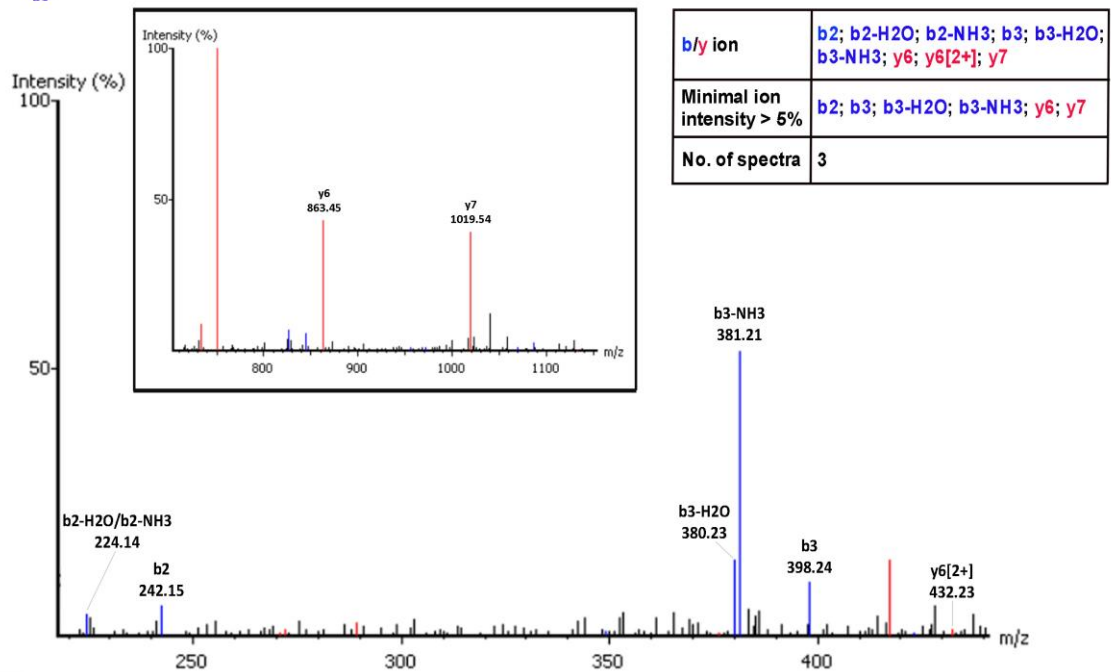
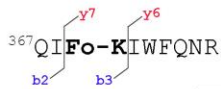
### B MS<sup>2</sup> *m/z* 737.37



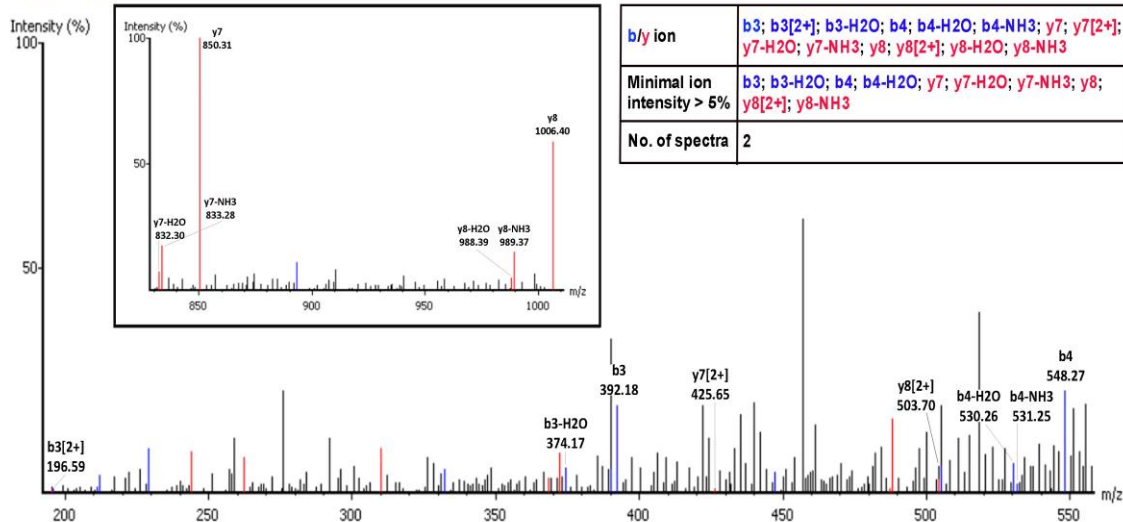
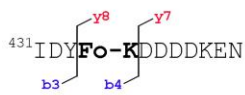
### C MS<sup>2</sup> *m/z* 571.3



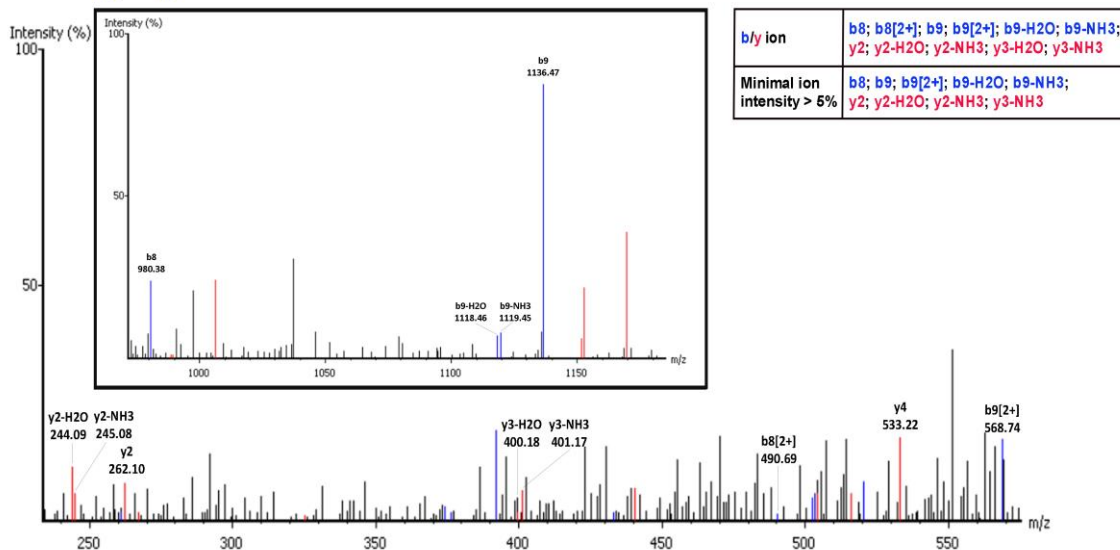
### D MS<sup>2</sup> *m/z* 630.85



**E MS<sup>2</sup> m/z 699.29**



**F MS<sup>2</sup> m/z 699.29**





**Figure 20. Formylation of Lysine 218, 309, 325, 369, 434 and 439 residues of**

**SCRTT.** MS<sup>2</sup> spectra of the peptide identified by LC-MS/MS is shown. A. Formylation of Lysine 218. The inset box shows fragment ions with  $m/z$  300 to 700. B. Formylation of Lysine 309. The inset box shows fragment ions with  $m/z$  1100 to 1300. C. Formylation of Lysine 325. The inset box shows fragment ions with  $m/z$  700 to 1400. D. Formylation of Lysine 369. The inset box shows fragment ions with  $m/z$  700 to 1100. E. Formylation of Lysine 434. The inset box shows fragment ions with  $m/z$  830 to 1010. F. Formylation of Lysine 439. The inset box shows fragment ions with  $m/z$  950 to 1200. The peptide sequence and  $m/z$  ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming formylation. The relevant fragment ions and their  $m/z$  ratios supporting formylation are labelled in the spectra.

### 3.8.4 Identification of methylated peptides of SCRTT

Four sites of methylation were identified by LC-MS/MS of SCRTT: Serine 19, Serine 166, Lysine 168 and Threonine 364 (Figure 21).

The evidence for methylation of S19 is an overall mass shift and b ions. The mass shift between the methylated and unmethylated peptide was 14 Da suggesting the attachment of a methyl group to the peptide. For assigning that S19 is methylated in the thermolytic peptide, <sup>17</sup>LAMe-S(+14.02)C(+57.02)YPQQ ( $m/z$  490.73<sup>2+</sup>), the mass shift of two fragment ions, b2 and b3 equaled the mass of a methylated serine residue (101 Da) (Figure 21A). The peptide score was 1.5-fold higher than the cut-off score and 3 independent spectra were obtained for the methylated peptide.

The evidence for methylation of S166 is an overall mass shift and b and y ions. The mass shift between the methylated and unmethylated peptide was 14 Da suggesting the attachment of a methyl group to the peptide. For assigning that S166 is methylated in the chymotryptic peptide,

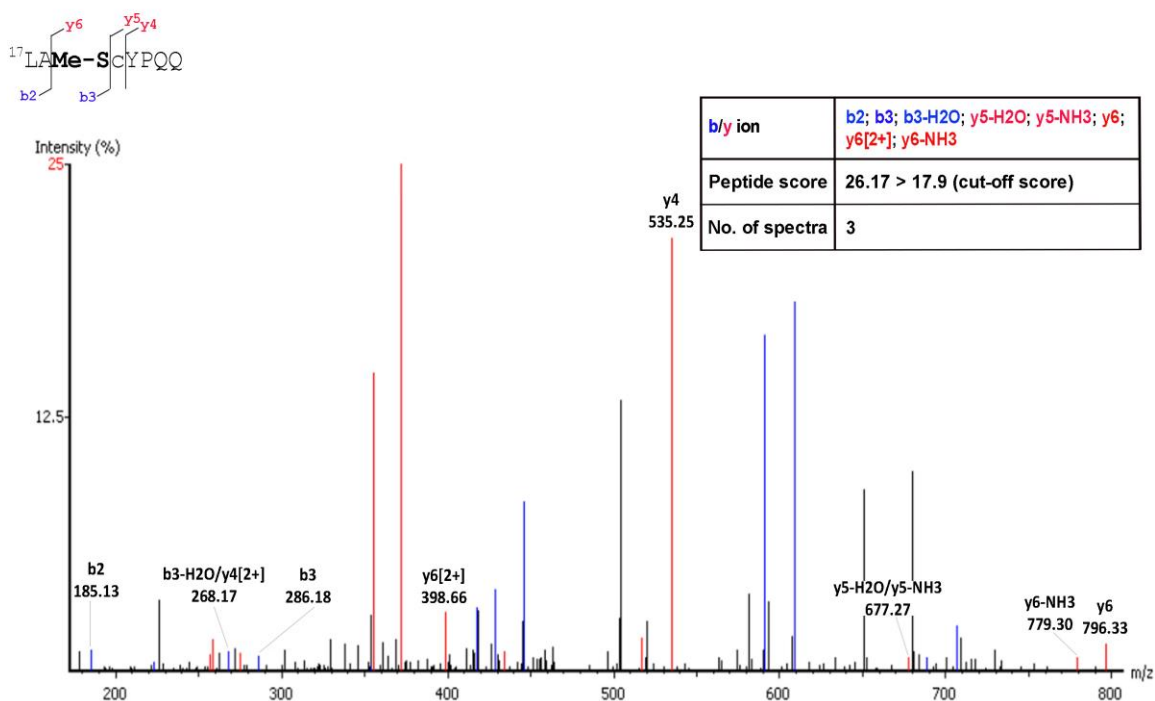
<sup>163</sup>ANIMe-S(+14.02)C(+57.02)KY ( $m/z$  435.21<sup>2+</sup>), the mass shift of b3 and b4 fragment

ions and  $y_3[2+]$  and  $y_4[2+]$  fragment ions equaled the mass of a methylated serine residue (101 Da) (Figure 21B). The fragment ion  $b_3$  supporting S166 methylation was detected at a minimal ion intensity of greater than 5%.

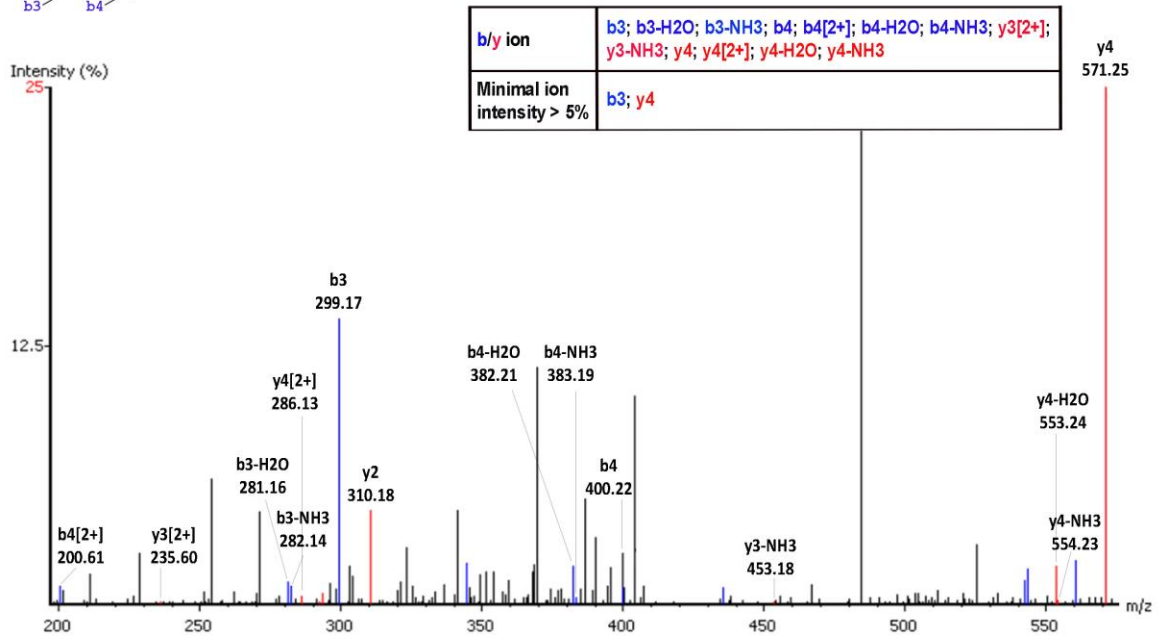
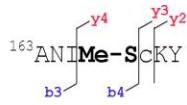
The evidence for methylation of K168 is identified  $y$  ions. For assigning that K168 is methylated in the chymotryptic peptide,  $^{156}\text{QQQQQQANISC}(+57.02)\text{Me-K}(+14.02)\text{Y}$  ( $m/z$  589.28<sup>3+</sup>), the mass shift of  $y_1$  and  $y_2$  fragment ions equaled the mass of a methylated lysine residue (142 Da) (Figure 21C). The peptide score was 2-fold higher than the cut-off score.

The evidence for methylation of T364 is an overall mass shift and identified  $b$  and  $y$  ions. For mapping methylation to T364 in the tryptic peptide,  $^{355}\text{IEIAHALC}(+57.02)\text{LMe-T}(+14.02)\text{ER}$  ( $m/z$  720.39<sup>2+</sup>), the mass shift of  $b_9[2+]$  and  $b_{10}[2+]$  fragment ions and  $y_2$  and  $y_3$  fragment ions equaled the mass of a methylated threonine residue (115 Da) (Figure 21D). For the methylated peptide, 3 independent spectra were obtained.

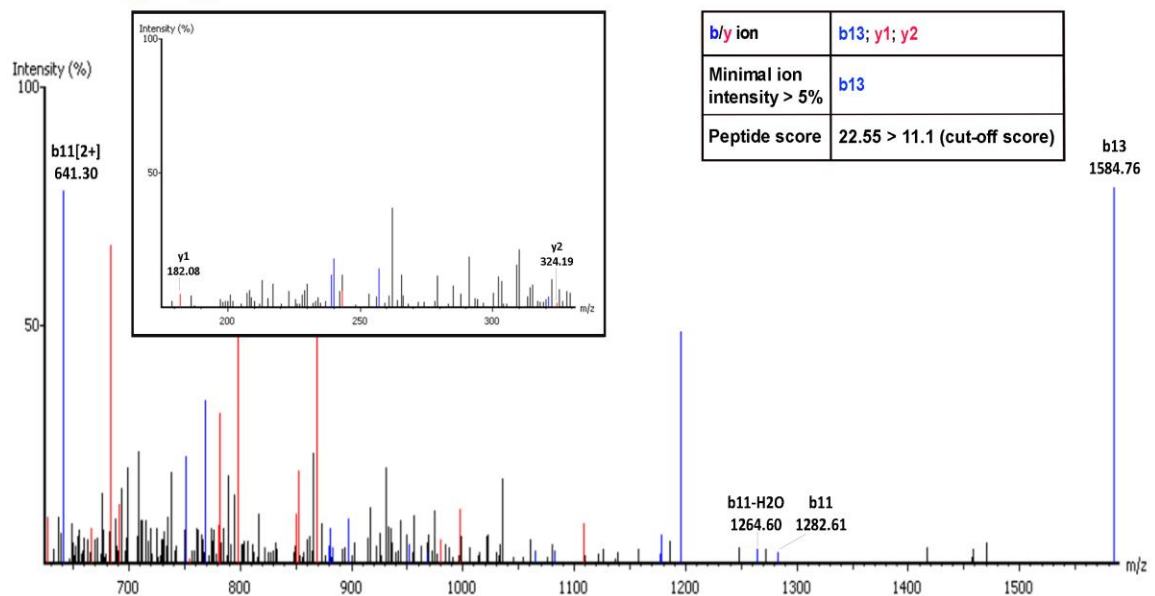
### A MS<sup>2</sup> $m/z$ 490.73



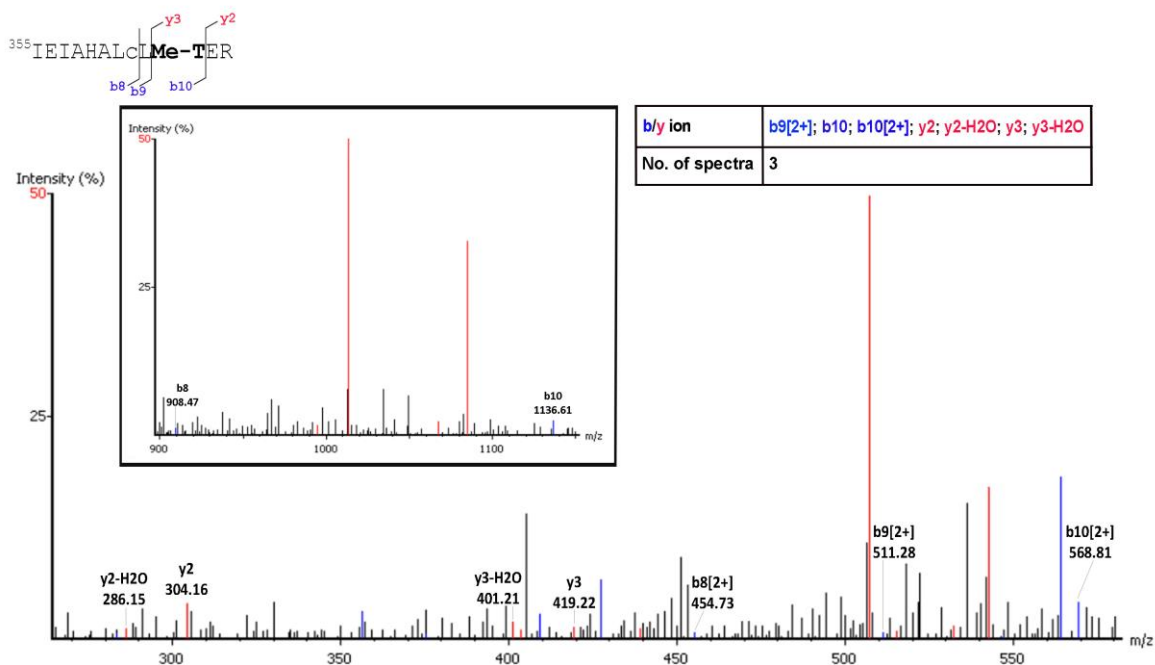
## B MS<sup>2</sup> *m/z* 435.21



## C MS<sup>2</sup> *m/z* 589.28



## D MS<sup>2</sup> *m/z* 720.39



**Figure 21. Methylation of Serine 19, Serine 166, Lysine 168 and Threonine 364 residues of SCRTT.** MS<sup>2</sup> spectra of the peptide identified by LC-MS/MS is shown. A. Methylation of Lysine 19. B. Methylation of Serine 166. C. Methylation of Lysine 168. The inset box shows fragment ions with *m/z* 180 to 330. D. Methylation of Threonine 364. The inset box shows fragment ions with *m/z* 900 to 1150. The peptide sequence and *m/z* ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming methylation. The relevant fragment ions and their *m/z* ratios supporting methylation are labelled in the spectra.

