DISSERTATION

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the
Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Dipl.-Biol. Sonja Matt
born in Gronau (Westphalia)

Oral Examination:

Topic

Polycomb Group Proteins in the freshwater polyp *Hydra*

Referees: Prof. Dr. Gabriele Petersen
PD Dr. Renate Voit

I hereby confirm that the presented dissertation is my own work; and that, to the best of my knowledge and belief, all published or other sources of material consulted have been acknowledged in notes to the text or the bibliography. I confirm that this dissertation has not been submitted for a comparable academic award to another department or university. I confirm that this dissertation has not been published and that I will not perform publication prior to completion of the examination procedure. I am aware of the regulations concerning the conferral of doctoral degrees.

Heidelberg,

Sonja Matt

for my parents

Seul l'inconnu épouvante les hommes. Mais, pour quiconque l'affronte, il n'est déjà plus l'inconnu.

Antoine de Saint-Exupéry (Terre des Hommes, 1939)

Only the unknown frightens men. But once a man has faced the unknown, that terror becomes the known.

Antoine de Saint-Exupéry (Wind, Sand and Stars, 1939)

Nur das Unbekannte ängstigt den Menschen. Sobald man ihm die Stirn bietet, ist es schon kein Unbekanntes mehr.

Antoine de Saint-Exupéry (Wind, Sand und Sterne, 1939)

Abstract

During the development of a multicellular organism, different cell types with specific functions have to be generated, although the cells contain the same genetic information. This is achieved by a cell type specific expression of genes. The gene expression pattern and the identity of the cell have to be maintained for several rounds of cell division. The growing field of epigenetics deals with the inheritance of traits, which are not determined in the underlying DNA sequence. The long-term maintenance of gene expression patterns, which can be described as a kind of a cellular memory, is counted among epigenetic mechanisms. Different epigenetic inheritance systems like chromatin remodeling exist. For instance, the chromatin structure can be altered by the histone modifying Polycomb Group (PcG) proteins. The PcG proteins are responsible for the inheritance of the repressed state. Many genes, which are controlling development, differentiation and the cell cycle, are regulated by PcG proteins. The proteins have been shown to act as large multiprotein complexes, so-called Polycomb Repressive Complexes (PRCs).

In the scope of this thesis, PcG and associated proteins have been characterized in the freshwater polyp *Hydra*. The PcG genes *Scm* and *Pho/YY1* could be identified in the *Hydra* genome and have been cloned. Together with the already identified proteins of the PRC1 and PRC2 complexes, *Hydra* possesses a complete set of PcG proteins. A comparison with homologous proteins of other species could demonstrate that the functional domains of the *Hydra* PcG proteins are highly conserved.

Interestingly, the genes of the PRC2 complex are exclusively expressed in one of the three cell lineages of *Hydra*. The genes are expressed in the so-called interstitial cell lineage. Studies with sexual polyps could show that the genes are also expressed in the gonads of the animals. The cell line specific expression of the genes in *Hydra* seems to separate interstitial cells from the epithelial cell lineages.

The DNA binding protein *Hy*YY1 has been of special interest in this study. It is a homolog of the *Drosophila* protein Pho and of the human transcription factor YY1. Pho can be found in a PcG complex termed PhoRC (Pleiohomeotic Repressive Complex) in combination with dSfmbt. This complex is responsible for the recruitment of the other PcG complexes to their target genes in the fruit fly. YY1 acts as a 'molecular adapter' and links the PcG complexes to so-called Polycomb Response Elements (PREs). These cis-regulatory elements can be found close to target genes and are important for the recruitment of the PcG complexes. The *Hydra* protein has both a zinc finger domain, which is responsible for the DNA binding, and a so-called REPO domain, which is necessary for the interaction with the PRCs. With the help of gel retardation assays, it could be demonstrated that *Hy*YY1 can effectively interact with the DNA in a sequence-specific manner. The consensus sequence is equivalent to the CCAT-motif found in *Drosophila* and mammals. In the fruit fly, this sequence can be found within PREs.

A polyclonal antibody against recombinantly expressed *Hy*YY1 has been generated during this study - making immunohistochemical analyses and immunoprecipitations (IPs) possible. An indication of the interaction of *Hy*YY1 with other PcG proteins could be provided by a Co-IP. However, this preliminary result has to be verified. In addition, it should be kept in mind that the

recruitment of PcG complexes in *Hydra* is not necessarily dependent on Pho/YY1 and PRE sites. At least for mammals, other recruiting mechanisms are known.
The existing antibodies against *Hy*YY1 and other*Hydra* PcG proteins render chromatin-IPs to identify target genes possible. In a preliminary experiment, sequences of the *Hydra* genome could be precipitated. Next-generation sequencing technologies will enable large-scale ChIP-Seq experiments and the future identification of PcG target genes in *Hydra*.

Zusammenfassung

Während der Entwicklung eines vielzelligen Organismus müssen unterschiedliche Zelltypen mit spezifischen Funktionen gebildet werden, obwohl alle Zellen dieselbe Erbinformation enthalten. Dies wird durch eine zelltypspezifische Expression von Genen erreicht. Das Genexpressionsmuster - und damit die Identität der Zelle, muss über viele Zellteilungen hinweg aufrecht erhalten werden. Das wachsende Feld der Epigenetik beschäftigt sich mit der Vererbung von Merkmalen, die nicht in der DNA-Sequenz festgelegt sind. Darunter fällt auch die langfristige Aufrechterhaltung von Genexpressionsmustern, welche auch als eine Art 'Zellgedächtnis' bezeichnet werden kann. Es existieren verschiedene epigenetische Vererbungssysteme, wie beispielsweise der Chromatinumbau ('chromatin remodeling'). Die Chromatinstruktur kann unter anderem durch die Histon-modifizierenden Polycomb Group (PcG) Proteine derart verändert werden, dass die Transkription verhindert wird. PcG Proteine sind für die Vererbung des reprimierten Zustandes verantwortlich. Viele Gene, welche die Entwicklung, die Differenzierung und den Zellzyklus kontrollieren, werden durch PcG Proteine reguliert. Die Proteine agieren dabei als große Multiproteinkomplexe, sogenannte Polycomb Repressive Complexes (PRCs).
Im Rahmen dieser Arbeit wurden Polycomb Group (PcG) und assoziierte Proteine im Süßwasserpolypen *Hydra* näher untersucht. Im Laufe dieser Arbeit konnten die PcG Gene *Scm* und *Pho/YY1* im *Hydra*-Genom identifiziert und kloniert werden. Zusammen mit den in früheren Studien identifizierten PcG Proteine der PRC1 und PRC2 Komplexe besitzt *Hydra* damit einen vollständigen Satz an PcG Proteinen. Durch den Vergleich mit homologen Proteinen anderer Organismen konnte gezeigt werden, dass vor allem die funktionellen Domänen der PcG Proteine in *Hydra* hoch konserviert sind.
Interessanterweise werden die Gene des PRC2 Komplexes ausschließlich in einer der drei Zelllinien von *Hydra* exprimiert. Die Gene werden in der sogenannten interstitiellen Zelllinie exprimiert. Anhand von Studien an geschlechtlichen Polypen konnte außerdem gezeigt werden, dass die Gene in den Gonaden der Tiere exprimiert werden. Die zelllinienspezifische Expression der Gene in *Hydra* scheint eine Abgrenzung der interstitiellen Zellen von den beiden Epithelzelllinien darzustellen.
Von besonderem Interesse war das DNA bindende Protein *Hy*YY1, welches ein Homolog des *Drosophila* Proteins Pho und des humanen Transkriptionsfaktors YY1 ist. In der Fruchtfliege bildet Pho zusammen mit dSfmbt einen PcG Komplex namens PhoRC (Pleiohomeotic Repressive Complex), welcher für die Rekrutierung der anderen PcG Protein-komplexe zu den Zielgenen verantwortlich ist. YY1 agiert als ein 'molekularer Adapter' und verbindet die PcG Komplexe mit sogenannten Polycomb Response Elements (PREs). Diese cis-regulatorischen Elemente werden in der Nähe von Zielgenen gefunden und sind für die Rekrutierung der PcG Proteine von Bedeutung. Das *Hydra* Protein besitzt sowohl vier Zinkfinger, welche für die DNS-Interaktion verantwortlich sind, als auch eine REPO-Domäne, welche für die Interaktion mit den PcG Komplexen notwendig ist. Anhand von Gelretardations-Analysen konnte gezeigt werden, dass *Hy*YY1 tatsächlich sequenzspezifisch mit der DNS interagieren kann. Die Konsensussequenz entspricht dem CCAT-Motiv von *Drosophila* und Säugern. In der Fruchtfliege wird diese Sequenz in PREs gefunden.

Während dieser Arbeit wurde ein polyklonaler Antikörper gegen rekombinant exprimiertes *Hy*YY1 hergestellt, welcher immunhistochemische Analysen und Methoden wie die Immunpräzipitation (IP) erlaubt. Durch eine Co-IP konnte ein anfänglicher Hinweis auf die Interaktion von *Hy*YY1 mit anderen PcG Proteinen erbracht werden. Dieses vorläufige Ergebnis muss allerdings verifiziert werden. Desweiteren sollte bedacht werden, dass die Rekrutierung der PcG Komplexe in *Hydra* nicht zwangsläufig über Pho/YY1 und PREs erfolgen muss. Zumindest in Säugern wurden weitere Mechanismen der Rekrutierung gefunden.
Die vorhandenen Antikörper gegen *Hy*YY1 und weitere *Hydra* PcG Proteine ermöglichen zudem Chromatin-IPs zur Identifizierung von Zielgenen. In einem Vorexperiment konnten bereits Sequenzen des *Hydra*-Genoms präzipitiert werden. Sequenziertechniken der neuen Generation erlauben die zukünftige Durchführung von ChIP-Seq Experimenten im größeren Maßstab und die Identifizierung von PcG-Zielgenen in *Hydra*.

Contents

List of abbreviations

General abbreviations

6His	hexa histidine
APS	ammonium persulfate
ATP	adenosine triphosphate
AXI	ampicillin-X-Gal-IPTG
B.C.	before Christ
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BLAST	Basic Local Alignment Search Tool
bp	base pair
BSA	bovine serum albumin, fraction V
ca.	circa
$CaCl_2$	calcium chloride
cDNA	complementary DNA
cfu	colony forming unit
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
ChIP	chromatin immunoprecipitation
CIP	calf intestine phosphatase
DAPI	4′,6-diamidino-2-phenylindole
DIG	digoxygenin
DMP	dimethyl pimelimidate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide
ds	double stranded
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EMSA	electrophoretic mobility shift assay
FBS	fetal bovine serum
GFP	green fluorescent protein
GTC	guanidinium thiocyanate
H_2O	water
HCl	hydrogen chloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	high fidelity
HRP	horseradish peroxidase
IP	immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISH	in situ hybridization
KCl	potassium chloride
LB	lysogeny broth
MAB	maleic acid buffer
$MgCl_2$	magnesium chloride
NaCl	sodium chloride
$NaHCO_3$	sodium bicarbonate
NaOH	sodium hydroxide
NBT	nitro blue tetrazolium
NTMT	sodium-tris-magnesium-Tween 20
OD	optical density
oligo	oligonucleotide

PAA	protein A agarose
PAGE	polyacrylamide gel electrophoresis
PAS	protein A sepharose
PBS	phosphate buffered saline
PBST	phosphate buffered saline-Tween 20
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pen-Strep	penicillin streptomycin
pfu	plaque forming unit
PMSF	phenylmethanesulfonylfluoride
PVDF	polyvinylidene fluoride
RACE	Rapid Amplification of cDNA Ends
RB	reaction buffer
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
rpm	revolutions per minute
RT-PCR	reverse transcription polymerase chain reaction
SDS	sodium dodecyl sulfate
ss	single stranded
SSC	saline-sodium citrate
TAE	tris-acetate-EDTA
TAP	tobacco acid pyrophosphatase
Taq	Thermus aquaticus
TBE	tris-borate-EDTA
TdT	terminal deoxynucleotidyl transferase
TEA	triethanolamine
TEMED	N,N,N′,N′-Tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
Tween 20	Polyoxyethylene (20) sorbitan monolaurate
U	unit
UV	ultraviolet
WB	Western blot
X-Gal	bromo-chloro-indolyl-galactopyranoside

Physical quantities

Ω	ohm
Å	Ångström
Amp	ampere
°C	degree Celsius
cm	centimetre
Da	dalton
fmol	femtomole
g	gram
g	gravitational force
Hz	hertz
kbp	kilo-base pair
kDa	kilo-dalton
kV	kilovolts
l	litre
M	molar
MDa	mega-dalton
mg	milligram
ml	millilitre
mM	millimolar
mV	millivolt
μF	microfarad
μg	microgram
μl	microlitre
ng	nanogram
nm	nanometre
pmol	picomole

V . . . . . volt

Symbols for nucleotides

A . . . . . adenine
C . . . . . cytosine
G . . . . . guanine
T . . . . . thymine
N . . . . . any base

Symbols for amino acids

A, Ala . . . . . alanine
C, Cys . . . . . cysteine
D, Asp . . . . . aspartic acid
F, Phe . . . . . phenylalanine
G, Gly . . . . . glycine
H, His . . . . . histidine
I, Ile . . . . . isoleucine
K, Lys . . . . . lysine
L, Leu . . . . . leucine
M, Met . . . . . methionine
N, Asn . . . . . asparagine
P, Pro . . . . . proline
Q, Gln . . . . . glutamic acid
R, Arg . . . . . arginine
S, Ser . . . . . serine
T, Thr . . . . . threonine
V, Val . . . . . valine
W, Trp . . . . . tryptophan
Y, Tyr . . . . . tyrosine

1. Introduction

1.1. The freshwater polyp *Hydra* as a model organism

The freshwater polyp *Hydra* was first described at the beginning of the 18th century by Antony van Leeuwenhoek in "Philosophical Transactions of the Royal Society" in 1702 as a kind of "animalcule": *"[...] Further, I discovered a little animal whose body was at times long, at times drawn up short, and to the middle of whose body (where I imagined the undermost part of its belly was) a still lesser animalcule of the same make seemed to be fixed fast by its hinder end. [...]"* (van Leeuwenhoek, 1702).
Some years later the Swiss naturalist Abraham Trembley studied *Hydra* in detail. In 1744 he published his findings in his book "Mémoires pour servir à l'histoire d'un genre de polypes d'eau douce, à bras en forme de cornes" which was translated into German in 1791 ("Abhandlungen zur Geschichte einer Polypenart des süssen Wassers mit hörnerförmigen Armen"). Not sure whether *Hydra* is an animal or plant, Trembley cut the polyp in two halves and discovered that it can regenerate a complete head with tentacles. Therefore he compared the freshwater polyp with the serpent-like water beast called Lernaean Hydra of the greek mythology. The mythological beast possessed nine heads which grew again after cutting them. It was a guardian of the entrance to the Underworld and was killed by Herakles as second task of his Twelve Labours. Trembley coined the name *Hydra* for the freshwater polyp. Furthermore, he was the first one to describe the locomotion of *Hydra*, its asexual reproduction by budding and he discovered a positive phototactic behavior in this eyeless animal. More remarkably, Trembley performed the first regeneration and transplantation experiments in *Hydra*: he discovered that *Hydra* can regenerate a whole animal from a piece of tissue and that a piece of tissue can be grafted from one polyp to another one (Trembley, 1744). In figure 1.1 an illustration of the ancient Greek mythological Hydra and drawings of *Hydra* from the 18^{th} century are depicted.

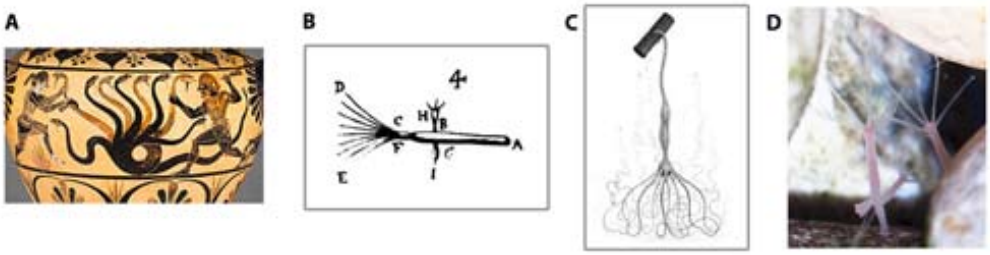


FIGURE 1.1 *The appearance of Hydra.* ***A*** *Ancient black-figure water jar depicting the Greek hero Herakles battling the Lernaean Hydra (date: about 525 B.C.; museum collection: The J. Paul Getty Museum, Malibu, California, USA).* ***B*** *Hydra as drawn by van Leeuwenhoek in 1702.* ***C*** *Observations made on the appearance of Hydra by Trembley in 1744.* ***D*** *A photograph of Hydra.*

In 1766 Peter Simon Pallas described in his book "Elenchus Zoophytorum" some more *Hydra* species. He distinguished between *Hydra oligactis*, *Hydra vulgaris*, *Hydra viridissima* and *Hydra attenuata* (Pallas, 1766).
About one and a half century later - in 1909, Ethel Nicholson Browne published her observations on grafting green *Hydra* (Browne, 1909). She was the first person who could demonstrate that a transplant of hypostomal tissue could induce a secondary axis in *Hydra* (figure 1.2). Some years later, Hans Spemann and Hilde Mangold did a very similar experiment in amphibian embryos: a part of the blastopore lip of a gastrula stage embryo induced a secondary axis when grafted into another gastrula stage embryo. The grafted tissue 'organized' its surrounding tissue (Spemann and Mangold, 1924). For their observation on the amphibian organizer, Spemann was awarded the Nobel Prize for Physiology or Medicine in 1935.

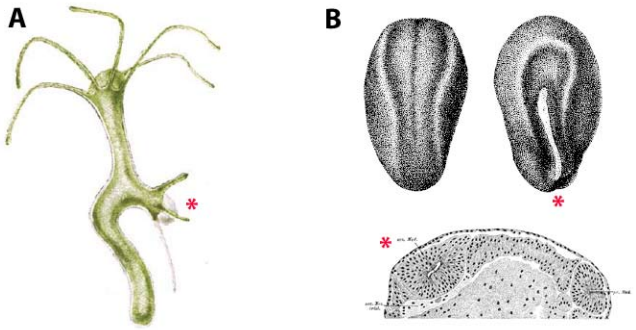


FIGURE 1.2 *Induction of secondary axes of polarity in Hydra and amphibians. (The induced secondary axes are marked by an asterisk.)* ***A*** *A transplant of hypostomal tissue (white) can induce a secondary axis in Hydra viridis (modified after Browne, 1909).* ***B*** *An eplant of a part of the blastopore lip of the newt Triturus cristatus can induce a secondary neural tube and a secondary embryo in Triturus taeniatus. The figure shows an embryo at the neural stage with primary and secondary axis (modified after Spemann and Mangold, 1924).*

In the mid-20^{th} century, the apparently simple freshwater polyp has been studied further. Studies on *Hydra's* stem cell lines and on the structure of its stinging cells, the nematocysts - have been performed (Chapman and Tilney, 1959a,b; David and Campbell, 1972; Campbell and David, 1974; David and Gierer, 1974). The high regenerative capacity and pattern formation of *Hydra* became more and more interesting for developmental biologists (for a review see Bode (2003); Fujisawa (2003); Holstein et al. (2003); Bosch (2007); Bosch et al. (2010)). Also its nerve net, which is one of the most simple neuronal systems in the animal kingdom - is of special interest for revealing the basic principles of neuronal differentiation (Davis, 1969; Dunne et al., 1985; Koizumi and Bode, 1986; Koizumi et al., 1988, 1990). In the 2000^{th} the genome of *Hydra* has been sequenced (Chapman et al., 2010) and *Hydra* can be easily cultured in the laboratory in artificial Hydra Medium (Lenhoff, 1983) making it an ideal model organism. However, compared to other classical model organisms like *Drosophila*, *C. elegans*, zebrafish or mice, *Hydra* has some major disadvatages as to functional studies:

- No *Hydra* cell culture system exists. Cells can be kept in culture for some month but they cannot be manipulated and they are not proliferating.
- There is still no possibility for knocking out genes. Although RNAi has been described in *Hydra*, the technique could not be reproduced efficiently (Lohmann et al., 1999; Galliot et al., 2007).
- Producing transgenic animals is feasible by now but still difficult (Wittlieb et al., 2006). To obtain homozygous transgenic lines is laborious and transgenic lines are often fragile in culture.

In the following sections, the characteristics of *Hydra* are described.

1.1.1. The position of *Hydra* in the metazoan phylogeny

Hydra belongs to the phylum Cnidaria. Besides the Porifera (sponges) and the Ctenophora (comb jellies), the Cnidaria belong to the most basal metazoan animals (figure 1.3). Both Ctenophora and Cnidaria are encompassed in the term Coelenterata. This term refers to the hollow body cavity of the animals. The phylum Cnidaria has evolved over 700 million years ago at the Precambrian time before the Cambrian explosion (Galliot and Miller, 2000; Chen et al., 2002). The following four classes belong to this phylum: Cubozoa (box jellyfish), Scyphozoa (true jellyfish), Anthozoa (sea anemones and corals) and Hydrozoa which includes *Hydra*. The phylum is one of the most species rich in the animal kingdom. It consists of about 9000 species whose animals are exclusively found in acquatic environments; most of the animals are marine. A common feature of all cnidarians are the eponymous cnidocytes. They are also called 'nettle cells' and are mainly used for catching prey. These specialized cells are exclusively found in the cnidarian phylum.
The classification of the Cnidaria within metazoan evolution is however discussed controversly (for a review see Ball et al., 2004; Martindale, 2005). There is a big discrepancy between morphological data which describes the animals as radially symmetric and molecular data describing them as bilateral symmetric. Interestingly the cnidarians already have a large subset of genes which is found in bilaterian animals. The simple, radial symmetric body of cnidarians is not reflected by the level of gene expression. Based on molecular data, the cnidarian phylum is classified to be a sister group to the Bilateria (Kim et al., 1999; Medina et al., 2001). In order to better understand metazoan evolution, the cnidarians are an interesting phylum to study.
Within the Cnidarian phylum, the Anthozoans are the most basal class (Kim et al., 1999). The Hydrozoans are a sister group and have diverged from the Anthozoans at least 560 million years ago (Chen et al., 2002). Wheras the Anthozoan species have a bilateral symmetric body, the Medusozoa (Scyphozoa, Cubozoa and Hydrozoa) have a radial symmetric body. Based on morphological data, it was proposed that the genus Hydra consists of four species groups: *viridissima*, *braueri*, *vulgaris* and *oligactis* group (Campbell, 1989). This classification was recently supported by molecular phylogenetic studies based on

mitochondrial DNA (Kawaida et al., 2010). The model organism used during this thesis, *Hydra magnipapillata 105* belongs to the *vulgaris* group.

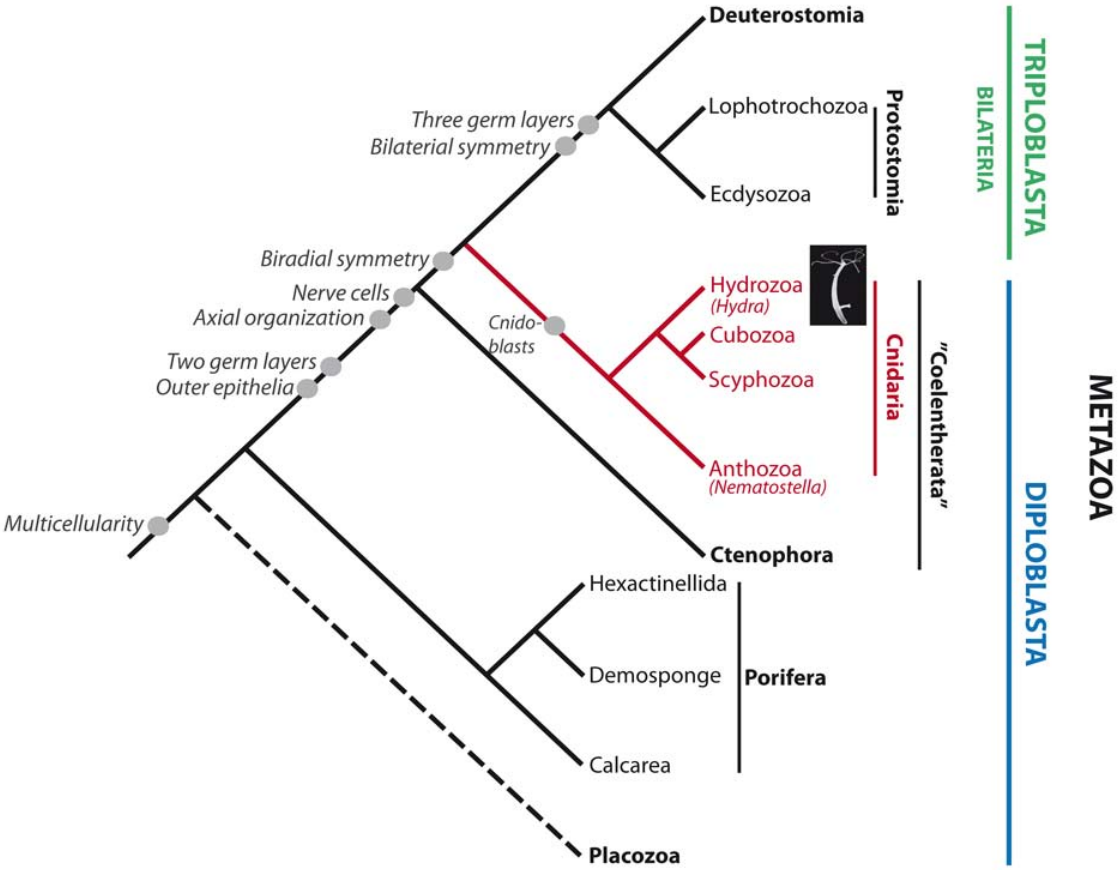


FIGURE 1.3 *Overview of metazoan phylogeny. The cnidarians (in red) are placed as a sister group to the Bilateria. They represent one of the most basal phylum in metazoan evolution. The position of the placozoans (dashed line) in metazoan evolution is discussed controversely. This phylogenetic tree refers to studies which position the placozoans basal relative to all other diploblast groups (Schierwater et al., 2009). The length and distances of the branches of the depicted tree have been chosen at random and do not allow a chronological or spacial classification. Innovations during evolution are indicated by grey dots.*

1.1.2. Characteristics of the *Hydra* genome

The genome of *Hydra magnipapillata 105* has been fully sequenced and published in 2010 (Chapman et al., 2010). It has been sequenced at the Joint Genome Institute using a whole genome shotgun approach. The genome has been assembled into contigs and can be accessed online at: *http://hydrazome.metazome.net/cgi-bin/gbrowse/hydra/*.

The *Hydra* genome is about 1.3 mega base pairs in size. It is very (A+T)-rich (over 70 %) and includes a large number of transposable elements (over 50 %); the most abundant is a non-long-terminal-repeat retroelement of the chicken repeat 1 family (CR1). *Hydra* probably contains about 20000 protein coding genes based on EST data, homology and gene prediction analysis (Table 1.1). An interesting feature of *Hydra* is the fact that genes

TABLE 1.1 *Some features of Hydra and its genome.*

Features	*Hydra magnipapillata*
class	Hydrozoa
symmetry	radial
medusa state	no (secondarily lost)
polyp state	solitary
habitat	fresh water ponds, lakes and streams
reproduction	asexual by budding and sexual
eggs per female	1 - 2
cell number in the adult polyp	50000 - 100000
diploid chromosome number	30
genome size	1.3 x 10^9 base pairs
protein coding genes	ca. 20000
GC content	29 %
transposable elements	57 % of the genome size
mitochondrial genome	two 8 kb linears

of *Hydra* undergo *trans*-splicing: A short *trans*-spliced leader sequence is added to the 5′ end of the messenger RNA which provides the opportunity to combine genes into operons (Stover and Steele, 2001). 32 potential operons are found in the *Hydra* genome. Amazingly, genes for all major bilaterian signaling pathways like Wnt (wingless/int-1), TGF-β (transforming growth factor beta), Hedgehog, RTK (receptor tyroseine kinase) and Notch are present in *Hydra*. There have also been found some non-metazoan, bacterial genes which are candidates for horizontal gene transfer.

1.1.3. Morphology of *Hydra*

Like all cnidarians *Hydra* has one oral-aboral body axis and two body layers: ectoderm and endoderm which are seperated by a acellular, jelly-like layer called mesoglea. The mesoglea of *Hydra* is similar to the vertebrate basement membrane and has the characteristics of an extracellular matrix (ECM). All major ECM proteins like collagen, laminin, fibronectin and heperan sulfate proteoglycans have been found in the mesoglea (Sarras et al., 1991a,b, 1993). The mesoglea is secreted by both epithelia. The column-like body can be arranged into a 'head' including the hypostome and tentacles and a body column which can be subdivided into a gastric region, a budding region and the peduncle. Typically, an adult polyp has about 4 - 6 tentacles. The basal disc of *Hydra* is considered to be its 'foot'. It can secrete a mucous by its ectodermal epithelial cells which allows the polyp to stick to the underground. It can also secrete a gas bubble which allows floating to the surface. *Hydra* is a rather small animal of about 0.5 - 2 centimetres in length. An overview of *Hydra's* morphology is depicted in figure 1.4.

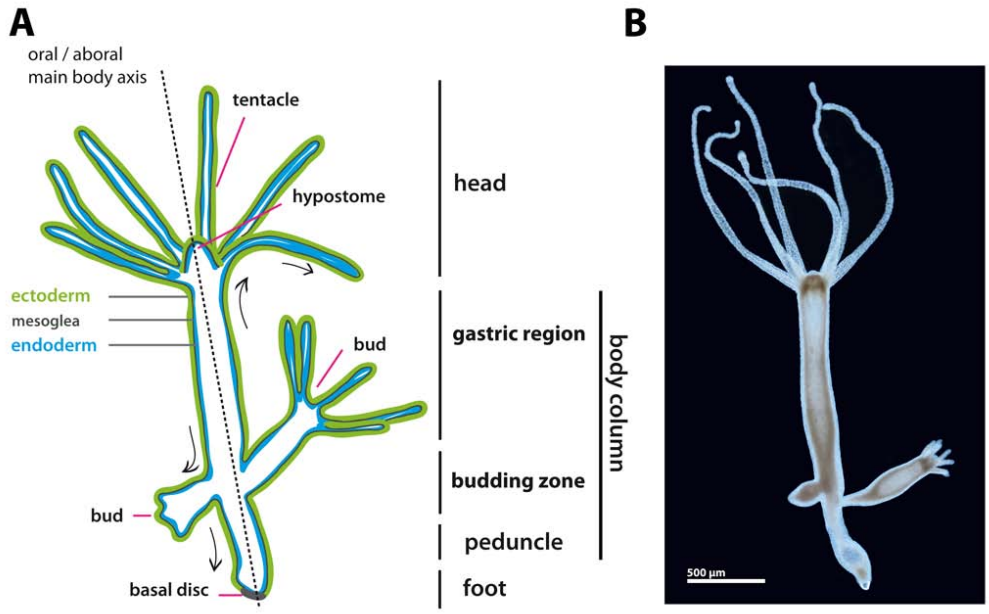


FIGURE 1.4 *Morphology of Hydra.* ***A*** *Schematic picture of Hydra. The two body layers are depicted in green (ectoderm) and blue (endoderm). The mesoglea is coloured in grey. The oral-aboral main body axis is indicated by a dashed line. The different regions of the body are marked. The arrows indicate the direction of Hydra's constant cell flow.* ***B*** *Dark field image of an adult polyp of Hydra magnipapillata 105. Scale bar: 500 µm.*

1.1.4. Cell types and capsules of *Hydra*

Hydra has three main stem cell lineages which are comprised by ectodermal epithelial stem cells, endodermal epithelial stem cells and interstitial stem cells, which are often abbreviated to I cells (for a review see Holstein, 1995). For an overview of the cellular organization of the two body layers see figure 1.5.
The different cell types of *Hydra* are well described. A technique called maceration can be used to separate the cells and to study the structure of the cells (Mundie, 1926; David, 1973). A maceration preparation and a schematic overview of *Hydra's* cells is depicted in figure 1.6.

Epithelial cells

The ectodermal epithelium secretes a glycocalyx made up of glycoproteins and polysaccharides and has a protective function. The ectodermal epithelial cells are responsible for osmoregulation (Prusch et al., 1976). The endodermal epithelial cells are used for ingestion by phagocytosis or pinocytosis and digestion. They possess two flagella to whirl food particles inside the gastric cavity. Both epithelial cell lines have muscle protrusions at their ends which are essential for contraction and elongation of the animal (Mueller, 1950; Haynes et al., 1968). Despite the fact that cnidarians are diploblastic animals that lack an obvious mesoderm, many of the animals have muscle formation and mesoderm marker genes like *mef2*, *twist* and *snail* (Martindale et al., 2004). The epithelial stem cell

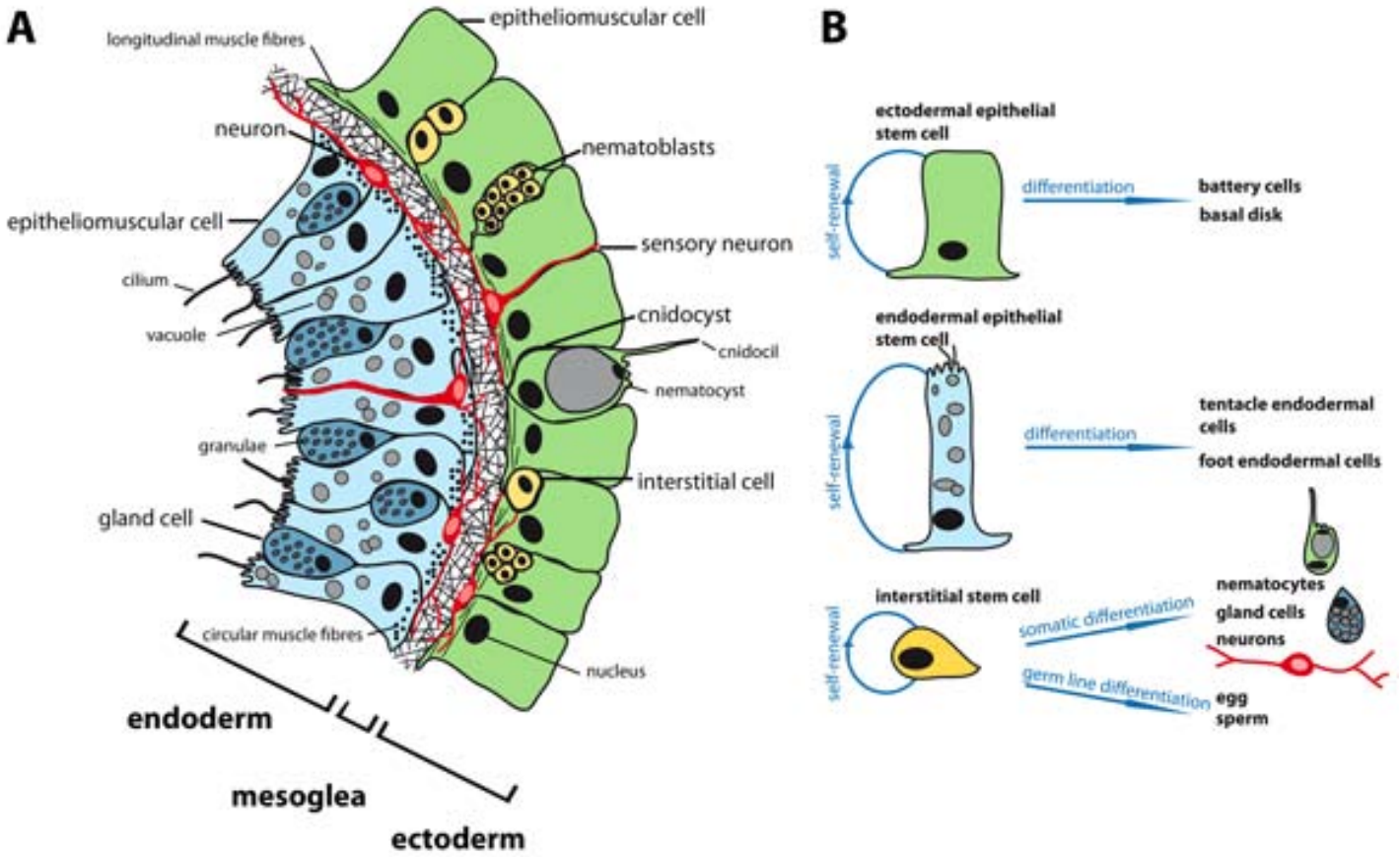


FIGURE 1.5 *Schematic overview of the two body layers of Hydra. On the left the cellular organisation of the ectodermal and endodermal layer is depicted. The different cell types are marked. The layers are seperated by the acellular mesoglea. On the right the cellular differentiation pathways for Hydra's cell lineages are shown.*

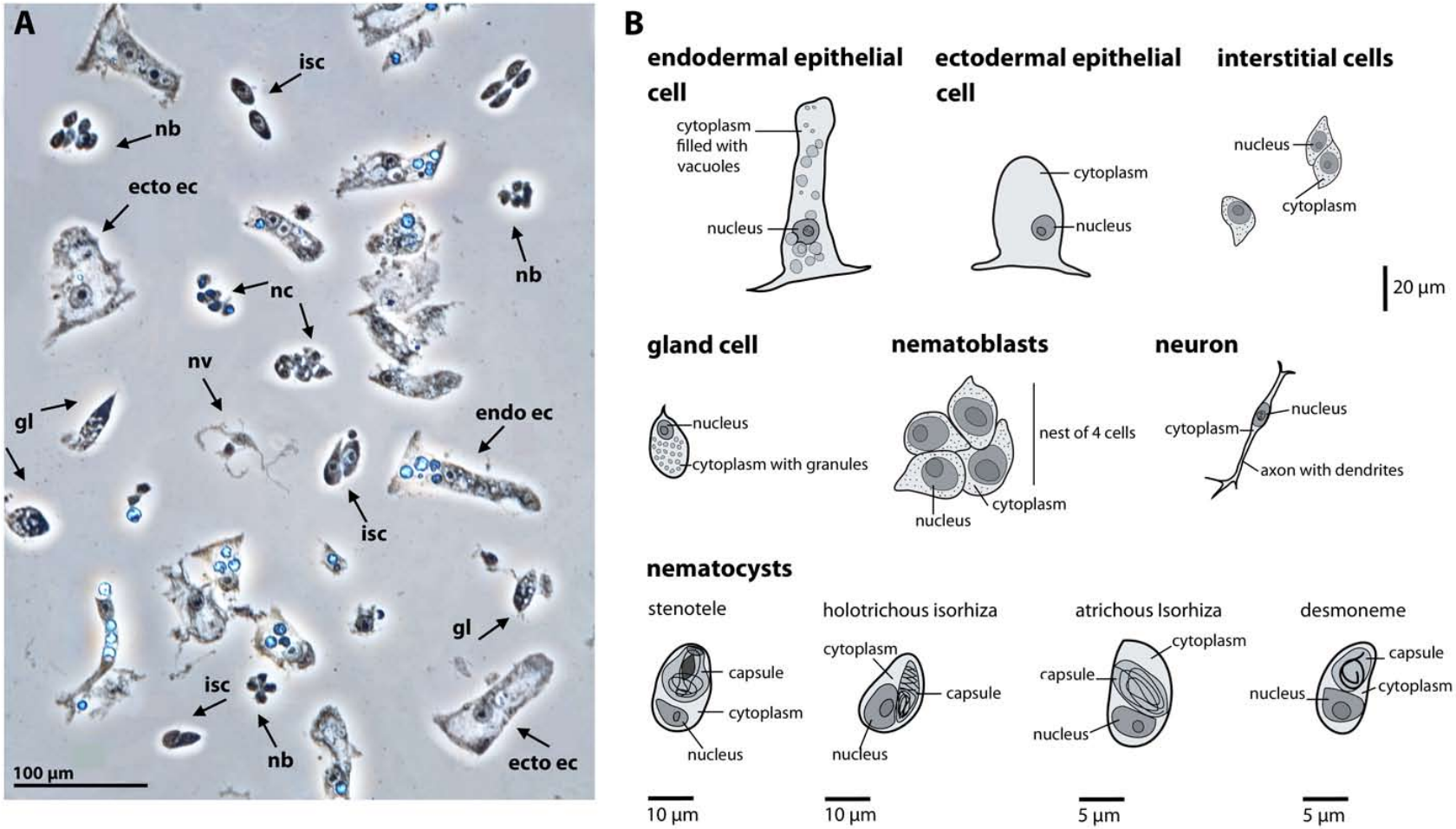


FIGURE 1.6 *Overview of the cell types of Hydra. In* ***A*** *a maceration preparation of Hydra is shown. The cells can be seperated by incubation in a mixture of formaldehyde, ethanol and glacial acetic acid. The different cell types can be easily distinguished. The ectodermal epithelial cells (ecto ec) are big cells with a uniform appearance, whereas the bottle shaped endodermal epithelial cells (endo ec) have big granules and vacuoles inside. The vacuoles appear bluish when observed under a phase contrast microscope. The I cells (Isc) exist either as single cells or as a pair of two cells which are linked by plasma bridges. The nuclei and nucleoli can be easily distinguished. Differentiated derivatives of the I cells like gland cells (gl), neurons (nv), nematoblasts (nb) and nematocytes (nc) can also be found. (Scale bar: 100 µm) In* ***B*** *the cell types are drawn schematically. The epithelial cells and the I cells with its derivatives are depicted. (The size of the cells is indicated by scale bars.)*

lines of *Hydra* are proliferating permanently generating a flow of cells from its body column to both the proximal and distal end of the animal and into its budding zone (Campbell, 1967; Otto and Campbell, 1977). Their mean cell cycle time is about three days (David and Campbell, 1972). Old cells are shedded at the tentacle tips and the basal disc. The ectodermal epithelial stem cells can differentiate into basal disc cells and so-called battery cells that can be found in the tentacles. The endodermal epithelial stem cells can differentiate into tentacle endodermal cells and foot endodermal cells.

Interstitial cells

The third pluripotent stem cell line of *Hydra*, the I cells are of special interest. They can either undergo germline differentiation into egg and sperm or somatic differentiation into neurons, *Hydra's* characteristic nematocytes and gland (zymogen) cells (Campbell and David, 1974; David and Gierer, 1974; Bode and Flick, 1976; Bode et al., 1976). The gland cells secrete proteolytic enzymes for digestion (Haynes and Burnett, 1963; Schmidt and David, 1986). The I cells constitute about 75 % of *Hydra's* cells. The mean cell cycle time of I cells is about one day. The interstitial stem cells are found throughout the gastric region of *Hydra* and seem to stay at their initial position, whereas the differentiating cells like nematoblasts and neuroblasts seem to migrate rapidly to their final destination (Bosch and David, 1990; Khalturin et al., 2007).
In *Hydra* one can distinguish between two classes of neurons: the ganglion cells and the sensory cells (Davis, 1969, 1971; David and Gierer, 1974). Ganglion cells are bipolar or multipolar nerve cells which are found throughout the body column and tentacles, sensory cells are unipolar and can be found in the body column. The nerve cells differentiate from neuronal precursor cells. After neural commitment, half of the cells migrate towards the proximal and distal end, whereas the other half stays at their position and is integrated into the nerve net (Bode and David, 1978; Heimfeld and Bode, 1985; Koizumi and Bode, 1986; Koizumi et al., 1988, 1990). The exact mechanisms of neural differentiation in *Hydra* are still unknown, but neuropeptides seem to play an important role (Takahashi et al., 2000; Hayakawa et al., 2007; Takahashi et al., 2008).
The nematocytes or cnidocytes are a characteristic feature of cnidarians. The nematocytes or stinging cells are regarded as mechanosensory or chemosensory neural cells (Hausmann and Holstein, 1985). The interstitial cells divide synchronously into clusters of 4, 8, 16 or 32 cells, the nematoblasts (David and Gierer, 1974). These cells are connected by cytoplasmic bridges and each cell produces a capsule inside a secretory vesicle (Holstein, 1981; Engel et al., 2002). After the capsule formation is completed, the cell clusters break into single cells and migrate into the tentacles. Four types of capsules can be distinguished in *Hydra*: stenoteles, atrichous isorhizas (stereoline glutinant), holotrichous isorhizas (streptoline glutinant) and desmonemes (volvent) (Chapman and Tilney, 1959a,b). In the tentacles, they are engulfed by specialized ectodermal epithelial cells, the battery cells. Normally, there are about 1 - 2 stenoteles in a battery cell, which

are surrounded by isorhizas and desmonemes. The stenotele is the most striking capsule of *Hydra*. Its stylet apparatus is armed with spines and it is used for penetrating and capturing prey. The holotrichous isorhiza has a tube populated with spinules and is supposedly used for defense. The atrichous isorhiza is used for attaching to the substrate during locomotion; the tube is devoid of spinules. Like stenoteles, the desmonemes are used for capturing prey. However, their tubes are unarmed and coiled after discharge. For an overview of *Hydra* nematocytes see figure 1.7.

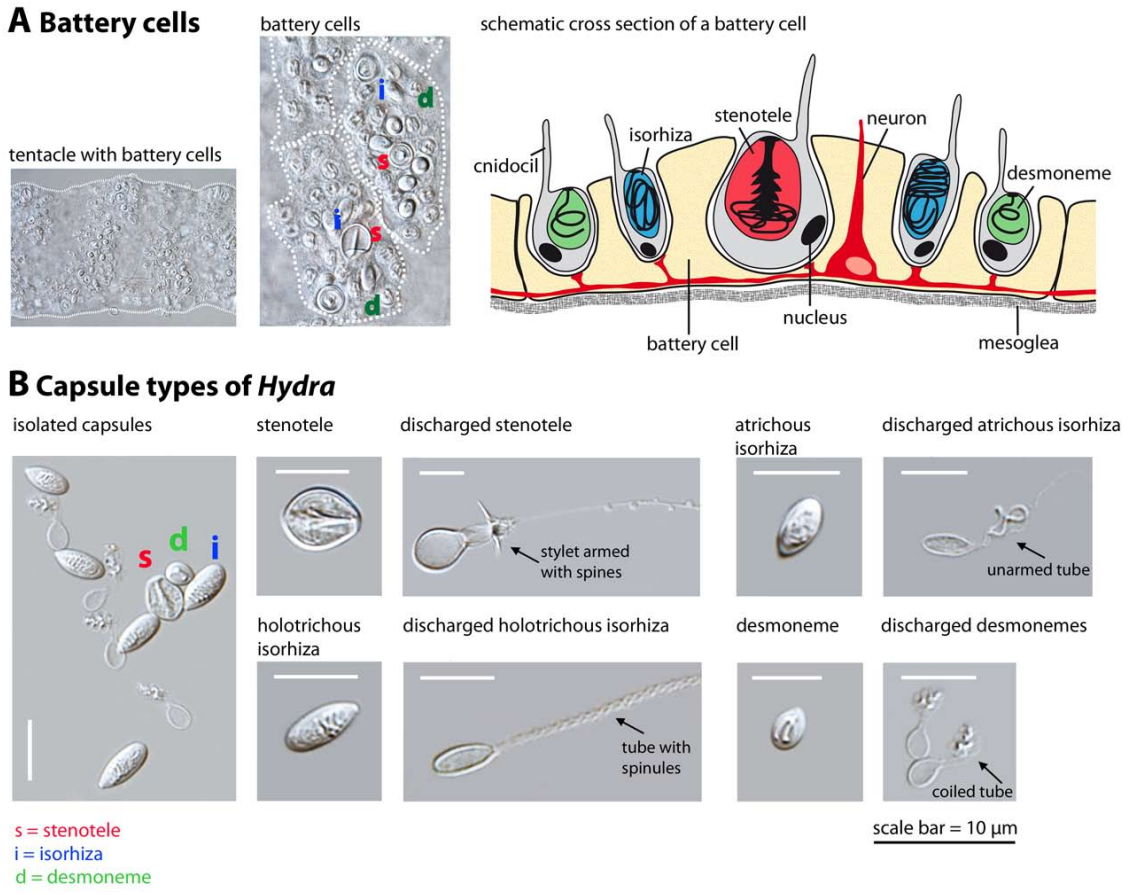


FIGURE 1.7 *Nematocytes of Hydra. In* ***A*** *a picture and a schematic overview of a battery cell are depicted. A section of a tentacle with battery cells is shown in different magnifications in both pictures on the left. The different capsule types can be easily distinguished. On the right, a schematic overview of a battery cell is shown (modified after Hufnagel et al., 1985; Hobmayer et al., 1990). In* ***B****, pictures of isolated capsules are shown. Both intact and discharged capsules are displayed. (Scale bar: 10 µm) (The capsule preparation for taking the pictures in* ***B*** *has been kindly provided by Anna Beckmann, Holstein lab.)*

The nematocyst is discharged upon mechanosensory or chemosensory stimulus; it is in a complex with a cnidocil apparatus which can be compared to a ciliary receptor. The discharge of nematocytes has been studied further in stenoteles (Tardent and Holstein, 1982; Holstein and Tardent, 1984; Nüchter et al., 2006). The process is completed within three milliseconds and the acceleration of the stilet is about 40000 g - which can be compared to a technical bullet. The wall of the *Hydra* nematocytes has to resist an enourmous internal osmotic pressure of 12.5 megapascals (Weber, 1989). To achieve its high tensile strength, the capsule wall is built up of a special type of collagens: the minicollagens (Kurz et al., 1991; Engel et al., 2001).

1.1.5. The life cycle of *Hydra* - living forever?

Hydra belongs to a family of hydrozoans which secondarily lost its medusa state. The solitary polyp state is its only appearance. It can proliferate either asexually by budding or sexually. However, not all *Hydra* species develop gonads. For an overview of *Hydra's* life cycle see figure 1.8.

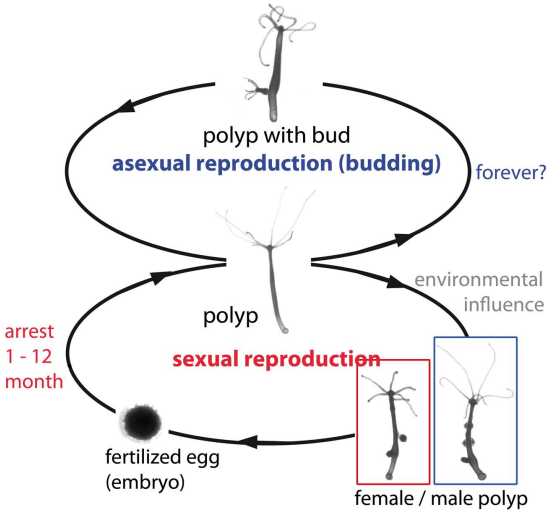


FIGURE 1.8 *The life cycle of Hydra. Hydra can either proliferate asexually by budding or sexually. Asexual reproduction by budding is the normal condition. Sexual reproduction is subjected to environmental influence.*

Polyps that reproduce asexually are potentially immortal. Asexual or vegetative reproduction takes place by budding. The budding process involves the simultaneous evagination of both ectodermal and endodermal layer. Tissue of the body column of the adult *Hydra* polyp gets transformed into tissue of the novel forming bud by morphallaxis (Holstein et al., 1991). The budding process is exclusively controlled by the epithelial stem cell lines. Polyps that lost their interstitial stem cell line were still able to form buds (Campbell, 1976; Sugiyama and Fujisawa, 1978). The buds form at the budding region of an adult polyp. An adult polyp can bear more than one bud at a time, whereat the oldest bud is next to the peduncle. The permanent proliferation of the epithelial cells and the constant cell flow of *Hydra* render the asexual reproduction by budding possible. Figure 1.9 illustrates the process.

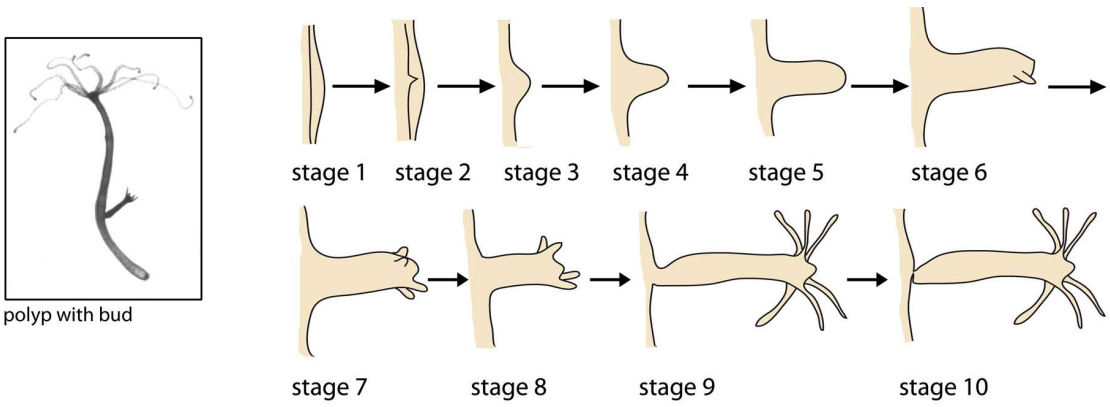


FIGURE 1.9 *The asexual or vegetative reproduction of Hydra by budding is depicted schematically. The arrangement of the different bud stages is adapted from Otto and Campbell (1977). In stages 1 - 5 the ectoderm and endoderm evaginate, in stage 6 - 8 the evagination of the tentacles occurs. In the last two stages (9 - 10) the peduncle region and the basal disc differentiate.*

Sexual reproduction in *Hydra* is possible but does not represent its standard form of reproduction. Normally, in higher metazoan animals a seperate germ line exists. However, *Hydra* is considered to be a metazoan lacking a distinct germ line. There is evidence that the interstitial stem cell line gives rise both to somatic cells and germ cells (Bosch and David, 1987). Cloning experiments showed that an interstitial cell population exists which could only differentiate into eggs (Littlefield, 1991) or sperm (Littlefield, 1985; Nishimiya-Fujisawa and Sugiyama, 1993). The sexual phenotype of *Hydra* is exclusively determined by the I cell line (Bosch and David, 1986); No sex chromosome was found in *Hydra*. There are *Hydra* species that are gonochroistic or hermaphroditic. In some *Hydra* species sex reversal can occasionally be observed. Figure 1.10 provides an overview of the sexual reproduction.

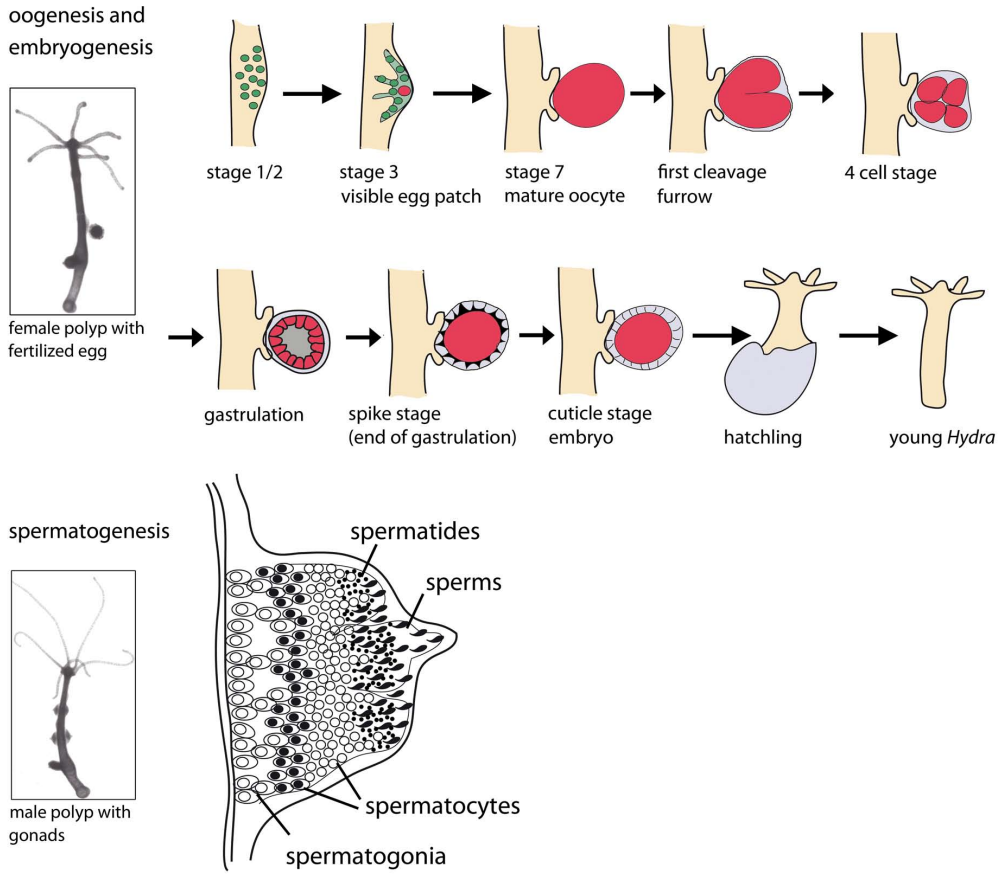


FIGURE 1.10 *Schematic overview of the sexual reproduction of Hydra. In the upper part, an overview of oogenesis and embryogenesis is given, the lower part depicts the spermatogenesis. At the beginning of oogenesis, numerous I cells accumulate in the inter-cellular space of the ectoderm (stage 1/2). One central I cell is commited to become the future oocyte (red), the surrounding I cells become nurse cells (green). A so-called egg patch becomes visible (stage 3). The oocyte increases in volume and phagocytizes adjacent apoptotic I cells. The mature oocyte breaks through the ectodermal layer and rests in the so-called egg cup (stage7). At this stage, the oocyte is ready to be fertilized. The different stages are adapted from Miller et al. (2000). The first cleavages after fertilization are unilateral, which means that the blastomeres are still interconnected by plasma bridges. During gastrulation, a small number of cells immigrates into the blastocoel to form the endodermal layer. After gastrulation (64-cell stage), the ectodermal embryonic cells secrete a two-cell layered cuticle to protect the embryo from dehydration (spike stage embryo). After an arrested stage which can last up to 12 month, the young polyp is hatching. Spermatogenesis in Hydra begins with a proliferatio of I cells in the region of the emerging gonad. After 3-4 days a mature gonad has formed in which the developmental stages of the spermatozoans exist in different layers. The mature sperms are released into the surrounding medium.*

1.1.6. Regeneration in *Hydra* - some remarkable features

Due to its highly dynamic tissue composition and the stem cell character of its cells, the freshwater polyp *Hydra* has a remarkable regenerative capacity: it can regenerate 'head' or 'foot' after amputation, and it can regenerate a whole animal from a small piece of tissue (figure 1.11). This exceptional high regenerative capacity was already detected 250 years ago by Abraham Trembley. During regeneration, the original polarity of the animal is retained: a head will regenerate at the apical end and a foot at the basal end. *Hydra's* epithelial stem cell lines are responsible for regeneration; Experiments with animals lacking interstitial cells showed normal regeneration (Holstein et al., 1991).
Even more striking is the fact that dissociated *Hydra* cells can reaggregate, reorganize and regenerate a normal animal after a short time (Noda, 1970; Gierer et al., 1972). In dissociated cell of *Hydra*, all positional information is lost and the process of regeneration and pattern formation can be studied from the very beginning (for a reviw see Holstein et al. (2003); Bode (2003, 2009). During *Hydra* regeneration no dedifferentiation or blastema formation occurs like in vertebrates. Instead, the process of regeneration is a morphogenetic process that involves subsequent cell adhesion, ecto-endo sorting, formation of the two epithelia and differentiation of head and foot (Technau and Holstein, 1992).

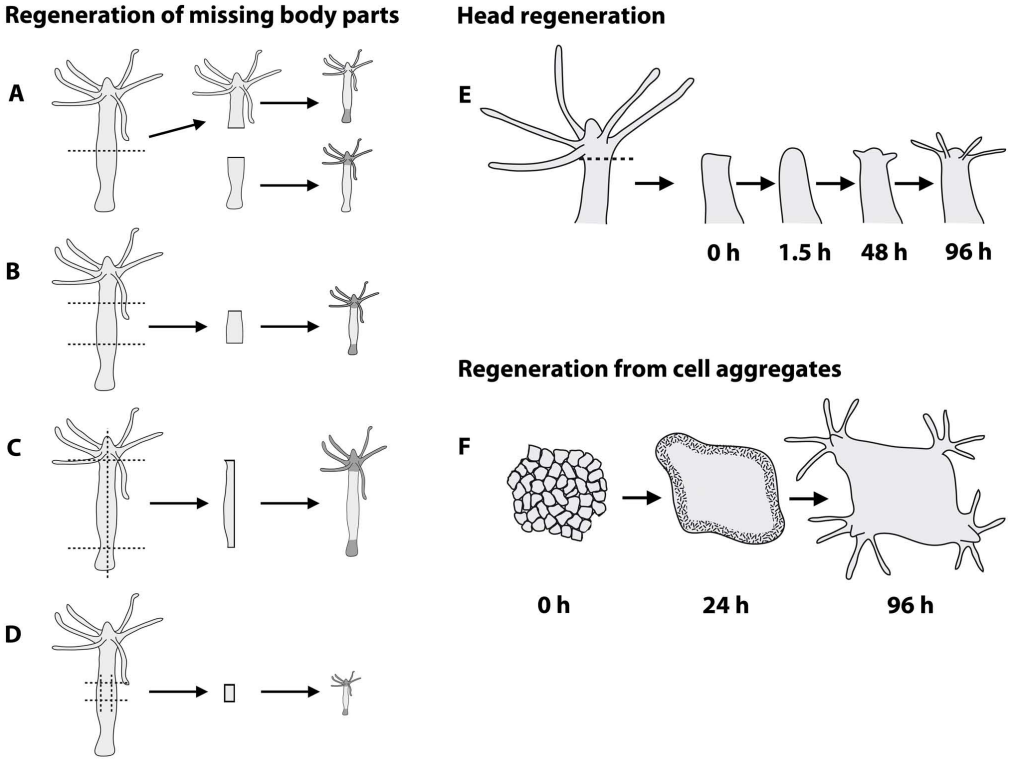


FIGURE 1.11 *Hydra has an almost unlimited regeneration capacity (after Bode (2003)). From **A** - **D** the regeneration of missing body parts is illustrated. Hydra can regenerate a new head or foot when bisected. It can regenerate both head and foot or even can regenerate from a small piece of tissue. In **E** the time course of head regeneration is depicted. Hydra can regenerate a new head within 48 hours. Head regeneration and the formation of a new organizing center is strongly dependent on the wnt pathway. **F** illustrates the process of reaggregation. Cell adhesion is followed by the formation of the two epithelial layers and the differentiation of head and foot structures.*

Alfred Gierer and Hans Meinhardt (Gierer and Meinhardt, 1972; Meinhardt and Gierer, 1974) set up a theoretical model for pattern formation processes: the activator/inhibitor model. This model is based on an auto-catalytically regulated activator, which also regulates the expression of its own inhibitor. A minimal excess of the activator can shift the balance, and both activator and inhibitor are expressed at elevated levels. Since the inhibitor has a longer diffusion range, the activator expression in the surrounding tissue is repressed and a defined signaling center/organizer is established. A version of this model can be applied to *Hydra* head formation and especially regeneration (figure 1.12). It has been proposed that a so-called 'head activator' is the key factor triggering head formation (Schaller and Gierer, 1973; Schaller, 1973). However, the 'head activator' molecule remains elusive. More recent studies provide evidence that the pattern formation process and the formation of new head organizing centers is controlled by the Wnt signaling pathway, which has been shown to be important for the function of vertebrate organizers (Hobmayer et al., 2000; Lengfeld et al., 2009). MacWilliams (1982) could show that the head regeneration is a two step process which is in accordance with the activator/inhibitor model. The simulation of Meinhardt (1993) assumes that the signals for the formation of the head, tentacles and foot are generated by seperate activator/inhibitor systems. A so-called source density gradient links the systems together. According to the model, tentacles are forming in the region of the highest source density/morphogen concentration that is not occupied by the head signal. Upon decapitation, the concentration begins to rise and temporarily reaches the threshold level of the tentacle zone. New tentacles are forming, the concentration increases further and reaches the hypostome level. Bode et al. (1988) could visualize the process by means of antibody stainings with a tentacle specific antibody (TS-19). The described setting comes true for an amputation at 80% body length. If the animal is amputated at 50% body length, the hypostome level is reached first, followed by the tentacle level.

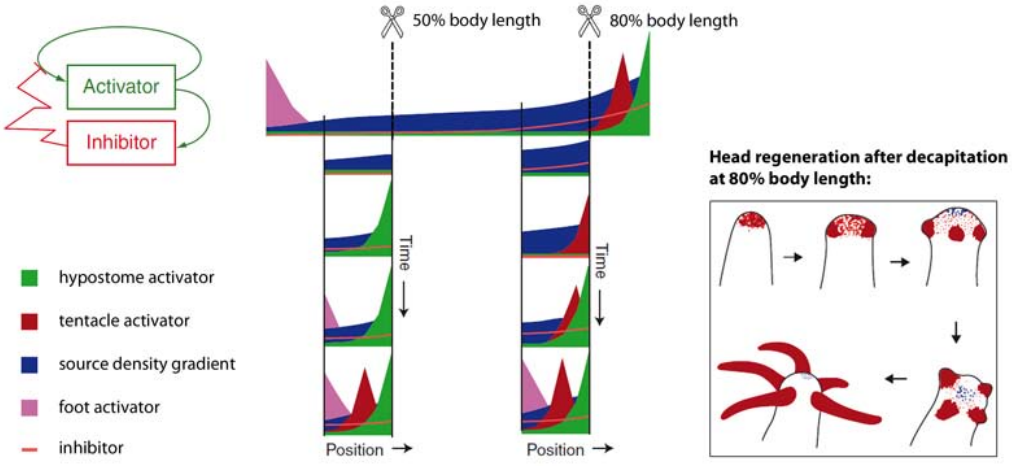


FIGURE 1.12 *'Activator/inhibitor' model for pattern formation in Hydra. Different activator/inhibitor systems are responsible for the formation of the head, tentacles and foot. The systems are coupled by the source density gradient(blue). Whereas the source density gradient is increased by the head system (green), it is decreased by the foot system (pink). The tentacles (red) are built in the region of the highest source density that is not occupied by the head signal. After amputation, the signals are regenerated with the correct polarity. The figure on the right depicts the process of head regeneration and tentacle formation as visualized by TS-19 monoclonal antibody (Bode et al., 1988). The signal appears at the most apical part of the tip and shifts to the tentacles at later stages. The figure was adapted from Meinhardt (1993, 2009)*

1.2. Polycomb Group proteins and epigenetics

1.2.1. What are epigenetics?

The term 'epigenetics' originates from the theory of 'epigenesis', which was described by Aristotle in the book "On the Generation of Animals". The term describes the development of an organism through a sequence of unfolding steps in which new structures and organs will differentiate which have not been existing in the original egg or spore. This definition of 'epigenesis' was used by developmental biologists and embryologists in the 17th century and was opposed by the theory of 'preformationism'. The preformationists held the view that the later form of an organism is already existing in the egg.
A more recent concept of epigenetics was first characterized by Conrad Hal Waddington to describe the entire developmental process from the fertilized egg to the mature organism as an interplay between genes and their surroundings to produce a phenotype. He coined the term 'epigenetics' to link the field of developmental biology and genetics.
"But the difference between an eye and a nose, for instance, is clearly neither genotypic nor phenotypic. It is due, as we have seen, to the different sets of developmental processes which have occurred in the two masses of tissue; and these again can be traced back to local interactions between the various genes of the genotype and the already differentiated regions of the cytoplasm in the egg. One might say that the set of organizers and organizing relations to which a certain piece of tissue will be subject during development make up its "epigenetic constitution" or "epigenotype"; [...]" (Waddington, 1939).
However, this concept of 'epigenetics' is still very broad and differs substantially from today´s acception of the word. In 1996, Riggs and Porter described epigenetics as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence" (Riggs and Porter, 1996). This definition can be seen as a working definition of epigenetics. Many apparently unexplainable phenomena contributed to this definition over time, like position effect variegation in *Drosophila* (Müller, 1930), transposable elements in maize (McClintock, 1950), X-chromosome inactivation (Lyon, 1961), genomic imprinting (Cattanach and Kirk, 1985) or DNA methylation (Holliday and Pugh, 1975).

1.2.2. Epigenetic modifications

Several epigenetic inheritance systems exist, including chromatin remodeling, RNA transcripts (e.g. RNAi), prions and cellular structures like ciliary rows in paramecium. Chromatin remodeling and epigenetic modifications on the chromatin level are further described in this section. Epigenetic modifications can be found on different levels of chromatin organization. The genetic information (DNA) in higher eukaryotes is about two meters long and needs to fit into a nucleus of about 10 micrometers. The secret lies in packaging: The DNA is wrapped around protein complexes, the histones, and is thus

maximally condensed (Kornberg, 1974; Luger et al., 1997). 147 base pairs of DNA are wrapped around a histone octamer, composed of two histone H2A/H2B dimers and one H3/H4 tetramer. This basic unit of chromatin is called nucleosome. Repeating units of nucleosomes constitute the so-called 10 nm-fiber, which is further compacted in the presence of the linker histone H1 and other scaffolding proteins. However, chromatin has not only packaging functions, but it plays an important role in gene regulation. Within the cell, chromatin is existent in two different states. Whereas actively transcribed genes can be found in open regions of euchromatin, the less active or repressed genes are at regions of more densly packed heterochromatin (Weintraub and Groudine, 1976). Mechanisms regulating the accessibility of the DNA template by stably altering higher-order chromatin structure have to exist. The chromatin structure is highly dynamic and modifications can be found on different levels of chromatin organization (figure 1.13):

- DNA modification (methylation of cytosine residues)
- histone modifications (methylation, acetylation, phosphorylation, ubiquitination)
- histone variants
- proteins that bind to histone or DNA modifications.

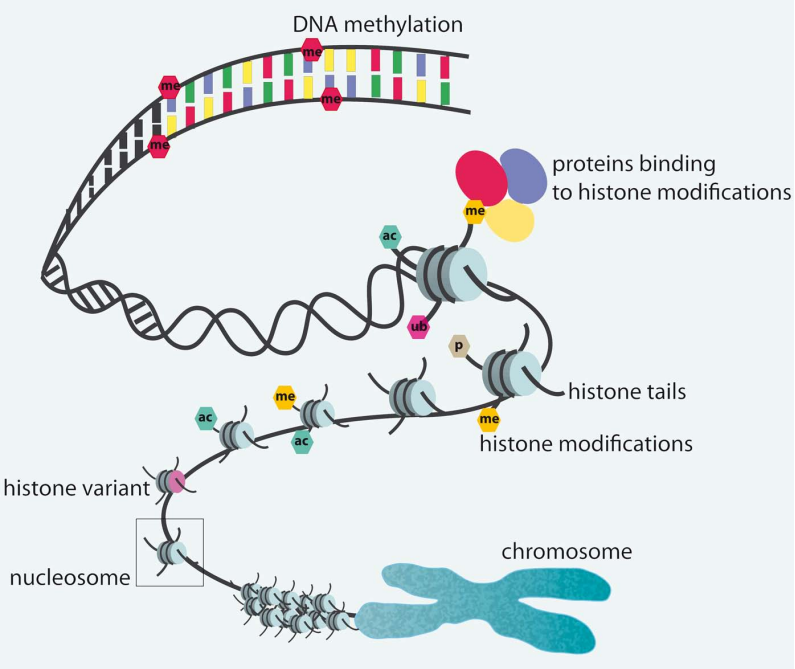


FIGURE 1.13 *Overview of chromatin modifications involved in epigenetic mechanisms.*

Modifications of the histone tails by chromatin modifying enzymes are of special interest in the context of epigenetic inheritance and gene regulation. In 1950, Stedman and Stedman proposed that histones could play a role in gene regulation. Allfrey et al. (1964) were the first ones to observe the relationship between histone modifications and gene activation or repression. The importance of histone modifications and histone modifying enzymes in gene regulation began to reveal. The first modification which was linked to gene activation was the acetylation of histones (Brownell et al., 1996). In the same

year, the first eukaryotic histone deacetylase (Rpd3) was discovered in yeast. In 1999, histone methyltransferases and their role in transcriptional regulation were described (Chen et al., 1999). All these discoveries led to the 'histone code' hypothesis of Jenuwein and Allis. They proposed that modifications of histone tails form a kind of language or cross-talk, which is read by other proteins facilitating distinct events like transcriptional activation or repression of target genes (figure 1.14).
One of the repressive histone marks is the trimethylation of histone H3 at lysine 27, which is set by the Polycomb Group (PcG) proteins. These proteins are introduced further in the next sections.

FIGURE 1.14 *The histone code: sites of histone tail modifications (adapted from Allis et al. (2006)). In the left corner a 2.5 Å structural model of a nucleosome is depicted. 147 base pairs of DNA (gray) are wrapped around a histone octamer composed of two copies of histones H2A (yellow), H2B (red), H3 (blue) and H4 (green), respectively. The picture was generated using a molecular graphics visualisation tool based on RasMol (by Roger Sayle). The underlying nucleosome structure was retrieved from the Protein Data Bank entry 3AFA (www.pdb.org/). The histone tails of the core histones with known sites of covalent modifications are illustrated. Active marks include acetylation, arginine methylation and some lysine methylations. Repressed marks include methylations on H3K9, H3K27 and H4K20.*

1.2.3. The identification of Polycomb Group (PcG) and Trithorax Group (TrxG) proteins

In a developing organism some genes have to be active whereas other genes have to be inactive to assure proper development. In multicellular organisms, different cell types have to perform varying functions, despite holding the same genetic information encoded in their DNA. The difference lies in the variable gene expression patterns. The particular challenge exists in maintaining the gene expression patterns during cell division after the initial stimulus has disappeared. A cellular or transcriptional memory system which

ensures the long-term maintenance of transcriptional states has to exist (Paro and Harte, 1996). The PcG and TrxG proteins have been first identified in the context of long-term maintenance of *HOX* (homeobox) gene expression in *Drosophila*. In metazoans, *HOX* gene expression defines the anterior-posterior axis of the animal and PcG mutants show severe homeotic transformations. In figure 1.15 mutant phenotypes caused by homeotic transformations are depicted.

FIGURE 1.15 *Examples for homeotic transformations in PcG mutants. Left: Homeotic transformations in a Drosophila Pc-mutant. A wild-type Drosophila larva with the correct sequence of segments is shown on the left. On the right site, a Pc-mutant larva is depicted. The segments are transformed and the larva shows severe brain defects (Müller et al., 1995). Middle: The axial skeleton of newborn wild-type (left) and homozygous bmi1 mutant (right) mice are shown. The mutant mouse shows defects in the sacral vertebra. The first sacral vertebra is transformed to the sixth lumbar vertebra (Alkema et al., 1995). Right: A wild-type (top) and a clf-2 mutant flower (bottom) of Arabidopsis. In the mutant flower, the petals are absent (Goodrich et al., 1997).*

In 1978 Lewis identified the *Polycomb* (*Pc*) gene as a negative trans-regulator of the *Drosophila* bithorax complex. Heterozygous and homozygous mutants were caused by ectopic expression of *HOX* genes. Some years later, in 1985, the *Trithorax* (*Trx*) gene has been identified as a trans-regulator of the *Drosophila* bithorax and Antennapedia complex (Ingham, 1985). Genes of the Trithorax Group have been shown to counteract the Polycomb Group genes (Kennison and Tamkun, 1988). PcG proteins propagate repressed chromatin states, whereas TrxG proteins propagate the active states. The first evidence for the direct interaction between Pc and homeotic loci was the binding pattern of Pc on *Drosophila* polytene chromosomes (Zink and Paro, 1989). Due to similarities to the *Drosophila* heterochromatin-associated protein HP1, it was hypothesized that the mechanism of action of the PcG proteins is an epigenetic mechanism by nature (Paro and Hogness, 1991). It has been discovered that PcG and TrxG proteins act in multimeric protein complexes to regulate higher order chromatin structure (Franke et al., 1992). The first mammalian PcG gene which has been discovered was the oncogene Bmi-1. It could be shown that Bmi-1 is a member of a vertebrate PcG complex and is involved in *HOX* gene repression as well (Brunk et al., 1991; Bunker and Kingston, 1994; Alkema et al., 1995). By substituting the *Drosophila* Pc protein for its mammalian counterpart, it could be demonstrated that the protein is functionally conserved (Müller et al., 1995). In 1997,

the first PcG member was identified in plants: CURLY LEAF (Goodrich et al., 1997).
In the following years, the PcG and TrxG proteins have been linked to various epigenetic phenomena like cosuppression, X-inactivation, genomic imprinting and RNAi (Pal-Bhadra et al., 1997; Wang et al., 2001; Mager et al., 2003; Plath et al., 2003; Silva et al., 2003; Grimaud et al., 2006). Furthermore the interaction of PcG proteins with general transcription factors and histone modifications has been shown (Cavalli and Paro, 1999; Breiling et al., 2001; Saurin et al., 2001).

1.2.4. The Polycomb Group response elements (PREs)

In *Drosophila* many of the genes regulated by PcG proteins are controlled by a so-called Polycomb response element (PRE) or cellular memory module (CMM). PREs are cis-regulatory elements that direct the genes to be active or repressed; a switch between the ON and OFF state can occur (reviewed in Ringrose and Paro (2001, 2004, 2007)). Besides genes involved in cell cycle regulation and senescence, developmental regulators, among which the homeotic genes can be found - have been shown to be regulated by PcG and TrxG proteins and controlled by PREs (Mihaly et al., 1998). Not only the homeotic genes are controlled by PREs, but also their initial transcriptional regulators, the segmentation genes (gap, pair rule and segment polarity genes).
PREs are switchable elements that can be turned from a silent to an active state during development. The new state can be maintained over several rounds of cell division. Busturia et al. (1997) showed that the PRE itself is directly involved in keeping the silent state during cell cycle. Removal of the PRE caused a derepression of its controlled gene. To further identify and study the nature of Polycomb response elements, PRE maintenance assays based on transgene reporter constructs have been performed in many classical studies. In figure 1.16 the assembly of such a transgene assay in *Drosophila* is depicted schematically.

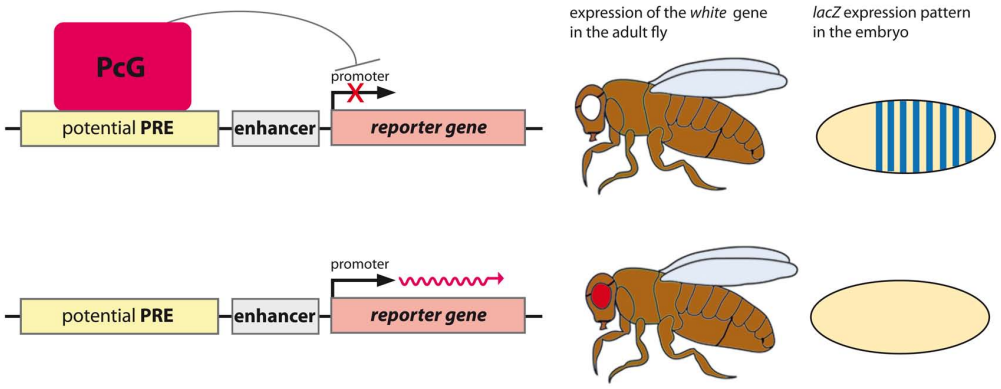


FIGURE 1.16 *PRE maintenance assay in Drosophila based on a transgenic reporter construct. A potential PRE fragment is flanked by an enhancer and a reporter gene. As reporter, lacZ or the Drosophila mini-white gene may be used. The transgene construct can be tronsformed into Drosophila by P-element transformation. In case the used PRE fragment functions as PRE, the PcG complex can bind and lacZ or mini-white are silenced. The resulting phenotype shows white eyes and no lacZ expression pattern, respectively. In case the potential PRE fragment shows no PRE function, the reporter genes are activated and expressed. The fly has wild-type red eyes or shows a lacZ expression pattern dependent on the enhancer used.*

It could be demonstrated that the initial configuration of a PRE is not necessarily persistent. PRE-directed gene silencing can be relieved and the resulting active state persists during cell division (Cavalli and Paro, 1998, 1999). By using a transgene carrying the Fab7-PRE from the *Drosophila* Bithorax complex flanked by a *lacZ* reporter gene driven by the GAL4 transcriptional activator, it could be shown that a burst of GAL4 transcription can switch the PRE from a silent configuration to an active one. Whether this active state is transient or persistent depends on the time point during development at which the transcriptional burst occurs. At least, for the transient activation, a memory system has to exist, allowing the PRE to 'remember' its initial state and to become inactivated again after several rounds of cell division.
PREs do not act alone but in groups. For instance it has been shown for the homeotic genes that one gene is controlled by two or more PREs (Simon et al., 1993). To identify the minimal required PRE, a reporter construct containing a PRE fragment flanking a tissue-specific enhancer for the *Ubx* gene followed by the Ubx promoter and a *LacZ* reporter gene has been used (Horard et al., 2000). In case the fragment has PRE activity, it should maintain the correct pattern of lacZ expression. Yet, a universal minimal PRE has not been identified up to now.
It has been shown that PcG proteins need to interact with the PRE to maintain its inactive configuration (Beuchle et al., 2001; Poux et al., 2001a). After knocking-out the PcG genes, a dramatic loss of silencing occurs.

1.2.5. The Polycomb Repressive Complexes (PRCs)

It has been suggested that the PcG proteins act in different complexes (Beuchle et al., 2001). There has to exist one complex forming the memory mark for maintaining silence during cell division and another, true repression complex which is responsible for gene silencing. These complexes have been termed Polycomb repressive complexes (PRCs). In *Drosophila* three major Polycomb Repressive Complexes (PRCs) have been described: PRC1, PRC2 and PhoRC. There is evidence that the composition of these complexes is dynamic and is dependent on tissue and target genes (Strutt and Paro, 1997; Orlando et al., 1998; Satijn and Otte, 1999). For most of the members of the complexes, mammalian homologues have been identified (Levine et al., 2002). For an overview of the members of the repressive complexes see table 1.2.

PRC2

The *Drosophila* PRC2 complex is a 600 kDa complex composed of the proteins E(z), Esc, Su(z)12 and the histone-binding protein p55. Tie et al. (2001) reported that the histone deacetylase Rpd3 co-fractionates with the PRC2 members. The PRC2 complex methylates histone H3 via the SET domain of its catalytic subunit E(Z) (Cao et al., 2002; Czermin et al., 2002; Müller et al., 2002; Kuzmichev et al., 2002). Proteins with SET domains

TABLE 1.2 *Drosophila PcG complexes and associated factors.*

***Drosophila* PcG gene**		**functional domain**	***Mus musculus* homologue**
Polycomb Repressive Complex 1 (PRC1)			
Pc	Polycomb	chromodomain	Cbx2 / M33 Cbx4 / MPc2 Cbx6 Cbx7 Cbx8 / MPc3
Ph-p, Ph-d	Polyhomeotic	zinc finger, SAM/SPM domain	Edr1 / Mph1 / Rae28 Edr2 / Mph2
Psc	Posterior Sex Combs	zinc finger, HTH domain	Bmi1 Rnf110 / Zfp144 / Mel18
dRing / Sce	dRing / Sex combs extra	RING zinc finger	Ring1 / Ring1a Rnf2 / Ring1b
Scm	Sex Combs on Mid-leg	SAM domain, MBT domain, zinc finger	Scmh1
Polycomb Repressive Complex 2 (PRC2)			
E(z)	Enhancer of zeste	SET domain	Ezh2 / Enx2 Ezh1 / Enx1
Esc **Escl**	Extra sex combs	WD40 repeats	Eed
Su(z)12	Suppressor of zeste 12	zinc finger, VEFS box	mSU(Z)12
p55 / Nurf-55 (Caf1)	p55	histone-binding domain	RbAp48 RbAp46
Pleiohomeotic Repressive Complex (PhoRC)			
Pho	Pleiohomeotic	zinc finger	Yy1
Phol	Pleiohomeotic-like	zinc finger	Yy2
dSfmbt	Scm-related gene containing four mbt domains	MBT domain, SAM domain	L3mbtl2
PcG recruiters			
Dsp1	Dorsal switch protein	HMG-box	Hmgb1 | 1
Grh	Grainyhead	CP2 domain	Grhl2
Gaf / Trl	Gaga factor / Trithorax-like	BTB domain, zinc finger	Zbtb7b
Lolal	Lola-like	BTB domain	?
Psq	Pipsqueak	HTH-psq domain	?
Zeste	Zeste	?	?
PRC2 associated factors			
Pcl	Polycomb-like	PHD zinc finger	Phf1
Rpd3	Rpd3	Histone deacetylase domain	Hdac2
Sir2	Sirtuin	SIR2 domain	Sirt1

have been discovered to possess a histone lysine methyltransferase (HKMT) activity (Rea et al., 2000). The PRC2 complex has been shown to methylate histone H3 at two lysine residues via E(z): at lysine 27 and to a lesser extent at lysine 9 (Cao et al., 2002; Czermin et al., 2002; Müller et al., 2002; Kuzmichev et al., 2002). Both methylation marks are associated with transcriptional silencing. However, E(z) cannot act on its own. The minimal PRC2 complex displaying HKMT activity is composed of E(z), Su(z)12 and Esc. In addition to its HKMT activity, the PRC2 complex could also possess other chromatin-modifying activities. In some variants, the histone deacetylase Rpd3/HDAC2 has been found (Tie et al., 2001).
Additional PRC2 related complexes like PRC3 and PRC4 have been identified. The PRC3 and PRC4 complexes can be distinguished from the PRC2 complex by their Eed isoforms (Kuzmichev et al., 2004, 2005). The PRC3 complex is composed of E(z), Su(z)12, p55 and two shorter Eed isoforms. This complex has been shown to methylate lysine 27 on histone 3 (H3K27), whereas the PRC2 complex with the longest isoform of Eed has been shown to methylate lysine 26 on histone H1 (H1K26).
Interestingly, the PRC2 complex can also be linked to gene silencing dependent on DNA methylation. Mammalian DNA can be modified by methylation at CpG dinucleotides. DNA methylation patterns are heritable epigenetic marks which are associated with gene silencing (for a review see Bird (2002)). For example X-inactivation and genomic imprinting have been linked to DNA methylation. Viré et al. (2006) could demonstrate that the Enhancer of zeste subunit in the context of PRC2/3 complexes can serve as a platform for the recruitment of DNA methyltransferases (DNMTs). It could be shown that E(z) is needed for the DNA methylation of its own target genes. Both histone methylation and DNA methylation could be connected with each other. This mechanism displays another possibility for PcG mediated gene silencing, which is only dependent on PRC2.

PRC1

The PRC1 complex was first purified from *Drosophila* and contains Pc, Psc, Ph-p, Ph-d, Scm and Sce/dRing (Shao et al., 1999). The complex is 1-2 MDa in size. A functional core complex composed of Pc, Psc, Ph and dRing could be reconstituted in vitro and has been shown to block transcription and nucleosome remodelling mediated by the SWI/SNF ATPase complex (Francis et al., 2001). The PRC1 complex can recognize the epigenetic methylation mark set by the PRC2 complex via the chromodomain of PC (Min et al., 2003; Fischle et al., 2003).
A human Polycomb Repressive Complex 1-like (hPRC1L) composed of Ring1, Ring2, Bmi1 and HPH2 has been shown to possess E3 ubiquitin ligase activity which is specific for lysine 119 on histone H2A. The monoubiquitination of H2AK119 could be dramatically decreased by knocking down Ring2 or Bmi1 leading to a derepression of target genes (Wang et al., 2004a; Cao et al., 2005). The E3-ligase activity seems to be critical for PcG-mediated silencing and the establishment of a repressed chromatin state.

General transcription factors like TBP and dTAFIIs have been co-purified with PRC1 (Breiling et al., 2001; Saurin et al., 2001). Apart from altering higher-order chromatin structure, one functional aspect of the PcG proteins is the direct inhibition of the transcriptional apparatus (Dellino et al., 2004).

PhoRC

The exact mechanism of interaction between the PcG proteins and their mode of action is still unclear. But it has been shown that PC (PRC1) and ESC (PRC2) need to interact transiently with each other to establish gene silencing (Poux et al., 2001b). Whereas the PRC2 complex is responsible for setting up the histone mark for a silent transcriptional state, the PRC1 complex is responsible for the establishment of the silent state. However, none of the members of the PRC1 and PRC2 complexes has been shown to bind to DNA in a sequence-specific manner. Consequently the question arises: How do PcG proteins find their target genes?

Several candidate proteins with a sequence-specific DNA binding activity have been linked to the targeting of the PcG complexes: Pleiohomeotic (Pho), Trithorax-like (Trl), Dorsal switch protein (Dsp1), Pipsqueak (Psq) or Zeste. Pleiohomeotic and the redundant Pleiohomeotic-like are the most promising candidates. Pho is the *Drosophila* homologue of the mammalian transcription factor Yin Yang-1 (YY1) and belongs to the PcG genes. Pho has been shown to interact with a number of PREs, like the *Drosophila engrailed* and *Ultrabithorax* PRE (Brown et al., 1998; Fritsch et al., 1999; Brown et al., 2003). It has been suggested that Trl or Dsp1 assist Pho in binding to PRE sites (Mahmoudi et al., 2003; Déjardin et al., 2005) Additionally, it could be demonstrated that the C-terminal binding protein (CtBP) plays a role in controlling the binding of Pho/YY1 and the recruitment of PcG proteins to PREs (Srinivasan and Atchison, 2004; Basu and Atchison, 2010).

Even though identified as a PcG protein, Pho could neither be found in the PRC1 nor the PRC2 complex. In lieu thereof, Pho could be detected in assemblies containing a dINO80 complex or dSfmbt. Following the terminology of the other repressive complexes, the Pho-dSfmbt complex was termed Pleiohomeotic Repressive Complex (PhoRC) (Klymenko et al., 2006). Besides PRC1 and PRC2, the PhoRC complex can be seen as the third major PcG complex with the following subunits: Pho, Phol and dSfmbt. Wang et al. (2004b) reported that Pho can directly interact with the PRC2 complex via E(z) and Esc and that Pho is required for PRE binding by E(z). Yet, in this study neither the PRC2 complex nor the PhoRC complex could recruit PRC1.

A hierarchical model of PcG mediated gene silencing can be set up: In a first step, the PhoRC complex binds to a PcG target site and recruits the PRC2 complex which methylates histone H3 on lysine 27. In a second step, the methylation mark is recognized by the PRC1 complex, which is recruited. Figure 1.17 illustrates the mechanism of action.

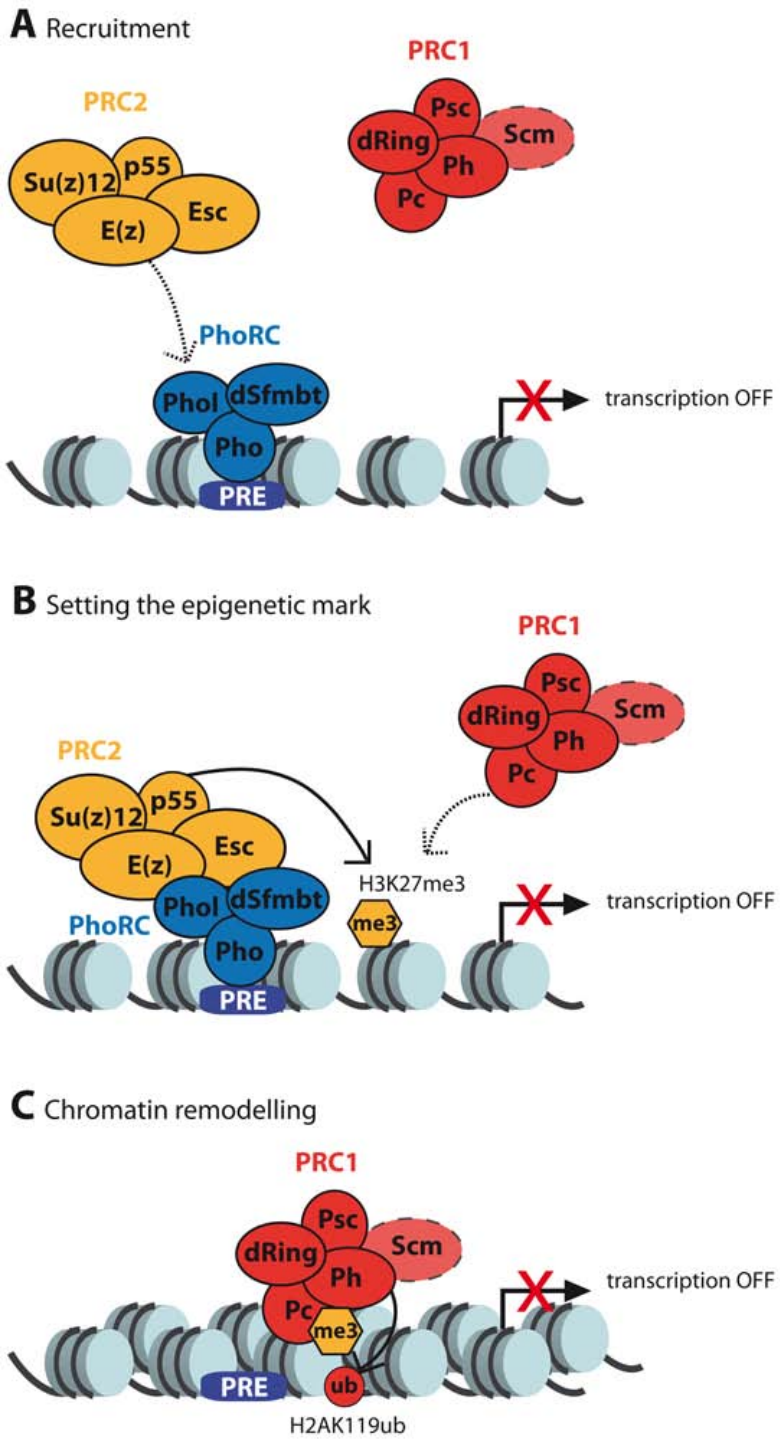


FIGURE 1.17 *Cartoon illustration the PcG dependent silencing mechanism in Drosophila.* ***A*** *To maintain the inactive transcriptional state, the PcG complexes have to inactivate the target gene PRE site. In a first step, the PhoRC complex with its subunits Pho, Phol and dSfmbt binds to the PRE site.* ***B*** *Through the interaction with PhoRC, the PRC2 complex composed of E(z), Esc, Su(z)12 and p55 is recruited to the target gene. The PRC2 complex establishes the H3K27me3 epigenetic mark via the histone methyltransferase E(z). In turn, the histone mark is recognized by the PRC1 complex. The PRC1 complex consists of Pc, Ph, Psc and Sce/dRing. Pc can interact with H3K27me3 via its chromodomain and thereby recruit the complex.* ***C*** *The PRC1 complex can establish transcriptional silencing of the target gene by ubiquitination of H2AK119 followed by chromatin remodeling.*

1.2.6. Transmission of the H3K27me3 epigenetic mark through the cell cycle

A big question concerns the transmission of the H3K27 epigenetic mark through the cell cycle. The silent chromatin state needs to persist during cell division for an effective long-term repression and the maintenance of cell identities. It has been proposed that the PRC2 complex can recognize and bind to its own methylation mark (H3K27me3) and copies this mark to neighbouring histones during DNA replication (Hansen et al., 2008; Hansen and Helin, 2009; Margueron et al., 2009). Another study suggested that the Polycomb core complexes are retained on their DNA templates during passage of the replication fork and that the H3K27me3 mark is not essential (Francis et al., 2009); the PRC1 complex remains continuously bound. Hence, the transmission of the epigenetic mark during cell cycle seems to be dependent on the Polycomb repressive complexes themselves.
Another question concerns the removal of the H3K27me3 epigenetic mark. Since proliferation and development are highly dynamic processes and genes are differentially expressed, a possibility to reset the H3K27me3 epigenetic mark has to exist. For a long time, no possibility for removing the epigenetic mark has been known. The mark seemed to be irreversibly bound. In 2007 a H3K27me3 demethylase has been discovered (Agger et al., 2007; Santa et al., 2007; Hong et al., 2007; Lan et al., 2007; Lee et al., 2007). The JmjC-domain-containing proteins UTX and JMJD3 have been shown to catalyze the demethylation of H3K27me3/2.