### 3.8.5 Identification of carboxylated peptides of SCRTT

Three sites of carboxylation were identified by LC-MS/MS of SCRTT: Tryptophan 307, Lysine 309 and Glutamic acid 323 (Figure 22).

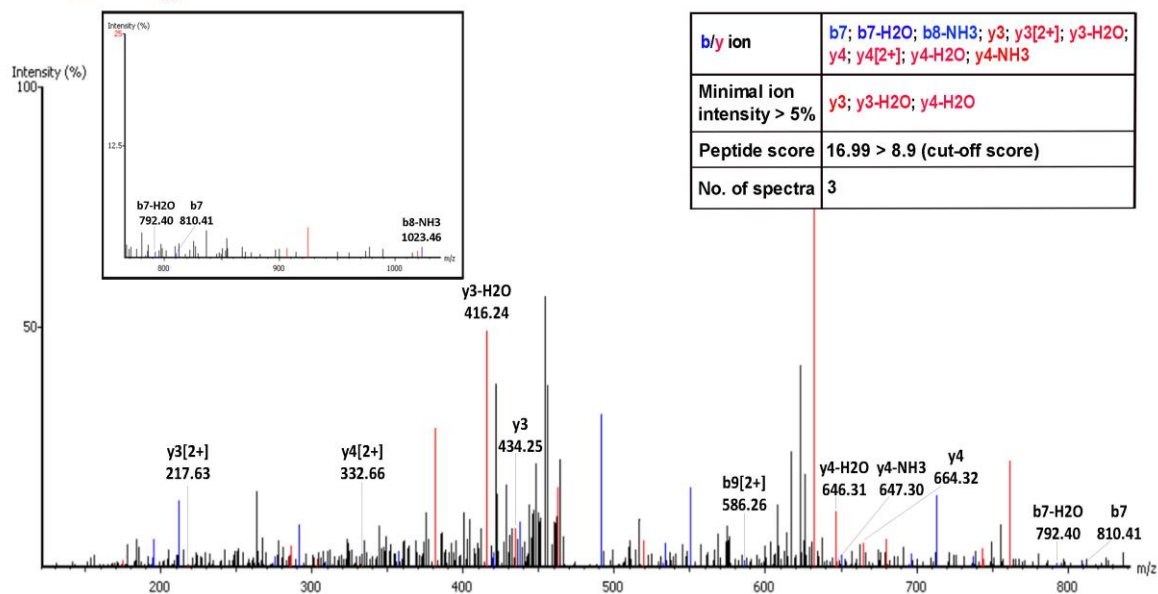
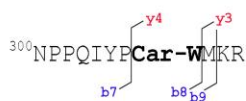
Three chemically distinct peptides carboxylated at W307 were identified by MS/MS of a tryptic digest of SCRTT (Table 8). The evidence for carboxylation of W307 is identified y ions. For assigning that W307 is carboxylated in the peptide,  $^{300}\text{NPPQIYPCar-W(+43.99)MKR}$  ( $m/z$  491.92 $^{3+}$ ), the mass shift of y3 and y4 fragment ions equaled the mass of a carboxylated tryptophan (230 Da) (Figure 22A). The fragment ion y3 supporting W307 carboxylation was detected at a minimal ion intensity of greater than 5%. The peptide score was 2-fold higher than the cut-of score and 3 independent spectra were obtained for the carboxylated peptide.

Three chemically distinct peptides carboxylated at K309 were identified by MS/MS of a tryptic digest of SCRTT (Table 8). The evidence for K309 carboxylation is identified b ions. For assigning that K309 is carboxylated in the peptide,  $^{300}\text{NPPQIYPWMCar-K(+43.99)R}$  ( $m/z$  737.37 $^{2+}$ ), the mass shift of b9[2+] and b10 fragment ions equaled the mass of a carboxylated lysine residue (171 Da) (Figure 22B). The peptide score is 2.2-fold higher than the cut-off score and 2 independent spectra were obtained for the carboxylated peptide.

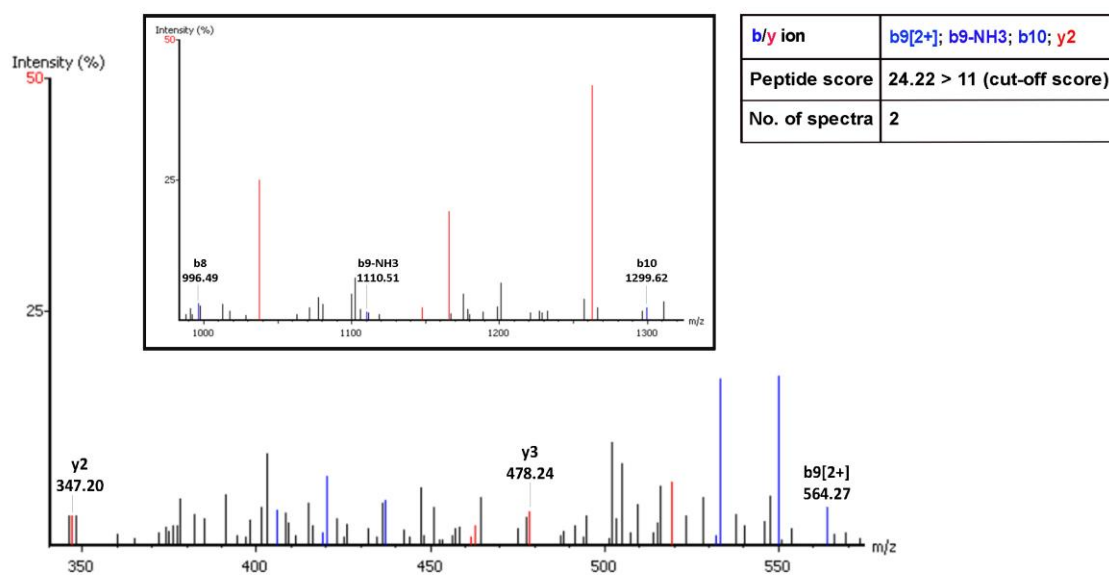
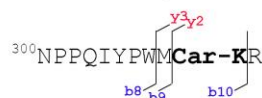
MS/MS analysis of SCRTT independently digested with trypsin and chymotrypsin identified two chemically distinct peptides carboxylated at E323 (Table 8). The evidence for E323 carboxylation is an overall mass shift and identified b and y ions. The tryptic peptide,

$^{311}\text{VHLGTSTVNANGCar-E(+43.99)TKR}$  ( $m/z$  576.63 $^{3+}$ ) was shifted from its non-carboxylated moiety by 44 Da, which is the mass of the carboxyl group, thereby suggesting the attachment of a carboxyl group to the peptide. For assigning that E323 is carboxylated in the tryptic peptide,  $^{311}\text{VHLGTSTVNANGCar-ETKR}$  ( $m/z$  576.63 $^{3+}$ ), the mass shifts of b12[2+] and b13[2+] fragment ions and y3 and y4 fragment ions equaled the mass of a carboxylated glutamic acid residue (173 Da) (Figure 22C). Fragment ions b12[2+], y3 and y4 were detected at a minimal ion intensity of greater than 5%. The peptide score was 1.8-fold higher than the cut-off score.

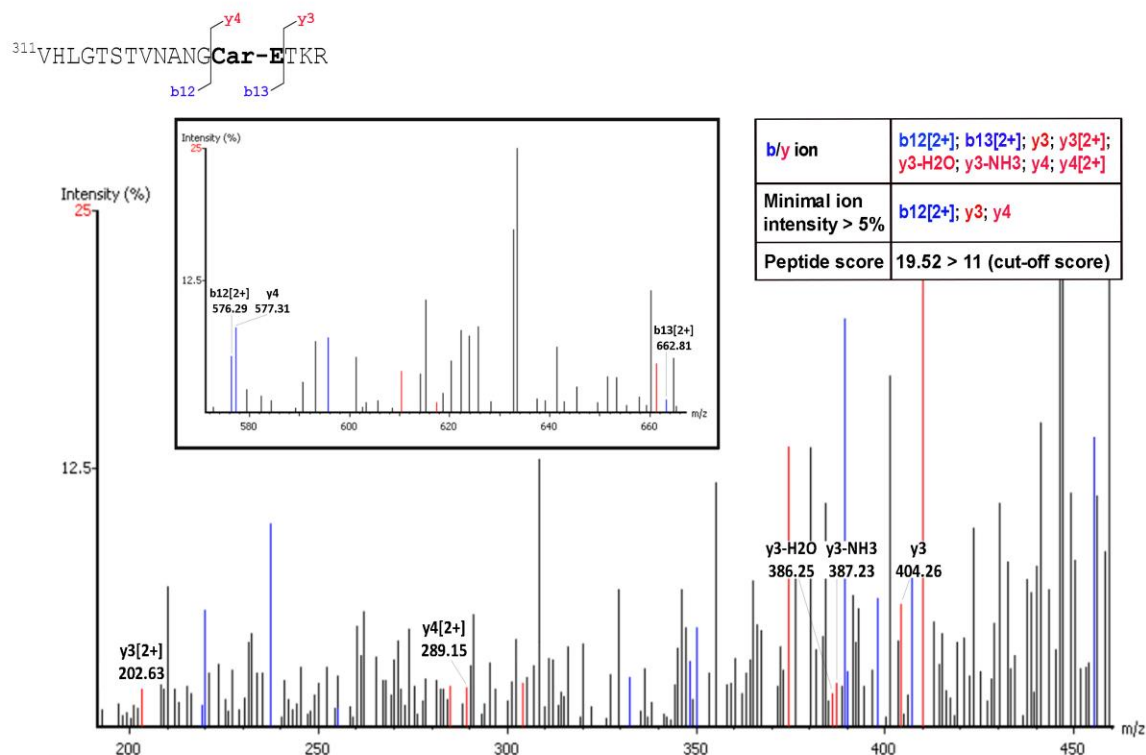
## A MS<sup>2</sup> *m/z* 491.92



## B MS<sup>2</sup> *m/z* 737.37



### C MS<sup>2</sup> *m/z* 576.63



**Figure 22. Carboxylation of Tryptophan 307, Lysine 309 and Glutamic acid 323 residues of SCRTT.** MS<sup>2</sup> spectra of the peptide identified by LC-MS/MS is shown. A. Carboxylation of Tryptophan 307. The inset box shows fragment ions with *m/z* 780 to 1030. B. Carboxylation of Lysine 309. The inset box shows fragment ions with *m/z* 990 to 1300. C. Carboxylation of Glutamic acid 323. The inset box shows fragment ions with *m/z* 570 to 670. The peptide sequence and *m/z* ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming carboxylation. The relevant fragment ions and their *m/z* ratios supporting carboxylation are labelled in the spectra.

### 3.8.6 Identification of hydroxylated peptides of SCRTT

Six sites of hydroxylation were identified by LC-MS/MS of SCRTT: Proline 22, Proline 107, Proline 269, Proline 306, Aspartic acid 108 and Aspartic acid 111 (Figure 23).

The evidence for hydroxylation of P22 is identified b and y ions. For assigning that P22 is hydroxylated in the thermolytic peptide,

$^{17}\text{LASC}(+57.02)\text{YHyd-P}(+15.99)\text{QQMNPQQNHPGAGNSS}$  ( $m/z$  801.35<sup>3+</sup>), the mass shift of b5 and b6 fragment ions and y16[2+] and y17[2+] fragment ions equaled 113 Da which is the mass of a hydroxylated proline (Figure 23A). Fragment ions b5 and y17[2+] were detected at a minimal ion intensity of greater than 5%.

Four chemically distinct peptides with P107 hydroxylated were identified by LC-MS/MS of which two were tryptic, one chymotryptic and one thermolytic (Table 8). The evidence of P107 hydroxylation is an overall mass shift and identified b and y ions. The thermolytic peptide,

$^{101}\text{YGGQGNHyd-P}(+15.99)\text{DMVDYTQ}$  ( $m/z$  780.82<sup>2+</sup>) was shifted by 16 Da relative to the mass of the non-hydroxylated peptide suggesting the attachment of a hydroxyl group to the peptide. For mapping hydroxylation to P107 in the peptide,  $^{101}\text{YGGQGNHyd-PDMVDYTQ}$  ( $m/z$  780.82<sup>2+</sup>), the mass shift of b6 and b7 fragment ions and y7 and y8 fragment ions was 113 Da, which is the mass of a hydroxylated proline (Figure 23B). All 4 b and y ions confirming P107 hydroxylation were detected at a minimal ion intensity of greater than 5%. For the hydroxylated peptide, 2 independent spectra were obtained.

Six chemically distinct peptides with D108 hydroxylated were identified by LC-MS/MS. Of the six, there were three tryptic, two chymotryptic and one thermolytic peptides (Table 8). The evidence of D108 hydroxylation is identified b and y ions. For mapping hydroxylation to D108 in the peptide,  $^{102}\text{GGQGNPHyd-D}(+15.99)\text{MVDYTQ}$  ( $m/z$  699.29<sup>2+</sup>), the mass shift of b6 and b7 fragment ions and y6 and y7 fragment ions was 131 Da, which is the mass of a hydroxylated aspartic acid (Figure 23C). The fragment ions b6, y6 and y7 were detected at a minimal ion intensity of greater than 5%. The peptide score was 2-fold higher than the cut-off score and 2 independent spectra were obtained for the hydroxylated peptide.



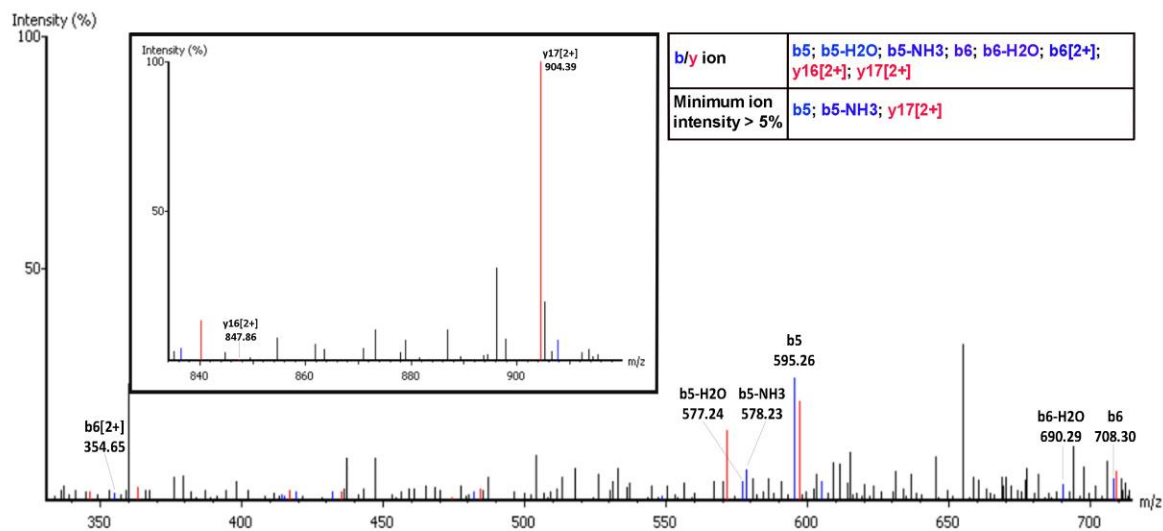
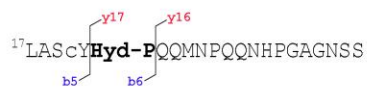
The evidence of D111 hydroxylation is identified b and y ions. For mapping hydroxylation to D111 in the peptide,

<sup>102</sup>GGQGNPDMVHyd-D(+15.99)YTQLQPQR ( $m/z$  673.98<sup>3+</sup>), the mass shift of b9 and b10 fragment ions and y8 and y9[2+] fragment ions was 131 Da, which is the mass of a hydroxylated aspartic acid (Figure 23D). Fragment ions b10, y8 and y9[2+] were detected at a minimal ion intensity of greater than 5%.

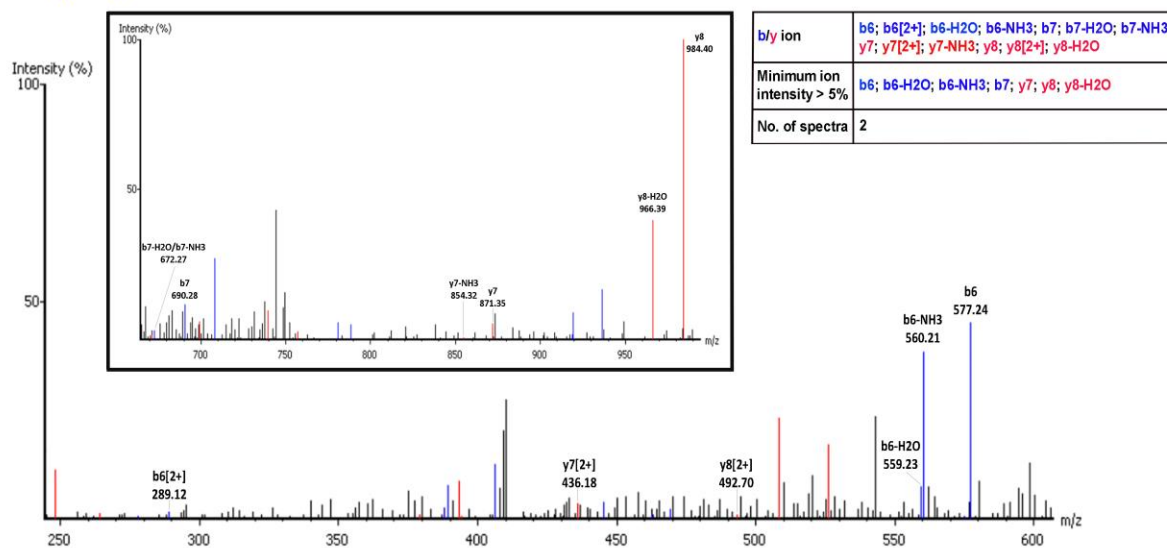
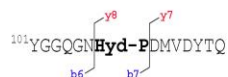
Three chemically distinct thermolytic peptides with P269 hydroxylated were identified in LC-MS/MS analysis of SCRIT (Table 8). The evidence of P269 hydroxylation is based on identified y ions. For mapping hydroxylation to P269 in the peptide, <sup>266</sup>VNVHyd-P(+15.99)MHSPGGGDSSESDSGNE ( $m/z$  1095.43<sup>2+</sup>), the mass shift of y18 and y19 fragment ions was 113 Da, which is the mass of a hydroxylated proline (Figure 23E). The fragment ion y19 confirming P269 hydroxylation was detected at a minimal ion intensity of greater than 5%. The peptide score was 2-fold higher than the cut-off score.