1.2.7. PcG targeting mechanisms in *Drosophila* and mammals

PRE-dependent PcG protein targeting in *Drosophila*

In *Drosophila*, PcG proteins are recruited to their target genes in a PRE-dependent manner. The PRE site is recognized by the PhoRC complex, which recruits the other repressive comolexes.
To predict and identify PcG target genes in *Drosophila*, the PRE elements were identified. About 100 PREs have been assumed in the *Drosophila* genome. The search has been exceedingly difficult in the beginning due to the lack of a PRE consensus sequence. The identification of PREs and PcG target genes was dependend on functional assays like transgene analysis or chromatin immunoprecipitation. With these classical approaches, the PREs at the following loci have been found in *Drosophila*: homeotic genes at the bithorax (BX-C) and Antennapedia complexes (ANT-C), the *engrailed*, *polyhomeotic* and *hedgehog* genes (Simon et al., 1993; Strutt and Paro, 1997; Brown et al., 1998; Americo et al., 2002; Maurange and Paro, 2002; Bloyer et al., 2003; DeVido et al., 2008). However, stainings on *Drosophila* polytene chromosomes revealed over 100 sites (Zink and Paro, 1989; Rastelli et al., 1993). In 2003 Ringrose et al. defined sequence criteria to distinguish PRE sites from non-PRE sites based on previously identified PREs. Binding motifs for GAF/Trl, Zeste and Pho have been shown to be required for the function of several PREs. A genome-

wide prediction identified 167 PRE sites in *Drosophila* which have been further analyzed concerning their functionality. PREs have been found in the vicinity of genes involved in development and cell proliferation like *hunchback*, *knirps*, *giant*, *even skipped*, *engrailed*, *hedgehog*, *armadillo*, *wingless*, *eyes absent*, *proliferation disrupter* or *lethal(2) giant larvae*. Five significant PRE sequence motifs have been identified in that study (table 1.3):

TABLE 1.3 *Left: DNA binding proteins which are known to play a role in PRE binding and their consensus motif. Right: motifs which have been found to be significantly enriched at PRE sites and DNA binding proteins known to interact with these sequences.*

DNA binding protein	**Sequence motif**
Pleiohomeotic/ Yin Yang-1	GCCAT
Trithorax-like/ GAGA factor	GAGAG
Zeste	(C/T)GAG(C/T)GAG

PRE motif	**DNA binding protein**
AGAGAGAGAGAG	Trxl/GAF
GGCCGCCATTTT	Pho/YY1
GTGTGTGTGTGT	?
TTTTTTTTGCTT	hunchback?
TGCTGCTGCTGC	?

Initially, PREs have been discovered in *Drosophila* (Simon et al., 1993). PREs in other species remained elusive for a long time. The first vertebrate PREs have been identified recently. Sing et al. (2009) identified a vertebrate PRE which regulates the *MafB/Kreisler* segmentation gene in mice. Inversion of *kreisler* (*kr*) shifts the MafB expression in the embryonic hindbrain anteriorly. It is postulated that the *kr* inversion antheriorizes MafB expression by affecting a PRE site. In M33 mutant mice the MafB expression is affected. By examining the *kr* inversion breakpoint, a conserved region, several Pho/YY1 binding sites and GAGAG motifs could be identified. Even a palindromic double Pho-binding site as in conserved *Drosophila* core elements has been found. The PRE was termed PRE*kr*. Functional transgene assays with PRE*kr* in *Drosophila* and mice revealed that the identified PRE can repress gene expression of a reporter gene. Knocking-down *Pho* or PRC1 components derepressed the reporter gene. A second mammalian PRE was identified by Woo et al. (2010). A region between HOXD11 and HOXD12 (D11.12) which is associated with PcG proteins has been identified. The D11.12 element has been shown to repress luciferase expression in a reporter construct. For a full repression of the reporter gene, a cluster of YY1 binding sites and a highly conserved region have been necessary. The conserved region seems to be required for the recruitment and/or stability of the Polycomb repressive complexes. After the knock-down of the PcG genes, the silencing function was lost. However, only isolated vertebrate PREs have been identified so far. In 2010 Liu et al. have followed a computational approach to predict PcG target genes in mouse embryonic stem cells. Surprisingly the YY1 motif could not be found in murine PREs at this approach. There is growing evidence for additional mechanisms in PcG recruitment in mammals.

PcG protein targeting by long non-coding RNAs in mammals

Mammalian PcG proteins are associated with X-inactivation. In female mammals, one of the two X chromosomes within a cell gets inactivated. Dosage compensation involves the silencing of almost all genes on the inactivated X chromosome (Xi). It has been shown that X inactivation is regulated by a locus called XIC (X inactivation center). This locus produces a long non-coding RNA molecule which accumulates at the future Xi. The *Xist* (X inactive specific transcript) RNA initiates gene silencing by nucleosome remodeling and DNA methylation. X inactivation may occur randomly or in some mammals like rodents, X inactivation has been shown to be imprinted.
It has been shown that the inactivation of one X chromosome in female cells is accompanied by H3K27 methylation mediated by the PRC2 complex (Wang et al., 2001; Plath et al., 2003; Silva et al., 2003). The recruitment of PRC2 to the future Xi appears to be dependent on *Xist* RNA. The association of PRC2 with the inactive X chromosome is transient and only required for establishing the epigenetic mark to maintain the silent state. Interestingly, PRC1-mediated H2A ubiquitination has been found associated with Xi, as well. This indicates that PRC1 may be involved in X inactivation, too (Fang et al., 2004; de Napoles et al., 2004).

PcG proteins are not only involved in X inactivation, but also in the process of genomic imprinting of autosomal genes. There are about 60 genes in mammals which have been shown to be paternally or maternally repressed. Mager et al. (2003) could show that in homozygous Eed mutant mouse embryos transcripts from paternally repressed genes are present. Changes in the DNA methylation state of the differentially methylated regions (DMR) on affected genes could be detected. The following paternally repressed genes got activated in the PcG mutant mice: *Cdkn1c*, *Ascl2*, *Grb10* and *Meg3*. Hence PcG silencing mechanisms are required for the regulation of autosomal imprinted loci.
Several studies showed that PcG mediated gene silencing of imprinted genes is dependent on the interaction of PRC2 with non-coding RNAs. Umlauf et al. (2004) and Terranova et al. (2008) revealed that the PRC2 complex is involved in silencing the paternally imprinted *Kcnq1* locus via its interaction with the *Kcnq1ot1* non-coding RNA. Rinn et al. (2007) demonstrated that an interaction between PRC2 and the *HOTAIR* non-coding RNA is needed for gene silencing. *HOTAIR* is a non-coding RNA in the HOXC locus and represses transcription along the HOXD locus in *trans*. The silencing mechanism of dosage compensation and genomic imprinting by PcG proteins in mammals can be compared with the mechanism of *HOX* gene repression in *Drosophila*.

2. Goals of this study

The present study was designed to identify and characterize PcG and associated proteins in the cnidarian *Hydra*. Special emphasis was placed on the PRC2 complex and the identification of DNA binding proteins which are important for targeting of the PcG complexes. Very little was found in the literature on PcG proteins in this 'basal' metazoan animal. In a phylogenetic study of Schuettengruber et al. (2007) the existence of an almost complete set of PcG proteins has been reported. However, the findings of this study are only based on *in silico* data. In a study by Genikhovich et al. (2006), two genes of the PRC2 complex from *Hydra vulgaris AEP* have been cloned and further characterized.
In the 'classical' *Drosophila* model, the PcG proteins are clustered into multimeric complexes which are recruited to their target genes in a hierarchical manner. The PRC2 core complex consisting of the proteins E(z), Esc, Su(z)12 and p55 is recruited to the target gene by the PhoRC complex consisting of the DNA binding protein Pho and dSfmbt. Surrounding nucleosomes are trimethylated at histone H3K27. This epigenetic mark is recognized by the PRC1 core complex consisting of Pc, Ph, Psc and dRing. Histone H2A gets ubiquitinated at K119 followed by chromatin remodeling and transcriptional silencing. However, recent studies on PcG proteins in mammals reveal a more complex situation. First, there is emerging evidence for various recruiting mechanisms of the PcG complexes to their target genes, in which long non-coding RNAs seem to play a role. Second, it becomes apparent that the PcG complexes can be assembled combinatorially depending on the target gene. Third, the actual process of gene silencing triggered by PcG proteins is dependent on downstream events like DNA methylation.
PcG and associated genes were analyzed in *Hydra* to gain inside into PcG based mechanisms in lower animals, which could help to better understand the more complex situation in higher animals. The goals of this study can be summarized in the following items:

1. The identification of PcG and associated genes in the *Hydra* genome, cloning of the candidate genes and identification of conserved functional domains.

2. The identification of the expression patterns of PcG genes in the freshwater polyp to get an idea of the processes, which might be regulated by PcG proteins: Common processes like stem cell regulation and embryogenesis or *Hydra*-specific pattern formation processes like budding and regeneration.

3. The identification of molecules, which are important for PcG recruitment and which could help to identify target genes in *Hydra*.

3. Materials and Methods

3.1. Materials

3.1.1. Chemicals and solutions

chemical	formula	MW (g/mol)	company
acetic anhydride	$C_4H_6O_3$	102.9	Riedel-de-Haen; Seelze (Germany)
acrylamide/bisacrylamide (Rotiphorese Gel 40 (29:1))	-	-	Roth; Karlsruhe (Germany)
agar	-	-	AppliChem; Darmstadt (Germany)
agarose	-	-	Roth; Karlsruhe (Germany)
ammonium acetate	$C_2H_3O_2NH_4$	77.08	VWR International; West Chester, Pennsylvania (USA)
ammonium persulfate	$(NH_4)_2S_2O_8$	228.2	Serva; Heidelberg (Germany)
ampicillin	$C_{16}H_{18}N_3NaO_4S$	371.39	Roth; Karlsruhe (Germany)
aprotinin	$C_284H_432N_84O_79S_7$	6511.51	AppliChem; Darmstadt (Germany)
β-mercaptoethanol	C_2H_6OS	78.13	Roth; Karlsruhe (Germany)
BCIP	$C_8H_6NO_4BrCIP{\cdot}C_7H_9N$	433.6	Roche Diagnostics GmbH; Mannheim (Germany)
benzamidine hydrochloride	$C_7H_8N_2 \cdot HCl$	156.62	AppliChem; Darmstadt (Germany)
BMPurple	-	-	Roche Diagnostics GmbH; Mannheim (Germany)
boric acid	H_3BO_3	61.83	Merck; Darmstadt (Germany)
bovine serum albumine, fraction V	-	-	Sigma-Aldrich; St. Louis, Missouri (USA)
bromophenol blue	$C_{19}H_{10}Br_4O_5S$	692.0	Serva; Heidelberg (Germany)
calcium chloride dihydrate	$CaCl_2 \cdot H_2O$	147.02	Merck; Darmstadt, Germany
CHAPS	$C_{32}H_{58}N_2O_7S$	614.88	Sigma-Aldrich; St. Louis, Missouri (USA)

chemical	formula	MW (g/mol)	company
chloramphenicol	$C_11H_12Cl_2N_2O_5$	323.1	Serva; Heidelberg (Germany)
chrome alum	$CrK(SO_4)_2 \cdot 12\ H_2O$	283.24	Roth; Karlsruhe (Germany)
Coomassie Brilliant Blue G 250	$C_{47}H48N_3NaO_7S_2$	877.0	Serva; Heidelberg (Germany)
dimethylformamide	C_3H_7NO	73.09	Serva; Heidelberg (Germany)
disodium EDTA	$C_{10}H_{14}O_8N_2Na_2 \cdot 2\ H_2O$	372.24	AppliChem; Darmstadt (Germany)
disodium hydrogen phosphate	Na_2HPO_4	141.96	AppliChem; Darmstadt (Germany)
DMP	$C_9H_{18}N_2O_2$	186.25	Sigma-Aldrich; St. Louis, Missouri (USA)
DMSO	C_2H_6OS	78.13	Serva; Heidelberg (Germany)
DTT	$C_4H_{10}O_2S_2$	154.3	Serva; Heidelberg (Germany)
ethanol	C_2H_6O	46.07	Sigma-Aldrich; St. Louis, Missouri (USA)
ethanolamine	C_2H_7NO	61.08	Roth; Karlsruhe (Germany)
ethidium bromide	$C_{21}H_{20}BrN_3$	394.29	Serva; Heidelberg (Germany)
ethyl carbamate	$C_3H_7NO_2$	89.09	Merck; Darmstadt (Germany)
ficoll	-	$\sim$ 400000	Fluka Biochemika; Buchs (Switzerland)
formaldehyde	CH_2O	30.03	Merck; Darmstadt (Germany)
formamide	CH_3NO	45.04	Sigma-Aldrich; St. Louis, Missouri (USA)
gelatin	-	-	Merck; Darmstadt (Germany)
gentamycin sulfate	$C_{19-21}H_{39-43}N_5O_7 \cdot 2.5\ H_2SO_4$	694.75 - 723.75	AppliChem; Darmstadt (Germany)
glacial acetic acid	$C_2H_4O_2$	60.05	Merck; Darmstadt (Germany)
glucose monohydrate	$C_6H_12O_6 \cdot H_2O$	198.17	Merck; Darmstadt (Germany)
glycerol	$C_3H_5(OH)_3$	92.09	AppliChem; Darmstadt (Germany)
glycine	$C_2H_5NO_22$	75.07	AppliChem; Darmstadt (Germany)
guanidin hydrochloride	CH_6ClN_3	95.53	AppliChem; Darmstadt (Germany)

chemical	formula	MW (g/mol)	company
guanidinium thiocyanate	$C_2H_6N_4S$	118.16	Merck-Schuchardt; Hohenbrunn (Germany)
Heparin sodium salt	-	-	Roth; Karlsruhe (Germany)
HEPES	$C_8H_{18}N_2O_4S$	238.3	Roth; Karlsruhe (Germany)
hydrochloric acid	HCl	36.46	VWR International; West Chester, Pennsylvania (USA)
hydrogen peroxide	H_2O_2	34.01	Merck; Darmstadt (Germany)
hydroxyurea	$CH_4N_2O_2$	76.05	Sigma-Aldrich; St. Louis, Missouri (USA)
imidazole	$C_3H_4N_2$	68.08	Merck; Darmstadt (Germany)
IPTG	$C_9H_{18}O_5S$	238.3	Serva; Heidelberg (Germany)
isopropyl alcohol	C_3H_8O	60.1	AppliChem; Darmstadt (Germany)
kanamycin sulfate	$C_{18}H_{36}N_4O_{11} \cdot H_2SO_4$	582.58	AppliChem; Darmstadt (Germany)
leupeptin hemisulfate	$C_20H_38N_6O_4$ $\cdot$ 0.5 H_2SO_4	475.60	AppliChem; Darmstadt (Germany)
levamisol hydrochloride	$C_11H_12N_2S \cdot HCl$	240.75	Sigma-Aldrich; St. Louis, Missouri (USA)
lithium chloride	LiCl	42.39	Roth; Karlsruhe (Germany)
luminol	$C_8H_7N_3O_2$	117.16	Fluka Biochemika; Buchs (Switzerland)
magnesium chloride hexahydrate	$MgCl_2 \cdot 6\ H_2O$	203.3	AppliChem; Darmstadt (Germany)
maleic acid	$C_4H_4O_4$	116.07	Roth; Karlsruhe (Germany)
methanol	CH_4O	32.04	J. T. Baker; Phillipsburg, New Jersey (USA)
milk powder (blotting grade)	-	-	Roth; Karlsruhe (Germany)
Mowiol 4-88	$(\text{-}CH_2CHOH\text{-})_n$	$\sim$ 31000	Roth; Karlsruhe (Germany)
NBT	$C_{40}H_{30}Cl_2N_{10}O_6$	817.7	Roche Diagnostics GmbH; Mannheim (Germany)
neutral red	$C_{15}H_{17}ClN_4$	288.7	Merck; Darmstadt (Germany)
nickel sulfate	$NiSO_4$	154.75	Merck; Darmstadt (Germany)

chemical	formula	MW (g/mol)	company
N-lauroyl-sarcosine	$C_{15}H_{29}NO_3$	271.4	Sigma-Aldrich; St. Louis, Missouri (USA)
Nonidet P40	-	-	AppliChem; Darmstadt (Germany)
paraformaldehyde	$OH(CH_2O)_nH$ (n = 8 - 100)		Sigma-Aldrich; St. Louis, Missouri (USA)
p-coumaric acid	$C_9H_8O_3$	164.16	Sigma-Aldrich; St. Louis, Missouri (USA)
pepstatine A	$C_{34}H_{63}N_5O_9$	685.89	AppliChem; Darmstadt (Germany)
PMSF	$C_7H_7FO_2S$	174.19	AppliChem; Darmstadt (Germany)
polyvinylpyrrolidone	$(C_6H_9NO)_n$	$\sim$ 24000	Roth; Karlsruhe (Germany)
potassium acetate	CH_3CO_2K	98.15	AppliChem; Darmstadt (Germany)
potassium chloride	KCl	74.56	J. T. Baker; Phillipsburg, New Jersey (USA)
pyruvic acid sodium salt	$C_3H_3NaO_3$	110.94	Roth; Karlsruhe (Germany)
sodium acetate trihydrate	$CH_3COONa \cdot 3\ H_2O$	136.08	Merck; Darmstadt (Germany)
sodium azide	NaN_3	65.0	Serva; Heidelberg (Germany)
sodium chloride	NaCl	58.44	VWR International; West Chester, Pennsylvania (USA)
sodium deoxycholate	$C_{24}H_{39}NaO_4$	414.55	Sigma-Aldrich; St. Louis, Missouri (USA)
sodium dihydrogen phosphate dihydrate	$NaH_2PO_4 \cdot 2\ H_2O$	156.01	Roth; Karlsruhe (Germany)
sodium dodecyl sulfate	$NaC_{12}H_{25}SO_4$	288.4	Serva; Heidelberg (Germany)
sodium hydrogen carbonate	$NaHCO_3$	84.01	J. T. Baker; Phillipsburg, New Jersey (USA)
TEMED	$C_6H_{16}N_2$	116.21	Roth; Karlsruhe (Germany)
tetracyclin	$C_{22}H_{24}N_2O_8 \cdot HCl$	480.9	Serva; Heidelberg (Germany)
triethanolamine hydrochloride	$C_6H_{16}ClNO_3$	185.65	Merck; Darmstadt (Germany)
Tris	$C_4H_{11}NO_3$	121.14	Roth; Karlsruhe (Germany)
trisodium citrate dihydrate	$C_6H_5Na_3O_7 \cdot 2\ H_2O$	294.10	GERBU Biochemicals GmbH; Gaiberg (Germany)

chemical	formula	MW (g/mol)	company
Triton X-100	$C_{14}H_{22}O(C_2H_4O)_n$	-	Merck; Darmstadt (Germany)
trypan blue	$C_{34}H_{28}N_6O_{14}S_4$	872.88	AppliChem; Darmstadt (Germany)
trypton	-	-	AppliChem; Darmstadt (Germany)
Tween 20	$C_{58}H_{114}O_{26}$	1227.54	Roth; Karlsruhe (Germany)
urea	CH_4N_2O	60.06	Roth; Karlsruhe (Germany)
X-Gal	$C_{14}H_{15}BrClNO_6$	408.63	AppliChem; Darmstadt (Germany)
xylene cyanol	$C_{25}H_{27}N_2NaO_6S_2$	538.6	Serva; Heidelberg (Germany)
yeast extract	-	-	Roth; Karlsruhe (Germany)
yeast RNA	-	-	Boehringer (Roche Diagnostics); Mannheim (Germany)
zinc chloride	$ZnCl_2$	136.28	Merck; Darmstadt (Germany)

3.1.2. Consumables

consumables	specification	company
12 ml polystyrene tubes	PS- Tube Screw Cap 12 ml	Greiner-Bio-One GmbH; Frickenhausen (Germany)
15 ml polypropylene tubes	tube, 15 ml, PP, conical bottom with blue screw cap	Greiner-Bio-One GmbH; Frickenhausen (Germany)
50 ml polypropylene tubes	tube, 50 ml, PP, conical bottom with blue screw cap	Greiner-Bio-One GmbH; Frickenhausen (Germany)
active ester agarose for coupling proteins	AffiGel 10 gel	BioRad Laboratories GmbH; München (Germany)
autoradiography films	Amersham Hyperfilm ECL (18 x 24 cm, 24 x 30 cm)	GE Healthcare UK Ltd; Little Chalfont, Buckinghamshire (England)
columns for protein purification	Econo-Pac, general purpose plastic column, 1-20 ml	BioRad Laboratories GmbH; München (Germany)
columns for protein purification	Poly-Prep, disposable plastic column, 2 ml	BioRad Laboratories GmbH; München (Germany)
cover glass	cover glass 24 x 32 mm	VWR International; West Chester, Pennsylvania (USA)
cuvettes	cuvettes 10 x 4 x45 mm, polystyrene	Sarstedt AG & Co.; Nümbrecht (Germany)
desalting columns	PD-10 desalting columns (packed with Sephadex G-25)	GE Healthcare UK Ltd; Little Chalfont, Buckinghamshire (England)

consumables	specification	company
dialysis tubing	Dialysis-Membrane type 27; MWCO 12000-16000; pore size 25A	BIOMOL GmbH; Hamburg (Germany)
dialysis tubing	dialysis tubing Membra-Cel; MWCO 10000-14000; type 36/32	Roth; Karlsruhe (Germany)
FBS	Fetal Bovine Serum Standard Quality	PAA Laboratories GmbH; Pasching (Austria)
Fixogum	Fixogum rubber cement	Marabu GmbH & Co.KG; Bietigheim-Bissingen (Germany)
folded filters	folded filters 595 1/2; ∅ 185 mm	Schleicher & Schuell BioScience GmbH; Dassel (Germany)
***Hydra* culture dishes**	polystyrene dishes (Rotilabo-"Gerda"-Frischhalteboxen)	Roth; Karlsruhe (Germany)
medium for insect cells	TC 100 Insect Medium with L-glutamine	Lonza Group Ltd.; Basel (Switzerland)
membrane filter	Minisart, single use syringe filter, sterile, 0.20 μm	Sartorius AG; Göttingen (Germany)
membrane filter	Rotilabo, single use syringe filter, sterile, 0.45 μm	Roth; Karlsruhe (Germany)
metal chelate affinity beads	Ni-NTA agarose beads	Qiagen GmbH; Hilden (Germany)
microscope slides	microscope slides 76 x 26 x 1 mm	Paul Marienfeld GmbH & Co.KG; Lauda-Königshofen (Germany)
pasteur pipettes	pasteur capillary pipettes, short size, 150 mm	WU; Mainz (Germany)
PCR tubes	0.2 ml PCR tubes with attached, flat cap	BioRad Laboratories GmbH; München (Germany)
Pen-Strep	Penicillin/Streptomycin (100x) (Penicillin 10.000 Units/ml; Streptomycin 10 mg/ml)	PAA Laboratories GmbH; Pasching (Austria)
petri dishes	petri dish 94 x 16 mm	Greiner-Bio-One GmbH; Frickenhausen (Germany)
pipette filter tips	micro filter tip 10 μl, filter tip 100 μl, filter tip 1000 μl	Greiner-Bio-One GmbH; Frickenhausen (Germany)
pipette tips	micropipette tip 0.5 - 10 μl, pipette tip 10-200 μl, pipette tip 201-1000 μl	Greiner-Bio-One GmbH; Frickenhausen (Germany)
positive charged nylon membrane	Amersham Hybond-N$^+$	GE Healthcare Europe GmbH; Freiburg (Germany)
protein A-agarose matrix	protein A-agarose	Roche Diagnostics GmbH; Mannheim (Germany)
protein A-sepharose matrix	protein A Sepharose CL-4B	GE Healthcare Europe GmbH; Freiburg (Germany)
protein concentrator	Vivaspin 2, 50000 MWCO PES	Sartorius AG; Göttingen (Germany)
protein concentrator	Vivaspin 15, 10000 MWCO PES	Sartorius AG; Göttingen (Germany)

consumables	specification	company
reaction tubes	SafeSeal MicroTube 2 ml, polypropylene	Sarstedt AG & Co.; Nümbrecht (Germany)
reaction tubes	MicroTube 1.5 ml, polypropylene	Sarstedt AG & Co.; Nümbrecht (Germany)
septa for sequencing sample tubes	Genetic Analyzer septa for 0.5 ml Sample Tubes	Applied Biosystems; Carlsbad, California (USA)
sequencing capillary	310 Genetic Analyzer Capillaries, 5-47 cm x 50 μm capillaries	Applied Biosystems; Carlsbad, California (USA)
sequencing polymer	POP-6, performance optimized polymer	Applied Biosystems; Carlsbad, California (USA)
sequencing running buffer	310 Running Buffer, 10X	Applied Biosystems; Carlsbad, California (USA)
sequencing sample tubes	Genetic Analyzer 0.5 ml Sample Tubes	Applied Biosystems; Carlsbad, California (USA)
syringe	luer slip syringe BD Discardit II: 10 ml, 20 ml	Becton Dickinson GmbH; Heidelberg (Germany)
syringe	luer-lok BD Plastipak: 50 ml	Becton Dickinson GmbH; Heidelberg (Germany)
tissue culture flasks	CELLSTAR cell culture flask with filter top: 250 ml (75 cm^2), 650 ml (175 cm^2)	Greiner-Bio-One GmbH; Frickenhausen (Germany)
tissue culture multiwell plates	CELLSTAR cell culture multiwell plate: 6 well (9.6 cm^2), 24 well (1.9 cm^2)	Greiner-Bio-One GmbH; Frickenhausen (Germany)
transfection reagent	cellfectin	Invitrogen GmbH; Darmstadt (Germany)
ultracentrifuge tubes	Ultra-Clear centrifuge tubes, 14 x 95 mm	Beckman Coulter GmbH; Krefeld (Germany)
Western blotting membrane	PVDF membrane	Millipore GmbH; Schwalbach (Germany)
Whatman paper	Whatman chromatography paper, 46 x 57 cm x 3 mm	Whatman International Ltd.; Maidstone (England)

3.1.3. Technical equipment

technical equipment	specification	company
autoclave	steam-sterilizer type W	Holzner Medizintechnik GmbH; Nußloch (Germany)
autoradiography cassette	Amersham Hypercassette Autoradiography Cassettes (18 x 24 cm, 24 x 30 cm)	GE Healthcare UK Ltd; Little Chalfont, Buckinghamshire (England)
benchtop centrifuge	Heraeus Biofuge pico	Heraeus Instruments (Thermo Fisher Scientific); Waltham, Massachusets (USA)
centrifuge	Beckman GS-6KR (swing-out rotor)	Beckman Coulter GmbH; Krefeld (Germany)
centrifuge for tissue culture	Beckman GS-63 (swing-out rotor)	Beckman Coulter GmbH; Krefeld (Germany)

technical equipment	specification	company
centrifuge	Sorvall RC5C (fixed angle rotors: GS-3, GSA, SA-600 and SS-34)	Sorvall Instruments (Thermo Fisher Scientific); Hamburg (Germany)
developer machine	tabletop developer machine Agfa Curix 60	Agfa-Gevaert N.V.; Mortsel (Belgium)
DNA gel system	in-house production	Institute of Zoology; Heidelberg (Germany)
electroporation apparatus	BioRad GenePulser II	BioRad Laboratories GmbH; München (Germany)
freezer -20 °C	Liebherr Premium	Liebherr-International Deutschland GmbH; Biberach an der Riss (Germany)
freezer -80 °C	Forma -86C ULT Freezer	Thermo Electron Corporation (Thermo Fisher Scientific); Waltham, Massachusets (USA)
gel documentation system	GelJetImager	Intas Science Imaging Instruments GmbH; Göttingen (Germany)
incubation shaker for cultivation of bacteria	Multitron Standard	Infors HT; Bottmingen (Switzerland)
incubator for cultivation of bacteria	incubator type B50	Memmert GmbH + Co. KG ; Schwabach (Germany)
incubator for cultivation of Sf9 insect cells	Heraeus B-5060-EK-CO2	Heraeus; Hanau (Germany)
inverted confocal microscope	Nikon A1R with Nikon Eclipse Ti	Nikon Instruments Europe B.V.; Kingston, Surrey (England)
inverted fluorescence microscope	Nikon Eclipse 80i (camera: Nikon Digital Sight DS-U1)	Nikon Instruments Europe B.V.; Kingston, Surrey (England)
magnetic stirrer	magnetic stirrer type MR1	Heidolph Elektro GmbH & Co. KG; Kelheim (Germany)
mixer	intelli-Mixer	neoLab; Heidelberg (Germany)
PCR machine	GeneAmp PCR system 2400	Perkin Elmer; Rodgau (Germany)
PCR machine	MyCycler	BioRad Laboratories GmbH; München (Germany)
PCR machine	PalmCycler CG1-96	Corbett Research; Cambridge (England)
pH-meter	Digital-pH-Meter 646 (pH-electrode: Mettler Toledo inLab 417)	Knick Elektronische Messgeräte GmbH & Co. KG; Berlin (Germany)
pipettes	2 μl, 20 μl, 200 μl and 1000 μl	Gilson Inc.; Middleton, Wisconsin (USA)
pipetting aid	pipetus-akku	Hirschmann Laborgeräte GmbH & Co. KG; Eberstadt (Germany)
power supply for DNA and protein gels	Gene Power Supply GPS 200/400	Pharmacia (GE Healthcare life sciences); Uppsala (Sweden)

technical equipment	specification	company
power supply for blotting	Power Pac HC	BioRad Laboratories GmbH; München (Germany)
protein blotting cell	Trans-Blot SD Semi-Dry Transfer Cell	BioRad Laboratories GmbH; München (Germany)
protein gel system	PerfectBlue Twin S	peqlab; Erlangen (Germany)
rocker	Duomax 1030	Heidolph Elektro GmbH & Co. KG; Kelheim (Germany)
roller mixer	CAT RM5	CAT Ingenieurbüro M. Zipperer GmbH; Staufen (Germany)
safety cabinet	Heraeus HeraSafe HS	Heraeus; Hanau (Germany)
sequencer	ABI PRISM 310 Genetic Analyzer	Applied Biosystems; Carlsbad, California (USA)
spectrophotometer	Nanodrop ND-1000	peqlab; Erlangen (Germany)
spectrophotometer	SmartSpecPlus Spectrophotometer	BioRad Laboratories GmbH; München (Germany)
thermomixer	Thermomixer comfort	Eppendorf AG; Hamburg (Germany)
ultracentrifuge	L8-55 ultracentrifuge (swing-out rotor: SW 40)	Beckman Coulter GmbH; Krefeld (Germany)
UV cross-linker	UV Stratalinker 1800	Stratagene Corporation; LaJolla, California (USA)
UV transilluminator	chromato-Vue Transilluminator Model TM-20 (220/240 V, 50 Hz, 1.5 Amp)	UVP; San Gabriel, California (USA)
vortexer	vortexer type REAX-1	Heidolph Elektro GmbH & Co. KG; Kelheim (Germany)
waterbath	JULABO type A6/U3	JULABO Labortechnik GmbH; Seelbach (Germany)
watersystem for ultrapure water	MilliQ	Millipore GmbH; Schwalbach (Germany)
watersystem for *Hydra* culture	Elix	Millipore GmbH; Schwalbach (Germany)
white light transilluminator	white light transilluminator (220 V, 50 Hz)	Rex Messinstrumentebau GmbH, Erlangen (Germany)

3.1.4. Primers

All primers for cloning were obtained from Operon. The oligonucleotides used for gel retardation assays are HPLC purified, all other primers are salt-free. Primer stock solutions of 100 pmol have been stored at -20 °C.

TABLE 3.4 *List of primers and oligonucleotides. (T_m = melting temperature; # = internal number)*

name	sequence (5′ → 3′)	for/rev	length	T_m (°C)	use	#
HyYY1						
YY1 CF for	CAA CCT TTG ATA GCA CTA CAA	for	21	56,7	Amplification of HyYY1 core fragment	203
YY1 CF rev	GAG TCT CTT TTA GAA TTT GAT TT	rev	29	53,9	Amplification of HyYY1 core fragment	204
YY1 5´Race	GAA TCT TCA AAG AAC CCA TCT GAA ACT GTT	rev	30	63,3	5′Race	199
YY1 5´Nested	GAG AAT TAG GTT GTA GTG CTA TCA AAG GTT	rev	30	63,3	5′Race (Nested)	200
YY1 3´Race	CAT TTG AGG GTT GTG GTA AAA GAT TTA GTC TA	for	32	63,4	3′Race	201
YY1 3´Nested	CGC AAT CTA CCA ACC TAA AAA GTC ATA TCT	for	30	63,3	3′Race (Nested)	202
YY1 FL for	ATG GCT GAC GAT TTA ACG ATA GA	for	23	59,2	Amplification of HyYY1 FL (translated region); Template for YY1 riboprobe	205
YY1 FL rev	CTA CAC AGC ATC TTC TTC TAT GTC T	rev	25	61,3	Amplification of HyYY1 FL (translated region, stop codon included); Template for YY1 riboprobe	206
YY1 Rp 1 for	GTT TGT GCT GAA TGC GGG A	for	19	60,16	Template for YY1 riboprobe RP1	224
YY1 Rp 1 rev	CTT CTA AGA TCG CTT AAA GTT TGC	rev	24	59,44	Template for YY1 riboprobe RP1	225
YY1 Rp 2 for	CGG TAC AAT TGG ACC TAA GTT AAT	for	24	59,44	Template for YY1 riboprobe RP2	226
YY1 Rp 2 rev	CAT TCC CAT CTG AAG CCC ATA	rev	21	60,61	Template for YY1 riboprobe RP2	227
XhoI-YY1 FL 1 for	GGG AAC TCG AGA TGG CTG ACG ATT TAA CGA	for	30	68,73	Introducing XhoI restriction site for cloning of HyYY1 into pET15b(+) vector	230
BamHI-YY1 FL rev	CCT TTG GAT CCC TAC ACA GCA TCT TCT TCT A	rev	31	67,29	Introducing BamHI restriction site for cloning of HyYY1 into pET15b(+) vector	248

name	sequence (5′ → 3′)	for/rev	length	T_m (°C)	use	#
YY1 FR-BamHI rev	CCT TTG GAT CCT GAA GCC CAT ACA ATA ACT GA	rev	32	67,21	Introducing BamHI restriction site for cloning of HyYY1ΔZnF into pET15b(+) vector	231
HindIII-rbs-YY1 FL	GGG AAA AGC TTA GAA GGA GAT ATA CAT ATG GCT GAC GAT TTA ACG A	for	46	69,96	Introducing HindIII restriction site and ribosomal binding site for cloning of HyYY1 into pET21(+) vector	
YY1FL-His-XhoI rev	CCT TTC TCG AGG TGA TGA TGA TGA TGA TGC ACA GCA TCT TCT TCT ATG T	rev	49	72,26	Introducing XhoI restriction site and 6His-tag for cloning of HyYY1 into pET21(+) vector	
YY1FR-His-XhoI rev	CCT TTC TCG AGG TGA TGA TGA TGA TGA TGT GAA GCC CAT ACA ATA ACT	rev	48	71,56	Introducing XhoI restriction site and 6His-tag for cloning of HyYY1ΔZnF into pET21(+) vector	
EcoRI-Koz-YY1 FL for	GGG AAG AAT TCG CCG CCA TGG CTG ACG ATT TAA CGA TAG AT	for	41	72,7	Introducing EcoRI restriction site and Kozak sequence for cloning of HyYY1 into pBac1 vector	
YY1 FL-KpnI rev	CCT TTG GTA CCC TAC ACA GCA TCT TCT TCT A	rev	31	67,28	Introducing KpnI restriction site for cloning of HyYY1 into pBac1 vector	292
YY1 FL-His-XhoI rev	CCT TTC TCG AGG TGA TGA TGA TGA TGA TGC TAC ACA GCA TCT TCT TCT ATG T	rev	52	72,63	Introducing XhoI restriction site and 6His-tag for cloning of HyYY1 into pBac1 vector	
YY1 Seq for	CGA TGT TCC TCG TAA ATG GTC A	for	22	60,8	Sequencing	207
YY1 Seq rev	CGC ATT CAG CAC AAA CAT GAA CA	rev	23	61	Sequencing	208
HySCM						
SCM CF for	CTT ATC CCC CAT CAA AT	for	17	52,37	Amplification of HySCM core fragment	85
SCM CF rev	CTT CAA ATC TTC AAT CAA AT	rev	20	50,15	Amplification of HyYY1 core fragment	86
SCM 5′Race	GCT TCA AGT TTC ATC CCA ACT GA	rev	23	60,99	5′Race	87
SCM 5′Nested	CGC ATT TGA TGG GGG ATA AGA A	rev	22	60,81	5′Race (Nested)	88
SCM 3′Race	GAT GGT AAA GCC CTA CTT CTT TT	for	23	59,2	3′Race	89
SCM 3′Nested	CTC GGA CCA GCC GTC AAA	for	18	62,18	3′Race (Nested)	90
SCM FL for 1	ATG GAA GTA CTT ATA TCA AAC ATA GT	for	26	56,7	Amplification of HySCM FL (translated region)	192

name	sequence (5′ → 3′)	for/rev	length	T_m (°C)	use	#
SCM FL for 2	ATG AGC TAC AGA ATA TGG TTT	for	21	55,2	Amplification of HySCM FL (translated region); Template for SCM riboprobe RP1	193
SCM FL for 3	ATG GAA ATG ATT AAC GCA A	for	19	48	Amplification of HySCM FL (translated region)	
SCM FL rev	CTA TGG AGT CAA ATA TCA ATA AA	rev	23	56	Amplification of HySCM FL (translated region, stop codon included); Template for SCM riboprobe RP3	194
SCM Seq for 1	CTC GGA CTA ATA ATG GTC ATT TTG A	for	25	59,66	Sequencing	133
SCM Seq for 2	CAG GCT ATA GGC ATG ACC CAA	for	21	62,57	Sequencing; Template for SCM riboprobe RP2	135
SCM Seq for 3	GAT ACT CCC CTA ATC TTC TTC A	for	22	61	Sequencing	195
SCM Seq for 4	CAA TAG GTG ATG AAG TCC ATG A	for	22	58,9	Sequencing	197
SCM Seq for 5	CGG AAA ACT AAT GAA GAT AAA AGT	for	24	56	Sequencing	198
SCM Seq rev 1	GTT CAT GGA CTT CAT CAC CTA T	rev	22	58,94	Sequencing	134
SCM Seq rev 2	CTG TGA GAC TGC ATT TAT GAC TAA	rev	24	59,44	Sequencing	136
SCM Seq rev 3	GTT TTA GGC TTC ATT GGA GTA AGT T	rev	25	59,66	Sequencing of SCM FL large	182
SCM Seq rev 4	GAC AAG AAG CAT AAT TTC AGT TGA	rev	24	59,7	Sequencing; Template for SCM riboprobe RP2	196
SCM Rp 1 rev	GAG CCA CGA TCA GTC CAA	rev	18	59,9	Template for SCM riboprobe RP1	228
SCM Rp 3 for	GAT GGT GAC ATT CAT TCA GTT A	for	22	57,08	Template for SCM riboprobe RP3	229
XhoI-SCM for	GGG AAC TCG AGA TGA GCT ACA GAA ATA TGG T	for	31	67,28	Introducing XhoI restriction site for cloning of HySCM into pET15b(+) vector	265
SCM FL-BamHI rev	CCT TTG GAT CCC TAT GGA GTC AAA GTA TCA ATA	rev	33	65,9	Introducing BamHI restriction site for cloning of HySCM into pET15b(+) vector	266
SCM FR-BamHI rev	CCT TTG GAT CCG AGA CTG CAT TTA TGA CTA A	rev	31	65,96	Introducing BamHI restriction site for cloning of HySCMΔMBT into pET15b(+) vector	264
HyEZH2						
EZH2 FL for	GTT GAT GAT GAA ACA GTT TTG CAT A	for	25	52,9	Amplification of HyEZH2 FL (translated region)	57
EZH2 FL rev	TCA CAC TGT CCA TTT CTT TG	for	20	45,4	Amplification of HyEZH2 FL (translated region)	60

name	sequence (5′ → 3′)	for/rev	length	T_m (°C)	use	#
EZH2 RP1 for	CTG AAG AAG CAA AGT CTA TTT ACA	for	24	57,3	Template for EZH2 riboprobe RP1	
EZH2 RP1 rev	CAA TGT ATG GTA AGA GTG CAT	rev	21	56,7	Template for EZH2 riboprobe RP1	
EZH2 RP2 for	CAT GTG AAC ATC CTG GGC A	for	19	60,16	Template for EZH2 riboprobe RP2	
EZH2 RP2 rev	CTG TCC ATT TCT TTG CAA CAC CT	rev	23	60,98	Template for EZH2 riboprobe RP2	
HyEED						
EED FL for	ATG GAT ACT GTT AAT GAT TTT AT	for	23	52,07	Amplification of HyEED FL (translated region); Template for EED riboprobe RP1	219
EED FL rev	TGT CTG CTT GTC CCA T	rev	16	54,15	Amplification of HyEED FL (translated region); Template for EED riboprobe RP2	220
EED RP1 rev	CTG TCT CAA GTG ACC AAA AT	rev	20	56,3	Template for EED riboprobe RP1	
EED RP2 for	CAG TGC AAG TAA AGA TCA CA	for	20	56,3	Template for EED riboprobe RP2	
HySUZ12						
SUZ FL for	ATG ACT TCT TTA AAT CAA AAG GTG GAT CTT CAC TTG G	for	37	65,79	Amplification of HySUZ12 FL (translated region)	306
SUZ FL rev	CAT TTT TGG TAA AAG TTC TTC TTT CAA CTT GTA G	rev	34	62,25	Amplification of HySUZ12 FL (translated region)	307
SUZ12 RP1 for	ATG ACT TCT TTA AAT CAA AAG GTG GA	for	26	58,28	Template for SUZ12 riboprobe RP1	
SUZ12 RP1 rev	CGA TGT AAT CCA TTT TCT TCA GT	rev	23	57,42	Template for SUZ12 riboprobe RP1	
SUZ12 RP2 for	CAG ACC GAA CTG AAG AAA AT	for	20	56,3	Template for SUZ12 riboprobe RP2	
SUZ12 RP2 rev	GTA AGG CCG TCC AAC AAT A	rev	19	58	Template for SUZ12 riboprobe RP2	
Hyp55						
P55 FL for	ATG GCT GCT TTA AGT GCT CCA AA	for	23	60,99	Amplification of Hyp55 FL (translated region)	180
P55 FL rev	AGG TTT CGA TTG GTG CCT T	rev	19	58	Amplification of Hyp55 FL (translated region); Template for p55 riboprobe RP1	181
P55 RP1 for	CAG GAC ATT TAT TAA GTG CAT CT	for	23	57,42	Template for p55 riboprobe RP1	

name	sequence (5′ → 3′)	for/rev	length	T_m (°C)	use	#
HyPC						
PC FL for	ATG CCT AAG GAC ACT GGT	for	18	57,62	Amplification of HyPC FL (translated region)	82
PC FL rev	TCA TGC AAA AAA ACC TTG GTC T	rev	22	57,08	Amplification of HyPC FL (translated region)	83
HyPH						
PH FL for	ATG GCT GTT TCA TTG GTG CAA	for	21	58,66	Amplification of HyPH FL (translated region)	213
PH FL rev	GGC GAT CTT AAA ACT ATG GTG A	rev	22	58,94	Amplification of HyPH FL (translated region)	215
HyPSC						
PSC FL for	ATG GAA GAA AAG AAC TCA CAT AA	for	23	55,64	Amplification of HyPSC FL (translated region)	277
PSC FL rev	CTT TAT CTT TAA CAA ACT ATT TCG A	rev	25	54,74	Amplification of HyPSC FL (translated region)	278
HyRING						
RING FL for	ATG TCA CAG TCT GCA CAA	for	18	55,34	Amplification of HyRING FL (translated region)	51
RING FL rev	TTA CCG ATT TTC ACT TGA	rev	18	50,79	Amplification of HyRING FL (translated region)	52
Gene race primers						
5′ GeneRacer	CGA CTG GAG CAC GAG GAC ACT GAC ATG GAC TGA AGG AGT AGA AA	/	23	66,3	5′ Race RNA oligo	210
5′ GR	CGA CTG GAG CAC GAG GAC ACT GA	rev	23	68,1	5′ Race primer	211
5′ GR Nested	GGA CAC TGA CAT GGA CTG AAG GAG TA	rev	26	66,2	5′ Race nested primer	212
QtdT	CCA GTG AGC AGA GTG ACG AGG ACT CGA GCT CAA GCT TTT TTT TTT TTT TTT T	/	52	71,84	3′ Race oligo for ligation	
Qo	CCA GTG AGC AGA GTG ACG	for	18	62,17	3′ Race primer	
Qi	GAG GAC TCG AGC TCA AGC	for	18	62,17	3′ Race nested primer	

name	sequence (5′ → 3′)	for/rev	length	T_m (°C)	use	#
Common primers						
M13 for	AGG GTT TTC CCA GTC ACG ACG TT	for	23	64,55	Sequencing of pGEM-T	
M13 rev	GAG CGG ATA ACA ATT TCA CAC AGG	rev	24	62,85	Sequencing of pGEM-T	
pET rev	GGA TAT AGT TCC TCC TTT CAG CAA	rev	24	61,15	Sequencing of pET-vectors	119
pET for	CCC GCG AAA TTA ATA CGA CTC A	for	22	60,81	Sequencing of pET-vectors	120
pFastBac1 Seq for	CAT ACC GTC CCA CCA TCG GGC	for	21	68,42	Sequencing of pFastBac1	299
PFastBac1 Seq rev	GAT CCT CTA GTA CTT CTC GAC AAG	rev	24	62,86	Sequencing of pFastBac1	300
HyEF1alpha for	GTT GGT CGT GTT GAA ACT GG	for	20	57,3	Amplification of HyEF1alpha as a control for semi-quantitative RT-PCR	
HyEF1alpha rev	TCC AGC AGC AAC ACC TGC TT	rev	20	59,4	Amplification of HyEF1alpha as a control for semi-quantitative RT-PCR	
Oligonucleotides for binding site selection and gel retardation assays						
EMSA random	CGT GAT TGG CTA CGC ATC GT CAG CAG CAG CAG CAG NNN NNN NNN NNN NNN CAG CAG CAG CAG CAG GTC ATA CTC ACT CAC GGA CT	/	85	82,41	Oligo with random core for HyYY1 binding site selection	
Klenow rev	AGT CCG TGA GTG AGT ATG AC	rev′	20	60,4	Primer used for annealing with EMSA random primer to generate a double-stranded oligo for binding site selection with the Klenow enzyme	
Klenow for	CGT GAT TGG CTA CGC ATC GT	for′	20	62,4	Control primer for the Klenow reaction	
***Hy*YY1 round 6 for**	GAT TGG CTA CGC ATC GTG TAA AGT TTT CTT GAG TCA TAC TCA CTC ACG GA	for	50	74,3	Oligo used for biotin-labeling and gel retardation assays; core sequence obtained in binding site selection with HyYY1 after round 6	
***Hy*YY1 round 6 rev**	TCC GTG AGT GAG TAT GAC TCA AGA AAA CTT TAC ACG ATG CGT AGC CAA TC	rev	50	74,3	Oligo used for biotin-labeling and gel retardation assays; core sequence obtained in binding site selection with HyYY1 after round 6	

name	sequence (5′ → 3′)	for/rev	length	T_m (°C)	use	#
***Hy*YY1 round 3 for**	GAT TGG CTA CGC ATC GTA GTT TTT TGT TGA ATG TCA TAC TCA CTC ACG GA	for	50	73,5	Oligo used for biotin-labeling and gel retardation assays; core sequence obtained in binding site selection with HyYY1 after round 3	
***Hy*YY1 round 3 rev**	TCC GTG AGT GAG TAT GAC ATT CAA CAA AAA ACT ACG ATG CGT AGC CAA TC	rev	50	73,5	Oligo used for biotin-labeling and gel retardation assays; core sequence obtained in binding site selection with HyYY1 after round 3	
control for	GAT TGG CTA CGC ATC GTC CTG CAG CAC TGT GTG TCA TAC TCA CTC ACG GA	for	50	≥75	Control oligo used for biotin-labeling and gel retardation assays; core sequence obtained in binding site selection with HyYY1ΔZnF after round 6	
control rev	TCC GTG AGT GAG TAT GAC ACA CAG TGC TGC AGG ACG ATG CGT AGC CAA TC	rev	50	≥75	Control oligo used for biotin-labeling and gel retardation assays; core sequence obtained in binding site selection with HyYY1ΔZnF after round 6	
YY1 consensus for	GAT TGG CTA CGC ATC GTG CCG CCA TTT TGG GCG TCA TAC TCA CTC ACG GA	for	50	≥75	Oligo used for biotin-labeling and gel retardation assays; core sequence corresponds to the known Drosophila and human consensus motif	
YY1 consensus rev	TCC GTG AGT GAG TAT GAC GCC CAA AAT GGC GGC ACG ATG CGT AGC CAA TC	rev	50	≥75	Oligo used for biotin-labeling and gel retardation assays; core sequence corresponds to the known Drosophila and human consensus motif	

3.1.5. Vectors

Cloning vectors

The pGEM-T vector is a linearized with 3′-T overhangs at both ends. The vector allows the efficient ligation of PCR products. A map of the pGEM-T vector and its multiple cloning site are depicted in figure 3.1.

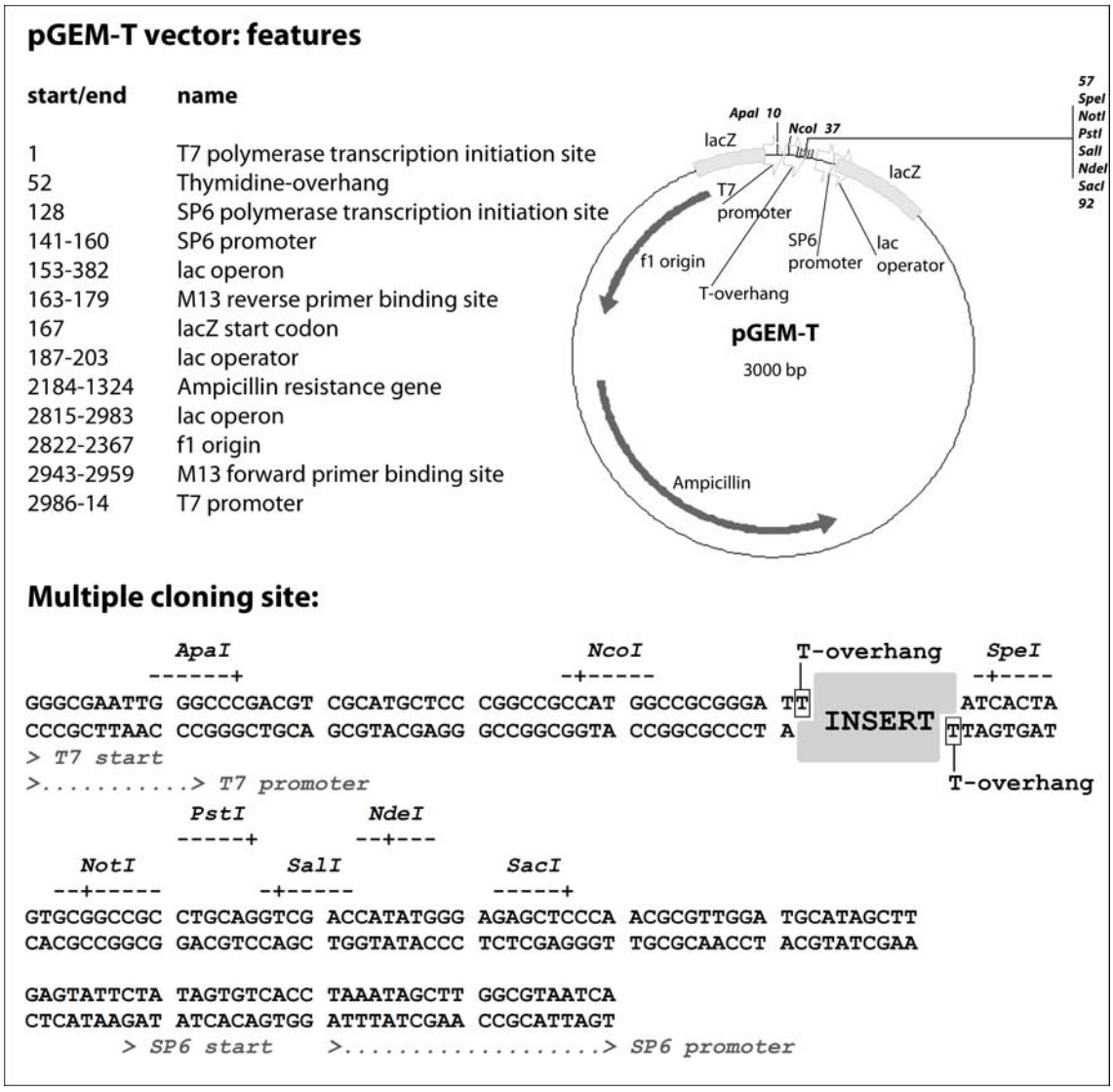


FIGURE 3.1 *pGEM-T vector: map, features and multiple cloning sites.*

TABLE 3.6 *List of cloning vectors. The genes that were cloned into the vector are indicated.*

plasmid	size (bp)	insert	cloning
pGEM-T	3000	-	commercially available from Promega
pGEM-T-*Hy*YY1	4088	*Hy*YY1 full length coding region	Sonja Matt (Holstein lab)
pGEM-T-*Hy*YY1 CF	3883	*Hy*YY1 'core fragment'	Sonja Matt (Holstein lab)
pGEM-T-*Hy*YY1 5′-end	3212	*Hy*YY1 fragment containing the 5′ UTR of the gene	Sonja Matt (Holstein lab)
pGEM-T-*Hy*YY1 3′-end small	3493	*Hy*YY1 fragment containing the 3′ UTR of the gene including the polyA signal	Sonja Matt (Holstein lab)
pGEM-T-*Hy*YY1 3′-end large	3634	*Hy*YY1 fragment containing the 3′ UTR of the gene including the polyA signal	Sonja Matt (Holstein lab)
pGEM-T-*Hy*YY1 RP1	3761	*Hy*YY1 template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*YY1 RP2	3480	*Hy*YY1 template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*SCM	5501	*Hy*SCM full length coding region	Sonja Matt (Holstein lab)
pGEM-T-*Hy*SCM CF	4705	*Hy*SCM 'core fragment'	Sonja Matt (Holstein lab)
pGEM-T-*Hy*SCM 5′-end	3875	*Hy*SCM fragment containing the 5′ UTR of the gene	Sonja Matt (Holstein lab)
pGEM-T-*Hy*SCM 3′-end	3204	*Hy*SCM fragment containing the 3′ UTR of the gene	Sonja Matt (Holstein lab)
pGEM-T-*Hy*PC	3678	*Hy*PC full length coding region	Anne Kuhn (Holstein lab)
pGEM-T-*Hy*PH	3990	*Hy*PH full length coding region	Anne Kuhn (Holstein lab)
pGEM-T-*Hy*PSC	3892	*Hy*PSC full length coding region	Anne Kuhn (Holstein lab)
pGEM-T-*Hy*RING	3980	*Hy*RIN full length coding region	Volker Lauschke (Holstein lab)
pGEM-T-*Hy*EZH2	5175	*Hy*EZH2 full length coding region	Christoph Herbel (Holstein lab)
pGEM-T-*Hy*EZH2 RP1	3745	*Hy*EZH2 template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*EH2 RP2	3713	*Hy*EZH2 template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*EED	4263	*Hy*EED full length coding region	Christoph Herbel (Holstein lab)
pGEM-T-*Hy*EED RP1	3744	*Hy*EED template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*EED RP2	3753	*Hy*EED template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*SUZ12	4916	*Hy*SUZ12 full length coding region	Volker Lauschke (Holstein lab)
pGEM-T-*Hy*SUZ12 RP1	3715	*Hy*SUZ12 template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*SUZ12 RP2	3764	*Hy*SUZ12 template for *ISH* riboprobe	Sonja Matt (Holstein lab)
pGEM-T-*Hy*p55	4305	*Hy*p55 full length coding region	Christoph Herbel (Holstein lab)
pGEM-T-*Hy*p55 RP	3721	*Hy*p55 template for *ISH* riboprobe	Sonja Matt (Holstein lab)

Prokaryotic expression vectors

The pET15b(+) vector is a prokaryotic expression vector that allows the expression of a fusion protein with an N-terminal 6His-tag. The pET21(+) vector is a transcription vector that lacks the ribosomal binding site and ATG start codon necessary for recombinant protein expression. It carries a C-terminal 6His-tag. Maps of the vectors and their multiple cloning sites are depicted in figures 3.2 and 3.3.

TABLE 3.8 *List of prokaryotic expression vectors. The expressed protein and the cloning vector are indicated.*

plasmid	size (bp)	insert	cloning
pET15b(+)	5708	-	available from Novagen
pET21(+)	5369	-	available from Novagen
pET15b(+)-*Hy*YY1	6795	*Hy*YY1 full length coding region	Sonja Matt (Holstein lab)
pET15b(+)-*Hy*YY1ΔZnF	6141	*Hy*YY1 lacking its C-terminal zinc finger domains	Sonja Matt (Holstein lab)
pET21(+)-*Hy*YY1	6477	*Hy*YY1 full length coding region	Sonja Matt (Holstein lab)
pET21(+)-*Hy*YY1ΔZnF	5826	*Hy*YY1 lacking its C-terminal zinc finger domains	Sonja Matt (Holstein lab)

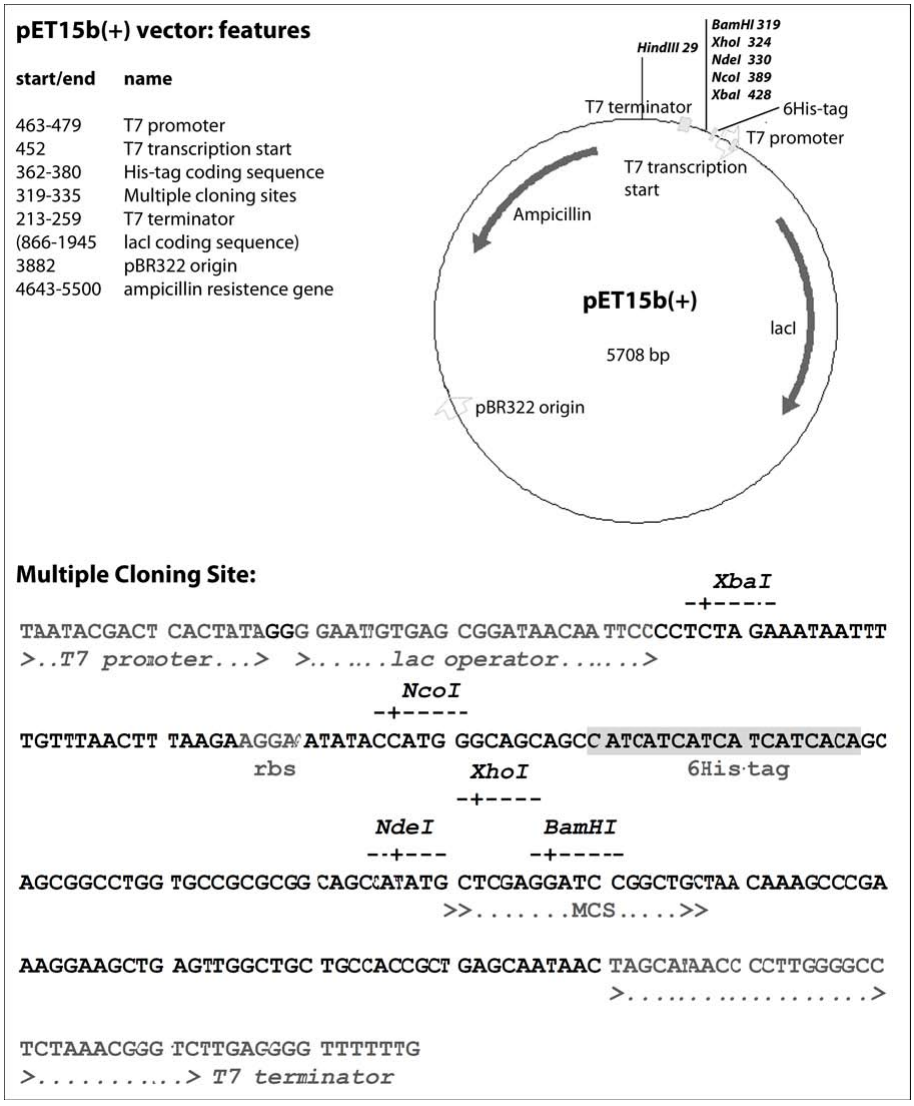


FIGURE 3.2 *pET15b(+) vector: map, features and multiple cloning sites.*

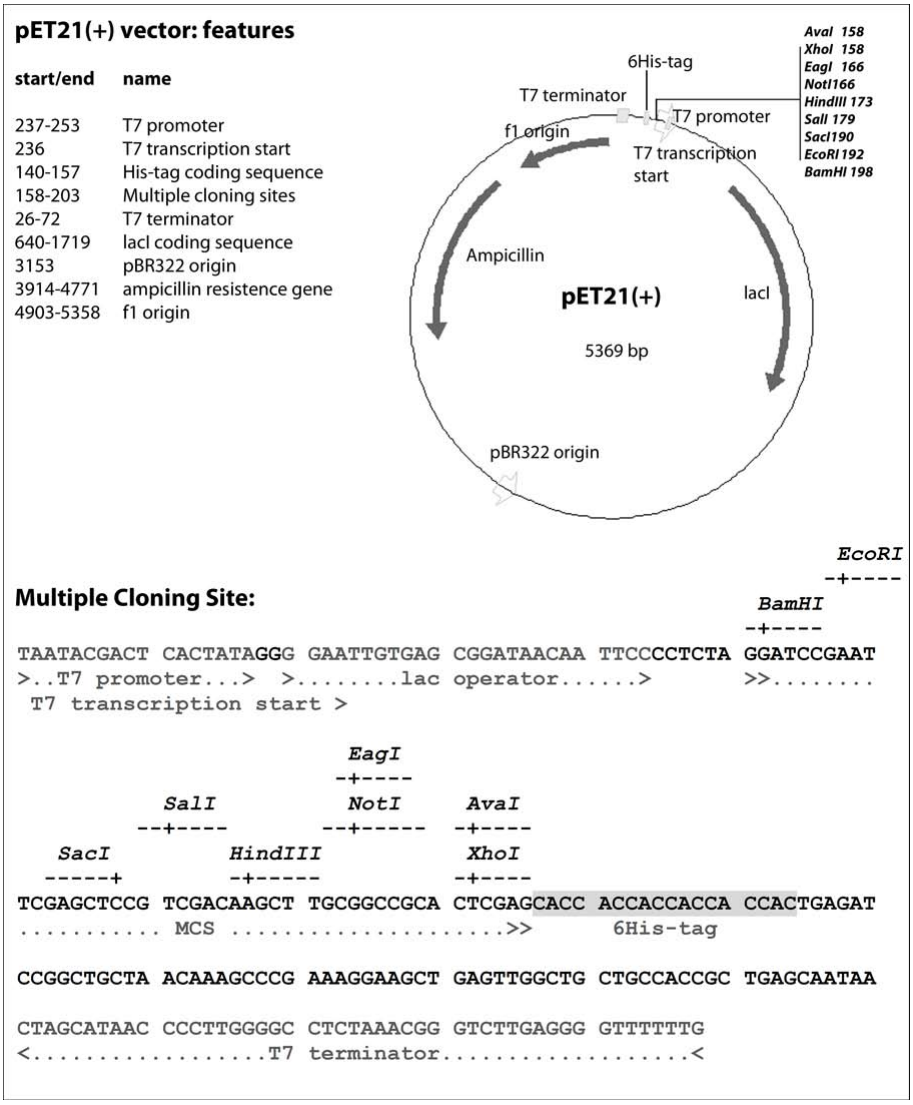


FIGURE 3.3 *pET21(+) vector: map, features and multiple cloning sites.*

Eukaryotic expression vectors

The pFastBac1 vector allows the generation of an expression construct containing the gene of interest, which is controlled by the polyhedrin promoter. The expression cassette is flanked by the left and the right arms of Tn7. A map of the pFastBac1 vector and its multiple cloning site are depicted in figure 3.4.

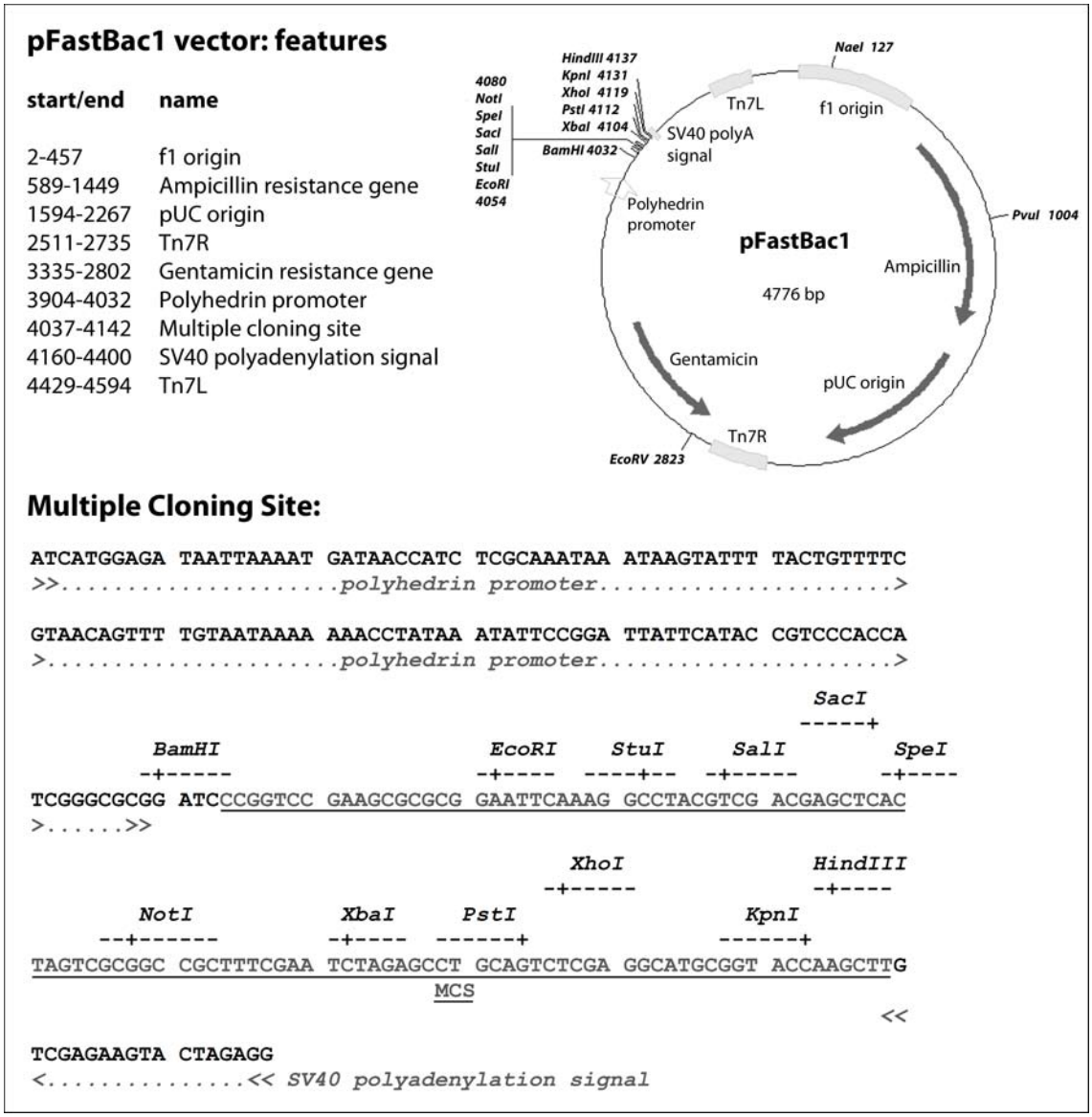


FIGURE 3.4 *pFastBac1 vector: map, features and multiple cloning sites.*

TABLE 3.10 *List of eukaryotic expression vectors. The expressed protein and the cloning vector are indicated.*

plasmid	size (bp)	insert	cloning
pFastBac1	4776	-	available from Invitrogen
pFastBac1-*Hy*YY1	5827	*Hy*YY1 full length coding region	Sonja Matt (Holstein lab)
pFastBac1-*Hy*YY1-6His	5824	*Hy*YY1 full length coding region with a C-terminal 6His-tag	Sonja Matt (Holstein lab)
pFastBac1-*Hy*EZH2	6941	*Hy*EZH2 full length coding region	Volker Lauschke (Holstein lab)
pFastBac1-*Hy*EED	6039	*Hy*EED full length coding region	Volker Lauschke (Holstein lab)
pFastBac1-*Hy*SUZ12	6687	*Hy*SUZ12 full length coding region	Volker Lauschke (Holstein lab)
pFastBac1-*Hy*p55	6070	*Hy*p55 full length coding region	Volker Lauschke (Holstein lab)

3.1.6. Enzymes

Restriction endonucleases

restriction enzyme	recognition sequence	temperature	buffer	manufacturer
ApaI	GGGCC↓C	37 °C	Buffer B	Fermentas
BamHI	G↓GATCC	37 °C	NEBuffer 3 + 100 μg/ml BSA	NEB
EcoRI	G↓AATTC	37 °C	NEBuffer 1, 2, 3 or 4	NEB
HindIII	A↓AGCTT	37 °C	NEBuffer 2	NEB
SacI	GAGCT↓C	37 °C	Buffer B	Fermentas
XhoI	C↓TCGAG	37 °C	NEBuffer 2, 3 or 4	NEB

Buffer composition:
NEBuffer 1: 10 mM Tris Propane-HCl, 10 mM $MgCl_2$, 1 mM DTT (pH 7.0 at 25 °C)
NEBuffer 2: 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT (pH 7.9 at 25 °C)
NEBuffer 3: 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT (pH 7.9 at 25 °C)
NEBuffer 4: 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT (pH 7.9 at 25 °C)
Buffer B: 10 mM tris-HCl, 10 mM $MgCl_2$, 0.1 mg/ml BSA (pH 7.5 at 37 °C)

Polymerases and other enzymes

enzyme	recognition sequence	manufacturer
Phusion	proofreading DNA polymerase	Finnzymes (No. F530S)
EuroTaq	DNA polymerase	BioCat (No. EME010001)
Klenow DNA polymeraseI	DNA polymerase lacking the 5′ - 3′ exonuclease activity	NEB (No. M0210S)
BigDye Terminator v3.1	ready reaction mix for cycle sequencing	Applied Biosystems (No. 4337457)
SP6 RNA polymerase	RNA polymerase	Roche (No. 10 810 274 001)
T7 RNA polymerase	RNA polymerase	Roche (No. 10 881 767 001)
DNaseI, RNase-free	deoxyribonuclease	Roche (No. 04 716 728 001)
RNaseA	endoribonuclease	Applichem (No. A3832)
RNaseH	ribonuclease	Invitrogen (No. 18080-051)
RNaseOUT	recombinant ribonuclease inhibitor	Invitrogen (No. 18080-051)
SuperScriptIII RT	reverse transcriptase	Invitrogen (No. 18080-051)
T4-PNK	polynucleotide kinase	Fermentas (EK0031)
TdT	Terminal deoxynucleotidyl transferase	Thermo Scientific (No. 89818)
proteinase K	serine protease	Roche (No. 1 092 766)
T4 RNA ligase	RNA ligase	Invitrogen (No.)
T4 DNA ligase	DNA ligase	Promega (No. M1804)

3.1.7. Antibodies

TABLE 3.12 *Primary and secondary antibodies. (WB = Western Blot, IH = Immunohistochemistry, IP = Immunoprecipitation, ChIP = Chromatin Immunoprecipitation)*

antibody	species	application	purity	manufacturer
primary antibodies:				
α**-Penta-His**	mouse IgG1	WB, IP	affinity pure	Qiagen (No. 34660)
α**-H3K27me3**	rabbit (polyclonal)	WB, IH, ChIP	affinity pure	Millipore (No. 07-449)
α**-*Hy*YY1**	rabbit (polyclonal)	WB, IH, IP, ChIP	crude serum, affinity pure	Holstein lab (Sonja Matt)
α**-*Hy*EZH2**	rabbit (polyclonal)	WB, IH, IP, ChIP	crude serum	Holstein lab (Christoph Herbel)
α**-*Hy*EED**	rabbit (polyclonal)	WB, IH, IP, ChIP	crude serum	Holstein lab (Volker Lauschke)
α**-*Hy*p55**	rabbit (polyclonal)	WB, IH, IP, ChIP	crude serum	Holstein lab (Volker Lauschke)
α**-*Hy*PC**	rabbit (polyclonal)	WB, IH, ChIP	crude serum	Holstein lab (Anne Kuhn)
α**-*Hy*RING**	rabbit (polyclonal)	WB, IH	crude serum	Holstein lab (Anne Kuhn)
α**-Digoxigenin-AP**	sheep (Fab fragments)	ISH	AP-conjugated	Roche (No. 11 093 274 910)
secondary antibodies:				
Goat-α**-Mouse HRP**	goat IgG (H+L)	WB	HRP conjugate	Dianova (No. 115-035-003)
Goat-α**-Rabbit HRP**	goat IgG (H+L)	WB	HRP conjugate	Dianova (No. 111-035-003)
Alexa Fluor 488 donkey-α**-rabbit**	donkey IgG (H+L)	IH	Alexa488-conjugated	Molecular Probes (No. A21206)

3.2. Methods: Laboratory animals

3.2.1. Cultivation of *Hydra*

The animals were kept in artificial Hydra Medium at 18 °C in polystyrene dishes. The initial culturing conditions of *Hydra* and the Hydra Medium recipe were described by Lenhoff in 1953. The *Hydra* polyps were fed two to three times a week with freshly hatched *Artemia salina nauplii*. The animals were washed twice after feeding and the medium was exchanged. Prior to RNA isolation or protein extraction, the animals were starved for 48 hours. The following *Hydra*-strains have been used to perform the experiments. Both strains belong to the *vulgaris* group:

Hydra magnipapillata 105:	also known as *Hydra magnipapillata* B8. This strain is the standard wild type strain whose genome was sequenced during the *Hydra* genome project. It originates from a single polyp collected by Dr. Tsutomo Sugiyama in a swamp adjacent to the National Institute of Genetics in Mishima, Japan.
Hydra vulgaris AEP:	This strain is a result of crosses of several American wild type strains. The founder strains are a male strain called CA7 which was collected from Boulder Creek (Susanville, California) and a female strain called PA1 which was collected from a pond on the Haverford College campus (Haverford, Pennsylvania). This strain produces eggs and sperm frequently. Its embryos are used for generating transgenic *Hydra*.

Hydra Medium (HM)	1 mM Tris-Cl pH 7.4 0.1 mM KCl 1 mM $CaCl_2$ 1 mM $NaHCO_3$ 0.1 mM $MgCl_2$

Preparation of artemia

Artemia were prepared by incubating two spoons of *Artemia* cysts (Inve) in 2 litres of salt water (32.5 g sea salt/l H_2O, pH 8.2) in aerated flasks at 32 °C for 24 hours. The hatched *Artemia* were seperated from unhatched cysts and loose eggshells. Hatched *Artemia* were kept in salt water at 18 °C in aerated flasks. Prior to feeding, the *Artemia* were washed with Hydra Medium to avoid changing the salt concentration in the *Hydra* culture during feeding.

3.2.2. Regeneration experiments with *Hydra*

For regeneration experiments polyps without buds have been used. For head regeneration experiments *Hydra* was cut with a scalpel just beneath its tentacle base. After decapitation the regenerating tip was transferred to a petri dish with fresh Hydra Medium. The polyp was allowed to regenerate for different periods of time. For RNA isolation and in situ hybridization, the polyps were allowed to regenerate for 1.5, 3, 6, 12, 24 and 48 hours.

3.2.3. Treatment of *Hydra* with hydroxyurea

It has been shown by Bode et al. (1976) that *Hydra* can be treated with hydroxyurea (HU) to selectively deplete its I-cell lineage. 50 polyps were transferred into a petri dish with fresh Hydra Medium and fed daily for three days. The Hydra Medium was then replaced with Hydra Medium containing 10 mM HU. The animals were kept in medium containing HU for four days with exchanging the medium daily. They were fed on the first two days of the treatment and then starved. On the fourth day, the animals were washed with normal Hydra Medium and used for RNA isolation and in situ hybridization immediately. As a control, animals were kept in the same way, except that the Hydra Medium was not exchanged with medium containing HU.

3.3. Methods: Microbiological techniques

3.3.1. Cultivation of *E. coli*

E. coli-strains were cultivated in LB-medium at 37 °C and 200 rpm in a shaker or on LB-agar plates at 37 °C in an incubator for 12 to 16 hours. To select for the desired recombinant bacteria antibiotics were added to the medium if necessary. During the thesis the follwing *E.coli*-strains have been used:

DH5α:	*F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG ϕdlacZΔM15 Δ(lacZYA-argF)U169 hsdR17(r_K- m_K-) λ-*
DH10Bac:	*F^- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL nupG/bMON14272/pMON7124*
BL21(DE3):	*F^- ompT gal dcm lon hsdS_B(r_B- m_B-) λ(DE3 [lacI lacUV5-T7 gene1 ind1 sam7 nin5])*
Rosettagami (DE3) pLysS:	*Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F'[lac^+ $lacI^q$ pro] gor522::Tn10 trxB pLysSRARE (Cam^R, Str^R, Tet^R)*

Media, solutions and antibiotics

LB-medium:
10 g tryptone
5 g yeast extract
5 g NaCl
Adjust volume to 1 liter with H_2O.
Sterilize by autoclaving.

LB-agar plates:
10 g tryptone
5 g yeast extract
5 g NaCl
15 g agar
Adjust volume to 1 liter with H_2O.
Sterilize by autoclaving.
Cool down to 50 °C and add the appropriate amount of antibiotic, X-Gal and IPTG. Mix the medium gently and pour into sterile plates. To remove air bubbles, flame the surface of the medium with a bunsen burner. After the medium hardened, the plates are stored at 4 °C.
For AXI-plates 100 μg/ml ampicillin, 40 μg/ml X-Gal and 0.2 M IPTG are added.
For DH10Bac-plates 50 μg/ml kanamycin, 7 μg/ml gentamycin, 10 μg/ml tetracyclin, 100 μg/ml X-Gal and 40 μg/ml IPTG are added.

SOC-medium:
20 g tryptone
5 g yeast extract
0.5 g NaCl
2.5 mM KCl
Adjust volume to 1 liter with H2O.
Sterilize by autoclaving.
Add $MgCl_2$ to a final concentration of 10 mM and glucose to 20 mM.

$CaCl_2$ solution:
60 mM $CaCl_2$
15% glycerol
10 mM PIPES pH 7.0

X-Gal stock-solution:
20 mg/ml in dimethylformamide
Sore at - 20 °C.

IPTG stock-solution:
1 M in H_2O
Store at - 20 °C.

Ampicillin stock-solution:
100 mg/ml in H_2O
Sterilize by filtration. Store at - 20 °C.
Working concentration: 100 μg/ml.

Kanamycin stock-solution:
100 mg/ml in H_2O
Sterilize by filtration. Store at - 20 °C.
Working concentration: 100 μg/ml.

Gentamycin stock-solution:
50 mg/ml in H_2O
Sterilize by filtration. Store at 4 °C.
Working concentration: 7 μg/ml.

Tetracyclin stock-solution:
6.1 mg/ml in ethanol
Store at - 20 °C.
Working concentration: 10 μg/ml for liquid cultures and 12.5 μg/ml for plates.

Chloramphenicol stock-solution:
34 mg/ml in ethanol
Store at - 20 °C.
Working concentration: 100 μg/ml.

3.3.2. Production of competent bacteria

Competent bacteria are able to take up DNA. Either chemocompetent bacteria or electrocompetent bacteria were produced. The competency of the cells is indicated by the transformation efficiency. The transformation efficiency was assessed by a test transformation and calculated as follows:

$$\text{transformation efficiency [cfu}/\mu\text{g]} = \frac{\text{number of colonies [cfu]}}{\text{amount of DNA [}\mu\text{g]}} \cdot \frac{\text{total volume}}{\text{plated volume}}$$

Electrocompetent bacteria

The production of electrocompetent bacteria was carried out as described by Dower et al. (1988). 5 ml LB-medium were inoculated with a single colony of *E.coli*-cells and incubated over night at 37 °C and 200 rpm in a shaker. 500 ml LB-medium were inoculated with the over night culture and incubated at 37 °C and 200 rpm in a shaker until an OD_{600} of 0.5 - 0.7 was reached. The following steps were carried out at 4 °C. The bacterial cells were incubated for 15 minutes on ice to cool down. Afterwards the bacteria were centrifuged for 20 minutes at 3000 g and 1 °C. The supernatant was discarded and the pellet was resuspended in 5 ml of ice-cold water. 500 ml ice-cold water were added and the suspension was centrifuged for 20 minutes at 3000 g and 1 °C. The supernatant was discarded. The bacteria were washed in ice-cold water twice. After the second wash, the bacterial cells were resuspended in the remaining supernatant and 40 ml of ice-cold 10% glycerol were added. The suspension was centrifuged for 10 minutes at 3000 g and 1 °C. The supernatant was discarded and the bacteria were resuspended in 1 volume of 10% glycerol. 55 μl aliquots were snap frozen in liquid nitrogen and stored at - 80 °C.
The transformation efficiency of the bacterial cells was determined by a test-transformation. For this purpose electrocompetent *E.coli* were transformed with a certain amount of DNA. To test the efficiency, an aliquot of electrocompetent cells was transformed with 1 ng pUC19-DNA. Different volumes (1, 10 and 100 μl) of the test-transformation were plated on LB-agar plates containing ampicillin and incubated over night at 37 °C. The number of colonies on the plates was counted and the transformation efficiency was calculated. Usually transformation efficiencies of about 10^9 cfu/μg were achieved for electrocompetent *E.coli*-BL21 (DE3) or *E.coli*-Rosettagami (DE3) pLysS.

Chemocompetent bacteria

Chemocompetent bacteria were produced based on the protocol of Mandel and Higa (1970). 50 ml LB-medium were inoculated with a single colony of *E.coli*-cells and grown over night at 37 °C and 200 rpm in a shaker. 400 ml LB-medium were inoculated with an OD_{600} of 0.05. The bacterial culture was incubated at 37 °C and 200 rpm in a shaker until an OD_{600} of 0.3 - 0.4 was reached. The bacteria were aliquoted in prechilled 50-ml tubes and cooled down on ice for 10 minutes. The cells were centrifuged for 15 minutes at 1600 g and 4 °C. The supernatant was discarded and the pellet was resuspended in 10 ml ice-cold $CaCl_2$ solution. The bacteria were centrifuged for 15 minutes at 1600 g and 4 °C. The supernatant was discarded and the pellet was again resuspended in 10 ml ice-cold $CaCl_2$ solution. The cells were kept on ice for 30 minutes. The bacteria were centrifuged for 15 minutes at 1600 g and 4 °C. The supernatant was discarded and the pellet was finally resuspended in 2 ml ice-cold $CaCl_2$ solution. 100 μl aliquots were snap frozen in liquid nitrogen and stored at - 80 °C.
The transformation efficiency of the chemocompetent bacteria were determined by a test-transformation.

An aliquot of chemocompetent cells was transformed with 10 ng pUC19-DNA. Different volumes (200 and 800 μl) of the test-transformation were plated on LB-agar plates containing ampicillin and incubated over night at 37 °C. The number of colonies on the plates was counted and the transformation efficiency was calculated. For chemocompetent *E.coli*-DH5α or *E.coli*-DH10Bac transformation efficiencies of about 10^7 cfu/μg were achieved.

3.3.3. Transformation of bacteria

E.coli-cells were either transformed by electroporation or calcium transformation. The exposure of bacterial cells to calcium ions enables them to take up DNA. Plasmid DNA is mixed with chemocompetent cells and heat shocking the cells allows the DNA to enter. Transformation by electroporation is achieved by briefly applying a high-voltage electrical field to the cells. Apparently transient holes in the cell membrane are produced and the DNA can enter the cell. Electroporation can also be used to transform eukaryotic or plant cells.

Transformation of electrocompetent bacteria

One aliquot of electrocompetent cells was thawed on ice and mixed gently with 1 ng - 200 ng of plasmid DNA. The DNA and cells were transferred to a prechilled electroporation cuvette. An electric pulse (25 μF, 2.5 kV, 200 Ω, field strength: about 12,5 kV/cm) was applied. Usually a time constant of 4.8 - 5.2 ms indicated that the procedure was successful. After transformation, the bacteria were mixed with 1 ml SOC-medium immediately and incubated at 37 °C in a shaker for 30 - 60 minutes. Different volumes of cells were either plated on LB-plates or the cells were directly used to inoculate a liquid culture. The appropriate antibiotics were added to the medium or plates to select for the desired transformants.

Transformation of chemocompetent bacteria

The chemocompetent cells were thawed rapidly and mixed gently with up to 500 ng of plasmid DNA. The DNA and cells were cooled on ice for 15 min. The cells were heat shocked for 50 seconds at 42 °C. After transformation, the bacteria were chilled on ice for 2 minutes and mixed with 1 ml SOC-medium. The bacteria were incubated at 37 °C in a shaker for 30 - 60 min. Different volumes of cells were either plated on LB-plates or the cells were directly used to inoculate a liquid culture. The appropriate antibiotics were added to the medium or plates to select for the desired transformants.

3.4. Methods: RNA

While working with RNA, special precautions have been taken to avoid the contamination with RNases. Fresh gloves have been worn and the workspace has been cleaned with 100% ethanol. The equipment like pipetmen and electrophoresis tanks have been cleaned before with 1% SDS, rinsed with H_2O followed by 100% ethanol and finally soaked in 3% H_2O_2 for 10 minutes and cleaned with H_2O. Bottles and flasks for preparing the solutions have been baked at 180 °C. Tubes and pipet tips from unopened bags have been used and working was carried out without speaking.

3.4.1. Isolation of RNA from *Hydra magnipapillata*

Total RNA of *Hydra* was isolated as described by Chomczynski and Sacchi (1987) with a mixture of acid guanidinium thiocyanate-phenol-chloroform. About 50 - 100 polyps were washed in Hydra Medium twice. The polyps were transferred to an Eppendorf tube and centrifuged at top speed in a tabletop centrifuge for 1 minute. The Hydra Medium was removed and the polyps were lysed completely in 500 μl of GTC-solution by pipetting them up and down several times on ice. 50 μl of 2 M sodium acetate (pH 4.0) and 550 μl of phenol-chloroform-isoamyl alcohol (25:24:1) were added. Mixing was achieved by vigorously shaking for 1 min by hand. The mixture was kept on ice for 10 minutes and centrifuged at top speed in a tabletop centrifuge for 5 minutes to separate the phases. The upper aqueous phase was transferred to a new Eppendorf tube. 250 μl of 1.2 M sodium chloride, 0.8 M sodium citrate and isopropanol were added, respectively. The mixture was incubated at room temperature for 15 minutes and the RNA was collected by centrifugation at top speed in a tabletop centrifuge for 15 minutes. The RNA pellet was washed three times with 250 μl of ice-cold 70% ethanol. Finally the RNA pellet was air-dried and solved in an appropriate volume of H_2O. To remove a possible contamination by genomic DNA, the RNA was treated with RNase-free DNaseI. The total RNA was mixed with 1 μl DNaseI (10 U/μl) and incubated for 20 minutes at 37 °C in 1 x Incubation Buffer. To remove the enzyme, the RNA was extracted with phenol-chloroform-isoamyl alcohol (25:24:1). The concentration and purity of the extracted RNA was determined by spectrophotometry and agarose gel electrophoresis (compare 3.5.3). The RNA was stored at - 80 °C.

Solutions for RNA extraction

GTC-solution:	4.2 M guanidinium thiocyanate (GTC) 25 mM sodium citrate (pH 7.0) 0.5% (w/v)N-lauroyl-sarcosine 5% (v/v)β-mercaptoethanol Dissolve GTC in a small volume of H_2O by heating. Add 1 M sodium citrate solution 1:40 and 10% sarcosyl solution 1:20. Add β-mercaptoethanol shortly before use. The GTC-solution can be stored at 4 °C for about four weeks.
1 M sodium citrate:	1 M sodium citrate Adjust the pH to 7.0. Sterilize by autoclaving.

10% N-lauroyl-sarcosine:	Dissolve 10% N-laroyl-sarcosine in H_2O. Sterilize by autoclaving.
2 M sodium acetate:	Dissolve 3 M sodium acetate in a small volume of H_2O. Adjust the pH to 4.0 by adding glacial acetic acid. Sterilize by autoclaving.
2.5 M sodium chloride:	Dissolve 2.5 M sodium chloride in H_2O. Sterilize by autoclaving.

3.4.2. Preparation of cDNA

cDNA was prepared for semiquantitative RT-PCR experiments and for 5′- and 3′-Race. 5′- and 3′-Race (rapid amplification of cDNA ends) PCR were performed to identify 5′- and 3′-ends of the *Hydra* Polycomb Group genes. The cDNA for 5′ and 3′ RACE was a courtesy of Dr. Anne Kuhn (Holstein lab).

Preparation of cDNA for 5′-Race

Hydra total RNA was reverse transcribed using the Invitrogen GeneRacer Kit according to the manufacturer´s instruction. To obtain cDNA for 5′-Race, first of all the 5′ cap structure from intact mRNA has to be removed and an RNA oligo has to be ligated to the 5′ end instead. 5 μg of total RNA was mixed with 1 μl RNaseOut (40 U/μl) and 1 μl CIP (10 U/μl) in a total volume of 10 μl in 1 x CIP Buffer and incubated at 50 °C for 1 hour to remove the 5′ phosphates. The RNA was extracted using phenol-chloroform-isoamyl alcohol (25:24:1) and precipitated by adding 2 μl 10 mg/ml mussel glycogen as a carrier, 10 μl 3 M sodium acetate (pH 5.2) and 220 μl 95% ethanol. The RNA was incubated at - 20 °C over night and centrifuged at top speed in a tabletop centrifuge for 20 minutes at room temperature. The RNA pellet was washed with 500 μl of ice-cold 70% ethanol. Finally the RNA pellet was air-dried and dissolved in 7 μl H_2O. The RNA was decapped using the tobacco acid pyrophosphatase. The RNA was mixed with 1 μl 10 x TAP Buffer, 1 μl RNaseOUT (40 U/μl) and 1 μl TAP (0.5 U/μl) and incubated at 37 °C for 1 hour. Afterwards the decapped RNA was phenol-chloroform extracted and precipitated again. Then the GeneRacer RNA Oligo (5′-CGACUGGAGCACGAGGACACUGACAUGGACUGAA GGAGUAGAAA-3′) was ligated to the 5′ end of the RNA. The decapped RNA was mixed with 0.25 μg GeneRacer RNA Oligo and incubated at 65 °C for 5 minutes. The reaction was chilled on ice and it was mixed with 1 μl 10 mM ATP, 1 μl RNaseOut (40 U/μl) and 1 μl T4 RNA ligase (5 U/μl) in a total volume of 10 μl in 1 x Ligase Buffer. The reaction was incubated for 1 hour at 37 °C. The RNA was phenol-chloroform extracted and precipitated once more. The RNA was transcribed into cDNA by reverse transcriptase. The RNA was mixed with 1 μl of 50 μM 5′ GeneRace primer and 1 μl 25 mM dNTP mix in a total volume of 12 μl. The mixture was incubated at 65 °C for 5 minutes and chilled on ice for 1 minute. The cDNA was synthesized by adding 4 μl of 5 x RT buffer, 1 μl Cloned AMV RT (15 U/μl), 2 μ H_2O and 1 μl RNaseOUT (40 U/μl). The reaction was incubated at 45 °C for 1 hour and terminated by inactivating the reverse transcriptase at 85 °C for 15 minutes. The cDNA was stored at - 20 °C. 5′-Race cDNA was diluted 1:10 for RACE-PCR.

Preparation of cDNA for 3′-Race

Hydra total RNA was reverse transcribed using the Invitrogen GeneRacer Kit according to the manufacturer´s instruction. Since the provided GeneRacer 3′ primer was not useful for obtaining correct 3′-ends of *Hydra* genes, the following primer was used instead:
QtdT 5′-CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC$(T)_{18}$-3′. This primer was described by Davis et al. (1995) for RACE-PCR in the flatworm *Schistosoma mansoni*. The 3′-Race cDNA was transcribed in the same way as described for the 5′-Race cDNA. 5 μg of total RNA was mixed with 1 μl of 50 μM QtdT primer and 1 μl 25 mM dNTP mix in a total volume of 12 μl. The mixture was incubated at 65 °C for 5 minutes to remove any RNA secondary structures and chilled on ice for 1 minute. The cDNA was synthesized by adding 4 μl of 5 x RT buffer, 1 μl Cloned AMV RT (15 U/μl), 2 μ H_2O and 1 μl RNaseOUT (40 U/μl). The reaction was incubated at 45 °C for 1 hour and terminated by inactivating the reverse transcriptase at 85 °C for 15 minutes. Finally the cDNA was stored at - 20 °C. The cDNA was diluted 1:10 for RACE-PCR.

3.4.3. In situ hybridization

Whole mount in situ hybridization with *Hydra* was performed as described by Grens et al. (1996) and Martinez et al. (1997).

Preparation of RNA probes for in situ hybridization

During in situ hybridization the transcripts of the gene of interest are detected using digoxygenin-labeled, single stranded "antisense" RNA probes. Purified PCR products were used as template for the *in vitro* run-off transcription of RNA probes. The "antisense" probe was obtained either with SP6 or T7 RNA polymerase.
The template was obtained by amplifying the sequence of interest, which has been cloned into pGEMT-vector - by PCR using M13 for and M13 rev primers. 1 ng of plasmid DNA were mixed with 1 μl 10 mM dNTP, 1 μl of 5 pmol M13 for primer, 1 μl of 5 pmol M13 rev primer, 1 μl Phusion DNA polymerase (2 U/μl) in a total volume of 50 μl in 1 x HF Buffer. The following PCR program was used for amplification:

initial denaturation	98 °C	3 minutes	
denaturation	95 °C	30 seconds	
annealing	56 °C	30 seconds	30 cycles
elongation	72 °C	30 seconds	
final elongation	72 °C	7 minutes	

Note: For the amplification of the template for the riboprobe YY1 fl, a longer annealing time of 45 seconds was used.

The PCR product was purified by Wizard SV Gel and PCR Clean-Up System from Promega according to the manufacturer´s instruction. The concentration and purity of the DNA was determined by spectrophotometry and agarose gel electrophoresis. The template DNA was stored at -20 °C.