The evidence of P306 hydroxylation is identified b and y ions. For mapping hydroxylation to P306 in the peptide, <sup>300</sup>NPPQIYHyd-P(+15.99)WMKR ( $m/z$  482.58<sup>3+</sup>), the mass shift of b6 and b7 fragment ions and y4 and y5 fragment ions was 113 Da, which is the mass of a hydroxylated proline (Figure 23F). Fragment ion y5 confirming P306 hydroxylation was detected at a minimal ion intensity of greater than 5%.

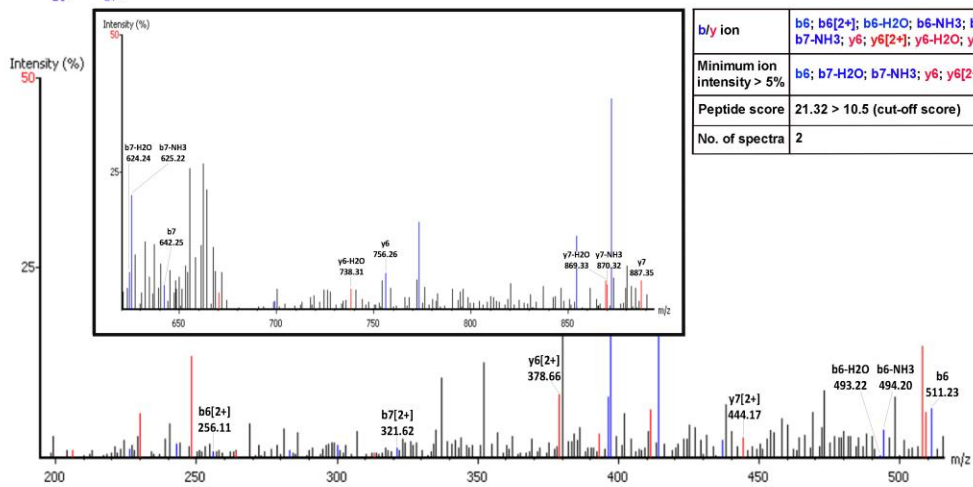
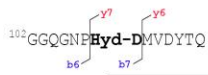
### A MS<sup>2</sup> m/z 801.35



### B MS<sup>2</sup> m/z 780.82

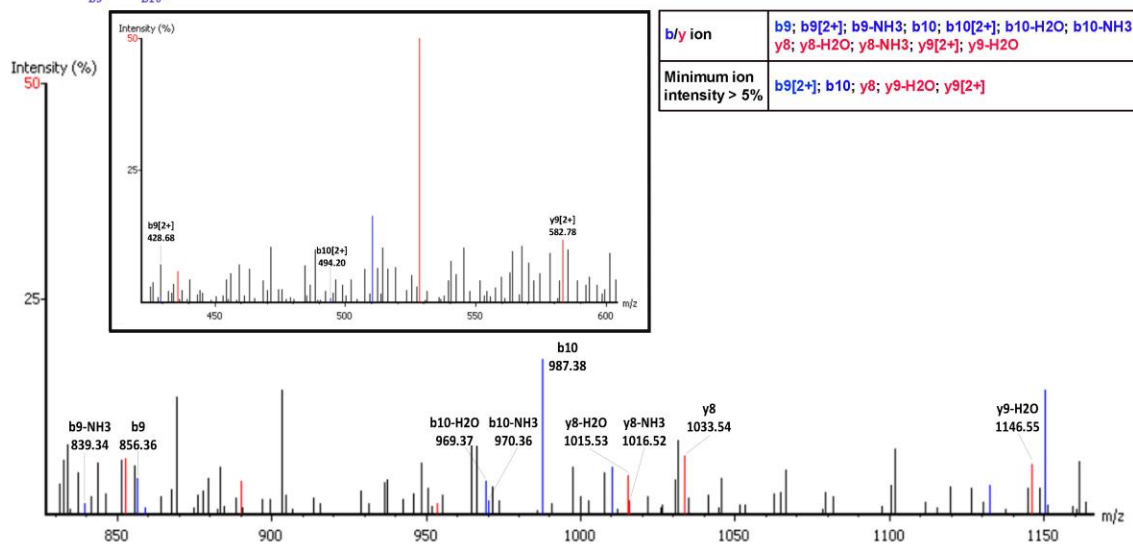
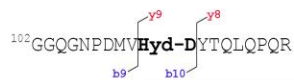


### C MS<sup>2</sup> m/z 699.29

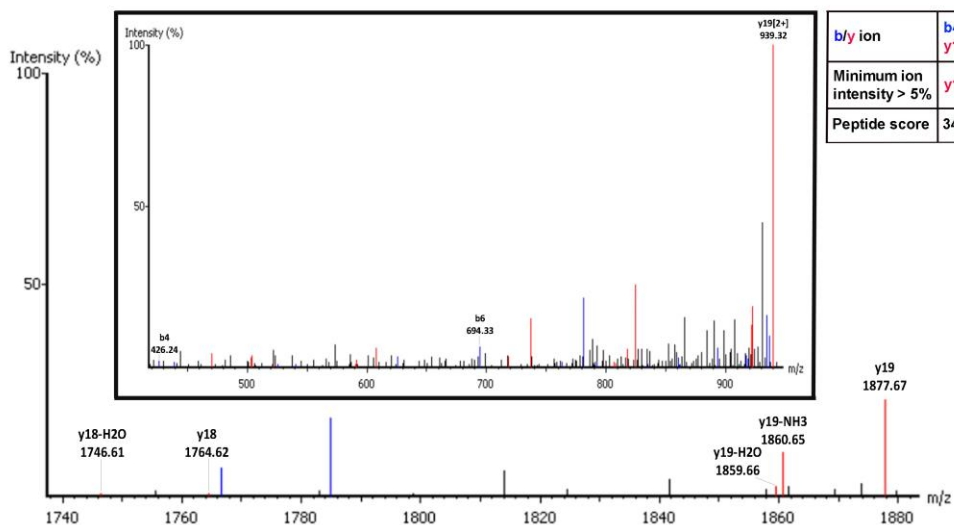
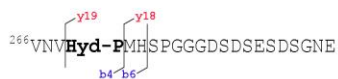
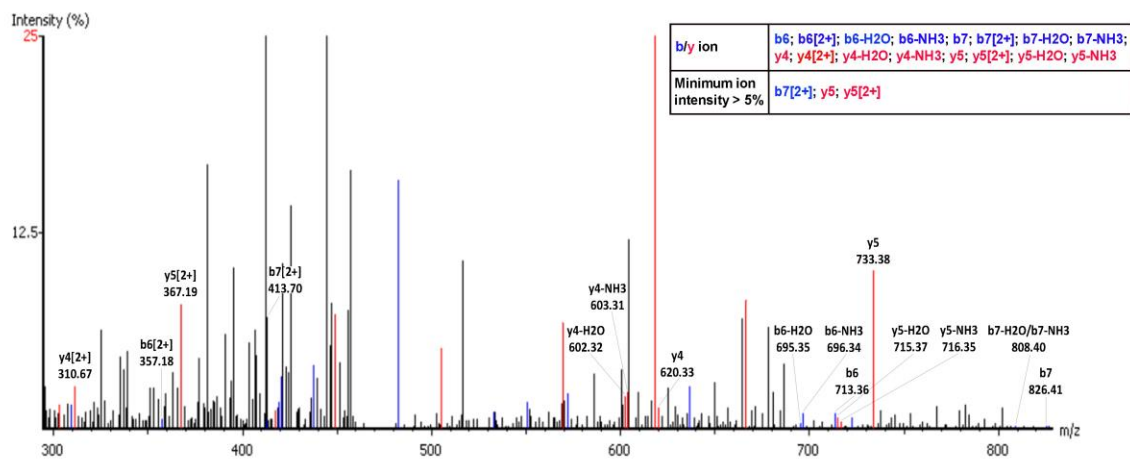
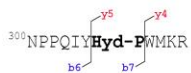


b/y ion	b6; b6[2+]; b6-H2O; b6-NH3; b7; b7[2+]; b7-H2O; b7-NH3; y6; y6[2+]; y6-H2O; y7; y7[2+]; y7-H2O; y7-NH3
Minimum ion intensity > 5%	b6; b7-H2O; b7-NH3; y6; y6[2+]; y7; y7-H2O
Peptide score	21.32 > 10.5 (cut-off score)
No. of spectra	2

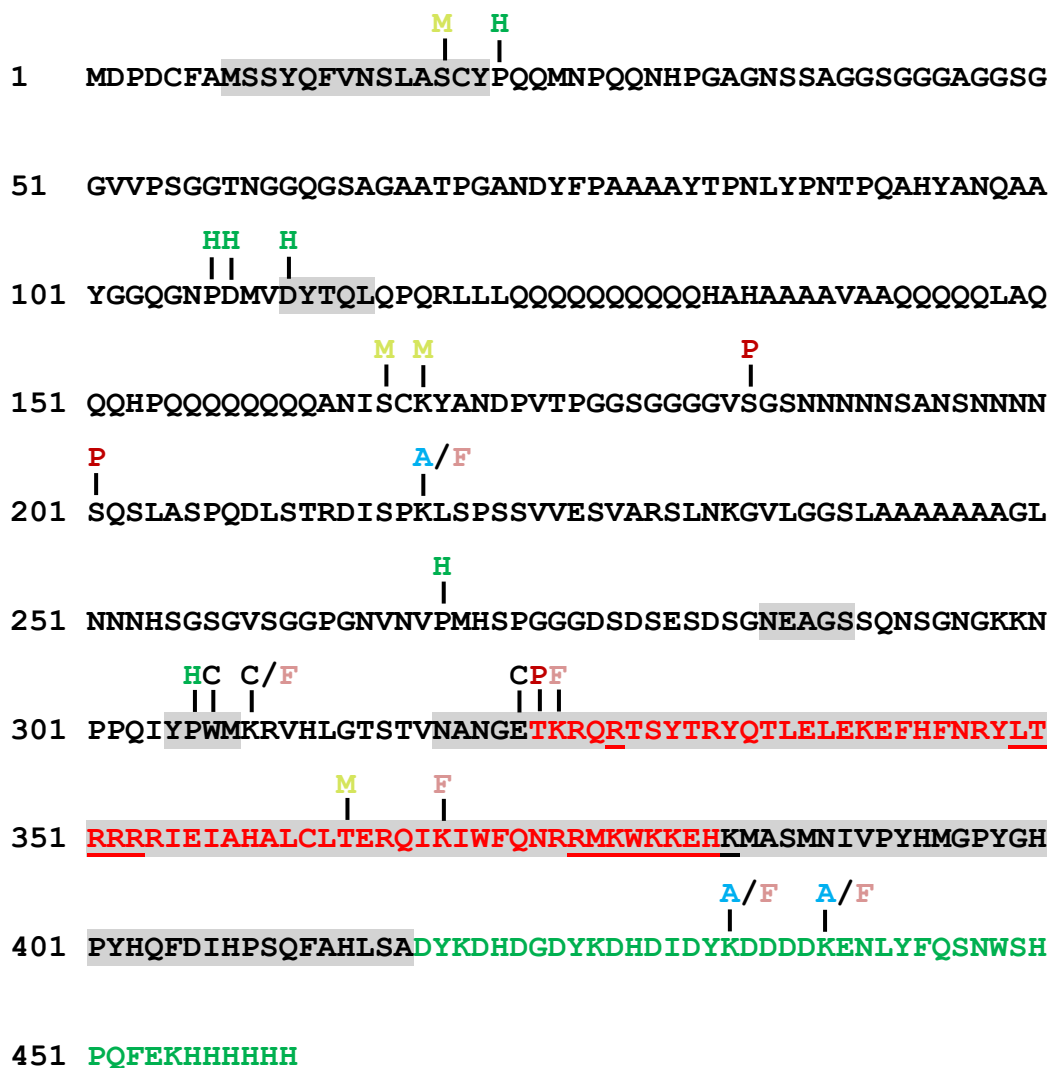
### D MS<sup>2</sup> m/z 673.98



b/y ion	b9; b9[2+]; b9-NH3; b10; b10[2+]; b10-H2O; b10-NH3; y8; y8-H2O; y8-NH3; y9[2+]; y9-H2O
Minimum ion intensity > 5%	b9[2+]; b10; y8; y9-H2O; y9[2+]

E MS<sup>2</sup> *m/z* 1095.43F MS<sup>2</sup> *m/z* 482.58

**Figure 23. Hydroxylation of Proline 22, 107, 269, 306, Aspartic Acid 108 and 111 residues of SCRTT.** MS<sup>2</sup> spectra of the peptide identified by LC-MS/MS is shown. A. Hydroxylation of Proline 22. The inset box shows fragment ions with  $m/z$  835 to 920. B. Hydroxylation of Proline 107. The inset box shows fragment ions with  $m/z$  670 to 1000. C. Hydroxylation of Aspartic acid 108. The inset box shows fragment ions with  $m/z$  620 to 900. D. Hydroxylation of Aspartic acid 111. The inset box shows fragment ions with  $m/z$  420 to 600. E. Hydroxylation of Proline 269. The inset box shows fragment ions with  $m/z$  420 to 940. F. Hydroxylation of Proline 306. The peptide sequence and  $m/z$  ratio are indicated at the top of the spectra. Positions of fragmentation are shown with vertical lines in the peptide sequence. The box on the right summarizes the evidence confirming hydroxylation. The relevant fragment ions and their  $m/z$  ratios supporting hydroxylation are labelled in the spectra.



**Figure 24. 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.

Few PTMs mapped by MS/MS analysis of bacterially isolated SCRTT were also detected in SCRTT isolated from developing embryos (Table 9).

**Table 9. Post-translational modifications common to bacterial and embryonic SCRTT**

Post-translational modifications	Site(s) modified
Acetylation	K439
Carboxylation	W307, K309, E323
Formylation	K309, K325, K439
Hydroxylation	P306, P107, D108, D111

### 3.9 Short Linear Motifs (SLiMs) in HOX and FTZ proteins

Besides the highly conserved 60 amino acid DNA binding homeodomain, HOX and FTZ transcription factors also possess short linear motifs (SLiMs) of several amino acids in length. SLiMs potentially act as protein interaction sites and sites of phosphorylation. To test whether SLiMs were preferential sites of HOX phosphorylation, I screened the Eukaryotic Linear Motif (ELM) resource (Dinkel *et al.*, 2016) for predicted SLiMs in *Drosophila melanogaster* HOX and FTZ proteins (Tables S1, S3, S5, S7, S9, S11, S13, S15, S17 & S19). Only the SLiMs found in proteins localized to the nucleus and cytoplasm were considered in this study, as HOX proteins are localized to the nucleus and have been shown to interact with both nuclear and cytoplasmic components (Merabet & Dard, 2014; Wiellette *et al.*, 1999). Some SLiMs also corresponded with regions of HOX proteins conserved at various taxonomic levels (Tables S2, S4, S6, S8, S10, S12, S14, S16 & S18; Figures S4-12). These SLiMs named ‘conserved SLiMs’, are a

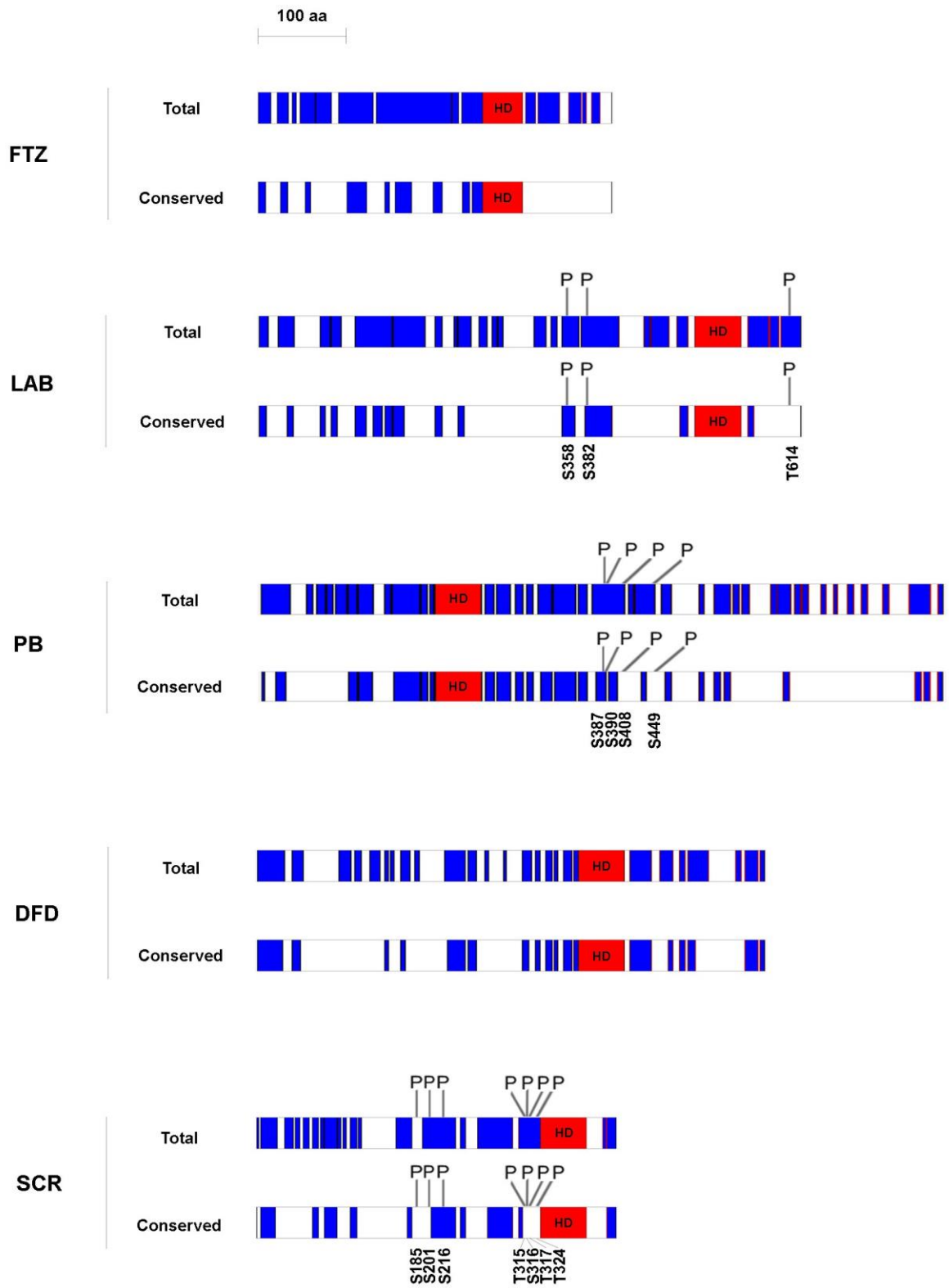
subgroup of the total SLiMs obtained from the ELM resource. For HOX proteins, 46-66% of the primary sequence were SLiM sequences and 8-47% were conserved SLiM sequences and for FTZ, 81% of the primary sequence was SLiM sequences and 27% was conserved SLiM sequences (Table 10 & Figure 25).