For the *in vitro* transcription of the RNA probe the DIG RNA Labeling Kit from Roche was used. The reaction was performed as follows: 250 ng template DNA were mixed with 1 μl 10 x Transcription Buffer, 1 μl 10 x DIG Labeling Mix, 0.5 μl RNaseH (40 U/μl) and 1 μl SP6 or T7 RNA polymerase (20 U/μl) in a total volume of 10 μl. The reaction was incubated at 37°C for 2 hours. 2 μl DNaseI (10 U/μl) were added and the mixture was incubated for additional 15 minutes at 37 °C. The RNA probe was precipitated by adding 2 μl 200 mM EDTA (pH 8.0), 100.5 μl H_2O, 135 μl 7.8 M ammonium acetate and 420 μl 100% ethanol. The probe was precipitated for 50 minutes at room temperature and centrifuged at top speed in a tabletop centrifuge for 20 minutes at room temperature. The pellet was washed with 250 μl of ice-cold 70% ethanol. Finally the pellet was air-dried and dissolved in 40 μl 50% formamide. The labeled RNA probe was stored at - 80 °C. The labeled probes were analyzed by agarose gel electrophoresis. Different volumes (1 μl, 2.5 μl and 5 μl) of a 1:10 dilution of the labeled probe were loaded on a 1% agarose gel in formamide loading dye (98% formamide, 10 mM EDTA (pH 8.0), 0.1% xylencyanol and 0.1% bromophenol blue). Prior to loading, the RNA was incubated at 65 °C for 10 min. The amount of labeled probe was estimated by comparison with a defined amount of control RNA which has been loaded onto the gel as well. Finally the labeled RNA probe was diluted to 5 ng/μl in 100% hybridization mix and stored at - 80 °C.

The following riboprobes have been generated:

riboprobe	length	gene
eed2 rp 1	742 bp	N-terminal fragment of the *Hydra* Eed gene
eed2 rp 2	751 bp	C-terminal fragment of the *Hydra* Eed gene
ez2 rp 1	743 bp	N-terminal fragment of the *Hydra* EZ2 gene
ez2 rp 2	711 bp	C-terminal fragment of the *Hydra* EZ2 gene
p55 rp 1	719 bp	C-terminal fragment of the *Hydra* p55 gene
suz12 rp 1	713 bp	N-terminal fragment of the *Hydra* SUZ12 gene
suz12 rp 2	767 bp	C-terminal fragment of the *Hydra* SUZ12 gene
yy1 rp 1	759 bp	C-terminal fragment of the *Hydra* YY1 gene
yy1 rp 2	478 bp	N-terminal fragment of the *Hydra* YY1 gene
yy1 fl rp	1131 bp	full length coding region of the *Hydra* YY1 gene

Whole mount in situ hybridization

For whole mount in situ hybridization about 300 *Hydra* polyps have been transferred to a 50 ml tube. The polyps were washed with fresh Hydra Medium twice by aspirating the old medium with a pasteur capillary pipette and adding fresh medium. To relax the animals, 20 ml 4% urethane were added to 20 ml Hydra Medium. The animals were incubated in 2% urethane for 1 - 2 minutes. The urethane was removed completely and the polyps were washed once in 4% formaldehyde. The formaldehyde was exchanged with 40 ml fresh 4% formaldehyde. The polyps were fixed over night at 4 °C on a roller mixer.

The fixative was removed from the animals and 40 ml 100% ethanol was added. Ethanol makes the membranes permeable and the probes can enter the tissue easier. The polyps were incubated for 5 minutes at room temperature on a roller mixer. This step was repeated four more times until the polyps have lost their colour and turned white. To rehydrate the animals, subsequent washes with 75%, 50% and 25% ethanol in 1 x PBST were performed. All steps were performed for 5 minutes at room temperature on a roller mixer. The animals were washed three times with 1 x PBST for 5 minutes. 1 x proteinase K in 1 x PBST was added to the polyps for 10 minutes. The fixed tissue is partially digested and gets permeabilized. The enzyme was removed completely from the polyps and the proteinase K treatment was terminated by the addition of 1 x glycine in 1 x PBST. The glycine solution was removed and the polyps were again washed in 1 x glycine in 1 x PBST for 5 minutes. Afterwards, the polyps were twice washed with 1 x PBST for 5 minutes, followed by two washes with 1 x TEA for minutes. To improve the signal to noise ratio, the animals were incubated for 5 minutes with a 1:400 dilution of acetic anhydride in 1 x TEA followed by an incubation for 5 minutes with a 1:200 dilution of acetic anhydride in 1 x TEA. The acetic anhydride probably neutralizes NH_2+ - groups which had been set free during the proteinase K treatment. The animals are washed in 1 x PBST for 5 minutes twice before they are fixed a second time with 4% formaldehyde. This second fixation step was performed for 20 minutes at room temperature. Once more the polyps were washed with 1 x PBST 5 times for 5 minutes. The animals were fixed for 30 minutes at 80 °C in 1 x PBST. They were washed once with a solution of 50% hybridization mix in 1 x PBST for 10 minutes before incubating the polyps in 100% hybridization mix for 10 minutes at room temperature. The animals were pre-hybridized in 100% hybridization mix for about 2 hours at 55 °C. After this step, the fixed animals can be stored at - 20 °C until use.

For hybridization, 10 - 15 animals per riboprobe were transferred to an Eppendorf tube. Hybridization with the riboprobes (0.2 ng/μl in 100% hybridization mix) was performed for 2.5 days at 55 °C.

The polyps were washed for minutes at 55 °C with decreasing concentrations (100%, 75%, 50% and 25%) of hybridization mix in 2 x SSC to remove unbound probe. Finally, the animals were incubated twice in 2 x SSC and 0.1% CHAPS for 30 minutes at 55 °C. The animals were washed twice with MAB for 10 minutes at room temperature on a rocker before blocking for 1 hour with MAB-B at room temperature. The animals were blocked for an additional 2 hours at room temperature with 1 x Blocking Solution in MAB. For detection of the DIG-labeled riboprobes, an anti-digoxigenin antibody conjugated to alkaline phosphatase (Fab fragment, Roche) was used. The polyps were incubated with a 1:4000 dilution of the antibody in 1 x Blocking Solution at 4 °C over night.

The animals were washed 8 times with MAB for 1 hour at room temperature on a rocker. Finally they were incubated with MAB over night at room temperature on a rocker.

The polyps were washed twice for 5 minutes at room temperature with NTMT followed by NTMT and Levamisol, respectively. The animals were transferred into a 24-well plate for staining. The staining reaction was either performed with NBT/BCIP or BMPurple in NTMT. NBT/BCIP or BMPurple serve as a substrate for the alkaline phosphatase; a purple indigo dye is formed and the transcript of the gene of interest can such be detected. 250 μl of substrate were added to each well. The polyps were stained for 30 minutes - 3 hours. The staining time is dependent on the riboprobe used and the unspecific background which is forming in the animal.

The staining reaction was stopped by removing the substrate and adding NTMT. The stained

polyps were incubated with 100% ethanol for 5 minutes until a colour change of the dye from purple to blue occured. The ethanol was removed and the polyps were washed with decreasing concentrations of ethanol (75%, 50% and 25%) for 5 min. Finally the animals were transferred into 1 x PBS.
The stained animals were mounted on microscope slides using 200 μl Mowiol per slide. After the Mowiol has hardened, the staining was examined under a microscope.

Solutions for in situ hybridization

Urethane:	4% urethane (w/v) in H_2O.
8% formaldehyde:	8% paraformaldehyde (w/v) in H_2O. Dissolve paraformaldehyde by heating to 60 °C for several hours until the solution clearifies. Adjust pH to 7.5 with 1 M sodium hydroxide. Always use freshly prepared formaldehyde.
10 x PBS:	1.5 M NaCl 0.08 M Na_2HPO_4 phosphate 0.02 M NaH_2PO4 Sterilize by autoclaving.
1 x PBST:	Dilute 100 ml 10 x PBS in 900 ml H_2O. Add 1 ml Tween20.
1000 x proteinase K:	10 mg/ml proteinase K in H_2O. Aliquot and store at - 80 °C.
10 x glycine:	40 mg/ml glycine in H_2O.
10 x TEA:	1 M triethanolamine in H_2O. Adjust pH to 7.8.
Hybridization mix:	50% formamide 5 x SSC 20 μg/ml yeast RNA 0.1% Tween20 0.1% Chaps 1 x Denhardt's 100 μg/ml heparin
20 x SSC:	3 M NaCl 0.3 M sodium citrate ad 900 ml H_2O and adjust the pH to 7.0. ad 1000 ml H_2O Sterilize by autoclaving.
Yeast RNA:	10 mg/ml yeast RNA in H_2O. Aliquot and store at - 20 °C.
1% Chaps:	1% Chaps (w/v) in H_2O.

50 x Denhardt's:
1% polyvinylpyrrolidone
1% ficol
1% bovine serum albumin (BSA), fraction V
Sterilize by filtration.
Aliquot and store at - 20 °C.

Heparin:
10 mg/ml in H_2O.
Aliquot and store at - 20 °C.

Maleic acid buffer (MAB):
100 mM maleic acid
150 mM NaCl
ad 900 ml H_2O and adjust the pH to 7.5.
ad 1000 ml H_2O
Sterilize by autoclaving.

MAB-B:
1% bovine serum albumine (fraction V)(w/v) in MAB.

10 x Blocking Solution:
10% Blocking Solution (Roche) (w/v) in MAB.
Aliquot and store at - 20 °C.

anti-digoxigenin antibody:
anti-digoxigenin antibody, alkaline phosphatase conjugated (Fab fragment, Roche) 1:4000 in 1 x Blocking Solution.

10 x sodium chloride:
1 M NaCl in H_2O.
Sterilize by autoclaving.

10 x Tris-HCl (pH 9.5):
1 M Tris in H_2O.
ad 900 ml H_2O and adjust the pH to 9.5.
ad 1000 ml H_2O
Sterilize by autoclaving.

10 x magnesium chloride:
0.5 M $MgCl_2$ in H_2O.
Sterilize by autoclaving.

NTMT:
100 mM sodium chloride
100 mM Tris-HCl (pH 9.5)
50 mM magnesium chloride
0.1% Tween20
Prepare shortly before use from stock solutions.

1 M levamisol:
1 M levamisol in H_2O.
Store at 4 °C.

NBT/BCIP substrate:
Add 50 μl NBT solution (Roche) and 37.5 μl BCIP (Roche) to 10 ml NTMT.
Prepare freshly before use.

BMPurple substrate:
Use BMPurple substrate (Roche) undiluted.

Mowiol:
Mix 6 g Mowiol with 6 g glycerol and dissolve for 1 hour at room temperature with frequent agitation.
Add 6 ml H_2O and stir for 1 hour at room temperature.
Add 12 ml 0.2 M Tris-HCl (pH 8.5) and incubate at 50 °C under periodical stirring until the Mowiol has dissolved.
Centrifuge for 30 minutes at 2000 g to remove undissolved material.
Aliquot the supernatant and store at - 20 °C.

3.5. Methods: DNA

3.5.1. Purification of plasmid DNA

DNA has been amplified in bacteria as 'high copy' (300 - >1000 copies per cell) or 'low copy' (5 - 20 copies per cell) extrachromosomal plasmid DNA. The plasmid DNA was isolated by alkaline lysis. The bacteria were lysed by the addition of sodium hydroxide. The cell membranes were dissolved and the proteins denatured. Neutralization led to precipitation of the denatured proteins and bacterial chromosomal DNA. The plasmid DNA, which stays in solution, was isolated.

DNA minipreparation

For the purification of plasmid DNA in a small scale, the Wizard Plus SV Minipreps DNA Purification System from Promega was used according to the manufacturer´s instruction. The system is based on DNA preparation by alkaline lysis. For plasmid DNA purification 5 ml LB medium with the appropriate antibiotics has been inoculated with a single colonie of *E.coli*-cells. The culture was incubated over night at 37 °C and 200 rpm in a shaker. The whole culture has been used for plasmid DNA purification. The bacteria were spun down for 5 minutes at full speed in a tabletop centrifuge. The bacterial pellet was resuspended in 250 μl Cell Resuspension Solution (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 μg RNaseA). 250 μl of Cell Lysis Solution (0.2 M NaOH, 1% SDS) and 10 μl of Alkaline Protease Solution were added. After incubation for 5 minutes at room temperature, cell lysis was stopped by the addition of 350 μl Neutralization Solution (4.09 M guanidine hydrochloride, 0.759 M potassium acetate, 2.12 M glacial acetic acid). The lysate was spun down for 10 minutes at top speed. The supernatant was decanted into a spin column with a silica-impregnated porous membrane to bind DNA. After centrifugation for 1 min at top speed, the bound DNA was washed with 750 μl of Wash Solution (60% ethanol, 60 mM potassium acetate, 8.3 mM Tris-HCl pH 7.5, 0.04 mM EDTA). The step was repeated with 250 μl of Wash Solution. After centrifuging for 2 minutes at top speed, the DNA was eluted with 250 μl - 250 μl H_2O. The purified plasmid DNA was analyzed by spectrophotometry and agarose gel electrophoresis.

3.5.2. Extraction and precipitation of DNA

Extraction of nucleic acids by phenol-chloroform

The phenol-chloroform extraction method is used to remove proteins from nucleic acids solutions (Chomczynski and Sacchi, 1987). Proteins are denatured by phenol and accumulate in the organic phase and at the interphase between the organic and the aqueous phase. Chloroform helps to seperate the two phases and isoamyl alcohol prevents the formation of foam.
An equivalent volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to the DNA solution. The solutions were vigourously mixed by shaking for 1 minute. The organic and the aqueous phase were seperated by centrifugation at top speed in a tabletop centrifuge for 5 minutes at room temperature. The upper, aqueous phase was transferred to a new tube and an equal volume of chloroform-isoamyl alcohol (24:1) was added. Again the solutions were mixed by shaking and the phases were seperated by centrifugation. The aqueous phase contains the nucleic acids.

Precipitation of DNA

Ethanol precipitation was used to purify and concentrate DNA from aqueous solutions. A mixture of salt and ethanol forces the DNA to precipitate out of solution by decreasing its solubility in water.
To an aqueous solution of DNA 1/10 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol were added. The DNA was centrifuged at top speed in a tabletop centrifuge for 20 minutes at room temperature. The pellet was washed with 70% ethanol. Finally the pellet was air-dried and dissolved in an appropriate amount of H_2O or TE-buffer.

3.5.3. Analysis of DNA

DNA was analyzed concerning both its quantity and quality. DNA quantity was determined by spectrophotometry, the quality was determined by gel electrophoresis. The DNA sequence was analyzed by automated sequencing.

Quantification of DNA

The concentration of DNA was determined by spectrophotometry using the Nanodrop spectrophotometer from Peqlab. 1 - 2 μl of DNA were used for analysis at an absorption of 260 nm. An OD_{260} of 1.0 equals a DNA concentration of 50 μg DNA per ml. (For RNA, an OD_{260} of 1.0 equals an RNA concentration of 40 μg RNA per ml.) Conclusions on the purity of the DNA were made by examining the ratio of OD_{260} to OD_{280}. A value between 1.8 and 2.0 represents a high-quality DNA sample.

Analysis of DNA by gel electrophoresis: agarose gel electrophoresis (AGE)

The quality and size of DNA was analyzed by agarose gel electrophoresis. Plasmid DNA preparations and PCR reactions were analyzed on 1% agarose gels using 1 x TAE or 1 x TBE buffer. To check the size of the DNA, additionally a DNA size marker has been loaded onto the gel. Electrophoresis was carried out at 100 mV in the same buffer which was used for the preparation of the gel. The DNA was visualised by staining for 5 minutes in an ethidium bromide solution and detection at 254 nm using a transilluminator. The gels were documented using the INTAS gel documentation system.

Solutions for gel electrophoresis

0.5 M EDTA:	186.12 g EDTA in 800 ml H_2O. Adjust pH to 8.0 with NaOH. Adjust volume to 1 litre with H_2O. Sterilize by autoclaving.
50 x TAE:	242 g Tris base 100 ml 0.5 M EDTA (pH 8.0) 57.1 ml glacial acetic acid Adjust volume to 1 litre with H_2O.

1 x TAE:	40 mM Tris acetate 1 mM EDTA (pH 8.0) Dilute 20 ml of 50 x TAE solution with 980 ml H_2O.
5 x TBE:	54 g Tris base 27.5 g boric acid 40 ml 0.5 M EDTA (pH 8.0) Adjust volume to 1 litre with H_2O.
1 x TBE:	44.5 mM Tris borate 2 mM EDTA (pH 8.0) Dilute 200 ml of 5 x TBE solution with 800 ml H_2O.
6 x DNA loading dye:	0.09% bromophenol blue (w/v) 0.09% xylencyanol (w/v) 60% glycerol (v/v) 60 mM EDTA (pH 8.0)
Ethidium bromide stock-solution:	10 mg/ml in H_2O. Working concentration: 100 ng/ml in H_2O.

Analysis of DNA by gel electrophoresis: polyacrylamide gel electrophoresis

DNA molecules of short length were resolved by polyacrylamide gel electrophoresis rather than agarose gel electrophoresis. The oligos used for band shift assays have been analyzed this way. For gel electrophoresis 10 - 12% polyacrylamide gels in 1 x TBE buffer have been prepared between two glass plates. Electrophoresis and detection of the DNA have been performed like agarose gel electrophoresis.

12% polyacrylamide gel:

H_2O	10 ml
40% polyacrylamide (acrylamide - bisacrylamide 29:1)	6 ml
5 x TBE	4 ml
10% ammonium persulfate	200 μl
TEMED	10 μl

Analysis of DNA by restriction cleavage

The DNA clones were analyzed by restriction enzyme cleavage, followed by agarose gel electrophoresis. To determine the correct insert size before sequencing, the plasmids were cut with two restriction enzymes simultaneously. The insert size of pGEM-T plasmids has been always analyzed by cleavage with ApaI and SacI, pET- and pFastBac1-clones have been analyzed by various enzymes. For analytical restriction enzyme cleavage, 300 ng of DNA have been incubated with an appropriate amount of enzyme in the recommended 1 x buffer in a total volume of 10 μl. After incubation for 1 hour at the appropriate temperature (usually 37 °C), 2 μl of 6 x DNA loading dye were added. The DNA fragments were analyzed by agarose gel electrophoresis. Clones of the correct insert size were further analyzed by sequencing.

(The restriction enzyme activity required to cut a certain amount of DNA can be exactly determined. A unit of restriction enzyme activity is the amount of restrition enzyme needed to completely cut 1 μg of bacteriophage lambda DNA in 1 hour. To cut a DNA fragment of x bp to completion, the following enzyme activity is required:

$$\text{enzyme activity [units]} = \frac{\text{x bp (size of DNA fragment)}}{\text{48000 bp (size of lambda genome)}} \cdot \frac{\text{number of restriction sites in the lambda genome}}{\text{number of restriction sites in the DNA fragment}})$$

DNA sequencing

DNA sequencing was carried out by the chain-termination method (Sanger et al., 1977). For the chain-termination reaction ddNTPs, each labeled with a different fluorescent dye - were used. The DNA sequence was detected by capillary electrophoresis. The DNA fragments of different length are seperated in a capillary filled with a gel matrix. At the exit from the gel matrix, the fluorescent dyes get excited by a laser and detected by a fluorescence detector. A data output is produced by a software in form of a chromatogram. All sequencing reactions were performed with the ABI PRISM 310 Genetic Analyzer from Applied Biosystems using the BigDye Terminator v1.0 Cycle Sequencing Kit.
For the sequencing reaction 200 - 400 ng of template DNA were mixed with 1 μl primer (5 pmol/μl) and 2 μl Big Dye in a total volume of 10 μl. The termination reaction was carried out in a thermal cycler (Gene Amp 2400, Perkin Elmer) using the following program:

initial denaturation	94 °C	5 minutes	
denaturation	96 °C	10 seconds	
annealing	58 °C	5 seconds	25 cycles
elongation	60 °C	4 minutes	

The reaction was precipitated using $\frac{1}{10}$ volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol. The DNA was centrifuged at top speed in a tabletop centrifuge for 20 minutes at room temperature. The pellet was washed with 250 μl of 70% ethanol. Finally the pellet was air-dried and dissolved in 25 μl H_2O. The DNA sample was automatically sequenced using the ABI PRISM 310 Genetic Analyzer.

3.5.4. Amplification of DNA by polymerase chain reaction (PCR)

Polymerase chain reaction was routinely used to amplify DNA templates for cloning, sequencing and for riboprobes. For all PCR reactions heat-stable Taq polymerases have been used: either the proof-reading Phusion High Fidelity DNA polymerase or the EuroTaq DNA polymerase.
For PCR amplification 1 μl of DNA was mixed with 1 μl 10 mM dNTPs, 10 pmol forward primer, 10 pmol reverse primer and 0.6 U of Phusion DNA polymerase in 1 x HF-Buffer or 1.25 EuroTaq polymerase in 1 x RB buffer containing 1.5 mM $MgCl_2$. The total reaction volume was 50 μl. Amplification was performed using the following cycling conditions:

initial denaturation	98 °C	3 minutes	
denaturation	98 °C	30 seconds	
annealing	X °C	30 seconds	x cycles
elongation	72 °C	1 minute / 1000 bp	
final elongation	72 °C	7 minutes	

Note: The annealing temperature depends on the melting temperature of the primers used.

All primers used for PCR amplification are listed in section 3.1.4 on page 40. The use of the primers and their respective melting temperatures are indicated.

Amplification of core fragments of the *Hydra* PcG genes by PCR

The *Hydra* Polycomb Group and associated genes have been identified by *in silico* analysis of the Hydra Genome Database followed by cloning of the genes. First, homologs of the *Hydra* Polycomb Group genes were identified by aligning protein sequences of Polycomb Group genes from other species with the *Hydra* genome (tBLASTn). Sequences with the highest similarity and the lowest E-value were double-checked by performing a BLASTp search with the most likely *Hydra* protein sequences against the NCBI database. Only if the result of the BLASTp search showed the correct protein, primers for the *Hydra* gene were designed. A fragment of the gene, which matches the sequence identified by the tBLASTn search was amplified from *Hydra* cDNA. The fragment does not correspond to the full length sequence of the gene, but to a part of the gene. The gene fragment was termed 'core fragment'.
For PCR amplification 1 μl of a 1:10 dilution of *Hydra* 3' RACE-cDNA was used. The core fragments of the genes were amplified using the following cycling conditions:

initial denaturation	98 °C	3 minutes	
denaturation	98 °C	30 seconds	
annealing	65 °C	30 seconds	5 cycles
elongation	72 °C	30 seconds	
denaturation	98 °C	30 seconds	
annealing	55 °C - 60 °C	30 seconds	25 cycles
elongation	72 °C	30 seconds	
final elongation	72 °C	7 minutes	

Usually the obtained amount of DNA has not been enough for further analysis of the DNA and for cloning. Therefore, the PCR reactions were reamplified with 1 μl of the initial PCR reaction. The core fragments of the genes were reamplified using the following cycling conditions:

initial denaturation	98 °C	3 minutes	
denaturation	98 °C	30 seconds	
annealing	55 °C - 60 °C	30 seconds	25 cycles
elongation	72 °C	30 seconds	
final elongation	72 °C	7 minutes	

Finally the PCR reactions have been purified, followed by cloning into pGEMT-vector and DNA sequencing of the amplified gene. The following core fragments of Polycomb Group associated genes have been amplified during this thesis:

gene	**length of the core fragment**	**region of the gene**
***Hy*YY1**	881 bp	The core fragment starts at amino acid position 33 of the full length gene and ends at amino acid 325.
***Hy*SCM**	1703 bp	The core fragment starts at amino acid position 180 of the full length gene and ends at amino acid 747.

RACE (rapid amplification of cDNA)-PCR

5'RACE- and 3'RACE-PCR have been performed to identify the 5' and 3' ends of the *Hydra* Polycomb Group genes. After the amplification of the core fragments, the full length genes and the 5' and 3' untranslated regions were identified by RACE-PCR followed by nested PCR.

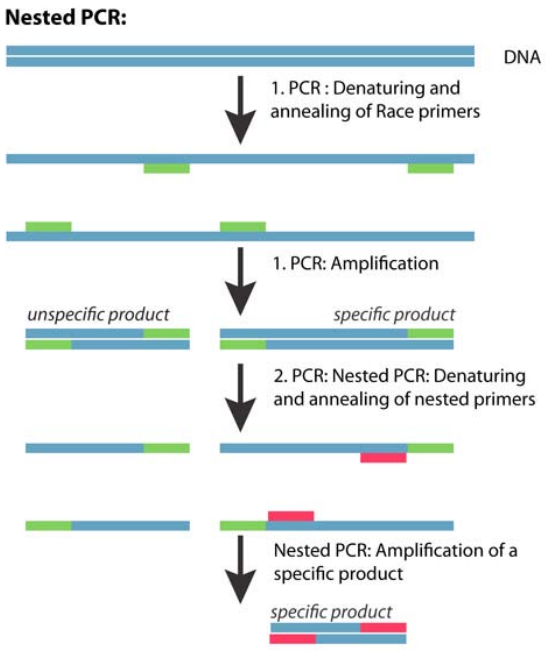


FIGURE 3.5 *Schematic overview of the amplification of a specific product by RACE-PCR followed by nested PCR.*

For PCR amplification 1 μl of a 1:10 dilution of *Hydra* 3′ or 5′ RACE-cDNA was used for PCR. The 5′ and 3′ ends of the genes were amplified using the same cycling conditions as for the amplification of the core fragments.
The PCR reactions have been used for reamplification of the DNA by nested PCR to reduce the background of unspecific products. For nested PCR, 1 μl of the RACE-PCR reaction was used for nested PCR. The nested PCR was carried out in the same way as the reamplification of the core fragments. The PCR reactions have been purified, followed by cloning into pGEMT-vector and DNA sequencing of the amplified fragment. The following 5′ and 3′ ends of Polycomb Group associated genes have been amplified:

gene	length of the fragment	5′ or 3′ end of the gene
***Hy*YY1**	185 bp	The 5′ end includes the first 42 amino acids of the protein coding region and spans the 5′ untranslated region of the gene.
***Hy*YY1**	632 bp	The 3′ end includes the last 55 amino acids of the protein coding region and spans the 3′ untranslated region of the gene and its 3′ end polyadenylation site.
***Hy*SCM**	203 bp	The 5′ end includes the first 186 amino acids of the protein coding region and spans the 5′ untranslated region of the gene.
***Hy*SCM**	203 bp	The 3′ end includes the last 33 amino acids of the protein coding region and spans the 3′ untranslated region of the gene and its 3′ end polyadenylation site.

Amplification of the full length *Hydra* PcG genes by PCR

After the amplification, cloning and sequencing of the core fragments, 5′ and 3′ ends of the *Hydra* genes, the full length coding region was amplified by PCR.
For PCR amplification 1 μl of a 1:10 dilution of *Hydra* 3′ RACE-cDNA was used for PCR. The full length coding region of the genes were amplified using the following cycling conditions:

initial denaturation	98 °C	3 minutes	
denaturation	98 °C	30 seconds	
annealing	60 °C	45 seconds	30 cycles
elongation	72 °C	30 seconds	
final elongation	72 °C	7 minutes	

The PCR reactions have been purified, followed by cloning into pGEMT-vector and DNA sequencing of the amplified full length gene. The following full length coding regions have been amplified:

gene	length	region
***Hy*YY1**	1131 bp	The fragment corresponds to the full length protein coding region of the gene.
***Hy*SCM**	2298 bp	The fragment corresponds to the full length protein coding region of the gene.

PCR-mediated restriction site engineering

For cloning of the *Hydra* Polycomb Group and associated genes in other vectors than pGEMT, new restriction sites have been attached to both ends of the genes by PCR-mediated restriction site engineering. The desired restriction sites were incorporated into the primers used for PCR. This method avoids blunt-end ligation of PCR products. The PCR product can be ligated into a vector in the favoured direction, which is especially important for expression vectors. Appropriate primer sites were added at the ends of gene specific primers. Since restriction endonucleases often fail to cleave these sites if they are located directly at the end of the DNA molecule, additional bases have been added to the end of the restriction sites.
The PCR was carried out at the same conditions just like the PCR amplification of the full length genes. Since the restriction sites do not anneal to the initial DNA template, the annealing temperature was calculated but for the matching part of the primers. As a template for the PCR, the pGEMT-vector containing the full length gene has been used. 1 ng DNA was mixed with 1 μl 10 mM dNTPs, 10 pmol of gene specific forward primer containg a new restriction site at its end, 10 pmol of gene specific reverse primer containing a new restriction site at its end and 0.6 U of Phusion DNA polymerase in 1 x HF-Buffer.
This method was used for adding additional histidine tags to the N-terminus of the *Hydra* Polycomb Group genes, as well.

Colony-PCR

Colony-PCR was performed to check the outcome of a ligation reaction in a simple and fast way. A DNA insert can be amplified from recombinant bacteria containing the desired DNA plasmid with insert specific primers. After the bacteria have been transformed with a ligation reaction and selected on agar plates containing antibiotics, single colonies have been used as template for PCR. A single colony has been transferred into a reaction tube with the help of a sterile tooth pick. After the tooth pick has been twisted in the reaction mixture, it has been transferred into 5 ml LB medium containing antibiotics. The bacterial colony has been transferred into the following reaction mixture: 1.5 mM $MgCl_2$, 1 μl of 10 mM dNTPs, 10 pmol forward and reverse primer and 1.25 units of EuroTaq DNA polymerase in 1 x RB buffer in a total volume of 50 μl. The following cycling conditions have been used: The PCR reaction was analyzed by agarose gel electrophoresis. In case a bacterial colony contained the desired DNA plasmid, the cells were cultured at 37 °C and 200 rpm over night and the plasmid DNA was isolated.

initial denaturation	95 °C	5 minutes	
denaturation	95 °C	30 seconds	
annealing	55 °C	30 seconds	30 cycles
elongation	72 °C	30 seconds	
final elongation	72 °C	7 minutes	

Note: Usually these cycling conditions were used for colony-PCR. In case a longer DNA fragment had to be amplified, both the annealing and elongation time were increased.

Purification of PCR-products

PCR products were purified using the Wizard SV Gel and PCR Clean-Up System from Promega according to the manufacturer's instructions. The PCR reaction was mixed with an equal volume of Membrane Binding Solution (4.5 M guanidinium isothiocyanate, 0.5 M potassium acetate pH 5.0) and incubated for 1 minute in a minicolumn with a silica-impregnated porous membrane to bind DNA. After centrifugation for 1 minute at top speed in a tabletop centrifuge, the membrane-bound DNA was washed with 700 μl of Membrane Wash Solution (10 mM potassium acetate, 80% ethanol, 16.7 μM EDTA). The wash solution was spin down for 1 minute at top speed and the membrane was washed for a second time with 500 μl of Membrane Wash Solution. After 5 minutes of centrifugation at top speed, the DNA was eluted with 30 - 50 μl H_2O. The purified plasmid DNA was analyzed by spectrophotometry and agarose gel electrophoresis.

3.5.5. Cloning

The *Hydra* Polycomb Group genes were cloned into diverse vectors for different purposes. The *Hydra* YY1 gene has been cloned into pET-vectors for recombinant protein expression and into the baculovirus expression vector pFastBac1 for the production of recombinant baculoviruses. For cloning, new restriction sites have been attached to both ends of the gene by PCR (see page 73). The obtained fragments were cut with the respective enzymes and ligated into the target vector.

Cloning into pGEM-T vector

The pGEM-T-vector was chosen for rapid cloning of PCR products. This vector exists as a linear DNA molecule with thymidine overhangs at both 3′ termini. Adding compatible deoxyadenosine overhangs to the PCR product allows the easy ligation into this vector. The pGEM-T-vector contains SP6 and T7 RNA polymerase promoters flanking its multiple cloning site. DNA fragments cloned into this vector have been used for the determination of the DNA sequence of the *Hydra* Polycomb Group genes by sequencing. Furthermore the inserted DNA has been used as a template for riboprobe production for in situ hybridization. Deoxyadenosine overhangs were added to purified PCR products by A-tailing. Some DNA polymerases add a single deoxyadenosine overhang in a template independent manner. However, the Polycomb Group genes have

been amplified using the Phusion High Fidelity proof-reading DNA polymerase which does not add a deoxyadenosine. Hence, the purified PCR product was A-tailed using the EuroTaq DNA polymerase. 800 ng purified PCR product were mixed with 0.3 μl of 10 mM dNTPS, 1.5 mM $MgCl_2$ and 0.15 units of EuroTaq DNA polymerase in 1 x RB buffer in a total volume of 15 μl. A-tailing was performed at 72 °C for 30 minutes in a thermal cycler. The A-tailed DNA molecule was ligated into pGEM-T-vector.

Ligation

PCR products were ligated into an adequate target vector with the help of the T4 DNA Ligase. During the ligation reaction, two ends of DNA molecules are covalently linked. The ligation reaction was carried out using a three times molar excess of insert DNA compared to vector DNA. The amount of insert DNA needed was calculated as follows:

$$\text{amount of insert DNA [ng]} = \frac{(\text{amount of vector DNA [ng]} \cdot \text{length of insert DNA [bp]})}{(\text{length of vector DNA [bp]})} \cdot \frac{3}{1}$$

The appropriate amounts of vector DNA and insert DNA were mixed with 3 units of T4-DNA-Ligase in 1 x Ligation buffer in a total volume of 10 μl. The ligation reaction was incubated over night at 4 - 16 °C. 4 μl of the ligation reaction were directly transformed into chemocompetent *E.coli*-DH5α. The cells were plated on LB agar plates containing the appropriate antibiotics. Cells transformed with pGEM-T were plated on AXI-plates. Insertion of a DNA fragment into the multiple cloning site of the pGEM-T-vector interrupts the coding sequence of the lacZ$alpha$ gene and disrupts the production of functional β-galactosidase, which leads to white coloured colonies on AXI-plates. Positive colonies containing the correct plasmid DNA were verified by colony-PCR or DNA isolation followed by restriction cleavage analysis.

3.6. Methods: Proteins

3.6.1. Bacterial expression of recombinant *Hy*YY1

The *Hy*YY1 protein was recombinantly expressed for different purposes. The protein was expressed as antigen for antibody production and purification. Furthermore, the protein was used for gel retardation assays. The prokaryotic vectors pET15b(+) and pET21(+) provide the basis for recombinant protein expression in bacteria. The pET15b(+)-vector allows the expression of an N-terminal 6His-fusion protein, whereas the pET21(+)-vector provides a C-terminal 6His-tag. pET21(+) is a transcription vector, which lacks the ribosomal binding site and the ATG start codon for efficient translation. For protein expression, the ribosomal binding site has been cloned into the vector, together with the *Hy*YY1 DNA. The *Hy*YY1 DNA and a truncated version lacking the C-terminal zinc finger domains have been cloned into both vectors. The proteins have been expressed as 6His-fusion proteins, which allows purification by metal ion affinity chromatography. Electrocompetent *Bl21DE3*- and *Rosettagami DE3 pLysS*-cells have been transformed with the plasmid DNA, respectively. The transformed cells were used to inoculate a liquid preculture, which was incubated in a shaker at 37 °C over night. 500 ml LB-medium containing 100 μg/ml ampicillin were inoculated with the preculture at an OD_{600} of 0.05 (500 ml : (OD_{600} preculture / OD_{600} 0.05) = x = volume needed to inoculate the expression culture). The expression culture was incubated in a shaker at 37 °C until an OD_{600} of 0.5 - 0.7 was reached. Recombinant protein synthesis was induced by the addition of IPTG at a final concentration of 1 mM. An induction with the lactose-analog IPTG is possible since the pET-expression vectors contain an inducible lac operon. The expression cultures have been incubated for 4 hours at either 37 °C (*Hy*YY1-fusion) or at 25 °C (*Hy*YY1ΔZnF-fusion). The bacterial cells were centrifuged at 5000 g and 4 °C for 10 minutes. The supernatant was discarded and the pellet was immediately used for protein purification or stored at -20 °C until use.
1 ml samples of the expression culture have been taken after 0, 1, 2, 3 and 4 hours to analyze the protein expression by SDS-PAGE. Equal amounts of cells were loaded onto the gel, to determine the progression of protein expression. The OD_{600} of each sample was determined by spectrophotometry. The bacterial cells were spun down for 1 minute at top speed in a tabletop centrifuge. The pellet was resuspended in ($OD_{600} \cdot 100$) μl of 1 x PBS. An equal volume of 1 x protein sample buffer was added. 8 μl of each sample were seperated by SDS-PAGE.

3.6.2. Purification of recombinant *Hy*YY1

The recombinant fusion proteins have been purified by metal ion affinity chromatography either under native or under denaturing conditions.

Protein purification under native conditions

The bacterial cell pellet was resuspended in 10 ml of a phosphate buffer (500 mM NaCl, 50 mM NaH_2PO_4, pH 7.8). The cells were lysed by sonication using a Branson sonifier 200 equipped with a microtip. Sonication was performed on ice for 3 times 60 seconds with 1 minute intervals on ice using the following settings: power 5, duty cycle 40%. The lysate was centrifuged for 30 - 45 minutes at 10000 g and 4 °C in a Sorvall RC-5B centrifuge (SA 600 rotor). The cleared lysate was transferred to a screw cap tube and mixed with 2 ml of a 50% Ni-NTA suspension, which has been

equilibrated with phosphate buffer. (Note: The insoluble pellet was used for protein purification under denaturing conditions.) The lysate was incubated with the agarose beads at 4 °C for 2 hours on a roller mixer to allow binding of the 6His-fusion protein to the beads. The suspension was transferred to a disposable plastic column and incubated for 10 minutes at 4°C until the bed of agarose beads has settled. The flow through was discarded. The column was subsequently washed with 20 ml of phosphate buffer containing increasing amounts of imidazole (0 mM, 20 mM, 50 mM, 100 mM). After the last wash step, the protein was incubated with 5 ml of phosphate buffer containing 250 mM imidazole for 30 minutes. The protein was eluted and fractions of 0.5 ml were collected.
The protein fractions were seperated by SDS-PAGE to monitor the purity of the preparation. The protein concentration was determined by spectrophotometry.

Protein purification under denaturing conditions using imidazole

The bacterial cell pellet was resuspended in 10 ml of lysis buffer containing guanidin (6 M guanidin-HCl, 100 mM NaH_2PO_4, 10 mM imidazole, 1 mM β-mercaptoethanol, pH 8.0). The cells were lysed for 30 minutes at room temperature on a roller mixer. The resultant lysate is very viscous and the DNA was sheared by sonication using a Branson sonifier 200 equipped with a microtip. Sonication was performed on ice for 3 times 10 seconds using the following settings: power 5, duty cycle 40%. The lysate was centrifuged for 30 - 45 minutes at 10000 g and 4 °C in a Sorvall RC-5B centrifuge (SA 600 rotor). Centrifugation was performed to remove bacterial remnants to avoid clogging of the column. The supernatant was transferred to a screw cap tube and mixed with 2 ml of a 50% Ni-NTA suspension, which has been equilibrated with lysis buffer. All further purification steps were performed in the same way as for the purification under native conditions, except that the lysis buffer was used as basis for the wash buffers and the elution buffer.

Protein purification under denaturing conditions using a stepwise decrease of the pH

The bacterial cell pellet was resuspended in 10 ml of lysis buffer pH 8.0 (6 M guanidin-HCl, 100 mM NaH_2PO_4, 10 mM Tris) and lysed over night at 4 °C on a roller mixer. The lysate was centrifuged for 30 - 45 minutes at 10000 g and 4 °C in a Sorvall RC-5B centrifuge (SA 600 rotor). The cleared lysate was transferred to a screw cap tube and mixed with 2 ml of a 50% Ni-NTA suspension, which has been equilibrated with lysis buffer. The lysate was incubated with the agarose beads at 4 °C for 2 hours on a roller mixer to allow binding of the 6His-fusion protein to the metal beads. The suspension was transferred to a disposable plastic column and incubated for 10 minutes at 4°C until the bed of agarose beads has settled. The flow through was discarded. The column was washed twice with 10 ml of lysis buffer pH 8.0. The washing steps were repeated with lysis buffer pH 6.3. After the last wash step, the protein was eluted with 5 ml of lysis buffer pH 5.9, followed by elution with 5 ml lysis buffer pH 4.5.
The protein fractions were seperated by SDS-PAGE to monitor the purity of the preparation. The protein concentration was determined by spectrophotometry.

3.6.3. Buffer exchange of recombinant proteins

The buffer of the purified *Hy*YY1 proteins were exchanged either by dialysis or by desalting using a PD10 column. For immunization and gel retardation assays, the proteins were buffered in 1 x

PBS.

Desalting using a PD10 column

To exchange the buffer of the purified *Hy*YY1 proteins and to remove remaining salts like imidazole and/or guanidinium, a PD10 desalting column was used according to the manufacturer's instructions (GE Healthcare). The PD10 columns are packed with a Sephadex G-25 medium, which allows the separation of high molecular weight substances (proteins) from low molecular weight substances (salts) by chromatography.
The desalting column was equilibrated with 25 ml 1 x PBS, the flow-through was discarded. 2.5 ml of sample were added to the column and the flow-through was discarded. The protein was eluted from the column with 3.5 ml of 1 x PBS. Fractions of 0.5 ml were collected and the protein concentration of each fraction was determined by OD_{280} measurement and Bradford assay. The first two fractions usually contain little protein and were discarded.

Dialysis

The buffers of the *Hy*YY1 proteins, purified under denaturing conditions using a decreasing pH, were exchanged by a stepwise dialysis procedure.
The proteins were transferred to a dialysis tubing (MWCO 5000) and dialyzed for 1 hour at 4 °C against two changes of dialysis buffer pH 6.3 (1 M guanidin-HCl, 100 mM NaH_2PO_4, 10 mM Tris). The pH was further increased by dialyzing the protein for 1 hour at 4 °C against two changes of dialysis buffer pH 8.0 (0.1 M guanidin-HCl, 100 mM NaH_2PO_4, 10 mM Tris). Finally, the protein was dialyzed for 1 hour at 4 °C against two changes of 1 x PBS. The protein solution was transferred to a 1.5 ml reaction tube. The protein concentration was determined by OD_{280} measurement and Bradford assay. Furthermore, the protein was analyzed by SDS-PAGE. The protein was used for gel retardation assay, immediatelly.
Preparing the membranes for dialysis:
To remove contaminants like heavy metals and sulfides, which may arise during the production process, the dialysis tubings were rinsed before use. The dialysis tubings were heated in a carbonate solution (10 mM EDTA, 10% sodium carbonate or bicarbonate) for 10 minutes at 60 - 70 °C. The tubings were washed in H_2O. They were again heated in the carbonate solution and washed in H_2O. Then, the tubings were heated to 70 °C in H_2O, twice. The tubings were stored in 70% ethanol at 4 °C.

3.6.4. Preparation of protein extracts from *Hydra magnipapillata*

Originally, the *Hydra* protein extracts were prepared for the establishment of an *in vitro transcription* system in *Hydra*. The protein extracts were used for antibody characterization and to determine the cellular compartment, in which the PcG proteins are located.

Preparation of Whole Cell Extracts (WCEs)

Hydra WCEs were prepared as described by Manley et al. (1980, 1983), yet on a smaller scale. All steps were carried out at 4 °C and on ice. 5 ml of densly packed *Hydra* (∼5000 polyps) were washed twice with Hydra Medium and twice with 1 x PBS. The animals were centrifuged for 5

minutes at 2000 g in a Beckman GS-6KR centrifuge. The resultant *Hydra* pellet was resuspended in four pellet volumes of 0.01 M Tris-HCl (pH 7.9 at 4 °C), 1 mM EDTA and 5 mM DTT. After 20 minutes, the cells were lysed in a Dounce homogenizer with 20 strokes using a B-type pestle. Four pellet volumes of 0.05 M Tris-HCl (pH 7.9 at 4 °C), 0.01 M $MgCl_2$, 2 mM DTT, 25% sucrose and 50% glycerol were added and the suspension was mixed gently. One pellet volume of saturated $(NH_4)_2SO_4$ was added in a dropwise fashion. Due to nuclear lysis, the suspension becomes more and more viscous during the procedure. The lysate was stirred for 30 minutes. The extract was centrifuged in a Beckman LB 55 ultracentrifuge at 45000 rpm in a SW 40Ti rotor for 3 hours. The supernatant was decanted. Proteins and nucleic acids were precipitated by the addition of solid $(NH_4)_2SO_4$ (0.33 g/ml). After the $(NH_4)_2SO_4$ has dissolved, 1 M NaOH (0.1 ml/10 g solid $(NH_4)_2SO_4$) was added and the suspension was stirred for 30 minutes. The suspension was centrifuged at 15000 g for 20 min in a SA600 rotor in a Sorvall RC5C centrifuge. The precipitate was resuspended with 5% of the volume of the supernatant with 0.025 M HEPES (pH 7.9), 0.1 M KCl, 0.012 M $MgCl_2$, 0.5 mM EDTA, 2 mM DTT and 17% glycerol. The suspension was dialyzed against three changes of the resuspension buffer for 2 hours. The dialyzate was centrifuged at 10000 g for 20 minutes to remove insoluble material. The supernatant was aliquoted and stored at -80 °C. The protein concentration was determined by Bradford assay and OD_{280} measurement. The protein concentration varies from extraction to extraction and is ranging from $\sim$2 mg/ml to $\sim$8 mg/ml.

Preparation of Nuclear Extracts (NEs) and Cytoplasmatic Extracts (CEs)

Hydra nuclear and cytoplasmatic extracts were prepared as described by (Dignam et al., 1983a,b). All steps were carried out at 4 °C and on ice. 5 - 15 ml of densly packed *Hydra* ($\sim$5000 - 15000 polyps) were washed twice with Hydra Medium and twice with 1 x PBS. The animals were centrifuged for 5 minutes at 2000 g in a Beckman GS-6KR centrifuge. The resultant *Hydra* pellet is resuspended in three pellet volumes of buffer A (10 mM HEPES (pH 7.9 at 4° C), 1.5 mM $MgCl_2$, 10 mM KCl, 0.25 mM DTT, 0.2 mM PMSF, 1 μM Pepstatin A, 2 x Roche complete EDTA-free). After 10 minutes, the cells were homogenized in a Dounce homogenizer with 20 strokes using a B-type pestle. The homogenate was filtered using a gauze (Miracloth) of 22 - 25 μm mesh size. The suspension was centrifuged at 3500 g for 15 min in a SA600 rotor in a Sorvall RC5C centrifuge. The resulting supernatant was used for the preparation of cytoplasmatic extracts, the pellet was used for the preparation of nuclear extracts. The supernatant was mixed with 0.1 volume of 10 x buffer B (0.3 M HEPES (pH 7.9 at 4° C), 1.4 M KCl, 30 mM EDTA, 0.25 mM DTT, 0.2 mM PMSF, 1 μM Pepstatin A, 2 x Roche complete EDTA-free) and centrifuged for 1 hour at 45000 rpm in a Beckman LB 55 ultracentrifuge in a SW 40Ti rotor. The supernatant is the cytoplasmatic extract. The cytoplasmatic extract was dialyzed like the nuclear extract. The nuclear pellet was resuspended in $^1/_2$ pellet volume of low salt buffer (20 mM HEPES (pH 7.9 at 4° C), 25% glycerol(v/v), 0.02 M KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.25 mM DTT, 0.2 mM PMSF, 1 μM Pepstatin A, 2 x Roche complete EDTA-free). $^1/_2$ pellet volume of high salt buffer (20 mM HEPES (pH 7.9 at 4° C), 25% glycerol (v/v), 1.6 M KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.25 mM DTT, 0.2 mM PMSF, 1 μM Pepstatin A, 2 x Roche complete EDTA-free) was added in a dropwise fashion to prevent clumping. The nuclei were extracted for 30 minutes with gentle stirring or mixing on a roller mixer. The suspension was centrifuged at 10000 g for 30 minutes. The supernatant was dialyzed against three changes of dialysis buffer (20 mM HEPES (pH 7.9 at 4° C), 20% glycerol (v/v), 0.1 M KCl, 0.2 mM

EDTA, 0.25 mM DTT, 0.2 mM PMSF, 1 μM Pepstatin A, 10 μM Leupeptin, 10 μM Aprotinin, 1 mM Benzamidine). The dialyzate was centrifuged for 20 minutes at 10000 g to remove insoluble material. The supernatant was aliquoted and stored at -80 °C. The concentration of the extracts is determined by Bradford assay and OD_{280} measurement. The protein concentration varies from extraction to extraction and is ranging from $\sim$2 mg/ml to $\sim$10 mg/ml.

3.6.5. Quantification of proteins

Bradford assay

The protein concentration of purified *Hy*YY1 protein and *Hydra* protein extracts were determined according to Bradford (1976). The Bradford assay is a colorimetric protein assay, which is based on an absorbance shift of the dye Coommassie Brilliant Blue G-250 upon binding to protein. The color of the dye changes from a brownish to a bluish color, the absorbance maximum shifts from 465 nm to 595 nm. Most of the buffer systems are compatible with the Bradford assay.
For each measurement, a complete calibration curve with BSA-standards ranging from 0 to 20 μg protein has been generated. The Bradford assay is almost linear over this range. The BSA-standards for calibration were measured in triplicate. For the determination of the protein concentration, a total of 800 μl sample in H_2O has been mixed with 200 μl of 5 x Bio-Rad Protein Assay Dye Reagent Concentrate. After incubation for 5-30 minutes at room temperature, the absorption at 595 nm has been determined using the BioRad SmartSpecPlus spectrophotometer. The protein concentration of the samples was calculated with the help of the calibration curve.

OD_{280} measurement

The protein concentration of purified proteins and cell extracts was determined by measuring the absorbance at 280 nm in a spectrophotometer versus a suitable control. An absorbance of 1.0 at 280 nm equates a protein concentration of 1 mg/ml. This method provides a rough approximation of the protein concentration.

3.6.6. SDS-PAGE

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) is a technique, which is used to separate proteins by size in a discontinuous gel system under denaturing conditions (Laemmli, 1970). Due to binding of SDS, the proteins are charged identically in proportion to their mass. The proteins are negatively charged and migrate to the anode during electrophoresis. For protein separation, 10 - 14% resolving gels and 5% stacking gels have been cast between two glass plates. After polymerization, the protein samples and a prestained molecular weight marker ('PageRuler Prestained Protein Ladder' from Fermentas or 'Prestained Protein Ladder, Broad Range' from NEB) have been loaded onto the gel. The molecular weight marker was used to monitor the progress of the electrophoretic run and to calculate the molecular weights of the samples. Gel electrophoresis was carried out at 20 mA/gel for 1.5 hours in 1 x SDS-running buffer. Afterwards, the polyacrylamide gel was stained with Coomassie Brilliant Blue G-250, or the gel was used for Western blotting.

Gels for SDS-PAGE (amounts are calculated for two gels of $\sim$ 9 x 8 cm)

10 / 12 / 14% resolving gel:

H_2O	9.8 / 8.8 / 7.8 ml
1.5 M Tris-HCl pH 8.8	5 ml
40% polyacrylamide (acrylamide - bisacrylamide 29:1)	5 / 6 / 7 ml
10% SDS	200 μl
10% APS	200 μl
TEMED	10 μl

5% stacking gel:

H_2O	4.3 ml
1.5 M Tris-HCl pH 6.8	0.75 ml
40% polyacrylamide (acrylamide - bisacrylamide 29:1)	0.75 ml
10% SDS	60 μl
10% APS	60 μl
TEMED	6 μl

Sample preparation for SDS-PAGE

Protein samples for SDS-PAGE have been mixed with an equal volume of 2 x SDS-PAGE sample buffer and boiled at 95 °C for 5 - 10 minutes shortly before loading onto the gel.
Protein samples in denaturing guanidinium buffer have been precipitated first to remove the salt: 225 μl of ice cold 100% ethanol have been added to 25 μl of protein sample. After vortexing for 10 seconds, the sample has been placed on -20 °C for 5 - 10 minutes. To precipitate the protein, the sample has been centrifuged for 5 minutes in a tabletop centrifuge at top speed. The supernatant was decanted and the pellet has been washed with 250 μl of 90% ethanol. After vortexing, the sample has been centrifuged again and the supernatant has been decanted carefully. The pellet has been resuspended in 25 μl of 1 x SDS-PAGE sample buffer and boiled for 5 - 10 minutes at 95 °C prior to loading onto the gel.

Rapid Coomassie staining

The dye Coomassie Brilliant Blue binds unspecifically to proteins and can be used to stain proteins (Wilson, 1983). For staining of proteins in polyacrylamide gels, the rapid Coomassie staining described by Sasse and Gallagher (2009) has been used.
After SDS-PAGE, the polyacrylamide gel was soaked in fixation solution (25% isopropanol, 10% acetic acid) and incubated for 30 minutes at room temperature. The gel was transferred to the staining solution (10% acetic acid, 0.006% Coomassie Brilliant Blue G-250), and staining was performed until distinct protein bands became clearly visible on the gel. The gel was documented using the INTAS gel doc system.

Solutions for SDS-PAGE and Coomassie staining

1.5 M Tris-HCl pH 8.8:	Dissolve 181.71 g Tris base in 900 ml H_2O. Adjust pH to 8.8 with HCl. Adjust volume to 1 litre with H_2O. Sterilize by autoclaving.
1.5 M Tris-HCl pH 6.8:	Dissolve 181.71 g Tris base in 900 ml H_2O. Adjust pH to 6.8 with HCl. Adjust volume to 1 litre with H_2O. Sterilize by autoclaving.
10% SDS:	Dissolve 10 g SDS in 90 ml H_2O. Adjust volume to 0.1 litre with H_2O. Store at room temperature.
10% APS:	Dissolve 0.1 g APS in 1 ml H_2O. Prepare freshly prior to use.
10 x SDS-PAGE running buffer:	10 g SDS 30.3 g Tris base 144.1 g glycine Dissolve in 800 ml H_2O. Adjust volume to 1 litre with H_2O.
2 x SDS-PAGE sample buffer:	10 ml 1.5 M Tris-HCl (pH 6.8) 12 ml 10% SDS 30 ml glycerol 15 ml β-mercaptoethanol 1.8 mg bromophenol blue Adjust volume to 100 ml with H_2O. Aliquot and store at -20 °C.
Fixation solution:	25% isopropanol 10% acetic acid
Staining solution:	10% acetic acid 0.006 g Coomassie Brilliant Blue G-250

3.7. Methods: Immunological techniques

3.7.1. Antibody production

Polyclonal antibodies against the *Hydra* PcG proteins were produced in New Zealand white rabbits at the IBF (Interfakultäre Biomedizinische Forschungseinrichtung = Interdisciplinary Research Facility for Biomedicine) at the university of Heidelberg. Handling of the animals and all injections were performed by the local staff. The rabbits were injected with 100 μg of purified recombinant *Hy*YY1 protein in PBS. The immune response was enhanced by mixing the protein antigen with Freund's adjuvant complete for the initial immunization and Freund's adjuvant incomplete for boosting. A total volume of 300 μl was injected intradermally in the back of the animal at several injection sites. Four weeks after the initial immunization, the rabbit has been injected regularly (every two weeks) to boost the immunization. Boosting was performed in the same way as the initial immunization. To monitor the progress of antibody production, samples of blood have been taken at regular intervals from the marginal ear vein and the serum was separated from the whole blood. The sera have been used for immunoblotting to probe different amounts of recombinantly expressed *Hy*YY1 protein and *Hydra* protein lysates.
In total, three rabbits were injected with recombinantly expressed *Hy*YY1 protein. Two animals were injected with purified *Hy*YY1ΔZnF, one animal was injected with the purified full length *Hy*YY1 protein. The antiserum obtained from the first rabbit injected with *Hy*YY1ΔZnF was affinity purified and used for all further studies.

Separation of serum from whole blood

The fresh blood samples were incubated at room temperature for 1 hour to clot. The clotted blood was separated from the wall of the reaction tube with the help of a clean pipette. The sample was incubated at 4°C over night to allow the contraction of the clotted blood. The serum was separated from the red blood cells and coagulation factors by centrifugation for 30 minutes at 3000 rpm (Beckman GS-6KR; swing-out rotor). The serum (supernatant) was transferred to a new reaction tube and centrifuged again for 30 minutes to remove residual red blood cells. The serum was aliquoted and stored at -20°C until use.

3.7.2. Antibody purification

The α*Hy*YY1 antibody was affinity purified using an antigen column. This method allows the purification of antigen-specific antibodies from a preparation of polyclonal antibodies. The pure antigen is bound to a solid support and the antibodies in the preparation, which are specific for the antigen will bind. Unspecific antibodies are removed by different washing steps.

Antigen coupling

Purified *Hy*YY1ΔZnF was coupled to Affi-Gel 10 Gel according to the manufacturer's instructions (BioRad). Affi-Gel 10 Gel consists of activated agarose beads containing a reactive ester (N-hydroxysuccinimide). This ester is highly selective for primary amino groups (NH_2), which form a stable amide bond with the terminal carboxyl group (COOH) of the spacer.

All following steps were performed at 4 °C in a cold room. 1 ml of Affi-Gel 10 Gel suspension was loaded into a disposable plastic column. After the gel has settled, the column was washed three times with 10 ml ice cold H_2O. The coupling efficiency of the protein is dependent on its isoelectric point (pI). Affi-Gel 10 is recommended for neutral to basic proteins with pIs ranging from 6.5 to 11. However, *Hy*YY1ΔZnF has a pI of 5.75. The coupling efficiency was enhanced by manipulating the coupling conditions by altering the pH and the salt concentration. $\sim$ 1.5 mg of purified protein were added to the column in a buffer containing 0.1 M MOPS pH 7.5 and 80 mM $CaCl_2$. The protein was incubated with the matrix for 4 hours at 4 °C. Uncoupled protein was removed by several washing steps with the same buffers used for antibody purification. The gel matrix was washed with 50 ml ice cold H_2O, followed by 20 ml ice cold 10 mM Tris-HCl (pH 7.5), 20 ml 100 mM glycine (pH 2.5) and 20 ml ice cold 10 mM Tris-HCl (pH 8.8). The pH was checked and the matrix was washed with 20 ml ice cold 100 mM triethylamine (pH 11.5). The pH was neutralized with 20 ml ice cold 10 mM Tris-HCl (pH 7.5). The gel was stored in 10 mM Tris-HCl (pH 7.5). For long-term storage, 0.2% sodium azide were added. To detect the coupling efficiency, samples of the different fractions during the coupling procedure were seperated by SDS-PAGE.

Antibody affinity purification

The antibody was purified according to a protocol in "Antibodies: A Laboratory Manual" (Harlow and Lane, 1988). 1 ml of the α*Hy*YY1 final bleed was diluted 1 in 10 in 10 mM Tris-HCl (pH 7.5). The antibody was passed through the column three times to ensure proper binding. The antigen-column was washed with 20 ml 10 mM Tris-HCl (pH 7.5), followed by 20 ml of 0.5 M NaCl and 10 mM Tris-HCl (pH 7.5). Antibodies, which were bound by acid-sensitive interactions were eluted with 10 ml 100 mM glycine (pH 2.5). 0.5 ml fractions of the eluate were collected in a tube containing an equal volume of 1 M Tris-HCl (pH 8.0) to neutralize the pH. The column was washed with 10 mM Tris-HCl (pH 8.8). The pH was checked and the antibodies, which were bound by base-sensitive interactions, were eluted with 10 ml 100 mM triethylamine (pH 11.5). 0.5 ml fractions of the eluate were collected in a tube containing an equal volume of 1 M Tris-HCl (pH 8.0). The column was washed with 10 mM Tris-HCl (pH 7.5), until the pH was 7.5. The column was stored at 4 °C. All eluate fractions were tested by Western blotting. The positive fractions were combined and concentrated by ultrafiltration.

Antibody concentration

After affinity purification, the antibodies were heavily diluted (1:40) compared to the starting serum. To concentrate the antibody solution, Vivaspin 15 columns (MWCO 10000) were used, which allow the concentration of macromolecules by ultracentrifugation. The Vivaspin 15 column was washed with 15 ml H_2O by centrifugation for 10 - 15 minutes at 2000 g. The membrane was blocked for 2 hours with 1 x PBS and 1% BSA. The blocking solution was discarded and the column was washed 3 times with H_2O by centrifugation at 2000 g. The combined antibody fractions were added subsequently to the column and concentrated at 2000 g, until a volume of $\sim$1 m was reached. 40 ml antibody solution was concentrated to 0.8 ml. The antibody solution was concentrated by a factor of 50.

3.7.3. Immunoblotting

Immunoblot or Western blot is a technique used to detect proteins in a sample. Proteins are blotted to a membrane and are probed with antibodies specific for the target protein (Burnette, 1981). Proteins separated by SDS-PAGE were transferred to a PVDF-membrane using the semi-dry blot technique. The PVDF-membrane was incubated in 100% methanol for 10 seconds, washing in H_2O for 5 minutes and equilibration in 1 x transfer buffer for at least 10 minutes. The protein gel was placed on top of the membrane. The assembly was sandwiched between two layers of 5 Whatman papers each, which have been soaked in transfer buffer. The assembly was placed on top of the anode of the blotting cell. The transfer was carried out at 15 V for 20 - 30 minutes. The transfer was monitored by a prestained protein marker.
To prevent unspecific binding of antibodies, the membrane was blocked for 2 hours at room temperature in blocking buffer. The membrane was incubated with the primary antibody in blocking buffer for 1 hour at room temperature or at 4 °C over night. To wash away excess antibody, the membrane was washed three times for 5 minutes with washing buffer at room temperature. Subsequently, the membrane was incubated with the secondary antibody diluted in blocking buffer for 0.5 - 1 hour at room temperature. The membrane was washed again three times for 5 minutes at room temperature with washing buffer.
The signal was detected by chemiluminescence. The ECL solutions 1 and 2 were prepared and mixed shortly before use, and the membrane was incubated for 1 - 5 minutes in the mixture. The membrane was exposed to an x-ray film. Exposure times varied between 5 seconds and 5 minutes. The films were developed in a developer machine (Agfa Curix 60).
All antibodies used for immunoblotting are listed in table 3.19.

TABLE 3.19 *Primary and secondary antibodies used for immunoblotting.*

antibody	species	dilution	purity
primary antibodies:			
α**Penta-His**	mouse IgG1	1:1000	affinity pure
α**H3K27me3**	rabbit (polyclonal)	1:10000	affinity pure
α***Hy*YY1**	rabbit (polyclonal)	1:1000	crude serum, affinity pure
α***Hy*EZH2**	rabbit (polyclonal)	1:500 - 1:1000	crude serum
α***Hy*EED**	rabbit (polyclonal)	1:1000	crude serum
α***Hy*p55**	rabbit (polyclonal)	1:1000	crude serum
α***Hy*PC**	rabbit (polyclonal)	1:1000	crude serum
α***Hy*RING**	rabbit (polyclonal)	1:1000	crude serum
secondary antibody:			
Goat-α-Mouse HRP	goat IgG (H+L)	1:4000	HRP conjugate
Goat-α-Rabbit HRP	goat IgG (H+L)	1:4000	HRP conjugate

Solutions for Immunoblotting

10 x transfer buffer:	250 mM Tris-base 1.5 M glycine
10 x PBS:	80 g NaCl (1.37 M) 2 g KCl (27 mM) 2.4 g KH_2PO_4 (18 mM) 26.8 g $Na_2HPO_4 \cdot 7H_2O$ (150 mM) Adjust the pH to 7.4. Sterilize by autoclaving.
Washing buffer:	0.2% Tween 20 in 1 x PBS
Blocking buffer:	2.5% milk powder in washing buffer
1 M Tris-HCl pH 8.5:	Dissolve 121.14 g Tris base in 900 ml H_2O. Adjust pH to 8.5 with HCl. Adjust volume to 1 litre with H_2O. Sterilize by autoclaving.
Luminol	250 mM Luminol in DMSO (0.44 g / 10 ml)
Coumaric acid	90 mM coumaric acid in DMSO (0.15 g / 10 ml)
ECL solution A:	Mix 12 μl 30% H_2O_2 with 20 ml 100 mM Tris-HCl (pH 8.5). Prepare freshly prior to use.
ECL solution B:	Mix 88 μl coumaric acid and 200 μl luminol with 19.8 ml 100 mM Tris-HCl (pH 8.5). Prepare freshly prior to use.

3.7.4. Immunohistochemistry

Immunofluorescence stainings with *Hydra* polyps were performed as described by Dunne et al. (1985). The polyps were stained with antibodies against PcG proteins and H3K27me3.
10 - 12 *Hydra* polyps were transferred into a 1.5 ml reaction tube. The animals were relaxed in 2% urethane for 1 - 2 minutes. The polyps were fixed in either Lavdovsky's fixative (ethanol : formaldehyde : acetic acid : H_2O (50 : 10 : 4 : 36)) or in 4% formaldehyde for 15 minutes at room temperature. The fixative was removed and the polyps were washed 3 times 10 minutes in 1 x PBS containing 0.1% Triton X-100 to permeabilize the membranes. For all tested antibodies, short fixation times proofed best. Longer fixation times resulted in a loss of the signal.
The polyps were incubated with the primary antibody diluted in 1 x PBS and 1% BSA at 4 °C over night. The primary antibody was removed and the polyps were washed 3 times for 10 minutes with 1 x PBS. The polyps were incubated with the secondary antibody for 1 - 2 hours at room temperature. To prevent unspecific binding, the secondary antibody had been preabsorped on fixed *Hydra* polyps over night at 4 °C. To co-stain the nuclei, 0.5 ng/μl of DAPI was added to the secondary antibody and incubated for additional 15 minutes.
The animals were washed 3 times for 10 minutes with 1 x PBS and mounted on microscope slides

using 200 μl Mowiol per slide. After the Mowiol has hardened, the staining was examined under a fluorescence or confocal microscope. All antibodies used for immunostainings are listed in table 3.3.

TABLE 3.3 *Primary and secondary antibodies used for staining Hydra.*

antibody	**species**	**dilution**	**purity**	**fixative**
primary antibodies:				
α**H3K27me3**	rabbit (polyclonal)	1:1000	affinity pure	4% formaldehyde
α***Hy*YY1**	rabbit (polyclonal)	1:200	crude serum, affinity pure	Lavdovsky
α***Hy*EZH2**	rabbit (polyclonal)	1:200	crude serum	Lavdovsky
α***Hy*EED**	rabbit (polyclonal)	1:200	crude serum	Lavdovsky
α***Hy*p55**	rabbit (polyclonal)	1:200	crude serum	Lavdovsky
α***Hy*PC**	rabbit (polyclonal)	1:200	crude serum	Lavdovsky
α***Hy*RING**	rabbit (polyclonal)	1:200	crude serum	Lavdovsky
secondary antibody:				
Alexa Fluor 488 donkey-α-rabbit	donkey	1:400	Alexa488-conjugated	Lavdovsky, 4% formaldehyde

3.8. Methods: Analysis of protein-DNA interactions

3.8.1. DNA binding site selection

The sequence specifity of the *Hy*YY1 protein was determined by incubating the immobilized 6His-fusion protein with a pool of random sequence oligonucleotides. Bound DNA was recovered and amplified by PCR. The DNA was used for further rounds of binding and finally the *Hy*YY1 consensus sequence was determined.

Oligonucleotides

To obtain a pool of double-stranded random sequence oligonucleotides for DNA binding site selection, a single-stranded 85-mer oligonucleotide was annealed with a 20-mer complementary primer and the Klenow fragment was used to fill in 3′ recessed ends. The following oligonucleotide were used:
EMSA random: 5′-CGT GAT TGG CTA CGC ATC GT CAG CAG CAG CAG CAG $(N)_{15}$ CAG CAG CAG CAG CAG GTC ATA CTC ACT CAC GGA CT-3′
Klenow rev: 5′-AGT CCG TGA GTG AGT ATG AC-3′

5 μg of the random oligonucleotide pool were combined with an equal amount of the complementary primer and incubated at 95 °C for 10 minutes. The oligonucleotides were cooled down slowly to 55 °C. The oligonucleotides were annealed at 55 °C for 30 minutes. To fill in the 3′ recessed ends, 2.5 μl of a 10 mM dNTP-mix and 1 μl of Klenow enzyme (5 U/μl) were added. The reaction was incubated at 37 °C for 30 minutes. The DNA was precipitated with ethanol and sodium acetate. The final DNA pellet was dissolved in 50 μl of H_2O. DNA concentration was determined by spectrophotometry and the DNA was analyzed by gel electrophoresis using a 12% polyacrylamide gel.

DNA binding site selection

The binding site selection (panning) was performed both with immobilized 6His-*Hy*YY1 fusion protein and with immobilized 6His-*Hy*YY1ΔZnF protein. The truncated protein served as a control for the binding site selection. The protein is lacking the C-terminal zinc fingers responsible for DNA binding and should not be able to bind to any specific DNA sequence.
100 μl of the immobilized protein (∼2.5 - 5 μg of protein) were incubated with 100 ng of the double-stranded random sequence oligonucleotide in 1 x binding buffer (10 mM HEPES-KOH (pH 7.9), 100 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA, 0.025% NP40, 10% glycerine, 100 μg/ml BSA, 2 mM spermidine, 0.5 mM DTT) and 1 μg of unspecific competitor (poly(dI-dC)) for 30 minutes at room temperature. The suspension was centrifuged in a tabletop centrifuge at 2000 rpm for 1 minute. The supernatant was discarded and the agarose beads were washed with 1 ml of 1 x binding buffer three times for 5 minutes. The suspension was centrifuged in a tabletop centrifuge at 2000 rpm for 1 minute. After the last washing step, the beads were resuspended in 50 μl of 1 x binding buffer and 150 μl of 0.1% SDS. The beads were incubated at 95 °C for 10 minutes. The DNA was extracted with phenol-chloroform-isoamylalcohol (25:24:1). 200 μl of phenol-chloroform-isoamylalcohol were added. After vigorous mixing, the phases were separated by centrifugation at top speed in a tabletop centrifuge for 5 minutes. The aqueous phase

was transfered to a new reaction tube and again extracted with 200 μl of phenol-chloroform-isoamylalcohol. The aqueous phase was precipitated with 1/10 volume of 3 M sodium acetate and 2.5 volumes of 100% ethanol. 20 μg of glycogen were added as carrier. After centrifugation in a tabletop centrifuge at top speed for 15 minutes, the supernatant was removed and the DNA pellet was washed with 750 μl of 70% ethanol. Finally, the DNA pellet was dissolved in 20 μl of H_2O.
The DNA was amplified by PCR. 1 μl of the DNA solution were mixed with 1 μl 10 mM dNTP-mix, 10 pmol forward primer (Klenow for), 10 pmol reverse primer (Klenow rev) and 0.5 μl (2.5 U) of EuroTaq DNA polymerase in 1 x RB-buffer containing 1.5 mM $MgCl_2$ (total volume: 50 μl). The DNA was amplified using the following cycling conditions:

initial denaturation	95 °C	5 minutes	
denaturation	95 °C	15 seconds	
annealing	55 °C	15 seconds	10 - 25 cycles
elongation	72 °C	15 seconds	
final elongation	72 °C	7 minutes	

The PCR products were analyzed by agarose gel electrophoresis on a 12% polyacrylamide gel. The amplified DNA was analyzed after different numbers of cycles (10, 15, 20 and 25). The DNA after cycle 10 - 15 was used for further rounds of binding site selection. A greater number of cycles resulted in the production of multimers from the original 85-mer oligonucleotide. 2 μl of the amplified DNA mixture were subjected to further rounds of selection, which were performed in the same way as the initial round of selection.
After rounds 3 and 6, the DNA was purified using Promega's 'Wizard SV Gel and PCR Clean-Up System'. The concentration was determined by spectrophotometry and the amplified DNA was ligated into pGEM-T vector at a vector : insert-ratio of $\sim$1 : 3. The ligation reaction was incubated over night at 4 °C. Chemocompetent *E.coli* DH5α-cells were transformed with 2 μl of the ligation reaction. Positive colonies were analyzed by colony-PCR. The DNA of positive clones was isolated using Promega's 'Wizard Plus SV Minipreps DNA Purification System'. The purified DNA was used for sequencing. For a detailed description of ligation, transformation, colony-PCR, DNA minipreparation and sequencing see section 3.5.

3.8.2. Gel retardation assay

Biotin-end-labeling of DNA

For labeling the target DNA sequence for gel retardation assays the terminal deoxynucleotidyl transferase (TdT) was used. The polymerase catalyzes the addition of deoxynucleotides to the 3 ' hydroxyl end of DNA. In this case biotinylated nucleotides were added to the DNA. For DNA labeling the 'Biotin 3' End DNA Labeling Kit' from Pierce was used according to the manufacturer's instructions. Two complementary single-stranded DNA oligonucleotides were both labeled and annealed afterwards. 50-mer oligonucleotides containing the following target sequences were used for gel retardation assays:

***Hy*YY1 round 6**: GTAAAGTTTTCTTGA
(sequence obtained after round 6 of the binding site selection with the full length *Hy*YY1 protein)
***Hy*YY1 round 3**: TTTTTGTTGAATGTC
(one of the sequences obtained after round 3 of the binding site selection with the full length *Hy*YY1 protein)
control: CCTGCAGCACTGTGT
(one of the sequences obtained after round 6 of the binding site selection with the truncated *Hy*YY1 protein)
YY1 consensus: GCCG**CCAT**TTTGGGC
(known Pho/YY1 binding sequence from human or *Drosophila*)
(The entire sequences of the respective oligonucleotides are specified in section 3.1.4).

For the labeling reaction 5 pmol 3′-OH ends of the oligonucleotide were combined with 5 μl of 5 μM Biotin-11-UTP and 0.2 units/μl of TdT in 1 x TdT Reaction Buffer in a total volume of 50 μl. The reaction was incubated at 37 °C for 30 minutes and terminated by the addition of 2.5 μl 0.2 M EDTA. The DNA was extracted by chloroform-isoamyl alcohol (24:1). The complementary single-stranded labeled DNA oligonucleotides were annealed before the application in a gel retardation assay. Annealing was carried out by denaturing a mixture of equal amounts of complementary single-stranded oligonucleotides at 90 °C for 1 minute. The mixture was allowed to cool down slowly and incubated at the melting temperature of the oligonucleotides for 30 minutes followed by cooling down slowly to room temperature. The double-stranded labeled oligonucleotide was stored at - 20 °C.
The labeling efficiency was determined by comparing the labeled oligonucleotide with labeled control oligo. 2 μl of samples and standards were spotted onto a positive charged nylon membrane. After absorption into the membrane, the dot blot was cross-linked by a Stratalinker cross-linking instrument using the auto cross-link function. Detection was carried out like for the gel retardation assay (see next section). Figure 3.6 shows the labeling efficieny of the DNA oligonucleotides used for gel retardation assays. All oligonucleotides could be labeled with a high efficiency of 75 - 100% compared to a labeled control oligonucleotide.

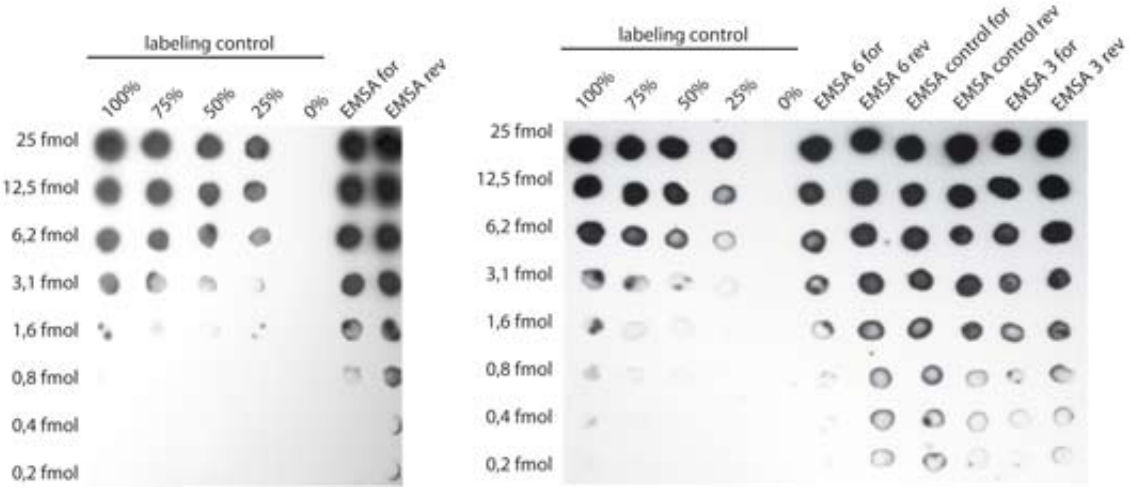


FIGURE 3.6 *Labeling efficiency. The names and the amounts of the spotted oligos are indicated. The biotin-labeled control oligonucleotide is on the left side on each blot.*

Gel retardation assay

To study protein-DNA interactions by gel retardation technique, the 'LightShift Chemiluminescent EMSA Kit' from Pierce was used according to the manufacturer's instructions.
The binding reaction was performed in 1 x binding buffer (10 mM HEPES-KOH (pH 7.9), 100 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA, 0.025% NP40, 10% glycerine, 100 μg/ml BSA, 2 mM spermidine, 0.5 mM DTT) containing 20 μM zinc chloride and 1 μg of unspecific competitor (poly(dI-dC)). The binding reactions were performed with 20 fmol of biotin-labeled oligonucleotide and protein amounts ranging from 50 ng to 5 μg. The binding reactions were incubated at room temperature for 30 minutes. 5 μl of 5 x loading buffer were added to each binding reaction and mixed by gently pipetting up and down. 20 μl of each sample were loaded onto a prepared and pre-run polyacrylamide gel. The gel was electrophoresed at 100 V in 0.5 x TBE buffer until the bromophenol blue dye had migrated $\sim 3/4$ of the distance of the gel.
The DNA was transferred to a positively charged nylon membrane. The membrane was soaked in 0.5 x TBE buffer for 10 minutes and the polyacrylamide gel was placed on top of the nylon membrane. The assembly was sandwiched between two layers of 5 Whatman papers each, which have been soaked in 0.5 x TBE buffer. The assembly was placed on top of the anode of the blotting cell. The transfer was carried out at 380 mA for 45 minutes. The transferred DNA was cross-linked to the nylon membrane using the auto-crosslink function of a Stratalinker cross-link instrument.
For detection, the membrane was blocked with 20 ml of 1 x blocking buffer for 15 minutes at room temperature. The blocking buffer was discarded and replaced with 20 ml of conjugate/blocking buffer (20 ml of 1 x blocking buffer containing 66.7 μl of Streptavidin-Horseradish Peroxidase Conjugate). The membrane was incubated in the conjugate/blocking buffer for 15 minutes at room temperature. The buffer was discarded and the membrane was rinsed in 20 ml of 1 x wash buffer. The membrane was washed four times 5 minutes with 20 ml of 1 x wash solution, followed by equilibration in 30 ml of substrate equilibration buffer for 5 minutes at room temperature. The substrate equilibration buffer was removed completely and the membrane was incubated in the freshly prepared substrate working solution for 5 minutes at room temperature. The substrate working solution was prepared by adding an equal volume of luminol/enhancer solution to the stable peroxide solution. The membrane was exposed to an x-ray film for 5 - 60 seconds and the film was developed using an Agfa Curix 60 developer machine.

Preparation of a native polyacrylamide gel:
A native 6% polyacrylamide gel was prepared in 0.5 x TBE buffer. The gel was cast one day in advance and stored at 4 °C. At the day of the experiment, the gel was pre-electrophoresed at 100 V in 0.5 x TBE buffer until the current did not change anymore. Before loading the samples, the buffer was exchanged and the wells were flushed to remove residual polyacrylamide.

Recipe for 2 gels of 8 x 8 cm x 0.8 mm	
H_2O	6.1 ml
5 x TBE	2 ml
100% glycerol	0.25 ml
40% polyacrylamide (acrylamide - bisacrylamide 29:1)	1.5 ml
10% APS	100 μl
TEMED	5 μl

3.9. Methods: Cell culture and Baculovirus production

3.9.1. Cultivation of Sf9 cells

The insect cell line Sf9 is a substrain from the Sf21 cell line. Sf21 is a continuous cell line that has been developed from ovaries of a moth species: the Fall Armyworm *Spodoptera frugiperda* (Vaughn et al., 1977). Both cell lines are often used for baculovirus production.
The Sf9 cells were kept in TC100 insect medium (Lonza), which has been supplemented with 10% FBS and 1% Pen/Strep at 28 °C in an incubator. The cells were grown in an adherent culture in a volume of 50 ml in a T-175 cm^2 flask with a filter cap. When the cells reached a density of $\sim 2 \cdot 10^6$ cells/ml, the cells were subcultured. Normally, the cells were subcultured twice a week. For subculturing, the cells were displaced from the flask's surface by tapping the flask, until the cells were detached. 10 ml of the cell suspension were transfered to 40 ml of pre-warmed fresh medium in a new flask. The mixture of fresh medium and conditioned medium helps to maintain the cells healthy.

Cell counting and determination of viability

To count the cells and to check the viability, 10 μl of the cell suspension were mixed with the same volume of 0.4% trypan blue, which selectively stains dead cells. 10 μl of the suspension were introduced into a Neubauer counting chamber and the number of live cells was determined under a microscope. The cell number in four large squares of the counting grid was determined. Each square has an area of 0.04 mm^2 and a depth of 0.1 mm. The number of cells per milliliter was calculated by multiplying the average number of cells by the chamber factor (10^4) and the dilution factor:

$$\text{cells/ml} = \frac{\text{total number of viable cells}}{4} \cdot 10^4 \cdot 2$$

Freezing of Sf9 cells

For long-term storage, the cells were frozen in liquid nitrogen. For freezing the cells, the Sf9 culture was expanded and the cells were harvested in exponential growth. The cells were displaced from the flask's surface and the suspension was transfered 50 ml tubes. The number of cells and the viability were determined. Only healthy cells were used for freezing and long-term storage. The suspension was centrifuged at 500 g and 4 °C for 5 minutes. The cell pellet was resuspended carefully in the desired volume of freezing medium (TC100 medium, 10% FBS, 1% Pen/Strep, 5% DMSO) to obtain cells at a density of $\sim 1 \cdot 10^7$ cells/ml. 1.5 ml of the cell suspension were aliquoted into 2 ml cryovials. The cells were frozen at -80 °C at a freezing rate of 1 °C per minute. After 24 hours, the cells were stored in liquid nitrogen, until use.
Thawing of the frozen cells was done rapidly at 37 °C in a water bath. The cell suspension was transfered into a cell culture flask with pre-warmed medium and incubated at 28 °C.

3.9.2. Baculovirus production

Baculoviruses are a family of viruses with rod-shaped capsids found exclusively in arthropods. They have a double-stranded, circular DNA genome with a length between 80 and 200 kilo-

basepairs. The DNAs of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) and the *Bombyx mori* nuclear polyhedrosis virus (BmNPV) are commonly used for expression vectors. The baculovirus expression system is widely used to generate properly post-translationally modified proteins in insect cells.

During this thesis, Invitrogen's Bac-to-Bac Baculovirus Expression System was used for recombinant baculovirus production and protein expression of *Hydra* PcG proteins. The method was developed by Ciccarone et al. (1997). This system provides a rapid method for the generation of recombinant baculoviruses. It is based on a site-specific transposition of an AcMNPV-based expression casette into a baculovirus shuttle vector called bacmid, which can be propagated in *E.coli*. An overview of the baculovirus expression system is depicted in figure 3.7.

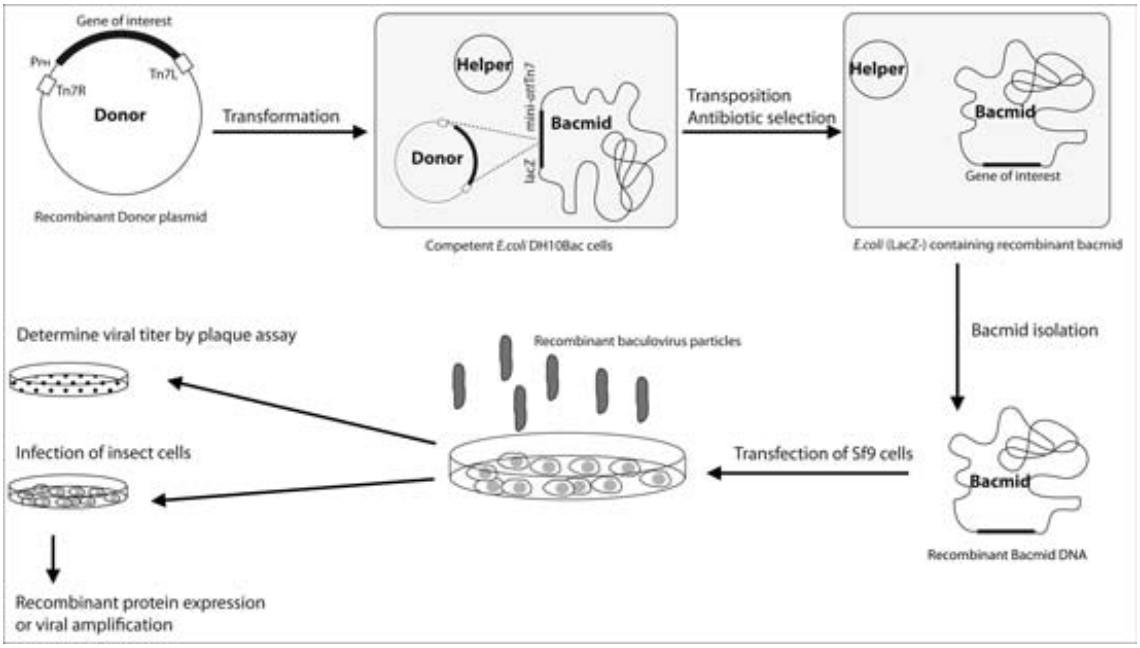


FIGURE 3.7 *Diagram of the Bac-to-Bac Baculovirus Expression System. The figure was adapted from Invitrogen's Bac-to-Bac manual.*

Production of recombinant bacmid DNA

For production of recombinant bacmid DNA, *E.coli*-DH10Bac-cells, which contain a baculovirus shuttle vector (bacmid) and a helper plasmid were used. The vector sequence which is flanked by mini-Tn7 sites allows the transposition into the mini-*att*Tn7 attachment site on the bacmid. Once the plasmid DNA is transformed into the DH10Bac-cells, the transposition occurs and a recombinant bacmid is produced. The insertion into the mini-*att*Tn7 site disrupts the expression of the LacZα peptide and positive colonies can be selected using LB-agar plates containing X-Gal and IPTG.

One aliquot of chemocompetent DH10Bac-cells was thawed on ice and transferred to a 15 ml polypropylene round-bottom tube. The cells were mixed gently with 200 ng of pFastBac1-plasmid DNA and incubated on ice for 30 minutes. The bacteria were heat shocked for 45 seconds at 42 °C and immediately chilled on ice for 2 minutes. 900 μl SOC-medium were added and the cells

were incubated at 37 °C and 200 rpm for 4 - 6 hours to allow transposition between the bacmid and the pFastBac1 vector. 10-fold serial dilutions of the cells were prepared and 100 μl of each dilution were plated on LB-agar plates. The plates were incubated at 37 °C for 48 hours. Positive white colonies were selected, restreaked on LB-plates and incubated at 37 °C for 48 hours. Only clear white colonies were taken for further analysis and recombinant bacmid DNA isolation.

Large plasmid purification

To isolate recombinant bacmid DNA, which is greater than 135 kb in length, a method originally developed for large plasmid DNA purification has been applied. The method was described in Invitrogen's manual of the Bac-to-Bac Baculovirus Expression System.
100 ml LB medium containing 50 μg/ml kanamycin, 7 μg/ml getamicin and 10 μg/ml tetracycline was inoculated with the desired recombinant *E.coli*-DH10Bac. The culture was grown over night at 37 °C and 200 rpm in a shaker. The cells were pelleted by centrifugation for 5 min at 5000 g. The supernatant was discarded and the cells were resuspended in 30 ml of Solution I. The resuspension was gently mixed with 30 ml of Solution II. The mixture was incubated for 5 minutes at room temperature and 30 ml of 3 M potassium acetate (pH 5.5) were added. The sample was chilled on ice for 10 minutes and centrifuged for 30 minutes at 10000 g and 4 °C. The supernatant was transferred to a new tube. In case the supernatant was still contaminated with a white precipitate of bacterial genomic DNA and proteins, it was cleared additionally by a folded filter. To precipitate the plasmid DNA, 80 ml of isopropanol were added and the mixture was kept on ice for 15 minutes. The DNA was centrifuged for 45 minutes at 3000 g. The pellet was washed with 10 ml 70% ethanol and air-dried. The DNA was dissolved in 500 μl TE-buffer (pH 8.0). The bacmid DNA was stored at 4 °C.

Solutions for large plasmid purification

Solution I:	15 mM Tris-HCl (pH 8.0) 10 mM EDTA 100 μg/ml RNase A Sterilize by filtration and store at 4 °C.
Solution II:	0.2 N NaOH 1% SDS Sterilize by filtration and store at room temperature.
3 M potassium acetate:	3 M potassium acetate Adjust the pH to 5.5 with glacial acetic acid. Sterilize by autoclaving.
TE-buffer:	10 mM Tris-HCl (pH 8.0) 1 mM EDTA (pH 8.0)

The recombinant bacmid DNA was analyzed by PCR using the M13 forward and M13 reverse primers or gene specific primers. If transposition has occured, a PCR product of ~2300 bp + size of the insert can be detected on an agarose gel. If no transposition occured, a PCR product of 300 bp can be detected. For PCR amplification 100 ng isolated bacmid DNA were mixed with 1 μl

10 mM dNTP-mix, 10 pmol forward primer, 10 pmol reverse primer and 2.5 U of EuroTaq DNA polymerase in 1 x RB-buffer containing 1.5 mM $MgCl_2$ (total volume: 50 μl). The insert DNA was amplified using the following cycling conditions:

initial denaturation	98 °C	3 minutes	
denaturation	98 °C	45 seconds	
annealing	55 °C	45 seconds	35 cycles
elongation	72 °C	5 minutes	
final elongation	72 °C	7 minutes	

Figure 3.8 gives an example of the analysis of recombinant *Hy*YY1-6His bacmid constructs by PCR. For the following transfection of insect cells, one of the correct constructs was chosen.

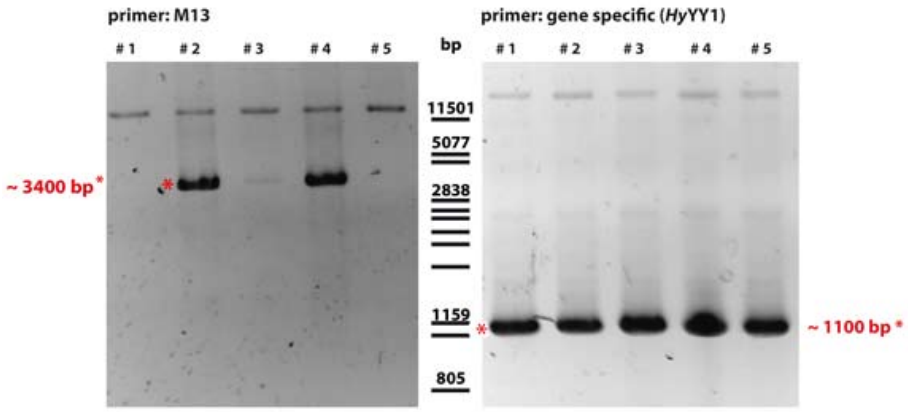


FIGURE 3.8 *1% agarose gels depicting the analysis of recombinant HyYY1-6His bacmid constructs. The left gel shows the result of the PCR with M13 primers, the right gel shows the result with gene specific primers. A PCR product of ∼3400 bp should be obtained with M13 primers, a PCR product of ∼1100 bp should be visible with gene specific primers. All analyzed constructs showed a band of the correct size with gene specific primers, but only clones # 2 and # 4 showed a band of the correct size with M13 primers. Clone # 3 showed a faint band with M13 primers. For the transfection of insect cells, only such constructs were used that showed a clear signal with both primer pairs.*

Transfection of Sf9 cells

To produce recombinant baculoviruses, Sf9 cells were transfected with the recombinant bacmid using a cationic lipid-based system. The cells were transfected in 6-well tissue culture plates.
For transfection, $9 \cdot 10^5$ cells were seeded per well in 2 ml of TC100 medium supplemented with 10% FBS and 1% Pen/Strep. The cells were allowed to attach to the surface at 28 °C for at least 1 hour. For transfection, 1 μg of recombinant bacmid DNA were diluted in 100 μl of unsupplemented TC100 medium. Furthermore, 6 μl of Cellfectin reagent were diluted in 100 μl of unsupplemented TC100 medium. The diluted bacmid DNA and the lipids were combined and incubated at room temperature for 30 - 60 minutes. Meanwhile, the medium was removed from the cells and the cells were washed with 2 ml of unsupplemented TC100 medium. 0.8 ml of unsupplemented medium were added to the DNA:lipid complexes. The medium was removed

from the cells and the transfection mix was added. The cells were incubated with the transfection mix at 28 °C for 5 hours. The DNA:lipid complexes were removed and 2 ml of supplemented TC100 medium were added to the cells. The cells were incubated at 28 °C in a humidified incubator, until signs of baculovirus infection became clearly visible. Typically, detachment of the cells from the surface and cell lysis became visible after 4 - 5 days post transfection.

Isolating P1 baculoviral stock

The P1 viral stock was isolated from the transfected cells showing signs of late stage infection. Both the cells and the supernatant were transfered to a 2 ml reaction tube and centrifuged at top speed in a tabletop centrifuge for 1 minute. The clarified supernatant was transfered to a fresh tube and stored as P1 viral stock at 4 °C protected from light.
However, the P1 viral stock is a small-scale viral stock with a low titer. The titer of the P1 viral stock was not determined. It can be assumed that the titer ranges from $1 \cdot 10^6$ - $1 \cdot 10^7$ pfu/ml. For further experiments, the stock needs to be amplified.

Amplification of the baculoviral stock

To amplify the viral stock, Sf9 cells were infected at a multiplicity of infection (MOI) ranging from 0.05 - 0.1. Sf9 cells at a density of $2 \cdot 10^6$ cells/ml were seeded in a volume of 10 ml supplemented TC100 medium in a T-75 cm^2 flask. The cells have been allowed to attach to the surface at 28 °C for at least 1 hour before the cells were infected with 0.4 ml of the P1 viral stock.

$$\text{inoculum required (ml)} = \frac{\text{MOI (pfu/cell)} \cdot \text{number of cells}}{\text{titer of viral stock (pfu/ml)}} = \frac{0.1 \cdot 2 \cdot 10^7}{0.5 \cdot 10^6} = 0.4 \text{ ml}$$

After the cells signs of late stage infection, the cells and the supernatant were transfered to a 15 ml tube and centrifuged at 500 g and 4 °C for 5 minutes. The clarified supernatant resembles the P2 viral stock. Again, the titer was not determined. Normally, the titer ranges from $1 \cdot 10^7$ - $1 \cdot 10^8$ pfu/ml. The P2 viral stock was used to generate a high-titer P3 viral stock. The P3 viral stock was produced in the same way as the P2 viral stock, except the scale was increased: Sf9 cells at a density of $2 \cdot 10^6$ cells/ml in a volume of 50 ml were infected with 0.1 ml of the P2 viral stock (MOI = 0.1). The titer of the P3 viral stock was determined. This stock was used for all further experiments.