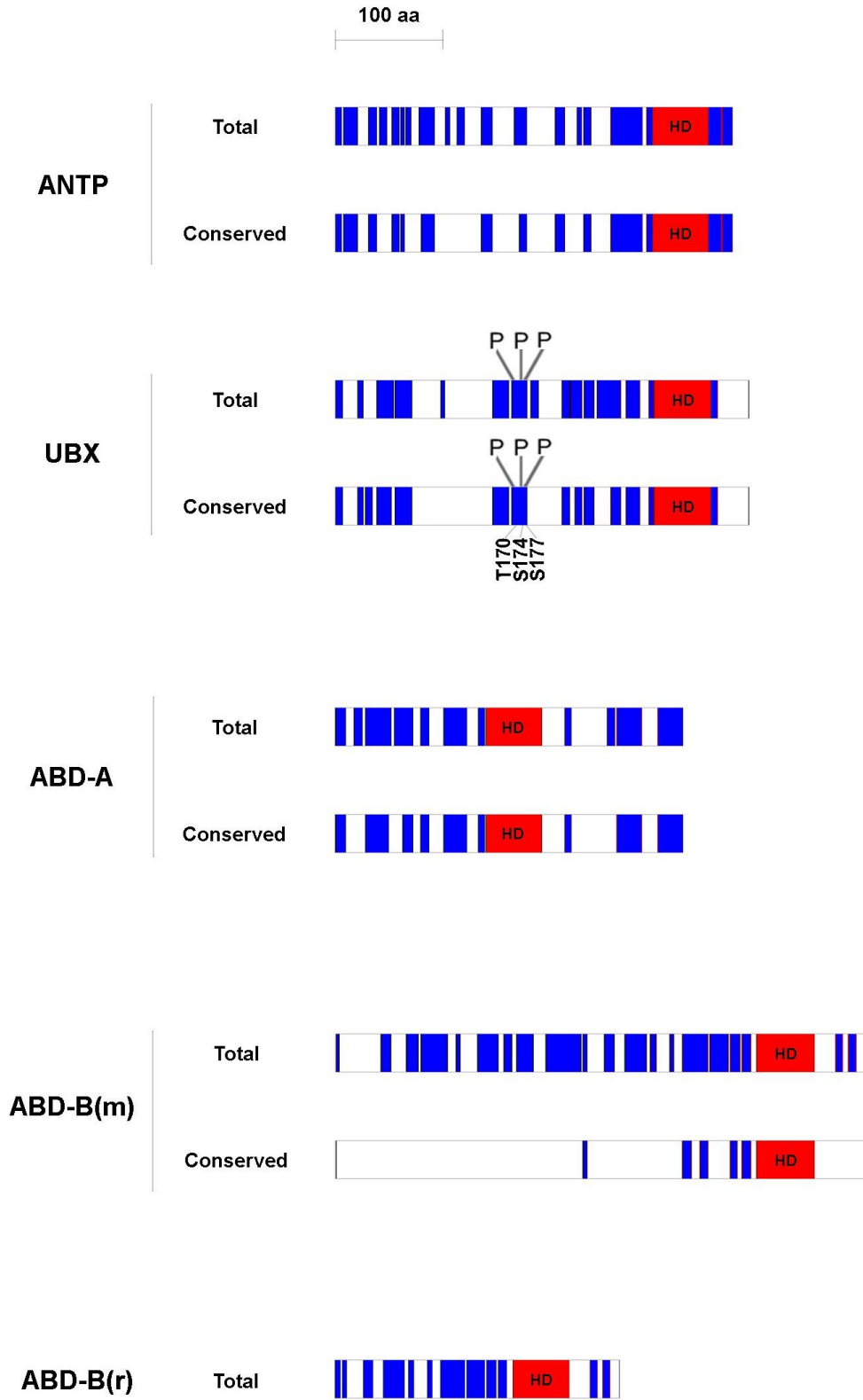
**Table 10. SLiMs in *Drosophila melanogaster* HOX and FTZ proteins**

Protein	Protein (disordered) size (aa)	No. of SLiMs	SLiM size (aa)	% SLiM	No. of conserved SLiMs	Conserved SLiM size (aa)	% Conserved SLiM
FTZ	364	137	295	81	41	100	27
LAB	576	139	366	64	57	149	26
PB	724	169	461	64	91	255	35
DFD	533	84	287	54	55	188	35
SCR	364	81	239	66	51	129	35
ANTP	325	57	175	54	48	140	43
UBX	336	51	156	46	42	133	40
ABD-A	277	45	156	56	35	131	47
ABD-B(m)	440	83	251	57	10	34	8
ABD-B(r)*	217	34	125	58	-	-	-

\* - Conserved SLiM analysis was not performed for ABD-B(r).







**Figure 25. Schematic of HOX and FTZ proteins 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 S1-19). For LAB, PB, SCR and UBX, phosphosites are marked to scale.

### 3.10 Distribution of S, T and Y is biased towards SLiMs

I tested whether the distribution of S, T and Y was biased toward regions of the proteins predicted to be SLiMs, that is S, T and Y residues are preferentially located in SLiMs. I determined the percentage of the primary amino acid sequence encompassed by SLiMs that were S, T and Y and the percentage of S, T and Y in non-SLiM regions (excluding the ordered portion of the HD) for all HOX and FTZ proteins (Table 11). I found consistently the percentage of S, T and Y to be higher in the SLiMs and conserved SLiMs than in non-SLiM portions of the HOX and FTZ proteins. This difference is significant for the LAB, PB, SCR, ANTP, UBX and ABD-A proteins for both total and conserved SLiMs. For FTZ, the difference was not significant for total and conserved SLiMs. For DFD and ABD-B(m), the total SLiM were significant but for the conserved SLiMs, the  $p$ -values were greater than 0.05 (Fisher, 1922). I conclude that the distribution of S, T and Y in HOX proteins is biased towards SLiMs.

**Table 11. S, T and Y distribution in SLiMs vs. non-SLiMs and their significance**

Protein	Percentage of amino acid that are S, T and Y				<i>p</i> -value	
	Total SLiMs		Conserved SLiMs		Total SLiMs	Conserved SLiMs
	SLiMs	Non-SLiMs	SLiMs	Non-SLiMs		
FTZ	32	23	34	29	0.19	0.371
LAB	33	6	34	19	$3.45 \times 10^{-15}$	$4.1 \times 10^{-4}$
PB	26	8	29	14	$5.73 \times 10^{-10}$	$2.62 \times 10^{-6}$
DFD	25	13	22	17	$3.9 \times 10^{-4}$	0.169
SCR	25	5	28	13	$3.38 \times 10^{-7}$	$5.7 \times 10^{-4}$
ANTP	23	5	25	8	$3.77 \times 10^{-6}$	$2 \times 10^{-5}$
UBX	26	9	28	10	$3 \times 10^{-5}$	$3 \times 10^{-5}$
ABD-A	21	4	22	6	$3 \times 10^{-5}$	$1.7 \times 10^{-4}$
ABD-B(m)	31	5	26	19	$4.02 \times 10^{-13}$	0.368
ABD-B(r)	30	12	-	-	$1.63 \times 10^{-3}$	-

### 3.11 Mapping of HOX phosphosites in SLiMs vs. non-SLiMs

To test whether SLiMs are preferential sites of phosphorylation, I collected all known phosphorylation sites in the *Drosophila* HOX proteins, LAB, PB, SCR and UBX and determined whether the phosphorylation sites were in SLiMs, conserved SLiMs or not in SLiMs (Table 12). I found that 16 out of 17 phosphosites were in a SLiM region of which 8 phosphosites were in conserved SLiMs (Table 12 & Figure 25). However, the Fisher's exact test (Fisher, 1922) revealed these 16:1 and 8:9 differences were not significant (Table 13). I conclude that there is no support for the hypothesis that SLiMs or conserved SLiMs are preferential sites of phosphorylation.

In addition, a minority of identified SLiMs were phosphorylated. In LAB, 20 out of 139 SLiMs and 11 out of 57 conserved SLiMs were phosphorylated (Tables S3 & S4). In PB, 8 out of 170 SLiMs and 5 out of 91 conserved SLiMs were phosphorylated (Tables S5 & S6). In SCR, 9 out of 81 SLiMs and 5 out of 51 conserved SLiMs were phosphorylated (Tables S9 & S10). In UBX, 6 out of 51 SLiMs and 5 out of 42 conserved SLiMs were phosphorylated (Tables S13 & S14). When the primary protein sequences of LAB, PB, SCR and UBX were conjugated and considered as one protein, 43 out of 441 SLiMs and 26 out of 241 conserved SLiMs were phosphorylated. This may indicate that only about 10% of SLiMs are *bona fida* sites of phosphorylation.

**Table 12. Mapping of HOX phosphosites in SLiMs vs. non-SLiMs**

<b>HOX Protein</b>	<b>Phosphosites</b>	<b>Map to SLiMs, conserved SLiMs or non-SLiMs</b>	<b>Extent of conservation</b>	<b>Relevant putative SLiM function (from ELM resource)</b>
LAB	S358*	Conserved SLiM	<i>Drosophila</i>	CK1, GSK3 and Proline-directed (eg., MAPK) phosphorylation site
	S382*	Conserved SLiM	<i>Drosophila</i>	CK1, GSK3 and Proline-directed (eg., MAPK) phosphorylation site
	T614*	SLiM	-	CK1, GSK3 and Proline-directed (eg., MAPK) phosphorylation site
PB	S387*	Conserved SLiM	<i>Drosophila</i>	CK1, GSK3 and Proline-directed (eg., MAPK) phosphorylation site
	S390*	Conserved SLiM	<i>Drosophila</i>	CK1, GSK3 and Proline-directed (eg., MAPK) phosphorylation site
	S408*	SLiM	-	PKA-type AGC kinase, CK1 and Plk1 phosphorylation site

	S449*	SLiM	-	GSK3 phosphorylation site
SCR	S185	Non-SLiM	-	-
	S201	SLiM	-	PIKK phosphorylation site
	S216 <sup>#</sup>	Conserved SLiM	<i>Drosophila</i>	GSK3, CDK and Proline-directed (eg., MAPK) phosphorylation site
	T315*	SLiM	-	NEK2 phosphorylation site
	S316*	SLiM	-	NEK2 phosphorylation site
	T317*	SLiM	-	NEK2 phosphorylation site
	T324	SLiM	-	-
UBX	T170 <sup>#</sup>	Conserved SLiM	<i>Drosophila</i>	GSK3 and NEK2 phosphorylation site
	S174 <sup>#</sup>	Conserved SLiM	Diptera & <i>Drosophila</i> <sup>^</sup>	CK1, GSK3 and Proline-directed (eg., MAPK) phosphorylation site
	S177 <sup>#</sup>	Conserved SLiM	Diptera & <i>Drosophila</i> <sup>^</sup>	CK1 and Proline-directed (eg., MAPK) phosphorylation site

\* - Phosphosites mapped by my supervisor, Dr. Anthony Percival-Smith of our laboratory (unpublished data)

<sup>#</sup> - Phosphosites mapped by a whole proteome analysis of *Drosophila* embryos (Zhai *et al.*, 2008)

<sup>^</sup> - Phosphosites S174 map to 4 conserved SLiM hits out of which 2 are Diptera-specific SLiMs and 2 are

*Drosophila*-specific SLiMs. Phosphosite S177 map to 2 conserved SLiM hits out of which 1 is Diptera-specific and the other *Drosophila*-specific (Table S14).

Data from Tables S3-6, S9, S10, S13 & S14 were used.

**Table 13. Phosphosite distribution in SLiMs vs. non-SLiMs and their significance**

Protein	Percentage of S, T and Y phosphorylated				<i>p</i> -value	
	Total SLiMs		Conserved SLiMs			
	SLiMs	Non-SLiMs	SLiMs	Non-SLiMs	Total SLiMs	Conserved SLiMs
LAB	3	0	4	1	1	0.558
PB	3	0	3	3	1	1
SCR	10	17	3	20	0.504	0.041
UBX	7	0	8	0	0.552	0.545
LAB+PB+SCR+UBX*	5	2	4	5	0.487	1

\* Considered as 1 protein by conjugating the primary protein sequence of all 4 proteins



## 4 Discussion

The objective of this study was to develop a comprehensive catalog of PTMs of tagged HOX and FTZ proteins extracted from the developing embryo and to test whether SLiMs are preferential sites of HOX and FTZ phosphorylation. In an effort to accomplish these objectives, ten transgenic fly lines each expressing a HOXTT protein or FTZTT were established. The full set of nine HOXTT proteins along with FTZTT were ectopically expressed from a heat-shock promoter. I showed that the HOXTT proteins were biologically active *in vivo*. I purified SCRTT from developing embryos and mapped PTMs of SCRTT by LC-MS/MS.

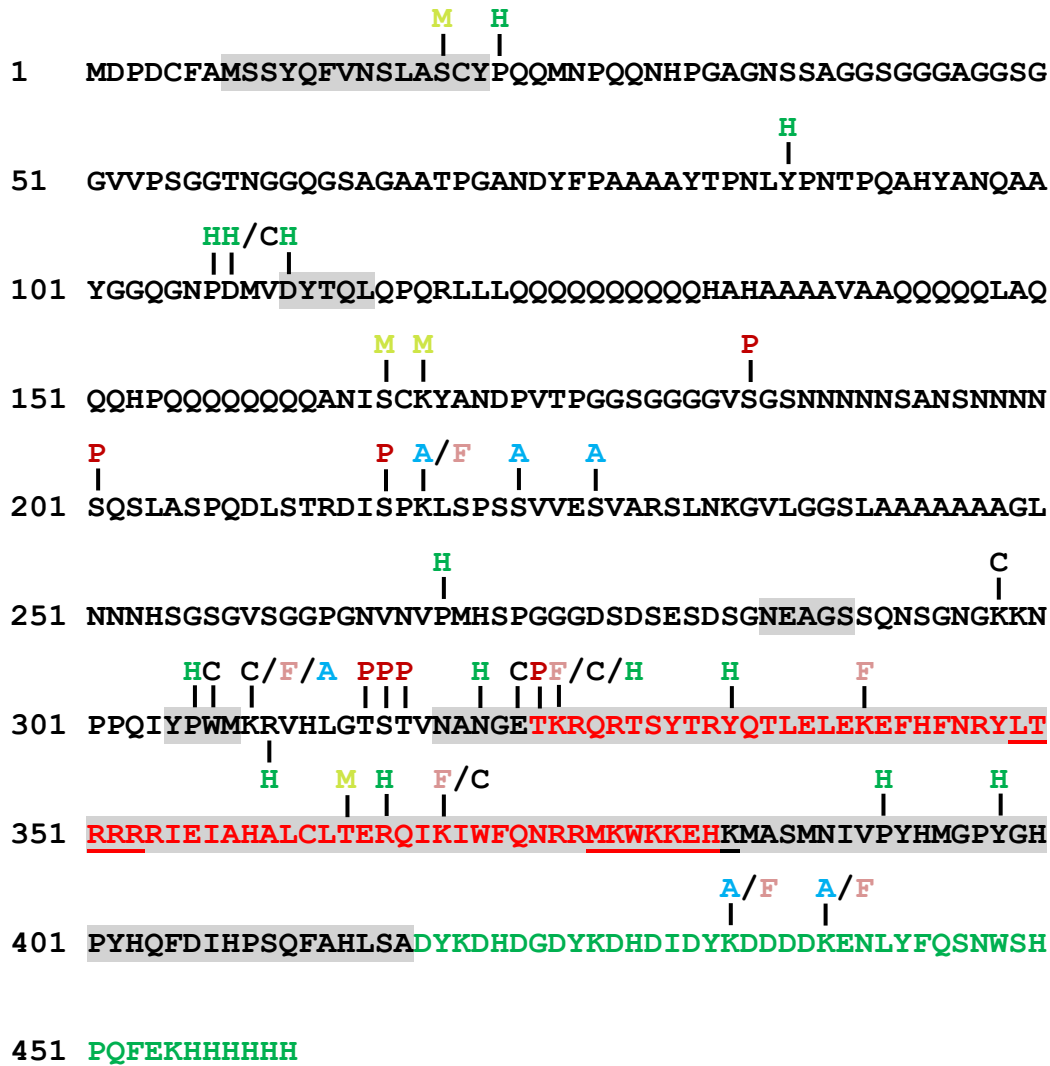
### 4.1 PTMs of embryonic SCRTT

Bottom-up MS/MS analysis of SCRTT purified from developing *D. melanogaster* embryos identified many amino acid residues that were covalently modified. This is the first report of the identification of PTMs, most importantly phosphorylation, of a HOX protein purified from developing embryos. It is important to distinguish between modifications that occurred *in vivo* which may be biologically relevant and potential *in vitro* artefactual modifications due to the chemistry of sample preparation. In this study, the identified covalent modifications that likely occurred due to sample preparation include carbamidomethylation of cysteines (Lapko *et al.*, 2000; Nielson *et al.*, 2006), carbamidomethylation of lysine, histidine, glutamic acid and aspartic acid and N-terminal amino groups (Boja & Fales, 2001), carbamylation or carbamoylation of lysine and other amino groups (Volkin *et al.*, 1997; Kollipara & Zahedi, 2013; Clauser *et al.*, 1995; Lipincott & Apostol, 1999), oxidation of methionine and tryptophan (Perdivara *et al.*, 2010; Stadtman & Levine, 2003), methyl esterification of glutamic and aspartic acid (Haebel *et al.*, 1998; Jung *et al.*, 2008; Sprung *et al.*, 2008; Chen *et al.*, 2010), acetylation at N-terminal of peptides (Choudhary *et al.*, 2009; Kelstrup *et al.*, 2014; Neilson *et al.*, 2006), formylation at N-terminal of peptides (Whitelegge *et al.*, 1998) and deamidation of asparagine and glutamine (Wright & Urry, 1991). Cysteines are deliberately carbamidomethylated by treatment with iodoacetamide prior to digestion with a protease in order to prevent formation of disulphide bonds between two cysteines (Lapko *et al.*,

2000). N- and O-carbamidomethylation of lysine, histidine, glutamic acid and aspartic acid residues and N-terminal amino groups were also identified which is probably an artifact caused by overalkylation in the presence of iodoacetamide (Boja & Fales, 2011). Carbamylation or carbamoylation of lysine and other amino groups were identified which are known to be induced by urea which was used as a denaturing agent during purification of HOXTT proteins (Volkin *et al.*, 1997; Kollipara & Zahedi, 2013; Clauser *et al.*, 1995; Lipincott & Apostol, 1999). Tryptophan oxidations were identified which are known to occur during separation of proteins by SDS-PAGE and in-gel digestion (Perdivara *et al.*, 2010). Reactive oxygen species oxidize methionine resulting in widespread methionine oxidation (Stadtman & Levine, 2003). Methyl esterification of aspartic and glutamic acid were identified which are known to occur due to reaction with methanol which is used as a cosolvent during Coomassie staining and destaining (Haebel *et al.*, 1998; Jung *et al.*, 2008; Sprung *et al.*, 2008; Chen *et al.*, 2010). Acetylation at the N-terminal of a peptide has been identified which is an artifact that occurs during SDS gel electrophoresis (Choudhary *et al.*, 2009; Kelstrup *et al.*, 2014; Neilson *et al.*, 2006). Many peptides were identified with N-terminal formylation, which might be due to the presence of formic acid (Whitelegge *et al.*, 1998). Asparagine and glutamine were found to be deamidated, which occurs nonenzymatically, and therefore, may be an artifact (Wright & Urry, 1991). Although all these modifications may be the result of sample handling, there is the potential that some of these PTMs, like spontaneous deamidation, occur *in vivo* and have an effect on protein activity. An example in SCR is the spontaneous deamidation of N321, the second N of the NANGE motif, which has been shown to affect the DNA binding of SCR (O'Connell *et al.*, 2015). This deamidation is observed in 28 of 75 peptides containing N321 (Figure S3).