Determination of the baculoviral titer

The titer of the P3 viral stock was determined by a viral plaque assay.
Sf9 cells were harvested and a cell suspension at $5 \cdot 10^5$ cells/ml in supplemented TC100 insect medium was prepared. 2 ml of the cell suspension were aliquoted into each well of two 6-well tissue culture plates. The cells were allowed to settle for 1 hour at room temperature. Only plates with attached cells and 50% confluence were used for the viral plaque assay. A logarithmic serial dilution (10^{-1} - 10^{-7}) of the P3 baculovirus stock was prepared in supplemented TC100 medium. The medium of each well was removed and replaced with 1 ml of the appropriate virus dilution. The partitioning of a 6-well plate is illustrated in figure 3.9. For each viral stock, two 6-well plates were used: one sample plate and one duplicate. The cells were incubated with the virus for 1 hour at room temperature. After 1 hour, the virus was removed from the cells completely and replaced

with 2 ml of plaquing medium, which was prepared by combining 1.5 ml Neutral Red Solution (1 mg/ml in TC100 medium) with 10.5 ml supplemented TC100 medium and 12 ml 2% agarose in unsupplemented TC100 medium. The plaquing medium has been kept at 40 °C in a water bath, until use. The agarose was allowed to harden for 10 - 20 minutes at room temperature. The cells were incubated at 28 °C in a humidified incubator for 7 - 10 days, until plaques became visible.

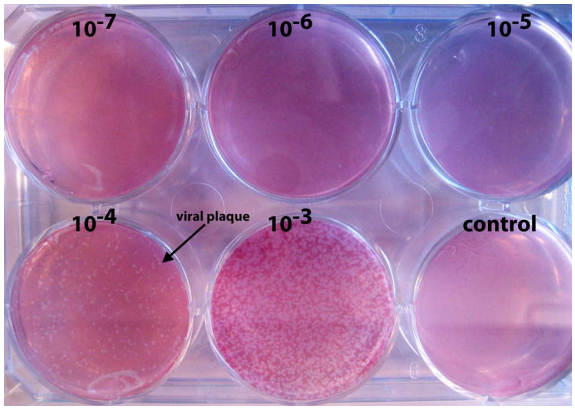


FIGURE 3.9 *Typical appearance of a viral plaque assay 7 - 10 days post infection. The partitioning of the 6-well plate is indicated.*

The plaques were counted and the titer was calculated as follows:

$$\text{titer (pfu/ml)} = \text{number of plaques} \cdot \text{dilution factor} \cdot \frac{1}{\text{ml of inoculum/well}}$$

The titers of the baculoviral stocks used for further experiments are given in table 3.5:

TABLE 3.5 *Titers of the P3 baculoviral stocks.*

recombinant baculovirus	titer (pfu/ml)	produced by
*Hy*YY1	not determined	Sonja Matt
*Hy*YY1-6His	$1.7 \cdot 10^7$	Sonja Matt
*Hy*EZH2	$4 \cdot 10^8$	Volker Lauschke
*Hy*EED	$3 \cdot 10^8$	Volker Lauschke
*Hy*SUZ12	$8.5 \cdot 10^6$	Volker Lauschke
*Hy*p55	$3 \cdot 10^8$	Volker Lauschke

3.9.3. Expression of recombinant *Hydra* PcG proteins in insect cells

Recombinant *Hydra* PcG proteins were expressed in Sf9 cells using the baculovirus expression system. The PcG proteins were expressed in insect cells to reconstitute the protein complexes and to perform protein-protein interaction studies with the help of immunoprecipitation techniques. To optimize the expression conditions, Sf9 cells were infected at varying MOIs (1, 2, 5, 10 and 20) and recombinant protein expression was assayed at different times after infection (24, 48, 72 and 96 hours). The optimization procedure was carried out in a 24-well format. For each recombinant protein, $6 \cdot 10^5$ cells were seeded into each well of a 24-well plate in TC100 medium

supplemented with 1% Pen/Strep. The cells were allowed to attach to the surface for at least one hour at 28 °C. The medium was removed from the cells, and the cells were washed with fresh medium once. The medium was replaced with 300 μl of fresh medium and the cells were infected with the baculovirus stock in each well at the desired MOI. Figure 4.22 depicts the partitioning of the 24-well plate. The infected cells were incubated at 28 °C in a humidified incubator and were harvested at the appropriate time. For harvesting, the cells and the supernatant were transfered to a 1.5 ml reaction tube and centrifuged at top speed in a tabletop centrifuge for 1 minute. The supernatant was removed and the cell pellet was rinsed with 1 x PBS. The cells were lysed with 400 μl 1 x protein sample buffer and the samples were boiled for 5 - 10 minutes at 95 °C. Until use, the samples were stored at -20 °C.

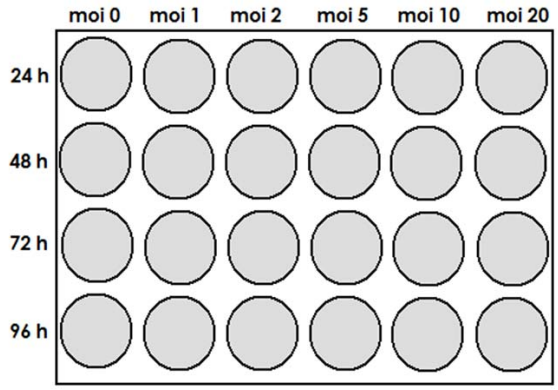


FIGURE 3.10 *Partitioning of a 24-well plate for optimizing the protein expression conditions.*

Samples of each protein expression condition were analyzed by immunoblotting. 20 μl of each protein sample were seperated by SDS-PAGE. The proteins were transfered to a PVDF-membrane and probed with a specific antibody to detect protein expression. For IP and Co-IP studies, the cells were infected or co-infected with recombinant baculovirus. For the *Hy*EZH2, *Hy*EED, *Hy*SUZ12 and *Hy*p55 baculoviruses a MOI of 5 was used for infection, for the HyYY1-6His baculovirus a MOI of 2 was used. The recombinant proteins were expressed for 48 hours. Table 3.6 gives an overview of the best conditions for recombinant protein expression. The conditions for the *Hy*EZH2, *Hy*EED, *Hy*SUZ12 and *Hy*p55 baculoviruses were determined by Volker Lauschke during his master thesis.

TABLE 3.6 *Optimal conditions for infection and protein expression.*

recombinant baculovirus	**titer (pfu/ml)**	**MOI**	**protein expression**
*Hy*YY1-6His	$1.7 \cdot 10^7$	1 - 5	24 - 48 hours
*Hy*EZH2	$4 \cdot 10^8$	2 - 10	48 - 96 hours
*Hy*EED	$3 \cdot 10^8$	2 - 10	24 - 96 hours
*Hy*SUZ12	$8.5 \cdot 10^6$	not determined	not determined (no specific antibody available)
*Hy*p55	$3 \cdot 10^8$	2 - 10	24 - 96 hours

3.10. Methods: Immunoprecipitation

Immunoprecipitation (IP) is a technique of isolating a protein antigen by binding to a specific antibody coupled to a solid-phase matrix. The technique is not only used to precipitate individual proteins, but intact protein complexes. An antibody against a known member of a protein complex is used to precipitate the entire complex. This method is called co-immunoprecipitation (Co-IP) and is used to analyze protein-protein interactions and to identify unknown members of protein complexes.
The immunoprecipitation technique is also used to locate DNA binding sites of a protein of interest on the genome. DNA binding proteins can be cross-linked to the DNA they are binding to, and a specific antibody against the DNA binding protein can be used to precipitate the DNA-protein complex. This variant is called chromatin immunoprecipitation (ChIP).

3.10.1. IP and Co-IP

Immunoprecipitation was used to characterize the generated antibodies against the *Hydra* PcG proteins and to check if they can be used for immunoprecipitation techniques like Co-IP and ChIP. IP and Co-IP studies were performed with recombinantly expressed *Hydra* PcG proteins in Sf9 insect cells and with endogenous *Hydra* proteins. For IP or Co-IP studies with recombinant proteins, $\sim 2 \cdot 10^7$ cells were infected or co-infected with the respective baculoviruses. The recombinant proteins were expressed using the conditions that have been previously tested. For studies with endogenous *Hydra* proteins, ~1 ml ($\simeq$ 1000 animals) of densly packed *Hydra magnipapillata 105* polyps were used. IP and Co-IP studies were performed according to the protocols by Harlow and Lane (1988) and Bonifacino et al. (2001). Figure 3.11 gives an overview of the different steps during IP or Co-IP.

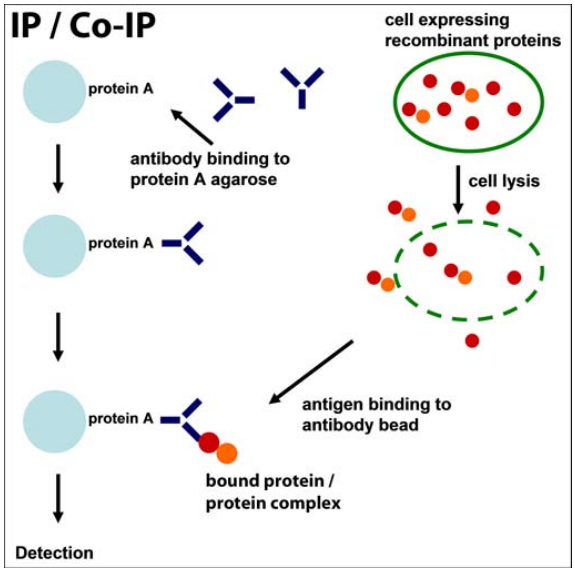


FIGURE 3.11 *Flow chart depicting the different steps during IP or Co-IP.*

Preparation of antibody-conjugated beads

30 μl of 50% protein A-agarose bead slurry were combined with 0.5 ml ice-cold 1 x PBS (137 mM NaCl, 2.7 mM KCl, 1.8 mM KH_2PO_4 15 mM Na_2HPO_4 (pH to 7.4)) and the respective amount of antibody in a 1.5 ml reaction tube. 1 μg of affinity purified antibody (e.g. α-*Hy*YY1, αPenta-His) or 5 μl of polyclonal antiserum have been used. It has to be kept in mind that antibodies from mouse bind to protein A with a lower affinity compared to antibodies from rabbit. The used αPenta-His antibody is an IgG1 from mouse, all other antibodies used for immunoprecipitation are from rabbit.
The suspension was mixed thoroughly and incubated over night at 4 °C in a rotator. The suspension was centrifuged in a tabletop centrifuge at top speed and 4 °C for 2 seconds. The supernatant containing the unbound antibody was removed. The beads were resuspended in 1 ml of nondenaturing lysis buffer (300 mM NaCl, 50 mM Tris-HCl (pH 7.4), 5 mM EDTA, 0.1% (w/v) Triton X-100, 0.02 mM sodium azide) and incubated on a rotator at 4 °C for 5 minutes. The suspension was centrifuged in a tabletop centrifuge at top speed and 4 °C for 2 seconds. The washing step was repeated two more times. The antibody-bound beads were stored at 4 °C until use.

Coupling antibodies to protein A-agarose beads

If the same antibody (or an antibody from the same species) was used for immunoprecipitation and detection, the IP antibody was cross-linked to protein A-agarose beads to avoid the contamination of the eluate with antibody. The heavy and the light chain of the antibody would be detected by the antibody used for immunoblotting and the antigen band may be masked.
50 μl of 50% protein A-agarose bead slurry were combined with 1 ml ice-cold 1 x PBS (137 mM NaCl, 2.7 mM KCl, 1.8 mM KH_2PO_4 15 mM Na_2HPO_4 (pH to 7.4)) and the respective amount of antibody in a 1.5 ml reaction tube. 2 μg of affinity purified antibody (e.g. α-*Hy*YY1) or 10 μl of polyclonal antiserum have been used. The suspension was mixed thoroughly and incubated for 1 hour at room temperature in a rotator. Unbound antibodies were removed by centrifugation in a tabletop centrifuge at top speed for 5 seconds. The beads were washed twice with 1 ml of 0.2 M sodium borate (pH 9.0) for 5 minutes on a rotator, followed by centrifugation in a tabletop centrifuge at top speed for 5 seconds. The antibodies were cross-linked to the protein A-agarose beads using dimethylpimelimidate (DMP). The beads were resuspended in 0.5 ml of sodium borate (pH 9.0) and 20 mM DMP. The DMP was added shortly before use as solid substance. The suspension was incubated for 30 minutes at room temperature on a rotator. The coupling reaction was stopped by washing the beads once in 0.2 M ethanolamine (pH 8.0), followed by incubation of the beads for 2 hours at room temperature on a rotator in 0.2 M ethanolamine. The suspension was centrifuged in a tabletop centrifuge at top speed for 5 seconds and the supernatant was removed. The beads were washed three times in 100 mM glycine (pH 3.0) for 5 minutes to remove any antibodies bound by noncovalent binding to the protein A-agarose. Finally, the suspension was centrifuged in a tabletop centrifuge at top speed for 5 seconds and the antibody coupled beads were resuspended in 1 x PBS with 0.01% merthiolate. The beads were stored at 4 °C until use.
The coupling procedure was monitored. Samples before and after cross-linking of the antibody were taken. The samples were eluted by boiling for 5 minutes at 95 °C and separated by SDS-PAGE. If the coupling procedure was successful, no antibody band should be visible on the SDS-

gel in the fraction after elution/boiling. In figure 3.12 the coupling procedure of α-*Hy*YY1 was monitored.

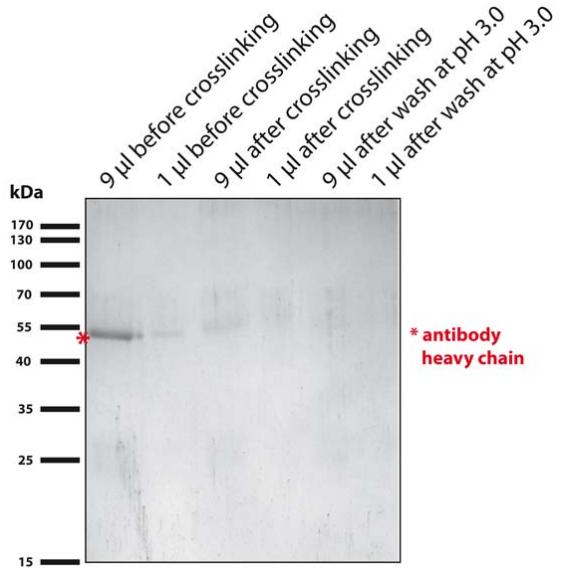


FIGURE 3.12 *12% SDS-gel visualizing the subsequent steps during cross-linking of the α-HyYY1 antibody to protein A-agarose beads. Different amounts of the beads were taken as sample to monitor the cross-linking procedure. Samples were taken before cross-linking with DMP, after cross-linking and after the final wash step with glycine. The beads were resuspended in an equal volume of 2 x protein sample buffer and boiled at 95 °C for 5 minutes prior to separation by SDS-PAGE. Whereas the band of the antibody heavy chain is clearly visible in the eluate before cross-linking, no band is visible after cross-linking and the final wash step.*

Immunoprecipitation

The insect cells or *Hydra* polyps were harvested by centrifugation at 500 g and 4 °C for 5 minutes. The supernatant was aspirated and the cells or *Hydra* polyps were washed in ice-cold 1 x PBS (137 mM NaCl, 2.7 mM KCl, 1.8 mM KH_2PO_4 15 mM Na_2HPO_4 (pH to 7.4)) twice. The cells or *Hydra* polyps were resuspended in 1 ml of ice-cold RIPA buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.1), 1% NP-40, 0.5% DOC, 0.1% SDS 1 mM PMSF and 1 x Roche complete protease inhibitor EDTA-free) per 0.5 - 2 $\cdot$ 10^7 cells or per $\sim$ 1000 *Hydra* polyps ($\sim$ 10^5cells per polyp). The suspension was incubated on ice for 15 - 30 min and transfered to a 1.5 ml reaction tube. The lysate was cleared by centrifugation in a tabletop centrifuge at top speed and 4 °C for 15 minutes. The supernatant was transfered to a fresh tube with 30 μl of 50 % protein A-agarose slurry for preclearing. The mixture was incubated on a rotator for 30 minutes at 4 °C. The protein A-sepharose was removed by centrifugation in a tabletop centrifuge at 4 °C for 5 minutes. The cleared lysate was transfrered to a fresh tube.

To immunoprecipitate, the entire lysate was transfered to a tube containing 10 μl of 10% BSA and μl specific antibody bound to protein A-agarose. The mixture was incubated for 2 hours at 4 °C on a rotator. Unbound proteins were removed by centrifugation in a tabletop centrifuge at top speed and 4 °C for 5 seconds. The supernatant containing unbound proteins was aspirated and the beads were washed with 1 ml of ice-cold wash buffer (300 mM NaCl, 50 mM Tris-HCl (pH 7.4), 5 mM EDTA, 0.1% (w/v) Triton X-100) for 5 minutes on a rotator. The protein A-agarose

beads were spun down in a tabletop centrifuge at top speed and 4 °C for 5 seconds. The supernatant was removed. This washing step was repeated four more times. The last washing step was performed with 1 ml of ice-cold 1 x PBS instead of wash buffer. The protein A-agarose pellet was resuspended in an equal volume of 2 x protein sample buffer and used for Western blot analysis. Samples of the different steps during immunoprecipitation were analyzed by Western blotting. Table 3.7 gives an overview of the conditions used for immunoprecipitation and co-immunoprecipitation.

TABLE 3.7 *Conditions used for immunoprecipitation and detection.*

immuno-precipitation technique	**system**	**antibody used for IP**	**antibody used for detection**
IP	Sf9 cells expressing 6His-*Hy*YY1 fusion protein	α-*Hy*YY1	αPenta-His
IP	Sf9 cells expressing 6His-*Hy*YY1 fusion protein	αPenta-His	α-*Hy*YY1
IP	Sf9 cells expressing *Hy*YY1 protein	α-*Hy*YY1 (cross-linked to protein A-agarose)	α-*Hy*YY1
Co-IP	Sf9 cells expressing 6His-*Hy*YY1 fusion protein and *Hy*EZH2, *Hy*EED, *Hy*SUZ12, *Hy*p55	αPenta-His	α-*Hy*EZH2, α-*Hy*EED, α-*Hy*p55
IP	*Hydra*	α-*Hy*YY1 (cross-linked to protein A-agarose)	α-*Hy*YY1
Co-IP	*Hydra*	α-*Hy*YY1 (cross-linked to protein A-agarose)	α-*Hy*EZH2, α-*Hy*EED, α-*Hy*p55

3.10.2. ChIP

To identify DNA binding sites and target genes of *Hy*YY1, chromatin immunoprecipitation (ChIP) followed by sequencing was performed (ChIP-Seq). Formaldehyde cross-linking and chromatin immunoprecipitation were adapted from a protocol optimized for *Drosophila* embryos (Cavalli et al., 1999). Figure 3.13 gives an overview of the ChIP/ChIP-Seq procedure.

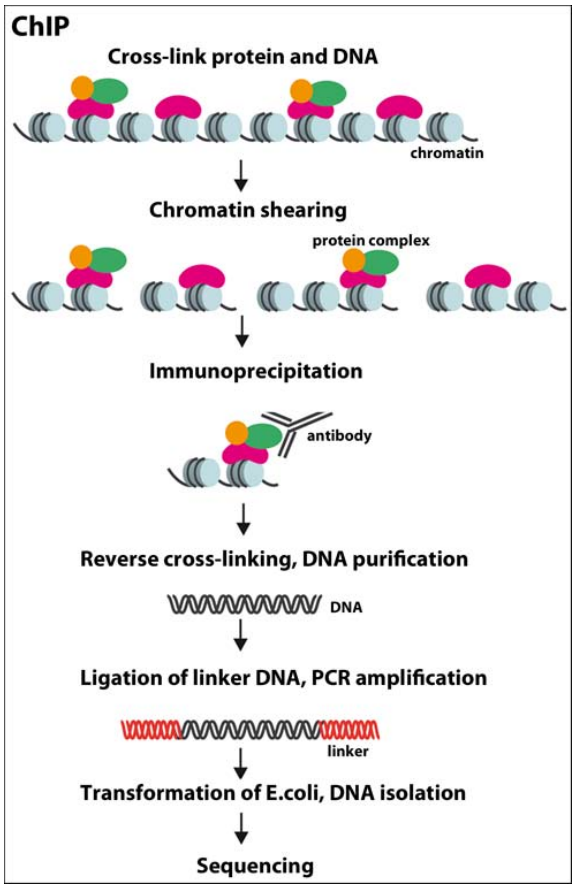


FIGURE 3.13 *Flow chart depicting the different steps of ChIP-Seq.*

Formaldehyde cross-linking of chromatin from *Hydra* polyps

~1 ml densly packed *Hydra* polyps were taken as starting material. The polyps were homogenized in a Dounce homogenizer with 20 strokes in 5 ml of buffer A1 and 1.8% formaldehyde (290 μl of 37% solution) at room temperature. Crosslinking was performed for 15 minutes at room temperature on a roller mixer. Crosslinking was stopped by the addition of glycine solution to 225 mM (540 μl of 2.5 M glycine solution for 6 ml of cross-linked material). The mixture was incubated for 5 minutes at room temperature and transfered to ice. The homogenate was centrifuged at 4000 g and 4 °C for 5 minutes. The supernatant was discarded and the pellet was resuspended in 3 ml of buffer A1. The washing step was repeated three more times. The pellet was washed in 3 ml of lysis buffer without SDS and centrifuged at 4000 g and 4 °C for 5 minutes. The pellet was resuspended in lysis buffer with SDS to 0.1% and N-lauroylsarcosine to 0.5 %. The suspension was incubated for 10 minutes at 4 °C on a rotator. The chromatin was sheared by sonication. The

suspension was sonicated for 8 times 30 seconds with 1 minute intervals on ice using a Branson sonifier 250 equipped with a microtip. The following settings for sonication were tested: power 2, duty cycle 20% or power 3, duty cycle 30%. After sonication, the suspension was rotated for 10 minutes followed by centrifugation for 5 minutes at top speed in a tabletop centrifuge. The supernatant was transfered to a new tube. The pellet was resuspended in 0.5 ml of lysis buffer and rotated for 10 minutes. The suspension was again centrifuged for 5 minutes. The supernatant was combined with the supernatant from the previous centrifugation step. The combined supernatants were centrifuged 2 times for 10 minutes at top speed. The resultant supernatant is the chromatin extract. At this step, the extract may be stored at -80 °C until use. The chromatin extract was washed 4 times with 1 ml of lysis buffer using a Vivaspin 50 column (MWCO 50000). The final volume of the chromatin extract was 1 ml. The chromatin extract was pre-incubated with 100 μl of PAS suspension for 2 hours at 4 °C. The sepharose was removed and the cross-linked chromatin was aliquoted (200 μl aliquots) and stored at -80 °C.

Preparation of protein A-sepharose (PAS) slurry:
100 mg of protein A-sepharose CL-4B (GE Healthcare, 17-0780-01) were resuspended in 1 ml of lysis buffer. The sepharose was washed three times with 1 ml of lysis buffer. The washing was completed within 1 hour. The suspension was centrifuged in a tabletop centrifuge at 2000 rpm for 1 minute. The supernatant was discarded and the PAS was blocked for 2 hours with blocking buffer. Again, the sepharose was washed three times with 1 ml of lysis buffer. The washing steps were completed within 1 hour. The suspension was centrifuged in a tabletop centrifuge at 2000 rpm for 1 minute. The supernatant was discarded and the PAS beads were finally resuspended in 1 ml of lysis buffer. The PAS slurry was aliquoted and stored at 4 °C.

Preparation of the Vivaspin 50 column:
The Vivaspin columns were washed with 2 ml of H_2O by centrifugation at 3000 rpm in a Beckman GS-6KR centrifuge with a swinging-bucket rotor, until the water has passed the column. The column was incubated with 2 ml of blocking buffer for 2 hours at room temperature. The blocking buffer was decanted and the column was washed four times with H_2O.

Purification of cross-linked *Hydra* chromatin

To analyze the amount of DNA and the size, the chromatin had to be reverse cross-linked first. To reverse cross-link the chromatin, proteinase K up to 100 μg/ml (0.5 μl) and SDS to a final concentration of 1% (5 μl) were added to 100 μl of chromatin extract. The mixture was reverse cross-linked for 6 hours at 60 °C, followed by 20 minutes at 70 °C. DNase-free RNase was added to a final concentration of 50 μg/ml (1 μl) and the mixture was incubated for an additional 2 hours at 37 °C. The DNA was extracted by phenol-chloroform. An equal amount of phenol-chloroform-isoamylalcohol (25:24:1) was addded to the chromatin extract. After vigorous mixing, the phases were separated by centrifugation at top speed in a tabletop centrifuge for 5 minutes. The upper aqueous phase was transfered into a new tube and again extracted with phenol-chloroform-isoamylalcohol. The DNA was precipitated with 1/10 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol. The DNA was precipitated for 30 minutes at -20 °C, followed by centrifugation at top speed in a tabletop centrifuge for 15 minutes. The resultant pellet was washed with 70% ethanol and again centrifuged at top speed for 5 minutes. The supernatant was

discarded and the DNA pellet was air dried. The pellet was resuspended in 12 μl of H_2O. The amount of DNA was determined by spectrophotometry using the Nanodrop spectrophotometer, the size of the chromatin fragments was determined by agarose gel electrophoresis using a 1.5 % TAE-agarose gel. Ideally, the chromatin should have a size ranging from 200 to 500 bp.

Chromatin immunoprecipitation

The prepared *Hydra* chromatin was immunoprecipitated by adding the desired antibody to 200 μl of the chromatin extract. Additionally, a control without antibody and a control using the α-*Hy*YY1 preimmuneserum were prepared. After incubation on a rotator for 4 hours at 4 °C, 50 μl of PAS suspension were added to each reaction. The suspension was incubated over night at 4 °C on a rotator. Table 4.8 gives an overview of the antibodies used for ChIP.

TABLE 3.8 *Antibodies used for ChIP.*

antibody	specifications	amount used for precipitation
α-H3K27me3	rabbit; affinity pure (1 mg/ml)	5 μl
α-*Hy*YY1	rabbit; affinity pure (1 mg/ml)	10 μl
α-*Hy*EZH2	rabbit; polyclonal antiserum	20 μl
α-*Hy*PC	rabbit; polyclonal antiserum	20 μl
control	α-*Hy*YY1 preimmuneserum	20 μl

Purification of immunoprecipitated *Hydra* chromatin

The PAS suspension was washed four times with lysis buffer, followed by two wash steps with TE-buffer. Each wash step was performed with 1 ml of the respective buffer for 5 minutes on a rotator at 4 °C. After each wash step, the PAS suspension was centrifuged in a tabletop centrifuge for 1 minute at 2000 rpm and 4 °C and the supernatant was discarded. After the final wash step, the beads were mixed with 100 μl of elution buffer 1. The suspension was incubated at 65 °C for 10 minutes, followed by centrifugation in a tabletop centrifuge for 1 minute at 2000 rpm and 4 °C. The supernatant was transfered to a new reaction tube. The beads were mixed with 150 μl of elution buffer 2. The suspension was centrifuged again and the supernatant was combined with the first supernatant. The combined supernatants are the chromatin precipitate.
The precipitate was reverse cross-linked for 6 hours at 65 °C. 250 μl of proteinase K solution were added and the mixture was incubated for 2 - 3 hours at 50 °C. 55 μl of 4 M LiCl and 500 μl of phenol-chloroform-isoamylalcohol were added. After vigorous mixing, the phases were separated by centrifugation at top speed in a tabletop centrifuge for 5 minutes. The aqueous phase was transfered to a new reaction tube and precipitated with 1 ml of 100% ethanol. After centrifugation in a tabletop centrifuge at top speed for 15 minutes, the supernatant was removed and the DNA pellet was washed with 750 μl of 70% ethanol. After centrifugation for 5 minutes, the precipitate was air dried and dissolved in 25 μl of H_2O.

Ligation of a phosphorylated linker DNA

At this stage, the DNA consists of fragments with various 5'- and 3'-ends, which result from the random shearing of the chromatin by sonication. For PCR amplification, the fragments were end-repaired and a short linker DNA with primer binding sites was ligated.
For the generation of blunt ends, 3 μl of T4 DNA ligase buffer and 1 μl (10 U) of polynucleotide kinase (PNK) were added to the DNA solution. The reaction was incubated at 37 °C for 30 minutes. The PNK was inactivated at 68 °C for 20 minutes. After the solution has cooled down to 37 °C, 1 μl of a 2 mM dNTP-mix and 1 μl (5 U) of Klenow fragment were added. The reaction was incubated at 37 °C for 30 minutes. The enzyme was inactivated for 10 minutes at 70 °C. The solution was cooled down slowly to 4 °C.
1 μl of 10 mM ATP, 1 μl of 1 μM linker DNA and 4 U of T4 DNA ligase were added to 9 μl of the end-repaired DNA. The ligation reaction was incubated at 4 °C over night.

Preparation of the linker DNA:
The linker DNA consists of a double-stranded oligonucleotide with a phosphorylated 5'-end. The following two oligonucleotides were annealed:
24-mer: 5'-AGA AGC TTG AAT TCG AGC AGT CAG-5' phosphorylated
20-mer: 5'-CTG CTC GAA TTC AAG CTT CT-3'
20 μl of 100 μM 24-mer oligo and 100 μM 20-mer oligo were mixed in 160 μl of TE-buffer. The mixture was incubated at 70 °C for 5 minutes and cooled down to 55 °C. The oligo solution was incubated for 5 minutes at 55 °C and cooled down slowly to 25 °C. The oligos were incubated at this temperature for 2 hours. After cooling down slowly to 4 °C, the linker DNA was stored at -20 °C until use.

Amplification of the immunoprecipitated DNA by PCR and ligation into pGEM-T vector

The ligation reaction was directly used for PCR amplification. For PCR amplification 12 μl of the ligation reaction were mixed with 2 μl 10 mM dNTP-mix, 2 μl of 1 μM 20-mer oligo and 0.5 μl (2.5 U) of EuroTaq DNA polymerase in 1 x RB-buffer containing 1.5 mM $MgCl_2$ (total volume: 80 μl). The DNA fragments were amplified using the following cycling conditions:

initial denaturation	94 °C	2 minutes	
denaturation	94 °C	1 minute	
annealing	55 °C	1 minute	35 cycles
elongation	72 °C	3 minutes	
final elongation	72 °C	7 minutes	

The PCR products were purified using Promega's 'Wizard SV Gel and PCR Clean-Up System'. The amount of DNA was determined by spectrophotometry using the Nanodrop spectrophotometer, the size of the PCR products was determined by agarose gel electrophoresis using a 1.5 % TAE-agarose gel.

The precipitated and amplified DNA was ligated into pGEM-T vector at a vector : insert-ratio of $\sim$1 : 3. The ligation reaction was incubated over night at 4 °C. Chemocompetent *E.coli* DH5α-cells were transformed with 4 μl of the ligation reaction. Positive colonies were analyzed by colony-PCR. The DNA of positive clones was isolated using Promega's 'Wizard Plus SV Minipreps DNA Purification System'. The purified DNA was used for sequencing.
For a detailed description of PCR product purification, ligation, transformation, colony-PCR, DNA minipreparation and sequencing see section 3.5.

Solutions for chromatin immunoprecipitation

37% formaldehyde:	Dissolve 11.1 g paraformaldehyde in H_2O at 80 °C. Add 40 μl 10 M KOH. Adjust volume to 30 ml with H_2O. Freeze 1 ml aliquots at -80 °C.
2.5 M glycine:	9.4 g glycine Adjust volume to 50 ml with H_2O.
10% N-lauroylsarcosine:	5 g N-lauroylsarcosine Adjust volume to 50 ml with H_2O. Sterilize by autoclaving.
1 M sodium butyrat:	Dissolve 1 M sodium butyrate in H_2O. Store at 4 °C until use.
10% sodium deoxycholate:	5 g sodium deoxycholate Adjust volume to 50 ml with H_2O. Mix until completely dissolved. Store protected from light.
3 M KCl:	11.18 g potassium chloride Adjust volume to 50 ml with H_2O. Sterilize by autoclaving.
2.5 M NaCl	7.31 g sodium chloride Adjust volume to 50 ml with H_2O. Sterilize by autoclaving.
1 M $MgCl_2$:	10.17 g magnesium chloride hexahydrate Adjust volume to 50 ml with H_2O. Sterilize by autoclaving.
1 M HEPES (pH 7.6):	119.1 g HEPES Adjust volume to 450 ml with H_2O. Adjust pH to 7.6 with 1 M KOH. Adjust volume to 500 ml with H_2O. Sterilize by autoclaving.
10% Triton X-100:	Mix 5 ml 100% Triton X-100 in 45 ml of H_2O.
0.5 M EDTA:	18.61 g disodium EDTA Adjust volume to 90 ml with H_2O. Adjust pH to 8.0 with 1 M NaOH. Adjust volume to 100 ml with H_2O. Sterilize by autoclaving.

0.5 M EGTA:	19.02 g EGTA Adjust volume to 90 ml with H_2O. Adjust pH to 8.0 with 1 M NaOH. Adjust volume to 100 ml with H_2O. Sterilize by autoclaving.
25 x Roche complete EDTA-free:	Dissolve 1 tablet in 2 ml of H_2O. Store at -20 °C until use.
Buffer A1:	60 mM KCl, 15 mM NaCl, 4 mM $MgCl_2$, 15 mM HEPES (pH 7.6), 0.5% Triton X-100, 0.5 mM DTT, 10 mM sodium butyrate, 1 x Roche complete EDTA-free Prepare freshly prior to use.
Lysis buffer:	140 mM NaCl, 15 mM HEPES (pH 7.6), 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% sodium deoxycholate, 0.05% SDS, 10 mM sodium butyrate, 1 x Roche complete EDTA-free Prepare freshly prior to use.
TE-buffer:	10 mM Tris-HCl (pH 8.0), 1 mM EDTA Sterilize by autoclaving.
Elution buffer I:	50 mM Tris (pH 8.0), 10 mM EDTA, 1% SDS
Elution buffer II:	10 mM Tris (pH 8.0), 1 mM EDTA, 0.67% SDS
Blocking buffer:	1 mg/ml BSA, 1 mg/ml yeast tRNA
glycogen:	20 mg/ml in H_2O Aliquot and store at -20 °C.
proteinase K:	20 mg/ml in H_2O Aliquot and store at -20 °C.
RNaseA:	Dissolve 5 mg RNaseA in 0.91 ml 0.001 M sodium acetate (pH 5.2). Inactivate DNases for 15 minutes at 99 °C. Cool slowly down to room temperature. Add 90 μl 1 M Tris-HCl (pH 7.4). Store aliquots at -20 °C.

4. Results and discussion

4.1. Identification of PcG-encoding genes in *Hydra*

PcG genes have been first identified in context of regulating homeotic gene expression in *Drosophila*. Subsequently, PcG proteins have been found in mammals and in plants. Recent data suggest that PcG proteins act in the regulation and maintenance of gene expression patterns involved in the majority of developmental and proliferation processes. PcG proteins are not only linked to homeotic gene silencing, but also to stem cell maintenance or silencing processes based on non-coding RNAs. To investigate the situation in a non-bilaterian animal, the PcG proteins in the cnidarian *Hydra* have been analyzed. During this study, the main focus was on identifying the genes coding for the DNA binding proteins which are important for recruiting the PcG complexes to their target genes. Candidate genes were: *Pleiohomeotic* (Pho/YY1), *Zeste* (Z), *Pipsqueak* (Psq) and *GAGA-factor/Trithorax-like* (Gaf/Trl). However, it has to be kept in mind, that Zeste, Psq and Gaf/Trl are only found in *Drosophila*. No homolog genes could be identified in vertebrates.
To identify the PcG proteins in *Hydra*, an approach based on genome database screens has been followed. Known PcG proteins from other species like fly, fish, frog, mouse or human were compared with the *Hydra* genome and the best hit sequences with the highest E-values were taken for further analysis. To identify the *Hydra* PcG genes, the intron-exon boundaries have been identified and the DNA sequences of 'core fragments' of the *Hydra* genes could be assembled. These core fragments were cloned from *Hydra* cDNA. The full length coding regions and the 5'- and 3'-ends of the genes were obtained by RACE-PCR.

4.1.1. Screening the *Hydra* genome for PcG genes

To identify genes that are homologous to vertebrate and fly *PcG* genes, the *Hydra* genome was browsed on the *Hydra* genome database Metazome (http://hydrazome.metazome.net/). Known protein sequences from different species have been retrieved from the NCBI protein database (http://www.ncbi.nlm.nih.gov/protein) and used to scan the *Hydra* genome with the help of a tBLASTn search using default parameters. The selected protein sequences were compared to the translated nucleotide database of the *Hydra* genome. Table 4.1 provides an overview of the protein sequences used in the search of PcG asso-

ciated genes and DNA binding factors.

TABLE 4.1 *Protein sequences from different species which have been used for the search of Hydra YY1, SCM and PcG associated DNA binding factors are indicated. The identification numbers of the NCBI or GenBank reference sequences are given. The sequences have been retrieved from the NCBI protein database. For Trl, Psq and Zeste only the known Drosophila proteins could be used.*

Protein	***Drosophila melanogaster***	***danio rerio***	***Xenopus laevis***	***mus musculus***	***homo sapiens***
YY1	pleiohomeotic GenBank AAC39123.1 520 aa	YY1 NCBI Ref Seq NP_997782.1 357 aa	YY1 NCBI Ref Seq NP_001087404.1 370 aa	YY1 NCBI Ref Seq NP_033563.2 414 aa	YY1 NCBI Ref Seq NP_003394.1 414 aa
Trl/Gaf	trithorax-like NCBI Ref Seq NP_001034015.1 567 aa	-	-	-	-
Psq	pipsqueak GenBank CAA62475.1 660 aa	-	-	-	-
Z	zeste GenBank AAF45783.1 575 aa	-	-	-	-
Grh	grainyhead GenBank AAM68470.2 784 aa	Grhl2 NCBI Ref Seq NP_001076541.1 554 aa	Grhl1 NCBI Ref Seq NP_00108907.1 609 aa	Grhl1 NCBI Ref Seq NP_001154878.1 618 aa	GRHL1 NCBI Ref Seq NP_937825.2 618 aa
Dsp1	dorsal switch protein 1 NCBI Ref Seq NP_727960.1 385 aa	Hmgb2 GenBank AAI65150.1 388 aa	Hmgb2 NCBI Ref Seq NP_001079387.1 211 aa	Hmgb2 NCBI Ref Seq NP_032278.1 210 aa	HMGB2 GenBank AAI00020.1 209 aa
Scm	sex combs on midleg GenBank AAF54419.2 877 aa	-	Scmh1 NCBI Ref Seq NP_00108740.8 602 aa	Scmh1 GenBank BAA90554 664 aa	SCMH1 NCBI Ref Seq NP_001026864 660 aa
Sfmbt	Scm-related gene containing four mbt domains GenBank AAF53249.2 1220 aa	-	-	Sfmbt1 NCBI Ref Seq NP_062333.1 863 aa	SFMBT1 NCBI Ref Seq NP_057413.2 866 aa

Hits in the *Hydra* genome could be found for almost all proteins, except for Zeste. However, hits with the highest significances and lowest E-values could be found for YY1, SFMBT, SCM and DSP1. All other proteins produced E-values higher than e^{-05}. For example, figure 4.1 shows the graphical presentation of the result of the tBLASTn search for YY1 and SCM, when comparing the respective *Drosophila* proteins with the *Hydra*

genome. For YY1 the best hit with the highest E-value could be found in Contig 38857 on the minus strand for all proteins subjected to the search. The *Drosophila* protein produced the best hit. For SCM, the most promising hits could be found in Contig 37649. These hits were not ranked the most significant hits by E-value, but the aligned sequences spanned a larger region of SCM than the best hit sequences. After the identification of the Contigs, the expressed regions/coding regions (exons) of the genes were identified.
The genomic region of the candidate gene was analyzed for open reading frames and for intron/exon boundaries. A 'core fragment' of the coding region of the candidate gene was assembled and the DNA sequence was translated into the protein sequence. The identified *Hydra* protein sequences, were reverse checked by a NCBI BLASTp search and cloned from *Hydra* cDNA. In table 4.1 a summary of the genome database screen and the reciprocal BLASTp search is given.

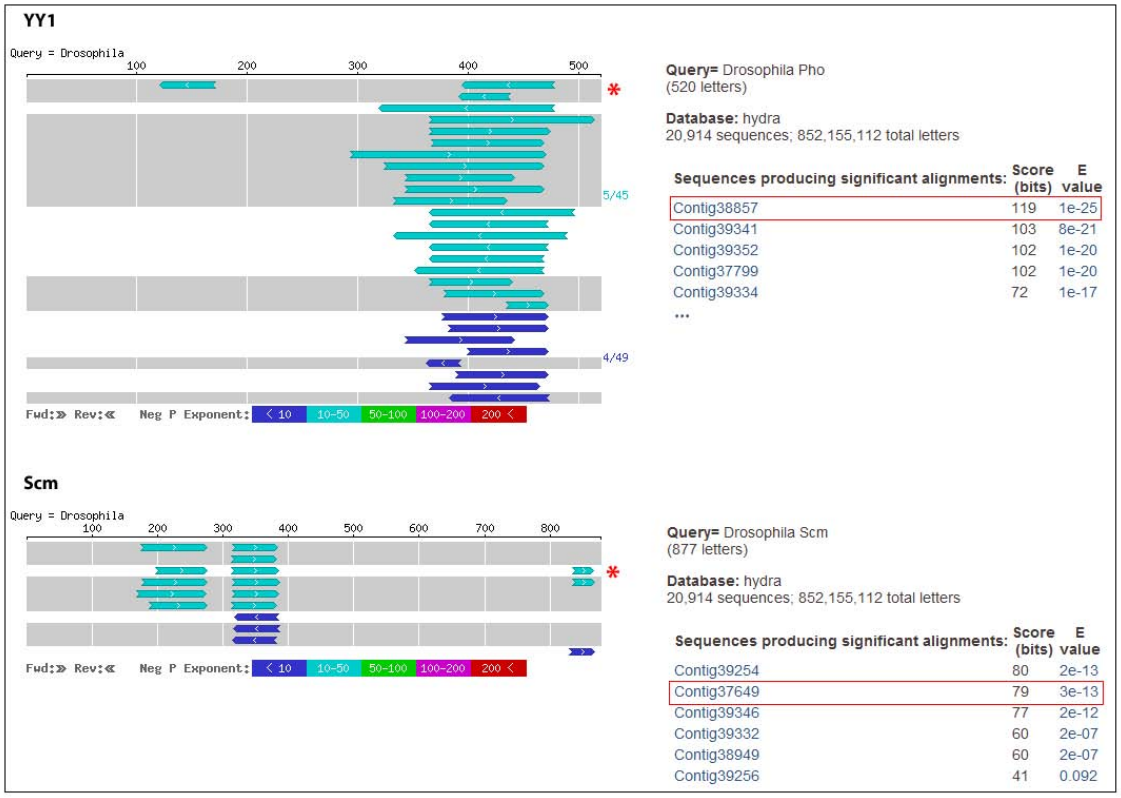


FIGURE 4.1 *Screenshot of the result of a tBLASTn search when browsing the Hydra genome for YY1 and SCM. The* ***top*** *part of the figure shows the result when using the Drosophila Pho protein sequence as a source for the search. In the* ***bottom*** *part of the figure, the genome was searched for Drosophila Scm. The graphical outputs on the left side of the figure show the range and distribution of HSPs (high scoring pairs) of query and target sequence. The texts on the right sides show a part of the summary of target sequences producing significant alignments.*

TABLE 4.2 *Overview of the reciprocal BLASTp search. Fragments of the putative Hydra PcG proteins were used to scan the NCBI database. The best hit sequences and the best hits for species like Drosophila, zebrafish, Xenopus, mouse and human are given.*

Protein	Contig	Best hit on reciprocal NCBI BLASTp					
		best hit	***Drosophila***	***Danio***	***Xenopus***	***mouse***	***human***
YY1	**Contig 38857**	PREDICTED: YY1 (*Hydra*)	pleiohomeotic	YY1	YY1	YY1	YY1
	Score $\geq$ 119	XP_002162851.1	NP_524630.1	NP_997782.1	NP_001087404.1	NP_033563.2	EAW81689.1
	E-value $\leq$ 1e^{-25}	Score =452	Score =329	Score =304	Score =296	Score =293	Score =295
		E-value = 2e^{-125}	E-value =3e^{-62}	E-value =1e^{-80}	E-value =2e^{-78}	E-value =1e^{-77}	E-value =6e^{-78}
Trl/Gaf	**Contig 38766**	PREDICTED: GI22247 (*Hydra*)	kelch	kelch-like 38	-	kelch-like 17	KIAA1129
	Score = 52	XP_002165897.1	NP_724095.1	NP_001038350.1		EDL15091.1	BAA86443.1
	E-value = 3e^{-05}	Score = 182	Score = 61.6	Score = 65.1		Score = 57.4	Score = 58.5
		E-value = 8e^{-45}	E-value = 3e^{-08}	E-value = 3e^{-09}		E-value = 5e^{-07}	E-value = 2e^{-07}
Psq	**Contig 34381**	pipsqueak (*Culex*)	PsqB	-	-	mKIAA1795	hypotetical protein
	Score = 46	XP_001864526.1	AAC47154.1			BAC98252.1	CAD91159.1
	E-value = 0.002	Score = 83.6	Score =82.8			Score = 41.2	Score = 41.2
		E-value = 3e^{-04}	E-value = 0.002			E-value = 0.043 5e^{-07}	E-value = 0.046
Z	no hits found	-	-	-	-	-	-
Grh	**Contig 39050**	PREDICTED: Grhl1 (*Hydra*)	grainy head	Grhl1	Grhl1	Grhl1	GRHL1
	Score $\geq$ 52	XP_002167163.1	NP_476844.2	XP_0012923763.1	AAI69982.1	NP_665897.2	AAH67519.1
	E-value $\leq$ 5e^{-05}	Score = 198	Score = 77.8	Score = 80.1	Score = 78.6	Score = 81.6	Score = 80.9
		E-value = 2e^{-49}	E-value = 4e^{-13}	E-value = 7e^{-14}	E-value = 2e^{-13}	E-value = 3e^{-14}	E-value = 4e^{-14}
Dsp1	**Contig 38846**	PREDICTED: HmgB3b (*Hydra*)	-	Hmgb3	Hmgb3	Hmgb3	HMGB3
	Score $\geq$ 64	NP_002166637.1		NP_001017769.1	NP_001089163.1	NP_032279.1	NP_005333.2
	E-value $\leq$ 2e^{-08}	Score = 373		Score = 122	Score = 191	Score = 179	Score = 177
		E-value = 3e^{-38}		E-value = 1e^{-11}	E-value = 3e^{-11}	E-value = 8e^{-11}	E-value = 2e^{-10}
Scm	**Contig 37649**	similar to predicted protein (*Hydra*)	l(3)mbt	-	-	l(3)mbt-like 4	L(3)MBT-like 4
	Score $\geq$ 67	XP_002160709.1	NP_524529.2			NP_796252.2	AAH39316.1
	E-value $\leq$ 1e^{-09}	Score = 455	Score = 237			Score = 335	Score = 308
		E-value = 2e^{-43}	E-value = 1e^{-18}			E-value = 2e^{-18}	E-value = 4e^{-20}
Sfmbt	**Contig 39254**	similar to predicted protein (*Hydra*)	l(3)mbt-like 2	mbtl1	mbt1	mCG1463	MBTD1
	Score $\geq$ 81	XP_002163058.1	EDL04567.1	NP_001085289.1	XP_001920015.2	EDL15904.1	AAH34364.1
	E-value $\leq$ 1e^{-13}	Score = 1159	403	Score = 366	Score = 316	Score = 315	Score = 315
		E-value = 0.0	E-value 3e^{-82}	E-value = 2e^{-85}	E-value = 6e^{-84}	E-value = 1e^{-83}	E-value = 1e^{-83}

Of all proteins searched for, the putative *Hydra* YY1 homolog gave the best result in the BLASTp analysis. The *Hydra* protein 'core fragment' was identified as the transcription factor *Pho/YY1* in *Drosophila*, *Danio rerio*, *Xenopus laevis*, mouse and human with satisfactory E-values. In *Drosophila*, Pho has been shown to form a third Polycomb repressive complex together with Phol and dSfmbt, which is called Pleiohomeotic Repressive Complex (PhoRC) (Klymenko et al., 2006). A putative homolog of the second member of this complex, dSfmbt, could be found in Contig 39254 in the *Hydra* genome. The sequence of the putative *HySFMBT* gene is distributed over two successive gene models according to the *Hydra* genome browser (http://hydrazome.metazome.net.): *Hma2.231754* and *Hma2.231755*. The gene models are hypothetical protein coding genes within the *Hydra* genome. Over 30000 gene models have been predicted by homology and *ab initio* based gene modeling during the *Hydra* genome project (Chapman et al., 2010). dSfmbt contains four MBT domains and displays a specific histone binding activity (Usui et al., 2000; Wu et al., 2007). The protein has been shown to directly interact with a second PcG protein containing MBT domains: Scm (Grimm et al., 2009). A candidate sequence for Scm could be found in Contig 37649. Scm has been shown to be a PcG protein and to be associated to the PRC1 complex (Peterson et al., 1997; Shao et al., 1999; Saurin et al., 2001). Loss of Scm in fly embryos leads to *HOX* gene misexpression and severe defects with phenotypes typical for the loss of PcG proteins. The reciprocal BLASTp search aligned the *Hydra* protein fragment with another MBT domain-containing protein called l(3)mbt.
Among the candidate genes of the PcG associated DNA binding proteins, only candidates for grainyhead and Dsp1 could be identified in the *Hydra* genome with satisfactory E-values. The putative *Hydra* Trl/Gaf protein turned out to be the *Drosophila* kelch or kelch-like protein during reciprocal BLASTp search. The kelch protein belongs to a member of a protein superfamily with kelch repeats required to maintain actin organization. It shares a BTB/POZ (broad domain, tramtrack, bric a brac) domain with Trl/Gaf. Trl is been classified as PcG and TrxG protein, likewise (Farkas et al., 1994; Mishra et al., 2001). A homolog of the *Drosophila* Zeste protein could not be identified in the *Hydra* genome. Like Trl/Gaf, Zeste is associated with both PcG and TrxG mediated gene regulation (Mulholland et al., 2003).
A similar sequence to the helix-turn-helix (HTH) motif of Pipsqueak could be found in the *Hydra* genome with an unconvincing E-value of 0.003. However, reciprocal BLASTp search aligned the *Hydra* sequence with pipsqueak from mosquito (*Culex quinquefasciatus*) and with Psq isoform B of *Drosophila* with good E-values. The pipsqueak protein has been co-purified from a PRC1 related complex termed CHRASCH (chromatin associated silencing complex for homeotics) (Huang et al., 2002; Huang and Chang, 2004). CHRASCH has been purified from *Drosophila* tissue culture cells and consists of Pc, Ph, Rpd3 and Psq. Psq has been shown to bind the same consensus sequence as Trl/Gaf. A vertebrate homolog is missing.
A candidate sequence for grainyhead could be identified in the *Hydra* genome and verified by reciprocal BLASTp. The BLASTp search aligned the *Hydra* protein fragment as

grainy head/grhl1. Grainyhead has been shown to interact with dRing/Sce of the PRC1 complex and the YY1 transcription factor (Tuckfield et al., 2002; Blastyák et al., 2006). Grainyhead is not essential for the recruitment of PcG proteins to PRE sites in *Drosophila*, but it can assist the binding. In contrast to Trl/Gaf, Psq and Zeste, homologs of grainyhead exist in other species.
For the last possible DNA binding protein Dsp1, a candidate sequence could be identified in the *Hydra* genome. Déjardin et al. (2005) showed that the binding of Dsp1 to PRE sites is important for Polycomb recruitment in *Drosophila*. Sequence motifs for the DNA binding proteins Dsp1, GAF and Pho were sufficient for efficient PcG recruitment to a PRE site. The BLASTp search identified the protein as HMGB3. However, the homolog of Dsp1 would be HMGB2. Both proteins share a so-called HMG (high mobility group) box. HMG box proteins are non-histone components in chromatin with DNA binding activity (for a review see (Stros, 2010)).

4.1.2. Cloning the *Hydra* PcG genes

Of all putative *Hydra* PcG and PcG associated genes, *HyYY1* and *HySCM* were further analyzed during this thesis. The candidate genes were cloned from complementary DNA (cDNA) which was synthesized from *Hydra* total RNA by reverse transcription. In figure 4.2 isolated total RNA which has been separated by agarose gel electrophoresis is depicted. The 28S rRNA of *Hydra* has a size of about 2400 bp, the 18S rRNA is of about 1300 bp in size. The cellular mRNA, which is less abundant and has no specific size, could only be detected as a smear. RACE (rapid amplification of cDNA ends) cDNA was synthesized from total RNA. The resulting pool of cDNAs is flanked by defined oligonucleotides and can be used for the amplification of both terminal ends of the genes. The *Hydra* cDNA used for RACE PCR was a courtesy of Dr. Anne Kuhn (Holstein lab). It was used to amplify *HyYY1* and *HySCM*. For details see materials and methods sections 3.4.2 and 3.5.4.

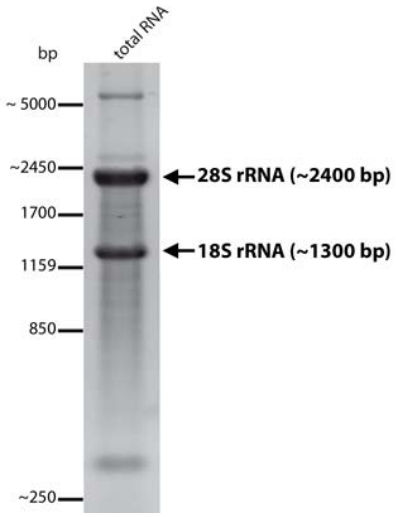


FIGURE 4.2 *Isolated total RNA from Hydra magnipapillata 105. The total RNA was seperated on a 1 % agarose gel and visualized by ethidium bromide staining. Both the 28S and 18S rRNA bands are visible. The size is indicated by a molecular weight marker (left side).*

HyYY1

For *HyYY1* a core fragment of 880 bp in size was amplified from the *Hydra* cDNA and for the 3'-end a fragment of ~ 500 bp in size was obtained. However, the amplified fragment containing the 5'-end was not homogeneous. Several fragments have been amplified by RACE and nested PCR. For the 5'-end DNA fragments ranging from 200 bp to 500 bp in size were obtained (see figure 4.3 A). Nevertheless, the DNA fragments were cloned into pGEM-T vector. The insert sizes were determined by restriction analysis with *ApaI* and *SacI* restriction endonucleases. The endonucleases cut the plasmids adjacent to both ends of the insert and were used to verify the correct insert size (see figure 4.3 B). The nucleotide composition was determined by sequencing. For the *HyYY1* 5'-end fragments, only the sequence of the smaller fragment was found in the *Hydra* genome. The larger fragment could not be identified. Two 3'-end fragments were identified. The smaller fragment is 139 nucleotides shorter and differs in its polyadenylation site.
A full length cDNA sequence of 1546 bp could be assembled for *HyYY1*. The full length sequence contains a 5' UTR of 37 bp, an open reading frame of 1083 bp coding for 361 amino acids and a 3' UTR of 85 bp and a polyadenylation signal. The coding region of *HyYY1* was amplified from cDNA with specific primers and cloned into pGEM-T vector. The pGEM-T-*HyYY1*-vector forms the basis for all further studies with the *HyYY1* gene. An overview of the amplification and cloning of *HyYY1* is given in figure 4.3 C.

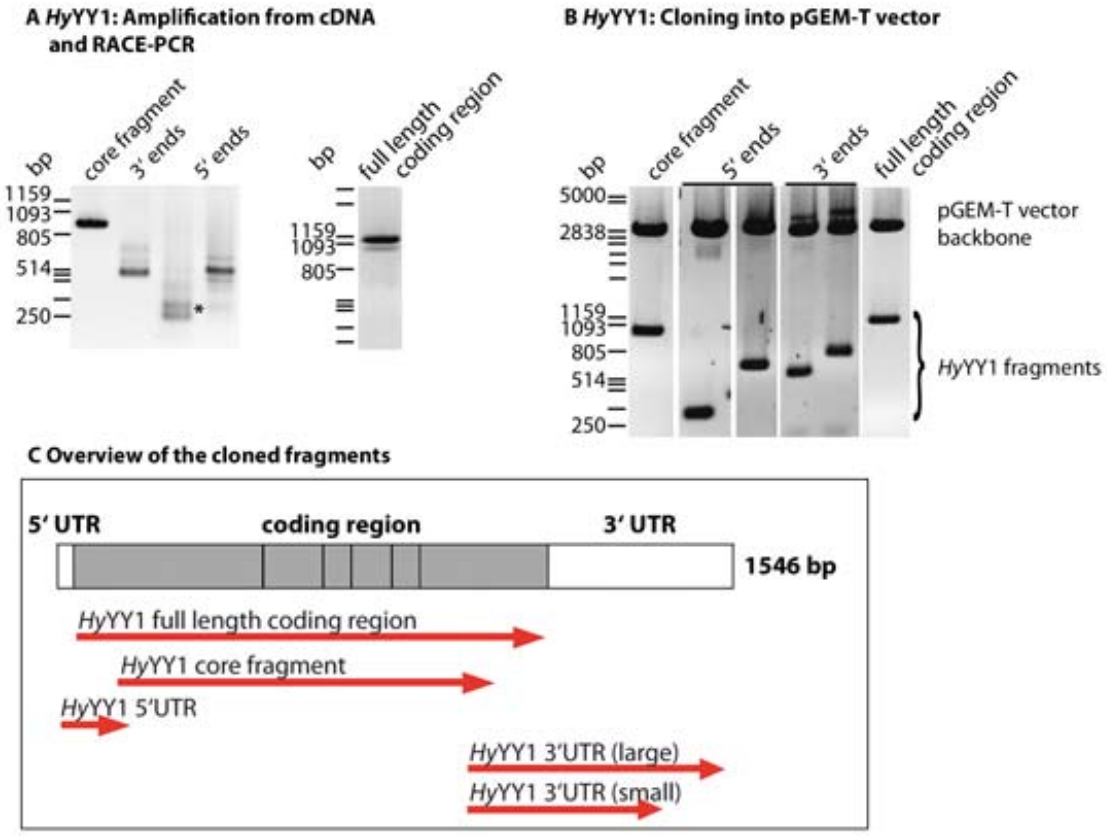


FIGURE 4.3 *Amplification and cloning of HyYY1 from cDNA. The obtained cDNA fragments were seperated by agarose gel electrophoresis.* ***A*** *1% agarose gel depicting the gene fragments of HyYY1 amplified from Hydra cDNA. The size of the fragments was estimated by comparison with a DNA molecular weight marker. Sizes of the marker bands are indicated on the left side of each gel.* ***B*** *1% agarose gels showing the result of the restriction analysis after cloning the HyYY1 gene fragments into pGEM-T vector.* ***C*** *Schematic overview of the cloned HyYY1 cDNA fragments after amplification.*

An overview of the genomic organisation and the nucleotide and amino acid sequences of *HyYY1* is given in figure 4.4. On genomic level, *HyYY1* is organized in 6 exons and spans a genomic region of ∼10.5 kilobases.

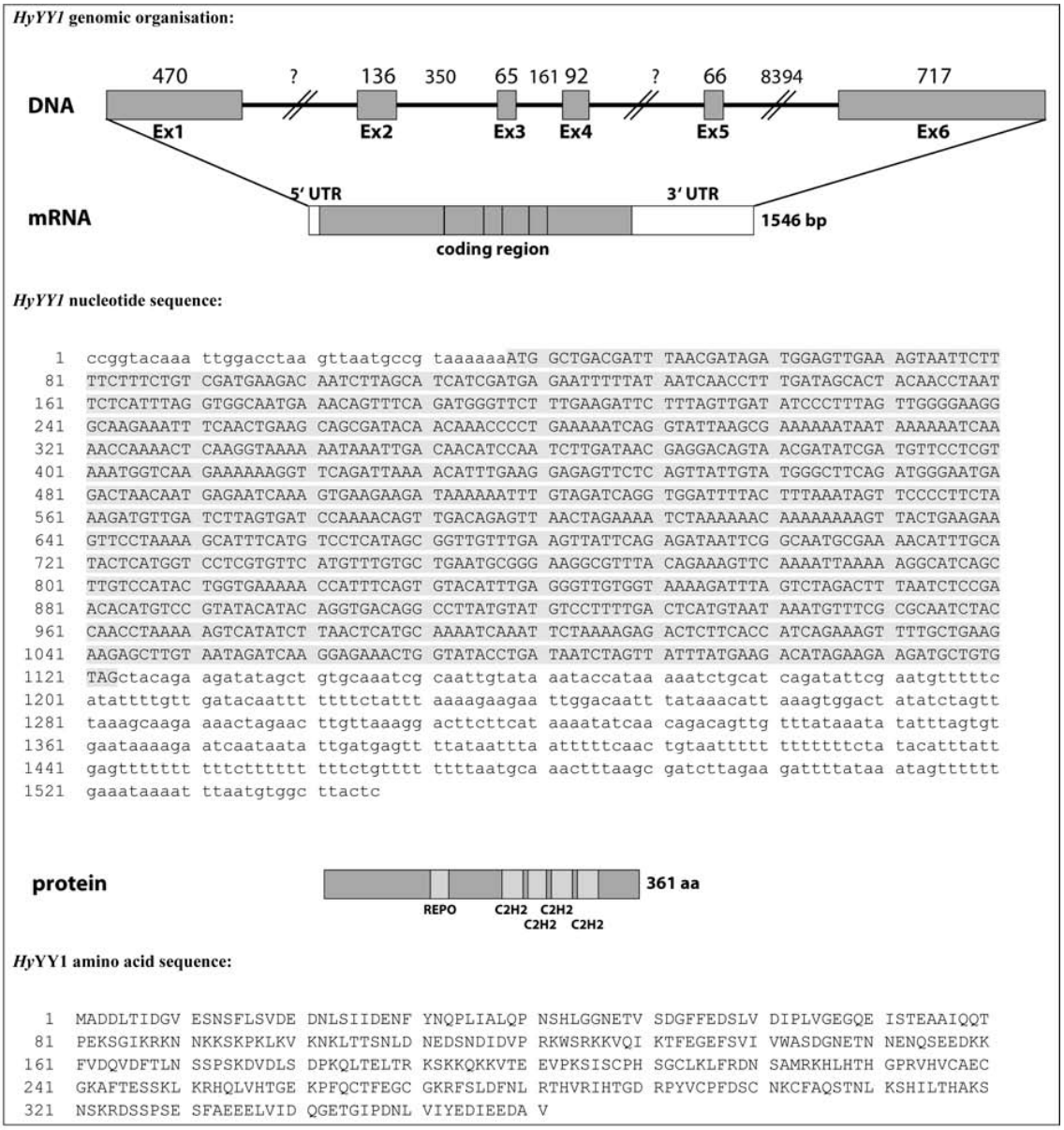


FIGURE 4.4 *Overview of the HyYY1 genomic organisation. Features are indicated: On the level of the genomic DNA, the number and length of the exons are indicated, on the mRNA level, the untranslated regions and the coding region are indicated. The conserved protein domains are displayed on the protein level. The HyYY1 DNA and protein sequence are shown. The DNA sequence includes the coding region of the gene (upper case letters) and the untranslated regions (lower case letters). The coding region is highlighted in gray.*

HySCM

The *HySCM* gene was cloned in the same way as *HyYY1*. The full length cDNA sequence of *HySCM* was obtained by means of PCR with specific primers against a conserved region of the candidate gene, followed by 5′- and 3′-RACE PCR. A core fragment of ∼1700 bp, a fragment containing the 3′-end of ∼200 bp and fragments of varying sizes containing the 5′-end were amplified from cDNA. Sequencing identified a ∼850 bp fragment as the correct fragment containing the 5′-end. All other sequenced 5′-end clones could not be found in the *Hydra* genome, at all.

A full length cDNA sequence of 2636 bp could be assembled for *HySCM*. The full length sequence contains a 5′ UTR of 85 bp, an open reading frame of 2496 bp coding for 832 amino acids and a 3′ UTR of 85 bp including a polyadenylation signal. The cloning and amplification process of *HySCM* is depicted in figure 4.5.

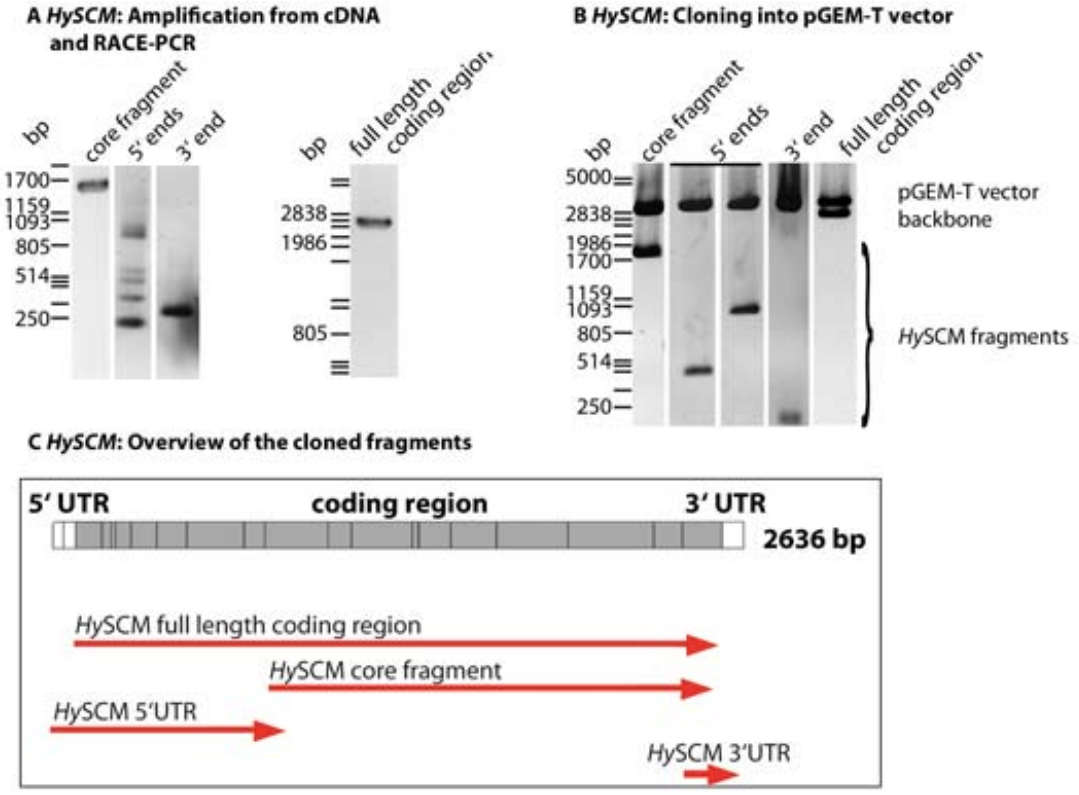
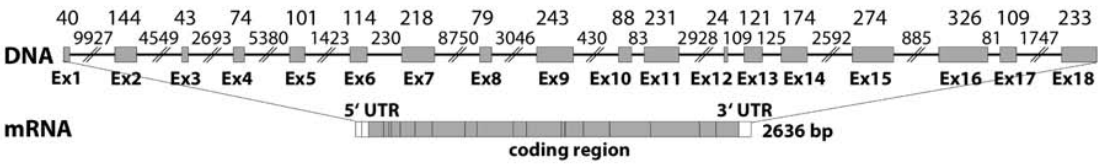


FIGURE 4.5 *Amplification and cloning of HySCM from cDNA. The obtained cDNA fragments were seperated by agarose gel electrophoresis.* ***A*** *1% agarose gel depicting the gene fragments of HySCM amplified from Hydra cDNA.* ***B*** *1% agarose gel showing the result of the restriction analysis with ApaI and SacI restriction endonucleases after cloning the gene fragments into pGEM-T vector.* ***C*** *Schematic overview of the cloned HySCM cDNA fragments after amplification.*

An overview of the genomic organisation and both the nucleotide and amino acid sequences of *HySCM* is given in figure 4.6. *HySCM* is organized in 18 exons and spans a genomic region of > 60 kb.

***HySCM* genomic organisation:**



***HySCM* nucleotide sequence:**

```
   1  aaccaaaaaa aaaatttaat ctatctattt tcttgagaaa gacttgttag gtacactagt aaatgttgct caaaatgatt
  81  gttctATGGA AATGATTAAC GCAAACACAA TTAATTTAAT AAAAAGTGAA CTTTCTTCTC CTCCTTTAAA AAGAAAAAAA
 161  TCTTTTATTG AAGATTCGAA CCAAGATGCA ACAGATTCAG AGCAGGTTTC TGAACACAAC ACATCTCGGA CTAATAATGG
 241  TCATTTTGAG ATGACAAATT TACCAAAGAA GCCTTTGACC CCATATATGA GCTACAGAAA TATGGTTTTC CAAGACATTA
 321  AACACAAACA CCCAGGTTGT TCACCATATG ACTTGGCTTG CGTCATAAGT GCAAAATGGA GTTTGTTAAG TGATGTAGAA
 401  AGAAAGGTAT GGGAAGTAAA GTATGAAAAA GAAATGAAAA GTTATGAAAA AAAGTTGAGA TTGTATCACC AAAAAGCTCT
 481  TAAACGGAAA ACTAATGAAG ATAAAAGTTT TTCACCAATA GGTGATGAAG TCCATGAACT AAACAAATTT GACACATTGG
 561  ATTATTGCTC AGGTTGTAAA AAACTTAAAG AGCTTCAAGG GTTTTCAGCA GAAATTAATG GCAGACAGGT ATATAGAGGG
 641  GATCTTTGTT CTAAAGCATG TTTTTTAGAA GCTTCGAAGC AGTTAAAATG CGTAGATGGA AGTACTTATA TCAAACATAG
 721  TAACAGCGAT CCAGTGTCAG TTAATTTATC ACCATTTAAT TGGGCAGAAT ACTTATCTGA AACAAATTCT GAACCTGCTG
 801  ATTGGATGTA CTGCAGACAG GCTTCTTATC CCCCATCAAA TGCGTTTTCA GTTGGGATGA AACTTGAAGC ACGAAGACAC
 881  GATCAAATAA AAGATTCCAA TCAAATTATT TTTAGTCTTG CTACAGTAGC TTTTATGATT GGTCCTAGGT TGCTTCTTCA
 961  CTTTGATGGC TCTTCTTCAA AGTATGATAT ATGGACACTA TGTGATTCAT GGGATATACA TCCTGTTGGT TGGACTGATC
1041  GTGGCTCTTT GCTTCCACCA ACAGGCTATA GGCATGACCC AAAAAACTAC ATGCAATTTT ATTCTTTAAA GATCCAGAAA
1121  GCAGAACTTG CAACTGCAAG ATGCTTCAAA AAAGTACCTT CAATGCCACC TAAAAATTGT TTTGAGGTTG GAATGAAGTT
1201  AGAGGCAGTA GATAAACAGA ATCCAAGTGT AATAGGTGTG GCCTCTGTTG TTAGAGTTCA AGACGAATTA GTGTATATAG
1281  AGTTTGATGG TTACCCTGGT TTAGGATATT CCTCTCATTA TGGAGATCGT GATTTTTTTC CAGTTGGCTG GTGCGCTCAT
1361  GCAAAGCATC CACTACGAAT TCCAGGTGAT GCAAAAAATA CTTTATCAAA TATTGGATCT TCTTGTTCTA GTTCCAGATA
1441  CTCCCCTAAC TCTTCTTCAA AACTTTCTAT GCATGAAATC GCTATGCATG CAGAAAAAAC TACTCCAATG AAGCCTAAAA
1521  CACTAAATCA AGCGGAAGTT TGTATCAATA TAGAATGTGT TTGTGGCCCT TACATTGATG TAGAAAAACT TAGAAAACTG
1601  ACTCCAATTC GCAAAGGAGA AATCATACAT GTTCTTAAAG GAGTTTTAGA AGACATCATT GCTTGTGCAA TATCACCAAA
1681  AACTGTTATA GGGTTTCTGA AGCCTGGTAG AGGAAATTGT GTAATCACTA CTGTTTGTGA AGGCACAACA TTTCAATGCT
1761  GTCTTACTGT TATTGATCAT GTTTCATCAT TATGGGATAT TCTTAAACAG TTTGCAGACA ATTTAAGGTC TTGTCCTAAC
1841  TTATTTTCAA CTGAAATGTA TGCTTCTTGT CCACGTTGTT TAAACAAACC TTCCACTGCC ATCTCCATAC CATTGATTTA
1921  TCAGAGGTTA ACATCAAACC CTGGTTCTTA TCTGGCTTCT TCTTTTCAAT ATCCTAAAAG CTCTAGGTTT ATTGAATTTG
2001  AAGATGGTGA CATTCATTCA GTTATGGATT TGGAAAAAAA TGAACCATTA AGTTTGAACC AGATAAAGTG TAATCAAGAC
2081  TCTCGGAATC CGGATAAAAA TACTAAAAAA CATGTAACTT GGCCGGCAAT ATCTCAGGCT GAAGTATTTA ATTATCAACC
2161  AGTTTATGCA GATACAGCAT CTCAAACAGA ATCCTTTTGT TTCACTTGTA ATTCTCTTAG TCATAAATGT AGTCTCACAG
2241  ACAAAAATAG GTTAAAAGAT GGTGGTGTTG AAATTGAAAG TGCAGAAATA TTGAGTAATA TCCAAGCTGC TGGTATCGAT
2321  ATTATGAATC GCCCTAGCAT GTGGAATGTG GATGATGTTG TTCTTTTTTT AAAGAAAACT TGCATTAAAG AATTTTCTCA
2401  AGTATTTAAA GACCATGAAA TTGATGGTAA AGCCCTACTT CTTTTGCAAA ATGAGCATAT TCTTCATCAT ATGGGTTTTA
2481  AACTCGGACC AGCCGTCAAA CTTCTGGATT TGATTGAAGA TTTGAAGGTT GCGGAACGTT CTTTTTCAAA AACGAAAGAG
2561  TTTATTGATA CTTTGACTCC ATAGgtgaat aagaaatcat ttttaaataa cagcaagcta atatatatga atataataaa
2641  gtaataaata aaaaaaaaaa aaaaaa
```

protein

***HySCM* amino acid sequence:**

```
   1  MEMINANTIN LIKSELSSPP LKRKKSFIED SNQDATDSEQ VSEHNTSRTN NGHFEMTNLP KKPLTPYMSY RNMVFQDIKH
  81  KHPGCSPYDL ACVISAKWSL LSDVERKVWE VKYEKEMKSY EKKLRLYHQK ALKRKTNEDK SFSPIGDEVH ELNKFDTLDY
 161  CSGCKKLKEL QGFSAEINGR QVYRGDLCSK ACFLEASKQL KCVDGSTYIK HSNSDPVSVN LSPFNWAEYL SETNSEPADW
 241  MYCRQASYPP SNAFSVGMKL EARRHDQIKD SNQIIFSLAT VAFMIGPRLL LHFDGSSSKY DIWTLCDSWD IHPVGWTDRG
 321  SLLPPTGYRH DPKNYMQFYS LKIQKAELAT ARCFKKVPSM PPKNCFEVGM KLEAVDKQNP SVIGVASVVR VQDELVYIEF
 401  DGYPGLGYSS HYGDRDFFPV GWCAHAKHPL RIPGDAKNTL SNIGSSCSSS RYSPNSSSKL SMHEIAMHAE KTTPMKPKTL
 481  NQAEVCINIE CVCGPYIDVE KLRKLTPIRK GEIIHVLKGV LEDIIACAIS PKTVIGFLKP GRGNCVITTV CEGTTFQCCL
 561  TVIDHVSSLW DILKQFADNL RSCPNLFSTE MYASCPRCLN KPSTAISIPL IYQRLTSNPG SYLASSFQYP KSSRFIEFED
 641  GDIHSVMDLE KNEPLSLNQI KCNQDSRNPD KNTKKHVTWP AISQAEVFNY QPVYADTASQ TESFCFTCNS LSHKCSLTDK
 721  NRLKDGGVEI ESAEILSNIQ AAGIDIMNRP SMWNVDDVVL FLKKTCIKEF SQVFKDHEID GKALLLLQNE HILHHMGFKL
 801  GPAVKLLDLI EDLKVAERSF SKTKEFIDTL TP
```

FIGURE 4.6 *Overview of the HySCM genomic organisation. Features are indicated. The number and length of the exons, the untranslated regions, the coding region and the conserved protein domains are indicated on the different levels of organisation. The HySCM DNA and protein sequence are shown. The DNA sequence includes the coding region of the gene (upper case letters) and the untranslated regions (lower case letters). The coding region is highlighted in gray.*

4.1.3. Conserved domains of the *Hydra* PcG proteins

During this study, the *Hydra* PcG genes *HyYY1* and *HySCM* have been identified in the genome and amplified from cDNA. Alltogether, the full length clones of 10 *Hydra* PcG genes are available: *HyPC*, *HyPH*, *HyPSC*, *HyRING* and *HySCM* from the *Hydra* PRC1 complex, *HyEZH2*, *HyEED*, *HySUZ12* and *Hyp55* from the PRC2 complex and *HyYY1* from the PhoRC complex. Figure 4.7 gives an overview of all *Hydra* PcG genes, which have been amplified from cDNA. The full length translated regions of each gene were amplified by PCR using specific primers and seperated by agarose gel electrophoresis. The complete sequence information for the *Hydra* PcG genes can be found in appendix A. The *Hydra* PcG genes were analyzed further by *in silico* based methods. The translated protein sequences were analyzed as to their domain composition and compared with PcG proteins from other species.

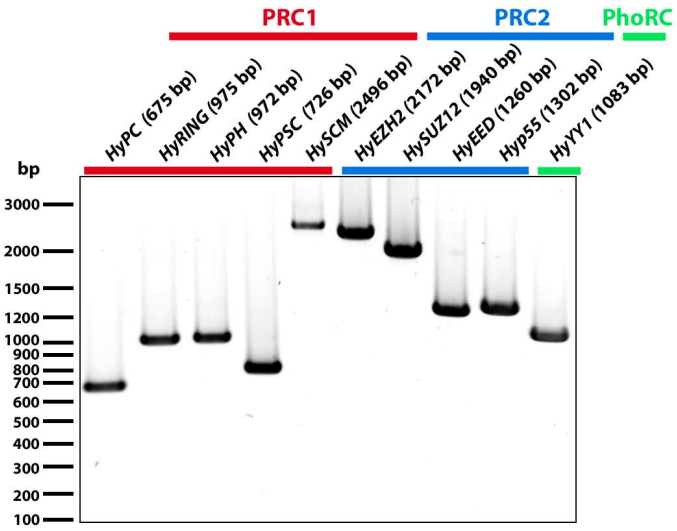


FIGURE 4.7 *Overview of the Hydra PcG genes. The 1% agarose gel shows the translated region of each of the Hydra PcG genes, which have been amplified from cDNA by PCR. The names of the genes and the correct sizes of the coding regions are indicated.*

The domain composition of the *Hydra* PcG proteins has been analyzed *in silico* by comparing the protein sequences with the SMART (Simple Modular Architecture Research Tool) database (Schultz et al., 1998; Letunic et al., 2009). An overview of the domain architecture of *Hydra* PcG protein homologs is depicted in figure 4.8. The *Hydra* PcG proteins are composed of the same functional domains as the human or *Drosophila* proteins, except *Hy*SCM, which has a HMG box at its N-terminus instead of a C2H2-type zinc finger. However, the similarity of the proteins can be mainly found within these domains; the other regions are not conserved. Protein sequence alignments of all identified *Hydra* PcG proteins with homolog proteins of other species like human, mouse, frog, fish or *Drosophila* can be found in appendix A section 3.

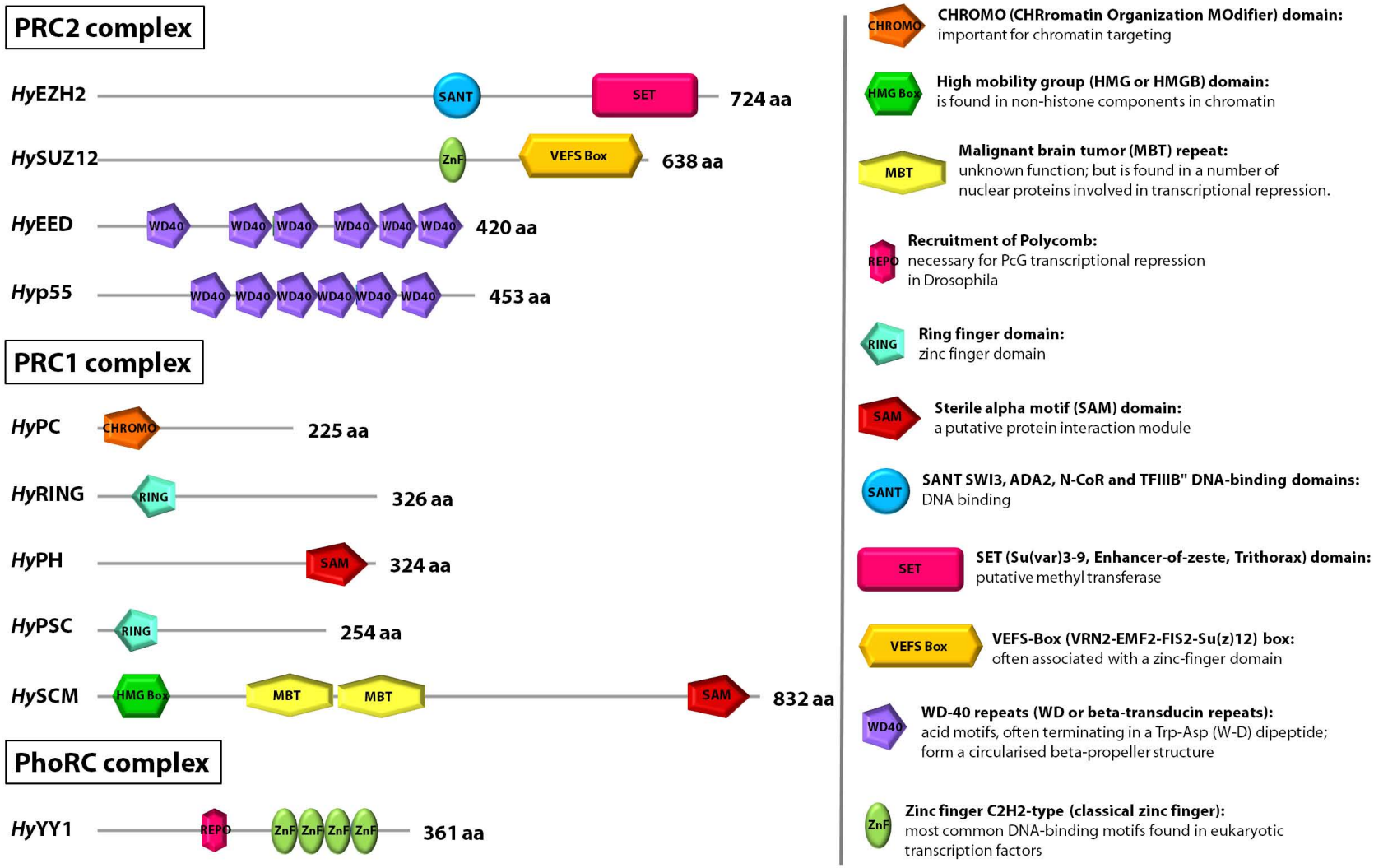


FIGURE 4.8 *Domain architecture of the Hydra PcG proteins.*

PRC1 conserved domains

The functional domains of *Hydra* PRC1 include the CHROMO domain of *Hy*PC, the RING fingers of *Hy*RING and *Hy*PSC and a SAM domain of *Hy*PH. A comparison of the domain architecture of *Hydra*, *Drosophila* and human PRC1 proteins is given in figure 4.9. The CHROMO (CHRomatin Organization MOdifier) domain was first identified in the heterochromatin-associated protein 1 (HP1) and in Pc (Paro and Hogness, 1991; Singh et al., 1991). The domain is necessary for chromatin binding; point mutations in the chromo domain of Pc disrupt chromatin binding (Messmer et al., 1992). Jacobs et al. (2001) could show that the chromo domain binds to methylated lysines on histone tails. Whereas HP1 has been shown to bind methylated H3K9, Fischle et al. (2003) and Min et al. (2003) could show that Pc preferentially binds trimethylated H3K27. Pc is the 'reader' of the epigenetic mark set by the PRC2 complex. The RING zinc fingers of dRing and Psc have been shown to possess H2AK119 E3 ubiquitin ligase activity (Wang et al., 2004c; Cao et al., 2005). The SAM domain which can be found in Polyhomeotic and Sex combs on midleg is a putative protein interaction module. Ph and Scm have been shown to interact with each other via their SAM domains *in vitro* (Peterson et al., 1997, 2004). Additionally, the PRC1 associated protein Scm has two MBT (Malignant Brain Tumor) domains, which could bind to various monomethylated lysines on histones H3 and H4 (H3K4mer, H3K9me, H3K27me, H3K36me and H4K20me) (Grimm et al., 2007; Santiveri et al., 2008).

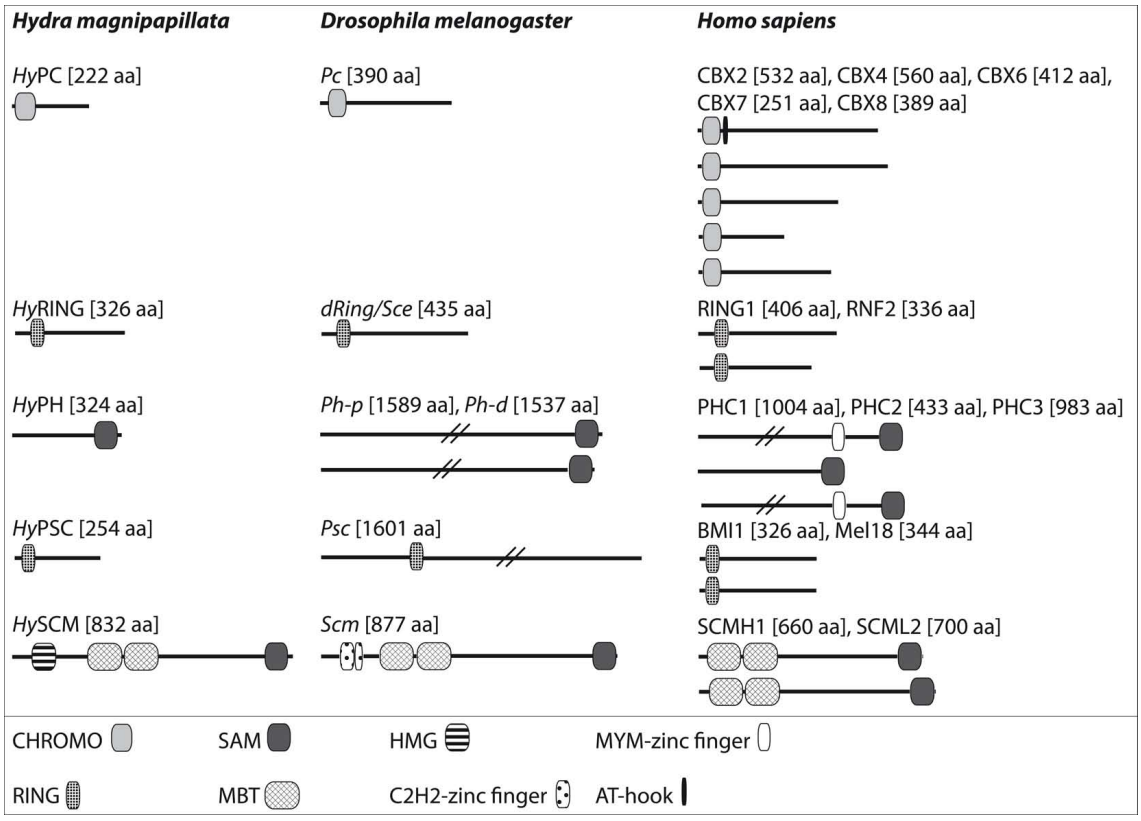


FIGURE 4.9 *Domain architecture of Hydra, Drosophila and human PRC1 proteins. The underlying protein sequences of the human and Drosophila proteins were retrieved from the NCBI database.*

PRC2 conserved domains

The functional domains of the PRC2 proteins include the SANT and SET domain of *Hy*EZH2, a zinc finger and VEFS box of *Hy*SUZ12 and stretches of WD40 repeats in both *Hy*EED and *Hy*p55. A comparison of the domain architecture of the *Hydra* PRC2 proteins with homologs from *Drosophila* and human is depicted in figure 4.10.

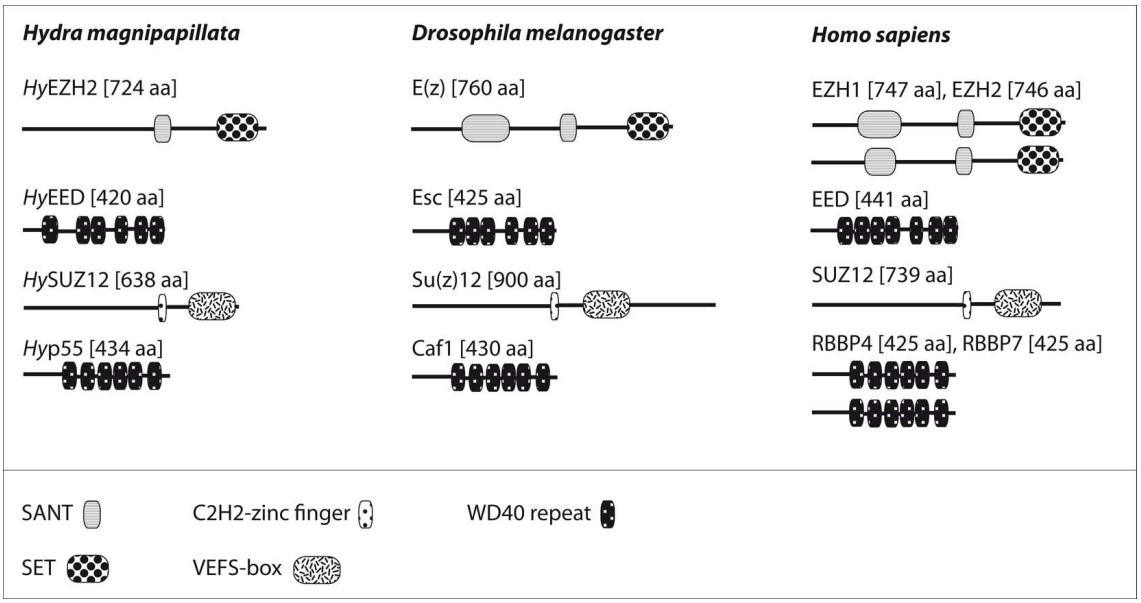


FIGURE 4.10 *Domain architecture of Hydra, Drosophila and human PRC2 proteins. The underlying protein sequences of the human and Drosophila proteins were retrieved from the NCBI database.*

The SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain is responsible for the histone lysine methyl transferase activity of Enhancer of zeste. E(z) has been shown to methylate both H3K9 and H3K27. Mutations in the SET domain disrupt its HKMTase activity (Czermin et al., 2002; Müller et al., 2002). The SANT (SWI3, ADA2, N-CoR and TFIIIB" DNA-binding) domain is similar to the DNA binding domain of c-Myb related proteins (for a review see: Boyer et al. (2004); de la Cruz et al. (2005)). The domain is involved in the recruitment of histone-modifying enzymes to histone tails and in the regulation of their enzymatic activity (Boyer et al., 2002). The exact function of the E(z) SANT domain is yet to be shown. The Supressor of zeste contains a VEFS (VRN2-EMF2-FIS2-Su(z)12) box composed of acidic clusters, which is associated with a zinc finger domain about 100 residues towards the N-terminus. The domain is exclusively found in Su(z)12 and the related plant PcG proteins VRN2, EMF2, FIS2 (Birve et al., 2001). In *Drosophila*, homozygous mutants expressing a truncated protein lacking the VEFS box show extensive misexpression of HOX genes. p55 and Esc/EED have several WD40 repeats. Proteins containing these repeats have diverse biological functions and are important for the assembly of multi-protein complexes (Li and Roberts, 2001). So far, the role of the WD40 domains in the assembly of the PRC2 complex is not known in detail. More recent studies

showed that both p55 and EED have an import function in histone recognition. p55 has been shown to bind to histone H4 via the β-propeller structure of its WD40 repeats (Song et al., 2010). EED has been shown to bind different trimethyllysine histone peptides on histone H3 and histone H1 via the β-propeller architecture of its WD40 repeats (Xu et al., 2010). Moreover, it could be shown that EED differentially regulates the histone methyltransferase activity of the PRC2 complex: binding of EED to H3K27me3 stimulates the HKMTase activity, whereas binding to H1K26me3 inhibits the methylation of H3K27.

PhoRC conserved domains

The conserved protein domains of the third PcG repressive complex, PhoRC, include zinc fingers, MBT and SAM domains. The DNA binding protein Pho/YY1 has four C2H2-type zinc fingers at its C-terminus, which can bind to DNA in a sequence specific manner. Additionally, Pho/YY1 has a short domain of 26 amino acids called REPO (REcruitment of POlycomb) domain. This domain has been shown to be necessary and sufficient for PcG recruitment in *Drosophila* (Wilkinson et al., 2006). The *Hydra* homolog also possesses this domain. A *Hydra* homolog of dSfmbt has been identified *in silico*, but has not been cloned, yet. Like the PRC1 associated protein Scm, dSfmbt contains MBT domains. In contrast to Scm, dSfmbt has four MBT domains and binds to both monomethylated and dimethylated lysine residues on histones H3 and H4 (Grimm et al., 2009). Several human homologs exist. A comparison of the domain architecture of the putative *Hydra* PhoRC proteins with homologs from *Drosophila* and human is depicted in figure 4.11.