After careful investigation to remove potential artefactual modifications, PTMs of potential biological importance were identified and reported. These include phosphorylation at S185, S201 and T324, acetylation at K218, K434 and K439, formylation at K218, K309, K325, K369, K434 and K439, methylation at S19, S166, K168 and T364, carboxylation at W307, K309 and E323 and hydroxylation at P22, P107, D108, D111, P269 and P306. Dr. Anthony Percival-Smith, my supervisor, independently identified PTMs in SCRTT. Figure 26 represents a summary map of all known PTMs of

SCR TT identified in our laboratory and includes a previously reported phosphosite, S216 (Zhai *et al.*, 2008).



**Figure 26. A summary map of post-translational modifications in SCR TT 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.

In this section of the discussion, these potentially biologically relevant PTMs will be correlated with information known about SCR. This information is 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.

The conserved domains/motifs in SCR that are post-translationally modified are the protostome-specific LASCY, insect-specific DYTQL, arthropod-specific SCKY, *Drosophila*-specific NSQSL, Diptera-specific DISPK, Diptera-specific VNVPM, Bilateran-specific YPWM, arthropod-specific NANGE, Bilateran-specific HD and insect-specific CTD (Sivanantharajah & Percival-Smith, 2015).

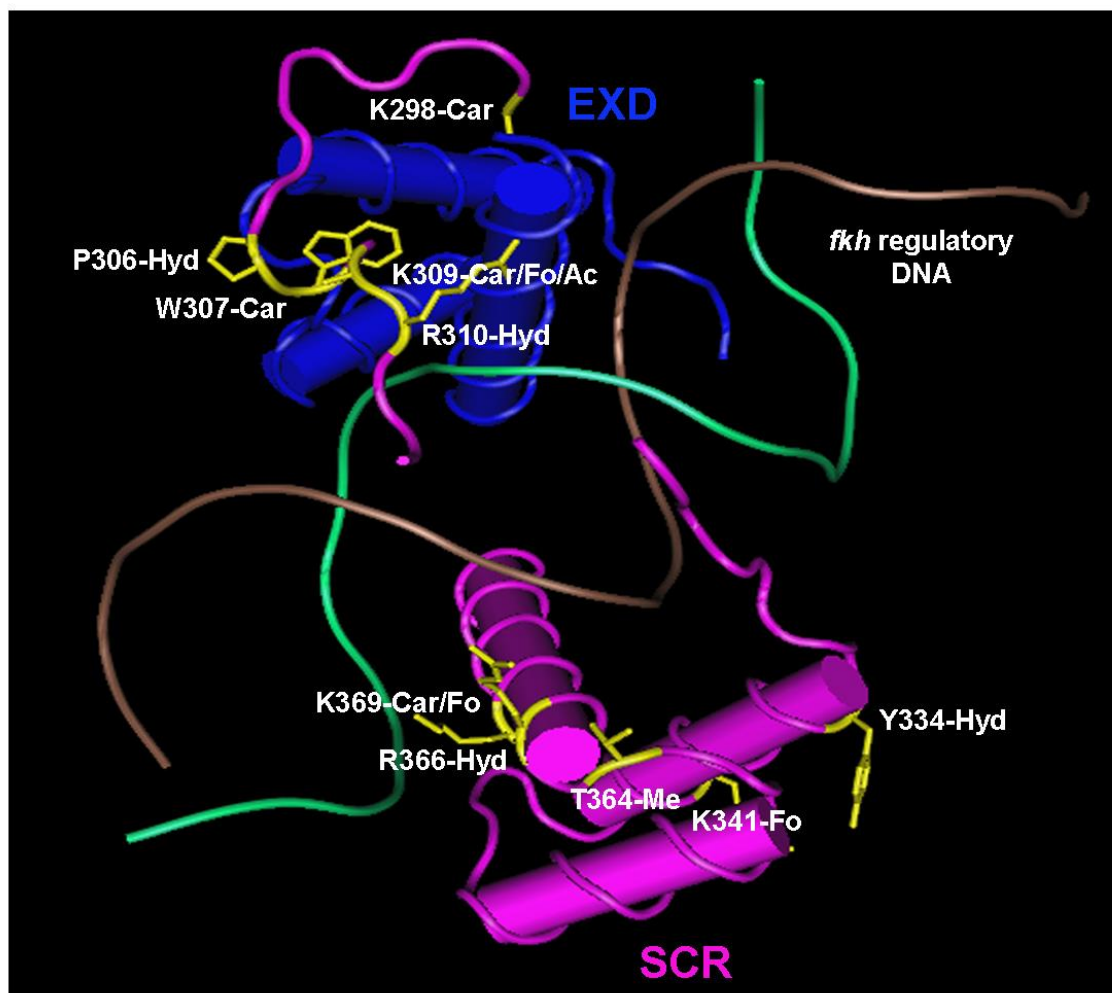
The conserved domains/motifs of SCR that are important for SCR activity are shown in Figure 4 (Sivanantharajah & Percival-Smith, 2009; Percival-Smith *et al.*, 2013; Sivanantharajah & Percival-Smith, 2014; Sivanantharajah, 2013). Of these functional domains/motifs of SCR, LASCY, DYTQL, YPWM, NANGE, HD and CTD are post-translationally modified. SCR has been proposed to have two functions, SCR<sup>T1</sup> and SCR<sup>lab</sup> and PB acts as the competence factor that switches SCR from determining T1 identity to labial identity (Percival-Smith *et al.*, 2013). The evolutionarily conserved motifs, LASCY, DYTQL, NANGE and the CTD of SCR are important for developmental competence by suppressing SCR<sup>lab</sup> activity in cells outside the competent primordial proboscis cells. These SCR motifs/domains mediate the suppression of ectopic proboscis formation, thereby negatively regulating SCR<sup>lab</sup> activity (Percival-Smith *et al.*, 2013) and therefore, PTMs in these motifs/domains might have a role in the regulation of SCR activity.

Many predicted conserved (the taxonomic level of conservation is indicated) and non-conserved SLiMs that are part of intrinsically disordered regions of SCR are post-translationally modified: methylation at S19 and hydroxylation at P22 of <sup>16</sup>SLASCYP (*Drosophila*-specific) and <sup>21</sup>YPQQ SLiMs (both map to protostome-specific LASCY motif), hydroxylation at Y87 of <sup>81</sup>AYTPNLY (*Drosophila*-specific), <sup>84</sup>PNLYP (*Drosophila*-specific), <sup>85</sup>NLYPNTP (*Drosophila*-specific), <sup>87</sup>YPNT, <sup>87</sup>YPNTPQ (*Drosophila*-specific) and <sup>87</sup>YPNTPQA (*Drosophila*-specific) SLiMs (all map to Diptera-

specific YTPNL motif), hydroxylation at D111 of <sup>110</sup>VDYTQLQ (Diptera-specific) and <sup>111</sup>DYTQL (insect-specific) SLiMs (both map to insect-specific DYTQL motif), methylations at S166 and K168 of <sup>163</sup>ANISCK SLiM (map to arthropod-specific SCKY motif), phosphorylation at S201 of the <sup>198</sup>NNNSQSL SLiM (map to *Drosophila*-specific NSQSL motif), phosphorylation at S216 (Zhai *et al.*, 2008), acetylation or formylation at K218, acetylation at S223 and S227 of <sup>209</sup>DLSTRDIS (*Drosophila*-specific), <sup>210</sup>LSTRDIS (*Drosophila*-specific), <sup>213</sup>RDISPK (*Drosophila*-specific), <sup>213</sup>RDISPKL (*Drosophila*-specific), <sup>213</sup>RDISPKLS (*Drosophila*-specific), <sup>217</sup>PKLSP (*Drosophila*-specific), <sup>217</sup>PKLSPS (*Drosophila*-specific), <sup>217</sup>PKLSPSS (*Drosophila*-specific), <sup>220</sup>SPSSVVE (*Drosophila*-specific) and <sup>224</sup>VVESVARS (*Drosophila*-specific) SLiMs (all map to Diptera-specific DISPK motif), hydroxylation at P269 of <sup>263</sup>PGNVNVP, <sup>269</sup>PMHSPG (*Drosophila*-specific) and <sup>269</sup>PMHSPGG (*Drosophila*-specific) SLiMs (all map to Diptera-specific VNVPM motif), hydroxylation at P306 and carboxylation at W307 of <sup>305</sup>YPWM SLiM (Bilateran-specific motif which becomes ordered/structured when bound to EXD), carboxylation or formylation or acetylation at K309, hydroxylation at R310, phosphorylations at T315, S316 and T317, hydroxylation at N321, carboxylation at E323, phosphorylation at T324, formylation or carboxylation or hydroxylation at K325 of <sup>313</sup>LGTSTV, <sup>313</sup>LGTSTVN and <sup>309</sup>KRVHLGTSTVNANGETKRQRT SLiMs (all map to arthropod-specific NANGE motif) (Sivanantharajah & Percival-Smith, 2015; Figure 26; Tables S9 & S10). Of the 33 identified post-translationally modified residues of SCR, 22 residues mapped to SLiMs (Figure 26), which agrees with the hypothesis that SLiMs are sites of PTMs.

The crystal structure of SCR-EXD bound to *fkh* regulatory DNA has been reported, and two regions of SCR are structured: residues 298 to 313 which contains the YPWM motif and 326 to 384 which contains the HD (Joshi *et al.*, 2007; Figure 27). Both structured regions are post-translationally modified.

First, I will discuss phosphorylation followed by other PTMs identified.



**Figure 27. 3-D crystal structure of SCR-EXD-DNA complex.** SCR is shown in pink and EXD is shown in blue. The two strands of *fkh* regulatory DNA is shown in brown and green. The modified amino acids of SCR along with their side chains are shown in yellow. Ac – acetylation, Car – carboxylation, Fo – formylation, Hyd – hydroxylation and Me – methylation. The structure coordinates with accession code 2R5Z (*fkh250*) were retrieved from RCSB Protein Data Bank (Joshi *et al.*, 2007). Cn3D 4.3.1 (NCBI) was used to annotate the 3-D structure.

#### 4.1.1 Phosphorylation of SCRTT

The most clustered set of phosphorylations identified were on the four amino acid residues, T315, S316, T317 and T324 which flank the NANGE motif (Figure 26). The phosphates may have a role in regulating the transformation of SCR from having T1 activity to having labial activity (Percival-Smith *et al.*, 2013). These phosphosites map to disordered regions of SCR and hence, are not observed in the SCR-EXD complex binding to *fkh* DNA (Figure 27).

Some of the 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). Although none of the amino acid residues of SCR directly involved in minor groove interaction were found to be modified, surrounding residues, T315, S316 and T317 that are part of the linker and are predicted targets of NEK2 kinase (Table 12); and T324 which 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. This may be a mechanism of regulation of the *in vivo* DNA-binding specificity of SCR.

The first ever reported SCR PTM is the phosphorylation of S216 (Zhai *et al.*, 2008) which maps to the *Drosophila*-specific DISPK SLiM. DISPK SLiM is a predicted target site of glycogen synthase kinase 3 (GSK3), cyclin-dependent kinase (CDK) and Proline-directed (eg., MAPK) phosphorylation (Table 12). The other two SCR phosphosites identified were S185 and S201 of which S201 mapped to the *Drosophila*-specific NSQSL motif and the predicted SLiM, NNNSQSL. S201 is a predicted target of phosphatidylinositol 3-kinase-related kinases (PIKK) (Table 12).

The 6<sup>th</sup> and 7<sup>th</sup> amino acid residues of the HD, T329 and S330 which were shown to be phosphorylated by cAMP-dependent protein kinase A (Berry & Gehring, 2000) were not identified as being phosphorylated in this study. The phosphorylation of these residues was proposed to inhibit SCR activity by inhibiting DNA binding and were proposed targets for the phosphatase PP2A,B'. However, these residues were ruled out as targets of

the phosphatase PP2A,B' (Moazzen *et al.*, 2009) and their phosphorylation may be an *in vitro* artifact.

#### 4.1.2 PTMs 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 27).

The Bilateran-specific YPWM motif of SCR which is also a predicted SLiM, binds to a hydrophobic pocket on the surface of EXD HD as seen in the case of UBX (Passner *et al.*, 1999). MS/MS analysis of SCR<sup>TT</sup> identified hydroxylation at P306 and carboxylation at W307 which renders the 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 its target, *fkh* gene which is required for the initiation of 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.*, 2009). 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 mapped to the HD of SCR are on the solvent



exposed surface. Although none of the amino acid residues of SCR that directly interact with the major groove of DNA were found to be modified, PTMs were detected in surrounding residues: hydroxylation at Y334, formylation at K341, methylation at T364, hydroxylation at R366 and formylation or carboxylation at K369. As the HD interacts with the negatively charged DNA, it is rich in basic amino acids, lysine, arginine and histidine. Hydroxylation of R366 and carboxylation of K369 gives a negative charge to the residues which might interfere with DNA binding. Since HD binds DNA, it is not unexpected to detect formylations at K341 and K369 of the HD of SCR as formylations often arise as a secondary modification due to oxidative DNA damage (Jiang *et al.*, 2007).

#### 4.1.3 PTMs in potentially disordered regions of SCR

This study identified many sites of SCR that were post-translationally modified and the majority of the sites (22 out of 33) mapped to potentially disordered regions of SCR (Figure 26; Tables S9 & S10). The post-translationally modified, disordered motifs, LASCY, DYTQL and SCKY will be discussed in detail.

The Protosome-specific LASCY functional motif of SCR which is part of a *Drosophila*-specific SLiM, SLASCYP is post-translationally modified. The serine residue of the LASCY motif, S19 is methylated and P22 which is the last residue of the SLASCYP SLiM is hydroxylated. The antimorphic allele, *Scr<sup>l4</sup>* with a Ser<sub>10</sub> to Leu change maps to the Bilateran-specific octapeptide motif and this results in the acquisition of a leucine zipper motif in the octapeptide and adjacent LASCY motifs of SCR (Sivanantharajah, 2013; Sivanantharajah & Percival-Smith, 2014). *SCR<sup>l4</sup>* inhibits the activity of endogenous *SCR<sup>+</sup>* in the developing embryo. Both the octapeptide and LASCY motifs together function as a transcriptional activation domain (Sivanantharajah, 2013; Sivanantharajah & Percival-Smith, 2014) which is probably thought to interact with proteins of the polycomb (Pc-G) and the trithorax group (Trx-G) which are known to regulate HOX activity by chromatin remodeling (Stankunas *et al.*, 1998; Crosby *et al.*, 1999; Kal *et al.*, 2000). Methylation of proteins, especially histone lysine and arginine methylation, are known to have an effect on chromatin remodeling and subsequent gene

expression (reviewed in Ng *et al.*, 2009). I identified a methylated serine (O-methylation) in the LASCY motif which might have a role in modulating the interaction between SCR<sup>14</sup> and Pc-G and Trx-G genes. Although N-methylations are more common, O-methylations also occur (Green & Garneau-Tsodikova, 2010). One might argue that S19 methylation is an artifact. A study in 2010 misassigned methylation on S28 of histone H3 from calf thymus. After careful investigation of the MS/MS and MS/MS/MS data, they found that K27 instead of S28 of histone H3 was methylated (Zhang *et al.*, 2010). In this study, the methylated peptide identified by MS/MS, LAS(+14.02)C(+57.02)YPQQ does not contain lysine or arginine which are commonly methylated and hence, S19 methylation is probably not an artifact. Hydroxylation at P22 makes the residue polar which may have a role in regulating the activity of the neighboring LASCY motif.

The insect-specific DYTQL motif and the CTD of SCR mediates a genetic interaction with *pb* and this is required for proboscis determination (Sivanantharajah & Percival-Smith, 2009; Sivanantharajah, 2013). D111 which is in the DYTQL motif and P392 and Y398 (which map to non-SLiMs) which are in the CTD of SCR are hydroxylated (Figure 26). The CTD of SCR is proposed to mask the octapeptide, and therefore, the octapeptide is no longer available to form a protein complex resulting in a loss of sex combs and pseudotrachea (Sivanantharajah & Percival-Smith, 2009). It is well known that hydroxyprolines are involved in the stability of collagen fibres (Berg & Prockop, 1973). Although addition of a polar hydroxyl group to polar amino acid residues like D and Y might not make a difference in polarity, hydroxylation of a non-polar residue like proline (P392) in the CTD of SCR is capable of altering its properties which might affect the 3-D conformation of this region of the protein, thereby inhibiting SCR<sup>T1</sup> activity. This might be a mechanism of regulation of the genetic interaction of the CTD of SCR with PB in determining proboscis, i.e., a switch to turn on SCR<sup>lab</sup> activity.

The arthropod-specific SCKY motif of SCR which is part of the SLiM, ANISCK is methylated at its serine and lysine residues. Although the function of the SCKY motif is not known, it might play a role in the suppression of SCR activity (Percival-Smith *et al.*, 2013). Assuming SCKY motif has a function, two methylations in the same motif may be

capable of perturbing the activity of the motif.

#### 4.1.4 PTMs in regions of SCR which are not a functional or a conserved domain/motif or a SLiM

A few PTMs were identified in SCR that did not map to a functional or a conserved domain/motif or a SLiM: hydroxylation at P107, carboxylation or hydroxylation at D108, phosphorylation at S185 and carboxylation at K298. The residues P107 and D108 of SCR, which maps to non-SLiMs, precede the insect-specific motif and SLiM, DYTQL. Hydroxylation of P107 and hydroxylation or carboxylation of D108 adds negative charges to this region of SCR which may have a role along with DYTQL motif in proboscis determination (Sivanantharajah & Percival-Smith, 2009; Sivanantharajah, 2013). The carboxylated residue, K298 precedes the structured, extended linker region of SCR (25 residues before the HD spanning from residues 299 to 323) and is observed in the SCR-EXD-DNA crystal structure (Joshi *et al.*, 2007; Figure 27). The carboxyl group adds negative charge to the K298 residue which might have an indirect role in regulating the DNA binding capacity of the linker region of SCR. Of the 7 known phosphosites of SCR (including the ones mapped by Dr. Anthony Percival-Smith and Zhai *et al.*, 2008; Figure 25), S185 is the only phosphosite which does not map to SLiMs. However, SLiMs neighboring S185 phosphosite are predicted kinase targets (Table S9). It can be concluded that S185 of SCR is a potential non-SLiM target of a kinase.

#### 4.1.5 Competition of acetylation and formylation observed in SCRTT

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 a previous report where lysine residues of core histone proteins were N<sup>ε</sup>-formylated and these lysine residues were also sites of frequent acetylation (Wiśniewski *et al.*, 2008).

Acetylation of histones by histone acetyltransferases is 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 (reviewed in Kouzarides, 2000; Bannister & Miska, 2000). An example of a non-histone protein that is acetylated is the tumor suppressor protein, p53 (Gu & Roeder, 1997).

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 with 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.