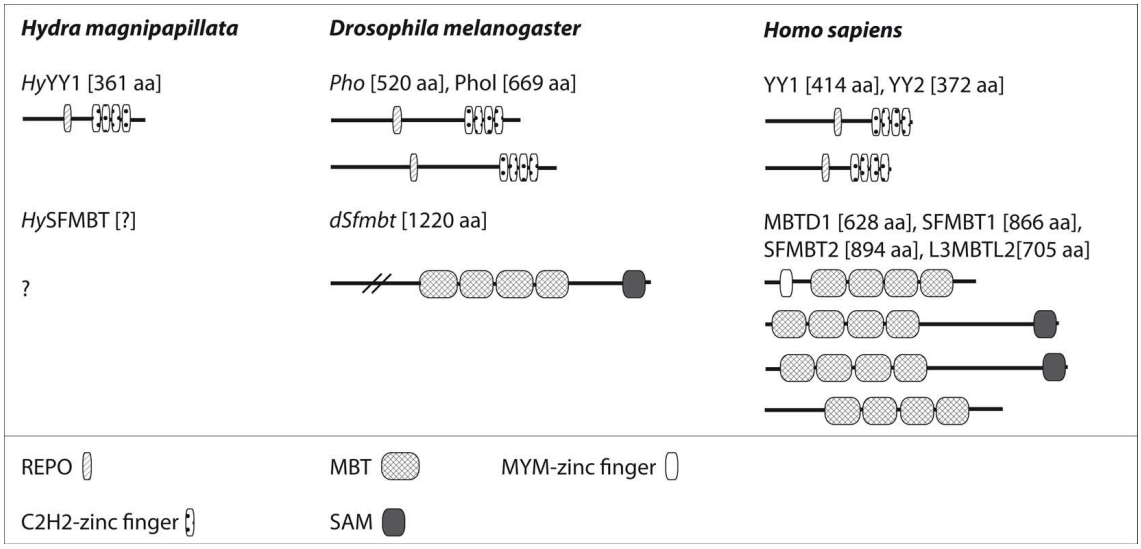


FIGURE 4.11 *Domain architecture of Hydra, Drosophila and human PhoRC proteins. The underlying protein sequences of the human and Drosophila proteins were retrieved from the NCBI database.*

4.1.4. SUMMARY

A complete set of PcG-encoding genes could be identified in *Hydra*

The results of the present thesis and previous studies of this lab show that the freshwater polyp *Hydra magnipapillata 105* has a complete set of PcG proteins. By comparing the sequences of the *Hydra* PcG proteins with known protein sequences from vertebrates, it could be shown that the key domains are conserved between species. However, the surrounding sequences are less similar. These findings are in accordance with a study from Whitcomb et al. (2007) who could show that the PcG proteins of organisms from plant to human differ significantly from each other outside of key domains. In the course of this study, homologs of the DNA binding protein *YY1* and the PRC1 associated gene *SCM* have been successfully identified in the *Hydra* genome and cloned from cDNA.
By *in silico* analysis of the *Hydra* genome only two of the PcG associated DNA binding proteins known from *Drosophila* could be found in *Hydra*: homologs of the *Drosophila* Pho and Dsp1. All other PcG associated DNA binding factors known from *Drosophila* are absent from *Hydra*. Neither Trl/Gaf nor Psq nor Zeste could be identified. Although a possible homolog of Dsp1 has been identified in the *Hydra* genome, it has not been cloned so far. Only one study could link the protein to PcG-mediated gene regulation: In *Drosophila*, Dsp1 has been shown to assist Pho in recruiting the PcG complexes to their target genes (Déjardin et al., 2005). PcG protein binding to the *Fab7* and *Engrailed* Polycomb response element was lost by abolishing the binding of Dsp1. Hence the identification and cloning of Pho/YY1 was preferential.
The human homolog of the *Hy*YY1 protein is a transcription factor that has been initially identified as a repressor of the P5 promoter of the adeno-associated virus (AAV) serotype 2 (Shi et al., 1991). Interestingly, its function could be altered from a repressor to a *trans*-activator by the viral protein E1A. Because of this dual function, the protein has been termed Yin Yang 1 (YY1). Around the same time, Hariharan et al. (1991) identified the protein as the transcription factor δ, which activates transcription of some murine ribosomal proteins. Furthermore, YY1 could be linked to transcription initiation (Seto et al., 1991). YY1 has been shown to interact with a number of transcriptional regulators, including Sp1, TFIIB, TAFII55, TBP, p300 and CBP (Seto et al., 1993; Usheva and Shenk, 1994; Austen et al., 1997). Usheva and Shenk could show that YY1 can direct basal transcription together with TFIIB and polymerase II. YY1 is involved in the regulation of promoters like c-myc, c-fos, Surf-2, IFN-γ, p53 or IL-4 (Riggs et al., 1993; Gualberto et al., 1992; Gaston and Fried, 1994; Ye et al., 1994; Furlong et al., 1996; Guo et al., 2001). It has also been found as a critical regulator of histone genes (Guo et al., 1995; Eliassen et al., 1998).
Satijn et al. (2001) provided the first evidence for an interaction of YY1 with PcG proteins. In *Xenopus* they could show that YY1 interacts with EED, a subunit of PRC2. No interaction between YY1 and PRC1 could be discovered in this study. In *Drosophila* S2 cells however, Mohd-Sarip et al. (2002) could demonstrate that YY1 can interact with

Pc and Ph, which belong to PRC1. A region within the protein that is important for the recruitment of Polycomb to DNA could be identified. Wilkinson et al. (2006) could narrow the region and identify a short conserved motif that is sufficient for PcG recruitment and transcriptional repression. The motif was termed REPO domain (REcruitment of POlycomb). Atchison et al. (2003) could show that the vertebrate YY1 can substitute for *Drosophila* Pho in transcriptional repression. The *Hydra* homolog of Pho/YY1 both exhibits the four conserved zinc fingers, which are important for DNA binding and the REPO domain, which is essential for PcG binding, as well. However, no N-terminal histidine stretches, transactivation or repression domain could be found in *Hy*YY1, when comparing it to the human or mouse homolog (Hariharan et al., 1991; Lee et al., 1994, 1995). In figure 4.12, schematic diagrams of the human YY1 protein and *Hy*YY1 are depicted.

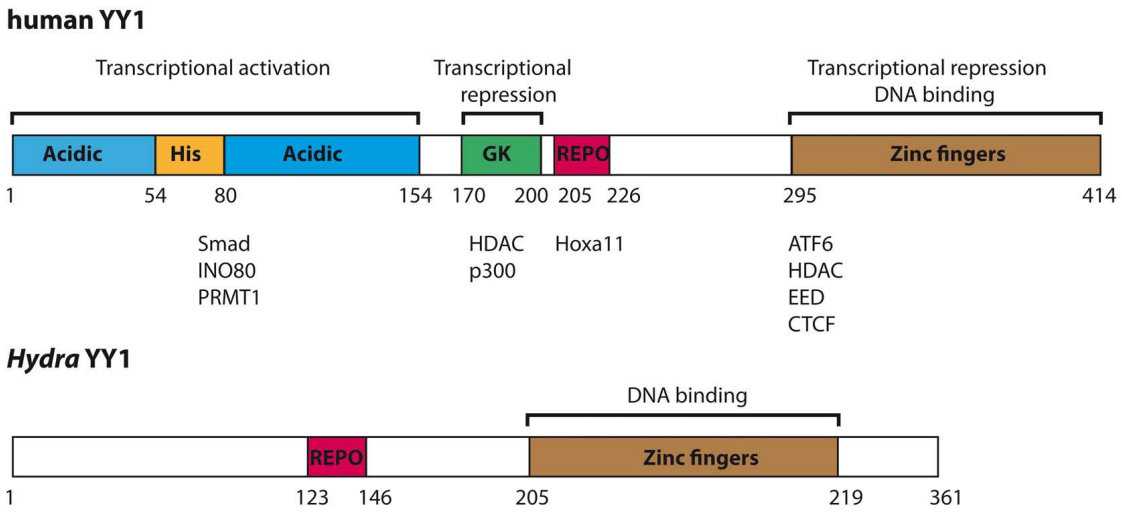


FIGURE 4.12 *Schematic representation of the human and Hydra YY1 protein. The diagram of the human protein has been adapted from (He and Casaccia-Bonnefil, 2008). The transcriptional activation and repression domains and the DNA binding domain are indicated. Acidic, acidic clusters; His, histidine-rich domain; GK, glycine-lysine-rich domain; REPO, recruitment of Polycomb domain. The proteins and complexes that have been demonstrated to interact with the human YY1 protein are shown below each region of interaction.*

In *Drosophila*, Pho can be found in a dINO80 complex or in an assembly together with dSfmbt called PhoRC (Klymenko et al., 2006). A possible homolog of dSfmbt could be found in the *Hydra* genome *in silico*, but has not been cloned from cDNA so far. In contrast, the closely related PcG protein Scm has been successfully cloned. Surprisingly the *Hydra* homolog of Scm has no N-terminal zinc finger domain like its human counterpart, but a HMG-box instead. As well as zinc fingers, HMG-boxes have been found in a wide variety of DNA binding proteins. Originally, the domain has been found in chromatin-associated proteins (for a review see Stros et al. (2007)).

In the past, Scm mutants have been shown to cause strong derepression of homeotic genes in *Drosophila* (Bornemann et al., 1996). Scm was classified as PcG gene and has

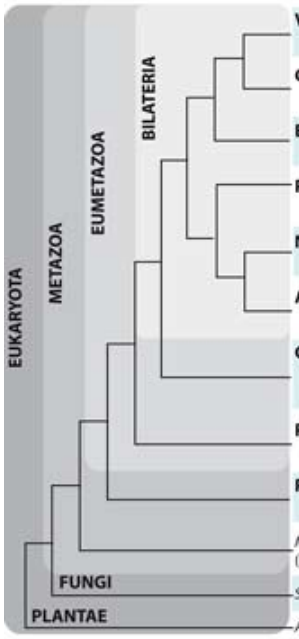
been co-purified with PRC1 in substoichiometric amounts (Saurin et al., 2001). It can interact with the PRC1 subunit Ph via its SAM domain. A PRC1 complex containing SCM could be assembled *in vitro* (Peterson et al., 2004). SCM contains two MBT domains and is related to L(3)MBT (three MBT domains) and SFMBT (four MBT domains). In two recent studies the importance of SCM and SFMBT in recruiting the PcG complexes to their targets begins to reveal. In *Drosophila* Wang et al. (2010) could show that Scm associates with a Polycomb response element upstream of *Ultrabithorax* which is also bound by PRC1, PRC2 and PhoRC. By knocking down Scm the binding of PRC1 and PRC2 to Polycomb response elements is affected. It is hypothesized that Scm may interact with a still unknown DNA binding factor and function in a complex like Pho and dSfmbt in the PhoRC complex. Grimm et al. (2009) could show that Scm and dSfmbt physically interact with each other. Furthermore by ChIP analysis it could be detected that both proteins co-occupy Polycomb response elements of target genes in *Drosophila*. It is suggested that the combined action of Scm and dSfmbt is important for PcG silencing. Results of this study and findings of previous studies of could show that both Scm and dSfmbt can bind to the same set of mono- and dimethylated histone lysine residues via their MBT domains. Both proteins have been found to bind to methylated H3K4, H3K9, H3K27, H3K36 and H4K20, respectively. However dSfmbt seems to bind the modified residues with a higher affinity. It is hypothesized that the heterodimerization of the proteins, each bound to a different histone - are important for chromatin compaction and maintenance of PcG repression of target genes. In this respect further work is required to clone the *Hydra* homolog of SFMBT. In future investigations, *Hy*SFMBT and *Hy*SCM might be interesting for the identification of PcG target genes in *Hydra*.

Polycomb Group proteins during animal evolution

The PcG proteins form large complexes which post-translationally modify histone tails. They act in transcriptional repression of target genes by altering higher order chromatin structure. The PcG protein complexes are evolutionary conserved. Components of the PRC2 complex can be found from plants to animals. However, the complex is absent from unicellular yeast. Components of the PRC1 complex are less conserved and absent from both yeast and plants. In the yeast *Schizosaccharomyces pombe* only one homolog of a protein from the PRC2 complex has been identified so far: Mis16, which has a strong similarity to the human RBBP4 and RBBP6 (Hayashi et al., 2004). In plants, a homolog to the PRC2 complex could be identified (for a review see Hennig and Derkacheva (2009)). An overview of the phylogenetic distribution of the PcG members is shown in figure 4.13. The figure was constructed on the basis of a phylogenetic study of Schuettengruber et al. (2007) and on the basis of genome data.
The phylogenetic tree shows that PcG-encoding genes can be already found in the single-celled choanoflagellate *Monosiga brevicollis*. This finding is supported by a recent study of Shaver et al. (2010) who could identify homologs of PcG proteins in quite a few uni-

cellular organisms. In the demosponge *Amphimedon queenslandica* all components of the PRC2 complex and some members of PRC1 can be found. In the placozoan *Trichoplax adherens* and the cnidarians *Nematostella vectensis* and *Hydra magnipapillata* the set of PcG-encoding genes is completed. However, some of the genes seem to have been lost during evolution in the nematode *Caenorhabditis elegans*. In vertebrates like mouse or human, some of the PcG-encoding genes have been duplicated repeatedly during evolution and multiple homologs exist, especially genes encoding the PRC1 proteins.

Phylogenetic distribution of the Polycomb Repressive Complexes PRC1 and PRC2



	PRC2				PRC1				Reference
	E(z)	Esc	Su(z)12	p55	Pc	Ph	Psc	dRing	
Vertebrate *(Homo sapiens)*	++	+	+	++	+++++	+++	++	++	Grossniklaus and Paro, 2006 *
Cephalochordata *(Branchiostoma lanceolatum)*	+	+	+	+	+	+	+	+	genome data 1
Echinodermata *(Strongylocentrotus purpuratus)*	+	+	+	+	+	+	+	+	Gustafson and Wessel, 2008 *
Platyhelminthes *(Schistosoma mansoni)*	+	+	+	+	-	+	+	+	genome data 2
Nematoda *(Caenorhabditis elegans)*	+	+	-	-	-	+	+	+	Karakuzu et al., 2009
Arthropoda *(Drosophila melanogaster)*	+	+	+	+	+	+	+	+	Grossniklaus and Paro, 2006 *
Cnidaria *(Hydra magnipapillata)*	+	+	+	+	+	+	+	+	this study
(Nematostella vectensis)	+	+	+	+	+	+	?	+	genome data 3
Placozoa *(Trichoplax adherens)*	+	+	+	+	+	+	?	+	genome data 4
Porifera *(Amphimedon queenslandica)*	+	+	+	+	-	+	?	+	genome data 5
Monosiga brevicollis (Choanoflagellate)	+	+	-	+	-	-	?	+	genome data 6
Schizosaccharomyces pombe	-	-	-	+	-	-	-	-	Hayashi et al., 2004
Arabidopsis thaliana	+	+	+	+	-	-	-	+	Hennig and Derkacheva, 2009 *

FIGURE 4.13 *Phylogenetic distribution of the Polycomb Repressive Complexes PRC1 and PRC2. The phylogenetic relationships are depicted on the left hand side and are according to the studies of Srivastava et al. (2010). On the right hand side, the presence or absence of the Polycomb Group proteins in the respective eukaryotic organism is indicated. The presence of the proteins is indicated by +, the absence by -. The data are either retrieved from literature or from the respective genome database. The asterisked references are reviews on the topic. The in silico data are retrieved from the following genome databases by performing a BLASTP or TBLASTN search with the respective human or Hydra protein:*
1 Branchiostoma: http://genome.jgi-psf.org/Brafl1/Brafl1.home.html; cited in Putnam et al. (2008)
2 Schistosoma: http://www.genedb.org/Homepage/Smansoni; cited in Berriman et al. (2009)
3 Nematostella: http://genome.jgi-psf.org/Nemve1/Nemve1.home.html; cited in Putnam et al. (2007)
4 Trichoplax: http://genome.jgi-psf.org/Triad1/Triad1.home.html; cited in Srivastava et al. (2008)
5 Amphimedon: http://spongezome.metazome.net/amphimedon; cited in Srivastava et al. (2010)
6 Monosiga: http://genome.jgi-psf.org/Monbr1/Monbr1.home.html; cited in King et al. (2008)
The retrieved Polycomb Group proteins are hypothetical protein models with a high similarity to the query sequence and the best E value.

4.2. Expression patterns of PcG and associated genes in *Hydra*

RNA *in situ* hybridizations (ISH) were performed to detect the transcript expression pattern of the PcG genes. The emphasis during this thesis was placed on *HyYY1* and the PRC2 complex. In situ hybridizations were performed using whole polyps of the *Hydra magnipapillata 105* strain. The transcripts were hybridized with digoxigenin-labeled RNA probes for 2.5 days at 55 °C. The antisense RNA-probes for *in situ* hybridization were produced by SP6-RNA polymerase from a DNA template. The sense control RNA-probes were reverse transcribed with T7-RNA polymerase. The concentration of the digoxigenin-labeled probes was analyzed by agarose gel electrophoresis since the label interferes with spectrophotometric concentration determination. An overview of the used riboprobes is given in table 4.3 and figure 4.14. The best conditions and probe concentrations have been determined beforehand. For *HyEZH2*, *HyEED* and *HySUZ12* two RNA-probes of about 700 - 800 bp in size have been used simultaneously. For *Hyp55* one probe of about 700 bp has been used, for *HyYY1* either one probe corresponding to the full length translated region of the gene or two probes of about 700 and 400 bp have been used, respectively. All probes have been used at a concentration of 0.2 ng/μl.

TABLE 4.3 *The riboprobes used for in situ hybridization are listed above. The nucleotide lengths and the estimated concentrations of the 'antisense' riboprobes and 'sense' controls are given.*

riboprobe	length	gene	concentration: 'antisense'	'sense' control
ezh2 rp 1	743 bp	*HyEZH2* N-terminus	400 ng/μl	400 ng/μl
ezh2 rp 2	711 bp	*HyEZH2* C-terminus	200 ng/μl	150 ng/μl
eed2 rp 1	742 bp	*HyEED* N-terminus	400 ng/μl	400 ng/μl
eed2 rp 2	751 bp	*HyEED* C-terminus	400 ng/μl	150 ng/μl
suz12 rp 1	713 bp	*HySUZ12* N-terminus	200 ng/μl	200 ng/μl
suz12 rp 2	767 bp	*HySUZ12* C-terminus	200 ng/μl	150 ng/μl
p55 rp 1	719 bp	*Hyp55* C-terminus	400 ng/μl	200 ng/μl
yy1 rp 1	759 bp	*HyYY1* C-terminus	400 ng/μl	200 ng/μl
yy1 rp 2	478 bp	*HyYY1* N-terminus	1000 ng/μl	250 ng/μl
yy1 rp fl	1131 bp	*HyYY1* coding region	200 ng/μl	100 ng/μl

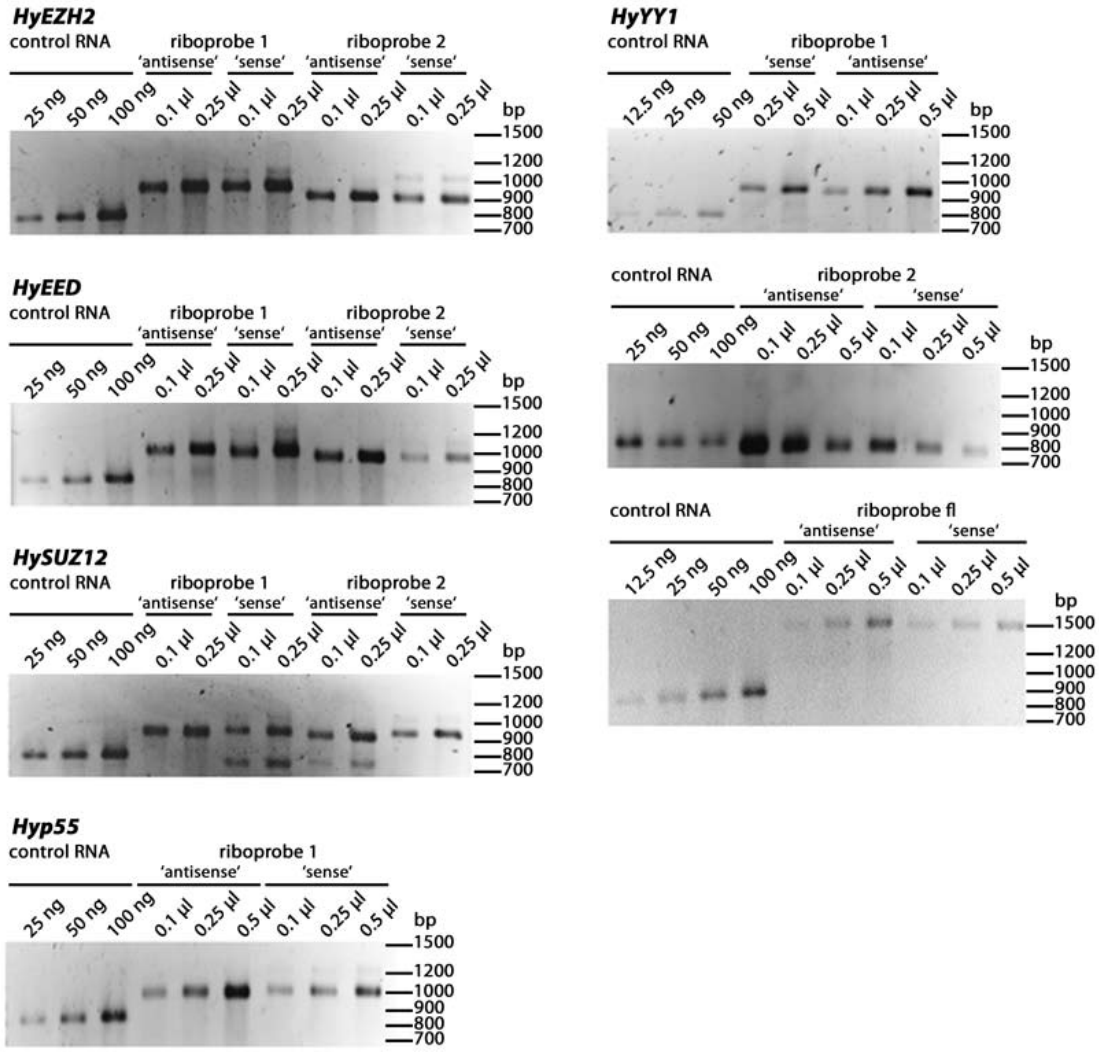


FIGURE 4.14 *Riboprobes for the Hydra PRC2 complex and YY1. Different amounts of RNA-probes have been analyzed on a 1 % agarose gel in 1 x TBE buffer. The concentration of the digoxigenin-labeled probes was estimated in comparison to a control RNA with known concentration. Some of the RNA-probes (ezh2 rp 1 'sense', ezh2 rp 2 'sense', suz12 rp 1 'sense' and suz12 rp 2 'antisense') show a double-band on the gel, probably due to RNA secondary structures.*

4.2.1. Localization of the PcG transcripts in adult *Hydra* polyps

The expression patterns of the PRC2 genes in *Hydra magnipapillata 105* adult polyps are restricted to the body column of *Hydra*. The foot and head region including tentacles and hypostome showed no expression of the PRC2 genes (figure 4.15). Staining could be observed in the interstitial cell lineage. Interstitial cells and clusters of nematoblasts are visible in the respective close-up views. Single cells can be clearly distinguished in polyps stained with freshly prepared NBT-BCIP substrate (figure 4.15 A, B, C and D). In samples stained with BMPurple substrate no single cells could be distinguished (figure 4.15 A', B', C' and D'). The stained region for both substrates remained the same, however with NBT/BCIP the pattern was clearer and the stained cells could be detected

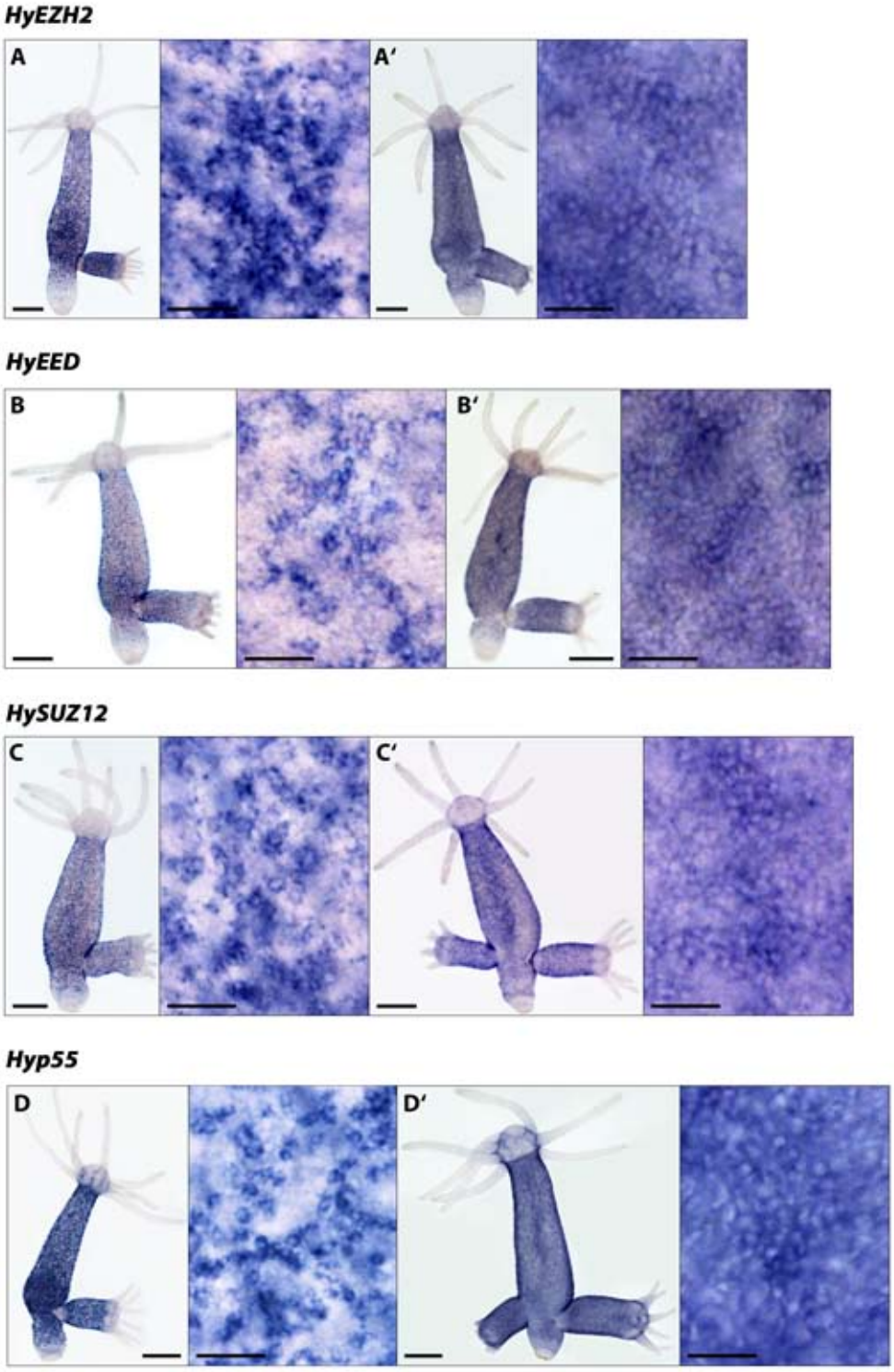


FIGURE 4.15 *In situ hybridization of whole-mount Hydra magnipapillata 105 polyps with anti-sense RNA probes directed against the mRNA species of the Hydra PRC2 genes. In* ***A*** *and* ***A'*** *the expression of HyEZH2 is shown,* ***B*** *and* ***B'*** *shows the expression of HyEED,* ***C*** *and* ***C'*** *shows the expression of HySUZ12 and* ***D*** *and* ***D'*** *the expression of Hyp55. An overview of the expression patterns in whole polyps is shown on the left side of the respective picture (scale bar: 500 μm), a close-up view is given on the right side (scale bar: 50 μm).*

more easily. The enhanced BMPurple substrate, which is also based on NBT/BCIP, led to a more diffuse staining pattern. The cell types could not be distinguished anymore. For the genes of the PRC2 complex, the NBT/BCIP solution has been utilised for all further expression analyses by *in situ* hybridization.
The result for *HyEED* is consistent with the results obtained for *Hydra vulgaris AEP* by Genikhovich et al. (2006). Compared to the other genes of the complex, the strongest expression/staining was observed when hybridizing with *Hyp55*.

The expression of *HyYY1* is limited to the region of the body column and the tentacle base. No expression could be detected in the head or foot region. No specific cell types could be distinguished in a close-up view, which indicates that *HyYY1* is ubiquitously expressed (figure 4.16; top). The expression is not limited to the interstitial cell lineage; *HyYY1* can also be localised in epithelial cells. During the budding process, an increase in the expression level can be detected in the developing bud. A stronger signal could be detected in the endodermal layer of the developing bud and in emerging tentacles (figure 4.16; bottom; red arrows). For these *in situ* hybridizations two riboprobes (yy1 rb1 and yy1 rb2) have been used simultaneously. Staining has been performed using the BMPurple substrate. Using NBT/BCIP substrate resulted in a very faint staining which could not be satisfactorily documented.

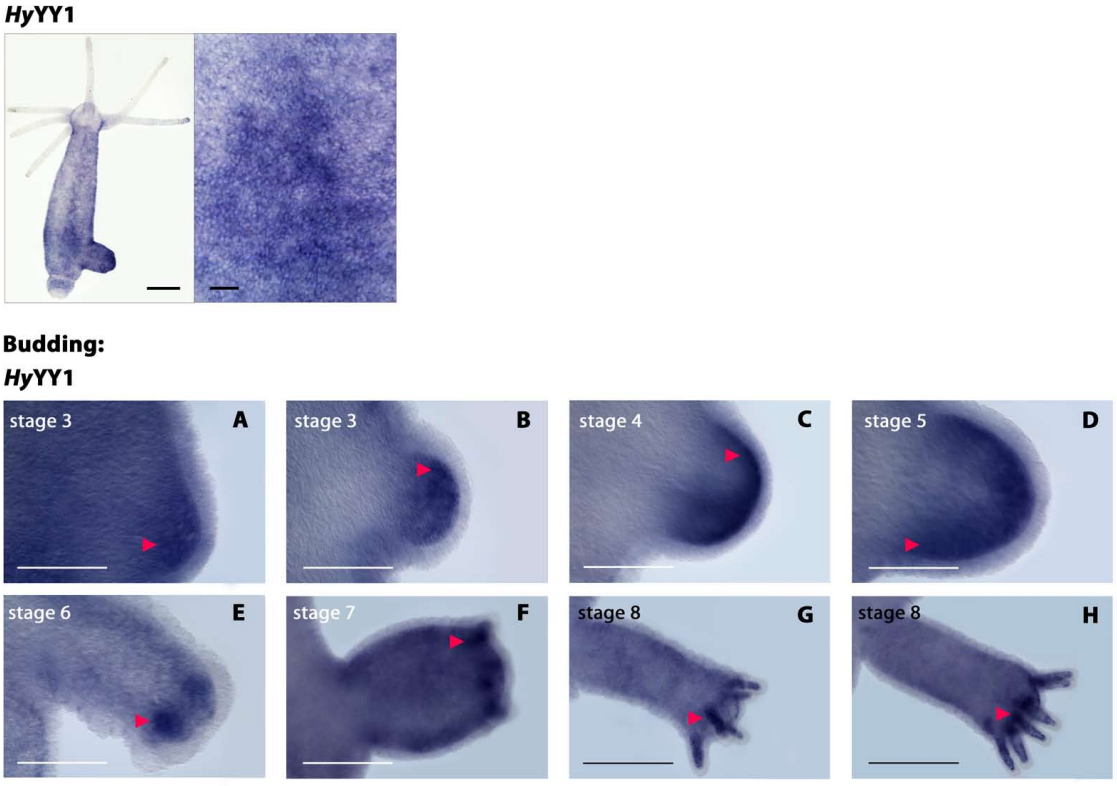


FIGURE 4.16 *In situ hybridization of whole-mount Hydra magnipapillata 105 polyps with anti-sense RNA probes directed against the mRNA species of HyYY1.* ***Top:*** *RNA expression pattern of HyYY1. An overview of the expression pattern in a whole polyps is shown on the left side of the picture (scale bar: 500 μm), a close-up view is given on the right side (scale bar: 50 μm).* ***Bottom:*** *Expression pattern of HyYY1 during the budding process. Red arrows indicate an increasing signal in the endodermal layer of the developing bud and the emerging tentacles.*

4.2.2. PcG gene expression in *Hydra* lacking interstitial cells

To further examine the transcript expression of the PRC2 genes in *Hydra*, polyps lacking interstitial cells have been used for *in situ* hybridization. In 1976 Bode et al. could show that treatment of *Hydra vulgaris* with hydroxyurea leads to a selective depletion of interstitial cells and the generation of so-called 'epithelial animals'. After three days of treatment with 10 mM hydroxyurea, the number of interstitial cells has been reduced significantly. By inhibiting the enzyme ribonucleotide reductase, hydroxyurea either kills the cells during S-phase or prevents the transition from G1 to S-phase (Sinclair, 1965).
According to this study, *Hydra magnipapillata 105* polyps have been treated with 10 mM hydroxyurea for four days to deplete the interstitial cells. Animals which have been treated in this way have been used for *in situ* hybridization with probes against the PRC2 genes and *HyYY1*. In case the expression pattern is restricted to interstitial cells, the expression level should be decreased - if not completely vanished.
In figure 4.17 the expression pattern of the *Hydra* PRC2 genes and *HyYY1* is shown in I cell-depleted polyps and control animals. After drug-induced depletion of the interstitial cells, the transcript expression of the *Hydra* PRC2 genes is dramatically decreased. By comparing the respective control with the interstitial cell-depleted animals, an obvious difference between the transcript expression level can be detected. For *HyEZH2* and *HyEED*, no expression can be detected in hydroxyurea treated polyps at all (figure 4.17 A', B'). For *HySUZ12* only a spot-like staining can be observed which corresponds to an unspecific accumulation of dye (figure 4.17 C'). After treatment with hydroxyurea, expression of *Hyp55* can still be found in some huge single cells throughout the body column (figure 4.17 D'), which resemble egg-restricted stem cells discovered by Littlefield (1991) in *Hydra oligactis*. Treatment of polyps of this species with hydroxyurea eliminated all 'little' interstitial stem cells undergoing somatic differentiation, nematoblasts and neurons. In contrast to these little I cells, the 'big' interstitial cells committed to the gamete lineages showed to be hydroxyurea-resistant. The germ line interstitial stem cells are cycling more slowly than the little interstitial cells. Hence the expression of *Hyp55* is not only present in interstitial cells of the of the somatic lineage but also in the big interstitial cells of the germ line. Even in non-sexual animals a subset of germ line cells can be found. In contrast to the other genes of the *Hydra* PRC2 complex (*HyEZH2*, *HyEED* and*HySUZ12*), *Hyp55* is consequently expressed in the entire interstitial cell lineage. The other genes are not expressed in the egg-restricted cells.
HyYY1 expression can be detected in the body column and in the tentacle base of the control animal ((figure 4.17 E)). An increased signal can be found in the youngest developing bud. The signal is decreased in hydroxyurea treated 'epithelial animals' (figure 4.17 E'). Transcript expression can still be found at epidermal cells of the tentacle base and in the budding region. Nevertheless, the expression level throughout the body column is obviously decreased. These findings are in concordance with the previous results. Whereas the genes of the PRC2 complex are restricted to the I cell lineage, *HyYY1* shows a weaker but broader expression.

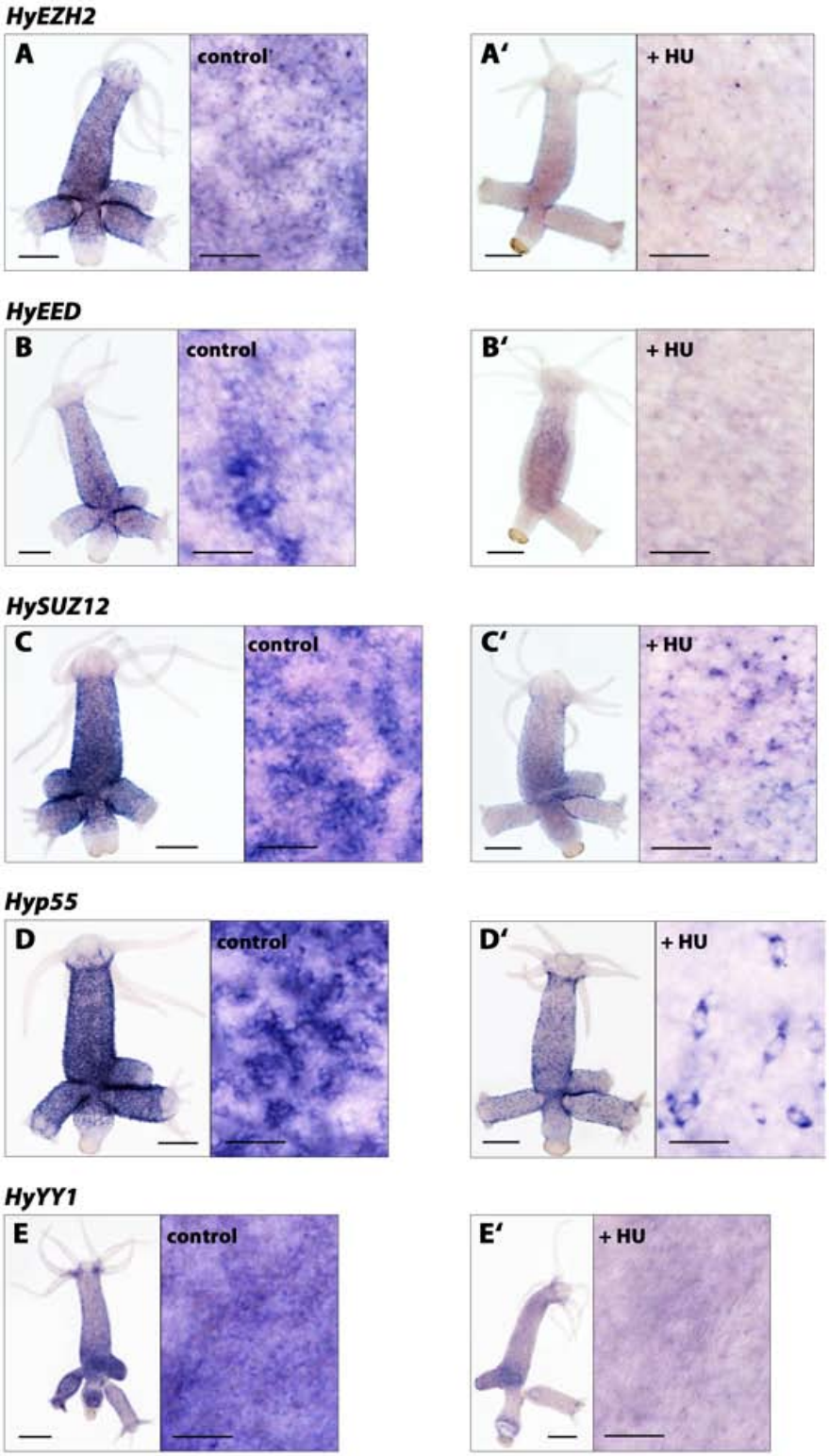


FIGURE 4.17 *In situ hybridization of whole-mount Hydra magnipapillata 105 with anti-sense RNA probes directed against the mRNA species of the Hydra PRC2 genes and HyYY1. The transcript expression is shown for interstitial cell-depleted polyps and control animals. The respective genes are indicated. An overview of the expression patterns in whole polyps is shown on the left side of the respective picture (scale bar: 500 μm) and a close-up view is given on the right side of each picture (scale bar: 50 μm). In **A**, **B**, **C**, **D** and **E** the expression of the genes in control animals is shown. The hydroxyurea-treated, interstitial cell-depleted polyps are shown in **A'**, **B'**, **C'**, **D'** and **E'**.*

4.2.3. PcG gene expression in *Hydra* during regeneration

The freshwater polyp *Hydra* has an enormous regeneration capacity. To test whether any of the *Hydra* PRC2 genes or *HyYY1* play a role in the process of regeneration, *in situ* hybridizations with animals undergoing regeneration have been performed.
For *in situ* hybridization, animals decapitated directly beneath the tentacle zone (80% body length) have been used. The animals have been allowed to regenerate a new head for different time periods (1.5 - 48 hours). Provided that the genes of the *Hydra* PRC2 complex and *HyYY1* are important for regeneration, an upregulation of the transcript expression should be visible at the regenerating tip at some time point after decapitation. Figure 4.18 shows the expression pattern of *Hyp55* and *HyYY1* during head regeneration. *Hyp55* serves as an example for the genes of the *Hydra* PRC2 complex in this context. The expression patterns for *HyEZH2*, *HyEED* and *HySUZ12* are comparable (data not shown). The expression of *Hyp55* is limited to the body column. At no time during the head regeneration process a signal can be detected in the apical tip. The expression of *HyYY1* is upregulated in the apical tip during regeneration. When the tentacles begin to emerge, the signal shifts into the base of the tentacles.

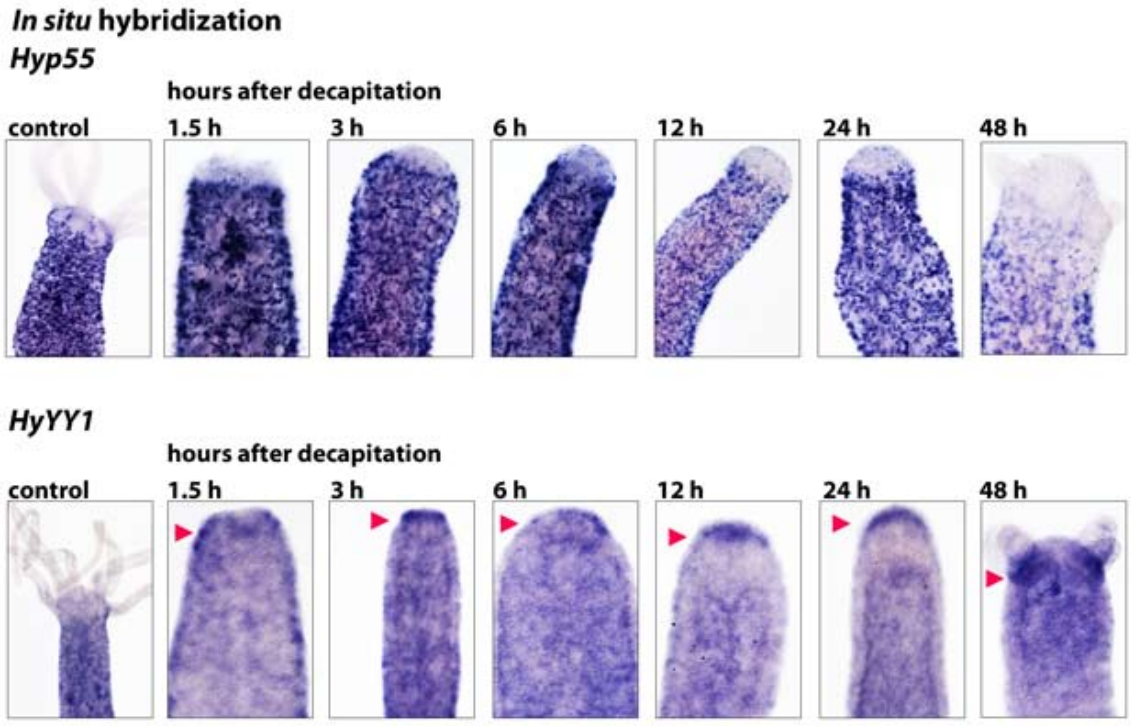


FIGURE 4.18 ***Top:*** *In situ hybridization of regenerating Hydra magnipapillata 105 polyps with antisense RNA probes directed against the mRNA species of Hyp55 and HyYY1. The area of the regenerating apical tip is shown. The different time points after decapitation are indicated.*

It has been shown that *Hydra* can regenerate missing body parts in a rather short period of time. Among the best studied regeneration processes is *Hydra* head regeneration. *Hydra* polyps are able to regenerate a new head structure in 48 - 72 hours. It could be shown that tissue movement of both epithelial layers is responsible for the formation of the new head. No cell proliferation seems to be involved (Cummings and Bode, 1984). The results of the *in situ* hybridization suggest that the genes of the PRC2 complex are not involved in the process of head regeneration. The genes are exclusively expressed in interstitial cells; no upregulation or shift of expression to epithelial cells could be detected. The apical tip of the regenerating animal shows no expression of the PRC2 genes (*HyEZH2*, *HyEED*, *HySUZ12* and *Hyp55*). In contrast, the expression of *HyYY1* is upregulated in the regenerating tip soon after decapitation. After two days of regeneration, new tentacles begin to form and the expression shifts from the apical tip towards the base of the tentacle zone. As could already be seen during the budding process, *HyYY1* is probably involved in tentacle formation.

4.2.4. PcG gene expression in *Hydra* during gametogenesis

To gain inside into the expression of PcG genes in *Hydra* during gametogenesis, sexual polyps of the *Hydra vulgaris AEP* strain have been used for *in situ* hybridization. This strain seasonally produces eggs and sperms. Provided that the genomes of both *Hydra magnipapillata 105* and *Hydra vulgaris AEP* are almost identical in their nucleotide composition, the same RNA-probes have been used. At least one of the PRC2 genes of *Hydra magnipapillata 105* could be compared with its homolog in *Hydra vulgaris AEP*. The *Hydra vulgaris AEP* nucleotide sequence of *HyEED* is available (GenBank AY347258) and the *HyEED* genes of both species show a nucleotide sequence identity of 96%.
The results of the *in situ* hybridization are depicted in figure 4.19. During the process of gametogenesis, the *Hydra* PRC2 genes and *HyYY1* are expressed in the bases of the testes of male polyps. The expression is probably restricted to spermatogonia and spermatocytes; no signal could be localized at the region of mature sperms. *Hyp55* and *HyYY1* are also visible in the egg patch of female polyps. The transcript expression is very intense within the gonads. Due to the lack of animal material, only few stages in female gametogenesis have been probed so far. Up to now, no stages of embryogenesis have been probed.

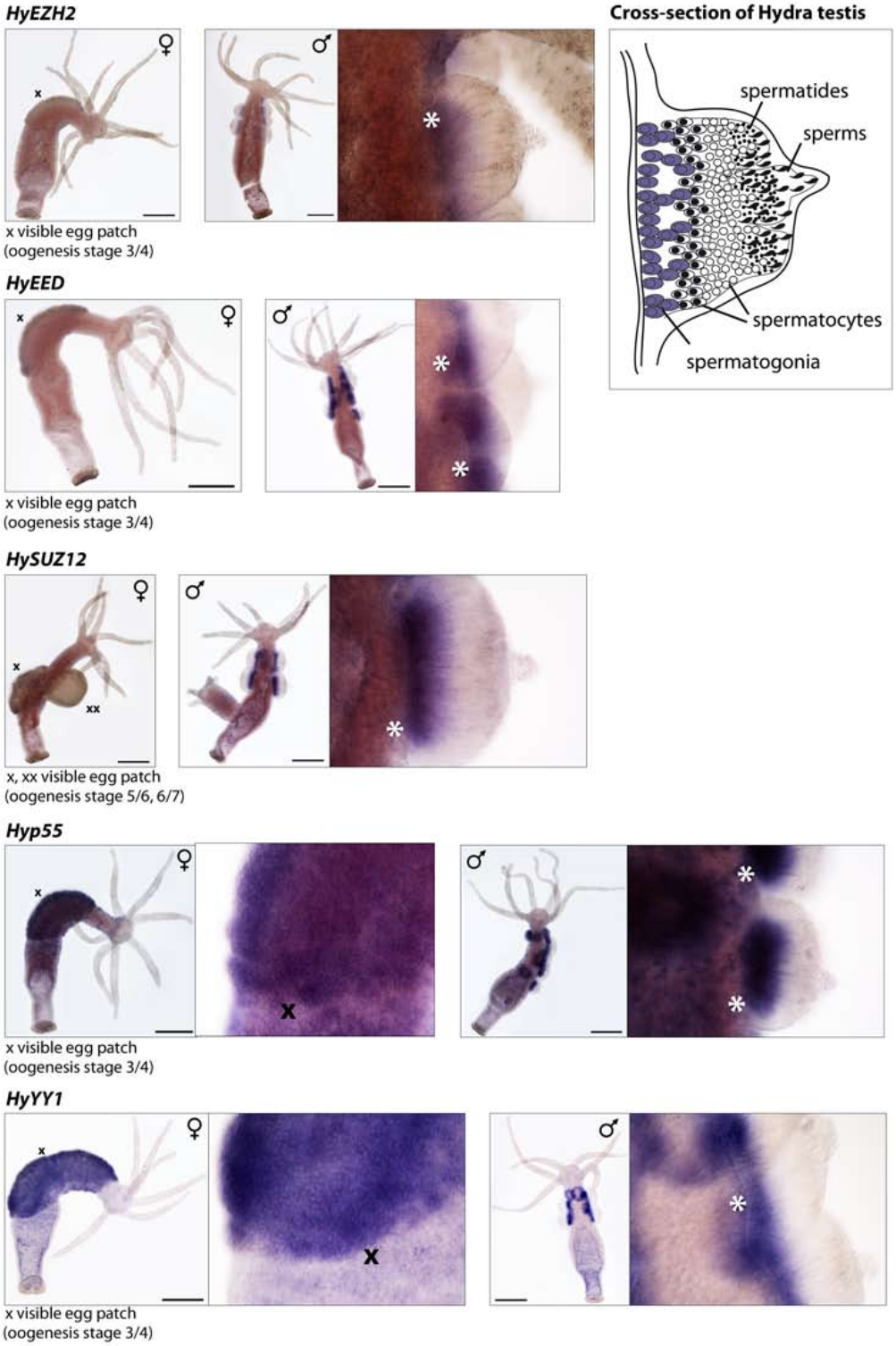


FIGURE 4.19 *In situ hybridization of sexual Hydra vulgaris AEP polyps with anti-sense RNA probes directed against the mRNA species of the Hydra PRC2 genes and HyYY1. The respective genes are indicated. The transcript expression is shown for both male (♂) and female (♀) polyps with clearly visible gonads. The female gonad is visible as a clustered egg patch (x), the male testes are visible as papilla-like protrusions (*). In the right upper corner, a schematic cross-section of the male testis is depicted. An overview of the expression patterns in whole polyps is shown on the left side of the respective picture (scale bar: 500 μm). In case the expression pattern of the genes is found in the gonads, a close-up view is given on the right side.*

4.2.5. SUMMARY

The expression of the *Hydra* PRC2 genes is restricted to the interstitial cell lineage

The freshwater polyp *Hydra* has three coexistent stem cell lines: namely two epithelial stem cell lines and the interstitial stem cells. However, genes of the PRC2 complex in *Hydra* are exclusively expressed in the interstitial cell lineage. No expression could be detected in the epithelial cell lineages. The expression is limited to interstitial stem cells, nematoblasts and germ cells. No expression can be detected in endodermal gland cells or in nerve cells, which accumulate in the tentacle and foot region. This raises the questions: Why are the PRC2 genes of *Hydra* restricted to the interstitial cell lineage?
In higher animals, epigenetic regulation by Polycomb Group proteins has been shown to be important for the maintenance of stem cell activity and for lineage commitment. In 2006 Boyer et al. and Lee et al. have shown that embryonic stem cells lacking PRC2 genes lose their ability to maintain their undifferentiated state. They lose their 'stemness'. Bernstein et al. (2006) and Mikkelsen et al. (2007) could show that the H3K27me3 histone mark set by the PcG proteins is coexistent with another histone mark: H3K4me3. H3K4me3 is set by the Trithorax Group proteins and is associated with active transcription. This 'bivalent' chromatin state in embryonic stem cells has been proposed to maintain PcG target genes in a poised state. Upon differentiation, the bivalent poised state shifts to an active (H3K4me3) or inactive state (H3K27me3). Target genes are expressed in a lineage-specific way. The model is depicted in figure 4.20.

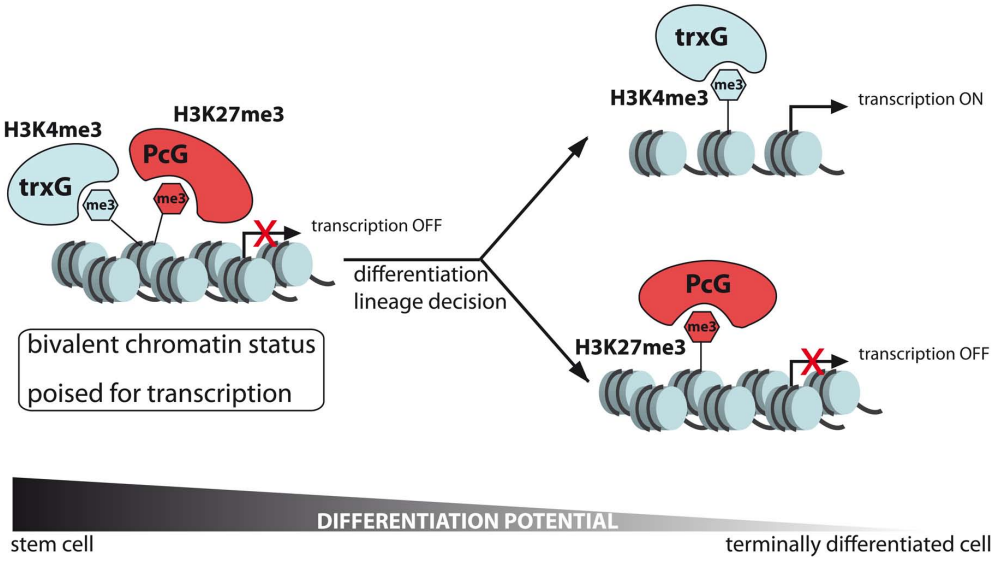


FIGURE 4.20 *The repressive H3K27me3 histone mark coincides with the active H3K4me3 histone mark. The bivalent chromatin state is released upon differentiation.*

In *Hydra*, the PcG genes are expressed in a cell lineage-specific way. They maintain the characteristics of the interstitial cell lineage in contrast to the epithelial cells. Only the terminally differentiated cells of the I cell lineage like neurons, nematocytes and gland cells show no PcG gene expression anymore. Which genes are expressed/repressed in

an I cell lineage-specific way is still unclear and has yet to be determined. Whether the PcG proteins are the only important regulators in the discrimination of the cell lineages is questionable. An interplay between PcG complexes and other regulators like chromatin-remodeling complexes is likely.

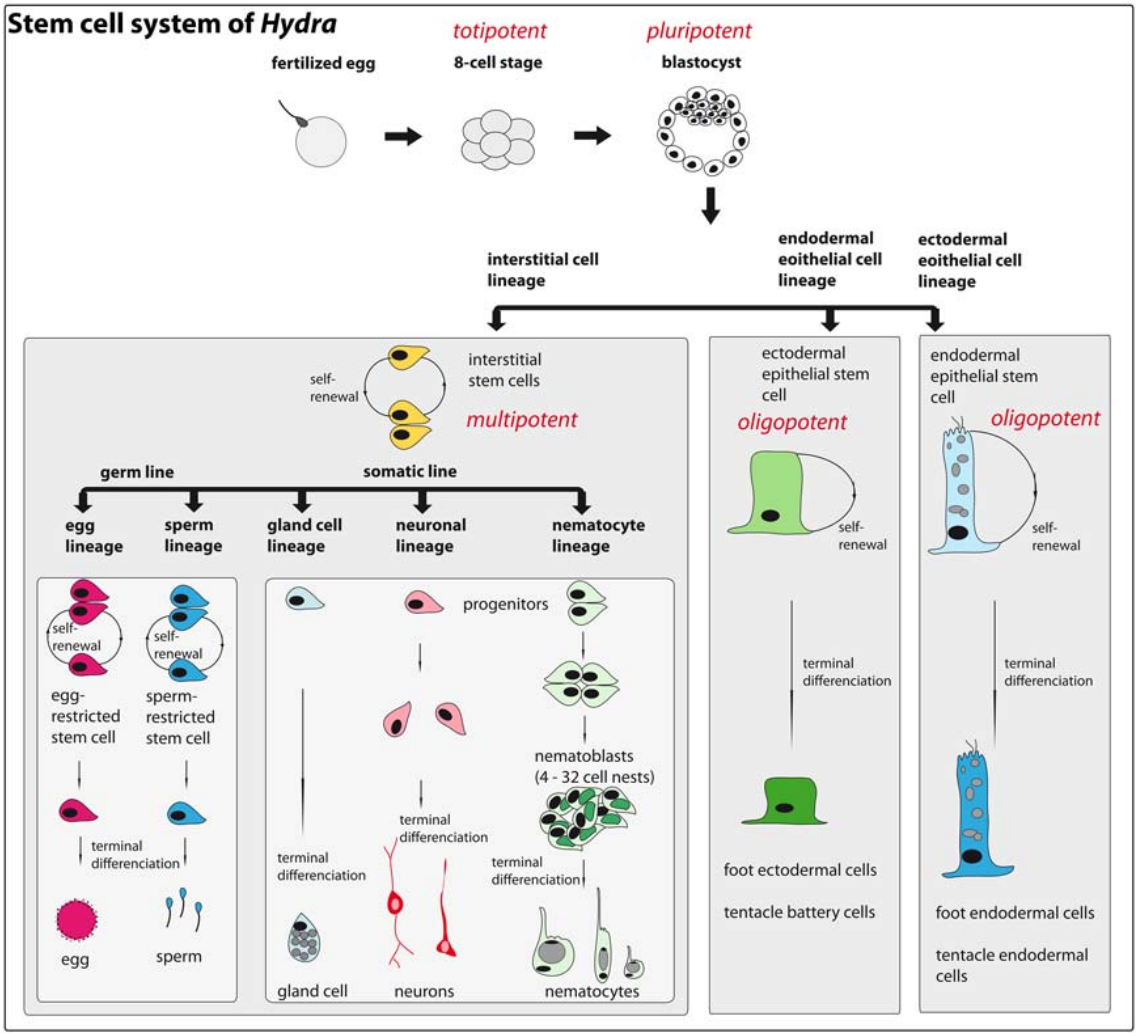


FIGURE 4.21 *Overview of the stem cell system of Hydra.*

Figure 4.21 depicts the stem cell system of *Hydra*. The epithelial stem cells of *Hydra* are oligopotent stem cells, which give rise to terminally differentiated cells of the head and foot region. In contrast to stem cells of higher animals, *Hydra's* epithelial stem cells do not solely act as stem cells but also take over physiological functions which are usually accomplished by differentiated cells. In this respect the epithelial cells are no 'true' stem cells. This properties of the epithelial stem cells in *Hydra* may reflect the evolutionary state (Bode, 2003). At early stages of metazoan evolution the stem cells possibly had a dual function. On the one hand they act as stem cells and on the other hand they still have a physiological function. In contrast, the interstitial stem cells of *Hydra* can be compared

to stem cells of higher animals. They act as 'true' stem cells. They are multipotent and give rise to different classes of differentiation products (gland cells, neurons, nematocytes and gametes).

Hyp55 is additionally expressed in the egg restricted cell lineage

It is interesting to note that *Hyp55* is also expressed in the egg restricted cell lineage. *In situ* hybridization with hydroxyurea treated animals of the *Hydra magnipapillata 105* strain and sexual animals from the *Hydra vulgaris AEP* strain showed that *Hyp55* is additionally expressed in the egg restricted stem cell lineage and in oocytes. In contrast, all other genes of the PRC2 complex are absent from the egg restricted cell lineage. A possible explanation for this result might be that p55 is not only a member of the PRC2 complex. p55 which is also known as Caf-1 (Chromatin assembly factor 1) or Nurf-55 (Nucleosome remodeling factor 55) can also be found in a chromatin-remodeling complex of the ISWI family: NURF. Chromatin-remodeling complexes alter the nucleosome structure in an ATP-dependent manner. There exist four main families: the SWI/SNF family, the ISWI family, the CHD and the INO80 family. Some of the chromatin-remodeling complexes belong to the Trithorax Group complexes which oppose the Polycomb Group complexes. The complexes are both associated with transcriptional activation and repression. The expansion of some gene families of the subunits during evolution allows the combinatorial assembly and formation of several hundred complexes. It seems that each complex functions in a different context. Depending on the composition of its subunits, each complex has only one distinct function (Ho and Crabtree, 2010).
The NURF complex is composed of the following subunits: the ATPase Iswi, Nurf-38, Nurf-55 (p55) and Nurf-301 in *Drosophila* and the ATPase SNF2L, RbBP4 or RbBP7 (p55) and BPTF in humans, respectively. The complex was first isolated from *Drosophila* embryos as a nucleosome remodeling factor which acts in concert with the transcription factor GAGA to alter chromatin structure (Tsukiyama and Wu, 1995). Martínez-Balbás et al. (1998) isolated Nurf-55 as a component of this complex. Some years later the human ortholog of the NURF complex has been purified and possibly plays a role in neuronal development by the regulation of *Engrailed* (Barak et al., 2003). The complex has been associated with the H3K4me3 methylation mark which can be recognized via the PHD finger of BPTF (Li et al., 2006; Wysocka et al., 2006). In *Drosophila* it has been shown that the NURF complex is involved in oogenesis. Deuring et al. (2000) could show that loss of maternal Iswi blocks oogenesis. Therefor the expression of *Hyp55* in the egg restricted cell lineagein *Hydra* is likely. It can therefore be assumed that *Hyp55* does not function as a member of the PRC2 complex in egg restricted cells but as a member of the chromatin-remodeling complex NURF. Possible homologs of the other members of NURF, like the ATPase subunit SNF2L and BPTF - could be identified in the *Hydra* genome *in silico*.

4.3. Antibody production and purification

PcG and associated proteins in *Hydra* have been identified, and the mRNA expression pattern of the *Hydra* PRC2 genes and *HyYY1* has been analyzed by means of *in situ* hybridization. To further characterize the *Hydra* PcG proteins, antibodies against the respective proteins have been produced. Antibodies against the *Hydra* proteins open up new possibilities in studying the intracellular protein localization and protein interactions. Methods like immunoprecipitation, co-immunoprecipitation and chromatin immunoprecipitation could be applied.
Antibodies against the core components of the *Hydra* PRC1 and PRC2 complexes have been produced by Christoph Herbel, Anne Kuhn and Volker Lauschke (Holstein lab). In the course of this study, an antibody against the DNA binding protein *Hy*YY1 has been produced. Polyclonal antibodies against the full length protein and against the first 144 amino acids of *Hy*YY1 have been produced. The truncated protein is lacking the C-terminal zinc finger region and was used as antigen to avoid cross-reactivity of the antibody with other zinc finger proteins. When performing a tBLASTn search against the *Hydra* genome, similar sequences could only be found in the *Hy*YY1 zinc finger region, but not for the N-terminal region.

4.3.1. Protein antigen expression

All immunizations and boosts have been performed with bacterially expressed protein antigens. Both the full length coding region and the truncated version of the *Hy*YY1 protein have been cloned into the pET15b(+)-vector and overexpressed in bacteria. The proteins have been expressed as His-tagged fusion proteins and purified by metal ion affinity chromatography (for details see materials and methods section 3.6). To determine the best conditions for protein expression, the proteins were expressed in different bacterial strains at different temperatures and protein expression was examined at various times after induction with IPTG. The time course of protein expression has been monitored by SDS-PAGE. The results are depicted in figure 4.22.
The full length *Hy*YY1 protein was expressed in two different bacterial strains: BL21DE3 and Rosettagami (DE3) pLysS. The Rosettagami-strain allows disulfide bond formation and the enhanced expression of eukaryotic proteins that contain rare codons. *Hy*YY1 is expressed sufficiently in both bacterial strains at a temperature of 37°C (figure 4.22 A). The protein migrates at ∼60 kDa in a 12% SDS-gel. This differs from the theoretical molecular weight of the protein, which is calculated to be 43 kDa. However, it is known from recombinantly expressed human YY1 protein that the experimentally observed molecular weight differs from the calculated molecular weight (Shi et al., 1997). The solubility of the recombinant protein varies. Whereas most of the protein can be found within the soluble fraction for BL21DE3, most of the protein is aggregated in Rosettagami (DE3) pLysS.

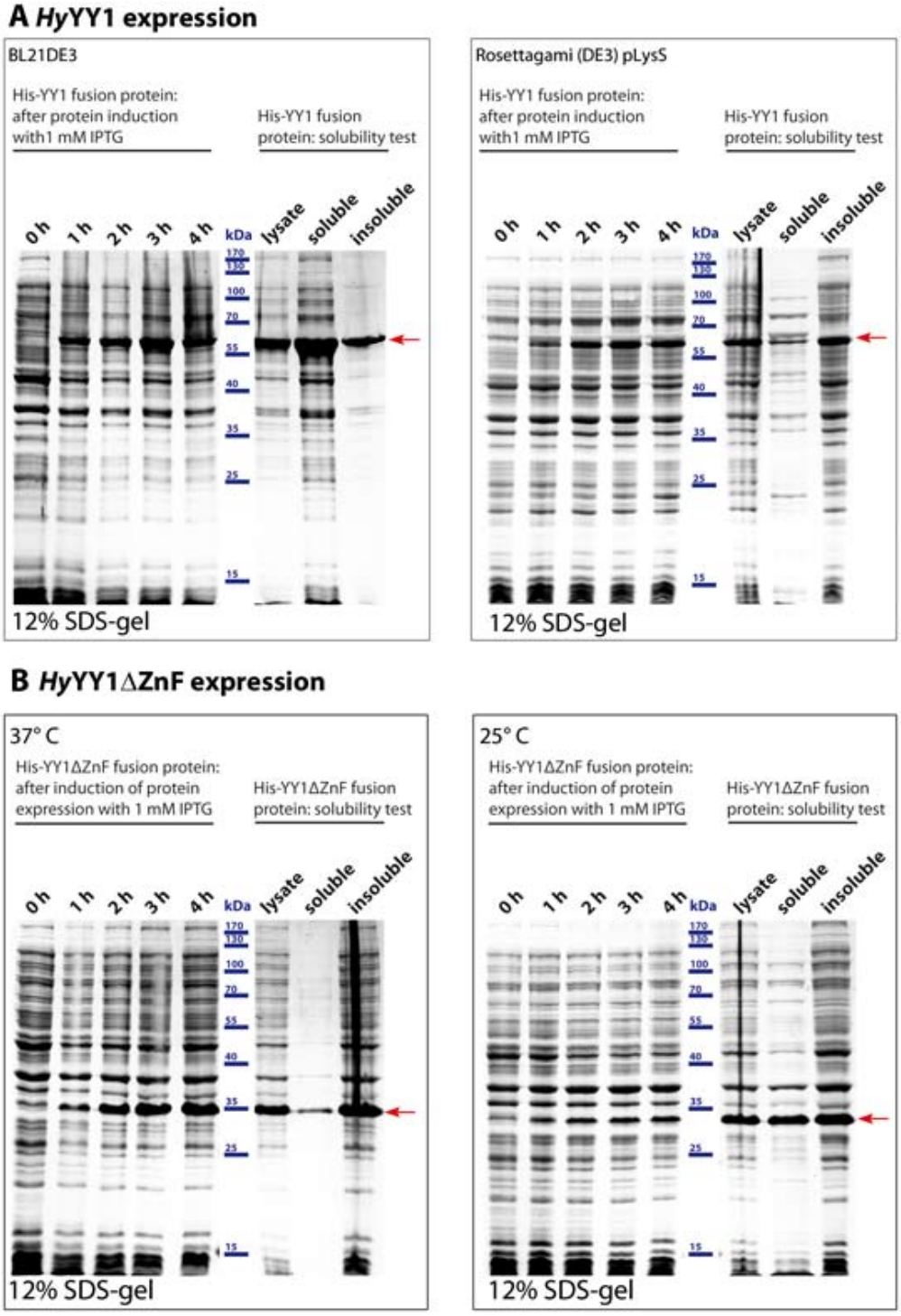


FIGURE 4.22 *SDS-gels showing the expression of HyYY1 and of HyYY1ΔZnF.* ***A*** *depicts a time course of HyYY1 protein expression in two different bacterial strains (BL21DE3 and Rosettagami (DE3) pLysS) at 37° C.* ***B*** *depicts a time course of HyYY1ΔZnF protein expression at different temperatures (25° C and 37° C) in BL21DE3. To compare the amount of expressed protein, an equal number of bacterial cells has been loaded onto the gels for each time point. To test whether the proteins are soluble, samples of the different fractions after sonication have been loaded onto the gels (whole lysate, soluble/supernatant, insoluble/pellet). Red arrow marks the recombinant fusion proteins.*

The *Hy*YY1ΔZnF protein was expressed in the bacterial strain BL21DE3 at different temperatures (figure 4.22 B). The solubility of the protein is greatly enhanced at 25°C compared to 37°C. Like the full length protein, the truncated protein migrates at a higher molecular weight during SDS-PAGE. The protein migrates at ∼35 kDa instead of the calculated 18,8 kDa.

In addition, immunoblots were performed to confirm that the protein bands on the SDS-gels are indeed the *Hy*YY1 fusion proteins and no artefacts. A time course of recombinant protein expression was performed and monitored by immunoblotting. As a negative control, protein expression was induced from the empty pET15b(+) vector and a time course was performed as well. The samples were probed with a specific antibody recognizing the 6His-tag of the fusion proteins (α-pentaHis antibody). Figure 4.23 shows the results of the immunoblots. The immunoblots confirm the identity of the fusion proteins and are in accordance with the SDS-gels. For the full length protein a specific band at ∼60 kDa is visible and several bands smaller than ∼60 kDa, which probably correspond to degradation products or not fully translated proteins (figure 4.23 A). For the *Hy*YY1ΔZnF protein a band of ∼35 kDa is visible (figure 4.23 B). For both proteins, a baseline expression can already be detected in BL21DE3 cells that have not been induced with IPTG. No proteins can be detected in the negative control samples.

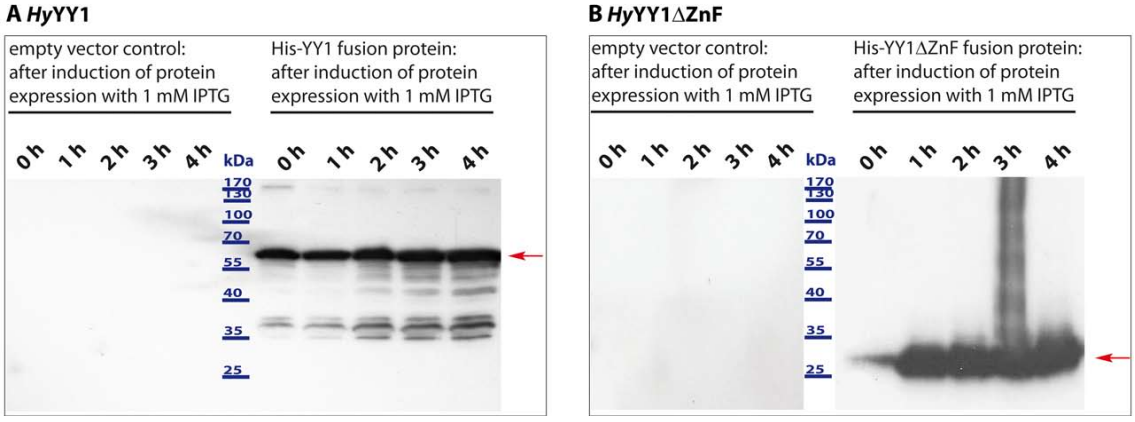


FIGURE 4.23 *Immunoblots showing the expression of HyYY1 and of HyYY1ΔZnF.* ***A*** *depicts a time course of HyYY1 protein expression in BL21DE3 at 37°C.* ***B*** *depicts a time course of HyYY1ΔZnF protein expression at 25°C in BL21DE3. To compare the amount of expressed protein, an equal number of bacterial cells has been loaded onto the gels for each time point. Red arrow marks the recombinant fusion proteins.*

For immunization both *Hy*YY1 and of *Hy*YY1ΔZnF were purified. After protein expression, the bacterial cells were lysed by sonication. The soluble fraction was separated from the insoluble fraction by centrifugation. The full length protein was purified as native protein from the soluble fraction by metal ion affinity chromatography and eluted from the matrix using increasing amounts of the histidine analog imidazole. An adequate

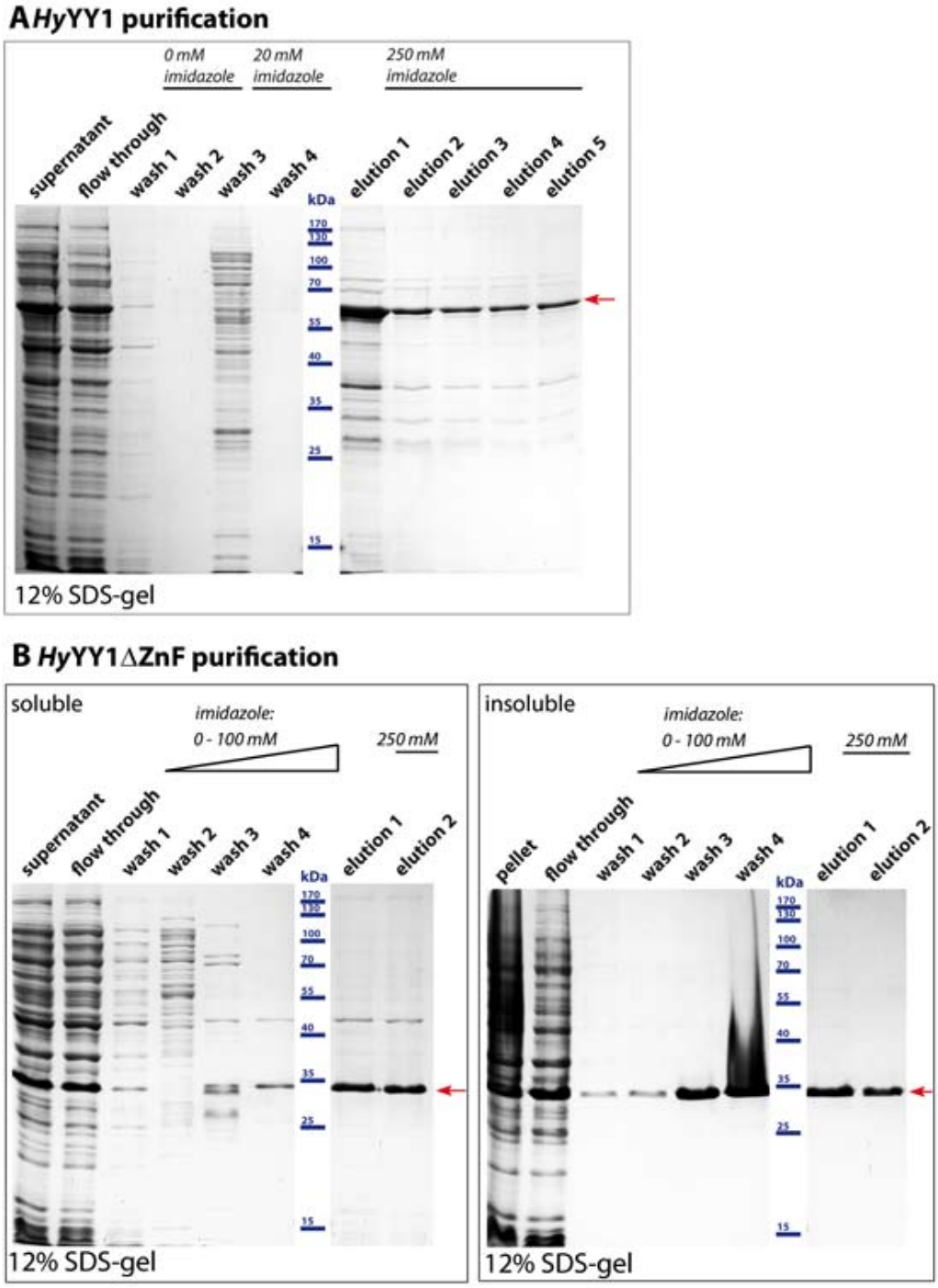


FIGURE 4.24 *Purification of HyYY1 and HyYY1ΔZnF monitored by SDS-PAGE.* ***A*** *depicts the purification of HyYY1 under native conditions from the soluble fraction. Samples from the different steps during protein purification have been loaded onto the gel.* ***B*** *depicts the purification of HyYY1ΔZnF under native and denaturing conditions. Samples from the different steps during protein purification have been loaded onto the gel. Red arrow marks the recombinant fusion proteins.*

amount of purified protein could be obtained for the immunization procedure. (figure 4.24 A). The truncated protein was purified both from the soluble and the insoluble fraction by metal ion affinity chromatography using increasing amounts of imidazole. A sufficiently purified protein was obtained with both purification methods (figure 4.24 B). Under denaturing conditions the *Hy*YY1ΔZnF protein seems to bind to the Ni-NTA beads with a lower affinity and is already eluted during the wash steps with low imidazole concentrations. By comparison, most of the native protein is not eluted from the beads until an imidazole concentration of 250 mM is reached. The amount of purified protein varies. A greater amount of the recombinant protein could be purified from the insoluble fraction under denaturing conditions.

4.3.2. Antibody production and characterization

Both the purified full length *Hy*YY1 protein and *Hy*YY1ΔZnF have been used as antigens for the production of polyclonal antibodies in rabbits. 100 μg of protein in Freund's adjuvant have been used for the initial immunization and each boost. Figure 4.25 shows the route of injections.

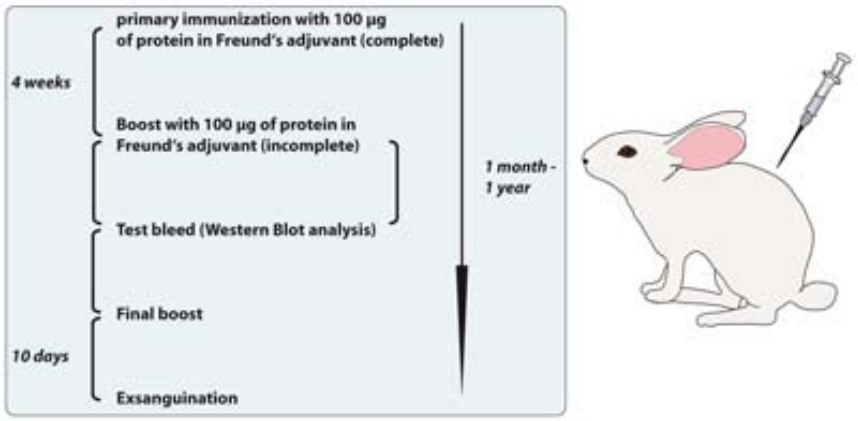


FIGURE 4.25 *Antibody production. Route of injections.*

During this study, two rabbits have been immunized with *Hy*YY1ΔZnF and one rabbit has been immunized with the full length protein. Table 4.4 gives an overview of the produced α-*Hy*YY1 antibodies. The antigen preparation used for the injections and the number of injections including the primary immunization are listed.

TABLE 4.4 *Antigens used for immunization.*

antibody	antigen	rabbit	number of injections
α-***Hy*YY1**	*Hy*YY1ΔZnF (pET15b(+)); native	# 1	4
α-***Hy*YY1 2**	*Hy*YY1ΔZnF (pET15b(+)); denatured	# 2	3
α-***Hy*YY1 3**	*Hy*YY1 (pET15b(+)); native	# 3	5

To monitor the progress of antibody production, samples of blood have been taken at regular intervals and analyzed by immunoblots. The blood samples have been used to probe different amounts of recombinantly expressed *Hy*YY1 protein and *Hydra* protein lysates. In figure 4.26 the analysis of the α-*Hy*YY1 preimmune serum and the first test serum taken four weeks after the initial immunization of rabbit # 1 is depicted exemplarily. It is well known that rabbits may contain cross-reactive antibodies in their preimmune serum, making it necessary to test the serum for reactivity as well. The sera were used to probe increasing amounts of the antigen. Whereas the recombinant protein is not detected by the preimmune serum (figure 4.26 A), the protein is detected when probed with the first test bleed (figure 4.26 B). Both the preimmune serum and the first α-*Hy*YY1 test serum cross-react with proteins from bacterial lysate.

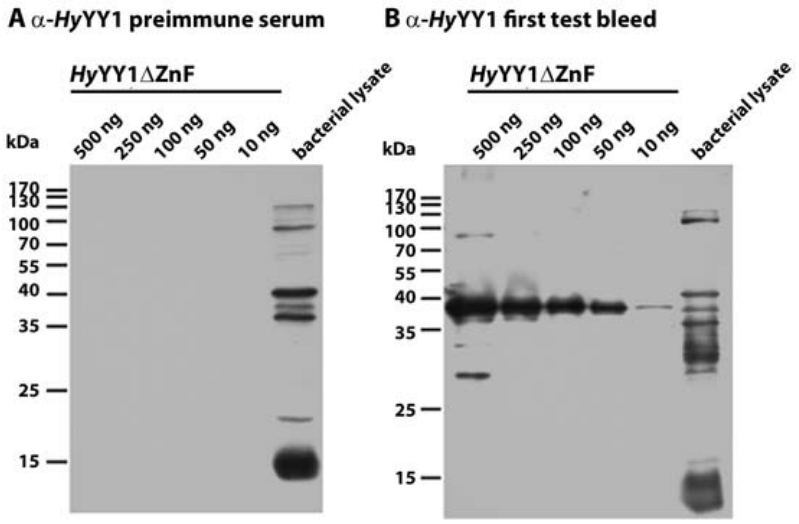


FIGURE 4.26 *Immunoblot. Different amounts of recombinant HyYY1ΔZnF protein and bacterial lysate were probed with the α-HyYY1 preimmune serum (**A**) and with the first α-HyYY1 test bleed after the initial immunization (**B**).*

Since all generated antibodies specifically recognize recombinant *Hy*YY1 protein, they were used to localize the endogenous protein. Therefore, all sera were tested on *Hydra* extracts. The sera were tested on nuclear extracts, cytoplasmatic extracts, whole cell extracts and *Hydra* polyps lysed in protein sample buffer. Simultaneously, the respective preimmune sera were tested on the *Hydra* extracts. The immunoblots in figure 4.27 give an overview of the analysis of the produced α-*Hy*YY1 sera in comparison to the respective preimmune sera. The results for the α-*Hy*YY1 antibody produced in rabbit # 1, # 2 and # 3 are depicted.

The α-*Hy*YY1 final serum produced in rabbit # 1 recognizes multiple bands in *Hydra* extracts. The most prominent bands in the nuclear, cytoplasmatic and whole cell extract migrate between $\sim$25 and 35 kDa and at $\sim$40 kDa. Additionally, a faint band can be detected at $\sim$70 kDa in the nuclear extract. When probing lysed *Hydra* polyps, proteins between $\sim$25 and 35 kDa, at $\sim$40, 55 and 70 kDa can be detected. The distribution of the

detected bands varies from the *Hydra* extracts to the freshly prepared *Hydra* lysate. The

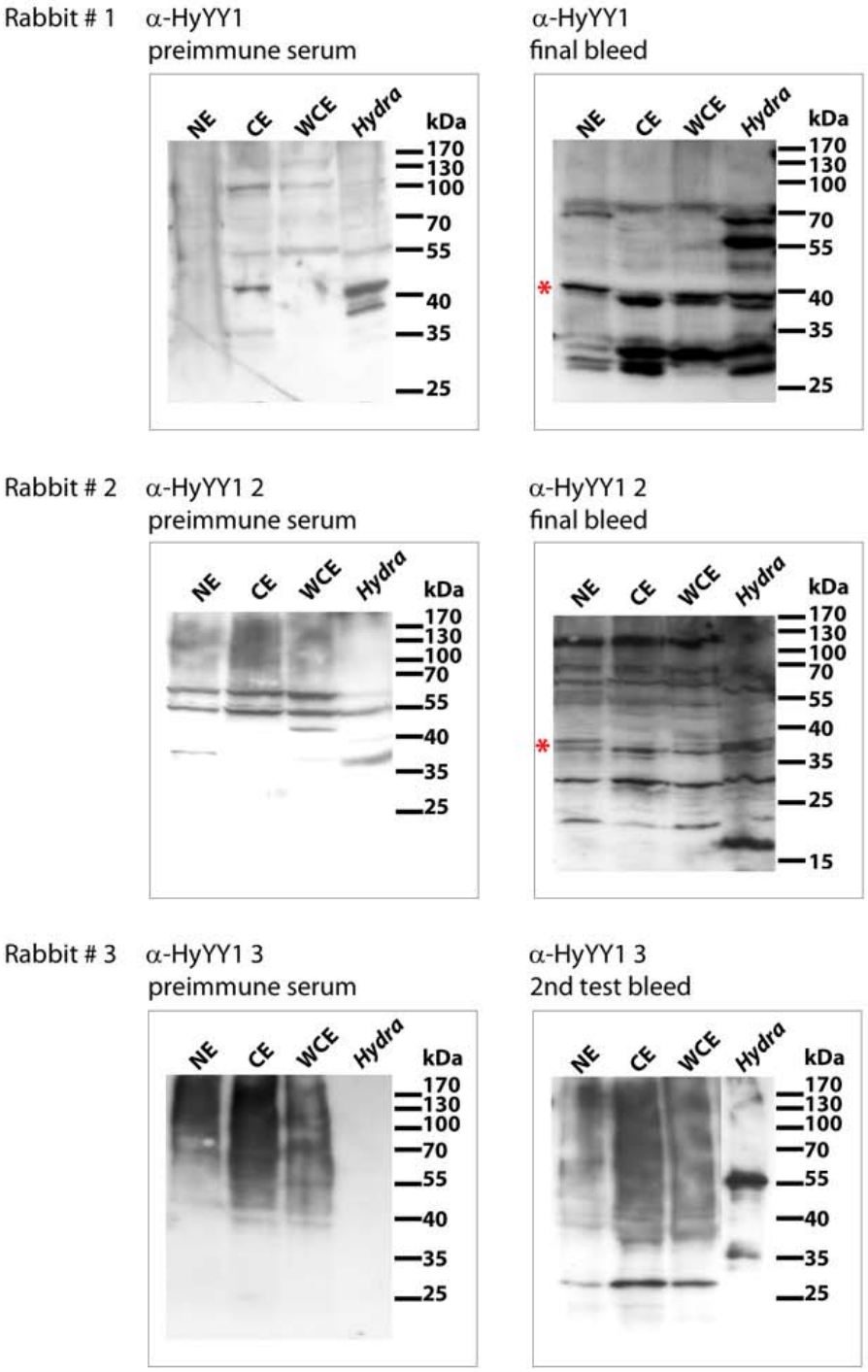


FIGURE 4.27 *Immunoblots. Hydra extracts were probed with different α-HyYY1 preimmune and test sera. The respective sera are indicated. All sera were diluted 1:1000 for immunoblotting.*
NE = nuclear extract (100 μg); CE = cytoplasmatic extract (100 μg); WCE = whole cell extract (100 μg); Hydra = 10 whole polyps lysed in protein sample buffer.

higher molecular weight bands at 55 and 70 kDa are not detected in the extracts. The full length protein may be subjected to degradation in the cause of the extraction procedure (for details see materials and methods section 3.6.4). Unfortunately, the preimmune serum cross-reacts with several proteins in the *Hydra* extracts. The preimmune serum detects bands at ∼40, 55 and 100 kDa. This background has to be considered for all further applications of the antibody. When comparing the final serum with the preimmune serum, the bands detected at ∼40 kDa (red asterisk) with the final serum are probably not specific and do not correspond to the *Hy*YY1 protein.
As for the first α-*Hy*YY1 antibody, the preimmune serum of the antibody produced in rabbit # 2 also cross-reacts with some *Hydra* proteins. The preimmune serum detects various bands between 35 and 70 kDa. Because of the increasing background, the immunization of rabbit # 2 was stopped after the second boost. The final serum recognizes bands migrating between 15 and 35 kDa and a high molecular weight band migrating between 100 and 130 kDa. These bands are not detected with the preimmune serum. Bands that are recognized by the preimmune serum are marked (red asterisk). The serum was not used for further experiments.
For the α-*Hy*YY1 antibody produced in rabbit # 3, an immunoblot with the second test serum retrieved two weeks after the first boost is depicted. A specific band between 25 and 35 kDa is recognized in all *Hydra* extracts. In lysed *Hydra* polyps two specific bands at 35 and 55 kDa can be detected. The preimmune serum detects no specific bands in freshly prepared *Hydra* lysate. However, there is a high background (smear) when probing *Hydra* extracts. Rabbit # 3 was boosted further with the full length *Hy*YY1 protein. The serum produced in rabbit # 3 is promising. However, the α-*Hy*YY1 3 antibody could not be characterized further in the scope of this studies because of time constrictions.
All tested sera detect multiple bands in different *Hydra* extracts and produce a high background. No definite statement on the intracellular localization of *Hy*YY1 can be made.

For all further experiments, the α-*Hy*YY1 antibody produced in rabbit # 1 has been used. The antibody recognizes specific bands between 25 and 35 kDa and at 55 and 70 kDa in immunoblots with *Hydra* extracts. From the overexpression of the *Hy*YY1 protein in bacteria, a protein of ∼60 kDa is expected for the full length protein. The band detected at 55 kDa in immunoblotting probably corresponds to the full length protein. The other bands may correspond to post-translationally modified protein (25-35 kDa and 70 kDa). The human YY1 protein is known to be post-translationally modified by phosphorylation, acetylation and caspase-dependent cleavage (He and Casaccia-Bonnefil, 2008).

4.3.3. Antibody purification

Since the crude serum produces a high background and recognizes multiple probably unspecific bands when probing *Hydra* extracts, the α-*Hy*YY1-antibody was affinity purified using an antigen column. The antigen was recombinantly expressed and covalently coupled to a solid support (AffiGel 10). Coupling was monitored by SDS-PAGE and the prepared matrix was used for the affinity purification of the α-*Hy*YY1 antibody (for details see materials and methods section 3.7.2).

For the purification of the antibody, two different antigen preperations were used. The antigen was either expressed in the pET15b(+)-vector or in the pET21(+)-vector. The antigen expressed in the pET21(+)-vector has no additional vector sequence at its N-terminus. Since several of the *Hydra* PcG proteins were expressed in the pET15b(+)-vector with the same N-terminal vector sequence (MGSSHHHHHHSSGLVPRGSHML), affinity purification of the antibody against an antigen expressed in a different vector can prevent cross-reactivity of the antibody with other recombinantly expressed *Hydra* PcG proteins.

The α-*Hy*YY1-antibody was purified either against the antigen expressed in the pET15b(+)-vector or in the pET21(+)-vector. The final serum was passed through the antigen column and unspecific antibodies were washed away. Cycles of low pH and high pH were used for elution. Antibodies bound by acid-sensitive interactions were eluted at pH 2.5, antibodies bound by base-sensitive interactions were eluted at pH 11.5. For each elution cycle, 20 fractions of 0.5 ml were collected and the pH was neutralized. The activity of the antibody in the different fractions was tested by immunoblotting. Each fraction was used to probe 20 ng of the antigen *Hy*YY1ΔZnF. For this purpose, the antigen was transfered to a PVDF membrane. The membrane was cut in stripes, and each stripe was used for testing one of the antibody fractions. For detection, the stripes were put together again. Exemplarily, the immunoblots for the purification of the α-*Hy*YY1 antibody against the antigen expressed in the pET15b(+)-vector are depicted in figure 4.28.

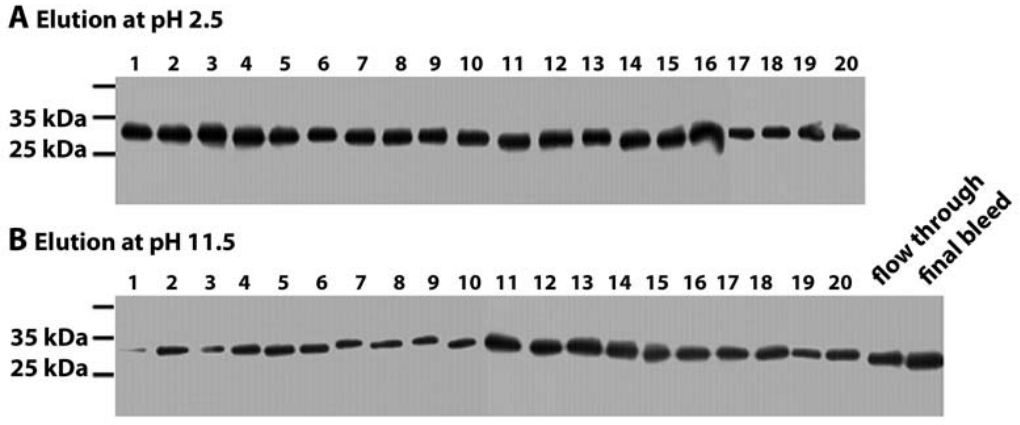


FIGURE 4.28 *Survey of the antibody affinity purification by immunoblotting. Each elution fraction was diluted 1:1000 and used to probe 20 ng of the antigen. The number of the fractions (1 - 20) of the elution cycles at low pH and high pH are indicated. In addition, the crude serum (final bleed) and the 'flow through' have been tested.*

The antibody is active in all fractions eluted from the antigen column at both low pH (figure 4.28 A) and high pH (figure 4.28 B). The activity is comparable with the the crude serum (final bleed). Additionally, the 'flow through'-fraction after binding of the antibody to the column has been tested. This fraction still shows antibody activity, which indicates that not all of the specific antibody could bind to the antigen column.
Positive fractions were combined and the purified antibody was concentrated. Table 4.5 gives an overview of the antibody concentrations after purification.

TABLE 4.5 *Affinity purified α-HyYY1 antibodies.*

antibody	antigen used for affinity purification	concentration
α-*Hy*YY1	*Hy*YY1ΔZnF (pET15b(+))	0.78 mg/ml
α-*Hy*YY1	*Hy*YY1ΔZnF (pET21(+))	0.98 mg/ml

Additionally, the purified antibodies were analyzed by SDS-PAGE in comparison to the crude serum (figure 4.29 A). Whereas the serum albumin band at 70 kDa is the predominant band in the unpurified serum, the strongest band in the affinity purified sera corresponds to the antibody heavy chain at 55 kDa. However, a small band at 70 kDa is still visible in the samples of the affinity purified antibodies.

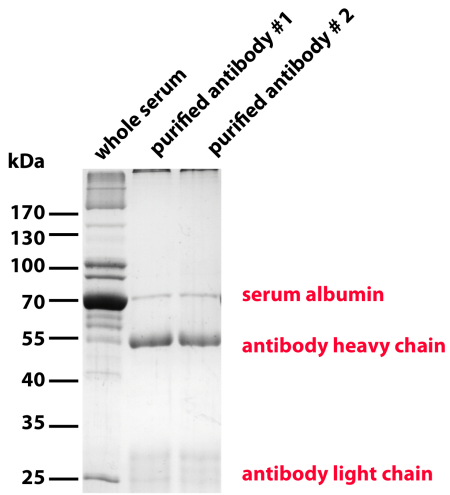


FIGURE 4.29 *Analysis of the affinity purified α-HyYY1 antibodies on a 12% SDS-PAGE gel. The antibody preparation and the respective molecular weights are indicated. Purified antibody # 1 = serum purified against HyYY1ΔZnF expressed in pET15b(+); purified antibody # 2 = serum purified against HyYY1ΔZnF expressed in pET21(+).*

The affinity purified sera were used to probe *Hydra* extracts. Immunoblots with the purified antibodies are depicted in figure 4.30. Figure 4.30 A shows an immunoblot with the unpurified serum, figures 4.30 B and C show immunoblots with the affinity purified sera. Both preparations of the affinity purified antibodies still recognize multiple bands in *Hy-*

dra extracts. No obvious difference can be detected in the distribution of the bands when probing the extracts with the purified sera in comparison to the unpurified final serum. The specificity of the antibody did not increase after purification. There are still several bands visible, which migrate between 25 and 35 kDa, two bands migrating at ~40 kDa and two higher molecular weight bands migrating between 50 and 70 kDa when probing whole *Hydra* animals. The higher molecular weight bands are not detected in the nuclear, cytoplasmatic and whole cell extracts. Whether all bands visible correspond to *Hy*YY1 or whether the antibody cross-reacts with other *Hydra* proteins could not be determined.

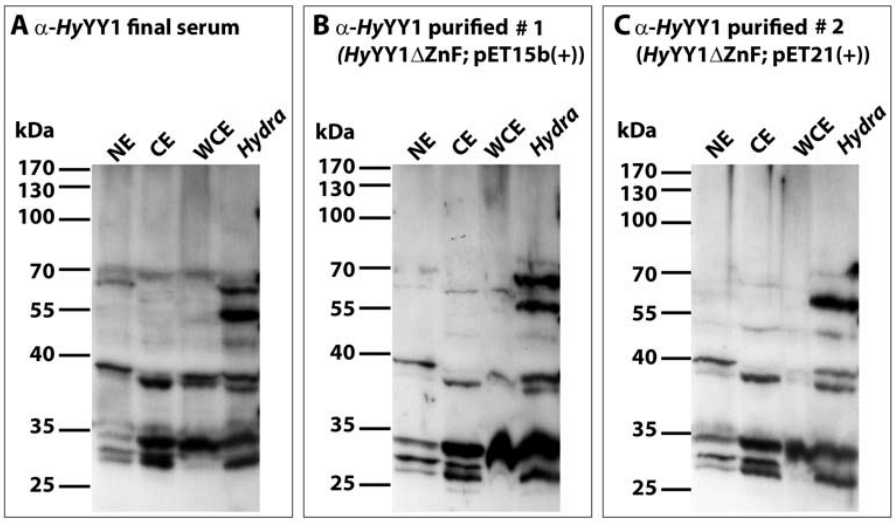


FIGURE 4.30 *Analysis of the affinity purified α-HyYY1 antibodies by immunoblotting. Hydra extracts were probed with the unpurified final serum (**A**), the serum purified against HyYY1ΔZnF expressed in pET15b(+) (**B**) and the serum purified against HyYY1ΔZnF expressed in pET21(+) (**C**). All sera were diluted 1:1000. NE = nuclear extract (100 μg); CE = cytoplasmatic extract (100 μg); WCE = whole cell extract (100 μg); Hydra = 10 whole polyps lysed in protein sample buffer.*

4.3.4. Localization of *Hydra* PcG proteins by immunostaining

The *Hy*YY1 protein was localized in *Hydra* by indirect immunofluorescence. Whole *Hydra* polyps were stained with the α-*Hy*YY1 final serum. The polyps were fixed with Lavdovsky fixative, a mixture of ethanol, formaldehyde and acetic acid. A short fixation time of 15 minutes proved best for the α-*Hy*YY1 antibody. Staining with the primary antibody was performed over night at 4°C, followed by the incubation with a fluorescent labeled secondary antibody (for details see materials and methods section 3.7.4). Nuclei were co-stained with DAPI. The staining pattern was detected by fluorescence microscopy.
In immunostainings with the α-*Hy*YY1 final serum, mainly the ectodermal epithelial cells are stained (figure 4.31). Staining can be observed in the membranes of epithelial cells and in the nuclei. In figure 4.31 E and E', some nuclei of ectodermal cells are marked in the α-*Hy*YY1 staining and the corresponding DAPI staining. Only the big nuclei of ectodermal epithelial cells are stained, the smaller nuclei interstitial cell nests are devoid of the antibody signal. In contrast to the *HyYY1 in situ hybridization*, the whole polyp is

stained. Whereas the *in situ hybridization* showed no *HyYY1* expression in the foot and the head region, the staining signal for the antibody can also be localized in these regions. As a control for the the α-*Hy*YY1 staining, polyps were stained with the *Hy*YY1 preimmune serum and the secondary antibody. Neither with the preimmune serum nor with the secondary antibody alone, a specific signal could be observed.

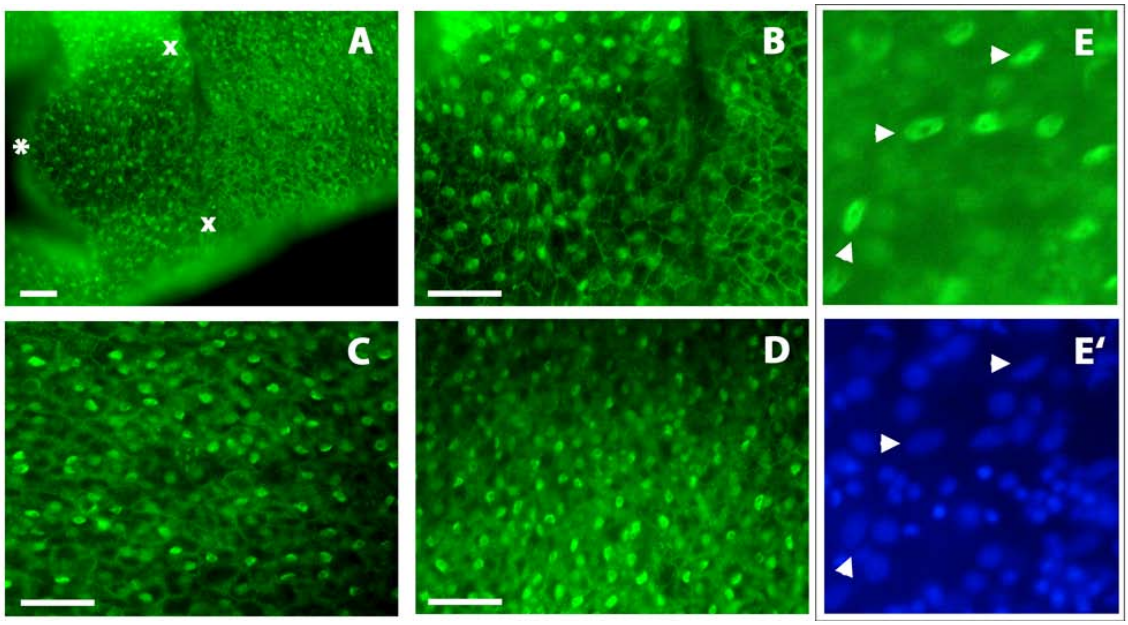


FIGURE 4.31 *Immunostaining. Whole Hydra polyps were co-stained with α-HyYY1 (green) and DAPI (blue). The head region of a polyp with hypostome (*) and tentacle base (x) is depicted in **A**, a close-up of the hypostome region is shown in **B**. In **C** and **D** a region of the body column of Hydra is depicted. Some nuclei of ectodermal cells are marked by a white arrowhead in the α-HyYY1 staining (**E**) and the corresponding DAPI staining (**E'**).*
Scale bars: 50 μm.

Immunostainings were also performed with the purified α-*Hy*YY1 serum. The immunofluorescent staining was detected by means of confocal microscopy (Nikon A1 confocal laser microscope system). The result is depicted in figure 4.32. At the top left side, an overview of the *Hy*YY1 staining pattern in a developing bud is depicted. The figure at the top right side gives an overview of the body regions of *Hydra* and serves for orientation. In figures 4.32 A - E representative immunofluorescence images of different sections of the *Hydra* polyp are shown at higher magnification. The staining pattern of the purified antibody does not differ from the pattern obtained with the crude final serum. Mainly the nuclei and the membranes of ectodermal epithelial cells are stained. In figure 4.32 A the nuclei are clearly visible, in 4.32 C a net-like structure corresponding to the membranes of the epithelial cells can be observerd. Additionally, some cells of the interstitial lineage are stained (figure 4.32 D, asterisks).

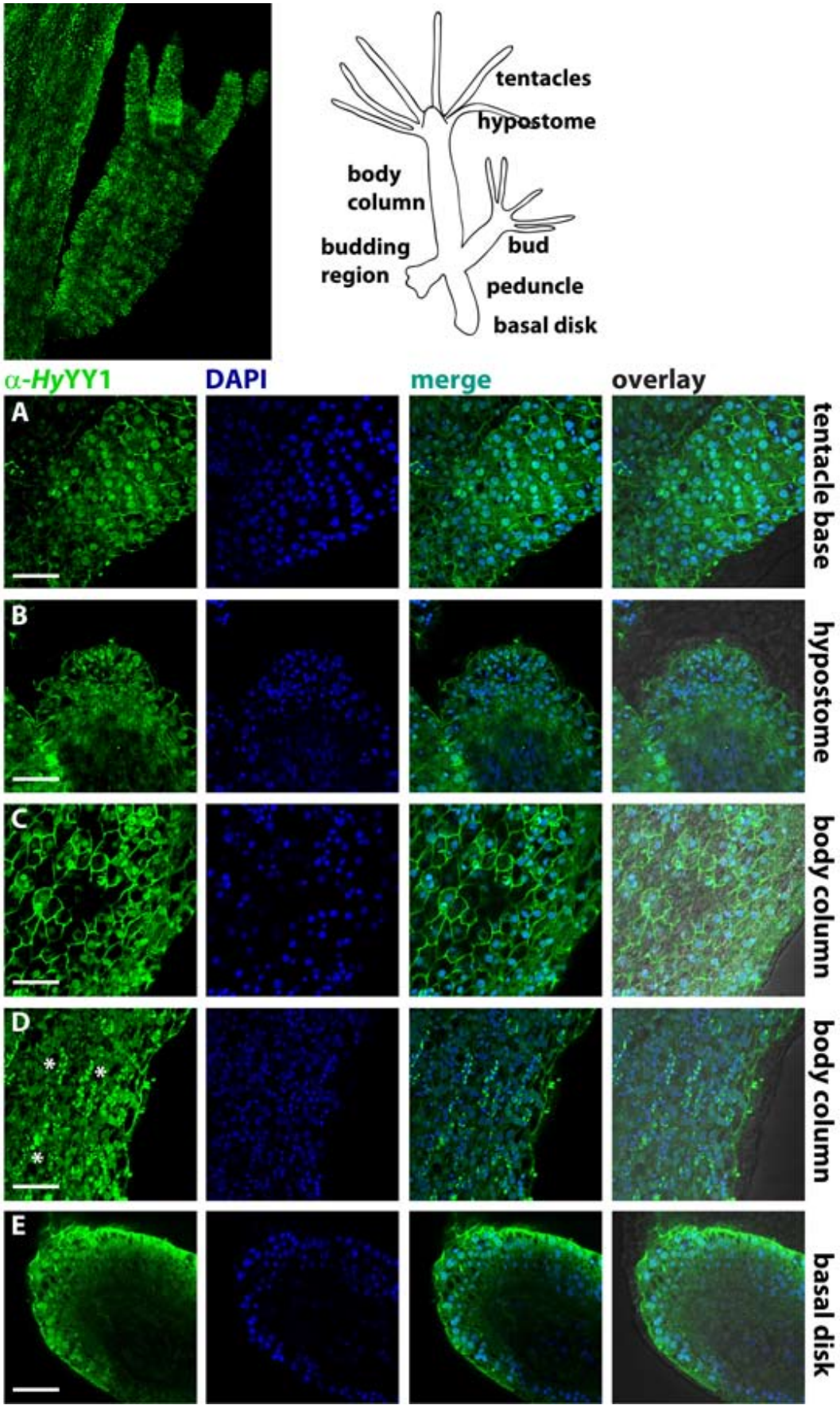


FIGURE 4.32 *Immunostaining. Whole Hydra polyps were co-stained with α-HyYY1 (green) and DAPI (blue). The two columns in the right show merged images and merged images overlaid with transmitted light images. Scale bars: 50 μm.*

Since antibodies against proteins of the *Hydra* PRC2 complex are available, indirect immunofluorescence stainings were performed with the α-*Hy*EZH2, α-*Hy*EED and α-*Hy*p55 antibodies. An antibody against *Hy*SUZ12 is not available. As for the α-*Hy*YY1 antibody, fixation with Lavdovsky for 15 minutes proofed best. The polyps were stained with the respective primary antibody over night and incubated with the fluorescent labeled secondary antibody. The nuclei were co-stained with DAPI. The staining pattern was detected by fluorescence microscopy using a confocal microscope. In addition to the application in immunostainings, the antibodies were used for immunoblotting to test their specificity.

The result for the α-*Hy*EZH2 antibody is depicted in figure 4.33. The antibody was used to probe *Hydra* lysate to test the specificity of the α-*Hy*EZH2 serum (figure 4.33 A). The antibody detects a double-band at $\sim$80 kDa and to a lesser extent several bands of lower molecular weights. The calculated molecular weight of *Hy*EZH2 is 83.6 kDa and it can be concluded that the band at $\sim$80 kDa corresponds to *Hy*EZH2. It is likely that the serum cross-reacts with other *Hydra* proteins. This fact has to be taken into account for all further applications. A characteristic result of the immunostaining is depicted in figure 4.33 B. As for the α-*Hy*YY1 antibody, the α-*Hy*EZH2 antibody mainly stains the nuclei of ectodermal epithelial cells, which can be clearly localized by co-staining with DAPI (figure 4.33 B). Furthermore, precipitates that are intensely stained can be observed all over the sample. These precipitates look different from the surrounding staining pattern and are probably artifacts. Staining can be observed in the entire polyp. The staining pattern obtained with the α-*Hy*EZH2 antibody is contradictory to the expression pattern for *HyEZH2* obtained by *in situ hybridization*. Whereas the *HyEZH2* mRNA is localized in interstitial cells (compare section 4.2.1), the *Hy*EZH2 protein could not be detected in the interstitial lineage.

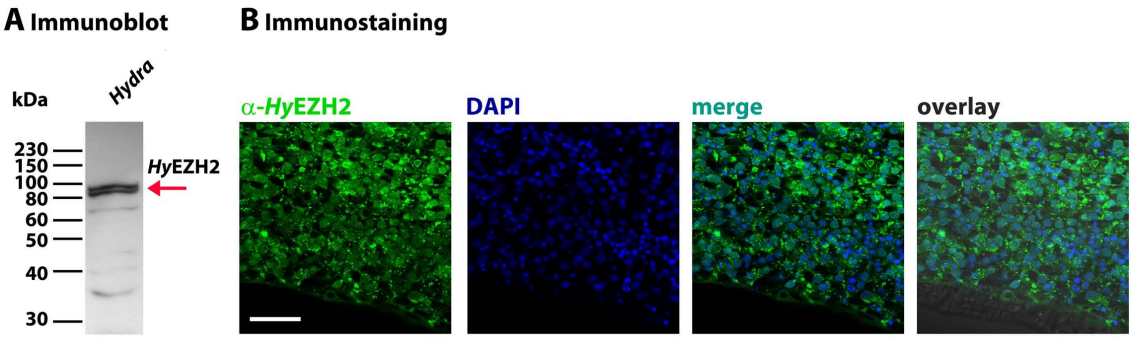


FIGURE 4.33 ***A*** *Immunoblot. Hydra lysate was probed with α-HyEZH2 final bleed.* ***B*** *Immunostaining. Whole Hydra polyps were co-stained with α-HyEZH2 (green) and DAPI (blue). A section of the Hydra body column is depicted. The two pictures in the right show a merged image and a merged image overlaid with the transmitted light image. Scale bar: 50 μm.*

Figure 4.34 A shows an immunoblot and figure 4.34 B shows a characteristic immunostaining result with the α-*Hy*EED antibody. For the immunoblot, *Hydra* lysate was probed with the α-*Hy*EED antibody. The antibody recognizes one distinct band at $\sim$70 kDa,

which migrates at a higher molecular weight than expected for the *Hy*EED protein. The calculated molecular weight of *Hy*EED is 47.9 kDa. A typical result for the immunostaining is depicted in figure 4.34 B. The antibody mainly stains the nuclei of ectodermal epithelial cells and cell membranes (figure 4.34 B). The whole polyp is stained, including head and foot region. As for *Hy*EZH2, the staining pattern is contrary to the result of the *in situ hybridization*. The *HyEED* mRNA is localized in interstitial cells (compare section 4.2.1), whereas the *Hy*EED protein is only detected in ectodermal epithelial cells.

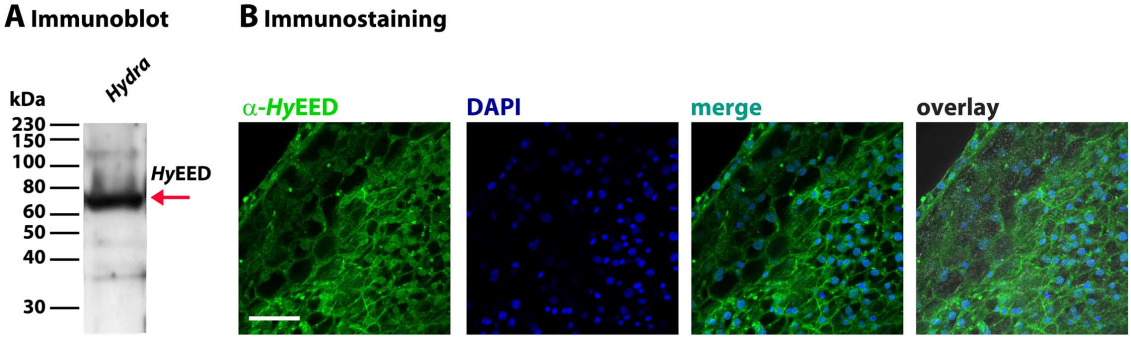


FIGURE 4.34 ***A*** *Immunoblot. Hydra lysate was probed with α-HyEED final bleed.* ***B*** *Immunostaining. Whole Hydra polyps were co-stained with α-HyEED (green) and DAPI (blue). A section of the Hydra body column is depicted. The two pictures in the right show a merged image and a merged image overlaid with the transmitted light image. Scale bar: 50 μm.*

The result for the α-*Hy*p55 antibody is depicted in figure 4.35. In immunoblotting, the α-*Hy*p55 serum detects a band at ∼55 kDa in *Hydra* lysate (4.35 A), which probably corresponds to the *Hy*p55 protein. The calculated molecular weight of *Hy*p55 is 49.1 kDa. Figure 4.35 B shows a typical result for the immunostaining of *Hydra* with the α-*Hy*p55 antibody. Staining can be observed in the nuclei of ectodermal epithelial cells and in nuclei of nematoblasts, which belong to the I cell lineage. However, the intensity of the staining is stronger in the nuclei of the nematoblasts (4.35 B, asterisk). Nuclei all over the polyp are stained with the α-*Hy*p55 antibody. Compared to the *Hyp55* mRNA, which is solely localized in the interstitial cell lineage (see section 4.2.1), the Hyp55 antibody can be additionally localized in ectodermal epithelial cells.

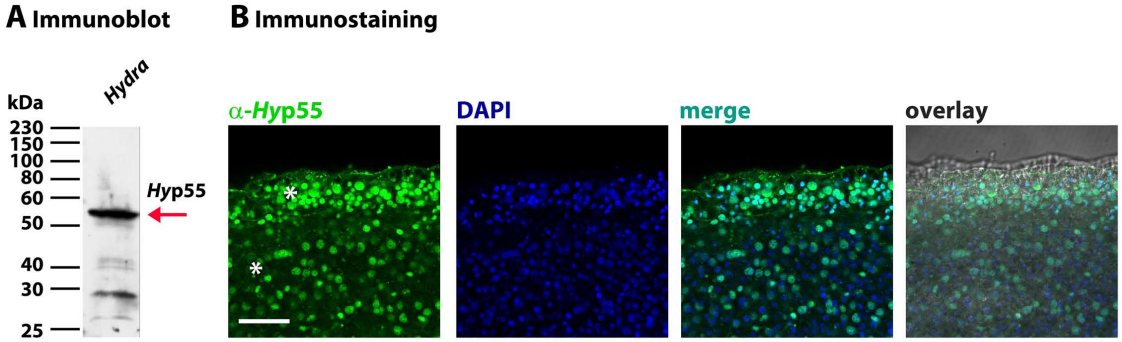


FIGURE 4.35 ***A*** *Immunoblot. Hydra lysate was probed with α-Hyp55 final bleed.* ***B*** *Immunostaining. Whole Hydra polyps were co-stained with α-Hyp55 (green) and DAPI (blue). A section of the Hydra body column is depicted. The two pictures in the right show a merged image and a merged image overlaid with the transmitted light image. Scale bar: 50 μm.*

4.3.5. Localization of the H3K27me3 epigenetic mark

The PRC2 complex catalyzes the trimethylation of histone H3 at lysine 27. To get an idea of the existence and the cellular localization of the epigenetic mark in *Hydra*, immunostainings and Western blots were performed with an antibody against H3K27me3.
Histone proteins are highly conserved and a putative *Hydra* histone H3 could be found in the genome (see appendix B.2). The N-terminal tail is 100% identical to the human histone, including the site of H3K27 trimethylation (figure 4.36 C). An antibody against human H3K27me3 was used to stain whole *Hydra* polyps and to probe *Hydra* lysate. For immunostaining, polyps were co-stained with DAPI and an antibody specifically recognizing H3K27me3. The antibody against the epigenetic mark set by the PRC2 complex is exclusively localized to the nuclei of the interstitial cell lineage (figure 4.36 A). With the help of Western blot analysis, it could be shown that the used antibody is specific for trimethylated H3K27. The antibody could not detect unmethylated H3K27 when probing unmodified recombinant histone octamers (figure 4.36 B). In *Hydra* lysates, a clear signal could be detected.

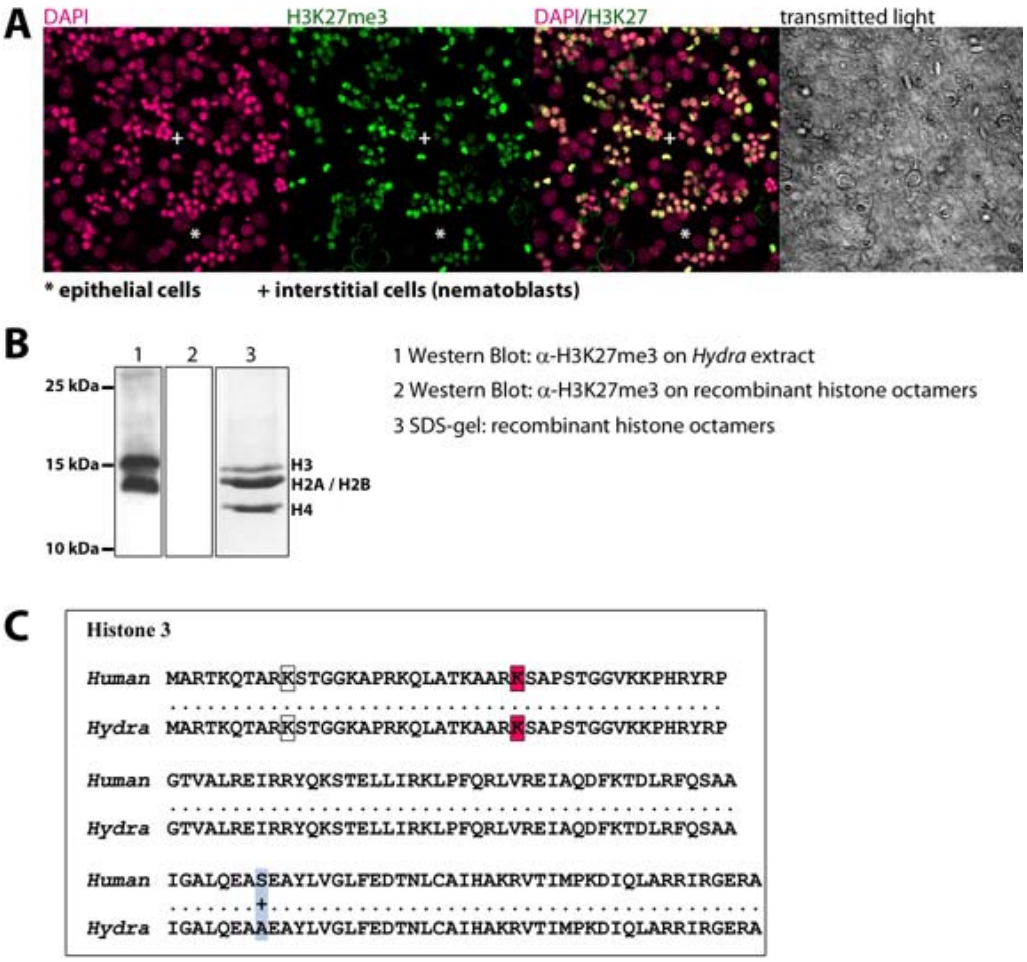


FIGURE 4.36 ***A*** *Immunostaining. Whole Hydra polyps were co-stained with α-H3K27me3 (green) and DAPI (magenta). A section of the Hydra body column is depicted. The two pictures in the right show a merged image and a transmitted light image.* ***B*** *Immunoblot. Hydra lysate and recombinant histone octamers were probed with α-H3K27me3. In lane 3 the recombinant octamer is separated by SDS-PAGE. The same amount of protein has been used for the immunoblotting.* ***C*** *Alignment of the amino acid sequences of the N-terminal tail of human histone H3 and a putative Hydra histone H3.*

4.3.6. SUMMARY

The intracellular localization of *Hy*YY1

Immunostainings suggest that the *Hy*YY1 protein is mainly localized in ectodermal epithelial cells. Furthermore, the stainings show that the generated α-*Hy*YY1 antibody is not only localized in the cell nucleus, but a clear signal can also be detected in cell membranes. The same result was obtained by Western blot analysis. When probing isolated cellular fractions of *Hydra*, a signal could be detected in all subcellular fractions (nuclear, cytoplasmatic and whole cell fractions).
For mammalian YY1, a clear nuclear localization has been demonstrated by performing immunostainings (Rylski et al., 2008). However, another study could show that the subcellular localization of YY1 changes in the course of the cell cycle (Palko et al., 2004). YY1 has been shown to interact with several cell cycle regulators including Rb, p53 and c-Myc. Whereas YY1 is primarily localized in the cytoplasm in late mitosis, it undergoes a switch in localization at the G1/S boundary of the cell cycle. During S-phase, YY1 is localized in the nucleus. Rizkallah and Hurt (2009) could show that the switch in localization is dependent on the phosphorylation status of YY1. Phosphorylation of YY1 abolishes its DNA binding and the protein is localized in the cytoplasm.
The result for *Hy*YY1 is not consistent with the literature. The protein could be localized in the cell nucleus and mainly in cell membranes. No study could localize the YY1 protein to be associated with the cell membrane. Since the generated α-*Hy*YY1 antibody recognizes quite a number of polypeptides in *Hydra* during Western blot analysis, it is most likely that the antibodies cross-react with *Hydra* proteins during immunostainings as well. Therefore, the generated α-*Hy*YY1 antibody is not useful for the intracellular localization of the endogenous *Hy*YY1 protein. The question whether the *Hy*YY1 protein is localized in the nucleus or in the cell membrane could not be solved. Nevertheless, the antibody is suitable for the detection of the recombinant *Hy*YY1 protein.

The intracellular localization of the *Hydra* PRC2 proteins

There is a discrepancy between the gene expression pattern of the PRC2 genes obtained by *in situ hybridizations* and the localization of the proteins by means of immunostainings. Whereas the results of *in situ hybridizations* suggest that the PRC2 genes are expressed in a lineage-specific way in cells of the interstitial lineage, the results of the antibody stainings suggest that the PRC2 proteins are additionally expressed in ectodermal epithelial cells. Surprisingly, the PRC2 antibodies could not only be localized to nuclear regions, but also to cytoplasmic regions and cell membranes.
However, in several studies with mammalian cells, the Polycomb Group proteins could be clearly localized to the cell nucleus. Whereas the proteins of the PRC2 complex are distributed throughout the whole nucleus, proteins of the PRC1 complex are localized to pericentromeric heterochromatin and form distinct compartments in the nucleus called

PcG bodies (Saurin et al., 1998; Sewalt et al., 1998). These PcG bodies even remain stably associated throughout mitosis, allowing the inheritance of the PcG complex through successive cell divisions. Ren et al. (2008) could show that the PcG bodies get more dispersed during embryonic stem cell differentiation, which is explained by the existence of different PcG target genes in differentiating cells. Hernández-Muñoz et al. (2005) could show that the recruitment of the PRC1 protein BMI1 is dependent on proteins of the PRC2 complex. By knocking down EED or EZH2, BMI1 could not be recruited to the PcG bodies. In a more recent study, it was suggested that the PcG bodies allow the spatial organization of Polycomb target genes in the cell nucleus and stabilize the maintenance of epigenetic gene silencing (Bantignies et al., 2011). Even PcG target genes that are separated by large genomic distances can come in close proximity in the PcG bodies and can be co-repressed. That study shows that the PcG bodies co-localize with *Antennapedia* and the bithorax complex gene *Abd-B* in *Drosophila*.

The results obtained with the help of antibody stainings contradict the immunostainings with the α-H3K27me3 antibody and the results obtained by *in situ hybridizations*. Both the RNA expression pattern of the *Hydra* PRC2 genes and the H3K27me3 epigenetic mark are clearly localized in the interstitial cell lineage. In contrast, all available antibodies against proteins of the *Hydra* PRC2 complex can additionally be localized to cells of the ectodermal epithelial lineage. Neither the RNA expression pattern nor the H3K27me3 modification do correlate with the protein expression profiles of the *Hydra* PRC2 proteins. In addition, the literature clearly localizes the PcG proteins to the cell nucleus. However, the only antibody, which is clearly found in the cell nucleus is α-*Hy*p55. Therefore, it is questionable whether the generated antibodies are useful for the intracellular localization of the *Hydra* PRC2 proteins by immunostainings.

4.4. The PcG protein *Hy*YY1 can interact with DNA

YY1 is a GLI-Krüppel related protein and has four C2H2-zinc fingers at its C-terminus. This C-terminal region is highly conserved among species and has a high similarity with the *Drosophila* PcG gene *Pleiohomeotic* (*Pho*), which has been shown to recruit the PcG complexes to their target genes in a sequence-specific manner (Brown et al., 1998; Mohd-Sarip et al., 2002).

Zinc finger proteins have been shown to interact with DNA following some chemical and stereochemical rules for the interaction between amino acid residues and nucleotides. With regard to these rules, the YY1-DNA interaction could be predicted (Suzuki et al., 1994): three of the four zinc fingers participate in the protein-DNA interaction. To identify a DNA consensus sequence for YY1, binding site selections with purified protein and random DNA oligonucleotides have been performed (Hyde-DeRuyscher et al., 1995; Yant et al., 1995). A core consensus sequence of four nucleotides could be identified: CCAT. In a study of do Kim and Kim (2009), this sequence could be expanded to a longer binding motif: GCCGCCATTTTG. Figure 4.37 provides a co-crystal structure and a schematic representation of the YY1-DNA interction. The co-crystal structure of YY1 bound to the AAV P5 initiator element, which has been resolved by Houbaviy et al. (1996) - is in accordance with the previous findings and predictions on YY1's DNA binding activity.

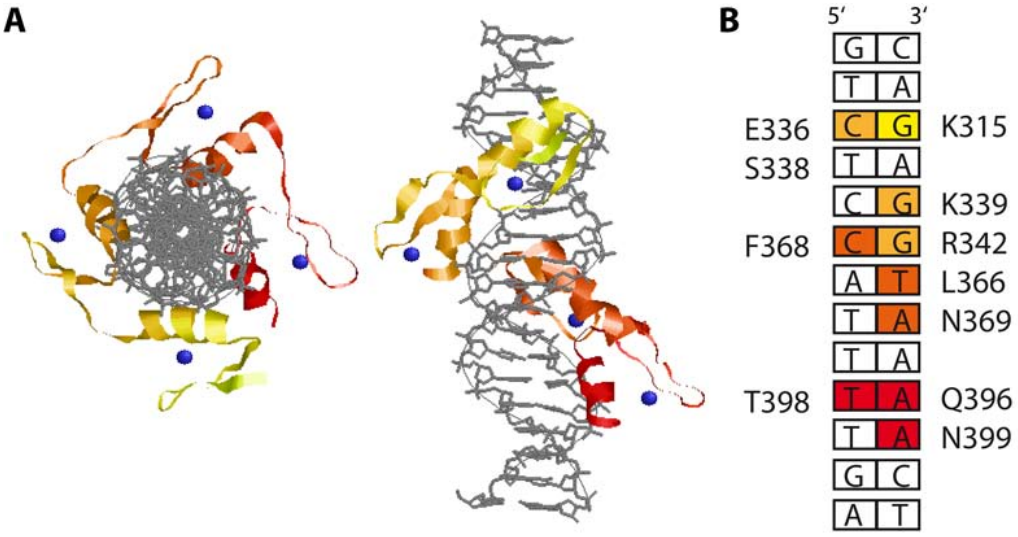


FIGURE 4.37 *Structure of YY1 bound to DNA.* ***A*** *presents the co-crystal structure of YY1 bound to the AAV P5 initiator element. The DNA is shown as a wireframe (gray), the YY1 zinc fingers are shown as a ribbon representation. Zinc fingers 1 - 4 are colored in yellow, orange, dark orange and red, respectively. The zinc ions are depicted as blue spheres. A top view of the model is depicted on the left, a lateral view on the right. The 2.5 Å structural model was retrieved from the Protein Data Bank entry 1UBD (www.pdb.org/). The picture was generated using a molecular graphics visualisation tool based on RasMol (by Roger Sayle).* ***B*** *represents a schematic model of the YY1-DNA interaction based on the crystal structure. The DNA strand is depicted and the protein-DNA contacts are colored dependent on their zinc finger of origin. Residues, which are responsible for the interaction are indicated. The model has been simplified; interactions with the DNA backbone are not shown. The co-crystal structure and the schematic model were taken from Houbaviy et al. (1996).*

YY1 is a sequence-specific DNA binding protein that has been found in a vast number of different contexts. The consensus sequence of the *Drosophila* and human protein is known. In this study, the DNA binding activity of the *Hydra* homolog has been examined. Since the four zinc fingers of *Hy*YY1 are highly similar to their counterparts in other species (see figure 4.46 on page 169), it is probable that *Hy*YY1 binds to the same consensus sequence. To provide evidence for this assumption, a DNA binding site selection followed by gel retardation assays has been performed with the *Hydra* protein.

4.4.1. DNA binding site selection

The DNA binding site selection for *Hy*YY1 was carried out in the same way as described for the human YY1 protein in the study by Hyde-DeRuyscher et al. (1995). In this study, a recombinantly expressed GST-YY1 fusion protein bound to glutathione-Sepharose beads was incubated with a double-stranded DNA containing a central random sequence. After the removal of unbound DNA, the protein-bound DNA has been eluted and amplified by PCR. The selected DNA was again incubated with the protein. After several rounds of selection, the sequences of the protein-bound DNA have been determined. In contrast to the mentioned approach, a His-YY1 fusion protein bound to agarose beads has been used for the binding site selection in this study. In figure 4.38 an overview of the procedure is depicted.

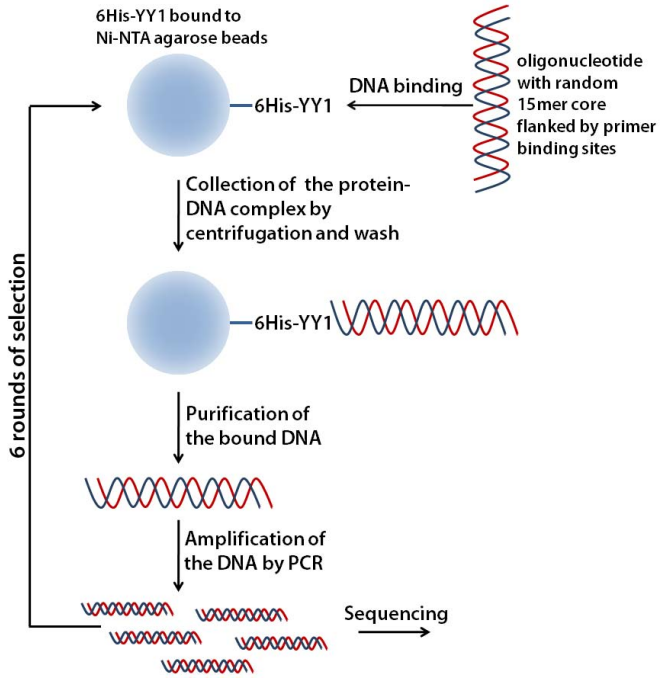


FIGURE 4.38 *Schematic overview of the binding site selection procedure.*

Oligonucleotide generation and protein immobilization

For the DNA binding site selection, a DNA of 85 nucleotides containing a random core sequence of 15 nucleotides and primer binding sites at both ends has been used. According to previous studies on YY1-DNA interactions, the zinc fingers contact a binding motif of 9 - 12 base pairs (Suzuki et al., 1994; Hyde-DeRuyscher et al., 1995; Yant et al., 1995; Houbaviy et al., 1996). Therefore, a random core of 15 nucleotides should be sufficient for the selection of the *Hy*YY1 consensus sequence. To generate a double-stranded oligonucleotide, the single-stranded DNA has been annealed with a short complementary DNA primer; the Klenow fragment was used to fill in recessed 3′ ends. The double-stranded DNA oligonucleotide was analyzed by gel electrophoresis (figure 4.39).

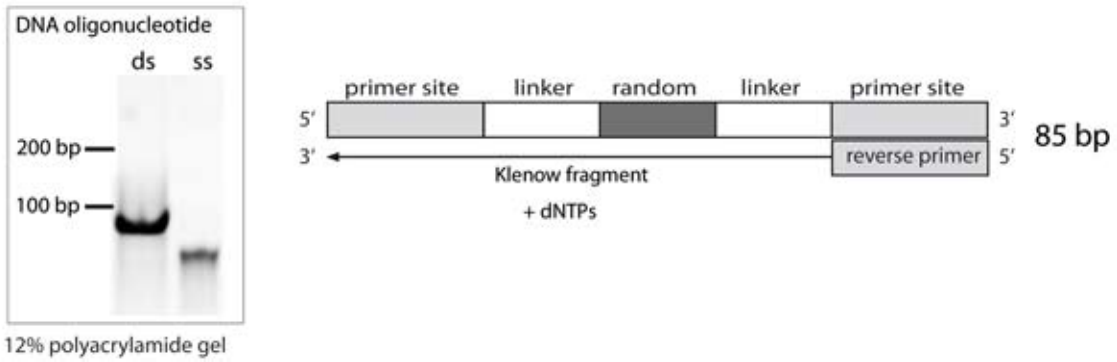


FIGURE 4.39 ***Left:*** *Analysis of the Klenow reaction on a 12% polyacrylamide gel. The generated double-stranded (ds) oligonucleotide migrates slower in the gel than the single-stranded (ss) DNA.* ***Right:*** *The 5′ –> 3′ polymerase activity of the Klenow fragment has been used to fill in the 3′ recessed ends of the single-stranded oligonucleotide.*

The generated oligonucleotides were incubated with recombinant *Hy*YY1 protein immobilized on agarose beads. The His-YY1 fusion protein was recombinantly expressed in bacteria and purified using metal ion affinity chromatography. The purified protein was not eluted from the nickel-NTA beads. Instead it was left bound to the beads. The purity and amount of the bound protein was analyzed by SDS-PAGE prior to the binding site selection. The *Hy*YY1ΔZnF protein was used as a background control for the binding site selection. The truncated protein should not have the ability to specifically bind to DNA, since the zinc fingers are responsible for its DNA binding activity. As shown in figure 4.40, both the full length recombinant protein and the truncated version could be sufficiently expressed and purified. 5 μl of the agarose-protein suspension were loaded onto the SDS-gel to estimate the amount of the bound protein in comparison to different amounts of BSA. The binding capacity of the Ni-NTA beads is up to 50 mg His-tagged protein per ml resin according to the manufacturer (Qiagen). The binding capacity varies and is strongly dependent on the characteristics of the fusion protein (for example amino acid composition, molecular weight, expression level). Whereas about 50 ng of the full length *Hy*YY1 fusion protein were bound to 1 μl of the Ni-NTA beads (figure 4.40 A), the truncated protein could be expressed more effectively and several microgram of the

protein were bound to 1 μl of the beads. The high protein amount resulted in a smear on the gel (figure 4.40 B).

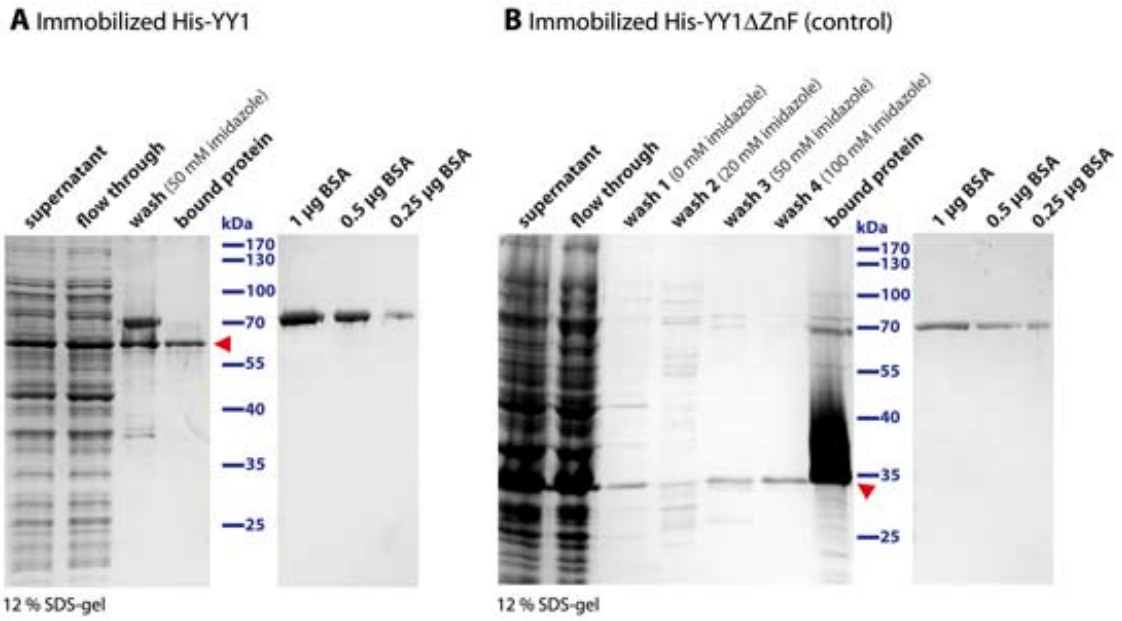


FIGURE 4.40 *Protein immobilization on Ni-NTA beads. The SDS-gel on the left presents the protein purification procedure of the recombinant HyYY1 fusion protein. The SDS-gel on the right shows the purification of the truncated control protein. 5 μl of each fraction have been applied to the gel. The respective protein fraction bound to the agarose beads is left to the protein size standard. The full length protein has a size of about 60 kDa, the truncated protein has a size of about 35 kDa. Decreasing amounts of BSA serve as a standard for estimating the concentration of the immobilized proteins. The red arrowheads indicate the positions of the recombinant proteins.*

Selection and analysis of *Hy*YY1 binding sites

For the first round of selection, 100 ng of the pool of random double-stranded oligonucleotides were incubated with an excess amount of immobilized protein (2.5 - 5 μg). Unbound DNA was removed and protein-bound DNA was retrieved and amplified by PCR. The amplified product was used for the next round of capture and amplification. A total of six rounds has been performed. After the third and the sixth round of selection, the amplified DNA was purified and ligated into pGEM-T vector. Competent bacterial cells were transformed with the ligation reaction. Positive clones were selected and the inserted DNA fragments have been sequenced to obtain the *Hy*YY1 consensus sequence. The captured DNA from the *Hy*YY1ΔZnF background control was amplified as well and the sequences have been determined.

The polyacrylamide gel in figure 4.41 shows the amplified DNA after each round of selection. The DNA captured by the truncated *Hy*YY1ΔZnF protein migrates at the expected 85 bp on a 12% polyacrylamide gel after each round of selection. However, the DNA captured by the full length protein changed in size. In the course of the binding site selection

two fragments of different sizes were amplified. The DNA amplified after round 4 of the selection procedure migrates at 85 bp and an additional band of a lower molecular weight appears on the gel. The lower band became more abundant during the following rounds of selection.

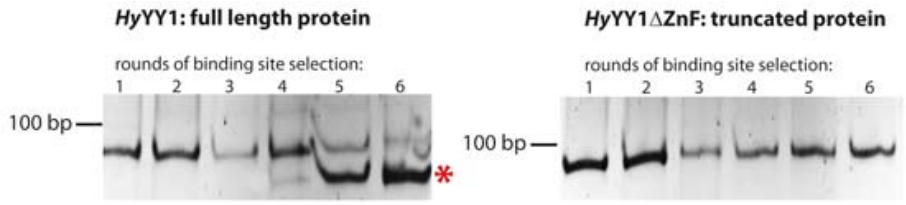


FIGURE 4.41 *The amplified DNA after each round of selection was seperated by gel electrophoresis.* ***Left:*** *DNA amplified during the binding site selection with the full length HyYY1 protein. The red asterisk indicates the position of the lower molecular weight oligonucleotide.* ***Right:*** *DNA amplified during the binding site selection with the truncated HyYY1ΔZnF protein.*

The tables on the following pages show the result of the binding site selection. The first table shows the sequences of oligonucleotides captured after rounds 3 and 6 of the selection procedure with the full length *Hy*YY1 protein (table 4.6), in the second table the sequences obtained with the truncated *Hy*YY1ΔZnF protein are depicted (table 4.7). The full length sequence of the captured oligonucleotides is shown. Random regions are indicated by underlined letters. There is an enrichment of the following sequence after round 6 of the selection procedure with the full length protein: GTAAAGTTTTCTTGA. 21 out of 21 sequenced clones contained the same conserved core sequence. No other sequence was otained during the binding site selection with the full length *Hy*YY1 protein. After round 3, no specific sequence was enriched. The captured oligonucleotide after round 6 is lacking 15 nucleotides in the constant region compared to the original oligonucleotide. This is in accordance with the observation during gel electrophoresis: The captured oligonucleotide shifts to a lower molecular weight in the course of the selection procedure. The sequences obtained with the truncated *Hy*YY1ΔZnF protein are completely random. As expected, no specific sequence was captured by the background control.

TABLE 4.6 *Sequences of oligonucleotides captured by HyYY1. Upper case letters indicate the constant regions, the random regions of the oligonucleotide are highlighted. In the first half of the table the sequences captured after round 3 of the binding site selection are listed, in the second half of the table the enriched sequences after round 6 of the selection procedure are listed.*

***Hy*YY1: Consensus sequences after 3 rounds of selection**

#	sequence	orientation
	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG NNNNNNNNNNNNNNN CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	
1	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TATCTATACATCGGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
2	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATCTTGTATATTTAT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
3	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATTGATCTTTAATTT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
4	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATTTAAAGAGTGTGG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
5	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TGGCGGGGAAACGCG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
6	...GATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATAGAAATTTTGGTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
7	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TATCCTTGGGAGTTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
8	......TGGCTACGCATCGTCAGCAGCAGCAGCAG GTCTAAATTTGAGGG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
9	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TATCTATACATCGGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
10	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG AGTTTTTTGTTGAAT CAG.....GCAGCAGGTCATACTCACTCACGGACT	−
11	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATTTCAGCTCTACAG CAGCAGCA.CAGCAGGTCATACTCACTCACGGACT	−
12	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG CCGTCTATATGGATT CAGCAGCAG.AGCAGGTCATACTCACTCACGGACT	−
13	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG AGTTTTTTGTTGAAT .....CAGCAGCTAGGTCATACTCACTCACGGACT	−
14	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GCGTCGACTCATTCA CAGCAGCAGCAGCAGGTCANACTCACTCACGGACT	−
15	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATTGGAAGGATATTT CAGCAG.......AGGTCATACTCACTCACGGACT	−
16	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TTGGCGAAGTTAAT. CAGCA.CAGCAGCAGGTCATACTCACTCACGGACT	−
17	..................CAGCAGCAGCAGCAG GGTTGTATGATATTA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
18	..TGATTGGCTACGCATCGTCAGCAGCAGCAG... CTACATGTTTTTGTT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
19	CGTGATTGGCT.CGCATCGTCAGCAGCAGCAGCAG CCGTGTCTATGTCAT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
20	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATATGCTAAGTATAA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
21	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TTGGACTGGCATATC ...CAGCAGCAGCAGGTCATACTCACTCACGGACT	−

***Hy*YY1: Consensus sequences after 6 rounds of selection**

#	sequence	orientation
	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG NNNNNNNNNNNNNNN CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	
1	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
2	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
3	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
4	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAG...GTCATACTCACTCACGGACT	−
5	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
6	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAG...GTCATACTCACTCACGGACT	−
7	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
8	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
9	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
10	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAG...GTCATACTCACTCACGGACT	−
11	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
12	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
13	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCGGGTCATACTCACTCACGGACT	−
14	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAG...GTCATACTCACTCACGGACT	−
15	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAG...GTCATACTCACTCACGGACT	−
16	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
17	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
18	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAG...GTCATACTCACTCACGGACT	−
19	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAG...GTCATACTCACTCACGGACT	−
20	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
21	CGTGATTGGCTACGCATCGT..... GTAAAGTTTTCTTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−

TABLE 4.7 *Sequences of oligonucleotides captured by HyYY1ΔZnF (background control). Upper case letters indicate the constant regions, the random regions of the oligonucleotide are highlighted. In the first half of the table the sequences captured after round 3 of the binding site selection are listed, in the second half of the table the sequences obtained after round 6 of the selection procedure are listed.*

HyYY1ΔZnF: Consensus sequences after 3 rounds of selection

#	sequence	orientation
	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG NNNNNNNNNNNNNNN CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	
1	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG AGCTGTGTCCTTTAT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
2	....ATTGGCTACGCATCGTCAGCAGCAGCAGCAG TAGACGGTGGACACT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
3	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GTGGTGGTTTAGTCC CAGCAGCAGCAG.AGGTCATACTCACTCACGGACT	−
4	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GTCGTCTGGTGCATA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
5	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GAGCTTGGATGAATT CAGAAGCAGCAGCAGGTCATACTCACTCACGGACT	+
6	...GATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATTACAAGTGGAGCG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
7	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GTACTCTTCACTTCT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
8	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG AAAGATATTTCTTGC CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
9	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG CCGTAAGCATGAATA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
10	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TAGTTAGCGTAGTAG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
11	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATATTATCCACCGGG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
12	....ATTGGCTACGCATCGTCAGCAGCAGCAGCAG TAGACGGTGGACACT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
13	CGTGATTGGCTACGCATCGTCAGCAGCAGAAG... CCGTGGGTCGCCTGA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
14	CGTGATTGGCTACGCATCGTCGGCAGCAGCAGCAG TCATCCGTCTGCGTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
15	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG AAAGATATTTCTTGC CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
16	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TTGGAACTCATAGCT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
17	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GGACAGGCGGATGTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
18	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG AGGGATTCGGGGGTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
19	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG AGGGGGGGGTTCCAC CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
20	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATCAGTTGATGGGCG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
21	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG CTTGGTGCCCTTCAT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
22	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GGCCCCGGTGGTGTA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+

HyYY1ΔZnF: Consensus sequences after 6 rounds of selection

#	sequence	orientation
	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG NNNNNNNNNNNNNNN CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	
1	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ACGTGGAGTATTAAT CAGCAGCAGCAACAGGTCATACTCACTCACGGACT	+
2	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GTGTGAAGCTTATGT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
3	CGTGATTGGCTACGCATCGTCAGTAGCAGCAGCAG TAGACGTTGTCGTCG CAGTAGCAGCAGCAGGTCATACTCACTCACGGACT	−
4	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GGTCGCGCAAGGTTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
5	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ACGTGGAGTATTAAT CAGCAGCAGCAACAGGTCATACTCACTCACGGACT	+
6	...GATTGGCTACGCATCGTCAGCAGCAGCAGCAG AATTTCTACGGAGTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
7	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GGCGGCGTGTCACTC CAGCAGCAGCAGCAGGTCATACTCACTCACGGA-T	−
8	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG CCTGCAGCACTGTGT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
9	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TAAATTCTAGAAGTC CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
10	...GATTGGCTACGCATCGTCAGCAGCAGCAGCAG GGTCTTGTTTTGTAG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
11	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TAAGAGGATGGTGCT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
12	CGTGATTGGCTACGCATCGTCAGCAGCAGCAACAG TTCGTCGTGCGACTT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
13	.GTGATTGGCTACGCATCGTCAGCAGCAGCAG... CGGTTGAGTTTAAAG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
14	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GAAGATAGTCGAGAT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
15	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TGATTTTCTGTCTAA CAGCAGCAGCAACAGGTCATACTCACTCACGGACT	−
16	CGTGATTGGCTACGCATCGTCAGCAG.AGCAGCAG ATTTGGTTTACTATA CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
17	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG ATCACGTTTTTCCCG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
18	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TTTGATTTCAATTGC CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
19	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG GAAGATAGTCGAGAT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
20	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TACCTAATAGTTAAC CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	+
21	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TACCCGTTTCCATTG CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−
22	CGTGATTGGCTACGCATCGTCAGCAGCAGCAGCAG TAAGAGGATGGTGCT CAGCAGCAGCAGCAGGTCATACTCACTCACGGACT	−

4.4.2. Gel retardation assays

To determine whether *Hy*YY1 can, indeed, bind to the selected site (GTAAAGTTTTCTTGA), a gel retardation assays have been performed. The gel retardation assay or electromobility shift assay (EMSA) is one possibility to analyze interactions between proteins and nucleic acids. Complexes of DNA and protein migrate more slowly compared to unbound, free DNA when subjected to gel electrophoresis. The protein-bound DNA is retarded / shifted in the gel. For gel retardation assays with the *Hy*YY1 protein a double stranded DNA oligonucleotide was used. The target DNA was end-labeled with biotin at its 3' termini. After the binding of the protein to the target DNA, the complex was resolved by non-denaturing gel electrophoresis using a TBE-polyacrylamide gel. The shift was detected by chemiluminescence using horseradish peroxidase coupled streptavidin.
The recombinant *Hy*YY1 and *Hy*YY1ΔZnF proteins for the gel retardation assays were expressed as 6His-fusion proteins and purified by metal ion affinity chromatography. The proteins were purified under the same conditions. In contrast to the proteins used for the binding site selection, the proteins were purified under denaturing conditions using a guanidin buffer. The proteins were eluted from the agarose beads using a buffer with a low pH. The buffer was exchanged to a neutral buffer and the proteins were renatured by a stepwise dialysis procedure (for details see materials and methods section 3.7.2). Finally the soluble proteins were buffered in 1 x PBS. The conditions for purification and the subsequent renaturing of the protein by dialysis were described for a human His-YY1 fusion protein by Shi et al. (1991).
The SDS-gel in figure 4.42 shows the purified proteins. With this purification method the protein yield was higher compared to the protein purification under native conditions. Protein amounts of 5 - 10 mg could be obtained compared to 1 - 2 mg.

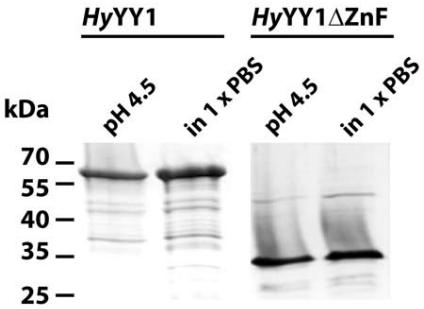


FIGURE 4.42 *Purification of HyYY1 and HyYY1ΔZnF monitored by SDS-PAGE. The proteins were purified under denaturing conditions and eluted using a stepwise decrease of the pH. The respective eluate fractions at pH 4.5 and the protein after dialysis against 1 x PBS are depicted.*

Gel retardation assay with *Hy*YY1 protein and oligonucleotides containing putative HyYY1 consensus sequences

To determine the ability of *Hy*YY1 to interact with DNA in a sequence-specific manner, a gel retardation assay with the purified recombinant protein and 50 bp oligonucleotides containing putative *Hy*YY1 consensus sequences was performed. 5 μg of recombinant

protein were incubated with 0.6 ng (= 20 fmol) of oligonucleotides containing the following 15 bp core sequences:

HyYY1 round 6: GTAAAGTTTTCTTGA
HyYY1 round 3: TTTTTGTTGAATGTC
control: CCTGCAGCACTGTGT
YY1 consensus: GCCG**CCAT**TTTGGGC

The consensus sequence of the 'YY1 consensus' oligonucleotide matches the known Pho/YY1 binding site, which has been identified by Hyde-DeRuyscher et al. (1995) and used for several YY1-DNA interaction studies. The 'HyYY1 round 3' oligonucleotide contains one of the sequences, which was obtained after round 3 of the binding site selection, the 'HyYY1 round 6' oligonucleotide contains the core sequence, which was captured by *Hy*YY1 after round 6 of the binding site selection. The core sequence of the 'control' oligonuleotide matches one of the sequences that was obtained after round 6 of the binding site selection with the truncated *Hy*YY1 protein. This oligonucleotide serves as a negative control. The DNA was seperated by gel electrophoresis on a 6% native polyacrylamide gel and the bands were visualized by chemiluminescence. The result is depicted in figure 4.43.

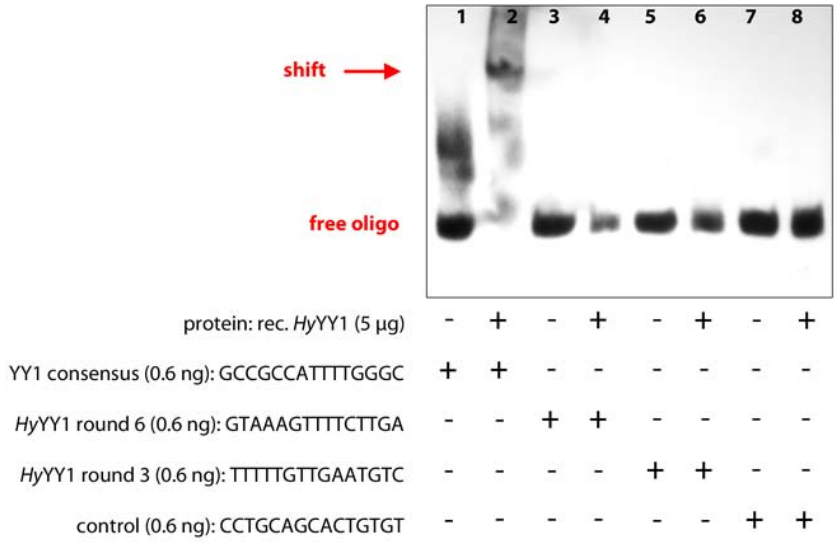


FIGURE 4.43 *Gel retardation assay with recombinant HyYY1 protein. 0.6 of a biotin-labeled 50 bp oligonucleotide were incubated with 5 µg of HyYY1 protein and seperated by gel electrophoresis on a 6% polyacrylamide gel. Oligonucleotides containing the core sequences obtained after rounds 3 and 6 of the HyYY1 binding site selection and an oligonucleotide containing the known Pho/YY1 binding site were used for the binding reaction. The bands were detected by HRP-conjugated streptavidin and visualized by chemiluminescence. A control with free oligonucleotide is shown in the first lane.*

The gel retardation assay revealed that the *Hy*YY1 protein can indeed interact with DNA. The protein efficiently forms a complex with the oligonucleotide containing the known Pho/YY1 consensus sequence. A shift of the DNA is clearly visible (figure 4.43 lane 2).

However, the recombinant protein could not bind to the oligonucleotide containing the putative *Hy*YY1 consensus sequence, which was obtained in the course of the binding site selection. No shift was visible (figure 4.43 lane 4). No DNA-protein complex was formed with the 'HyYY1 round 3' oligonucleotide and the control DNA (figure 4.43 lane 6 and 8). The *Hy*YY1 protein could only interact with the oligonucleotide containing the known Pho/YY1 consensus sequence: GCCATNTT. The other oligonucleotides were level with the free oligonucleotide on the gel (figure 4.43 band 1). The protein-DNA interaction of *Hy*YY1 is sequence-specific.

Gel retardation assay with increasing amounts of *Hy*YY1 and *Hy*YY1ΔZnF protein

Since the truncated *Hy*YY1ΔZnF protein could bind to the random core oligonucleotide during the binding site selection quite well, it was determined whether the protein can interact with DNA in an unspecific manner. A gel retardation assay with different amounts of either *Hy*YY1 or *Hy*YY1ΔZnF protein was performed. The recombinant proteins were incubated with 20 fmol of the biotin labeled DNA oligonucleotide containing the known Pho/YY1 consensus sequence. Figure 4.44 shows that only the full length protein containing the four C2H2-type zinc fingers can bind to DNA. The protein lacking the zinc finger region could not form a protein-DNA complex and shift the DNA in the gel (figure 4.44 lanes 7 - 10). The outcome of the experiment is not surprising since several studies showed that the DNA binding activity of YY1 is dependent on its zinc fingers.
The binding reaction was performed with increasing amounts of recombinant protein ranging from 0.5 - 5 μg. A protein amount of 2.5 μg could shift the DNA completely (figure 4.44 lane 3). This is equivalent to a $\sim$4 fold molar excess of protein. Lower protein amounts (0.5 - 1 μg) led to an incomplete shift of the oligonucleotide. Both free oligonucleotide and complexed DNA are visible on the gel (figure 4.44 lane 4 and 5).

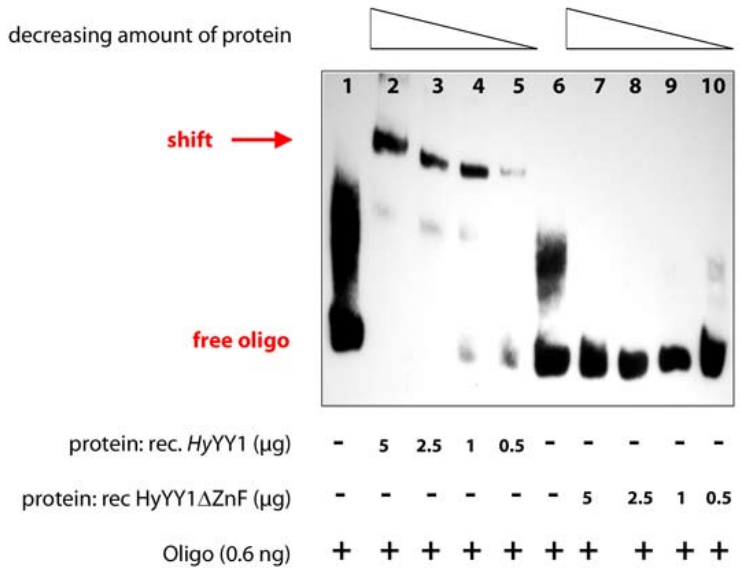


FIGURE 4.44 *Gel retardation assay with recombinant HyYY1 and HyYY1ΔZnF. 0.6 ng of the biotin-labeled oligonucleotide containing the known Pho/YY1 binding site were incubated with different amounts of HyYY1 protein ranging from 0.5 - 5 μg and seperated by gel electrophoresis on a 6% polyacrylamide gel. The bands were detected by HRP-conjugated streptavidin and visualized by chemiluminescence.*

4.4.3. SUMMARY

The DNA binding motif of the *Hydra* YY1 homolog corresponds to the motif of the human YY1 and the *Drosophila* Pho

With the help of the gel retardation assays it could be shown that the *Hy*YY1 protein can indeed interact with DNA like its human or *Drosophila* homolog. *Hy*YY1 binds to double-stranded DNA in a sequence-specific manner and its consensus motif corresponds to the known Pho/YY1 consensus sequence. The DNA binding activitiy of *Hy*YY1 exclusively depends on its C-terminal zinc finger domains. A truncated protein without these domains could not form a complex with DNA during gel retardation compared to the full length *Hy*YY1 protein.
Obviously, *Hy*YY1 binds to the known Pho/YY1 consensus motif. The sequence enriched during the binding site selection was not bound by *Hy*YY1 in the course of the gel retardation assays. The captured sequence (GTAAAGTTTTCTTGA) differs from the known consensus sequence of human YY1 or *Drosophila* Pho. The known consensus sequence is: GCCATNTT (Hyde-DeRuyscher et al., 1995; do Kim and Kim, 2009). Sequence logos of the captured sequence and the mammalian consensus are depicted in figure 4.45.

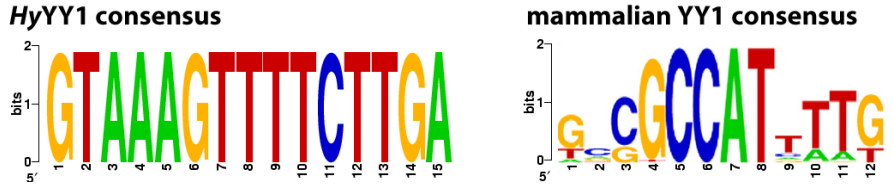


FIGURE 4.45 *Sequence logos. The sequence captured by HyYY1 in the course of the binding site selection is depicted on the left and was generated with the help of the sequence logos program (http://weblogo.berkeley.edu/) The consensus sequence of the human YY1 is depicted on the right. The logo was taken from (do Kim and Kim, 2009).*

It is very unlikely that the *Hydra* YY1 protein binds to a different consensus, since the amino acid sequence of the *Hy*YY1 zinc fingers is 94% identical to the human protein sequence and 92% identical to the zinc fingers of Pho. Especially, the amino acid residues, which are responsible for the DNA contact are identical (Houbaviy et al., 1996; Kim et al., 2007). A sequence alignment of the zinc finger domains of *Hy*YY1 and YY1 proteins from other species is depicted in figure 4.46. The amino acid residues, which are known to contact the target DNA are highlighted.
Neither the residues that directly contact the bases, nor the residues that contact the DNA backbone are exchanged in the *Hydra* protein. However, there are several amino acid substitutions in residues that are not known to contact the DNA. When compared to the human YY1 protein, there are four substitutions in the first zinc finger domain (Ala→Ser209, Lys→Ser213, Thr→Leu216, Met→Leu218), there is one substitution in the second zinc finger (Val→Thr296) and there are two amino acid substitutions in the forth zinc finger domain (Gly→Ser360, Lys→Cys364). Maybe these residues alter the structure

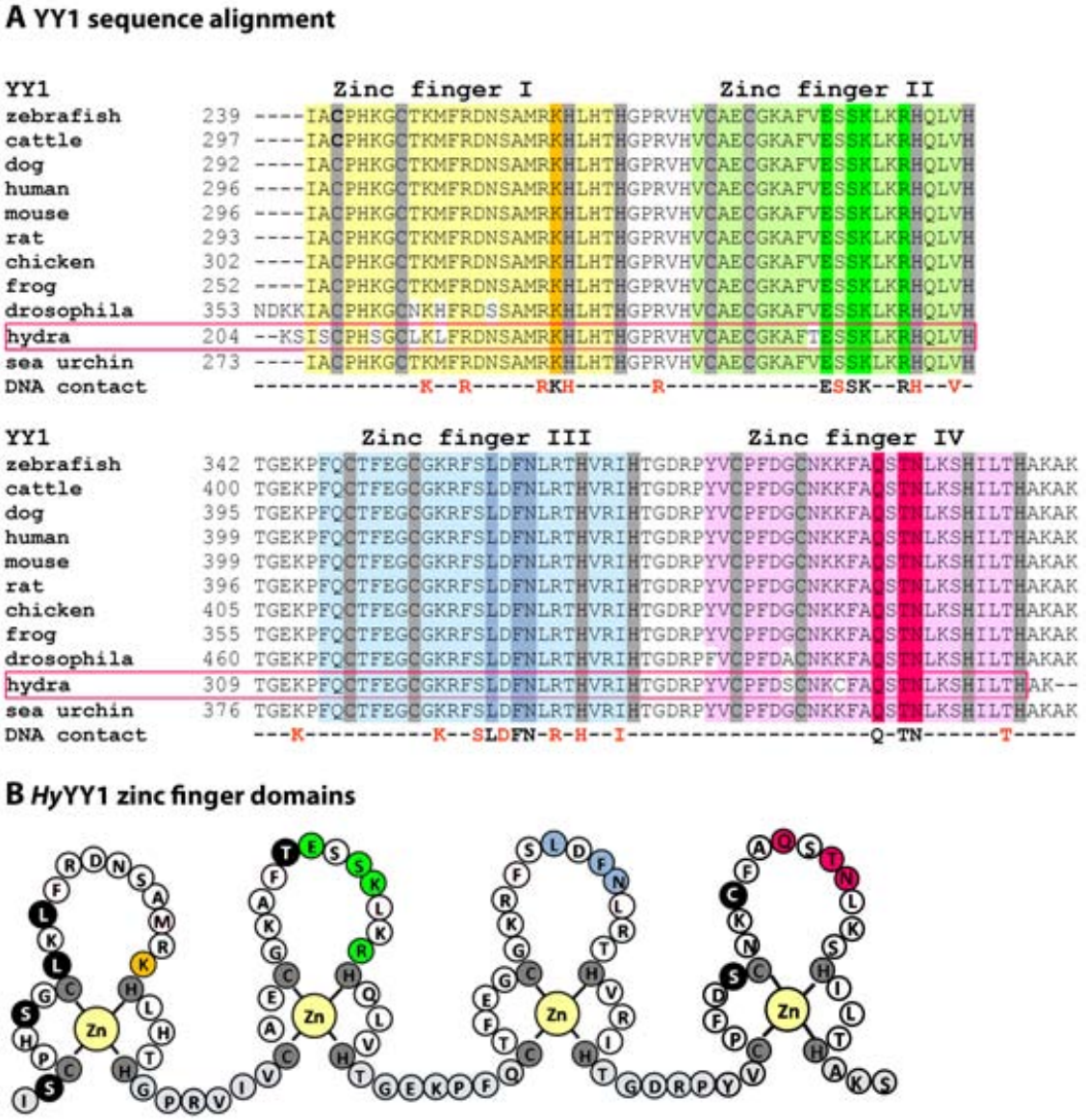


FIGURE 4.46 *Sequence alignment of the YY1 zinc finger domains.* ***A*** *The zinc finger domains of HyYY1 are compared to the zinc fingers of YY1 proteins from different species. The HyYY1 amino acid sequence is framed in red. The zinc finger domains are indicated and highlighted in yellow, green, blue and red. The cysteine and histidine residues of the C2H2-type zinc fingers are highlighted in gray, residues known to contact the bases of the target DNA are highlighted and indicated on the last row (black letters). Residues that contact the DNA backbone are indicated on the last row (red letters).* ***B*** *Overview of the Hydra YY1 zinc fingers. The color code is the same as for the alignment. Non-conserved residues in comparison to the human YY1 protein are highlighted in black.*

of the affected zinc fingers and decrease the binding affinity of the *Hydra* protein to the DNA.

Kim et al. (2007) could show that YY1 has been duplicated by retroposition events in the course of time in *Drosophila* and mammals. In fly, Phol (Pleiohomeotic like) has been duplicated from Pho and in mammals YY2 and REX1 (Reduced Expression 1) have been duplicated from YY1. The duplicated proteins all show amino acid substitutions in their

zinc finger regions and have been shown to bind DNA with a lower affinity. But even REX1, which shows some of the amino acid exchanges of *Hy*YY1, has a consensus motif with a conserved core sequence: CCAT. This sequence can be found in the YY1 motif and in the consensus motifs of all duplicated proteins.

The results of the *Hy*YY1 binding site selection are not in accordance with the gel retardation assays

There is a discrepancy between the results of the *Hy*YY1 binding site selection and the gel retardation assays. The question arises: Why is the captured sequence not shifted by *Hy*YY1 during the gel retardation assay? One possibility would be that the known consensus sequence is not found in the pool of random oligonucleotides used for the binding site selection at all, another possibility would be that the enriched consensus sequence is a PCR artifact.
The first possibility is very unlikely - in case the four nucleobases get equally incorporated in the oligonucleotide during the synthesis. $1.07 \cdot 10^9$ (= 4^{15}) different molecules varying in their random core sequences would exist. At the beginning of the binding site selection, 100 ng of the pool of random oligonucleotides has been used for the first round of selection. 100 ng correspond to $1.96 \cdot 10^{12}$ molecules (n = m/M = $100 \cdot 10^{-9}$ g / 30730 g/mol = $3.25 \cdot 10^{-12}$ mol; N = $N_A \cdot$ n = $6.022 \cdot 10^{23}$ $mol^{-1} \cdot 3.25 \cdot 10^{-12}$ mol = $1.96 \cdot 10^{12}$ molecules). Since the number of molecules is about 2000 times higher than the number of possible different molecules, each molecule should exist in the pool of oligonucleotides in an adequate amount.
Probably, the enriched sequence is a PCR artifact. The sequence captured by the full length *Hy*YY1 protein decreases in size in the course of the binding site selection. The smaller oligonucleotide may be a PCR artifact, which competes with the original 85 bp oligonucleotide during amplification. The truncated oligonucleotide gets enriched and adulterates the result. However, the sequence never gets enriched by the control protein. This contradicts the notion that the enrichment of the truncated sequence is just an artifact during PCR amplification. Maybe the *Hy*YY1-DNA interaction is specific but weak and the protein-DNA complex is not stable during electrophoresis.

4.5. Protein-protein interactions and targeting

In *Drosophila*, the PcG proteins are recruited to their target genes via the interaction of the Pleiohomeotic repressive complex (PhoRC) with so-called Polycomb response elements (PREs). Pho is a member of the PhoRC complex and serves as a molecular linker between the PcG repressive complexes and their target genes (Klymenko et al., 2006). A direct interaction between Pho and proteins of the *Drosophila* PRC1 and PRC2 complexes could be demonstrated (Mohd-Sarip et al., 2002; Poux et al., 2001b).
The existence of a vertebrate PhoRC complex and its role in recruiting the PcG complexes to their target genes is still to be shown. However, there is growing evidence for a role of YY1 in PcG target gene recruitment and the interaction of YY1 with proteins of the PRC1 and PRC2 complexes in vertebrates. Satijn et al. (2001) could identify YY1 in a complex together with EED and EZH, and they could demonstrate that an interaction between YY1 and EED is important for *Xenopus* neural induction. Caretti et al. (2004) could demonstrate the interaction of YY1 with EZH2 during muscle differentiation in mouse embryos and Kobrossy et al. (2006) could show that YY1 interacts with the PRC1 protein Mel18 (Psc) during neural differentiation in mouse.
During this study, it could already be shown that *Hy*YY1 interacts with DNA in a sequence-specific manner, which is a prerequisite for the role of *Hy*YY1 as a molecular adapter. The interaction between *Hy*YY1 and other *Hydra* PcG proteins is examined with the help of co-immunoprecipitation studies. In figure 4.47 a possible model of YY1 as a molecular adapter linking the PRC1 and PRC2 complexes to DNA is depicted.

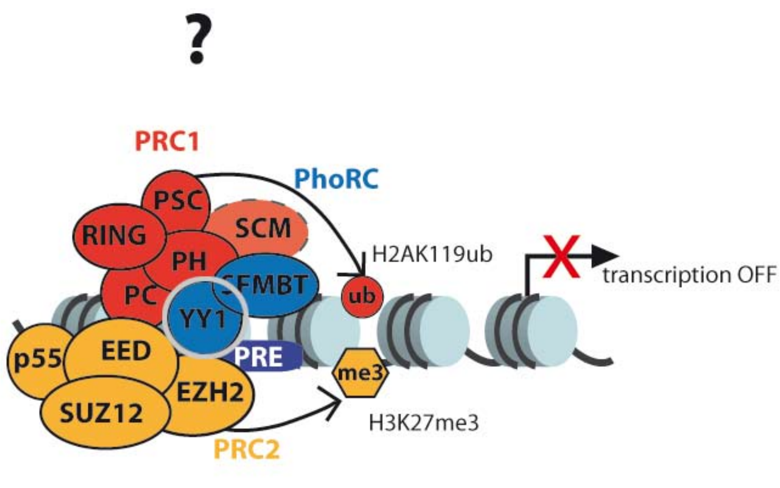


FIGURE 4.47 *YY1 as a molecular adapter linking the PcG complexes to DNA.*

4.5.1. Interactions between *Hy*YY1 and the *Hydra* PRC2 complex

The interaction between *Hy*YY1 and the *Hydra* PRC2 proteins was examined with the help of immunoprecipitation techniques. Co-immunoprecipitations were performed both with recombinant proteins expressed in insect cells and with endogenous *Hydra* proteins. The recombinant system has the advantage that some of the *Hydra* PcG proteins can be expressed as tagged fusion proteins, which allows the use of antibodies raised in different species for precipitation and detection, since all generated antibodies against the *Hydra* PcG proteins were produced in rabbit. Using antibodies of the same species can result in the cross-reactions of the detection antibody with the heavy and light chain of the precipitation antibody. Non-covalently bound antibodies will elute from the solid support together with the precipitated proteins and appear on Western blot gels. Mainly the cross-reaction with the antibody's heavy chain may cause a problem. The heavy chain is detected at ∼55 kDa during Western blotting, which is in the range of some of the PcG proteins (*Hy*EED: ∼70 kDa , *Hy*p55: ∼55 kDa and *Hy*YY1: ∼60 kDa on SDS-gels). The proteins of interest may be masked by the heavy chain. To avoid this cross-reactivity in the endogenous system using *Hydra* extracts, the precipitating antibody was immobilized on protein A-agarose beads by cross-linking the antibody to the solid support.
For immunoprecipitation studies with recombinant proteins, *Hy*YY1 and the *Hydra* PRC2 proteins were expressed in Sf9 insect cells. Compared with a bacterial expression system, the insect cells are a higher eukaryotic system, which allows post-translational modifications and the correct folding of the recombinant proteins. It has already been demonstrated that functional PcG complexes can be reconstituted in Sf9 insect cells - both for the *Drosophila* PRC1 complex (Francis et al., 2001) and for the *Hydra* PRC2 complex (Lauschke, 2009).

*Hy*YY1 protein expression in insect cells

In a first step, *Hy*YY1 had to be expressed in Sf9 insect cells and the conditions for recombinant protein expression had to be optimized.
The *Hy*YY1 protein was expressed using the baculovirus expression system. For this purpose, the full length coding sequence of *Hy*YY1 was cloned into the baculovirus expression vector pFastBac1 and a recombinant bacmid was produced. Sf9 insect cells were transfected with the bacmid to produce recombinant *Hy*YY1-6His-baculoviruses. The baculovirus stock was amplified to obtain a high titer P3 viral stock, which was used for all further experiments. The conditions for *Hy*YY1-6His protein expression were optimized by infecting insect cells with increasing MOIs. The amplification procedure from the generation of the low titer P1 viral stock to the high titer P3 viral stock is depicted in figure 4.48 A. In figure 4.48 B and C the optimization procedure is shown. Recombinant protein expression was monitored by immunoblotting.

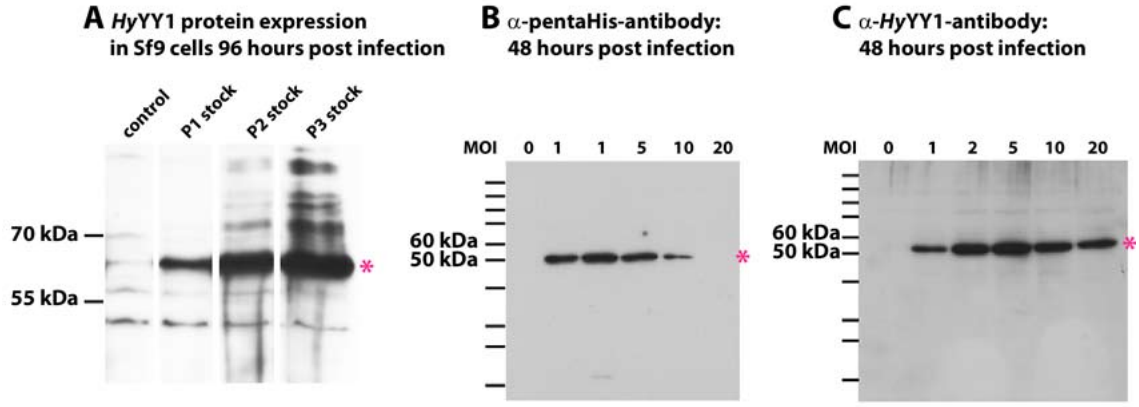


FIGURE 4.48 *Recombinant HyYY1-6His protein expression in Sf9 insect cells detected by immunoblotting using the α-HyYY1-antibody. The red asterisks marks the height of the recombinant HyYY1-6His protein.* ***A*** *Amplification of the recombinant HyYY1-6His baculovirus. Uninfected Sf9 cells served as a control.* ***B, C*** *Optimization of the infection and protein expression conditions. Sf9 cells were infected with the P3 viral stock using MOIs ranging from 1 to 20.*

The *Hy*YY1 protein could be successfully expressed in Sf9 cells. Infecting Sf9 cells with a MOI ranging from 1 - 5 for 24 - 48 hours were the best conditions for *Hy*YY1-6His protein expression. Infection of the insect cells with higher MOIs (10 - 20) resulted in a decreased protein expression. Longer expression periods ($>$ 48 hours) resulted in an increase of protein degradation.

Immunoprecipitation of *Hy*YY1

Different antibodies were tested for the immunoprecipitation of the *Hy*YY1-6His fusion protein from Sf9 cells. Both the generated α-*Hy*YY1-antibody (rabbit, polyclonal) and the commercially available α-pentaHis-antibody (mouse, monoclonal) were tested for the application in immunoprecipitation studies. For each approach, $2 \cdot 10^7$ Sf9 cells were infected with the generated P3 stock of the recombinant *Hy*YY1-6His baculovirus using a MOI of 2. The recombinant protein was expressed for 48 hours. The cells were lysed and immunoprecipitated with the respective antibodies bound to immobilized protein A-agarose beads. The precipitated protein was eluted by boiling the beads in protein sample buffer and detected by Western blotting. Figure 4.49 shows the result of the immunoprecipitations.

The *Hy*YY1-6His fusion protein can be successfully precipitated with both the α-*Hy*YY1-antibody and the α-pentaHis-antibody (figure 4.49 A, B). When precipitating with α-*Hy*YY1, the α-pentaHis-antibody was used for detection and vice-versa. Since the antibodies were raised in different species, no cross-reactivity of the antibodies interfered with the detection of the fusion protein.

Figure 4.49 C shows the result of the immunoprecipitation of the *Hy*YY1-6His fusion protein using the α-*Hy*YY1-antibody for both precipitation and Western blotting. The precipitating antibody was cross-linked to the immobilized protein A-agarose beads to circumvent or minimize the cross-reactivity during detection. Cross-linking was per-

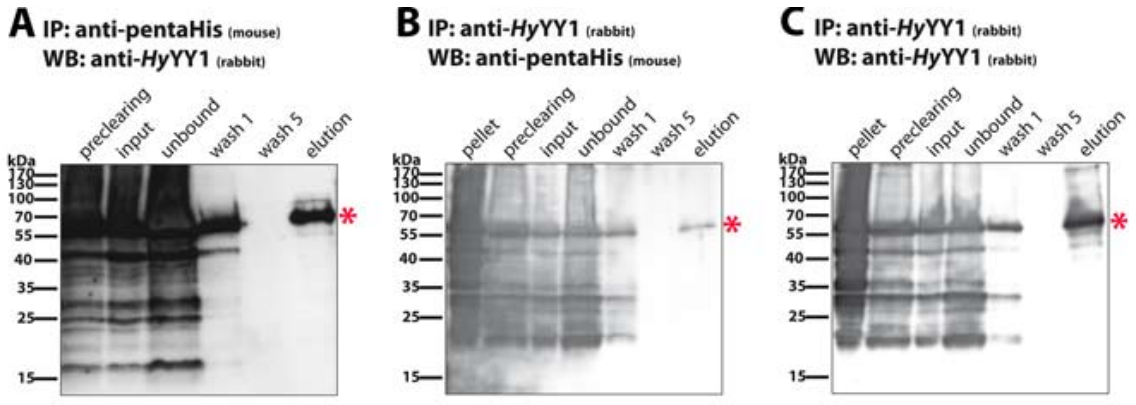


FIGURE 4.49 *HyYY1-immmunoprecipitations using antibodies either against the His-tag of the 6His-HyYY1 fusion protein (α-pentaHis-antibody; mouse) or against HyYY1 (α-HyYY1-antibody; rabbit). Western blots of the immunoprecipitations were probed with the same antibodies. The red asterisks marks the height of the recombinant HyYY1-6His protein.*
A *Antibody used for IP: α-pentaHis; antibody used for WB: α-HyYY1.*
B *Antibody used for IP: α-HyYY1; antibody used for WB: α-pentaHis.*
C *Antibody used for IP: α-HyYY1; antibody used for WB: α-HyYY1.*

formed with the bi-functional coupling reagent dimethyl pimelimidate (DMP). The successful cross-linking of the antibody could be detected by the absence of the antibody light chain in the elution fraction. One would expect to see the heavy and light chains of the antibody in the elution fraction if the cross-linking failed. In the case of the immunoprecipitation of the *Hy*YY1-6His fusion protein, no definite conclusion could be drawn for the heavy chain since both migrate on the same height on the gel.
Immunoprecipitation studies are strongly dependent on the abundance of the antigen and the affinity of the precipitating antibody to its antigen. This initial experiment demonstrates that the *Hy*YY1 protein is expressed in Sf9 cells in an adequate amount for immunoprecipitation and that the generated α-*Hy*YY1-antibody is useful for precipitation.

Co-immunoprecipitation of recombinant *Hy*YY1 and *Hydra* PRC2 proteins: first results

To study possible protein-protein interactions between *Hy*YY1 and the *Hydra* PRC2 proteins, a co-immunoprecipitation of the proteins was performed. An antibody against *Hy*YY1 was used to target a putative *Hy*YY1-PRC2 multiprotein complex and pull the entire complex out of solution. During his master thesis, Volker Lauschke could already demonstrate an interaction between *Hy*EED and *Hy*EZH2 with this technique.
For the co-immunoprecipitation of *Hy*YY1 with the *Hydra* PRC2 proteins, Sf9 insect cells were co-infected with the respective baculoviruses. The cells were co-infected with the following viruses: *Hy*YY1-6His, *Hy*EZH2, *Hy*EED, *Hy*SUZ12 and *Hy*p55. For all viruses, except for *Hy*SUZ12, a MOI of 1 was used for infection. For *Hy*SUZ12 a MOI of 2 was used. The proteins were expressed for 48 hours. The cells were lysed and the

cleared lysate was used for immunoprecipitation. For precipitation, an antibody against *Hy*YY1 (α-pentaHis) was used, for detection antibodies against *Hy*EZH2, *Hy*EED and *Hy*p55 were used. In figure 4.50 the result of the co-immunoprecipitations of *Hy*YY1 and *Hy*EZH2 is depicted.

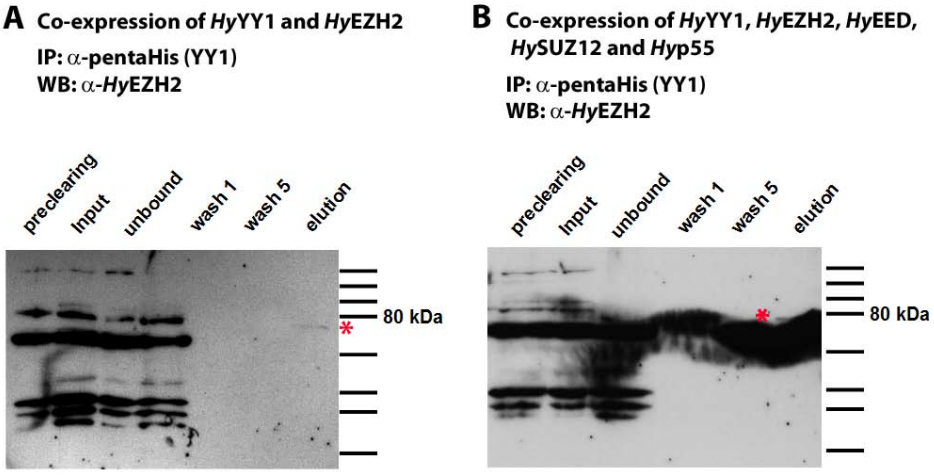


FIGURE 4.50 *Co-immmunoprecipitations using an antibody against the HyYY1-6His fusion protein (α-pentaHis-antibody; mouse). Western blots of the immunoprecipitations were probed with an antibody against HyEZH2 (α-HyEZH2-antibody; rabbit).* ***A*** *Co-immunoprecipitation using Sf9 cells co-expressing HyYY1 and HyEZH2.* ***B*** *Co-immunoprecipitation using Sf9 cells co-expressing HyYY1, HyEZH2, HyEED, HySUZ12 and Hyp55.*

The co-immunoprecipitation of *Hy*YY1 and the *Hydra* PRC2 proteins indicate that *Hy*YY1 physically interacts with the *Hy*EZH2 protein. A weak band of the correct size ($\sim$80 kDa) for *Hy*EZH2 is visible in the elution fraction. The *Hy*EZH2 protein could be co-immunoprecipitated with *Hy*YY1 from infected Sf9 cells either expressing *Hy*YY1 and *Hy*EZH2 (figure 4.50 A) or expressing *Hy*YY1, *Hy*EZH2, *Hy*EED, *Hy*SUZ12 and *Hy*p55 (figure 4.50 B). These co-immunoprecipitations give a first evidence for the interaction of *Hy*YY1 with one of the proteins of the *Hydra* PRC2 complex. However, the obtained band is weak and at the blot in figure 4.50 B shows a strong background masking a part of the band.

The presented co-immunoprecipitations should be seen as a preliminary experiment, which has to be repeated to assure the interaction between *Hy*YY1 and the *Hydra* PRC2 proteins. Two other antibodies were used for the detection of *Hy*EED and *Hy*p55. An antibody against *Hy*SUZ12 was not available. However, no interaction between *Hy*YY1 and *Hy*EED or *Hy*p55 could be detected, so far. Furthermore, the interaction between *Hy*YY1 and proteins of the *Hydra* PRC1 complex has yet to be demonstrated.

For future experiments, could be useful to perform the co-immunoprecipitation using varying conditions. The precipitation was performed with cells lysed under denaturing conditions using ionic detergents like SDS and sodium deoxycholate, which may contribute to the denaturing of multiprotein complexes and protein-antibody complexes.

Instead, a milder, non-denaturing buffer, which allows efficient cell lysis and keeps the protein complexes intact, could be applied for future experiments, since immunoprecipitations rely on the formation of a stable antigen-antibody complex.

Co-immunoprecipitation of endogenous *Hydra* PcG proteins: first results

Co-immunoprecipitation using endogenous *Hydra* proteins were performed in the same way as co-immunoprecipitations with recombinant proteins expressed in Sf9 cells. about 1000 *Hydra* polyps ($\sim 10^8$ cells) were used for each immunoprecipitation approach. The precipitating antibodies were cross-linked to the protein A-agarose beads. Immunoprecipitations of endogenous *Hydra* proteins were performed with the following precipitating antibodies: α-*Hy*YY1, α-*Hy*EED and α-*Hy*p55. The suitability of these antibodies for immunoprecipitation has been tested before. During this study, the α-*Hy*YY1-antibody has been successfully tested, the α-*Hy*EED and α-*Hy*p55 antibodies have been tested by Volker Lauschke during his master thesis. However, the antibodies were used for the precipitation of recombinantly overexpressed proteins in insect cells, no immunoprecipitations of endogenous *Hydra* proteins were performed, yet. Figure 4.51 shows exemplary results of the co-immunoprecipitation of endogenous *Hydra* proteins. The result for the co-immunoprecipitation of *Hy*YY1 and *Hy*EED is depicted in figure 4.51 A, the result of the co-immunoprecipitation of *Hy*EED and *Hy*p55 in figure 4.51 B.

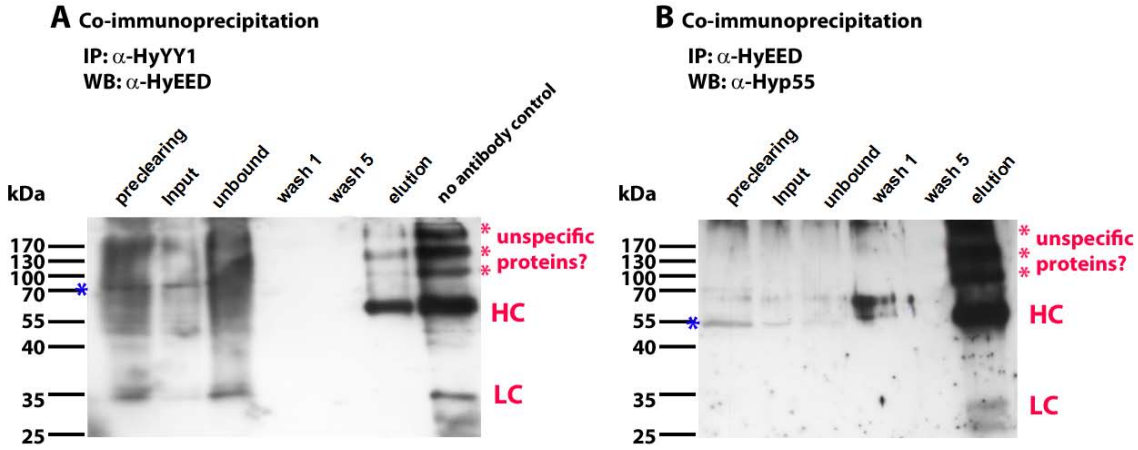


FIGURE 4.51 *Immunoprecipitation of endogenous Hydra proteins.* ***A*** *Co-immmunoprecipitations using an antibody against HyYY1. Western blots of the immunoprecipitations were probed with an antibody against HyEED.* ***B*** *Immunoprecipitation using an antibody against HyEED. Western blots of the immunoprecipitations were probed with an antibody against Hyp55.*

No protein-protein interactions could be detected for the endogenous *Hydra* PcG proteins. Neither for *Hy*YY1 and *Hy*EED, nor for *Hy*EED and *Hy*p55. In the elution fractions, no bands at the expected size of the probed *Hydra* PcG proteins are visible. Instead, clear bands at 55 kDa and 25 kDa are visible. These bands correspond to the heavy and light chains of the precipitating antibodies. This indicates that the cross-linking of the precip-

itating antibody to the matrix was not successful. The antibody is still eluted from the beads and interferes with the detection of the target proteins. In addition, higher molecular weight bands greater than 100 kDa are visible (red asterisks) - no matter, which combination of the antibodies was used for precipitation and detection. The bands even appeared in the control reaction, when precipitating was performed with protein A-agarose beads only. It seems that some of the endogenous *Hydra* proteins bind to the protein A-agarose matrix, even though a pre-clearing step with protein A-agarose beads was performed. Protein bands at the expected sizes for *Hy*EED and *Hy*p55 can be detected in the 'preclearing', 'input' and 'unbound' fraction (blue asterisks). For *Hy*EED a band at $\sim$70 kDa is visible, for *Hy*p55 a band at $\sim$55 kDa can be detected.
The results for both co-immunoprecipitations are negative. Either the target proteins are masked in the elution fraction, or the abundance of the proteins is too low and they could not be precipitated at all.
For further immunoprecipitations using *Hydra* lysates, the different steps of the procedure need to be otimized. Optimization could include the use of a different cell lysis buffer, more thorough washing steps and milder elution conditions. The RIPA buffer used for cell lysis may denature the multiprotein complexes and interfere with the formation of a stable protein-antibody complex. Furthermore, the elution by boiling the beads in SDS is a rather harsh method and the proteins could be eluted at a low pH using a glycine buffer, instead. Due to the lack of material and time constrictions, no further experimental conditions could be tested during this study.

4.5.2. Identification of PcG target genes in *Hydra* by ChIP-Seq

In 2006 the first genome-wide analyses of PcG binding patterns by ChIP-on-chip have been reported in *Drosophila* (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). In the same year, first ChIP-on-chip analyses have been carried out in murine and human embryonic stem cells and human embryonic fibroblasts, respectively (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). An alternative to the ChIP-on-chip technology is ChIP-Seq. Since no microarray for ChIP-on-chip is available for the *Hydra* genome, the ChIP-Seq method has been chosen to identify PcG target genes in *Hydra*.

Small-scale ChIP-Seq: initial settings

During this study, an initial small-scale ChIP-Seq experiment was performed with the generated antibodies against the *Hydra* PcG proteins and a commercially available α-H3K27me3 antibody. The α-H3K27me3 antibody is directed against the epigenetic mark set by the PRC2 complex: trimethylation of lysine 27 in histone H3. A putative *Hydra* histone H3 can be found in the genome, and the amino acid sequence of the N-terminal histone tail is 100% identical to the human histone. See appendix B for more details on *Hydra* histones.

For ChIP-Seq, ∼5000 *Hydra* polyps were used as starting material. The protein-DNA complexes were cross-linked with paraformaldehyde and chromatin was sheared by sonication. Different sonication settings were tested to obtain DNA fragment sizes ranging from 200 - 500 basepairs, which are optimal for chromatin immunoprecipitation. Figure 4.52 depicts the *Hydra* chromatin preparations. Chromatin preparation 2 was used for precipitation. Most of the chromatin fragments are ranging from ∼300 - 1000 basepairs. The size of the obtained chromatin fragments is still very big. Nevertheless, the prepared chromatin was immunoprecipitated with the following antibodies: α-*Hy*EZH2, α-*Hy*PC, α-*Hy*YY1 and the commercially available α-H3K27me3 antibody. The DNA was purified from the precipitated chromatin and amplified. The obtained DNA was cloned into pGEM-T vector, isolated and further analyzed by sequencing. For details see materials and methods section 3.10.2. Table 4.8 gives an overview of the sequenced clones.