#### 4.1.6 Reversible PTMs identified in SCRTT

Cell signaling is highly dependent on the reversible nature of PTMs of proteins (Theillet *et al.*, 2012; Walsh *et al.*, 2005). Of all the types of PTMs mapped in embryonic SCRTT, phosphorylation, acetylation and methylation are the reversible PTMs which have been studied extensively. Phosphorylation is a key reversible PTM in eukaryotes (Li *et al.*, 2013; Sacco *et al.*, 2012) where kinases add phosphate groups to S, T and Y residues of proteins and phosphatases remove phosphates, thereby regulating protein activity. Lysine acetyltransferases catalyze the transfer of an acetyl group onto a lysine residue of a protein whereas lysine deacetylases catalyze the removal of the acetyl group and this is known to regulate chromatin remodeling and subsequent transcription of genes (Allfrey *et al.*, 1964; reviewed in Roth *et al.*, 2001; Grozinger & Schreiber, 2002). Methylation is also a reversible PTM (Shi *et al.*, 2004) where methylation and demethylation are catalyzed by methyltransferases and demethylases, respectively. Histone methylation and demethylation is known to affect chromatin remodeling and gene expression (Bannister & Kouzarides, 2011). The reversible nature of these PTMs might have a role in regulating SCR activity.

## 4.2 Affinity purification of SCRTT

SCRTT extracted from developing *D. melanogaster* embryos were affinity purified using Ni-NTA sepharose beads, which bind the 6X His tag. The SCRTT signal in various fractions collected during nuclear fractionation and affinity purification were estimated from the signal on a Western Blot (Figure 16D). During nuclear fractionation, the SCRTT signal in the lysate and pellet 1 fraction is lower than that of pellet 2 and pellet 3 fractions (Figure 16D). This anomaly is also reflected in the purification table as a 50-fold increase of yield during nuclear fractionation (Table 6). The likely explanation of this anomaly is that the lysate contains a large amount of total protein that competes with SCRTT for binding to the membrane, and as the total protein decreases from lysate to pellet 1 to pellet 2 and onwards, there is less competition and more SCRTT is bound to the membrane.

HOXTT proteins could not be efficiently eluted from the Ni-NTA sepharose beads. Multiple attempts were made to elute HOXTT proteins from the Ni-NTA beads under both aqueous and denaturing conditions using different final concentrations of imidazole upto 1M. Hence, the final fold purification of SCRTT could not be estimated (Table 6). Although His-tagged proteins and peptides expressed in bacteria can be readily eluted from Ni beads, proteins extracted from *Drosophila* seem to stick to the beads. The likely reasons why embryonic HOXTT proteins could not be efficiently eluted from the Ni-NTA beads are that the Ni beads were not accessible to the imidazole or the Ni coordination with the His tag was very strong and therefore, imidazole could not displace the His tag from Ni (Bauman Jr. & Wang, 1964).

## 4.3 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  $\alpha$ -casein which is a known heavily phosphorylated protein (Larsen *et al.*, 2005) (Figure S2). Of the 61 identified phosphorylated peptides that cover the region 61-70 of  $\alpha$ -casein, 24 peptides were phosphorylated at two amino acid residues (Figure 14). 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,  $\alpha$ -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<sub>2</sub> beads to enrich trypsinized  $\alpha$ -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  $\alpha$ -casein post-TiO<sub>2</sub> treatment (84 were phosphopeptides out of 91 peptides detected), only 54 phosphopeptides were detected post-TiO<sub>2</sub> out of 158 phosphopeptides initially detected without TiO<sub>2</sub> enrichment which indicates loss of phosphopeptides. No phosphopeptides were detected upon TiO<sub>2</sub> enrichment of trypsinized SCRTT. Even the MS/MS analysis of the TiO<sub>2</sub> 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.

#### 4.4 SLiM analysis and their significance

The second objective of this study was to test the hypothesis that SLiMs of FTZ and HOX proteins are preferential sites of phosphorylation (Sivanantharajah & Percival-Smith, 2015). It is essential to know whether the phosphorylatable amino acid residues, S, T and Y are concentrated to SLiMs than non-SLiMs as this would explain why SLiMs are predicted sites of phosphorylation. I found that the location of S, T and Y were biased towards SLiMs (Table 11). Out of 17 sites of HOX phosphorylation mapped, 16 were in SLiMs and 8 were in conserved SLiMs (Table 12). Although apparently it seems that almost all phosphosites map to SLiMs, the 16:1 or 8:9 differences were not statistically significant (Table 13). The 16:1 difference can be explained by the concentration of S, T and Y in SLiMs vs. non-SLiMs.

I found that only 10% SLiMs had phosphates in SCRTT. This may indicate that a minority of SLiMs are phosphosites. However, there are many predicted SLiMs of FTZ and HOX proteins which are putative targets of various kinases (Tables S1-19) and hence, more SLiMs are likely to be phosphorylated. In addition, phosphorylations may

not have been detected due to their substoichiometric nature which explains the low percentage of SLiMs that have been found to be phosphorylated in this study.

SLiMs are sequence targets of other PTMs (Punternvoll *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. For example, I identified a methylation at S19 and a hydroxylation at P22 of the *Drosophila*-specific SLiM, SLASCYP. Another example would be the ANISCK SLiM which has its serine and lysine methylated.

SLiMs are sites of protein-protein interaction (Davey *et al.*, 2012). I have identified SLiMs in regions of SCR which are known to interact with other proteins in the developing embryo, for example, the octapeptide and YPWM. I identified SLiMs, FAMSSY, AMSSY, SSYQF, SSYQFV, SYQFV, YQFV, FVNLSLA and SLASCYP (Tables S9 & S10) in the octapeptide of SCR which is involved in protein-protein interaction important for sex comb and pseudotrachea formation (Sivanantharajah & Percival-Smith, 2009). The SCR motif, YPWM is a predicted SLiM which interacts with EXD and the SCR-EXD complex binds *fkh* DNA, thereby initiating salivary gland development (Ryoo & Mann, 1999; Joshi *et al.*, 2007).

## 4.5 Future directions

A survey map of PTMs in SCRTT protein extracted from developing *D. melanogaster* embryos is reported in this thesis. Dr. Anthony Percival-Smith also identified phosphorylation in two other tagged HOX proteins, LABTT and PBTT (Percival-Smith, unpublished). A future goal would be completion of the analysis of FTZTT and the other HOXTT proteins for PTMs.

This study identified PTMs on overexpressed SCRTT protein. One of the potential limitations of detecting PTMs on overexpressed proteins is whether the same PTMs would be detected on endogenous proteins. One approach to avoid this limitation is to immunoprecipitate endogenous HOX proteins using anti-HOX antibodies followed by



PTM detection by MS/MS analysis. Moreover, from a pool of endogenous proteins, phosphorylated proteins can be immunoprecipitated using phosphospecific antibodies which bind to phosphoserine (pS), phosphothreonine (pT) and phosphotyrosine (pY) residues following which protein identification and PTM detection by MS/MS analysis of the immunoprecipitated proteins can be done.

Phosphorylation is proposed to be a major reversible regulatory mechanism of protein function (Li *et al.*, 2013; Sacco *et al.*, 2012; Mylin *et al.*, 1989; Hunter & Karin, 1992; Ardito *et al.*, 2017). A future goal would be a detailed analysis of the function of phosphosites identified in SCR<sup>TT</sup>, LAB<sup>TT</sup>, PB<sup>TT</sup> and UB<sup>X</sup>TT (Table 12). Genetic modification of phosphosites that map to structurally and functionally important regions of HOX proteins is one approach. For instance, the SCR phosphorylation sites, T315, S316, T317 and T324 that map to the linker and the N-terminal of the HD might play a role in the interaction with the minor groove of *fkh* DNA (Joshi *et al.*, 2007). Using PCR-based site-directed mutagenesis, the identified S and T residues that are phosphorylated could be changed to an alanine residue which would abolish phosphorylation. S and T residues could also be changed to an acidic aspartic acid residue which mimics phosphorylation due to the presence of the negative charge on the acidic amino acids. Following the genetic modification, the mutant and the control constructs will be microinjected into the fly embryo and stable germ-line transformants will be obtained. Detailed analysis of the larval and adult phenotypes upon ectopic expression of the mutant construct will reveal if the phosphorylation site has a function *in vivo*.

Another interesting future goal would be to determine whether hydroxylation of P306 and carboxylation of W307 in the Bilateran-specific YPWM motif of SCR interferes with EXD binding. These two modifications render the YPWM motif hydrophilic which might not be able to bind the hydrophobic pocket of EXD and this might inhibit the expression of the target gene, *fkh*, thereby, inhibiting the initiation of salivary gland development (Ryoo & Mann, 1999). This would require the isolation of modified proteins for analysis in a DNA binding assay.

A potential future experiment would be to study the genetic interaction of *Hox* genes, *Scr* and *pb* in flies expressing a mutant form of SCR where the phosphorylatable serine and threonine residues flanking the NANGE motif are mutated to alanine. This would suggest if phosphorylation at a site flanking the NANGE motif is required for turning on the activity of the NANGE motif, thereby mediating a genetic interaction with *pb*. Four sites of phosphorylation flank the NANGE motif of SCR (Figure 26) and one phosphorylatable residue at a time can be mutated to investigate the role of phosphorylation at that particular site. The DYTQL motif and CTD of SCR mediates a genetic interaction with *pb* (Sivanantharajah & Percival-Smith, 2009; Sivanantharajah, 2013). It is important to find out if the PTMs of these two motifs/domains play a role in the genetic interaction with *pb*. Similar experiments can be designed to decipher the role of PTMs in HOX activity.

In summary, this study identified sites of phosphorylation and other PTMs in a tagged HOX protein, SCR<sup>TT</sup> 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 work involved the optimization of the protocol for SCR<sup>TT</sup> purification from developing embryos and this can be extended to the purification of FTZ<sup>TT</sup> and other HOX<sup>TT</sup> proteins.

## References

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., ... Venter, J. C. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185-2195.
- Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198-207.
- Akam, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development*, 101(1), 1-22.
- Akam, M. (1989). *Hox* and HOM: homologous gene clusters in insects and vertebrates. *Cell*, 57(3), 347-349.
- 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.
- Andrew, D. J. (1998). Regulation and Formation of the *Drosophila* Salivary Glands. *Annals of the New York Academy of Sciences*, 842(1), 55-69.
- 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.

- Ashburner, M. (1989). *Drosophila: A Laboratory Handbook and Manual*. (Cold Spring Harbor Laboratory Press, New York)
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381-395.
- 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.
- Barford, D., & Johnson, L. N. (1989). The allosteric transition of glycogen phosphorylase. *Nature*, *340*(6235), 609-616.
- Barford, D., Hu, S. H., & Johnson, L. N. (1991). Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. *Journal of Molecular Biology*, *218*(1), 233-260.
- Barnett, A. A., & Thomas, R. H. (2013). Posterior Hox gene reduction in an arthropod: *Ultrabithorax* and *Abdominal-B* are expressed in a single segment in the mite *Archezogozetes longisetosus*. *EvoDevo*, *4*, 23.
- Bateson, W. (1894). *Materials for the study of variation: treated with special regard to discontinuity in the origin of species*. (Macmillan and Co. Ithaca, New York)
- Bath, D. (2010). The role of post-translational modification in regulating the activities of Fushi Tarazu, a pair-rule segmentation protein in *Drosophila melanogaster*. M.Sc. Thesis. The University of Western Ontario, London, Ontario, Canada.
- Bauman Jr., J. E., & Wang, J. C. (1964). Imidazole complexes of nickel (II), copper (II), zinc (II), and silver (I). *Inorganic Chemistry*, *3*(3), 368-373.
- 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.

- Beeman, R. W. (1987). A homeotic gene cluster in the red flour beetle. *Nature*, 327(6119), 247-249.
- Berg, R. A., & Prockop, D. J. (1973). The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochemical and Biophysical Research Communications*, 52(1), 115-120.
- 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.
- Bibikova, M., Golic, M., Golic, K. G., & Carroll, D. (2002). Targeted Chromosomal Cleavage and Mutagenesis in *Drosophila* Using Zinc-Finger Nucleases. *Genetics*, 161(3), 1169–1175.
- Biggs, W. H. 3<sup>rd</sup>, & Zipursky, S. L. (1992). Primary structure, expression and signal-dependent tyrosine phosphorylation of a *Drosophila* homolog of extracellular signal-regulated kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 89(14), 6295-6299.
- Biggs, W. H. 3<sup>rd</sup>, Zavitz, K. H., Dickson, B., van der Straten, A., Brunner, D., Hafen, E., & Zipursky, S. L. (1994). The *Drosophila* rolled locus encodes a MAP kinase required in the Sevenless signal transduction pathway. *The EMBO Journal*, 13(7), 1628-1635.
- 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 <sup>1</sup>H nuclear magnetic resonance spectroscopy. *Journal of Molecular Biology*, 214(1),183-197.
- Boja, E. S., & Fales, H. M. (2001). Overalkylation of a protein digest with iodoacetamide. *Analytical chemistry*, 73(15), 3576-3582.
- 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.
- Bram, R. J., & Kornberg, R. D. (1985). Specific protein binding to far upstream activating sequences in polymerase II promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 82(1), 43-47.
- 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.
- Brent, R., & Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell*, 43(3 Pt 2), 729-736.
- Brooke, N. M., Garcia-Fernàndez, J., & Holland, P. W. H. (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature*, 392(6679), 920-922.
- Brunner, D., Dücker, K., Oellers, N., Hafen, E., Scholz, H., & Klämbt, C. (1994). The ETS domain protein pointed-P2 is a target of MAP kinase in the sevenless signal transduction pathway. *Nature*, 370(6488), 386-389.
- Campos-Ortega, J., & Hartenstein, V. (1985). The embryonic development of *Drosophila melanogaster*. Springer-Verlag, New York, NY.
- Capovilla, M., Brandt, M., & Botas, J. (1994). Direct Regulation of *decapentaplegic* by *Ultrabithorax* and Its Role in *Drosophila* Midgut Morphogenesis. *Cell*, 76(3), 461–475.
- Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2005). From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design. Second Edition. (Oxford: Blackwell Publishing)

- Carroll, S. B., & Vavra, S. H. (1989). The zygotic control of *Drosophila* pair-rule gene expression. II. Spatial repression by gap and pair-rule gene products. *Development*, *107*(3), 673-683.
- Casanova, J., Sánchez-Herrero, E., & Morata, G. (1986). Identification and characterization of a parasegment specific regulatory element of the abdominal-B gene of *Drosophila*. *Cell*, *47*(4), 627-636.
- Chen, G., Liu, H., Wang, X., & Li, Z. (2010). *In vitro* methylation by methanol: proteomic screening and prevalence investigation. *Analytica Chimica Acta*, *661*(1), 67-75.
- Choudhary, C., Kumar, C., Gnad, F., Nielson, M. L., Rehman, M., Walther, T. C., ... Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, *325*(5942), 834-840.
- Clauser, K. R., Hall, S. C., Smith, D. M., Webb, J. W., Andrews, L. E., Tran, H. M., ... Burlingame, A. L. (1995). Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(11), 5072-5076.
- Cohen, P. (2000). The regulation of protein function by multisite phosphorylation – a 25 year update. *Trends in Biochemical Sciences*, *25*(12), 596–601.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819-823.
- Crosby, M. A., Miller, C., Alon, T., Watson, K. L., Verrijzer, C. P., Goldman-Levi, R., & Zak, N. B. (1999). The *trithorax* group gene *moira* encodes a brahma-associated putative chromatin-remodeling factor in *Drosophila melanogaster*. *Molecular and Cellular Biology*, *19*(2), 1159-1170.

- 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.
- D'Avino, P. P., & Thummel, C. S. (1999). Ectopic expression systems in *Drosophila*. *Methods in Enzymology*, 306, 129-142.
- De Oliveira, A. L., Wollesen, T., Kristof, A., Scherholz, M., Redl, E., Todt, C., ... Wanninger, A. (2016). Comparative transcriptomics enlarges the toolkit of known developmental genes in mollusks. *BMC Genomics*, 17(1), 905.
- Devenport, M. P., Blass, C., & Eggleston, P. (2000). Characterization of the Hox gene cluster in the malaria vector mosquito, *Anopheles gambiae*. *Evolution & Development*, 2(6), 326-339.
- 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.
- Doe, C. Q., Hiromi, Y., Gehring, W. J., & Goodman, C. S. (1988). Expression and function of the segmentation gene *fushi tarazu* during *Drosophila* neurogenesis. *Science*, 239(4836), 170-175.
- 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.
- Dong, J., & Krause, H. M. (1999). Tissue-specific requirements for a phosphorylation site in the Fushi tarazu homeodomain. *Development Genes and Evolution*, 209(7), 427-431.
- Driever, W., & Nüsslein-Volhard, C. (1988). The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell*, 54(1), 95-104.



- 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.
- Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development*, 134(14), 2549–2560.
- Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3), 197–208.
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1990). Electrospray ionization - principles and practice. *Mass Spectrometry Reviews*, 9(1), 37–70.
- Ferrier, D. E. K., & Akam, M. (1996). Organization of the Hox gene cluster in the grasshopper, *Schistocerca gregaria*. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23), 13024–13029.
- Ferrier, D. E. K., & Holland, P. W. H. (2001). Ancient origin of the Hox gene cluster. *Nature Reviews Genetics*. 2(1), 33–38.
- Finnerty, J. R., & Martindale, M. Q. (1999). Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. *Evolution & Development*, 1(1), 16–23.
- 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.
- Frasch, M., & Levine, M. (1987). Complementary patterns of *even-skipped* and *fushi tarazu* expression involve their differential regulation by a common set of segmentation genes in *Drosophila*. *Genes & Development*, 1(9), 981–985.
- Freeman, R., Ikuta, T., Wu, M., Koyanagi, R., Kawashima, T., Tagawa, K., ... Gerhart, J. (2012). Identical genomic organization of two hemichordate Hox clusters. *Current Biology*, 22(21), 2053–2058.

- Fröblius, A. C., Matus, D. Q., & Seaver, E. C. (2008). Genomic organization and expression demonstrate spatial and temporal *Hox* gene colinearity in the lophotrochozoan *Capitella* sp. I. *PLoS ONE*, 3(12), e4004.
- Gaunt, S. J. (2015). The significance of *Hox* gene collinearity. *The International Journal of Developmental Biology*, 59(4-6), 159-170.
- 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., Kloter, U., & Suga, H. (2009). Evolution of the Hox gene complex from an evolutionary ground state. *Current Topics in Developmental Biology*, 88, 35-61.
- 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. & Gehring, W. J. (1988). Head and thoracic transformations caused by ectopic expression of *Antennapedia* during *Drosophila* development. *Development*, 102, 657-675.
- 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.
- Giniger, E., Varnum, S. M., & Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, 40(4), 767-774.

- González-Reyes, A., & Morata, G. (1991). Organization of the *Drosophila* head as revealed by the ectopic expression of the Ultrabithorax product. *Development*, *113*(4), 1459-1471.
- 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.
- Gramates, L. S., Marygold, S. J., Santos, G. D., Urbano, J. M., Antonazzo, G., Matthews, B. B., ... the Flybase Consortium. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Research*, *45*(D1), D663-D671.
- Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., ... O'Connor-Giles, K. M. (2013). Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics*, *194*(4), 1029–1035.
- Grenier, J. K., & Carroll, S. B. (2000). Functional evolution of the Ultrabithorax protein. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(2), 704-709.
- Green, K. D., & Garneau-Tsodikova, S. (2010). Posttranslational modification of proteins. *Comprehensive Natural Products II*, *5*, 433-468.
- Greenspan, R. J. (2004). Fly pushing: the theory and practice of *Drosophila* genetics (2<sup>nd</sup> edition). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
- Groban, E. S., Narayanan, A., & Jacobson, M. P. (2006). Conformational changes in protein loops and helices induced by post-translational phosphorylation. *PLoS Computational Biology*, *2*(4), e32.
- Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic *Drosophila* by using the site-specific integrase from phage  $\Phi$ C31. *Genetics*, *166*(4), 1775-1782.