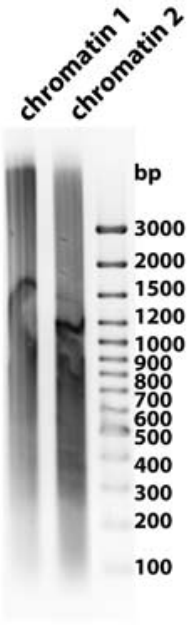


FIGURE 4.52 *Sheared Hydra chromatin. Chromatin was sheared with the help of a Branson Sonifier 250 sonicator. Different settings were used for shearing. Chromatin 1: power 2, duty cycle 20%; Chromatin 2: power 3, duty cycle 30%. For analysis, 1 μg of DNA was loaded per lane on a 1% agarose gel.*

precipitating antibody	number of sequenced clones	sequences found in the *Hydra* genome	clone	sequences found in the *Hydra* genome
α-*Hy*Pc	6	6 / 6	# 20	CTGCTCGAATTCAAGCTTCTGCTGTTGAGATAAAGCGCCAGGCAATATCTATGACAGCCGGTATAAATTAATTAATGAACAAGTTA GTTCAGTCACTAAATAATAAAAAACGTTAATAGTAAGCTTTCTTGGTTTTTATCTTGTTTGTTATTGCTTCTAAGTCGACCGATTG CAAGTTTAATTTGAAGCTTTTAACTGTAGTAGATTGAACAACTTCGGCTGGTAGACTATTCCATCCATCGACAACTCTATTGCTGA AAAAATATTGCCTTTTCAAACAATTTCTGACCAATTCAGGTTTCATTCGATGTTTTGCTCCTCTAATGTTAGCTGCAGGACCTGAT GAGCAAATAGAGTTCATTGTTTGATTTTGTTAATATCAATGATTTCAATACCTTTAATATTTTGAATTGTGGATAAATCGCC...
			# 25	AGAAGCTTGAATTCGAGCAGTCAGCATTACGACGATGTGACAGAGGTTGAAGGTTGGCTGCAAGAGCAGGTCCAACTATATTTACA ATGTGTTTTTGCACCTTGTCTAAAAGAGAAAGGGCATCATTAGAAGATCCACCCCAGATATGACAACAGTATTCCATACAAGGCCG GATTTAGAAGCTTGAATTCGAGCAGA
			# 26	...TTGAATTCGAGCAGTCAGCATTACNACGATGTGCACAGAGGTTGAAGGTTGGCTCGCAAGCAGCAGGTCCAACTATATTTACA ATGTGTTTTTGCACCTTGTCTAAAAGAGAAAGGGCATCATTAGAAGATCCACCCCAGATATGACAACAGTATTCCATACAAGGCCG GATTTACAAGCTTGAATTCGAGCAGA
			# 57	CAAGCTTCTAGAAGCTTGAATTCGAGCAGTCAGTGTTAACAAAAACAAGCTAGACTCAAAGTGTAACAGGCTCATGGTGTCCGTTG GCTTCCTGTTTAGTTTGGTACATAGCATTCCCCTCGAGTCCTTTGACATTGTGAGATCTAGGAAGTCAATCGTATATGGCGAACTT CTGAAGGTAAATTTTATGCTGTTGTGCAGCCCATTCATGAAAATTNGAATTCTCTCGAGCTCTTCCGCCGTCCCATTGAAAATAGA AAAGAGATCATCAATGAATCTTTTATAAAACTTTATATTGTTCTGGGAAGTTGTTNGTAATTTCAGTGTCCAGAATTCCAGAANCT TGAATTCGAGCAGA
			# 58	AGAAGCTTGAATTCGAGCAGTCAGGCCAAGGACCATAATGTAAAAGTTTTGGTGAGATCAGTTCATGATCTGAGTCATGTTAAGGA AACAAGGAGAACTTAAACACAAGGTAAGGTCAAAAACAAGACACAAATAGAAGCTTGAATTCGAGCAG
			# 59	AGAAGCTTGAATTCGAGCAGTCAGGCATTTGATACTGTAAACCACAAGATTCTTATCTACAAACTACGTAGTTACGGAGTAGTTCA CTCTAATTTAAAGTGGCTTCAATGTTATTTAGAAAACCGTAAACAAGGTGTACCATATGATCTAACATGCTCCCCATTTGAATCAA TAAGTTGCGGAGTTCCACAAGGATCAATACTTGGACCTTTATTGTTTTTAAGTTTATATCAATGATATTTTTTTGTCTTCAAAAAT ACTTAATTATACTATTTTTGCTGATGACACTAATGTTTTCAGAAGCTTGAATTCGAGCAGA
α-*Hy*EZH2	6	1 / 6	# 70	GGGGAAAAAANCTAGATGAATAAAAGTTTTTAGAGCATGCAGGGACAGATACAGTAAACGAATGAGACTTAGCTGAATGACGAGTA AAACGAGAATGAATTTTAGTTGATGGAACTAGAGATGATAACTCCTTTGAGCAGCGACCATGATAGTATTTGTAGCAAAAGAGAAA GAGATGCAACTTTACGACGATGGGAAAGAGGTTCAAGCTTAGCAGATAGTCCAGGTCCAACAACATTTACAATGCATTTTTGCAGA GAAGAGAAAGTGAATCGTTAGAAGAACCAGCCCAAATATGACAACAGTATTCCATTCAGGGACGAATAAGAGATTTGTAGAGGTAG AGAGTGGAATCAGAAGTAAGGAAATGGCGAGCACGATAAAGAGAAGCCACCTTAGCAGAAGCTATTTACATCGTTATTATATG
α-*Hy*YY1	6	0 / 6	-	-
α-H3K27me3	6	5 / 6	# 2	CTCGAATTCAAGCTTCTAGGATTGAAAACGAAATCCAGCGGGAGCGACAAAACGCCAACCCGAATTTAGATAGCATCAAAGAAAAA CGCGATGCTCTTTTCTTGCTCCATTTCCAAAGAGCGTTTATTCGCACGAAGCANAAACTAGTTGAAGAAGGAGAAAAACCTTCAAA ATTTTTGTACAGTTTGGAAAAAACTCAACAGAGGAACAAGTCAATTACCAAAGTTCGCAAGAACGACGGAACATTGACGAAGTGAA CCTGGTGAGATTTTAAACTCCCTTGTTTCTTATTACAAAAACATCTACACCAAAACGAATTTATGCANAGACAGCCAAACTATGTC TTGTCACCATCACTAAACATTTGAGTGATGATGAAACCAATTATTAAACACCCGCCTAACCGGCGCGGAGTTAAAAGAACCCTTTT AAATTTGAAACGGAAATCCCGGCTATGCGACTTCCGCTGATTTATAAACTTTTG
			# 8	GAAGCTTGAATTCGAGCAGTCNGAAAAGTATATCAGCTTTTCTAAAGAACTTAAAGTAAATGAATTTACGAATAAAGAAAATAAAG TTATTGAAGTTAAACTACAACTACGTTTTCTTGATAGTTTTAGATTTATGGCTTCTAGTTTAGATAGTTTAACTAAAAATCTAACA AAGGAGCAATGTAAGAATGTTAGTAGATACTATTCAGAAGCTTGAATTCGAGCAGA
			# 12	AGAAGCTGAATTCGAGCAGTCAGCGTTGGTTTTGCATCGGAAAAAAGAGCCGACGTTTGCGATGGTTTTACATACGTTTGCCCAAT GAAAGTGTGCTAGCTGGGATGATGAGTGTATTTTTCTTATAAAATAGAAAAAAAAGTCAAAAATTACATGNTTTACTTGGTTTTGA GCAAAATGTGGTTACATTACCACACGCAGGGTAACAAAACAATGTTTGCTAAAAAGACGGTTAATCTCTATAGAAGCTTGAATTCG AGCAG
			# 13	CTGCTCGAATTCAAGCTTCTCTGACTGCTCGAATTCAAGCTTCTAAATCACAATGAATTTTAAGTGATAAAGCTTGTTCTGCAGAC ATCATAGTTGGCTGATGTTTGTCCTTTTTTTTTTGTTTTGCTTGTTTGAACTTGAGAAAACATTCCCAAATCCAATTTTTTAGGACC TGACTGCTCGAATTCAAGCTTCTAAAACAGGGATAACAGATGCCGCTTTCCAGCATGCTGGAAAACAAGACTCTGATAAGCACTTA ATAAATAATTTTGAAAGTATAGATGACAGCTCGGGAGAAAACCTCTGCAAGACAATAACAGGTATGCTGTCTGGACCACAAGCTGT AGAAGAGTCTAAGGGGATTCCATAAATGTCGCACTCTGACTCTCGAATTCAAGCTTCT
			# 17	TTACCAGCTGCCTGTATCGGACCGAACAATCCCACGGACGTGCTGTTGCACCCGGATTGTTTTCTCACCGCCCTGCACCGAAACCG CTGATTTCACTAACGCAACCATAAGAAAGGACTCCTGTTTGCTACATCTCATATTGACTACCAATGTACAGAACTGGCGAAGTGTC TTGTTCAGTGATAAATTGAAGTTCACTATCATTGGGAGCGATGGCATTTGCCGTGTACGTCGACCGGCTGGAAAATGCCTCGATTT ACGTTACTTGTCATACCACCGTGAAAGCATGGTCGGAGGCAATTGTAATGGTCTGGGGCTTCTGAAAGATGTTATGTTCCTATGTC GAATGGAATATGCCAGTAAAATGGATTTTCACCACAAAGCTTGATTCCAGCAAATCATAGTGCGGCCGCTTGCCGCC

TABLE 4.8 *Overview of the precipitated sequences by the α-HyPC, α-HyEZH2, α-HyYY1 and α-H3K27me3 antibody.*

This initial experiment shows that the generated antibodies against the *Hydra* PcG proteins are suitable for chromatin immunoprecipitation experiments. Sequencing of the precipitated DNA revealed that indeed *Hydra* chromatin was precipitated. *Hydra* chromatin could be precipitated with the α-*Hy*PC and α-*Hy*EZH2 antibodies. In addition, chromatin could be precipitated with the commercial α-H3K27me3 antibody. Most of the precipitated sequences are found within repetitive elements in the *Hydra* genome. Unfortunatelly, none of the precipitated sequences with the α-*Hy*YY1 antibody can be found in the genome. With this small-scale ChIP-Seq experiment no PcG target genes could be identified in *Hydra*, yet. However, the outcome of the initial experiment is promising.

For a high throughput ChIP-Seq experiment, the DNA fragments derived from ChIP are sequenced by next-generation methods. Protein-DNA interaction sites can be profiled by massively parallel sequencing. The resulting sequences are mapped to the reference genome and peaks at specific genomic regions are obtained with the most abundant sequenced fragments.

There are several steps, which have to be improved for high-throughput ChIP-Seq experiments. First of all, the shearing of the chromatin has to be improved. For ChIP-Seq, the size of the chromatin is important. Ideally, the fragments should have a length of about 200 bp for a good resolution. The sheared *Hydra* chromatin used for precipitation was most abundant at 600 - 700 bp. This is too large and the sonication settings need to be optimized. Second, an appropriate background control has to be included. For example, a mock IP without any antibody or an IP with an unspecific antibody can be performed as a background control.

However, only a sensitive antibody will enrich specific sequences compared to the background control. Whether the generated antibodies against the *Hydra* PcG proteins are sensitive enough, could not be determined.

4.5.3. SUMMARY

YY1 and the targeting of Polycomb Repressive Complexes

The *Hydra* homolog of the DNA binding protein YY1 was of special interest during this thesis. The *Drosophila* homolog Pho is the only DNA binding protein that is found in a PcG complex. Pho is a member of the third major Polycomb repressive complex PhoRC, which recruits the other Polycomb repressive complexes to their target genes via the interaction with a so-called Polycomb response element (PRE). In *Drosophila*, PcG target genes could be mapped to the genome by the identification of PRE sites. However, the existence of the PhoRC complex and PRE sites in species other than *Drosophila* is still unclear. There is emerging evidence that PcG proteins can be recruited to their target genes in a PRE-independent manner by long non-coding RNAs.
Preliminary results of co-immunoprecipitation studies at least suggest a direct interaction between the *Hy*YY1 protein and a protein of the *Hydra* PRC2 complex. However, unlike the genes of the *Hydra* PRC2 complex, the expression of *HyYY1* is not limited to the interstitial cell lineage. The gene has been found to be mainly expressed in ectodermal epithelial cells. Like *Hyp55*, *HyYY1* is additionally expressed in cells of the future oocyte. The broader expression of *Hy*YY1 can be explained by the more general function of YY1. YY1 does not solely act as a PcG protein. Whether *Hy*YY1 really acts as a molecular adapter in linking the Polycomb repressive complexes to their target DNA in *Hydra* has yet to be shown. For the future, it is planned to identify PcG target genes in *Hydra* by an approach based on ChIP-Seq. ChIP-Seq experiments are made possible, since antibodies against some of the *Hydra* PcG proteins are available now. During this thesis only an initial small-scale ChIP-Seq experiment with some of the generated antibodies could be performed. This initial experiment should be seen as 'proof of principle'. More work needs to be accomplished and large-scale ChIP-Seq experiments need to be performed to identify the *Hydra* PcG target genes.

ChIP-based identification of Polycomb Group target genes

In 2006 the first genome-wide analyses of PcG binding patterns by ChIP-on-chip have been reported in *Drosophila* (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). In the same year, first ChIP-on-chip analyses have been carried out in murine and human embryonic stem cells and human embryonic fibroblasts, respectively (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). Binding patterns for components of the PRC1 and PRC2 complexes and the H3K27me3 mark have been analyzed. A strong coincidence for the binding patterns could be observed. Among the bound genes are many genes important for differentiation and development (see figure 4.53).
Components of all major signaling pathways that are regulated by PcG-mediated silencing mechanisms in higher animals have been found in *Hydra*. Homologs of almost all vertebrate genes involved in developmental processes and pattern formation could be

found in the freshwater polyp, including genes of the wnt-, TGFβ, RTK-, hedgehog- and notch/delta-signalling pathway (Hobmayer et al., 2000, 2001; Miller and Steele, 2000; Käsbauer et al., 2007; Münder et al., 2010; Kaloulis, 2000). In contrast, most of the *HOX* genes are absent or secondarily lost (Garcia-Fernàndez, 2005; Steele et al., 2010).

Drosophila target gene	*Human target gene*
armadillo	*β-catenin*
ventral veins lacking	*POU3F1/Oct6*
eyes absent	*EYA2*
pannier	*GATA4*
disco	*BNC1*
cousin of atonal	*ATOH1*
ladybird early	*LBX1*
pdp1	*HLF*
caudal	*CDX2*
Rx	*RAX*
B-H1	*BARHL1, BARHL2*
H15	*TBX20*
vestigial	*VGLL1, VGLL2*
dachshund	*Dachshund homolog 1*
engrailed	*EN1, EN2*
even skipped	*Even-skipped homeobox1*
eyes absent	*Eyes absent*
giant	*Hepatic leukemia factor*
gooseberry	*Paired box gene 3*
hairy	*HES1, HES2*
knirps	*Estrogen receptor 1*
tailless	*NR2E1*
wingless	*WNT1*
homothorax	*MEIS2*
Ubx	*HOXA7, HOXB7*
abd-A	*HOXB-D8*
abd-B	*HOXA-D9*

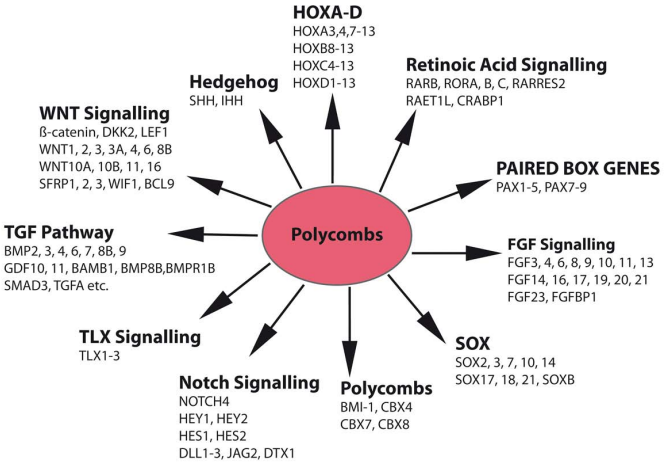


FIGURE 4.53 *Polycomb Group target genes detected in human embryonic fibroblast.* ***Left:*** *Polycomb Group target genes are conserved in Drosophila and humans. Confirmed or predictet Drosophila target genes are in accordance with the identified human targets.* ***Right:*** *PcG target genes which are involved in major signaling pathways are depicted. The figure has been adapted from Bracken et al. (2006).*

Bibliography

Agger K, **Cloos P A C**, **Christensen J**, **Pasini D**, **Rose S**, **Rappsilber J**, **Issaeva I**, **Canaani E**, **Salcini A E** and **Helin K** (2007): UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449 (7163): 731–734.

Alkema M J, **van der Lugt N M**, **Bobeldijk R C**, **Berns A** and **van Lohuizen M** (1995): Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. *Nature* 374 (6524): 724–727.

Allfrey V G, **Faulkner R** and **Mirsky A E** (1964): Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A* 51: 786–794.

Allis C D, **Jenuwein T** and **Reinberg D** (editors), *Epigenetics* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2006).

Americo J, **Whiteley M**, **Brown J L**, **Fujioka M**, **Jaynes J B** and **Kassis J A** (2002): A complex array of DNA-binding proteins required for pairing-sensitive silencing by a polycomb group response element from the Drosophila engrailed gene. *Genetics* 160 (4): 1561–1571.

Atchison L, **Ghias A**, **Wilkinson F**, **Bonini N** and **Atchison M L** (2003): Transcription factor YY1 functions as a PcG protein in vivo. *EMBO J* 22 (6): 1347–1358.

Austen M, **Lüscher B** and **Lüscher-Firzlaff J M** (1997): Characterization of the transcriptional regulator YY1. The bipartite transactivation domain is independent of interaction with the TATA box-binding protein, transcription factor IIB, TAFII55, or cAMP-responsive element-binding protein (CPB)-binding protein. *J Biol Chem* 272 (3): 1709–1717.

Ball E E, **Hayward D C**, **Saint R** and **Miller D J** (2004): A simple plan–cnidarians and the origins of developmental mechanisms. *Nat Rev Genet* 5 (8): 567–577.

Bantignies F, **Roure V**, **Comet I**, **Leblanc B**, **Schuettengruber B**, **Bonnet J**, **Tixier V**, **Mas A** and **Cavalli G** (2011): Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell* 144 (2): 214–226.

Barak O, **Lazzaro M A**, **Lane W S**, **Speicher D W**, **Picketts D J** and **Shiekhattar R** (2003): Isolation of human NURF: a regulator of Engrailed gene expression. *EMBO J* 22 (22): 6089–6100.

Basu A and **Atchison M L** (2010): CtBP levels control intergenic transcripts, PHO/YY1 DNA binding, and PcG recruitment to DNA. *J Cell Biochem* .

Bernstein B E, **Mikkelsen T S**, **Xie X**, **Kamal M**, **Huebert D J**, **Cuff J**, **Fry B**, **Meissner A**, **Wernig M**, **Plath K**, **Jaenisch R**, **Wagschal A**, **Feil R**, **Schreiber S L** and **Lander E S** (2006): A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125 (2): 315–326.

Berriman M, **Haas B J**, **LoVerde P T**, **Wilson R A**, **Dillon G P**, **Cerqueira G C**, **Mashiyama S T**, **Al-Lazikani B**, **Andrade L F**, **Ashton P D**, **Aslett M A**, **Bartholomeu D C**, **Blandin G**, **Caffrey C R**, **Coghlan A**, **Coulson R**, **Day T A**, **Delcher A**, **DeMarco R**, **Djikeng A**, **Eyre T**, **Gamble J A**, **Ghedin E**, **Gu Y**, **Hertz-Fowler C**, **Hirai H**, **Hirai Y**, **Houston R**, **Ivens A**, **Johnston D A**, **Lacerda D**, **Macedo C D**, **McVeigh P**, **Ning Z**, **Oliveira G**, **Overington J P**, **Parkhill J**, **Pertea M**, **Pierce R J**, **Protasio A V**, **Quail M A**, **Rajandream M A**, **Rogers J**, **Sajid M**, **Salzberg S L**, **Stanke M**, **Tivey A R**, **White O**, **Williams D L**, **Wortman J**, **Wu W**, **Zamanian M**, **Zerlotini A**, **Fraser-Liggett C M**, **Barrell B G** and **El-Sayed N M** (2009): The genome of the blood fluke Schistosoma mansoni. *Nature* 460 (7253): 352–358.

Beuchle D, **Struhl G** and **Müller J** (2001): Polycomb group proteins and heritable silencing of Drosophila Hox genes. *Development* 128 (6): 993–1004.

Bird A (2002): DNA methylation patterns and epigenetic memory. *Genes Dev* 16 (1): 6–21.

Birve A, **Sengupta A K**, **Beuchle D**, **Larsson J**, **Kennison J A**, **A R L** and **Müller J** (2001): Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. *Development* 128 (17): 3371–3379.

Blastyák A, **Mishra R K**, **Karch F** and **Gyurkovics H** (2006): Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. *Mol Cell Biol* 26 (4): 1434–1444.

Bloyer S, **Cavalli G**, **Brock H W** and **Dura J M** (2003): Identification and characterization of polyhomeotic PREs and TREs. *Dev Biol* 261 (2): 426–442.

Bode H R (2003): Head regeneration in Hydra. *Dev Dyn* 226 (2): 225–236.

Bode H R (2009): Axial patterning in hydra. *Cold Spring Harb Perspect Biol* 1 (1): a000 463.

Bode H R and **David C N** (1978): Regulation of a multipotent stem cell, the interstitial cell of hydra. *Prog Biophys Mol Biol* 33 (2): 189–206.

Bode H R and **Flick K M** (1976): Distribution and dynamics of nematocyte populations in Hydra attenuata. *J Cell Sci* 21 (1): 15–34.

Bode H R, **Flick K M** and **Smith G S** (1976): Regulation of interstitial cell differentiation in Hydra attenuata. I. Homeostatic control of interstitial cell population size. *J Cell Sci* 20 (1): 29–46.

Bode P M, **Awad T A**, **Koizumi O**, **Nakashima Y**, **Grimmelikhuijzen C J** and **Bode H R** (1988): Development of the two-part pattern during regeneration of the head in hydra. *Development* 102 (1): 223–235.

Bonifacino J S, **Dell'Angelica E C** and **Springer T A** (2001): Immunoprecipitation. *Curr Protoc Protein Sci* Chapter 9: Unit 9.8.

Bornemann D, **Miller E** and **Simon J** (1996): The Drosophila Polycomb group gene Sex comb on midleg (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development* 122 (5): 1621–1630.

Bosch T C and **David C N** (1990): Cloned interstitial stem cells grow as contiguous patches in hydra. *Dev Biol* 138 (2): 513–515.

Bosch T C G (2007): Why polyps regenerate and we don't: towards a cellular and molecular framework for Hydra regeneration. *Dev Biol* 303 (2): 421–433.

Bosch T C G, **Anton-Erxleben F**, **Hemmrich G** and **Khalturin K** (2010): The Hydra polyp: nothing but an active stem cell community. *Dev Growth Differ* 52 (1): 15–25.

Bosch T C G and **David C N** (1986): Male and female stem cells and sex reversal in Hydra polyps. *Proc Natl Acad Sci U S A* 83 (24): 9478–9482.

Bosch T C G and **David C N** (1987): Stem cells of Hydra magnipapillata can differentiate into somatic cells and germ line cells. *Dev Biol* 121: 182–191.

Boyer L A, **Langer M R**, **Crowley K A**, **Tan S**, **Denu J M** and **Peterson C L** (2002): Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol Cell* 10 (4): 935–942.

Boyer L A, **Latek R R** and **Peterson C L** (2004): The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol* 5 (2): 158–163.

Boyer L A, **Plath K**, **Zeitlinger J**, **Brambrink T**, **Medeiros L A**, **Lee T I**, **Levine S S**, **Wernig M**, **Tajonar A**, **Ray M K**, **Bell G W**, **Otte A P**, **Vidal M**, **Gifford D K**, **Young R A** and **Jaenisch R** (2006): Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441 (7091): 349–353.

Bracken A P, **Dietrich N**, **Pasini D**, **Hansen K H** and **Helin K** (2006): Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* 20 (9): 1123–1136.

Bradford M M (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254.

Breiling A, **Turner B M**, **Bianchi M E** and **Orlando V** (2001): General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* 412 (6847): 651–655.

Brown J L, **Fritsch C**, **Mueller J** and **Kassis J A** (2003): The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development* 130 (2): 285–294.

Brown J L, **Mucci D**, **Whiteley M**, **Dirksen M L** and **Kassis J A** (1998): The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell* 1 (7): 1057–1064.

Browne E N (1909): The production of new hydranths in hydra by the insertion of small grafts. *The Journal of Experimental Zoology* 8: 1–23.

Brownell J E, **Zhou J**, **Ranalli T**, **Kobayashi R**, **Edmondson D G**, **Roth S Y** and **Allis C D** (1996): Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84 (6): 843–851.

Brunk B P, **Martin E C** and **Adler P N** (1991): Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. *Nature* 353 (6342): 351–353.

Bunker C A and **Kingston R E** (1994): Transcriptional repression by Drosophila and mammalian Polycomb group proteins in transfected mammalian cells. *Mol Cell Biol* 14 (3): 1721–1732.

Burnette W N (1981): "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate–polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem* 112 (2): 195–203.

Busturia A, **Wightman C D** and **Sakonju S** (1997): A silencer is required for maintenance of transcriptional repression throughout Drosophila development. *Development* 124 (21): 4343–4350.

Campbell R D (1967): Tissue dynamics of steady state growth in Hydra littoralis. II. Patterns of tissue movement. *J Morphol* 121 (1): 19–28.

Campbell R D (1976): Elimination by Hydra interstitial and nerve cells by means of colchicine. *J Cell Sci* 21 (1): 1–13.

Campbell R D (1989): Taxonomy of the European Hydra (Gnidaria: Hydrozoa): a re-examination of its history with emphasis on the species H. vulgaris Pallas, H. attenuata Pallas and H. circumcincta Schulze. *Zoological Journal of the Linnean Society* 95: 219–244.

Campbell R D and **David C N** (1974): Cell cycle kinetics and development of Hydra attenuata. II. Interstitial cells. *J Cell Sci* 16 (2): 349–358.

Cao R, **Tsukada Y I** and **Zhang Y** (2005): Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* 20 (6): 845–854.

Cao R, **Wang L**, **Wang H**, **Xia L**, **Erdjument-Bromage H**, **Tempst P**, **Jones R S** and **Zhang Y** (2002): Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298 (5595): 1039–1043.

Caretti G, **Padova M D**, **Micales B**, **Lyons G E** and **Sartorelli V** (2004): The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev* 18 (21): 2627–2638.

Cattanach B M and **Kirk M** (1985): Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 315 (6019): 496–498.

Cavalli G, **Orlando V** and **Paro R** (1999): Mapping DNA target sites of chromatin-associated proteins by formaldehyde cross-linking in Drosophila embryos. *Chromosome Structural Analysis: A Practical Approach* : 20–37.

Cavalli G and **Paro R** (1998): The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93 (4): 505–518.

Cavalli G and **Paro R** (1999): Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* 286 (5441): 955–958.

Chapman G B and **Tilney L G** (1959a): Cytological studies of the nematocysts of Hydra. I. Desmonemes, isorhizas, cnidocils, and supporting structures. *J Biophys Biochem Cytol* 5 (1): 69–78.

Chapman G B and **Tilney L G** (1959b): Cytological studies of the nematocysts of Hydra. II. The stenoteles. *J Biophys Biochem Cytol* 5 (1): 79–84.

Chapman J A, **Kirkness E F**, **Simakov O**, **Hampson S E**, **Mitros T**, **Weinmaier T**, **Rattei T**, **Balasubramanian P G**, **Borman J**, **Busam D**, **Disbennett K**, **Pfannkoch C**, **Sumin N**, **Sutton G G**, **Viswanathan L D**, **Walenz B**, **Goodstein D M**, **Hellsten U**, **Kawashima T**, **Prochnik S E**, **Putnam N H**, **Shu S**, **Blumberg B**, **Dana C E**, **Gee L**, **Kibler D F**, **Law L**, **Lindgens D**, **Martinez D E**, **Peng J**, **Wigge P A**, **Bertulat B**, **Guder C**, **Nakamura Y**, **Ozbek S**, **Watanabe H**, **Khalturin K**, **Hemmrich G**, **Franke A**, **Augustin R**, **Fraune S**, **Hayakawa E**, **Hayakawa S**, **Hirose M**, **Hwang J S**, **Ikeo K**, **Nishimiya-Fujisawa C**, **Ogura A**, **Takahashi T**, **Steinmetz P R H**, **Zhang X**, **Aufschnaiter R**, **Eder M K**, **Gorny A K**, **Salvenmoser W**, **Heimberg A M**, **Wheeler B M**, **Peterson K J**, **Böttger A**, **Tischler P**, **Wolf A**, **Gojobori T**, **Remington K A**, **Strausberg R L**, **Venter J C**, **Technau U**, **Hobmayer B**, **Bosch T C G**, **Holstein T W**, **Fujisawa T**, **Bode H R**, **David C N**, **Rokhsar D S** and **Steele R E** (2010): The dynamic genome of Hydra. *Nature* 464 (7288): 592–596.

Chen D, **Ma H**, **Hong H**, **Koh S S**, **Huang S**, **Schurter B T**, **Aswad D W** and **Stallcup M R** (1999): Regulation of transcription by a protein methyltransferase. *Science* 284: 2174–2177.

Chen J Y, **Oliveri P**, **Gao F**, **Dornbos S Q**, **Li C W**, **Bottjer D J** and **Davidson E H** (2002): Precambrian animal life: probable developmental and adult cnidarian forms from Southwest China. *Dev Biol* 248 (1): 182–196.

Chomczynski P and **Sacchi N** (1987): Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162 (1): 156–159.

Ciccarone V C, **Polayes D A** and **Luckow V A**, *Molecular Diagnosis of Infectious Diseases: Generation of Recombinant Baculovirus DNA in E.coli Using a Baculovirus Shuttle Vector*, vol. 13 of *Methods in Molecular Medicine* (Humana Press, Totowa, NJ 1997).

Clapier C R and **Cairns B R** (2009): The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78: 273–304.

Cummings S G and **Bode H R** (1984): Head regeneration and polarity reversal in Hydra attenuata can occur in the absence of DNA synthesis. *Roux's Arch Dev Biol* 194: 79–86.

Czermin B, **Melfi R**, **McCabe D**, **Seitz V**, **Imhof A** and **Pirrotta V** (2002): Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 111 (2): 185–196.

David C N (1973): A quantitative method for maceration of hydra tissue. *Wilhelm Roux' Archiv* 171: 259–268.

David C N and **Campbell R D** (1972): Cell cycle kinetics and development of Hydra attenuata. I. Epithelial cells. *J Cell Sci* 11 (2): 557–568.

David C N and **Gierer A** (1974): Cell cycle kinetics and development of Hydra attenuata. III. Nerve and nematocyte differentiation. *J Cell Sci* 16 (2): 359–375.

Davis L E (1969): Differentiation of neurosensory cells in Hydra. *J Cell Sci* 5 (3): 699–726.

Davis L E (1971): Differentiation of ganglionic cells in Hydra. *J Exp Zool* 176 (1): 107–128.

Davis R E, **Hardwick C**, **Tavernier P**, **Hodgson S** and **Singh H** (1995): RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, Schistosoma mansoni. *J Biol Chem* 270 (37): 21 813–21 819.

de la Cruz X, **Lois S**, **Sánchez-Molina S** and **Martínez-Balbás M A** (2005): Do protein motifs read the histone code? *Bioessays* 27 (2): 164–175.

de Napoles M, **Mermoud J E**, **Wakao R**, **Tang Y A**, **Endoh M**, **Appanah R**, **Nesterova T B**, **Silva J**, **Otte A P**, **Vidal M**, **Koseki H** and **Brockdorff N** (2004): Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* 7 (5): 663–676.

Dellino G I, **Schwartz Y B**, **Farkas G**, **McCabe D**, **Elgin S C R** and **Pirrotta V** (2004): Polycomb silencing blocks transcription initiation. *Mol Cell* 13 (6): 887–893.

Deuring R, **Fanti L**, **Armstrong J A**, **Sarte M**, **Papoulas O**, **Prestel M**, **Daubresse G**, **Verardo M**, **Moseley S L**, **Berloco M**, **Tsukiyama T**, **Wu C**, **Pimpinelli S** and **Tamkun J W** (2000): The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* 5 (2): 355–365.

DeVido S K, **Kwon D**, **Brown J L** and **Kassis J A** (2008): The role of Polycomb-group response elements in regulation of engrailed transcription in Drosophila. *Development* 135 (4): 669–676.

Dignam J D, **Lebovitz R M** and **Roeder R G** (1983a): Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11 (5): 1475–1489.

Dignam J D, **Martin P L**, **Shastry B S** and **Roeder R G** (1983b): Eukaryotic gene transcription with purified components. *Methods Enzymol* 101: 582–598.

Déjardin J, **Rappailles A**, **Cuvier O**, **Grimaud C**, **Decoville M**, **Locker D** and **Cavalli G** (2005): Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. *Nature* 434 (7032): 533–538.

do Kim J and **Kim J** (2009): YY1's longer DNA-binding motifs. *Genomics* 93 (2): 152–158.

Dower W J, **Miller J F** and **Ragsdale C W** (1988): High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res* 16 (13): 6127–6145.

Dunne J F, **Javois L C**, **Huang L W** and **Bode H R** (1985): A subset of cells in the nerve net of Hydra oligactis defined by a monoclonal antibody: its arrangement and development. *Dev Biol* 109 (1): 41–53.

Eliassen K A, **Baldwin A**, **Sikorski E M** and **Hurt M M** (1998): Role for a YY1-binding element in replication-dependent mouse histone gene expression. *Mol Cell Biol* 18 (12): 7106–7118.

Engel U, **Ozbek S**, **Streitwolf-Engel R**, **Petri B**, **Lottspeich F**, **Holstein T W**, **Oezbek S** and **Engel R** (2002): Nowa, a novel protein with minicollagen Cys-rich domains, is involved in nematocyst formation in Hydra. *J Cell Sci* 115 (Pt 20): 3923–3934.

Engel U, **Pertz O**, **Fauser C**, **Engel J**, **David C N** and **Holstein T W** (2001): A switch in disulfide linkage during minicollagen assembly in Hydra nematocysts. *EMBO J* 20 (12): 3063–3073.

Fang J, **Chen T**, **Chadwick B**, **Li E** and **Zhang Y** (2004): Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. *J Biol Chem* 279 (51): 52 812–52 815.

Farkas G, **Gausz J**, **Galloni M**, **Reuter G**, **Gyurkovics H** and **Karch F** (1994): The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* 371 (6500): 806–808.

Fischle W, **Wang Y**, **Jacobs S A**, **Kim Y**, **Allis C D** and **Khorasanizadeh S** (2003): Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* 17: 1870–1881.

Francis N J, **Follmer N E**, **Simon M D**, **Aghia G** and **Butler J D** (2009): Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell* 137 (1): 110–122.

Francis N J, **Saurin A J**, **Shao Z** and **Kingston R E** (2001): Reconstitution of a functional core polycomb repressive complex. *Mol Cell* 8 (3): 545–556.

Franke A, **DeCamillis M**, **Zink D**, **Cheng N**, **Brock H W** and **Paro R** (1992): Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. *EMBO J* 11 (8): 2941–2950.

Fritsch C, **Brown J L**, **Kassis J A** and **Müller J** (1999): The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development* 126 (17): 3905–3913.

Fujisawa T (2003): Hydra regeneration and epitheliopeptides. *Dev Dyn* 226 (2): 182–189.

Furlong E E, **Rein T** and **Martin F** (1996): YY1 and NF1 both activate the human p53 promoter by alternatively binding to a composite element, and YY1 and E1A cooperate to amplify p53 promoter activity. *Mol Cell Biol* 16 (10): 5933–5945.

Galliot B, **Miljkovic-Licina M**, **Ghila L** and **Chera S** (2007): RNAi gene silencing affects cell and developmental plasticity in hydra. *C R Biol* 330 (6-7): 491–497.

Galliot B and **Miller D** (2000): Origin of anterior patterning. How old is our head? *Trends Genet* 16 (1): 1–5.

Garcia-Fernàndez J (2005): The genesis and evolution of homeobox gene clusters. *Nat Rev Genet* 6 (12): 881–892.

Gaston K and **Fried M** (1994): YY1 is involved in the regulation of the bi-directional promoter of the Surf-1 and Surf-2 genes. *FEBS Lett* 347 (2-3): 289–294.

Genikhovich G, **Kürn U**, **Hemmrich G** and **Bosch T C G** (2006): Discovery of genes expressed in Hydra embryogenesis. *Dev Biol* 289 (2): 466–481.

Gierer A, **Berking S**, **Bode H**, **David C N**, **Flick K**, **Hansmann G**, **Schaller H** and **Trenkner E** (1972): Regeneration of hydra from reaggregated cells. *Nat New Biol* 239 (91): 98–101.

Gierer A and **Meinhardt H** (1972): A theory of biological pattern formation. *Kybernetik* 12 (1): 30–39.

Goodrich J, **Puangsomlee P**, **Martin M**, **Long D**, **Meyerowitz E M** and **Coupland G** (1997): A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. *Nature* 386 (6620): 44–51.

Grens A, **Gee L**, **Fisher D A** and **Bode H R** (1996): CnNK-2, an NK-2 homeobox gene, has a role in patterning the basal end of the axis in hydra. *Dev Biol* 180 (2): 473–488.

Grimaud C, **Bantignies F**, **Pal-Bhadra M**, **Ghana P**, **Bhadra U** and **Cavalli G** (2006): RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* 124 (5): 957–971.

Grimm C, **de Ayala Alonso A G**, **Rybin V**, **Steuerwald U**, **Ly-Hartig N**, **Fischle W**, **Müller J** and **Müller C W** (2007): Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein Sex comb on midleg. *EMBO Rep* 8 (11): 1031–1037.

Grimm C, **Matos R**, **Ly-Hartig N**, **Steuerwald U**, **Lindner D**, **Rybin V**, **Müller J** and **Müller C W** (2009): Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J* 28 (13): 1965–1977.

Gualberto A, **LePage D**, **Pons G**, **Mader S L**, **Park K**, **Atchison M L** and **Walsh K** (1992): Functional antagonism between YY1 and the serum response factor. *Mol Cell Biol* 12 (9): 4209–4214.

Guo B, **Odgren P R**, **van Wijnen A J**, **Last T J**, **Nickerson J**, **Penman S**, **Lian J B**, **Stein J L** and **Stein G S** (1995): The nuclear matrix protein NMP-1 is the transcription factor YY1. *Proc Natl Acad Sci U S A* 92 (23): 10 526–10 530.

Guo J, **Casolaro V**, **Seto E**, **Yang W M**, **Chang C**, **Seminario M C**, **Keen J** and **Georas S N** (2001): Yin-Yang 1 activates interleukin-4 gene expression in T cells. *J Biol Chem* 276 (52): 48 871–48 878.

Hansen K H, **Bracken A P**, **Pasini D**, **Dietrich N**, **Gehani S S**, **Monrad A**, **Rappsilber J**, **Lerdrup M** and **Helin K** (2008): A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* 10 (11): 1291–1300.

Hansen K H and **Helin K** (2009): Epigenetic inheritance through self-recruitment of the polycomb repressive complex 2. *Epigenetics* 4 (3): 133–138.

Hariharan N, **Kelley D E** and **Perry R P** (1991): Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. *Proc Natl Acad Sci U S A* 88 (21): 9799–9803.

Harlow E and **Lane D** (editors), *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1988).

Hausmann K and **Holstein T** (1985): Bilateral symmetry in the cnidocil-nematocyst complex of the freshwater medusaCraspedacusta sowerbii Lankester (Hydrozoa, Limnomedusae). *Journal of Ultrastructure Research* 90 (1): 89–104.

Hayakawa E, **Takahashi T**, **Nishimiya-Fujisawa C** and **Fujisawa T** (2007): A novel neuropeptide (FRamide) family identified by a peptidomic approach in Hydra magnipapillata. *FEBS J* 274 (20): 5438–5448.

Hayashi T, **Fujita Y**, **Iwasaki O**, **Adachi Y**, **Takahashi K** and **Yanagida M** (2004): Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* 118 (6): 715–729.

Haynes J and **Burnett A L** (1963): Dedifferentiation and redifferentiation of cells in Hydra viridis. *Science* 142: 1481–1483.

Haynes J F, **Burnett A L** and **Davis L E** (1968): Histological and ultrastructural study of the muscular and nervous systems in Hydra. I. The muscular system and the mesoglea. *J Exp Zool* 167 (3): 283–293.

He Y and **Casaccia-Bonnefil P** (2008): The Yin and Yang of YY1 in the nervous system. *J Neurochem* 106 (4): 1493–1502.

Heimfeld S and **Bode H R** (1985): Growth regulation of the interstitial cell population in hydra. I. Evidence for global control by nerve cells in the head. *Dev Biol* 110 (2): 297–307.

Hennig L and **Derkacheva M** (2009): Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet* 25 (9): 414–423.

Herbel C (2009): *Identification of Polycomb Repressive Complex 2 target genes in Hydra magnipapillata*. Master's thesis, Faculty for Biosciences of the Ruperto-Carola University of Heidelberg.

Hernández-Muñoz I, **Taghavi P**, **Kuijl C**, **Neefjes J** and **van Lohuizen M** (2005): Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol Cell Biol* 25 (24): 11 047–11 058.

Ho L and **Crabtree G R** (2010): Chromatin remodelling during development. *Nature* 463 (7280): 474–484.

Hobmayer B, **Rentzsch F** and **Holstein T W** (2001): Identification and expression of HySmad1, a member of the R-Smad family of TGFbeta signal transducers, in the diploblastic metazoan Hydra. *Dev Genes Evol* 211 (12): 597–602.

Hobmayer B, **Rentzsch F**, **Kuhn K**, **Happel C M**, **von Laue C C**, **Snyder P**, **Rothbächer U** and **Holstein T W** (2000): WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. *Nature* 407 (6801): 186–189.

Hobmayer E, **Holstein T W** and **David C N** (1990): Tentacle morphogenesis in hydra. II. Formation of a complex between a sensory nerve cell and a battery cell. *Development* 109: 897–904.

Holliday R and **Pugh J E** (1975): DNA modification mechanisms and gene activity during development. *Science* 187 (4173): 226–232.

Holstein T (1981): The morphogenesis of nematocytes in Hydra and Forskålia: an ultrastructural study. *J Ultrastruct Res* 75 (3): 276–290.

Holstein T, *Cnidaria: Hydrozoa*, vol. 1 of *Süßwasserfauna von Mitteleuropa* (J. Schwoerbel and P. Zwick, Stuttgart, Jena, New York 1995).

Holstein T and **Tardent P** (1984): An ultrahigh-speed analysis of exocytosis: nematocyst discharge. *Science* 223 (4638): 830–833.

Holstein T W, **Hobmayer E** and **David C N** (1991): Pattern of epithelial cell cycling in hydra. *Dev Biol* 148 (2): 602–611.

Holstein T W, **Hobmayer E** and **Technau U** (2003): Cnidarians: an evolutionarily conserved model system for regeneration? *Dev Dyn* 226 (2): 257–267.

Hong S, **Cho Y W**, **Yu L R**, **Yu H**, **Veenstra T D** and **Ge K** (2007): Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc Natl Acad Sci U S A* 104 (47): 18 439–18 444.

Horard B, **Tatout C**, **Poux S** and **Pirrotta V** (2000): Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. *Mol Cell Biol* 20 (9): 3187–3197.

Houbaviy H B, **Usheva A**, **Shenk T** and **Burley S K** (1996): Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc Natl Acad Sci U S A* 93 (24): 13 577–13 582.

Huang D H and **Chang Y L** (2004): Isolation and characterization of CHRASCH, a polycomb-containing silencing complex. *Methods Enzymol* 377: 267–282.

Huang D H, **Chang Y L**, **Yang C C**, **Pan I C** and **King B** (2002): pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. *Mol Cell Biol* 22 (17): 6261–6271.

Hufnagel L A, **Kass-Simon G** and **Lyon M K** (1985): Functional Organization of Battery Cell Complexes in Tentacles of Hydra attenuata. *Journal of Morphology* 184: 323–341.

Hyde-DeRuyscher R P, **Jennings E** and **Shenk T** (1995): DNA binding sites for the transcriptional activator/repressor YY1. *Nucleic Acids Res* 23 (21): 4457–4465.

Ingham P W (1985): A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila. *J Embryol Exp Morphol* 89: 349–365.

Jacobs S A, **Taverna S D**, **Zhang Y**, **Briggs S D**, **Li J**, **Eissenberg J C**, **Allis C D** and **Khorasanizadeh S** (2001): Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J* 20 (18): 5232–5241.

Jenuwein T and **Allis C D** (2001): Translating the histone code. *Science* 293 (5532): 1074–1080.

Kaloulis K (2000): *Molecular Basis of Morphogenetic Events in Hydra: Study of the CREB and Hedgehog Pathways during Budding and Regeneration.* Ph.D. thesis, Faculty of Sciences, University of Geneva.

Kawaida H, **Shimizu H**, **Fujisawa T**, **Tachida H** and **Kobayakawa Y** (2010): Molecular phylogenetic study in genus Hydra. *Gene* .

Kennison J A and **Tamkun J W** (1988): Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proc Natl Acad Sci U S A* 85 (21): 8136–8140.

Khalturin K, **Anton-Erxleben F**, **Milde S**, **Plötz C**, **Wittlieb J**, **Hemmrich G** and **Bosch T C G** (2007): Transgenic stem cells in Hydra reveal an early evolutionary origin for key elements controlling self-renewal and differentiation. *Dev Biol* 309 (1): 32–44.

Kim J, **Kim W** and **Cunningham C W** (1999): A new perspective on lower metazoan relationships from 18S rDNA sequences. *Mol Biol Evol* 16 (3): 423–427.

Kim J D, **Faulk C** and **Kim J** (2007): Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Res* 35 (10): 3442–3452.

King N, **Westbrook M J**, **Young S L**, **Kuo A**, **Abedin M**, **Chapman J**, **Fairclough S**, **Hellsten U**, **Isogai Y**, **Letunic I**, **Marr M**, **Pincus D**, **Putnam N**, **Rokas A**, **Wright K J**, **Zuzow R**, **Dirks W**, **Good M**, **Goodstein D**, **Lemons D**, **Li W**, **Lyons J B**, **Morris A**, **Nichols S**, **Richter D J**, **Salamov A**, **Sequencing J G I**, **Bork P**, **Lim W A**, **Manning G**, **Miller W T**, **McGinnis W**, **Shapiro H**, **Tjian R**, **Grigoriev I V** and **Rokhsar D** (2008): The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451 (7180): 783–788.

Klymenko T, **Papp B**, **Fischle W**, **Köcher T**, **Schelder M**, **Fritsch C**, **Wild B**, **Wilm M** and **Müller J** (2006): A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* 20 (9): 1110–1122.

Kobrossy L, **Rastegar M** and **Featherstone M** (2006): Interplay between chromatin and transacting factors regulating the Hoxd4 promoter during neural differentiation. *J Biol Chem* 281 (36): 25 926–25 939.

Koizumi O and **Bode H R** (1986): Plasticity in the nervous system of adult hydra. I. The position-dependent expression of FMRFamide-like immunoreactivity. *Dev Biol* 116 (2): 407–421.

Koizumi O, **Heimfeld S** and **Bode H R** (1988): Plasticity in the nervous system of adult hydra. II. Conversion of ganglion cells of the body column into epidermal sensory cells of the hypostome. *Dev Biol* 129 (2): 358–371.

Koizumi O, **Mizumoto H**, **Sugiyama T** and **Bode H R** (1990): Nerve net formation in the primitive nervous system of Hydra–an overview. *Neurosci Res Suppl* 13: S165–S170.

Kornberg R D (1974): Chromatin structure: a repeating unit of histones and DNA. *Science* 184 (139): 868–871.

Käsbauer T, **Towb P**, **Alexandrova O**, **David C N**, **Dall'armi E**, **Staudigl A**, **Stiening B** and **Böttger A** (2007): The Notch signaling pathway in the cnidarian Hydra. *Dev Biol* 303 (1): 376–390.

Kurz E M, **Holstein T W**, **Petri B M**, **Engel J** and **David C N** (1991): Mini-collagens in hydra nematocytes. *J Cell Biol* 115 (4): 1159–1169.

Kuzmichev A, **Jenuwein T**, **Tempst P** and **Reinberg D** (2004): Different Ezh2-Containing Complexes Target Methylation of Histone H1 or Nucleosomal Histone H3. *Molecular Cell* 14: 183–193.

Kuzmichev A, **Margueron R**, **Vaquero A**, **Preissner T S**, **Scher M**, **Kirmizis A**, **Ouyang X**, **Brockdorff N**, **Abate-Shen C**, **Farnham P** and **Reinberg D** (2005): Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. *Proc Natl Acad Sci U S A* 102 (6): 1859–1864.

Kuzmichev A, **Nishioka K**, **Erdjument-Bromage H**, **Tempst P** and **Reinberg D** (2002): Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 16 (22): 2893–2905.

Laemmli U K (1970): Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227 (5259): 680–685.

Lan F, **Bayliss P E**, **Rinn J L**, **Whetstine J R**, **Wang J K**, **Chen S**, **Iwase S**, **Alpatov R**, **Issaeva I**, **Canaani E**, **Roberts T M**, **Chang H Y** and **Shi Y** (2007): A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* 449 (7163): 689–694.

Lauschke V (2009): *Functional characterization of the Polycomb repressive complex 2 (PRC2) in Hydra magnipapillata*. Master's thesis, Faculty for Biosciences of the Ruperto-Carola University of Heidelberg.

Lee J S, **See R H**, **Galvin K M**, **Wang J** and **Shi Y** (1995): Functional interactions between YY1 and adenovirus E1A. *Nucleic Acids Res* 23 (6): 925–931.

Lee M G, **Villa R**, **Trojer P**, **Norman J**, **Yan K P**, **Reinberg D**, **Croce L D** and **Shiekhattar R** (2007): Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* 318 (5849): 447–450.

Lee T C, **Zhang Y** and **Schwartz R J** (1994): Bifunctional transcriptional properties of YY1 in regulating muscle actin and c-myc gene expression during myogenesis. *Oncogene* 9 (4): 1047–1052.

Lee T I, **Jenner R G**, **Boyer L A**, **Guenther M G**, **Levine S S**, **Kumar R M**, **Chevalier B**, **Johnstone S E**, **Cole M F**, **ichi Isono K**, **Koseki H**, **Fuchikami T**, **Abe K**, **Murray H L**, **Zucker J P**, **Yuan B**, **Bell G W**, **Herbolsheimer E**, **Hannett N M**, **Sun K**, **Odom D T**, **Otte A P**, **Volkert T L**, **Bartel D P**, **Melton D A**, **Gifford D K**, **Jaenisch R** and **Young R A** (2006): Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125 (2): 301–313.

Lengfeld T, **Watanabe H**, **Simakov O**, **Lindgens D**, **Gee L**, **Law L**, **Schmidt H A**, **Ozbek S**, **Bode H** and **Holstein T W** (2009): Multiple Wnts are involved in Hydra organizer formation and regeneration. *Dev Biol* 330 (1): 186–199.

Lenhoff H (editor), *Hydra:Research Methods* (Springer 1983).

Letunic I, **Doerks T** and **Bork P** (2009): SMART 6: recent updates and new developments. *Nucleic Acids Res* 37 (Database issue): D229–D232.

Levine S S, **Weiss A**, **Erdjument-Bromage H**, **Shao Z**, **Tempst P** and **Kingston R E** (2002): The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* 22 (17): 6070–6078.

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

Li D and **Roberts R** (2001): WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell Mol Life Sci* 58 (14): 2085–2097.

Li H, **Ilin S**, **Wang W**, **Duncan E M**, **Wysocka J**, **Allis C D** and **Patel D J** (2006): Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442 (7098): 91–95.

Littlefield C L (1985): Germ cells in Hydra oligactis males. I. Isolation of a subpopulation of interstitial cells that is developmentally restricted to sperm production. *Dev Biol* 112 (1): 185–193.

Littlefield C L (1991): Cell lineages in Hydra: isolation and characterization of an interstitial stem cell restricted to egg production in Hydra oligactis. *Dev Biol* 143 (2): 378–388.

Liu Y, **Shao Z** and **Yuan G C** (2010): Prediction of Polycomb target genes in mouse embryonic stem cells. *Genomics* .

Lohmann J U, **Endl I** and **Bosch T C** (1999): Silencing of developmental genes in Hydra. *Dev Biol* 214 (1): 211–214.

Luger K, **Mäder A W**, **Richmond R K**, **Sargent D F** and **Richmond T J** (1997): Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389 (6648): 251–260.

Lyon M F (1961): Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* 190: 372–373.

MacWilliams H K (1982): Pattern Formation in Hydra Head Regeneration. *American Zoologist* 22: 17–26.

Mager J, **Montgomery N D**, **de Villena F P M** and **Magnuson T** (2003): Genome imprinting regulated by the mouse Polycomb group protein Eed. *Nat Genet* 33 (4): 502–507.

Mahmoudi T, **Zuijderduijn L M P**, **Mohd-Sarip A** and **Verrijzer C P** (2003): GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. *Nucleic Acids Res* 31 (14): 4147–4156.

Mandel M and **Higa A** (1970): Calcium-dependent bacteriophage DNA infection. *Journal of Molecular Biology* 53: 159–162.

Manley J L, **Fire A**, **Cano A**, **Sharp P A** and **Gefter M L** (1980): DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. *Proc Natl Acad Sci U S A* 77 (7): 3855–3859.

Manley J L, **Fire A**, **Samuels M** and **Sharp P A** (1983): In vitro transcription: whole-cell extract. *Methods Enzymol* 101: 568–582.

Margueron R, **Justin N**, **Ohno K**, **Sharpe M L**, **Son J**, **Drury W J**, **Voigt P**, **Martin S R**, **Taylor W R**, **Marco V D**, **Pirrotta V**, **Reinberg D** and **Gamblin S J** (2009): Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461 (7265): 762–767.

Martindale M Q (2005): The evolution of metazoan axial properties. *Nat Rev Genet* 6 (12): 917–927.

Martindale M Q, **Pang K** and **Finnerty J R** (2004): Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). *Development* 131 (10): 2463–2474.

Martinez D E, **Dirksen M L**, **Bode P M**, **Jamrich M**, **Steele R E** and **Bode H R** (1997): Budhead, a fork head/HNF-3 homologue, is expressed during axis formation and head specification in hydra. *Dev Biol* 192 (2): 523–536.

Martínez-Balbás M A, **Tsukiyama T**, **Gdula D** and **Wu C** (1998): Drosophila NURF-55, a WD repeat protein involved in histone metabolism. *Proc Natl Acad Sci U S A* 95 (1): 132–137.

Maurange C and **Paro R** (2002): A cellular memory module conveys epigenetic inheritance of hedgehog expression during Drosophila wing imaginal disc development. *Genes Dev* 16 (20): 2672–2683.

McClintock B (1950): The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* 36 (6): 344–355.

Medina M, **Collins A G**, **Silberman J D** and **Sogin M L** (2001): Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proc Natl Acad Sci U S A* 98 (17): 9707–9712.

Meinhardt H (1993): A model for pattern formation of hypostome, tentacles, and foot in hydra: how to form structures close to each other, how to form them at a distance. *Dev Biol* 157 (2): 321–333.

Meinhardt H (2009): Models for the generation and interpretation of gradients. *Cold Spring Harb Perspect Biol* 1 (4): a001 362.

Meinhardt H and **Gierer A** (1974): Applications of a theory of biological pattern formation based on lateral inhibition. *J Cell Sci* 15 (2): 321–346.

Messmer S, **Franke A** and **Paro R** (1992): Analysis of the functional role of the Polycomb chromo domain in Drosophila melanogaster. *Genes Dev* 6 (7): 1241–1254.

Mihaly J, **Mishra R K** and **Karch F** (1998): A conserved sequence motif in Polycomb-response elements. *Mol Cell* 1 (7): 1065–1066.

Mikkelsen T S, **Ku M**, **Jaffe D B**, **Issac B**, **Lieberman E**, **Giannoukos G**, **Alvarez P**, **Brockman W**, **Kim T K**, **Koche R P**, **Lee W**, **Mendenhall E**, **O'Donovan A**, **Presser A**, **Russ C**, **Xie X**, **Meissner A**, **Wernig M**, **Jaenisch R**, **Nusbaum C**, **Lander E S** and **Bernstein B E** (2007): Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448 (7153): 553–560.

Miller M A and **Steele R E** (2000): Lemon encodes an unusual receptor protein-tyrosine kinase expressed during gametogenesis in Hydra. *Dev Biol* 224 (2): 286–298.

Miller M A, **Technau U**, **Smith K M** and **Steele R E** (2000): Oocyte development in Hydra involves selection from competent precursor cells. *Dev Biol* 224 (2): 326–338.

Min J, **Zhang Y** and **Xu R M** (2003): Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* 17 (15): 1823–1828.

Mishra R K, **Mihaly J**, **Barges S**, **Spierer A**, **Karch F**, **Hagstrom K**, **Schweinsberg S E** and **Schedl P** (2001): The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol Cell Biol* 21 (4): 1311–1318.

Müller H J (1930): Types of visible variations induced by X-rays in Drosophila. *Journal of Genetics* 22: 299–334.

Müller J, **Gaunt S** and **Lawrence P A** (1995): Function of the Polycomb protein is conserved in mice and flies. *Development* 121 (9): 2847–2852.

Müller J, **Hart C M**, **Francis N J**, **Vargas M L**, **Sengupta A**, **Wild B**, **Miller E L**, **O'Connor M B**, **Kingston R E** and **Simon J A** (2002): Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111 (2): 197–208.

Münder S, **Käsbauer T**, **Prexl A**, **Aufschnaiter R**, **Zhang X**, **Towb P** and **Böttger A** (2010): Notch signalling defines critical boundary during budding in Hydra. *Dev Biol* 344 (1): 331–345.

Mohd-Sarip A, **Venturini F**, **Chalkley G E** and **Verrijzer C P** (2002): Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. *Mol Cell Biol* 22 (21): 7473–7483.

Mueller J F (1950): Some Observations on the Structure of Hydra, with Particular Reference to the Muscular System. *Transactions of the American Microscopical Society* 69 (2): 133–147.

Mulholland N M, **King I F G** and **Kingston R E** (2003): Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes Dev* 17 (22): 2741–2746.

Mundie J R (1926): Maceration of green hydra. *Science* 64 (1667): 579.

Nüchter T, **Benoit M**, **Engel U**, **Ozbek S** and **Holstein T W** (2006): Nanosecond-scale kinetics of nematocyst discharge. *Curr Biol* 16 (9): R316–R318.

Nègre N, **Hennetin J**, **Sun L V**, **Lavrov S**, **Bellis M**, **White K P** and **Cavalli G** (2006): Chromosomal distribution of PcG proteins during Drosophila development. *PLoS Biol* 4 (6): e170.

Nishimiya-Fujisawa C and **Sugiyama T** (1993): Genetic analysis of developmental mechanisms in hydra. XX. Cloning of interstitial stem cells restricted to the sperm differentiation pathway in Hydra magnipapillata. *Dev Biol* 157 (1): 1–9.

Noda K (1970): The Fate of Aggregates formed by Two Species of Hydra (Hydra magnipapillata and Pelmatohydra robusta). *Journal of the Faculty of Science Hokkaido University Series ??* 17 (3): 432–439.

Orlando V, **Jane E P**, **Chinwalla V**, **Harte P J** and **Paro R** (1998): Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. *EMBO J* 17 (17): 5141–5150.

Otto J J and **Campbell R D** (1977): Tissue economics of hydra: regulation of cell cycle, animal size and development by controlled feeding rates. *J Cell Sci* 28: 117–132.

Pal-Bhadra M, **Bhadra U** and **Birchler J A** (1997): Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. *Cell* 90 (3): 479–490.

Palko L, **Bass H W**, **Beyrouthy M J** and **Hurt M M** (2004): The Yin Yang-1 (YY1) protein undergoes a DNA-replication-associated switch in localization from the cytoplasm to the nucleus at the onset of S phase. *J Cell Sci* 117 (Pt 3): 465–476.

Pallas P S, *Elenchus zoophytorum sistens generum adumbrationes generaliores et specierum cognitarum succintas descriptiones, cum selectis auctorum synonymis* (Hagae-Comitum :Apud Petrum van Cleef 1766).

Paro R and **Harte P J**, *The role of Polycomb and Trithorax group chromatin complexes in the maintenance of determined cell states*. Epigenetic mechanisms of gene regulation (Cold Spring Harbor Laboratory Press 1996).

Paro R and **Hogness D S** (1991): The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc Natl Acad Sci U S A* 88 (1): 263–267.

Peterson A J, **Kyba M**, **Bornemann D**, **Morgan K**, **Brock H W** and **Simon J** (1997): A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Mol Cell Biol* 17 (11): 6683–6692.

Peterson A J, **Mallin D R**, **Francis N J**, **Ketel C S**, **Stamm J**, **Voeller R K**, **Kingston R E** and **Simon J A** (2004): Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression. *Genetics* 167 (3): 1225–1239.

Plath K, **Fang J**, **Mlynarczyk-Evans S K**, **Cao R**, **Worringer K A**, **Wang H**, **de la Cruz C C**, **Otte A P**, **Panning B** and **Zhang Y** (2003): Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300 (5616): 131–135.

Poux S, **McCabe D** and **Pirrotta V** (2001a): Recruitment of components of Polycomb Group chromatin complexes in Drosophila. *Development* 128 (1): 75–85.

Poux S, **Melfi R** and **Pirrotta V** (2001b): Establishment of Polycomb silencing requires a transient interaction between PC and ESC. *Genes Dev* 15 (19): 2509–2514.

Prusch R D, **Benos D J** and **Ritter M** (1976): Osmoregulatory control mechanisms in freshwater coelenterates. *Comp Biochem Physiol A Comp Physiol* 53 (2): 161–164.

Putnam N H, **Butts T**, **Ferrier D E K**, **Furlong R F**, **Hellsten U**, **Kawashima T**, **Robinson-Rechavi M**, **Shoguchi E**, **Terry A**, **Yu J K**, **Benito-Gutiérrez E L**, **Dubchak I**, **Garcia-Fernàndez J**, **Gibson-Brown J J**, **Grigoriev I V**, **Horton A C**, **de Jong P J**, **Jurka J**, **Kapitonov V V**, **Kohara Y**, **Kuroki Y**, **Lindquist E**, **Lucas S**, **Osoegawa K**, **Pennacchio L A**, **Salamov A A**, **Satou Y**, **Sauka-Spengler T**, **Schmutz J**, **Shin-I T**, **Toyoda A**, **Bronner-Fraser M**, **Fujiyama A**, **Holland L Z**, **Holland P W H**, **Satoh N** and **Rokhsar D S** (2008): The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453 (7198): 1064–1071.

Putnam N H, **Srivastava M**, **Hellsten U**, **Dirks B**, **Chapman J**, **Salamov A**, **Terry A**, **Shapiro H**, **Lindquist E**, **Kapitonov V V**, **Jurka J**, **Genikhovich G**, **Grigoriev I V**, **Lucas S M**, **Steele R E**, **Finnerty J R**, **Technau U**, **Martindale M Q** and **Rokhsar D S** (2007): Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317 (5834): 86–94.

Rastelli L, **Chan C S** and **Pirrotta V** (1993): Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function. *EMBO J* 12 (4): 1513–1522.

Rea S, **Eisenhaber F**, **O'Carroll D**, **Strahl B D**, **Sun Z W**, **Schmid M**, **Opravil S**, **Mechtler K**, **Ponting C P**, **Allis C D** and **Jenuwein T** (2000): Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406 (6796): 593–599.

Ren X, **Vincenz C** and **Kerppola T K** (2008): Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol Cell Biol* 28 (9): 2884–2895.

Riggs A D and **Porter T N**, *Historical Overview of epigenetic mechanisms*. Epigenetic mechanisms of gene regulation (Cold Spring Harbor Laboratory Press 1996).

Riggs K J, **Saleque S**, **Wong K K**, **Merrell K T**, **Lee J S**, **Shi Y** and **Calame K** (1993): Yin-yang 1 activates the c-myc promoter. *Mol Cell Biol* 13 (12): 7487–7495.

Ringrose L and **Paro R** (2001): Remembering silence. *Bioessays* 23 (7): 566–570.

Ringrose L and **Paro R** (2004): Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* 38: 413–443.

Ringrose L and **Paro R** (2007): Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 134 (2): 223–232.

Ringrose L, **Rehmsmeier M**, **Dura J M** and **Paro R** (2003): Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. *Dev Cell* 5 (5): 759–771.

Rinn J L, **Kertesz M**, **Wang J K**, **Squazzo S L**, **Xu X**, **Brugmann S A**, **Goodnough L H**, **Helms J A**, **Farnham P J**, **Segal E** and **Chang H Y** (2007): Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129 (7): 1311–1323.

Rizkallah R and **Hurt M M** (2009): Regulation of the transcription factor YY1 in mitosis through phosphorylation of its DNA-binding domain. *Mol Biol Cell* 20 (22): 4766–4776.

Rylski M, **Amborska R**, **Zybura K**, **Konopacki F A**, **Wilczynski G M** and **Kaczmarek L** (2008): Yin Yang 1 expression in the adult rodent brain. *Neurochem Res* 33 (12): 2556–2564.

Sanger F, **Nicklen S** and **Coulson A R** (1977): DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74 (12): 5463–5467.

Santa F D, **Totaro M G**, **Prosperini E**, **Notarbartolo S**, **Testa G** and **Natoli G** (2007): The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* 130 (6): 1083–1094.

Santiveri C M, **Lechtenberg B C**, **Allen M D**, **Sathyamurthy A**, **Jaulent A M**, **Freund S M V** and **Bycroft M** (2008): The malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine. *J Mol Biol* 382 (5): 1107–1112.

Sarras M P, **Madden M E**, **Zhang X M**, **Gunwar S**, **Huff J K** and **Hudson B G** (1991a): Extracellular matrix (mesoglea) of Hydra vulgaris. I. Isolation and characterization. *Dev Biol* 148 (2): 481–494.

Sarras M P, **Meador D** and **Zhang X M** (1991b): Extracellular matrix (mesoglea) of Hydra vulgaris. II. Influence of collagen and proteoglycan components on head regeneration. *Dev Biol* 148 (2): 495–500.

Sarras M P, **Zhang X**, **Huff J K**, **Accavitti M A**, **John P L S** and **Abrahamson D R** (1993): Extracellular matrix (mesoglea) of Hydra vulgaris III. Formation and function during morphogenesis of hydra cell aggregates. *Dev Biol* 157 (2): 383–398.

Sasse J and **Gallagher S R** (2009): Staining proteins in gels. *Curr Protoc Mol Biol* Chapter 10: Unit 10.6.

Satijn D P, **Hamer K M**, **den Blaauwen J** and **Otte A P** (2001): The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in Xenopus embryos. *Mol Cell Biol* 21 (4): 1360–1369.

Satijn D P and **Otte A P** (1999): Polycomb group protein complexes: do different complexes regulate distinct target genes? *Biochim Biophys Acta* 1447 (1): 1–16.

Saurin A J, **Shao Z**, **Erdjument-Bromage H**, **Tempst P** and **Kingston R E** (2001): A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412 (6847): 655–660.

Saurin A J, **Shiels C**, **Williamson J**, **Satijn D P**, **Otte A P**, **Sheer D** and **Freemont P S** (1998): The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol* 142 (4): 887–898.

Schaller H and **Gierer A** (1973): Distribution of the head-activating substance in hydra and its localization in membranous particles in nerve cells. *J Embryol Exp Morphol* 29 (1): 39–52.

Schaller H C (1973): Isolation and characterization of a low-molecular-weight substance activating head and bud formation in hydra. *J Embryol Exp Morphol* 29 (1): 27–38.

Schierwater B, **Eitel M**, **Jakob W**, **Osigus H J**, **Hadrys H**, **Dellaporta S L**, **Kolokotronis S O** and **Desalle R** (2009): Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol* 7 (1): e20.

Schmidt T and **David C N** (1986): Gland cells in Hydra: cell cycle kinetics and development. *J Cell Sci* 85: 197–215.

Schuettengruber B, **Chourrout D**, **Vervoort M**, **Leblanc B** and **Cavalli G** (2007): Genome regulation by polycomb and trithorax proteins. *Cell* 128 (4): 735–745.

Schultz J, **Milpetz F**, **Bork P** and **Ponting C P** (1998): SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95 (11): 5857–5864.

Schwartz Y B, **Kahn T G**, **Nix D A**, **Li X Y**, **Bourgon R**, **Biggin M** and **Pirrotta V** (2006): Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat Genet* 38 (6): 700–705.

Seto E, **Lewis B** and **Shenk T** (1993): Interaction between transcription factors Sp1 and YY1. *Nature* 365 (6445): 462–464.

Seto E, **Shi Y** and **Shenk T** (1991): YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* 354 (6350): 241–245.

Sewalt R G, **van der Vlag J**, **Gunster M J**, **Hamer K M**, **den Blaauwen J L**, **Satijn D P**, **Hendrix T**, **van Driel R** and **Otte A P** (1998): Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. *Mol Cell Biol* 18 (6): 3586–3595.

Shao Z, **Raible F**, **Mollaaghababa R**, **Guyon J R**, **Wu C T**, **Bender W** and **Kingston R E** (1999): Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98 (1): 37–46.

Shaver S, **Casas-Mollano J A**, **Cerny R L** and **Cerutti H** (2010): Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics* 5 (4): 301–312.

Shi Y, **Lee J S** and **Galvin K M** (1997): Everything you have ever wanted to know about Yin Yang 1..... *Biochim Biophys Acta* 1332 (2): F49–F66.

Shi Y, **Seto E**, **Chang L S** and **Shenk T** (1991): Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* 67 (2): 377–388.

Silva J, **Mak W**, **Zvetkova I**, **Appanah R**, **Nesterova T B**, **Webster Z**, **Peters A H F M**, **Jenuwein T**, **Otte A P** and **Brockdorff N** (2003): Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* 4 (4): 481–495.

Simon J, **Chiang A**, **Bender W**, **Shimell M J** and **O'Connor M** (1993): Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Dev Biol* 158 (1): 131–144.

Sinclair W K (1965): Hydroxyurea: differential lethal effects on cultured mammalian cells during the cell cycle. *Science* 150 (704): 1729–1731.

Sing A, **Pannell D**, **Karaiskakis A**, **Sturgeon K**, **Djabali M**, **Ellis J**, **Lipshitz H D** and **Cordes S P** (2009): A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* 138 (5): 885–897.

Singh P B, **Miller J R**, **Pearce J**, **Kothary R**, **Burton R D**, **Paro R**, **James T C** and **Gaunt S J** (1991): A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. *Nucleic Acids Res* 19 (4): 789–794.

Song W, **Tao K**, **Li H**, **Jin C**, **Song Z**, **Li J**, **Shi H**, **Li X**, **Dang Z** and **Dou K** (2010): Bmi-1 is related to proliferation, survival and poor prognosis in pancreatic cancer. *Cancer Sci* 101 (7): 1754–1760.

Spemann H and **Mangold H** (1924): Über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. *Archiv für Mikroskopische Anatomie und Entwicklungsmechanik* 100: 599–638.

Srinivasan L and **Atchison M L** (2004): YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev* 18 (21): 2596–2601.

Srivastava M, **Begovic E**, **Chapman J**, **Putnam N H**, **Hellsten U**, **Kawashima T**, **Kuo A**, **Mitros T**, **Salamov A**, **Carpenter M L**, **Signorovitch A Y**, **Moreno M A**, **Kamm K**, **Grimwood J**, **Schmutz J**, **Shapiro H**, **Grigoriev I V**, **Buss L W**, **Schierwater B**, **Dellaporta S L** and **Rokhsar D S** (2008): The Trichoplax genome and the nature of placozoans. *Nature* 454 (7207): 955–960.

Srivastava M, **Simakov O**, **Chapman J**, **Fahey B**, **Gauthier M E A**, **Mitros T**, **Richards G S**, **Conaco C**, **Dacre M**, **Hellsten U**, **Larroux C**, **Putnam N H**, **Stanke M**, **Adamska M**, **Darling A**, **Degnan S M**, **Oakley T H**, **Plachetzki D C**, **Zhai Y**, **Adamski M**, **Calcino A**, **Cummins S F**, **Goodstein D M**, **Harris C**, **Jackson D J**, **Leys S P**, **Shu S**, **Woodcroft B J**, **Vervoort M**, **Kosik K S**, **Manning G**, **Degnan B M** and **Rokhsar D S** (2010): The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466 (7307): 720–726.

Stedman E and **Stedman E** (1950): Cell specificity of histones. *Nature* 166: 780–781.

Steele R E, **David C N** and **Technau U** (2010): A genomic view of 500 million years of cnidarian evolution. *Trends Genet* .

Stover N A and **Steele R E** (2001): Trans-spliced leader addition to mRNAs in a cnidarian. *Proc Natl Acad Sci U S A* 98 (10): 5693–5698.

Stros M (2010): HMGB proteins: interactions with DNA and chromatin. *Biochim Biophys Acta* 1799 (1-2): 101–113.

Stros M, **Launholt D** and **Grasser K D** (2007): The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci* 64 (19-20): 2590–2606.

Strutt H and **Paro R** (1997): The polycomb group protein complex of Drosophila melanogaster has different compositions at different target genes. *Mol Cell Biol* 17 (12): 6773–6783.

Suganuma T and **Workman J L** (2008): Crosstalk among Histone Modifications. *Cell* 135 (4): 604–607.

Sugiyama T and **Fujisawa T** (1978): Genetic analysis of developmental mechanisms in hydra. V. Cell lineage and development of chimera hydra. *J Cell Sci* 32: 215–232.

Suzuki M, **Gerstein M** and **Yagi N** (1994): Stereochemical basis of DNA recognition by Zn fingers. *Nucleic Acids Res* 22 (16): 3397–3405.

Takahashi T, **Hayakawa E**, **Koizumi O** and **Fujisawa T** (2008): Neuropeptides and their functions in Hydra. *Acta Biol Hung* 59 Suppl: 227–235.

Takahashi T, **Koizumi O**, **Ariura Y**, **Romanovitch A**, **Bosch T C**, **Kobayakawa Y**, **Mohri S**, **Bode H R**, **Yum S**, **Hatta M** and **Fujisawa T** (2000): A novel neuropeptide, Hym-355, positively regulates neuron differentiation in Hydra. *Development* 127 (5): 997–1005.

Tardent P and **Holstein T** (1982): Morphology and morphodynamics of the stenotele nematocyst of Hydra attenuata Pall. (Hydrozoa, Cnidaria). *Cell Tissue Res* 224 (2): 269–290.

Technau U and **Holstein T W** (1992): Cell sorting during the regeneration of Hydra from reaggregated cells. *Dev Biol* 151 (1): 117–127.

Terranova R, **Yokobayashi S**, **Stadler M B**, **Otte A P**, **van Lohuizen M**, **Orkin S H** and **Peters A H F M** (2008): Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 15 (5): 668–679.

Tie F, **Furuyama T**, **Prasad-Sinha J**, **Jane E** and **Harte P J** (2001): The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* 128: 275–286.

Tolhuis B, **de Wit E**, **Muijrers I**, **Teunissen H**, **Talhout W**, **van Steensel B** and **van Lohuizen M** (2006): Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. *Nat Genet* 38 (6): 694–699.

Trembley A, *Mémoires pour servir à l'histoire d'un genre de polypes d'eau douce, à bras en forme de cornes* (Jean Verbeek and Herman Verbeek, Leiden, the Netherlands 1744).

Tsukiyama T and **Wu C** (1995): Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* 83 (6): 1011–1020.

Tuckfield A, **Clouston D R**, **Wilanowski T M**, **Zhao L L**, **Cunningham J M** and **Jane S M** (2002): Binding of the RING polycomb proteins to specific target genes in complex with the grainyhead-like family of developmental transcription factors. *Mol Cell Biol* 22 (6): 1936–1946.

Umlauf D, **Goto Y**, **Cao R**, **Cerqueira F**, **Wagschal A**, **Zhang Y** and **Feil R** (2004): Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat Genet* 36 (12): 1296–1300.

Usheva A and **Shenk T** (1994): TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* 76 (6): 1115–1121.

Usui H, **Ichikawa T**, **Kobayashi K** and **Kumanishi T** (2000): Cloning of a novel murine gene Sfmbt, Scm-related gene containing four mbt domains, structurally belonging to the Polycomb group of genes. *Gene* 248 (1-2): 127–135.

van Leeuwenhoek A (1702): Part of a Letter from Mr Antony van Leeuwenhoek, F. R. S. concerning Green Weeds Growing in Water, and Some Animalcula Found about Them. *Philosophical Transactions of the Royal Society* 23: 1304–1311.

Vaughn J L, **Goodwin R H**, **Tompkins G J** and **McCawley P** (1977): The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro* 13 (4): 213–217.

Viré E, **Brenner C**, **Deplus R**, **Blanchon L**, **Fraga M**, **Didelot C**, **Morey L**, **Eynde A V**, **Bernard D**, **Vanderwinden J M**, **Bollen M**, **Esteller M**, **Croce L D**, **de Launoit Y** and **Fuks F** (2006): The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439 (7078): 871–874.

Waddington C H, *An introduction to modern genetics* (The Macmillan Company, New York 1939).

Wang H, **Wang L**, **Erdjument-Bromage H**, **Vidal M**, **Tempst P**, **Jones R S** and **Zhang Y** (2004a): Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431 (7010): 873–878.

Wang J, **Mager J**, **Chen Y**, **Schneider E**, **Cross J C**, **Nagy A** and **Magnuson T** (2001): Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat Genet* 28 (4): 371–375.

Wang L, **Brown J L**, **Cao R**, **Zhang Y**, **Kassis J A** and **Jones R S** (2004b): Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14 (5): 637–646.

Wang L, **Jahren N**, **Miller E L**, **Ketel C S**, **Mallin D R** and **Simon J A** (2010): Comparative analysis of chromatin binding by Sex Comb on Midleg (SCM) and other polycomb group repressors at a Drosophila Hox gene. *Mol Cell Biol* 30 (11): 2584–2593.

Wang S, **Robertson G P** and **Zhu J** (2004c): A novel human homologue of Drosophila polycomb-like gene is up-regulated in multiple cancers. *Gene* 343 (1): 69–78.

Weber J (1989): Nematocysts (stinging capsules of Cnidaria) as Donnan-potential-dominated osmotic systems. *Eur J Biochem* 184 (2): 465–476.

Weintraub H and **Groudine M** (1976): Chromosomal subunits in active genes have an altered conformation. *Science* 193 (4256): 848–856.

Whitcomb S J, **Basu A**, **Allis C D** and **Bernstein E** (2007): Polycomb Group proteins: an evolutionary perspective. *Trends Genet* 23 (10): 494–502.

Wilkinson F H, **Park K** and **Atchison M L** (2006): Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proc Natl Acad Sci U S A* 103 (51): 19 296–19 301.

Wilson C M (1983): Staining of proteins on gels: comparisons of dyes and procedures. *Methods Enzymol* 91: 236–247.

Wittlieb J, **Khalturin K**, **Lohmann J U**, **Anton-Erxleben F** and **Bosch T C G** (2006): Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. *Proc Natl Acad Sci U S A* 103 (16): 6208–6211.

Woo C J, **Kharchenko P V**, **Daheron L**, **Park P J** and **Kingston R E** (2010): A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140 (1): 99–110.

Wu S, **Trievel R C** and **Rice J C** (2007): Human SFMBT is a transcriptional repressor protein that selectively binds the N-terminal tail of histone H3. *FEBS Lett* 581 (17): 3289–3296.

Wysocka J, **Swigut T**, **Xiao H**, **Milne T A**, **Kwon S Y**, **Landry J**, **Kauer M**, **Tackett A J**, **Chait B T**, **Badenhorst P**, **Wu C** and **Allis C D** (2006): A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442 (7098): 86–90.

Xu C, **Bian C**, **Yang W**, **Galka M**, **Ouyang H**, **Chen C**, **Qiu W**, **Liu H**, **Jones A E**, **MacKenzie F**, **Pan P**, **Li S S C**, **Wang H** and **Min J** (2010): Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc Natl Acad Sci U S A* 107 (45): 19 266–19 271.

Yant S R, **Zhu W**, **Millinoff D**, **Slightom J L**, **Goodman M** and **Gumucio D L** (1995): High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster. *Nucleic Acids Res* 23 (21): 4353–4362.

Ye J, **Ghosh P**, **Cippitelli M**, **Subleski J**, **Hardy K J**, **Ortaldo J R** and **Young H A** (1994): Characterization of a silencer regulatory element in the human interferon-gamma promoter. *J Biol Chem* 269 (41): 25 728–25 734.

Zink B and **Paro R** (1989): In vivo binding pattern of a trans-regulator of homoeotic genes in Drosophila melanogaster. *Nature* 337 (6206): 468–471.

Appendix

A. Molecule information on *Hydra* PcG genes

A.1. Information on *Hydra* PcG genes

TABLE A.1 Information on Hydra PcG genes. *The position of the genes in the Hydra genome is indicated by the Contig in which the sequence can be found and the genomic span. The number of exons is given. Gene models which resemble parts of the sequences are indicated. The genomic information was retrieved from the Hydra genome browser: http://hydrazome.metazome.net/cgi-bin/gbrowse/hydra/.*

Gene	Contig number	Genomic span	Exons	Coding region	Comments
Polycomb Repressive Complex 1 (PRC1)					
HyPC	29278 (- strand) 34962 (- strand)	10507 - 9850 34963 - 29278	1 - 2 3 - 6	675 nt	The sequence is distributed over 2 Contigs. Parts can also be found in the following gene models: *Hma2.202781* and *Hma2.210328*.
HyPH	39256 (+ strand)	101145 - 110657	1 - 6	972 nt	The sequence can also be found in the following gene models: *Hma2.2231779* and *Hma1.135095*.
HyPSC	38639 (+ strand)	88289 - 94536	1 - 10	762 nt	Part of the sequence can also be found in the following gene model: *Hma1.128026*.
HyRING	35899 (- strand)	62102 - 39091	1 - 4	975 nt	Parts of the sequence can also be found in the following gene models: *Hma2.212836*, *Hma2.212837*, *Hma1.13039* and *Hma1.102404*.
HySCM	37649 (+ strand)	71904 - 119561	1 - 18	2496 nt	Parts of the sequence can also be found in the following gene models: *Hma2.219746*, *Hma2.219747*, *Hma1.111023*, *Hma1.111024* and *Hma1.131847*.
Polycomb Repressive Complex 2 (PRC2)					
HyEZH2	38429 (- strand) 18810 (- strand)	83750 - 525 576 - 481	1 - 7, 9 - 22 8	2172 nt	The sequence is distributed over 2 Contigs. Contig 38429 seems to be misarranged. The spatial order of the exons is: 1 - 7, 9 - 11, followed by 18 - 22 and 12 - 17 at the end. Parts of the sequence can also be found in the following gene models: *Hma1.108906*, *Hma1.132922*, *Hma2.224311*, *Hma2.224312* and *Hma1.112227*.

Gene	Contig number	Genomic span	Exons	Coding region	Comments
HyEED	38295 (- strand)	73097 - 34821	1 - 17	1260 nt	Parts of the sequence can also be found in the following gene models: *Hma1.123942* and *Hma2.223416*.
HySUZ12	37069 (+ strand) 37069 (- strand) 38304 (+ strand)	74040 - 76899 101530 - 79197 7582 - 11070	1 - 2 3 - 9 11 - 14	1914 nt	The sequence is distributed over 2 Contigs. Contig 37069 seems to be misarranged. The spatial order of the exons is the following: Contig 37069 starts with exons 3 - 9 on the - strand, which are followed by exons 1 - 2 on the + strand. Exons 11 - 14 can be found in Contig 38304. 'Exon 10' could not be found in any Contig. Parts of the sequence can also be found in the following gene models: *Hma1.10833*, *Hma1.121070*, *Hma2.217024* and *Hma2.223484*.
Hyp55	39183 (+ strand) 38803 (- strand)	290158 - 290218 191502 - 162173	1 3 - 14	1302 nt	The sequence is distributed over 2 Contigs. Parts of the sequence can also be found in the following gene models: *Hma1.133695*, *Hma1.120922* and *Hma2.227018*.
Pleiohomeotic Repressive Complex (PhoRC)					
HyYY1	38857 (- strand) 24938 (- strand)	65227 - 35086 2875 - 2072	1, 5 - 6 2 - 4	1083 nt	The sequence is distributed over 2 Contigs. Parts of the sequence can also be found in the following gene models: *Hma2.227444*, *Hma2.200971* and *Hma1.128713*

A.2. Information on *Hydra* PcG proteins

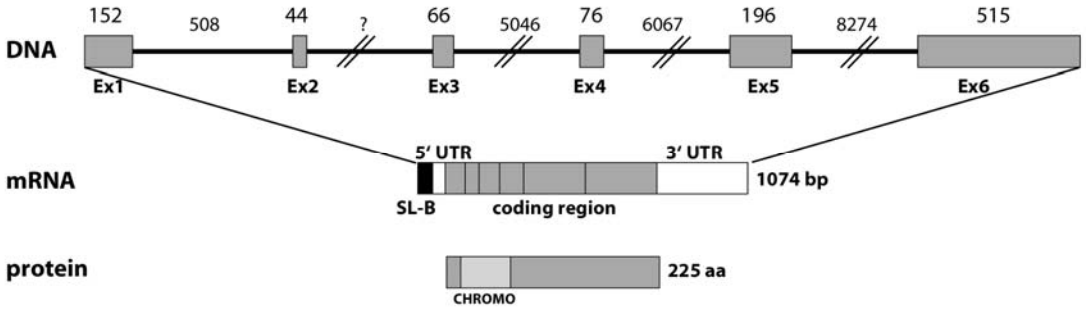
TABLE A.2 Information on Hydra PcG proteins. *The amino acid sequence lengths are given. The respective calculated molecular weights (MW) and isoelectric points (pI) of the proteins are indicated. Conserved domains found in the Hydra proteins are indicated. Information on the molecular weights and isoelectric points have been calculated using the ProtParam tool: http://www.expasy.ch/tools/protparam.html. Information on conserved protein domains was obtained from SMART (Simple Modular Architecture Research Tool): http://smart.embl-heidelberg.de/.*

Protein	Length	MW	pI	Domains
Polycomb Repressive Complex 1 (PRC1)				
***Hy*PC**	225 aa	25.1 kDa	5.76	CHROMO domain
***Hy*PH**	324 aa	36.5 kDa	6.89	SAM domain
***Hy*PSC**	254 aa	30.1 kDa	7.87	RING zinc finger
***Hy*RING**	325 aa	37.2 kDa	6.47	RING zinc finger
***Hy*SCM**	832 aa	94.0 kDa	7.66	HMG box, MBT domains and SAM domain
Polycomb Repressive Complex 2 (PRC2)				
***Hy*EZH2**	724 aa	83.7 kDa	7.94	SANT and SET domain
***Hy*EED**	420 aa	47.9 kDa	6.24	WD 40 repeats
***Hy*SUZ12**	638 aa	74.1 kDa	8.43	zinc finger domain and VEFS box
***Hy*p55**	434 aa	49.1 kDa	4.99	WD 40 repeats
Pleiohomeotic Repressive Complex (PhoRC)				
***Hy*YY1**	361 aa	40.5 kDa	5.54	REPO domain and C2H2-type zinc fingers

A.3. Sequence information on *Hydra* PcG genes

HyPC

From gene to protein:
The *HyPC* gene is composed of 6 exons and spans a genomic region of > 20 kb. The respective exon and intron lengths are indicated. Undetermined lengths are indicated by a question mark. Based on a ~ 1 kb mRNA, a protein of 225 aa is translated.



Sequence information:
Nucleotide and derived amino acid sequence of *HyPC* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The untranslated regions are highlighted in grey; the polyadenylation site is indicated by underlined nucleotides. The trans-spliced leader sequence is marked by a dashed line. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The CHROMO domain is highlighted in yellow.

```
Exon 1 →                          3' UTR
  1  acggaaaaaa acacatactg aaacttttta gtccctgtgt aataagtttt
 51  acattctgct gtccccctca aatacaatta tcttaaATGC CTAAGGACAC
                                              M   P   K   D

                                    CHROMO domain
101  TGGTCCTGGA GTTTTTGCTG CTGAAAAAAT ATTGAAAAAG AGATATAAAA
      T   G   P   G   V   F   A   A   E   K   I   L   K   K   R   Y   K

     Exon 2 →                                          Exon 3 →
151  AG|GGGCGAGC TGAGTATTTA ATAAAATGGC AAGGTTATTC AGCT|AAGTAT
       K   G   R   A   E   Y   L   I   K   W   Q   G   Y   S   A   K   Y

201  AACACATGGG AACCAGCTGC AAACATACTT GATGAAAGAT TACTTCAATG
       N   T   W   E   P   A   A   N   I   L   D   E   R   L   L   Q

                Exon 4 →
251  TTACAAAAGT TT|ACAATCAG CTTCAAGAGG GAGAAAAAGA AAAAAAGGAT
     C   Y   K   S   L   Q   S   A   S   R   G   R   K   R   K   K   G

                                                Exon 5 →
301  CAATAAGCAA GTCAACAAAA AAATCGAAAC ATGATGAG|TC TTTTGAAGTG
       S   I   S   K   S   T   K   K   S   K   H   D   E   S   F   E   V

351  AATGATTCAA CAGCAGAAGA TAGTGTAACG CAAGATGAAG ATGAAATTGT
       N   D   S   T   A   E   D   S   V   T   Q   D   E   D   E   I
```

```
 401  TTCAGTTGGA TCAGTAGAAG ATTCCATTCA ATCATCCATT CAGGATGATT
      V  S  V  G   S  V  E   D  S  I   Q  S  S  I   Q  D  D

 451  TGACCAATGA ACAAAATGAG ACAGCAACAC TTAGTGACAG TTCTAATAAG
       L  T  N   E  Q  N  E   T  A  T   L  S  D   S  S  N  K

                                                Exon 6 →
 501  AATGGAAAAA AACTTGGTAA AGATTTAAAA TTGG|GTAACG GAGATAACGA
        N  G  K   K  L  G   K  D  L  K   L   G  N   G  D  N

 551  TACGTCTGAC ACTATGATTC ATCATAATAC ACAAAAGATG TCTTCTACTG
      D  T  S  D   T  M  I   H  H  N   T  Q  K  M   S  S  T

 601  TTCGTTTAAC AGAAAAAACT TTTAGCAAAA CTGAAACAGA AGAAGTAGAA
       V  R  L   T  E  K  T   F  S  K   T  E  T   E  E  V  E

 651  TGCATATGTT GGAAAAAACC GCTGATTGAT CAAGTCACAA TCACTGACGT
        C  I  C   W  K  K   P  L  I  D   Q  V  T   I  T  D

 701  CACACTAAAT GGATTAACAG CAACTTTCAC AGAAGCTTCA ACAGACCAAG
      V  T  L  N   G  L  T   A  T  F   T  E  A  S   T  D  Q

                                  5' UTR
 751  GTTTTTTTGC ATGAaatatt attacaaaga aatctcataa gtaattcagg
       G  F  F   A  *

 801  gatcatgttc agctatttat tattattatt ttgctgtttt ggatgtctag
 851  ttggctttaa atacgcattt tatagttgag ttccttattt gttattaatt
 901  atgagtaatc taccaattcc aacgaataat gatagataac ccaacgatgt
 951  aatattacga tgttgtgtat atattatagt atctattagc gtgtgtaata
1001  tattttctgt attgtttgcg ttatgtgtaa tatgaaataa accaagatca
          polyA site
1051  aaacaaaaaa aaaaaaaaaa aaaa
```

Protein alignment:

Comparison of *Hy*PC with PC protein sequences from other species reveals a conservation on the amino acid level of only 13%. Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

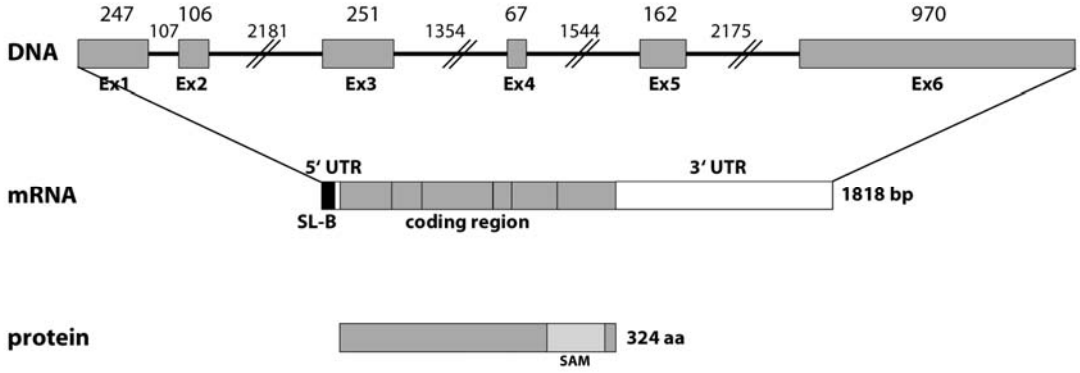
```
                                          CHROMO domain
hydra PC           1 -----MPKDTGPG-----------VFAAEKILKKRYKKGRAEYLIKWQGYSAKYNTWEPA
zebrafish CBX2     1 -----MEGLSAVGEQ---------VFDAECILNKRTRKGKLEYLVKWRGWSSKHNSWEPQ
mouse Cbx2         1 -----MEELSSVGEQ---------VFAAECILSKRLRKGKLEYLVKWRGWSSKHNSWEPE
human CBX2         1 -----MEELSSVGEQ---------VFAAECILSKRLRKGKLEYLVKWRGWSSKHNSWEPE
xenopus CBX2       1 -----MEELSAVGEQ---------VFAAECILSKRLRKGTAEYLVKWRGWSSKHNSWEPE
drosophila Pc      1 MTGRGKGSKGKLGRDNATDDPVDLVYAAEKIQKRVKKGVVEYRVKWKGWNQRYNTWEPE
                     CHROMOdomain
hydra PC          45 ANILDERLLQCY------KSLQSASRGRK-----RKKGSISKSTKKSKHDE--------S
zebrafish CBX2    47 ENLLDPRLLVAFNKREQEKELLISKKGKRPRGRPRK--IMETIPVVSK----SSSSSSSS
mouse Cbx2        47 ENILDPRLLLAFQKKEHEKEVQNRKRGKRPRGRPRKHTVTSSCSRRSKLKEPDAPSKSKS
human CBX2        47 ENILDPRLLLAFQKKEHEKEVQNRKRGKRPRGRPRKLTAMSSCSRRSKLKEPDAPSKSKS
xenopus CBX2      47 ENILDPRLLVAFQKREQEKELRNRKRGKRPRGRPRKNV------------ETDIPLKAKS
drosophila Pc     61 VNILDRRLIDIYEQT------------NKSSGTPSKR------GIKKKEKEPD-------

hydra PC          86 FEVNDSTAEDSVTQDEDE-----------------------------IVSVGSVEDSI-
zebrafish CBX2   101 SSSGSSSSSSSSSSTDDDDEDDHNMTPKPI-PRPREHLPVPQKKAQIVVAKPGPP----
mouse Cbx2       107 SSSSSSSTSSSSSSDEEE---DDSDLDSK-RGPRGRETHPVPQKKAQILVAKPELKDPIR
human CBX2       107 SSSSSSSTSSSSSSD-EE---DDSDLDAK-RGPRGRETHPVPQKKAQILVAKPELKDPIR
xenopus CBX2      95 SSSSSSSSSSSSSSSSSSSDDSDAETQQRNPRPRDSHPVPQKKAPAVLARTELKEPVR
drosophila Pc     96 ---------PEPESEEDEYTFTENDVDTH-QATTSSATHDKESKKE---------KKHHH
```

```
hydra PC                ------------------------------------------------------------
zebrafish CBX2      156 KKRGRKALPPELKAIRQVKGTRKILKPISRDSDLRGIKKPLMPASFTYTGL---------
mouse Cbx2          163 KKRGRKPLPPEQKAARRPVSLAKVLKTTRKDLGTSAAKLP-PPLSAPVAGLAALKAHTKE
human CBX2          162 KKRGRKPLPPEQKATRRPVSLAKVLKTARKDLGAPASKLP-PPLSAPVAGLAALKAHAKE
xenopus CBX2        155 KKRGRKPLPPEQKLPRR----AKGAKPGPK---GSMNKLQ-P--GHNIQGFQALKTHSKD
drosophila Pc       137 HHHHHHHHIKSERNSGRRSES-------------------------PLTHHHHHHHH---

hydra PC            115 ------------------------------------------QSSI---QDDLTNEQNET
zebrafish CBX2      207 ------NRTSGREPMAMHNRGSFTHKSSLSSLGRSIGSVSSPPT-----LNRSPQTKSAS
mouse Cbx2          222 AC-------GGPSTMATPENLASLMKGMAGSPSRGG---I-WQSSIVHYMNRMSQSQVQA
human CBX2          221 AC-------GGPSAMATPENLASLMKGMASSPGRGG---ISWQSSIVHYMNRMTQSQAQA
xenopus CBX2        205 MHASSNNRPGGLSA----ELLSSIAKNSPTQPNGSSPRSLSWQSSIVHYMNRINQNSSQP
drosophila Pc       168 ------------------ESKRQRIDHSSSS-----------NSSFTH--NSFVPEPDSN

hydra PC            130 AT-----LSDSSNKNGKKLGKDLKL-----------------------------------
zebrafish CBX2      256 DFKLSVSDMNS--------GLDPKTPTCKSPGVAALNLHSSNGQTCPQLSPTVPKDQTLL
mouse Cbx2          271 ASRLALKA-QATNKCG--LGLDLKVRTQKGGELGG----SPAGGKVPKAPGGGAAEQQRG
human CBX2          271 ASRLALKA-QATNKCG--LGLDLKVRTQK-GELGM----SPPGSKIPKAPSGGAVEQKVG
xenopus CBX2        261 GRKPASSTFNAKRSC-----LDAKSLFKPRSEAEI----SPA---MPKT-------SKLH
drosophila Pc       197 SSSSEDQPLIGTKRKAEVL--------KESGKIGVTIKTSPDGPTIKP------------

hydra PC                ------------------------------------------------------------
zebrafish CBX2      308 QR-----SASLPKSPSSSFSSLKTPS---SLQALNLQSVNKTVQGNGTDLKTSPHSGRKS
mouse Cbx2          324 N-----HS--------GSPGAQLAPTQELSLQVLDLQSVKNGVPGVGL---LARHAPA-K
human CBX2          323 NTGGPPHTHGASRVPAGCPGPQPAPTQELSLQVLDLQSVKNGMPGVGL---LARHATATK
xenopus CBX2        302 ENEEQTHSHTVQPAPTVAAGSNENPSQN-----APVQTKTQGM----------------R
drosophila Pc       237 -----------------QPTQQVTPSQQQPFQ-----------------------------

hydra PC            150 --------------GNGD------------NDTSDTMIHHNTQKMSSTVRLTEK------
zebrafish CBX2      360 SGFNTSSAPNTPSKFQTSQQALKSPQKLKADDL-AERLGKKSQARTEKILPTEGRD----
mouse Cbx2          367 AIPATNPATG---KGPGSGPTGANMTNAPTDNNKGEKLTCKATALP---APSVKRDTVKS
human CBX2          380 GVPATNPAPG---KGTGSGLIGASGATMPTDTSKSEKLASRAVAPP---TPASKRDCVKG
xenopus CBX2        341 AVVVTSPTSQNTQKSNGSHAVGV------ATSCKGDKIGKKTGVVT---EP---------
drosophila Pc       252 ------------------------------DQQQAEKIASEAAT----QLKSEQQ-----

hydra PC            178 ----------------------TFSKTETEEVECIC---------------------------
zebrafish CBX2      415 -SQPA-QDR-PSSKDPSKQSKTLSELSTGEEGSSSDTDHDS-SFPRDSHDL-----SISV
mouse Cbx2          421 VAASGGQEG-HTAPGEGRKPPALSELSTGEENSSSDSDPDSTSLPSAAQNL-----SVAI
human CBX2          434 SATPSGQES-RTAPGEARKAATLPEMSAGEESSSSDSDPDSASPPSTGQNP-----SVSV
xenopus CBX2        383 TAHPATERA-Q--PTEGQRDVA--DLSTGDD-SSLDSDHDS-SLSSQ---------DMAV
drosophila Pc       273 -ATPLATEAINTTPAESG--------AEEEEVANEEGNQQAPQVPSENNNIPKPCNNLAI

hydra PC            192 ------------------WK--KPLIDQ-VTITDVTLNGLTATFTEASTDQGFFA-----
zebrafish CBX2      466 QAGQD-------------WRPTRSLIEH-VFVTDVTANLVTVTVKESPTSVGFFSIRNY-
mouse Cbx2          475 QTSQD-------------WKPTRSLIEH-VFVTDVTANLITVTVKESPTSVGFFNLRHY-
human CBX2          488 QTSQD-------------WKPTRSLIEH-VFVTDVTANLITVTVKESPTSVGFFNLRHY-
xenopus CBX2        427 QASQD-------------WKPARSLLEH-VFVTDVTANLITVTVKESPTSVGFFNMRHF-
drosophila Pc       324 NQKQPLTPLSPRALPPRFWLPAKCNISNRVVITDVTVNLETVTIRECKTERGFFRERDMK

hydra PC                -------
zebrafish CBX2          -------
mouse Cbx2              -------
human CBX2              -------
xenopus CBX2            -------
drosophila Pc       384 GDSSPVA
```

HyPH

From gene to protein:

The *HyPH* gene is composed of 6 exons and spans a genomic region of ~ 9 kb. The respective exon and intron lengths are indicated. Based on a ~ 1.8 kb mRNA, a protein of 324 aa is translated.