- Grozinger, C. M., & Schreiber, S. L. (2002). Deacetylase Enzymes: Biological Functions and the Use of Small-Molecule Inhibitors. *Chemistry & Biology*, 9(1), 3-16.
- 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.
- Guerrera, I. C., & Kleiner, O. (2005). Application of mass spectrometry in proteomics. *Bioscience Reports*, 25(1-2), 71-93.
- Haebel, S., Albrecht, T., Spärbier, K., Walden, P., Körner, R., & Steup, M. (1998). Electrophoresis-related protein modification: alkylation of carboxy residues revealed by mass spectrometry. *Electrophoresis*, 19(5), 679-686.
- Hafen, E., Kuroiwa, A., & Gehring, W. J. (1984). Spatial Distribution of Transcripts from the Segmentation Gene *fushi tarazu* during *Drosophila* Embryonic Development. *Cell*, 37(3), 833-841.
- Han, X., Aslanian, A., Yates, J. R. 3<sup>rd</sup> (2008). Mass spectrometry for proteomics. *Current Opinion in Chemical Biology*, 12(5), 483-490.
- Harding, K., Wedeen, C., McGinnis, W., & Levine, M. (1985). Spatially regulated expression of homeotic genes in *Drosophila*. *Science*, 229(4719), 1236-1242.
- Hiromi, Y., & Gehring, W. J. (1987). Regulation and Function of the *Drosophila* Segmentation Gene *fushi tarazu*. *Cell*, 50(6), 963-974.
- 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.
- Hombria, J. C., & Lovegrove, B. (2003). Beyond homeosis--HOX function in morphogenesis and organogenesis. *Differentiation*, *71*(8), 461-476.
- Horan, G. S. B., Wu, K., Wolgemuth, D. J., & Behringer, R. R. (1994). Homeotic transformation of cervical vertebrae in *Hoxa-4* mutant mice. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(26), 12644–12648.
- Hueber, S. D., Rauch, J., Djordjevic, M. A., Gunter, H., Weiller, G. F., & Frickey, T. (2013). Analysis of central Hox protein types across bilateran clades: On the diversification of central Hox proteins from an Antennapedia/Hox7-like protein. *Developmental Biology*, *383*(2), 175-185.
- Hughes, C. L., & Kaufman, T. C. (2002). Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. *Development*, *129*(5), 1225-1238.
- 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.
- Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., Goswami, T., Rad, R., Beausoleil, S. A., ... Gygi, S. P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*, *143*(7), 1174-1189.
- 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.

- Ikuta, T., & Saiga, H. (2005). Organization of Hox genes in ascidians: present, past, and future. *Developmental Dynamics*, 233(2), 382-389.
- Irvine, S. Q., & Martindale, M. Q. (2000). Expression patterns of anterior Hox genes in the polychaete *Chaetopterus*: correlation with morphological boundaries. *Developmental Biology*, 217(2), 333-351.
- Ish-Horowicz, D., Pinchin, S. M., Ingham, P. W., & Gyurkovics, H. G. (1989). Autocatalytic *ftz* activation and metameric instability induced by ectopic *ftz* expression. *Cell*, 57(2), 223–232.
- Ishiwata, K., Sasaki, G., Ogawa, J., Miyata, T., & Su, Z. H. (2011). Phylogenetic relationships among insect orders based on three nuclear protein-coding gene sequences. *Molecular Phylogenetics and Evolution*, 58(2), 169-180.
- Jäckle, H., Hoch, M., Pankratz, M. J., Gerwin, N., Sauer, F., & Brönnner, G. (1992). Transcriptional control by *Drosophila* gap genes. *Journal of Cell Science, Supplement 16*, 39-51.
- 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.
- Janssen, R., Eriksson, B. J., Tait, N. N., & Budd, G. E. (2014). Onychophoran Hox genes and the evolution of arthropod Hox gene expression. *Frontiers in Zoology*, 11(1), 22.
- Jeannotte, L., Lemieux, M., Charron, J., Poirier, F., & Robertson, E. J. (1993). Specification of axial identity in the mouse: Role of the *Hoxa-5* (*Hox1.3*) gene. *Genes & Development*, 7(11), 2085–2096.

- 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.
- 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.
- Jung, S. Y., Li, Y., Wang, Y., Chen, Y., Zhao, Y., & Qin, J. (2008). Complications in the assignment of 14 and 28 Da mass shift detected by mass spectrometry as in vivo methylation from endogenous proteins. *Analytical Chemistry*, *80*(5), 1721-1729.
- Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., & Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* - II. Zygotic loci on the third chromosome. *Wilhelm Roux's Archives of Developmental Biology*, *193*(5), 283–295.
- Kal, A. J., Mahmoudi, T., Zak, N. B., & Verrijzer, C. P. (2000). The *Drosophila* Brahma complex is an essential coactivator for the *trithorax* group protein Zeste. *Genes & Development*, *14*(9), 1058-1071.
- Kanca, O., Bellen, H. J., & Schnorrer, F. (2017). Gene tagging strategies to assess protein expression, localization, and function in *Drosophila*. *Genetics*, *207*(2), 389-412.
- Karr, T. L., & Kornberg, T. B. (1989). *fushi tarazu* protein expression in the cellular blastoderm of *Drosophila* detected using a novel imaging technique. *Development*, *106*(1), 95–103.
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, bbx108.

- Kaufman, T. C., Lewis, R., & Wakimoto, B. (1980). Cytogenetic analysis of chromosome 3 in *DROSOPHILA MELANOGASTER*: the homoeotic gene complex in polytene chromosome interval 84A-B. *Genetics*, *94*(1), 115-133.
- Kavi, H. H., Fernandez, H., Xie, W., & Birchler, J. A. (2008). Genetics and Biochemistry of RNAi in *Drosophila*. *Current Topics in Microbiology and Immunology*, *320*, 37-75.
- Keegan, L., Gill, G., & Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science*, *231*(4739), 699-704.
- Kelstrup, C. D., Frese, C., Heck, A. J. R., Olsen, J. V., & Nielson, M. L. (2014). Analytical utility of mass spectral binning in proteomic experiments by SPECTral Immonium Ion Detection (SPIID). *Molecular & Cellular Proteomics*, *13*(8), 1914-1924.
- 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, D. H., Lee, B. Y., Kim, H. S., Jeong, C. B., Hwang, D. S., Kim, I. C., & Lee, J. S. (2018). Identification and characterization of homeobox (*Hox*) genes and conservation of the single *Hox* cluster (324.6 kb) in the water flea *Daphnia magna*. *Journal of Experimental Zoology. Part B. Molecular and Developmental Evolution*, *330*(2), 76-82.
- Kim, J., Petritis, K., Shen, Y., Camp, D. G. 2<sup>nd</sup>, Moore, R. J., & Smith, R. D. (2007). Phosphopeptide elution times in reversed-phase liquid chromatography. *Journal of Chromatography A*, *1172*(1), 9-18.
- Kollipara, L., & Zahedi, R. P. (2013). Protein carbamylation: in vivo modification or in vitro artefact? *Proteomics*, *13*(6), 941-944.



- Kourakis, M. J., & Martindale, M. Q. (2001). Hox gene duplication and deployment in the annelid leech *Helobdella*. *Evolution & Development*, 3(3), 145-153.
- Kourakis, M. J., Master, V. A., Lokhorst, D. K., Nardelli-Haeffliger, D., Wedeen, C. J., Martindale, M. Q., & Shankland, M. (1997). Conserved anterior boundaries of Hox gene expression in the central nervous system of the leech *Helobdella*. *Developmental Biology*, 190(2), 284-300.
- 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.
- Kuziora, M. A. (1993). *Abdominal-B* protein isoforms exhibit distinct cuticular transformations and regulatory activities when ectopically expressed in *Drosophila* embryos. *Mechanisms of Development*, 42(3), 125-137.
- Kuziora, M. A., & McGinnis, W. (1988). Autoregulation of a *Drosophila* homeotic selector gene. *Cell*, 55(3), 477-485.
- Lamka, M. L., Boulet, A. M., & Sakonju, S. (1992). Ectopic expression of UBX and ABD-B proteins during *Drosophila* embryogenesis: competition, not a functional hierarchy, explains phenotypic suppression. *Development*, 116(4), 841-854.
- Lapko, V. N., Smith, D. L., & Smith, J. B. (2000). Identification of an artifact in the mass spectrometry of proteins derivatized with iodoacetamide. *Journal of Mass Spectrometry*, 35(4), 572-575.

- 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.
- Laughon, A., & Scott, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature*, 310(5972), 25–31.
- Lawrence, P. A., & Johnston, P. (1989). Pattern formation in the *Drosophila* embryo : allocation of cells to parasegments by *even-skipped* and *fushi tarazu*. *Development*, 105(4), 761–767.
- Lehmann, R., & Nüsslein-Volhard, C. (1991). The maternal gene *nanos* has a central role in posterior pattern formation of the *Drosophila* embryo. *Development*, 112(3), 679-691.
- Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature*, 276(5688), 565-570.
- Lewis, E. B. (1985). Regulation of the genes of the bithorax complex in *Drosophila*. *Cold Spring Harbor Symposia on Quantitative Biology*, 50, 155-164.
- 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., Murre, C., & McGinnis, W. (1999). Activity regulation of a Hox protein and a role for the homeodomain in inhibiting transcriptional activation. *The EMBO Journal*, 18(1), 198-211.
- Li, X., Wilmanns, M., Thornton, J., & Köhn, M. (2013). Elucidating human phosphatase-substrate networks. *Science Signaling*, 6(275), rs10.

- Lippincott, J., & Apostol, I. (1999). Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea. *Analytical Biochemistry*, *267*(1), 57-64.
- Löhr, U., & Pick, L. (2005). Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of *Drosophila melanogaster*. *Current Biology*, *15*(7), 643-649.
- Löhr, U., Yussa, M., & Pick, L. (2001). *Drosophila fushi tarazu*: a gene on the border of homeotic function. *Current Biology*, *11*(18), 1403-1412.
- Lovegrove, B., Simões, S., Rivas, M. L., Sotillos, S., Johnson, K., Knust, E, ... Hombria, J. C. (2006). Coordinated control of cell adhesion, polarity, and cytoskeleton underlies Hox-induced organogenesis in *Drosophila*. *Current Biology*, *16*(22), 2206-2216.
- Lu, X., Melnick, M. B., Hsu, J-C., & Perrimon, N. (1994). Genetic and molecular analyses of mutation involved in *Drosophila raf* signal transduction. *The EMBO Journal*, *13*(11), 2592-2599.
- Malicki, J., Schughart, K., & McGinnis, W. (1990). Mouse *Hox-2.2* Specifies Thoracic Segmental Identity in *Drosophila* Embryos and Larvae. *Cell*, *63*(5), 961-967.
- Mann, R. S., & Chan, S. K. (1996). Extra specificity from *extradenticle*: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends in Genetics*, *12*(7), 258-262 and *12*(8), 328.
- Mann, M., & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, *21*(3), 255-261.
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. *Science*, *298*(5600), 1912-1934.

- Martinez-Arias, A., & Lawrence, P. A. (1985). Parasegments and compartments in the *Drosophila* embryo. *Nature*, *313*(6004), 639–642.
- Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, E. R. 3<sup>rd</sup>, Hurov, K. E., Luo, J., ... Elledge, S. J. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*, *316*(5828), 1160-1166.
- McGinnis, N., Kuziora, M. A., & McGinnis, W. (1990). Human *Hox-4.2* and *Drosophila Deformed* Encode Similar Regulatory Specificities in *Drosophila* Embryos and Larvae. *Cell*, *63*(5), 969–976.
- McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., & Gehring, W. J. (1984). A Homologous Protein-Coding Sequence in *Drosophila* Homeotic Genes and Its Conservation in Other Metazoans. *Cell*, *37*(2), 403–408.
- McGinnis, W., & Kuziora, M. (1994). The Molecular Architects of Body Design. *Scientific American*, *270*(2), 58–61, 64-66.
- McPherson, M. J., & Møller, S. G. (2000). PCR. Springer-Verlag, New York, NY.
- 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.
- Merabet, S., Saadaoui, M., Sambrani, N., Hudry, B., Pradel, J., Affolter, M., & Graba, Y. (2007). A unique Extradenticle recruitment mode in the *Drosophila* Hox protein Ultrabithorax. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 16946-16951.
- Meusemann, K., von Reumont, B. M., Simon, S., Roeding, F., Strauss, S., Kück, P., ... Misof, B. (2010). A phylogenomic approach to resolve the arthropod tree of life. *Molecular Biology and Evolution*, *27*(11), 2451-2464.

- 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.
- Monteiro, A. S., & Ferrier, D. E. K. (2006). Hox genes are not always Colinear. *International Journal of Biological Sciences*, 2(3), 95-103.
- Moodie, S. A., & Wolfman, A. (1994). The 3Rs of life: Ras, Raf and growth regulation. *Trends in Genetics*, 10(2), 44-48.
- Murtha, M. T., Leckman, J. F., & Ruddle, F. H. (1991). Detection of homeobox genes in development and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 88(23), 10711-10715.
- 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. (2005). Linear motifs: evolutionary interaction switches. *FEBS Letters*, 579(15), 3342-3345.
- Neduva, V., & Russell, R. B. (2006). Peptides mediating interaction networks: new leads at last. *Current Opinion in Biotechnology*, 17(5), 465-471.
- Negre, B., Casillas, S., Suzanne, M., Sánchez-Herrero, E., Akam, M., Nefedov, M., ... Ruiz, A. (2005). Conservation of regulatory sequences and gene expression patterns in the disintegrating *Drosophila Hox* gene complex. *Genome Research*, 15(5), 692–700.

- Ng, S. S., Yue, W. W., Oppermann, U., & Klose, R. J. (2009). Dynamic protein methylation in chromatin biology. *Cellular and Molecular Life Sciences*, 66(3), 407-422.
- Nielson, M. L., Savitski, M. M., & Zubarev, R. A. (2006). Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Molecular & Cellular Proteomics*, 5(12), 2384-2391.
- Nolting, D., Malek, R., Makarov, A. (2017). Ion traps in modern mass spectrometry. *Mass Spectrometry Reviews*, doi: 10.1002/mas.21549.
- Nüsslein-Volhard, C. (1979). Maternal effect mutations that alter the spatial coordinates of the embryo of *Drosophila melanogaster*. In Determination of spatial organization. (ed. G. Subtelney and I.R. Koenigsberg), pp. 185-211. Academic Press, New York.
- Nüsslein-Volhard, C. (1991). Determination of the embryonic axes of *Drosophila*. *Development*, 113(Issue Supplement 1), 1-10.
- Nüsslein-Volhard, C., Frohnhofer, H. G., & Lehmann, R. (1987). Determination of anteroposterior polarity in *Drosophila*. *Science*, 238(4834), 1675-1681.
- Nüsslein-Volhard, C., & Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287(5785), 795-801.
- Nüsslein-Volhard, C, Wieschaus, E., & Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* - I. Zygotic loci on the second chromosome. *Wilhelm Roux's Archives of Developmental Biology*, 193, 267-282.
- O'Connell, N. E., Lelli, K., Mann, R. S., & Palmer, A. G. 3<sup>rd</sup> (2015). Asparagine deamidation reduces DNA-binding affinity of the *Drosophila melanogaster* Scr homeodomain. *FEBS Letters*, 589(21), 3237-3241.
- O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry*, 250(10), 4007-4021.

- O'Farrell, P. Z., Goodman, H. M., & O'Farrell, P. H. (1977). High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell*, *12*(4), 1133-1141.
- Ojala, P. M., Yamamoto, K., Castaños-Vélez, E., Biberfeld, P., Korsmeyer, S. J., & Mäkelä, T. P. (2000). The apoptotic v-cyclin-CDK6 complex phosphorylates and inactivates Bcl-2. *Nature Cell Biology*, *2*(11), 819-825.
- Olivier, J. P., Raabe, T., Henkemeyer, M., Dickson, B., Mbamalu, G., Margolis, B., ... Pawson, T. (1993). A *Drosophila* SH2-SH3 adaptor protein implicated in coupling the Sevenless tyrosine kinase to an activator of Ras guanidine nucleotide exchange, SOS. *Cell*, *73*(1), 179-191.
- Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., ... Mann, M. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science Signaling*, *3*(104), ra3.
- O'Neill, E. M., Rebay, I., Tjian, R., & Rubin, G. M. (1994). The activities of two Ets-related transcription factors required for *Drosophila* eye development are modulated by the Ras/MAPK pathway. *Cell*, *78*(1), 137-147.
- 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.
- Pace, R. M., Grbić, M., & Nagy, L. M. (2016). Composition and genomic organization of arthropod Hox clusters. *Evo Devo*, *7*, 11.
- 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.

- Pascual-Anaya, J., Adachi, N., Álvarez, S., Kuratani, S., D'Aniello, S., & Garcia-Fernández, J. (2012). Broken colinearity of the amphioxus *Hox* cluster. *EvoDevo*, 3(1), 28.
- 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.
- Pearson, J. C., Lemons, D., & McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nature Reviews Genetics*, 6(12), 893-904.
- Peifer, M., & Wieschaus, E. (1990). Mutations in the *Drosophila* gene *extradenticle* affect the way specific homeo domain proteins regulate segmental identity. *Genes & Development*, 4(7), 1209–1223.
- 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 melanogaster*. *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.
- Perdivara, I., Deterding, L. J., Przybylski, M., & Tomer, K. B. (2010). Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: chemical artifact or post-translational modification? *Journal of the American Society for Mass Spectrometry*, 21(7), 1114-1117.
- Powers, T. P., Hogan, J., Ke, Z., Dymbrowski, K., Wang, X., Collins, F. H., & Kaufman, T. C. (2000). Characterization of the Hox cluster from the mosquito *Anopheles gambiae* (Diptera: culicidae). *Evolution & Development*, 2(6), 311-325.



- 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.
- Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. *Nature*, *322*(6081), 697-701.
- Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature*, *335*(6192), 683-689.
- 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.
- Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Müller, M., Gehring, W. J., & Wüthrich, K. (1994). Nuclear magnetic resonance solution structure of the *fushi tarazu* homeodomain from *Drosophila* and comparison with the *Antennapedia* homeodomain. *Journal of Molecular Biology*, *238*(3), 333-345.
- 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.
- Rebay, I., & Rubin, G.M. (1995). Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell*, *81*(6), 857-866.
- 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.

- Rogge, R. D., Karlovich, C. A., & Banerjee, U. (1991). Genetic dissection of a neurodevelopmental pathway: Son of sevenless functions downstream of the sevenless and EGF receptor tyrosine kinases. *Cell*, *64*(1), 39-48.
- Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., & Honig, B. (2009). 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.
- Ryder, E., & Russell, S. (2003). Transposable elements as tools for genomics and genetics in *Drosophila*. *Briefings in Functional Genomics and Proteomics*, *2*(1), 57-71.
- 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.
- Samadi, L., & Steiner, G. (2010). Expression of *Hox* genes during the larval development of the snail, *Gibbula varia* (L.)--further evidence of non-colinearity in molluscs. *Development Genes and Evolution*, *220*(5-6), 161-172.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.

- Sánchez-Herrero, E., Guerrero, I., Sampedro, J., & Gonzáles-Reyes, A. (1994). Developmental consequences of unrestricted expression of the *abd-A* gene of *Drosophila*. *Mechanisms of Development*, 46(3), 153-167.
- Sánchez-Herrero, E., Vernós, I., Marco, R., & Morata, G. (1985). Genetic organization of *Drosophila* bithorax complex. *Nature*, 313(5998), 108-113.
- Schiemann, S. M., Martín-Durán, J. M., Børve, A., Vellutini, B. C., Passamaneck, Y. J., & Hejnov, A. (2017). Clustered brachiopod Hox genes are not expressed collinearly and are associated with lophotrochozoan novelties. *Proceedings of the National Academy of Sciences of the United States of America*, 114(10), E1913-E1922.
- Schierwater, B., Murtha, M., Dick, M., Ruddle, F. H., & Buss, L. W. (1991). Homeoboxes in Cnidarians. *The Journal of Experimental Zoology*, 260(3), 413-416.
- Schneuwly, S., Klemenz, R., & Gehring, W. J. (1987). Redesigning the body plan of *Drosophila* by ectopic expression of the homeotic gene *Antennapedia*. *Nature*, 325(6107), 816-818.
- Schummer, M., Scheurlen, I., Schaller, C., & Galliot, B. (1992). HOM/HOX homeobox genes are present in hydra (*Chlorohydra viridissima*) and are differentially expressed during regeneration. *The EMBO Journal*, 11(5), 1815-1823.
- Scott, M. P., Weiner, A. J., Hazelrigg, T. I., Polisky, B. A., Pirrotta, V., Scalenghe, F., & Kaufman, T. C. (1983). The Molecular Organization of the *Antennapedia* Locus of *Drosophila*. *Cell*, 35(3 Pt 2), 763-776.
- Seetharam, A. S., & Stuart, G. W. (2013). Whole genome phylogeny for 21 *Drosophila* species using predicted 2b-RAD fragments. *PeerJ*, 1, e226.
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., ... Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell*, 119(7), 941-953.

- Shuldiner, A. R., Scott, L. A., & Roth, J. (1990). PCR-induced (ligase-free) subcloning: a rapid reliable method to subclone polymerase chain reaction (PCR) products. *Nucleic Acids Research*, *18*(7), 1920.
- Shuldiner, A. R., Tanner, K., Scott, L. A., Moore, C. A., & Roth, J. (1991). Ligase-free subcloning: A versatile method to subclone polymerase chain reaction (PCR) products in a single day. *Analytical Biochemistry*, *194*(1), 9-15.
- 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.
- Simon, M. A., Bowtell, D. D., Dodson, G. S., Lavery, T. R., & Rubin, G. M. (1991). Ras1 and putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase. *Cell*, *67*(4), 701-716.
- Simon, M. A., Dodson, G. S., & Rubin, G. M. (1993). An SH3-SH2-SH3 protein is required for p21<sup>Ras1</sup> activation and binds to Sevenless and SOS proteins *in vitro*. *Cell*, *73*(1), 169-177.
- Sindona, G. & Taverna, D. (2012). Modern Methodologies to Assess the Olive Oil Quality. Olive Germplasm - The Olive Cultivation, Table Olive and Olive Oil Industry in Italy, Dr. Innocenzo Muzzalupo (Ed.), InTech, DOI: 10.5772/52142. Available from: <https://www.intechopen.com/books/olive-germplasm-the-olive-cultivation-table-olive-and-olive-oil-industry-in-italy/modern-methodologies-to-assess-the-olive-oil-quality>.
- Singhal, N., Kumar, M., Kanaujia, P. K., & Viridi, J. S. (2015). MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology*, *6*, 791.
- Sivanantharajah, L. (2013). Investigating the functional significance of evolutionarily conserved protein motifs of the *Drosophila melanogaster* HOX protein, Sex combs

reduced. Ph.D. Thesis. The University of Western Ontario, London, Ontario, Canada.

- 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.
- Sprang, S. R., Acharya, K. R., Goldsmith, E. J., Stuart, D. I., Varvill, K., Fletterick, R. J., ... Johnson, L. N. (1988). Structural changes in glycogen phosphorylase induced by phosphorylation. *Nature*, 336(6196), 215-221.
- Sprung, R., Chen, Y., Zhang, K., Cheng, D., Zhang, T., Peng, J., & Zhao, Y. (2008). Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. *Journal of Proteome Research*, 7(3), 1001-1006.
- Stadtman, E. R., & Levine, R. L. (2003). Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. *Amino Acids*, 25(3-4), 207-218.
- Stankunas, K., Berger, J., Ruse, C., Sinclair, D. A. R., Randazzo, F., & Brock, H. W. (1998). The *Enhancer of Polycomb* gene of *Drosophila* encodes a chromatin protein conserved in yeast and mammals. *Development*, 125(20), 4055-4066.
- 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.

- Steen, H., & Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews Molecular Cell Biology*, 5(9), 699-711.
- St Johnston, D., & Nüsslein-Volhard, C. (1992). The Origin of Pattern and Polarity in the *Drosophila* Embryo. *Cell*, 68(2), 201–219.
- 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.
- Struhl, G. (1985). Near-reciprocal phenotypes caused by inactivation or indiscriminate expression of the *Drosophila* segmentation gene *ftz*. *Nature*, 318(6047), 677–680.
- Struhl, G., & Basler, K. (1993). Organizing Activity of Wingless Protein in *Drosophila*. *Cell*, 72(4), 527–540.
- 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.
- 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.
- Swalla, B. J., & Smith, A. B. (2008). Deciphering deuterostome phylogeny: molecular, morphological and paleontological perspectives. *Philosophical transactions of the Royal Society of London. Series B. Biological Sciences*, 363(1496), 1557-1568.
- Theillet, F. X., Smet-Nocca, C., Liokatis, S., Thongwichian, R., Kosten, J., Yoon, M. K., ... Selenko, P. (2012). Cell signaling, post-translational protein modifications and NMR spectroscopy. *Journal of Biomolecular NMR*, 54(3), 217-236.
- 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.
- Tomancak, P., Beaton, A., Weiszmam, R., Kwan, E., Shu, S., Lewis, S. E., ... Rubin, G. M. (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biology*, 3(12), RESEARCH0088.
- Tomancak, P., Berman, B. P., Beaton, A., Weiszmam, R., Kwan, E., Hartenstein, V., ... Rubin, G. M. (2007). Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biology*, 8(7), R145.
- Tomba, P., Fuxreiter, M., Oldfield, C. J., Simon, I., Dunker, A. K., & Uversky, V. N. (2009). Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 31(3), 328-335.
- Torgerson, D. G., Boyko, A. R., Hernandez, R. D., Indap, A., Hu, X., White, T. J., ... Clark, A. G. (2009). Evolutionary Processes Acting on Candidate *cis*-Regulatory Regions in Humans Inferred from Patterns of Polymorphism and Divergence. *PLoS Genetics*, 5(8), e1000592.
- Tour, E., Hittinger, C. T., & McGinnis, W. (2005). Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development*, 132(23), 5271-5281.
- Tsuchimoto, J., & Yamaguchi, M. (2014). *Hox* expression in the direct-type developing sand dollar *Peronella japonica*. *Developmental Dynamics*, 243(8), 1020-1029.
- Tsuda, L., Inoue, Y. H., Yoo, M. A., Mizuno, M., Hata, M., Lim, Y. M., ... Nishida, Y. (1993). A protein kinase similar to MAP kinase activator acts downstream of the Raf kinase in *Drosophila*. *Cell*, 72(3), 407-414.
- 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.
- Vavra, S. H., & Carroll, S. B. (1989). The zygotic control of *Drosophila* pair-rule gene expression. I. A search for new pair-rule regulatory loci. *Development*, 107(3), 663-672.
- Venken, K. J. T., & Bellen, H. J. (2005). Emerging technologies for gene manipulation in *Drosophila melanogaster*. *Nature Reviews Genetics*, 6(3), 167–178.
- Venken, K. J. T., & Bellen, H. J. (2007). Transgenesis upgrades for *Drosophila melanogaster*. *Development*, 134(20), 3571–3584.
- Venken, K. J. T., He, Y., Hoskins, R. A., & Bellen, H. J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science*, 314(5806), 1747-1751.
- Volkin, D. B., Mach, H., & Middaugh, C. R. (1997). Degradative covalent reactions important to protein stability. *Molecular Biotechnology*, 8(2), 105-122.
- Wakimoto, B. T., & Kaufman, T. C. (1981). Analysis of Larval Segmentation in Lethal Genotypes Associated with the Antennapedia Gene Complex in *Drosophila melanogaster*. *Developmental Biology*, 81(1), 51–64.
- Wakimoto, B. T., Turner, F. R., & Kaufman, T. C. (1984). Defects in embryogenesis in mutants associated with the antennapedia gene complex of *Drosophila melanogaster*. *Developmental Biology*, 102(1), 147-172.
- Walsh, C. T., Garneau-Tsodikova, S., & Gatto Jr., G. J. (2005). Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angewandte Chemie*, 44(45), 7342-7372.
- Wassarman, D. A., Therrien, M., & Rubin, G. M. (1995). The Ras signaling pathway in *Drosophila*. *Current Opinion in Genetics and Development*, 5(1), 44-50.



- Weiner, A. J., Scott, M. P., & Kaufman, T. C. (1984). A Molecular Analysis of *fushi tarazu*, a Gene in *Drosophila melanogaster* That Encodes a Product Affecting Embryonic Segment Number and Cell Fate. *Cell*, 37(3), 843–851.
- Wells, J. M., & McLuckey, S. A. (2005). Collision-induced dissociation (CID) of peptides and proteins. *Methods in Enzymology*, 402, 148-185.
- Whitelegge, J. P. (2009). Protein mass spectrometry. (Amsterdam; London: Elsevier)
- Whitelegge, J. P., Gunderson, C. B., & Faull, K. F. (1998). Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Science*, 7(6), 1423-1430.
- Wiellette, E. L., Harding, K. W., Mace, K. A., Ronshaugen, M. R., Wang, F. Y., & McGinnis, W. (1999). *spen* encodes an 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., & Noell, E. F. (1986). Specificity of embryonic lethal mutations in *Drosophila* analyzed in germ line clones. *Wilhelm Roux's Archives of Developmental Biology*, 195, 63-73.
- Wieschaus, E., & Nüsslein-Volhard, C. (1986). Looking at embryos, pp. 199-228 in *Drosophila: A Practical Approach*, edited by D. B. ROBERTS. IRL, Oxford.
- Wieschaus, E., Nüsslein-Volhard, C., & Jürgens, G. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* - III. Zygotic loci on the X-chromosome and fourth chromosome. *Wilhelm Roux's Archives of Developmental Biology*, 193(5), 296–307.
- Wiśniewski, J. R., Zougman, A., & Mann, M. (2008). N<sup>ε</sup>-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.

- Witze, E. S., Old, W. M., Resing, K. A., & Ahn, N. G. (2007). Mapping protein post-translational modifications with mass spectrometry. *Nature Methods*, 4(10), 798–806.
- Wolff, T. and Ready, D. F. (1993). *Pattern formation in the Drosophila retina*. In: *The Development of Drosophila melanogaster*. Cold Spring Harbor Laboratory Press. Vol. 2 Pp. 1277-1325.
- Wright, H. T., & Urry, D. W. (1991). Nonenzymatic deamidation of asparaginyl and glutaminyl residues in protein. *Critical Reviews in Biochemistry and Molecular Biology*, 26(1), 1-52.
- Xu, C., Kauffmann, R. C., Zhang, J., Kladny, S., Carthew, R. W. (2000). Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the *Drosophila* eye. *Cell*, 103(1), 87-97.
- 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., Chen, Y., Zhang, Z., Xing, G., Wysocka, J., & Zhao, Y. (2010). MS/MS/MS reveals false positive identification of histone serine methylation. *Journal of Proteome Research*, 9(1), 585-594.
- 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.
- Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A. N., Heck, A. J., & Mohammed, S. (2013). Toward a comprehensive characterization of a human cancer cell phosphoproteome. *Journal of Proteome Research*, 12(1), 260-271.

## Appendices

### Appendix A. Permission to use Figure 1

Re: Thesis figure...

Hi Anirban,

Glad you hear you're finishing up. I'd be happy to have you use those images.

Good luck with the writing!

Lovesha

**Subject:** Thesis figure...

Hi Lovesha,

Hope you are well. I am writing my thesis now and was wondering if I could use one of the figures from your PhD thesis. It is figure 1.1A and B which shows the hierarchy of proteins expressed to control development and their mutant phenotypes represented pictorially. If you permit me, it will be of great help. Please let me know. Thank you.

## Appendix B. Permission to use Figure 2

Dear, Anirban Banerjee

- 1) You have a permission to use the illustration for your PhD/thesis.  
(PhD/thesis only)
- 2) You have a permission to re-design fig.Drosophila Embryo.(on p.61)  
(PhD/thesis only)

Re: SciAm/art

Anirban Banerjee

Wed 2017-10-04 8:04 PM

Hello,

Thanks for your reply and also for your permission. Since, I have to attach a written permission from you at the end of my PhD thesis, could you please write a formal email to me or a letter permitting me to use the figure on Page 61 of the Scientific American article (link below and paper attached)

McGinnis, W., & Kuziora, M. (1994). The Molecular Architects of Body Design. *Scientific American*, 270(2), 58-66. Retrieved from <http://www.jstor.org.proxy1.lib.uwo.ca/stable/24942591>

Please mention article and figure details in that email or letter.

I am sorry to bother you. But this will be of help to me. Thanks.

## Appendix C. Permission to use Figure 5

Re: Regarding permission to use a figure..

Dear Anirban,

My name is Ivan Šuljić and I am writing you from InTech's Publishing Ethics and Legal Affairs Department.

You may freely use the figure in your publication as long as you cite the source chapter of the original figure, and include the following notice in the figure caption:

© 2012 Sindona G, Taverna D. Published in [short citation] under CC BY 3.0 license. Available from: <http://dx.doi.org/10.5772/52142>

If you need any other help feel free to contact me.

Kind regards,

Ivan

Regarding permission to use a figure..

Anirban Banerjee

Sat 2017-09-16 5:56 AM

Hello,

I am Anirban Banerjee, a PhD candidate in Biology at Western University, London, Ontario, Canada and I would like to use a figure from one of your publications in my PhD thesis.

The link to the publication is:

<https://www.intechopen.com/books/olive-germplasm-the-olive-cultivation-table-olive-and-olive-oil-industry-in-italy/modern-methodologies-to-assess-the-olive-oil-quality>

I would like to use Figure 4 from subsection 1.2.3 in my PhD thesis.

If you could give me an official written permission to use this figure, I would be grateful to you. Please let me know if you need anything. Thanks.

## Curriculum Vitae

**Name:** Anirban Banerjee

**Post-secondary Education and Degrees:**

West Bengal University of Technology  
Kolkata, West Bengal, India  
2004-2008 B.Tech.

University of Saskatchewan  
Saskatoon, Saskatchewan, Canada  
2008-2011 M.Sc.

The University of Western Ontario  
London, Ontario, Canada  
2013-2018 Ph.D.

**Honors and Awards:**

Thesis Completion Fund Scholarship (UWO)  
2017-18

Western Graduate Research Scholarship (UWO)  
2013-17

Graduate Student Travel Award (UWO)  
2017

Faculty of Science Graduate Student Teaching Award (UWO)  
2015-16

Graduate Teaching Assistant Award (UWO)  
2014-15

Dr. René R. Roth Memorial Award (UWO)  
2014

University of Saskatchewan Graduate Teaching Assistantship  
2010-11

University of Saskatchewan Graduate Teaching Assistantship  
2009-10

Zacharuk Scholarship (Mohyla Institute, Saskatoon)  
2009-10

**Related Work Experience:** Teaching Assistant  
The University of Western Ontario  
2013-2018

Instructor  
Prep101  
2016-2018

Teaching Assistant  
University of Saskatchewan  
2008-2011

### **Publications:**

**Banerjee, A.**, & Percival-Smith, A. (2018). Post-translational modifications of *Drosophila melanogaster* HOX protein, Sex combs reduced. Manuscript in preparation.

Hayes, S., Wang, W., Rajamanickam, K., Chu, A., **Banerjee, A.**, & Hayes, C. (2016). Lambda gpP-DnaB Helicase Sequestration and gpP-RpoB Associated Effects: On Screens for Auxotrophs, Selection for Rif<sup>R</sup>, Toxicity, Mutagenicity, Plasmid Curing. *Viruses*, 8, 172.

### **Selected Presentations:**

**Banerjee, A.**, & Percival-Smith, A. (2017). Mapping patterns of phosphorylation of HOX proteins in *Drosophila melanogaster*. Canfly XIV. Banff, Alberta, Canada. Oral Presentation.

**Banerjee, A.**, Bath, D., & Percival-Smith, A. (2015). Mapping patterns of phosphorylation of proteins important in *Drosophila melanogaster* development. Canfly XIII. Montreal, Québec, Canada. Poster Presentation.

Hayes, S., Marciniuk, K., **Banerjee, A.**, Hayes, C., Chu, A., Horbay, M., & Erker, C. (2009). Lethal/inhibitory and mutator activities of  $\lambda$  replication initiation gene. 18<sup>th</sup> Evergreen Phage Biology Meeting. Olympia, Washington, USA. Oral Presentation.

**Advisory/Supervisory Role:**

## Research Project Advisor:

2017-18: Lama Mouneimne, Honors thesis student, The University of Western Ontario  
 Thesis title: All-in-one CRISPR/Cas9 and donor DNA minicircles for safe and efficient *in vivo* targeted “knock-in” cancer gene therapy

2016-17: Lindsay Hilton, Honors thesis student, The University of Western Ontario  
 Thesis title: Temporal characterization of Wnt target genes, crosstalk with hedgehog signaling, and their involvement in F9 cell differentiation

2015-16: Taylor Lyons, Honors thesis student, The University of Western Ontario  
 Thesis title: Functional analysis of rare disease genes in learning and memory, using *Drosophila melanogaster* as a model organism

## Volunteer Student Supervisor:

2016-17: Stuart Cameron, The University of Western Ontario

2016-17: Irina Sukhova, The University of Western Ontario

2016-17: Kartik Gupta, The University of Western Ontario

2016-17: Nikola Vladicic, The University of Western Ontario

2016: Eisha Gupta, The University of Western Ontario

2015-17: Sina Fararooie, The University of Western Ontario

2014-16: Mohammad Danial Gill, The University of Western Ontario

2015: Jennifer Lee-Nguyen, The University of Western Ontario

2014-15: Claire Newman, The University of Western Ontario

**Academic Service:**

2015-17: Founder Member of Western Synthetic Biology Research (WSBR) Program at  
 The University of Western Ontario

2015-16: Departmental Steward of Biology for PSAC Local 610 (TA Union) of The  
 University of Western Ontario

2015-16: Advisor of the Innovation Committee of The University of Western Ontario  
 chapter of the international organization, Universities Allied for Essential  
 Medicines (UAEM)

2014-15: Committee Member of the International Graduate Student Issues Committee  
 (IGSIC) of the Society of Graduate Students (SOGS) at The University of  
 Western Ontario