Sequence information:

Nucleotide and derived amino acid sequence of *HyPH* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The untranslated regions are highlighted in grey; the polyadenylation site is indicated by underlined nucleotides. The trans-spliced leader sequence is marked by a dashed line. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The SAM domain is highlighted in light blue.

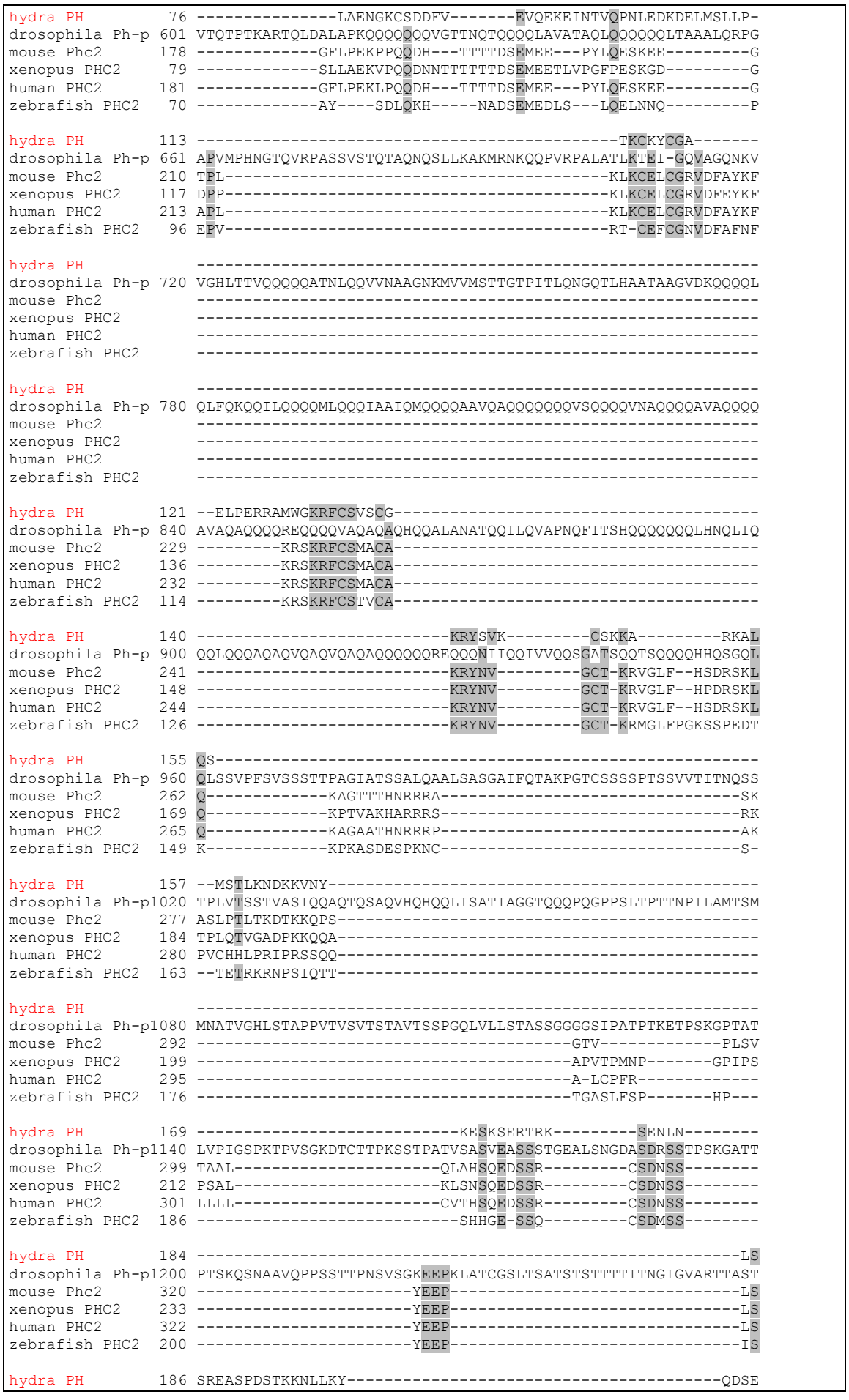
```
Exon 1 →
  1  acggaaaaaa acacatactg aaacttttta gtccctgtgt aataagcttc
 51  tataagttta aattATGGCT GTTTCATTGG TGCAAAAAAG TAATTGGACT
                    M  A   V  S  L   V  Q  K   S  N  W  T

101  AATATTACAG ATACTTATAG TTCTGCTGAT AATTGTATAA AATCGAAATC
       N  I  T   D  T  Y   S  S  A  D   N  C  I   K  S  K

151  AGATCTTTTT TGTAATAAGC ATGATGAATC GAAAGTTACT TTGTTGATAC
     S  D  L  F   C  N  K   H  D  E   S  K  V  T   L  L  I

                                                  Exon 2 →
201  ATGTTATTGA TGGTTGGATA ATTGAAGAAA GCAATCAACC ATTTAAG|GAT
      H  V  I   D  G  W  I   I  E  E   S  N  Q   P  F  K   D

251  TCAAATATTG ATGAAAATAT AAATGGGGGT GTTAAGTCAT TAGCAGAGAA
       S  N  I   D  E  N   I  N  G  G   V  K  S   L  A  E

301  TGGTAAATGT AGTGATGATT TTGTAGAAGT TCAGGAAAAA GAAATTAACA
     N  G  K  C   S  D  D   F  V  E   V  Q  E  K   E  I  N

      Exon 3 →
351  CTG|TTCAACC TAATCTTGAG GATAAAGATG AGCTTATGTC ACTCTTACCA
      T   V  Q   P  N  L  E   D  K  D   E  L  M   S  L  L  P

401  ACAAAATGCA AATATTGTGG AGCTGAGCTA CCAGAGAGAC GGGCTATGTG
       T  K  C   K  Y  C   G  A  E  L   P  E  R   R  A  M

451  GGGAAAACGT TTTTGTAGTG TATCATGTGG TAAAAGGTAC AGTGTAAAGT
     W  G  K  R   F  C  S   V  S  C   G  K  R  Y   S  V  K
```

```
501  GTTCAAAAAA AGCTAGGAAA GCTCTTCAAA GTATGTCTAC TTTAAAGAAT
       C  S  K   K  A  R  K   A  L  Q   S  M  S   T  L  K  N

551  GATAAAAAGG TAAACTATAA GGAAAGTAAA TCCGAAAGAA CTAGGAAATC
       D  K  K   V  N  Y   K  E  S  K   S  E  R   T  R  K

     Exon 4 →

601  TGAG|AATTTG AATCTTTCAA GTAGAGAAGC TTCACCGGAT AGCACAAAGA
     S  E   N  L   N  L  S   S  R  E   A  S  P  D   S  T  K

                          Exon 5 →

651  AAAATTTGTT AAAATACCAA G|ATTCTGAAT CCACATTATG CTTTCCTCCA
      K  N  L   L  K  Y  Q    D  S  E   S  T  L   C  F  P  P

701  CGAGATCCTT TGGGTTTTGG TCTTGATGTT TTTGATTTTG AAGATGAACT
       R  D  P   L  G  F   G  L  D  V   F  D  F   E  D  E

751  TTATGATGGC ATTGAGCCTA TCCAAAATTA CTCATTTTTA CCATTAGTAA
     L  Y  D  G   I  E  P   I  Q  N   Y  S  F  L   P  L  V

                                        Exon 6 →

801  CATGGTCTGT TAACCAAGTT TCGGACTATA TAAG|CACGAT TCCCGGATGC
      T  W  S   V  N  Q  V   S  D  Y   I   S  T   I  P  G  C

                          SAM domain

851  GCTCAATACG TGCCAGTTTT TGAAGCTGAA GAGATTGACG GTCAGGCATT
       A  Q  Y   V  P  V   F  E  A  E   E  I  D   G  Q  A

901  ATTGTTATTA AAGGTAGAAC ATATGGTTCA CGGTATGAAT ATCAAGGTTG
     L  L  L  L   K  V  E   H  M  V   H  G  M  N   I  K  V

951  GACCTGCTAT TAAAATTGCT GCCACTATAC GTTCAATCAA GTTAAAGTAT
      G  P  A   I  K  I  A   A  T  I   R  S  I   K  L  K  Y

1001 GGAATTAAAA CTAGATCGAA GTATTTAAGT TCACCATAGt tttaagatcg
       G  I  K   T  R  S   K  Y  L  S   S  P  *

1051 ccatagcttt gcgcaaactc taatgcttca gttttaatca acaacggtgt
1101 ctttcatgat atataggagg aacaagattc aaatatcttt gttttgattg
1151 aaaacttctt ggctgctatt ggattgttca tagaatattg attttaatgg
1201 aagtttagat attaaatttc ataaaattcg aaaatacatg tagtgaatag
1251 acaacagaaa gtgcaacaga gtatgttaaa atgcgtaatg cgctagtgaa
1301 tatttttcga gtttattatt cattgttgca aataaaaggt tcactaaatt
1351 tgacttgctt ctgaatttat ggttaaagtc taaaaactgc tacaagagca
1401 ttttttttt tacttgcaac caaattgagc gatgttgacc agttaaccat
1451 taagattctt ttaatccact gacattttga actacaaatg taatttctta
1501 ttttgtagtt tgagttagta ttaactagcc ttagtgaaaa aatcctgaaa
1551 gatattatat aaaaatctct aaggactgtt ttatatgata ttggaacaag
1601 ttatgtttag aataaatgca ataggacgtg gcatatatgg ttaactatac
1651 gttgattcct agaggagttt tatacaatct tgctcaaaag tattatagtt
1701 taatgtattt ttatttttgaa attttaacaa atatttaaaa tgtacagtta
1751 gcaattaaat gtaataagta tttatatata tgaatgttat aagaaaatta
1801 aaataaaaaa aaaaaaaa
```

Protein alignment:

Comparison of *Hy*PH with PH protein sequences from other species reveals a conservation on the amino acid level of only 16%. Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

```
hydra PH              ------------------------------------------------------------
drosophila Ph-p     1 MDRRALKFMQKRADTESDTTTPVSTTASQGISASAILAGGTLPLKDNSNIREKPLHHNYN
mouse Phc2          1 ----------------------------DTGSQNGHPEGGSHPPQRRFQHTSAVILQ---
xenopus PHC2          ------------------------------------------------------------
human PHC2          1 ---------------------------------MCLRGGCSP------RAPAAAPQ---
zebrafish PHC2        ------------------------------------------------------------

hydra PH              ------------------------------------------------------------
drosophila Ph-p    61 HNNNNSSQHSHSHQQQQQQQQVGGKQLERPLKCLETLAQKAGITFDEKYDVASPPHPGIAQ
mouse Phc2         30 ------------------------------------------------VQPASPVTPQ
xenopus PHC2          ------------------------------------------------------------
human PHC2         18 ------------------------------------------------PRPPPALPPR
zebrafish PHC2        ------------------------------------------------------------

hydra PH            1 -----------MAVSLVQKSNWTN-------------------------------------
drosophila Ph-p   121 QQATSGTGPATGSGSVTPTSHRHGTPPTGRRQTHTPSTPNRPSAPSTPNTNCNSIARHTS
mouse Phc2         40 QCAPDDWKEVVPAEKSVPVXR------PGPSPHQQAIIPAIP-----GG-----------
xenopus PHC2          ------------------------------------------------------------
human PHC2         28 PRAP------------VPASR------PG----RPLLTPARPCGRMRRG-----------
zebrafish PHC2        ------------------------------------------------------------

hydra PH              ------------------------------------------------------------
drosophila Ph-p   181 LTLEKAQNPGQQVAATTTVPLQISPEQLQQFYASNPYAIQVKQEFPTHTTSGSGTELKHA
mouse Phc2         78 -------LPGPK---SPNIQQCPAH------------------------ETGQG--IVHA
xenopus PHC2          ------------------------------------------------------------
human PHC2         55 -------SPGPRLGGSRGERRRPAGRDPARVGPGQGLRRPARPGPAAWTETGQG--IVHA
zebrafish PHC2        ------------------------------------------------------------

hydra PH           14 ITD-------------TYSSADNCIKSKS-------------------------------
drosophila Ph-p   241 TNIMEVQQQLQLQQQLSEANGGGAASAGAGGAASPANSQQSQQQQHSTAISTMSPMQLAA
mouse Phc2        102 LTDLS-------SPGMTSGNGNSA--SSIAGTA---------------------------
xenopus PHC2        1 ---------------MTSGNGSSPVPTAATGNR---------------------------
human PHC2        106 LTDLS-------IPGMTSGNGNSA--SSIAGTA---------------------------
zebrafish PHC2      1 ---------------MTSGSGNNA--PTVTGSA---------------------------

hydra PH           30 --------------------------------DLFCNKHDESKVT---------------
drosophila Ph-p   301 ATGGVGGDWTQGRTVQLMQPSTSFLYPQMIVSGNLLHPGGLGQQPIQVITAGKPFQGNGP
mouse Phc2        126 ----------------------------PQ----------NGENKPP-QAIVK--------P
xenopus PHC2       19 ----------------------------TQ----------NGENKPP-QAVVK--------P
human PHC2        130 ----------------------------PQ----------NGENKPP-QAIVK--------P
zebrafish PHC2     17 ----------------------------PQ----------NGESKPPPQAVVK--------P

hydra PH           43 -LLIHVIDGW------------------------------------------------------
drosophila Ph-p   361 QMLTTTTQNAKQMIGGQAGFAGGNYATCIPTNHNQSPQTVLFSPMNVISPQQQQNLLQSM
mouse Phc2        141 QILTHVIE----------GF-------------------------------------------
xenopus PHC2       34 QILTHFIE----------GF-------------------------------------------
human PHC2        145 QILTHVIE----------GF-------------------------------------------
zebrafish PHC2     33 QILTHVIE----------GF-------------------------------------------

hydra PH           52 ------------------------------------------------IIEESN-------
drosophila Ph-p   421 AAAAQQQQLTQQQQQFNQQQQQQLTQQQQQLTAALAKVGVDAQGKLAQKVVQKVTTTSSA
mouse Phc2        151 ------------------------------------------------VIQE---------
xenopus PHC2       44 ------------------------------------------------VIQE---------
human PHC2        155 ------------------------------------------------VIQE---------
zebrafish PHC2     43 ------------------------------------------------VIQE---------

hydra PH              ------------------------------------------------------------
drosophila Ph-p   481 VQAATGPGSTGSTQTQQVQQVQQQQQQTTQTTQQCVQVSQSTLPVGVGGQSVQTAQLLNA
mouse Phc2            ------------------------------------------------------------
xenopus PHC2          ------------------------------------------------------------
human PHC2            ------------------------------------------------------------
zebrafish PHC2        ------------------------------------------------------------

hydra PH           58 ------------------QPFKDSNIDENIN---------------GGVKS---------
drosophila Ph-p   541 GQAQQMQIPWFLQNAAGLQPFGPNQIILRNQPDGTQGMFIQQQPATQTLQTQQNQIIQCN
mouse Phc2        155 ----------------GAEPF-P---------VGRSSLLV------GNLKKKYAQ-----
xenopus PHC2       48 ----------------GAQPF-PSHRSRAVLEVGHSSLLT------G-AQEKYQQ-----
human PHC2        159 ----------------GAD-----------VSRWDARLLV------GNLKKKYAQ-----
zebrafish PHC2     47 ----------------GAEPF-P---------VERPSLLI------ENLKQKQHH-----
```

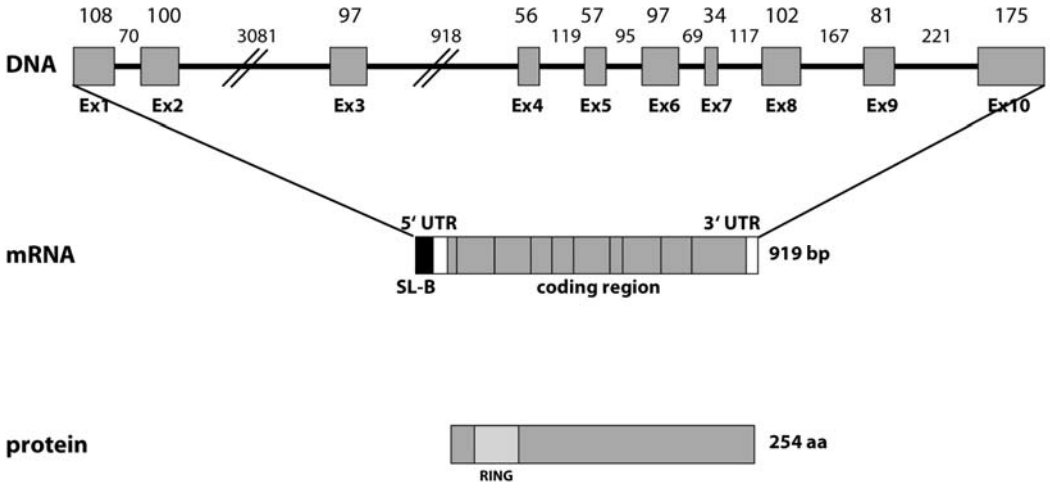

hydra PH 76 ----------LAENGKCSDDFV-------EVQEKEINTVQPNLEDKDELMSLLP-
drosophila Ph-p 601 VTQTPTKARTQLDALAPKQQQQQQQVGTTNQTQQQQLAVATAQLQQQQQQLTAAALQRPG
mouse Phc2 178 ----------GFLPEKPPQQDH---TTTTDSEMEE---PYLQESKEE---------G
xenopus PHC2 79 ----------SLLAEKVPQQDNNTTTTTTDSEMEETLVPGFPESKGD---------G
human PHC2 181 ----------GFLPEKLPQQDH---TTTTDSEMEE---PYLQESKEE---------G
zebrafish PHC2 70 ----------AY----SDLQKH-----NADSEMEDLS---LQELNNQ---------P
hydra PH 113 ----------TKCKYCGA-------
drosophila Ph-p 661 APVMPHNGTQVRPASSVSTQTAQNQSLLKAKMRNKQQPVRPALATLKTEI-GQVAGQNKV
mouse Phc2 210 TPL----------KLKCELCGRVDFAYKF
xenopus PHC2 117 DPP----------KLKCELCGRVDFEYKF
human PHC2 213 APL----------KLKCELCGRVDFAYKF
zebrafish PHC2 96 EPV----------RT-CEFCGNVDFAFNF
hydra PH ----------
drosophila Ph-p 720 VGHLTTVQQQQQATNLQQVVNAAGNKMVVMSTTGTPITLQNGQTLHAATAAGVDKQQQQL
mouse Phc2 ----------
xenopus PHC2 ----------
human PHC2 ----------
zebrafish PHC2 ----------
hydra PH ----------
drosophila Ph-p 780 QLFQKQQILQQQQMLQQQIAAIQMQQQQAAVQAQQQQQQQVSQQQQVNAQQQQAVAQQQQ
mouse Phc2 ----------
xenopus PHC2 ----------
human PHC2 ----------
zebrafish PHC2 ----------
hydra PH 121 --ELPERRAMWGKRFCSVSCG----------
drosophila Ph-p 840 AVAQAQQQQREQQQQVAQAQAQHQQALANATQQILQVAPNQFITSHQQQQQQQLHNQLIQ
mouse Phc2 229 ---------KRSKRFCSMACA----------
xenopus PHC2 136 ---------KRSKRFCSMACA----------
human PHC2 232 ---------KRSKRFCSMACA----------
zebrafish PHC2 114 ---------KRSKRFCSTVCA----------
hydra PH 140 ----------KRYSVK---------CSKKA---------RKAL
drosophila Ph-p 900 QQLQQQAQAQVQAQVQAQAQQQQQQREQQQNIIQQIVVQQSGATSQQTSQQQQHHQSGQL
mouse Phc2 241 ----------KRYNV---------GCT-KRVGLF--HSDRSKL
xenopus PHC2 148 ----------KRYNV---------GCT-KRVGLF--HPDRSKL
human PHC2 244 ----------KRYNV---------GCT-KRVGLF--HSDRSKL
zebrafish PHC2 126 ----------KRYNV---------GCT-KRMGLFPGKSSPEDT
hydra PH 155 QS----------
drosophila Ph-p 960 QLSSVPFSVSSSTTPAGIATSSALQAALSASGAIFQTAKPGTCSSSSPTSSVVTITNQSS
mouse Phc2 262 Q----------KAGTTTHNRRRA----------SK
xenopus PHC2 169 Q----------KPTVAKHARRRS----------RK
human PHC2 265 Q----------KAGAATHNRRRP----------AK
zebrafish PHC2 149 K----------KPKASDESPKNC----------S-
hydra PH 157 --MSTLKNDKKVNY----------
drosophila Ph-p1020 TPLVTSSTVASIQQAQTQSAQVHQHQQLISATIAGGTQQQPQGPPSLTPTTNPILAMTSM
mouse Phc2 277 ASLPILTKDTKKQPS----------
xenopus PHC2 184 TPLQTVGADPKKQQA----------
human PHC2 280 PVCHHLPRIPRSSQQ----------
zebrafish PHC2 163 --TETRKRNPSIQTT----------
hydra PH ----------
drosophila Ph-p1080 MNATVGHLSTAPPVTVSVTSTAVTSSPGQLVLLSTASSGGGGSIPATPTKETPSKGPTAT
mouse Phc2 292 ----------GTV----------PLSV
xenopus PHC2 199 ----------APVTPMNP-------GPIPS
human PHC2 295 ----------A-LCPFR----------
zebrafish PHC2 176 ----------TGASLFSP-------HP---
hydra PH 169 ----------KESKSERTRK---------SENLN--------
drosophila Ph-p1140 LVPIGSPKTPVSGKDTCTTPKSSTPATVSASVEASSSTGEALSNGDASDRSSTPSKGATT
mouse Phc2 299 TAAL----------QLAHSQEDSSR---------CSDNSS--------
xenopus PHC2 212 PSAL----------KLSNSQEDSSR---------CSDNSS--------
human PHC2 301 LLLL----------CVTHSQEDSSR---------CSDNSS--------
zebrafish PHC2 186 ----------SHHGE-SSQ---------CSDMSS--------
hydra PH 184 ----------LS
drosophila Ph-p1200 PTSKQSNAAVQPPSSTTPNSVSGKEEPKLATCGSLTSATSTSTTTTITNGIGVARTTAST
mouse Phc2 320 ----------YEEP----------LS
xenopus PHC2 233 ----------YEEP----------LS
human PHC2 322 ----------YEEP----------LS
zebrafish PHC2 200 ----------YEEP----------IS
hydra PH 186 SREASPDSTKKNLLKY----------QDSE


drosophila Ph-p1260 AVSTASTTTTSSGTFITSCTSTTTTTTSSISNGSKDLPKAMIKPNVLTHVIDGFIIQEAN
mouse Phc2 326 PISASSSTSRR----------RQGQ
xenopus PHC2 239 PMSASSSLSRA----------RQ-E
human PHC2 328 PISASSSTSAG----------DKAS
zebrafish PHC2 206 PLS-NSSFGAP----------IEHE
hydra PH 206 STLCFPP---RDPLGFGL----------
drosophila Ph-p1320 EPFPVTRQRYADKDVSDEPPKKKATMQEDIKLSGIASAPGSDMVACEQCGKMEHKAKLKR
mouse Phc2 341 RDLDLPDMHMRD----------LVGV----------GH----------
xenopus PHC2 253 HNVEPPNLHSRD----------PIAM----------SQ----------
human PHC2 343 GTWSSP----------TCI----------CG----------
zebrafish PHC2 220 ESFD----HSRE----------LTPL----------LT----------
hydra PH 221 ----------DVFDFEDELYDGIEPI
drosophila Ph-p1380 KRYCSPGCSRQAKNGIGGVGSGETNGLGTGGIVGVDAMALVDRLDEAMAEEKMQTEATPK
mouse Phc2 359 ----------H-FLPSEP--------
xenopus PHC2 271 ----------D-FLPSDP--------
human PHC2 354 ----------T-WWAWDT--------
zebrafish PHC2 234 ----------QHFLASDP--------
hydra PH 237 QNYSFLPL----------
drosophila Ph-p1440 LSESFPILGASTEVPPMSLPVQAAISAPSPLAMPLGSPLSVALPTLAPLSVVTSGAAPKS
mouse Phc2 ----------
xenopus PHC2 ----------
human PHC2 ----------
zebrafish PHC2 ----------
SAM domain
hydra PH 245 -----VTW------SVNQVSDYISTIPGCAQYVPVFEAEEIDGQALLLLKVEHMVHGMNI
drosophila Ph-p1500 SEVNGTDRPPISSWSVDDVSNFIRELPGCQDYVDDFIQQEIDGQALLLLKEKHLVNAMGM
mouse Phc2 366 -----TKW------NVEDVYEFIRSLPGCQEIAEEFRAQEIDGQALLLLKEDHLMSAMNI
xenopus PHC2 278 -----TKW------NVEDVYDFVRSLPGCQEISEEFRAQEIDGQALLLLKEDHLMSAMNI
human PHC2 361 -----TSCQVSHQVNVEDVYEFIRSLPGCQEIAEEFRAQEIDGQALLLLKEDHLMSVMNI
zebrafish PHC2 242 -----TKW------KVEDVYEFICSLPGCHEIAEEFRSQEIDGQALMLLKEDHLMSTMNI
SAM domain
hydra PH 294 KVGPAIKIAATIRSIKLKYGIKTRSKYLSSP
drosophila Ph-p1560 KLGPALKIVAKVESIKEVPPPGEAKDPGAQ-
mouse Phc2 415 KLGPALKIYARISMLKDS----------
xenopus PHC2 327 KLGPALKLYARISMLKDS----------
human PHC2 416 KLGPALKIYARISMLKDS----------
zebrafish PHC2 291 KLGPALKIFARISMLKDS----------

HyPSC

From gene to protein:

The *HyPSC* gene is composed of 10 exons and spans a genomic region of ~ 6 kb. The respective exon and intron lengths are indicated. Based on a ~ 1 kb mRNA, a protein of 254 aa is translated.



Sequence information:

Nucleotide and derived amino acid sequence of *HyPSC* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The untranslated regions are highlighted in grey; the polyadenylation site is indicated by underlined nucleotides. The trans-spliced leader sequence is marked by a dashed line. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The RING finger is highlighted in light red.

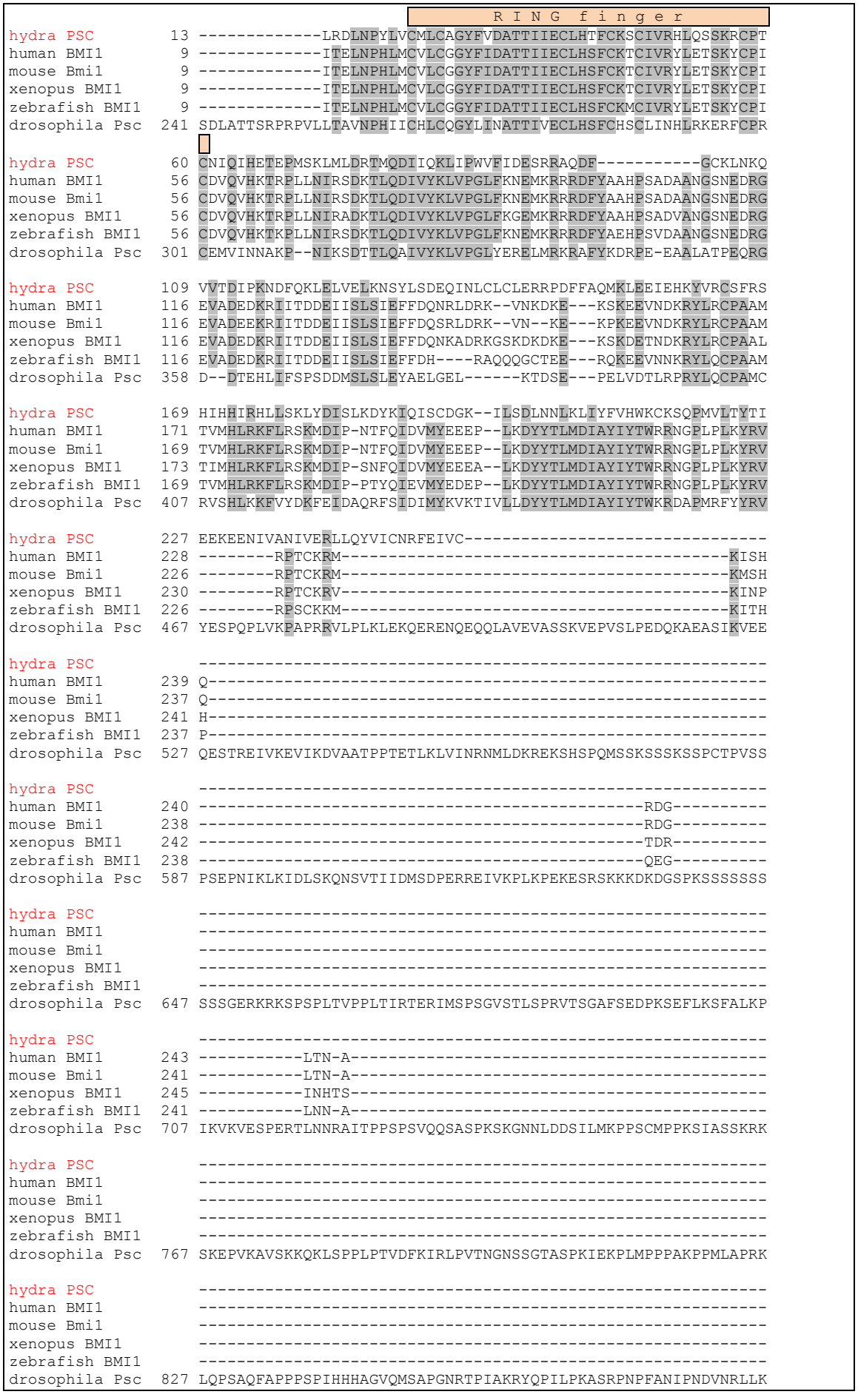
```
Exon 1 →
  1  ggacactgac atggactgaa ggagtagaaa acggaaaaaa acacatactg
 51  aaacttttta gtccctgtgt aataagtaaa gctataaata tatttgtaaa
                                                Exon 2 →
101  ctattaccaa aaatATGGAA GAAAAGAACT CACATAAG|GT TTCCATCAAG
                     M  E   E  K  N   S  H  K    V  S  I  K
151  TTACGAGATC TGAATCCTTA TTTGGTTTGT ATGCTTTGTG CTGGATATTT
       L  R  D   L  N  P   Y  L  V  C   M  L  C   A  G  Y
          RING finger                          Exon 3 →
201  TGTAGATGCC ACAACTATTA TAGAATGCTT GCATACAT|TT TGCAAAAGTT
     F  V  D  A   T  T  I   I  E  C   L  H  T   F  C  K  S
251  GCATAGTTCG ACATCTTCAA TCCAGTAAAC GATGTCCTAC TTGTAATATT
      C  I  V   R  H  L  Q   S  S  K   R  C  P   T  C  N  I
                                              Exon 4 →
301  CAAATTCACG AGACTGAACC CATGTCTAAA CTTAT|GCTTG ACCGAACCAT
       Q  I  H   E  T  E   P  M  S  K   L   M  L   D  R  T
                                                 Exon 5 →
351  GCAAGATATC ATTCAAAAGT TGATACCATG GGTTTTTATT G|ATGAAAGTC
     M  Q  D  I   I  Q  K   L  I  P   W  V  F  I    D  E  S
```

```
                                                                  Exon 6 →
401  GTCGTGCACA AGACTTTGGA TGTAAATTAA ACAAGCAAGT TGTGACAG|AT
       R  R  A   Q  D  F  G   C  K  L   N  K  Q   V  V  T   D
451  ATACCTAAAA ATGATTTTCA AAAACTTGAG TTGGTAGAAC TAAAGAATAG
       I  P  K   N  D  F   Q  K  L  E   L  V  E   L  K  N
                                                  Exon 7 →
501  TTATTTAAGC GATGAACAGA TCAATTTGTG TTTATGTTTG GAAAG|ACGTC
     S  Y  L  S   D  E  Q   I  N  L   C  L  C  L   E   R  R
                                  Exon 8 →
551  CGGATTTTTT CGCTCAAATG AAATTGGAG|G AAATCGAACA TAAATACGTG
       P  D  F   F  A  Q  M   K  L  E    E  I  E   H  K  Y  V
601  CGCTGTTCAT TCCGATCGCA TATACACCAT ATTCGCCATT TATTGAGTAA
       R  C  S   F  R  S   H  I  H  H   I  R  H   L  L  S
                                      Exon 9 →
651  ACTATACGAC ATTTCTCTCA AAGACTACAA G|ATTCAAATC TCGTGTGATG
     K  L  Y  D   I  S  L   K  D  Y   K   I  Q  I   S  C  D
701  GAAAAATATT ATCTGATTTA AATAATCTGA AATTGATATA TTTTGTACAT
       G  K  I   L  S  D  L   N  N  L   K  L  I   Y  F  V  H
                Exon 10 →
751  TGGAAATGCA AG|TCTCAACC AATGGTCTTA ACTTATACTA TAGAAGAAAA
       W  K  C   K   S  Q   P  M  V  L   T  Y  T   I  E  E
801  AGAAGAAAAC ATCGTCGCTA ATATTGTAGA AAGATTGCTT CAATACGTCA
     K  E  E  N   I  V  A   N  I  V   E  R  L  L   Q  Y  V
851  TATGTAATAG ATTCGAAATA GTTTGTTAAa gataaagtta taactgtttt
       I  C  N   R  F  E  I   V  C  *
901  tactggcaaa aaaaaaaaa
```

Protein alignment:

Comparison of *Hy*PSC with PSC protein sequences from other species reveals a conservation on the amino acid level of 21%. Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

```
hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc     1 MMTPESKAIQPAAATTKQTAEATATTTMAHTQQKSQLSTLAKTTTTTATNKAAKSVVSNA

hydra PSC            ----------------------------------------MEEKNSHKVS-IK------
human BMI1           ------------------------------------------------MHRTTRIK------
mouse Bmi1           ------------------------------------------------MHRTTRIK------
xenopus BMI1         ------------------------------------------------MHRTTRIK------
zebrafish BMI1       ------------------------------------------------MHRTTRIK------
drosophila Psc    61 NSSGNNSSKKLALSQSQKTTTTTTPPTTTTTTTAAAAAEATTNADKMQKQQQLKQQLFAA

hydra PSc            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc   121 CSIKVKSENTLATTANAALAAATTTTTTATPALATGKAAKTILENGIKKESTPPAVESVE

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc   181 ASSSSSSSSSSSSSSSSSWPTTRRATSEDASSNGGASADEEKSEEDPTAAVAASSTATTT
```


R I N G f i n g e r
hydra PSC 13 ----------LRDLNPYLVCMLCAGYFVDATTIIECLHTFCKSCIVRHLQSSKRCPT
human BMI1 9 ----------ITELNPHLMCVLCGGYFIDATTIIECLHSFCKTCIVRYLETSKYCPI
mouse Bmi1 9 ----------ITELNPHLMCVLCGGYFIDATTIIECLHSFCKTCIVRYLETSKYCPI
xenopus BMI1 9 ----------ITELNPHLMCVLCGGYFIDATTIIECLHSFCKTCIVRYLETSKYCPI
zebrafish BMI1 9 ----------ITELNPHLMCVLCGGYFIDATTIIECLHSFCKMCIVRYLETSKYCPI
drosophila Psc 241 SDLATTSRPRPVLLTAVNPHIICHLCQGYLINATTIVECLHSFCHSCLINHLRKERFCPR
hydra PSC 60 CNIQIHETEPMSKLMLDRTMQDIIQKLIPWVFIDESRRAQDF----------GCKLNKQ
human BMI1 56 CDVQVHKTRPLLNIRSDKTLQDIVYKLVPGLFKNEMKRRRDFYAAHPSADAANGSNEDRG
mouse Bmi1 56 CDVQVHKTRPLLNIRSDKTLQDIVYKLVPGLFKNEMKRRRDFYAAHPSADAANGSNEDRG
xenopus BMI1 56 CDVQVHKTRPLLNIRADKTLQDIVYKLVPGLFKGEMKRRRDFYAAHPSADVANGSNEDRG
zebrafish BMI1 56 CDVQVHKTKPLLNIRSDKTLQDIVYKLVPGLFKNEMKRRRDFYAEHPSVDAANGSNEDRG
drosophila Psc 301 CEMVINNAKP--NIKSDTTLQAIVYKLVPGLYERELMRKRAFYKDRPE-EAALATPEQRG
hydra PSC 109 VVTDIPKNDFQKLELVELKNSYLSDEQINLCLCLERRPDFFAQMKLEEIEHKYVRCSFRS
human BMI1 116 EVADEDKRIITDDEIISLSIEFFDQNRLDRK--VNKDKE---KSKEEVNDKRYLRCPAAM
mouse Bmi1 116 EVADEEKRIITDDEIISLSIEFFDQSRLDRK--VN--KE---KPKEEVNDKRYLRCPAAM
xenopus BMI1 116 EVADEDKRIITDDEIISLSIEFFDQNKADRKGSKDKDKE---KSKDETNDKRYLRCPAAL
zebrafish BMI1 116 EVADEDKRIITDDEIISLSIEFFDH----RAQQQGCTEE---RQKEEVNNKRYLQCPAAM
drosophila Psc 358 D--DTEHLIFSPSDDMSLSLEYAELGEL------KTDSE---PELVDTLRPRYLQCPAMC
hydra PSC 169 HIHHIRHLLSKLYDISLKDYKIQISCDGK--ILSDLNNLKLIYFVHWKCKSQPMVLTYTI
human BMI1 171 TVMHLRKFLRSKMDIP-NTFQIDVMYEEEP--LKDYYTLMDIAYIYTWRRNGPLPLKYRV
mouse Bmi1 169 TVMHLRKFLRSKMDIP-NTFQIDVMYEEEP--LKDYYTLMDIAYIYTWRRNGPLPLKYRV
xenopus BMI1 173 TIMHLRKFLRSKMDIP-SNFQIDVMYEEEA--LKDYYTLMDIAYIYTWRRNGPLPLKYRV
zebrafish BMI1 169 TVMHLRKFLRSKMDIP-PTYQIEVMYEDEP--LKDYYTLMDIAYIYTWRRNGPLPLKYRV
drosophila Psc 407 RVSHLKKFVYDKFEIDAQRFSIDIMYKVKTIVLLDYYTLMDIAYIYTWKRDAPMRFYYRV
hydra PSC 227 EEKEENIVANIVERLLQYVICNRFEIVC----------
human BMI1 228 --------RPTCKRM----------KISH
mouse Bmi1 226 --------RPTCKRM----------KMSH
xenopus BMI1 230 --------RPTCKRV----------KINP
zebrafish BMI1 226 --------RPSCKKM----------KITH
drosophila Psc 467 YESPQPLVKPAPRRVLPLKLEKQERENQEQQLAVEVASSKVEPVSLPEDQKAEASIKVEE
hydra PSC ----------
human BMI1 239 Q----------
mouse Bmi1 237 Q----------
xenopus BMI1 241 H----------
zebrafish BMI1 237 P----------
drosophila Psc 527 QESTREIVKEVIKDVAATPPTETLKLVINRNMLDKREKSHSPQMSSKSSSKSSPCTPVSS
hydra PSC ----------
human BMI1 240 ----------RDG----------
mouse Bmi1 238 ----------RDG----------
xenopus BMI1 242 ----------TDR----------
zebrafish BMI1 238 ----------QEG----------
drosophila Psc 587 PSEPNIKLKIDLSKQNSVTIIDMSDPERREIVKPLKPEKESRSKKKDKDGSPKSSSSSSS
hydra PSC ----------
human BMI1 ----------
mouse Bmi1 ----------
xenopus BMI1 ----------
zebrafish BMI1 ----------
drosophila Psc 647 SSSGERKRKSPSPLTVPPLTIRTERIMSPSGVSTLSPRVTSGAFSEDPKSEFLKSFALKP
hydra PSC ----------
human BMI1 243 ----------LTN-A----------
mouse Bmi1 241 ----------LTN-A----------
xenopus BMI1 245 ----------INHTS----------
zebrafish BMI1 241 ----------LNN-A----------
drosophila Psc 707 IKVKVESPERTLNNRAITPPSPSVQQSASPKSKGNNLDDSILMKPPSCMPPKSIASSKRK
hydra PSC ----------
human BMI1 ----------
mouse Bmi1 ----------
xenopus BMI1 ----------
zebrafish BMI1 ----------
drosophila Psc 767 SKEPVKAVSKKQKLSPPLPTVDFKIRLPVTNGNSSGTASPKIEKPLMPPPAKPPMLAPRK
hydra PSC ----------
human BMI1 ----------
mouse Bmi1 ----------
xenopus BMI1 ----------
zebrafish BMI1 ----------
drosophila Psc 827 LQPSAQFAPPPSPIHHHAGVQMSAPGNRTPIAKRYQPILPKASRPNPFANIPNDVNRLLK

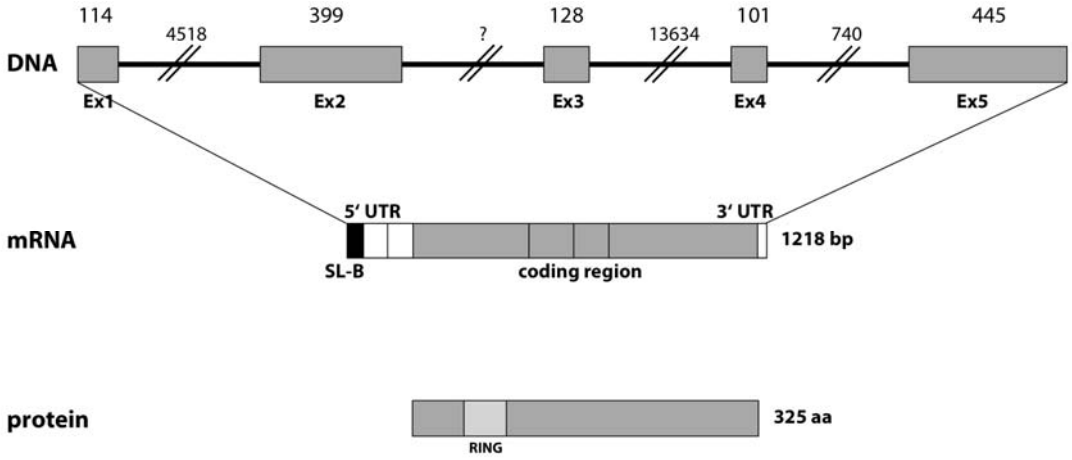
```
hydra PSC            ------------------------------------------------------------
human BMI1       247 ----------GELESDS-----------GSDKANSPA----------------GGIP
mouse Bmi1       245 ----------GELESDS-----------GSDKANSPA----------------GGVP
xenopus BMI1     250 ----------GDMESDS-----------GSDKAGSLG----------------VVIP
zebrafish BMI1   245 ----------NRSESDS-----------ASDKACSPA----------------GVP
drosophila Psc   887 DAGTEIKSIGGGSVENNSNAAQKPHLYGPKGESKMGPPALPATTPSQGNKNVGKQAGNLP

hydra PSC            ------------------------------------------------------------
human BMI1       267 --------------------STSSC------------LPSPSTPVQSPH----------P
mouse Bmi1       265 --------------------STSSC------------LPSPSTPVQSPH----------P
xenopus BMI1     270 --------------------STSSC------------IPSP--PVQSPH----------P
zebrafish BMI1   264 --------------------STSSP------------LPSPSTLVQPSQ----------P
drosophila Psc   947 MSAPPNKGNSSNNYLNLALFNSSKCKGKEAPPGCRTPMYTPNSPIYSPSSPQYVPSYNIP

hydra PSC            ------------------------------------------------------------
human BMI1       285 QFPHISSTMNGTSNSPSGNHQSS------------------FANRPRKSSVN--------
mouse Bmi1       283 QFPHISSTMNGTSNSPSANHQSS------------------FASRPRKSSLN--------
xenopus BMI1     286 HFPHISSTINGTSSS-SSSHQNP------------------FTNRARKISLN--------
zebrafish BMI1   282 HFTHISSPINGTTMT-SPNRQFN------------------FS-KVRKSALN--------
drosophila Psc  1007 TMPTYKYTPKPTPNSGSGNGGSGSYLQNMLGGGNGGSLGGLFPSPPTKSDQNTNPAQGGG

hydra PSC            ------------------------------------------------------------
human BMI1       319 -GSSATSSG------------------------------------------------------
mouse Bmi1       317 -GSSATSSG------------------------------------------------------
xenopus BMI1     319 -GVSAISSG------------------------------------------------------
zebrafish BMI1   314 -GSS-TSSG------------------------------------------------------
drosophila Psc  1067 GSSSATQSGGNNGIVNNNIYMPNEDAPEKQQVKVKSLLNSCNINIPSSLSITISRDNGDS

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc  1127 SSPNNGQHPKHKSPVNNYIEIVKLPDQPQDQVQAAKEAQKRQSPPAAVPGHLAAKLPPPP

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc  1187 PSKAIPSPQHLVSRMTPPQLPKVATPPPPSSPRVITPPKTSPPANAAKVTPLKPVLTPTQ

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc  1247 VDKKTPSPEKRTAAQMGSHSPTASENKSPKGGPAGVANSTGGAQNGDPAAKKFRPILPRQ

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc  1307 NGMPELAPKLPTLAPFVGFNPLQNPAAGKKVPPSKKSPNAGAAAHQSGQQKLVNGGQSQP

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc  1367 AQQKTSPPAQKNQQQVKKVSKNPTPPPPSLPAVGKMMPHPVMHSQNAPLSIASSASAAAV

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc  1427 ASGQLDLSNFLKENLRRVHAAQAAQAAQVAAAANQSNMMYNLAQMGHMTPAMYNYQQAYF

hydra PSC            ------------------------------------------------------------
human BMI1           ------------------------------------------------------------
mouse Bmi1           ------------------------------------------------------------
xenopus BMI1         ------------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------------
drosophila Psc  1487 REQLSRMQRVGNEVFNDYLQKLKTAAATGGGGPVEGELKPMLPTVTLPSPGATPPAASPK

hydra PSC            ------------------------------------------------------
```

```
human BMI1           ------------------------------------------------------
mouse Bmi1           ------------------------------------------------------
xenopus BMI1         ------------------------------------------------------
zebrafish BMI1       ------------------------------------------------------
drosophila Psc 1547 TSPLPAGKLTAAATAPQTKGNSSSGAANARQQTAATGNNGATVPAASLPPATKSK
```

HyRING

From gene to protein:

The *HyRING* gene is composed of 5 exons and spans a genomic region of > 20 kb. The respective exon and intron lengths are indicated. Based on a ~ 1.2 kb mRNA, a protein of 325 aa is translated.



Sequence information:

Nucleotide and derived amino acid sequence of *HyRING* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The untranslated regions are highlighted in grey; the polyadenylation site is indicated by underlined nucleotides. The trans-spliced leader sequence is marked by a dashed line. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The RING finger is highlighted in light red.

```
Exon 1 →
  1  acggaaaaaa acacatactg aaacttttta gtccctgtgt aataagctgt
 51  ttaaaataac gatgcaaaat ttttcgcata taattcagta ccatataatc
                  Exon 2 →
101  agttgcatat aatc|agtttc gaagctagct aacccaaaat ttgtgtctgc
151  ataaaaaata tttgtgatat gtaaatgtag tctatgATGT CACAGTCTGC
                                          M   S  Q  S

201  ACAAGGTCTT CAACAACCAA AATCTTGGGA ACTGAGTTTG TATGAGCTTC
     A  Q  G  L   Q  Q  P   K  S  W   E  L  S  L   Y  E  L

251  ATCGAACACC ACAAGAAGTT ATTACAGATG ATACAGAAAT TGCTATATCA
      H  R  T  P  Q  E  V   I  T  D   D  T  E   I  A  I  S

301  CCGAGATCTC TTCATAGCGA ACTTATGTGT CCAATTTGTT TAGATATGTT
       P  R  S  L  H  S   E  L  M  C  P  I  C  L  D  M

                           RING finger
351  AACTAATACT ATGACAACAA AGGAATGCTT ACATAGGTTT TGTCATGATT
     L  T  N  T  M  T  T  K  E  C  L  H  R  F  C  H  D

401  GTATAATAAC CGCTTTAAGA GCTGGGAACA AAGAGTGTCC TACATGTAGA
      C  I  I  T  A  L  R  A  G  N  K  E  C  P  T  C  R

451  AAAAAGTTAA TATCAAAAAG ATCACTGAGA CCCGATCCTA ACTTTGATTC
       K  K  L  I  S  K  R  S  L  R  P  D  P  N  F  D
```

```
                    Exon 3 →
501  TCTAATTGCA AAG|ATATATC CTAATCGAGA CGAGTATGAA GCTCATCAAG
       S  L  I  A   K   I  Y   P  N  R   D  E  Y  E   A  H  Q

551  AACAGGCACT TGAAAAGTTA CATCAACATA CCAATCAGCA AGCACTGACC
       E  Q  A   L  E  K  L   H  Q  H   T  N  Q   Q  A  L  T

                                                  Exon 4 →
601  AACAGTATAG AAGAAGGTCT TCGTATTCAA GCTAGAAACA G|ACTGAACCG
       N  S  I   E  E  G   L  R  I  Q   A  R  N    R  L  N

651  AATGCTTAAA CAACAGCGCA AGCAAGCAGA ACATGCTCTA GTTGATCATT
       R  M  L  K   Q  Q  R   K  Q  A   E  H  A  L   V  D  H

                                                  Exon 5 →
701  CAGCGACTGT TAAAGAGGAA GATAGCCCAG TAATTTCACA AG|ATTCTGGT
       S  A  T   V  K  E  E   D  S  P   V  I  S   Q   D  S  G

751  GAAATGTCGA CAGAAAATAA CGATTCAGCA GTTATGACTG GAAAACATAA
       E  M  S   T  E  N   N  D  S  A   V  M  T   G  K  H

801  AGTTTCGAAG AGAAAACACA AAGATTCTAC TGGTTCTCTT ACCGATAATT
       K  V  S  K   R  K  H   K  D  S   T  G  S  L   T  D  N

851  CAAACTTTTT AAATAATGAA GTAAACGAGG TAGAAGTGCT GTTTCACCCT
       S  N  F   L  N  N  E   V  N  E   V  E  V   L  F  H  P

901  CATCCTGCTT GTGATATTCT TAAACACTTT AAATCGCGAT ATTTAAAAAC
       H  P  A   C  D  I   L  K  H  F   K  S  R   Y  L  K

951  TACGGGAATT GCAACTGTTG CTCATTTGTG TCAATTTTTA GTTACAAGAA
       T  T  G  I   A  T  V   A  H  L   C  Q  F  L   V  T  R

1001 TTTCTCTAGA AGAAATAACT CCGTGTATCG ACTTTGAAGT GTTGGTACGT
       I  S  L   E  E  I  T   P  C  I   D  F  E   V  L  V  R

1051 ACGGTCGGGG AATATCAGAA ACTTGAAAAA CATTTAACTC TTGCAGAAAT
       T  V  G   E  Y  Q   K  L  E  K   H  L  T   L  A  E

1101 TAAACAGCTT TATTCAAAAG ATTCGACACC ACTTGAACTT TTTTTTTCAA
       I  K  Q  L   Y  S  K   D  S  T   P  L  E  L   F  F  S

1151 GTGAAAATCG GTAAAtttgtt taaaaactaa atagtttaaa caaaaaaaaa
       S  E  N   R  *

1201 aaaaaaaaaa aaaaaaaa
```

Protein alignment:

Comparison of *Hy*RING with RING protein sequences from other species reveals significant conservation on the amino acid level (44%). Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

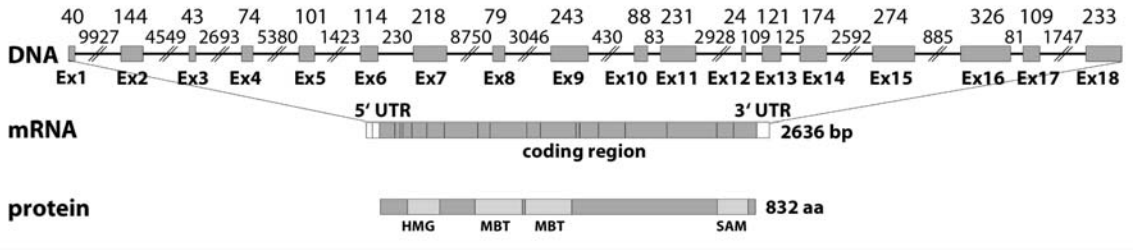
```
                                                                  
hydra RING          1 MSQSAQ--GLQQ-PKSWELSLYELHRTPQEVITDDTEIAISPRSLHSELMCPICLDMLTN
drosophila Sce      1 -----MTSLDPAPNKTWELSLYELQRKPQEVITDSTEIAVSPRSLHSELMCPICLDMLKK
mouse Rnf2          1 MSQAVQTNGTQPLSKTWELSLYELQRTPQEAITDGLEIVVSPRSLHSELMCPICLDMLKN
human RNF2          1 MSQAVQTNGTQPLSKTWELSLYELQRTPQEAITDGLEIVVSPRSLHSELMCPICLDMLKN
xenopus RNF2        1 MAQAVQTNGAQPLSKTWELSLYELQRTPQEAITDGLEIVVSPRSLHSELMCPICLDMLKN
zebrafish RNF2      1 MTQTVQTNGVQPLSKTWELSLYELQRTPQEAITDGLEIAVSPRSLHSELMCPICLDMLKN
                        R I N G  f i n g e r                      
hydra RING         58 TMTTKECLHRFCHDCIITALRAGNKECPTCRKKLISKRSLRPDPNFDSLIAKIYPNRDEY
drosophila Sce     56 TMTTKECLHRFCSDCIVTALRSGNKECPTCRKKLVSKRSLRADPNFDLLISKIYPSREEY
mouse Rnf2         61 TMTTKECLHRFCADCIITALRSGNKECPTCRKKLVSKRSLRPDPNFDALISKIYPSRDEY
human RNF2         61 TMTTKECLHRFCADCIITALRSGNKECPTCRKKLVSKRSLRPDPNFDALISKIYPSRDEY
xenopus RNF2       61 TMTTKECLHRFCADCIITALRSGNKECPTCRKKLVSKRSLRPDPNFDALISKIYPSRDEY
zebrafish RNF2     61 TMTTKECLHRFCADCIITALRSGNKECPTCRKKLVSKRSLRPDPNFDALISKIYPSRDEY
```

```
hydra RING       118 EAHQEQALEKLHQHTNQQALTNSIEEGLRIQARNRLNRMLKQQRKQAEHA----------
drosophila Sce   116 EAIQEKVMAKFNQTQSQQALVNSINEGIKLQSQNRPQRFRTKGGGGGGGGGGNGNGAANV
mouse Rnf2       121 EAHQERVLARINKHNNQQALSHSIEEGLKIQAMNRLQRGKKQQIENGSGAEDNGD-----
human RNF2       121 EAHQERVLARINKHNNQQALSHSIEEGLKIQAMNRLQRGKKQQIENGSGAEDNGD-----
xenopus RNF2     121 EAHQERVLARINKHNNQQALSHSIEEGLKIQALNRLPRGKKQQVENGSGAEDNAD-----
zebrafish RNF2   121 EAHQERVLARISKHNNQQALSHSIEEGLKIQAMNRLQRGKKHQIENGSGAEDNGD-----

hydra RING       168 ------------------------LVDHSATVKEEDSPVISQDSGEMSTENNDSAVMTGK
drosophila Sce   176 AAPPAPGAPTAVGRNASNQMHVHDTASNDSNSNTNSIDRENRDPGHSGTSAASAITSASN
mouse Rnf2       176 --------------------------SSHCSNASTHS----NQEAG------------PSN
human RNF2       176 --------------------------SSHCSNASTHS----NQEAG------------PSN
xenopus RNF2     176 --------------------------SSHCSNASVHS----NQEAG------------PSN
zebrafish RNF2   176 --------------------------SSHCSNASVHS----NQEAG------------PSI

hydra RING       204 HKVSKRKHKDSTGSLT-------DNSN---------------------------------
drosophila Sce   236 ---AAPSSSANSGASTSATRMQVDDASNPPSVRSTPSPVPSNSSSSKPKRAMSVLTSERS
mouse Rnf2       195 ---KRTKTSDDSG--------LELDNNNAAVAI----DPVMDGAS----------------
human RNF2       195 ---KRTKTSDDSG--------LELDNNNAAMAI----DPVMDGAS----------------
xenopus RNF2     195 ---KRTKTSDDSG--------LELDTNNETASM----DSVLDGAS----------------
zebrafish RNF2   195 ---KRTKTSDDSG--------LDMDNATENGGG----DIALDGVS----------------

hydra RING       224 --------------------------FLNNEVNEVEVLFHPHP--------ACDILKHFK
drosophila Sce   293 EESESDSQMDCRTEGDSNIDTEGEGNGELGINDEIELVFKPHPTEMSADNQLIRALKENC
mouse Rnf2       225 --------------------------------EIELVFRPHPTLMEKDDSA-------Q
human RNF2       225 --------------------------------EIELVFRPHPTLMEKDDSA-------Q
xenopus RNF2     225 --------------------------------EIELVFRPHPTLMEKDDSA-------Q
zebrafish RNF2   225 --------------------------------EIELVFRPHPTLMEKEDAA-------Q

hydra RING       250 SRYLKTTGIATVAHLCQFLVTRISLEEITPCID-------------FEVLVRTV-GEYQK
drosophila Sce   353 VRYIKTTANATVDHLSKYLAMRLTLDLGADLPE-------ACRVLNFCIYVAPQPQQLVI
mouse Rnf2       245 TRYIKTSGNATVDHLSKYLAVRLALEELRSKGESNQMNLDTASEKQYTIYIATASGQFTV
human RNF2       245 TRYIKTSGNATVDHLSKYLAVRLALEELRSKGESNQMNLDTASEKQYTIYIATASGQFTV
xenopus RNF2     245 TRYIKTSGNATVDHLSKYLAVRLALEEFRSKGENNEMSLSAASEKQYTIYIATANGQFTV
zebrafish RNF2   245 TRYIKTSGNATVDHLSKYLAVRLALEEMRKNGEASPINVEAASEKQYTIYIPTASNQFTV

hydra RING       296 LEKHLTLAEIKQLYSKDSTPLELFFSSENR--
drosophila Sce   406 LNGNQTLHQVNDKFWKVNKPMEMYYSWKKT--
mouse Rnf2       305 LNGSFSLELVSEKYWKVNKPMELYYAPTKEHK
human RNF2       305 LNGSFSLELVSEKYWKVNKPMELYYAPTKEHK
xenopus RNF2     305 LNGSFSLELVSEKYWKVNKPMELYYAPTKEHK
zebrafish RNF2   305 LNGSFSLELVSEKYWKVNKPMELYFAPTKEHK
```

HySCM

From gene to protein:

The *HySCM* gene is composed of 18 exons and spans a genomic region of > 60 kb. The respective exon and intron lengths are indicated. Based on a ~ 2.6 kb mRNA, a protein of 832 aa is translated.

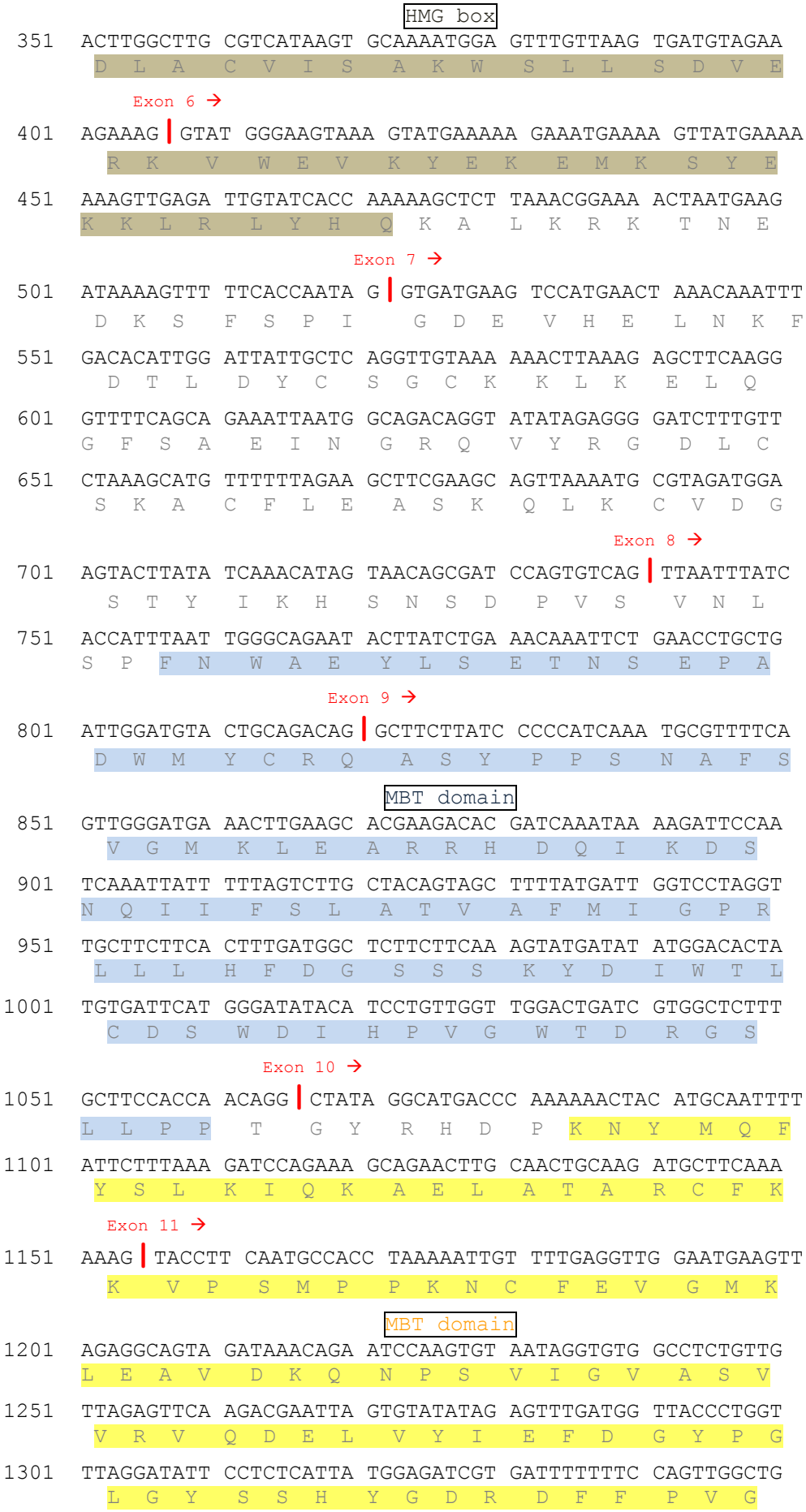


Sequence information:

Nucleotide and derived amino acid sequence of *HySCM* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The untranslated regions are highlighted in grey; the polyadenylation site is indicated by underlined nucleotides. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The HMG-box is highlighted in brown, the MBT domains are highlighted in light blue and yellow, the SAM domain is highlighted in green.

Note: In the sequenced clones there is probably a nucleotide-deletion in the 5´UTR after position 261. One adenine is missing (framed A). This leads to a shorter coding sequence for *Hy*SCM! The coding region of the sequenced clones starts at position 287 (highlighted in light red), whereas the correct coding region according to the Contig starts at position 86 (highlighted in red). The correct sequence according to the Contig is given below.

```
Exon 1 →                                                    Exon 2 →
   1  aaccaaaaaa aaaatttaat ctatctattt tcttgagaa | a gacttgttag
  51  gtacactagt aaatgttgct caaaatgatt gttctATGGA AATGATTAAC
                                                  M   E  M  I  N
 101  GCAAACACAA TTAATTTAAT AAAAAGTGAA CTTTCTTCTC CTCCTTTAAA
        A  N  T   I  N  L   I  K  S  E   L  S  S   P  P  L
                                                Exon 3 →
 151  AAGAAAAAAA TCTTTTATTG AAGATTCGAA CCAAG | ATGCA ACAGATTCAG
      K  R  K  K   S  F  I   E  D  S   N  Q   D  A   T  D  S
                                        Exon 4 →
 201  AGCAGGTTTC TGAACACAAC ACATCTCGG | A CTAATAATGG TCATTTTGAG
       E  Q  V   S  E  H  N   T  S  R    T  N  N   G  H  F  E
 251  ATGACAAATT TACCAAAGAA GCCTTTGACC CCATATATGA GCTACAGAAA
        M  T  N   L  P  K   K  P  L  T   P  Y  M   S  Y  R
                                        Exon 5 →
 301  TATGGTTTTC CAAGACATTA AACAC | AAACA CCCAGGTTGT TCACCATATG
      N  M  V  F   Q  D  I   K  H   K   H  P  G  C   S  P  Y
```


HMG box
351 ACTTGGCTTG CGTCATAAGT GCAAAATGGA GTTTGTTAAG TGATGTAGAA
D L A C V I S A K W S L L S D V E
Exon 6
401 AGAAAG | GTAT GGGAAGTAAA GTATGAAAAA GAAATGAAAA GTTATGAAAA
R K V W E V K Y E K E M K S Y E
451 AAAGTTGAGA TTGTATCACC AAAAAGCTCT TAAACGGAAA ACTAATGAAG
K K L R L Y H Q K A L K R K T N E
Exon 7
501 ATAAAAGTTT TTCACCAATA G | GTGATGAAG TCCATGAACT AAACAAATTT
D K S F S P I G D E V H E L N K F
551 GACACATTGG ATTATTGCTC AGGTTGTAAA AAACTTAAAG AGCTTCAAGG
D T L D Y C S G C K K L K E L Q
601 GTTTTCAGCA GAAATTAATG GCAGACAGGT ATATAGAGGG GATCTTTGTT
G F S A E I N G R Q V Y R G D L C
651 CTAAAGCATG TTTTTTAGAA GCTTCGAAGC AGTTAAAATG CGTAGATGGA
S K A C F L E A S K Q L K C V D G
Exon 8
701 AGTACTTATA TCAAACATAG TAACAGCGAT CCAGTGTCAG | TTAATTTATC
S T Y I K H S N S D P V S V N L
751 ACCATTTAAT TGGGCAGAAT ACTTATCTGA AACAAATTCT GAACCTGCTG
S P F N W A E Y L S E T N S E P A
Exon 9
801 ATTGGATGTA CTGCAGACAG | GCTTCTTATC CCCCATCAAA TGCGTTTTCA
D W M Y C R Q A S Y P P S N A F S
MBT domain
851 GTTGGGATGA AACTTGAAGC ACGAAGACAC GATCAAATAA AAGATTCCAA
V G M K L E A R R H D Q I K D S
901 TCAAATTATT TTTAGTCTTG CTACAGTAGC TTTTATGATT GGTCCTAGGT
N Q I I F S L A T V A F M I G P R
951 TGCTTCTTCA CTTTGATGGC TCTTCTTCAA AGTATGATAT ATGGACACTA
L L L H F D G S S S K Y D I W T L
1001 TGTGATTCAT GGGATATACA TCCTGTTGGT TGGACTGATC GTGGCTCTTT
C D S W D I H P V G W T D R G S
Exon 10
1051 GCTTCCACCA ACAGG | CTATA GGCATGACCC AAAAAACTAC ATGCAATTTT
L L P P T G Y R H D P K N Y M Q F
1101 ATTCTTTAAA GATCCAGAAA GCAGAACTTG CAACTGCAAG ATGCTTCAAA
Y S L K I Q K A E L A T A R C F K
Exon 11
1151 AAAG | TACCTT CAATGCCACC TAAAAATTGT TTTGAGGTTG GAATGAAGTT
K V P S M P P K N C F E V G M K
MBT domain
1201 AGAGGCAGTA GATAAACAGA ATCCAAGTGT AATAGGTGTG GCCTCTGTTG
L E A V D K Q N P S V I G V A S V
1251 TTAGAGTTCA AGACGAATTA GTGTATATAG AGTTTGATGG TTACCCTGGT
V R V Q D E L V Y I E F D G Y P G
1301 TTAGGATATT CCTCTCATTA TGGAGATCGT GATTTTTTTC CAGTTGGCTG
L G Y S S H Y G D R D F F P V G

Exon 12 →

1351 GTGCGCTCAT GCAAAGCATC CACTACGAAT TCCAGG|TGAT GCAAAAAATA
W C A H A K H P L R I P G D A K N

Exon 13 →

1401 CTTTATCAAA|TATTGGATCT TCTTGTTCTA GTTCCAGATA CTCCCCTAAC
T L S N I G S S C S S S R Y S P N

1451 TCTTCTTCAA AACTTTCTAT GCATGAAATC GCTATGCATG CAGAAAAAAC
S S S K L S M H E I A M H A E K

Exon 14 →

1501 TACTCCAATG AAGCCTAAAA CACTAAATCA AGC|GGAAGTT TGTATCAATA
T T P M K P K T L N Q A E V C I N

1551 TAGAATGTGT TTGTGGCCCT TACATTGATG TAGAAAAACT TAGAAAACTG
I E C V C G P Y I D V E K L R K L

1601 ACTCCAATTC GCAAAGGAGA AATCATACAT GTTCTTAAAG GAGTTTTAGA
T P I R K G E I I H V L K G V L

1651 AGACATCATT GCTTGTGCAA TATCACCAAA AACTGTTATA GGGTTTCTGA
E D I I A C A I S P K T V I G F L

Exon 15 →

1701 AGCCTGG|TAG AGGAAATTGT GTAATCACTA CTGTTTGTGA AGGCACAACA
K P G R G N C V I T T V C E G T T

1751 TTTCAATGCT GTCTTACTGT TATTGATCAT GTTTCATCAT TATGGGATAT
F Q C C L T V I D H V S S L W D

1801 TCTTAAACAG TTTGCAGACA ATTTAAGGTC TTGTCCTAAC TTATTTTCAA
I L K Q F A D N L R S C P N L F S

1851 CTGAAATGTA TGCTTCTTGT CCACGTTGTT TAAACAAACC TTCCACTGCC
T E M Y A S C P R C L N K P S T A

1901 ATCTCCATAC CATTGATTTA TCAGAGGTTA ACATCAAACC CTGGTTCTTA
I S I P L I Y Q R L T S N P G S

Exon 16 →

1951 TCTGGCTTCT TCTTTTCAAT ATCCTAAAAG|CTCTAGGTTT ATTGAATTTG
Y L A S S F Q Y P K S S R F I E F

2001 AAGATGGTGA CATTCATTCA GTTATGGATT TGGAAAAAAA TGAACCATTA
E D G D I H S V M D L E K N E P L

2051 AGTTTGAACC AGATAAAGTG TAATCAAGAC TCTCGGAATC CGGATAAAAA
S L N Q I K C N Q D S R N P D K

2101 TACTAAAAAA CATGTAACTT GGCCGGCAAT ATCTCAGGCT GAAGTATTTA
N T K K H V T W P A I S Q A E V F

2151 ATTATCAACC AGTTTATGCA GATACAGCAT CTCAAACAGA ATCCTTTTGT
N Y Q P V Y A D T A S Q T E S F C

2201 TTCACTTGTA ATTCTCTTAG TCATAAATGT AGTCTCACAG ACAAAAATAG
F T C N S L S H K C S L T D K N

2251 GTTAAAAGAT GGTGGTGTTG AAATTGAAAG TGCAGAAATA TTGAGTAATA
R L K D G G V E I E S A E I L S N

Exon 17 →

2301 TCCAAG|CTGC TGGTATCGAT ATTATGAATC GCCCTAGCAT GTGGAATGTG
I Q A A G I D I M N R P S M W N V

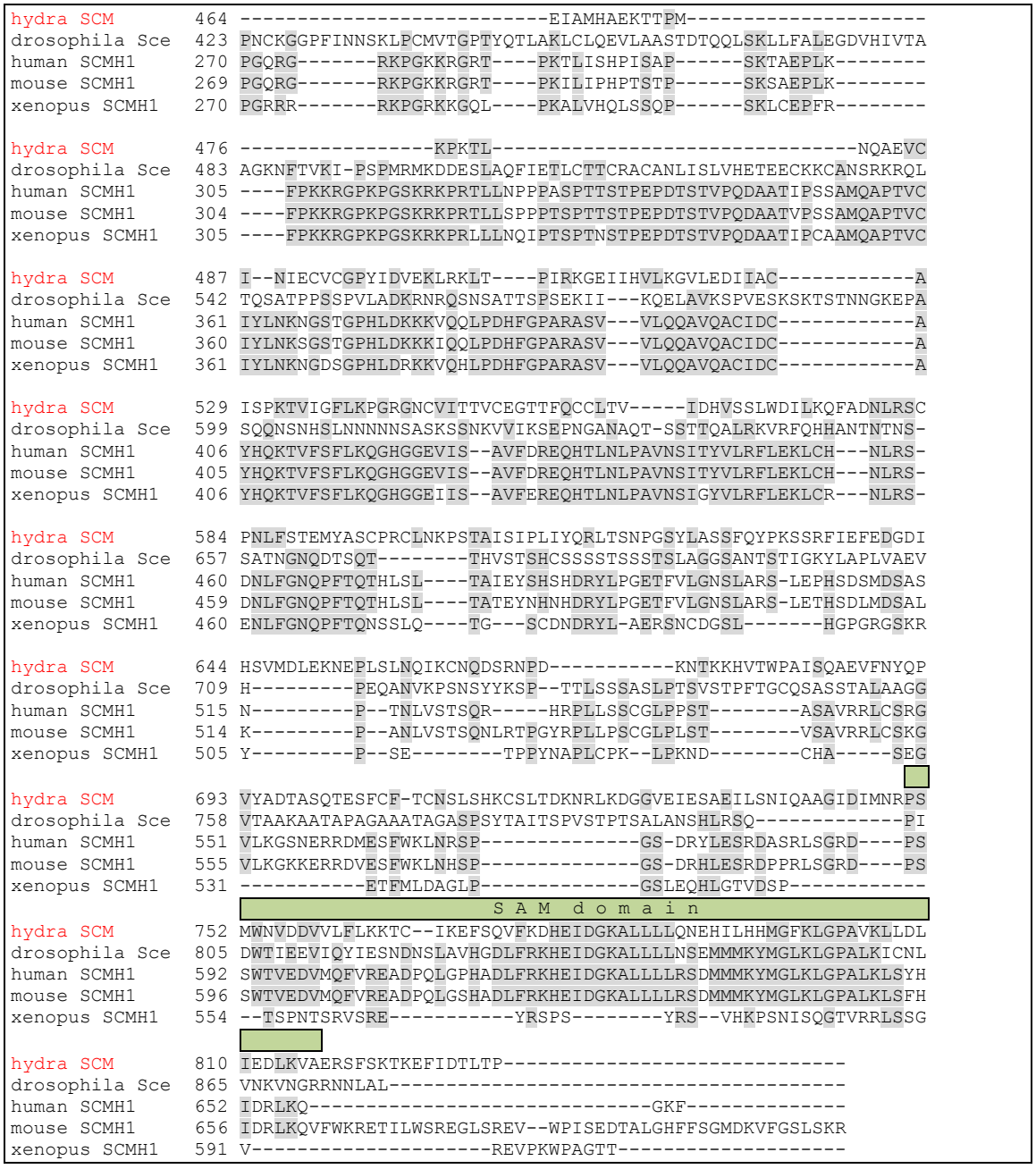
2351 GATGATGTTG TTCTTTTTTT AAAGAAAACT TGCATTAAAG AATTTTCTCA
D D V V L F L K K T C I K E F S

```
                            Exon 18 →
2401  AGTATTTAAA GACCAT | GAAA TTGATGGTAA AGCCCTACTT CTTTTGCAAA
      Q   V   F   K    D   H    E   I   D   G    K   A   L   L    L   L   Q

                                         SAM domain
2451  ATGAGCATAT TCTTCATCAT ATGGGTTTTA AACTCGGACC AGCCGTCAAA
        N   E   H    I   L   H   H    M   G   F    K   L   G    P   A   V   K

2501  CTTCTGGATT TGATTGAAGA TTTGAAGGTT GCGGAACGTT CTTTTTCAAA
         L   L   D    L   I   E    D   L   K   V    A   E   R    S   F   S

2551  AACGAAAGAG TTTATTGATA CTTTGACTCC ATAGgtgaat aagaaatcat
      K   T   K   E    F   I   D    T   L   T    P   *
2601  ttttaaataa cagcaagcta atatatatga atataataaa gtaataaata
2651  aaaaaaaaaa aaaaaa
```

Protein alignment:

Comparison of *Hy*SCM with SCM protein sequences from other species reveals a conservation on the amino acid level of 20%. Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

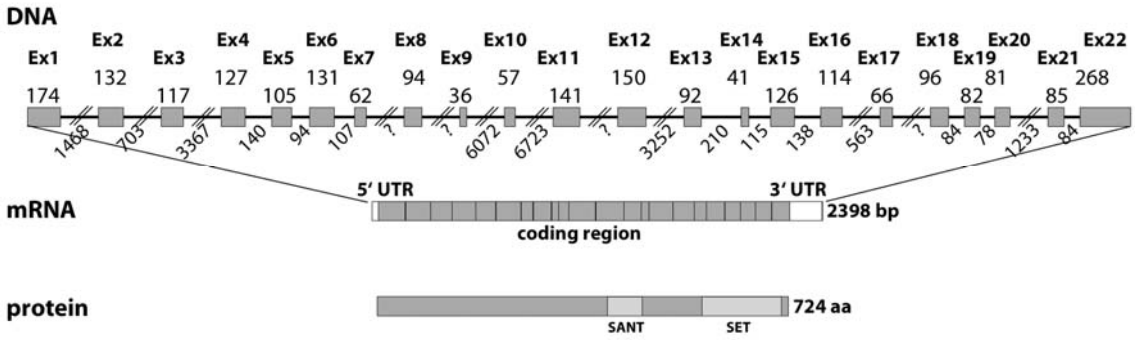
```
hydra SCM          1 -MEMINANTINLIKSELSSPPLKRKKSFIEDSNQDATDSEQVSEHNTSRTNNGHFEMTNL
drosophila Sce     1 MSGGRDSSTSSGSNSAAPGASTNATSSASASASSTSTSASPGSTTSPASTQRQRGRPAKR
human SCMH1        1 -----------------------------------------------------------M
mouse SCMH1        1 -----------------------------------------------------------M
xenopus SCMH1      1 -----------------------------------------------------------M
                                             H M G   b o x
hydra SCM         60 PKKPLTPYMSYRNMVFQDIKHKHPGCSPYDLACVISAKWSLLS--DVERKVWEVKYEKEM
drosophila Sce    61 ATCTWCGEGKLPLQYVLPTQTGKKEFCSETCIAEFRKAYSKGACTQCDNVIRDGAPNKEF
human SCMH1        2 LVC-----------------------------------YSVLAC----EILWD------L
mouse SCMH1        2 LVC-----------------------------------YSVLAC----ESLWD------L
xenopus SCMH1      2 RKP-----------------------------------GHLKAT----D--WKDGRRHGR

hydra SCM        118 KSYEKKLRLYHQKALKRKTNEDKSFSPIGDEVHELNKFDTLDYCSGCKKLKELQGFSAEI
drosophila Sce   121 -C---SIMCMNKHQKKNCSTRHSGGSASG-------------------------------
human SCMH1       17 PC---SIM----------------GSPLG-------------------------------
mouse SCMH1       17 PC---SIM----------------GSPLG-------------------------------
xenopus SCMH1     21 PS---DTH----------------ANQ-G-------------------------------

hydra SCM        178 NGRQVYRGDLCSKACFLEASKQLKCVDGSTYIKHSNSDPVSVNLSPFNWAEYLSETNSEP
drosophila Sce   146 -----------------KGLAESERKLLASGAPAPTGPFQYESFHVFDWDAYLEETGSEA
human SCMH1       27 -----------------H----------------------------FTWDKYLKETCSVP
mouse SCMH1       27 -----------------H----------------------------FTWDKYLKETCSVP
xenopus SCMH1     30 -----------------H----------------------------FSWEKYLKETGSQP
                                             M B T   d o m a i n
hydra SCM        238 ADWMYCRQASY-PPSNAFSVGMKLEARRHDQIKDSNQIIFSLATVAFMIGPRLLLHFDGS
drosophila Sce   189 AP-AKCFKQAQNPPNNDFKIGMKLEAL-----DPRNVTSTCIATVVGVLGSRLRLRLDGS
human SCMH1       42 AP-VHCFKQSYTPPSNEFKISMKLEAQ-----DPRNTTSTCIATVVGLTGARLRLRLDGS
mouse SCMH1       42 AP-VHCFKQSYTPPSNEFKISMKLEAQ-----DPRNTTSTCIATVVGLTGARLRLRLDGS
xenopus SCMH1     45 AP-PHCFRQSCSPPSNDYKIGMKLEAQ-----DPRNTTSTCIASVVGLTGARLRLRLDGS
                           M B T   d o m a i n                 M B T domain
hydra SCM        297 SSKYDIWTLCDSWDIHPVGWTDR--GSLLPPTGYRHDPKNYMQFYSLKIQKAELATARCF
drosophila Sce   243 DSQNDFWRLVDSTEIHAIGHCEKNGGMLQPPLGFRMNASSWPGYLCKILNNAMVAPEEIF
human SCMH1       96 DNKNDFWRLVDSAEIQPIGNCEKNGGMLQPPLGFRLNASSWPMFLLKTLNGAEMAPIRIF
mouse SCMH1       96 DNKNDFWRLVDSSEIQPIGNCEKNGGMLQPPLGFRLNASSWPMFLLKTLNGAEMAPIKIF
xenopus SCMH1     99 DNKNDFWRLVDSSEIQPIGTCEKSGGMLQPPLGFRLNASSWPMFLLKTLNGAEMAPAKVF
                                             M B T   d o m a i n
hydra SCM        355 KKVPSMPPKNCFEVGMKLEAVDKQNPSVIGVASVVRVQDELVYIEFDGYPG-LGYSSHYG
drosophila Sce   303 QPEPPEPEENLFKVGQKLEAVDKKNPQLICCATVDAIKDDQIHVTFDGWRGAFDYWCNYR
human SCMH1      156 HKEPPSPSHNFFKMGMKLEAVDRKNPHFICPATIGEVRGSEVLVTFDGWRGAFDYWCRFD
mouse SCMH1      156 HKEPPSPSHNFFKMGMKLEAVDRKNPHFICPATIGEVRGAEVLVTFDGWRGAFDYWCRFD
xenopus SCMH1    159 HKEPPSPPQNYFEIGMKLEAVDRKNPHFICPATIGELRGSEILVTFDGWRGAFDYWCRYD

hydra SCM        414 DRDFFPVGWCAHAKHPLRIPGDAKNTLSNIGSSCSSSRYSPNSSSKLSMH----------
drosophila Sce   363 SRDIFPAGWCARSCHPMQPPGHKSRMDSSSSKQRCPRPRYTVVAESEAMVPASPATAHFH
human SCMH1      216 SRDIFPVGWCSLTGDNLQPPGTKVVIPKN------PYPASDVNTEKPSIHSSTKTVLEHQ
mouse SCMH1      216 SRDIFPVGWCSLTGDNLQPPGTKVVIPKN------PSPSSDVSTEKPSIH-STKTVLEHQ
xenopus SCMH1    219 SRDIFPVGWCSLTGDNLQPPGTKVAIPKS------PLPAAGIAT---SMMRGPKNLLPVV
```


hydra SCM 464 ----------EIAMHAEKTTPM----------
drosophila Sce 423 PNCKGGPFINNSKLPCMVTGPTYQTLAKLCLQEVLAASTDTQQLSKLLFALEGDVHIVTA
human SCMH1 270 PGQRG-------RKPGKKRGRT----PKTLISHPISAP------SKTAEPLK--------
mouse SCMH1 269 PGQRG-------RKPGKKRGRT----PKILIPHPTSTP------SKSAEPLK--------
xenopus SCMH1 270 PGRRR-------RKPGRKKGQL----PKALVHQLSSQP------SKLCEPFR--------
hydra SCM 476 ----------KPKTL----------NQAEVC
drosophila Sce 483 AGKNFTVKI-PSPMRMKDDESLAQFIETLCTTCRACANLISLVHETEECKKCANSRKRQL
human SCMH1 305 ----FPKKRGPKPGSKRKPRTLLNPPPASPTTSTPEPDTSTVPQDAATIPSSAMQAPTVC
mouse SCMH1 304 ----FPKKRGPKPGSKRKPRTLLSPPPTSPTTSTPEPDTSTVPQDAATVPSSAMQAPTVC
xenopus SCMH1 305 ----FPKKRGPKPGSKRKPRLLLNQIPTSPTNSTPEPDTSTVPQDAATIPCAAMQAPTVC
hydra SCM 487 I--NIECVCGPYIDVEKLRKLT----PIRKGEIIHVLKGVLEDIIAC----------A
drosophila Sce 542 TQSATPPSSPVLADKRNRQSNSATTSPSEKII---KQELAVKSPVESKSKTSTNNGKEPA
human SCMH1 361 IYLNKNGSTGPHLDKKKVQQLPDHFGPARASV---VLQQAVQACIDC----------A
mouse SCMH1 360 IYLNKSGSTGPHLDKKKIQQLPDHFGPARASV---VLQQAVQACIDC----------A
xenopus SCMH1 361 IYLNKNGDSGPHLDRKKVQHLPDHFGPARASV---VLQQAVQACIDC----------A
hydra SCM 529 ISPKTVIGFLKPGRGNCVITTVCEGTTFQCCLTV-----IDHVSSLWDILKQFADNLRSC
drosophila Sce 599 SQQNSNHSLNNNNNSASKSSNKVVIKSEPNGANAQT-SSTTQALRKVRFQHHANTNTNS-
human SCMH1 406 YHQKTVFSFLKQGHGGEVIS--AVFDREQHTLNLPAVNSITYVLRFLEKLCH---NLRS-
mouse SCMH1 405 YHQKTVFSFLKQGHGGEVIS--AVFDREQHTLNLPAVNSITYVLRFLEKLCH---NLRS-
xenopus SCMH1 406 YHQKTVFSFLKQGHGGEIIS--AVFEREQHTLNLPAVNSIGYVLRFLEKLCR---NLRS-
hydra SCM 584 PNLFSTEMYASCPRCLNKPSTAISIPLIYQRLTSNPGSYLASSFQYPKSSRFIEFEDGDI
drosophila Sce 657 SATNGNQDTSQT--------THVSTSHCSSSSTSSSTSLAGGSANTSTIGKYLAPLVAEV
human SCMH1 460 DNLFGNQPFTQTHLSL----TAIEYSHSHDRYLPGETFVLGNSLARS-LEPHSDSMDSAS
mouse SCMH1 459 DNLFGNQPFTQTHLSL----TATEYNHNHDRYLPGETFVLGNSLARS-LETHSDLMDSAL
xenopus SCMH1 460 ENLFGNQPFTQNSSLQ----TG---SCDNDRYL-AERSNCDGSL-------HGPGRGSKR
hydra SCM 644 HSVMDLEKNEPLSLNQIKCNQDSRNPD----------KNTKKHVTWPAISQAEVFNYQP
drosophila Sce 709 H---------PEQANVKPSNSYYKSP--TTLSSSASLPTSVSTPFTGCQSASSTALAAGG
human SCMH1 515 N---------P--TNLVSTSQR-----HRPLLSSCGLPPST--------ASAVRRLCSRG
mouse SCMH1 514 K---------P--ANLVSTSQNLRTPGYRPLLPSCGLPLST--------VSAVRRLCSKG
xenopus SCMH1 505 Y---------P--SE--------TPPYNAPLCPK--LPKND--------CHA-----SEG
hydra SCM 693 VYADTASQTESFCF-TCNSLSHKCSLTDKNRLKDGGVEIESAEILSNIQAAGIDIMNRPS
drosophila Sce 758 VTAAKAATAPAGAAATAGASPSYTAITSPVSTPTSALANSHLRSQ----------PI
human SCMH1 551 VLKGSNERRDMESFWKLNRSP----------GS-DRYLESRDASRLSGRD----PS
mouse SCMH1 555 VLKGKKERRDVESFWKLNHSP----------GS-DRHLESRDPPRLSGRD----PS
xenopus SCMH1 531 ----------ETFMLDAGLP----------GSLEQHLGTVDSP----------
SAM domain
hydra SCM 752 MWNVDDVVLFLKKTC--IKEFSQVFKDHEIDGKALLLLQNEHILHHMGFKLGPAVKLLDL
drosophila Sce 805 DWTIEEVIQYIESNDNSLAVHGDLFRKHEIDGKALLLLNSEMMMKYMGLKLGPALKICNL
human SCMH1 592 SWTVEDVMQFVREADPQLGPHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLSYH
mouse SCMH1 596 SWTVEDVMQFVREADPQLGSHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLSFH
xenopus SCMH1 554 --TSPNTSRVSRE----------YRSPS--------YRS--VHKPSNISQGTVRRLSSG
hydra SCM 810 IEDLKVAERSFSKTKEFIDTLTP----------
drosophila Sce 865 VNKVNGRRNNLAL----------
human SCMH1 652 IDRLKQ----------GKF----------
mouse SCMH1 656 IDRLKQVFWKRETILWSREGLSREV--WPISEDTALGHFFSGMDKVFGSLSKR
xenopus SCMH1 591 V----------REVPKWPAGTT----------

HyEZH2

From gene to protein:

The *HyEZH2* gene is composed of 22 exons and spans a genomic region of > 83 kb. The respective exon and intron lengths are indicated. Undetermined lengths are indicated by a question mark. Based on a ~ 2.4 kb mRNA, a protein of 724 aa is translated.

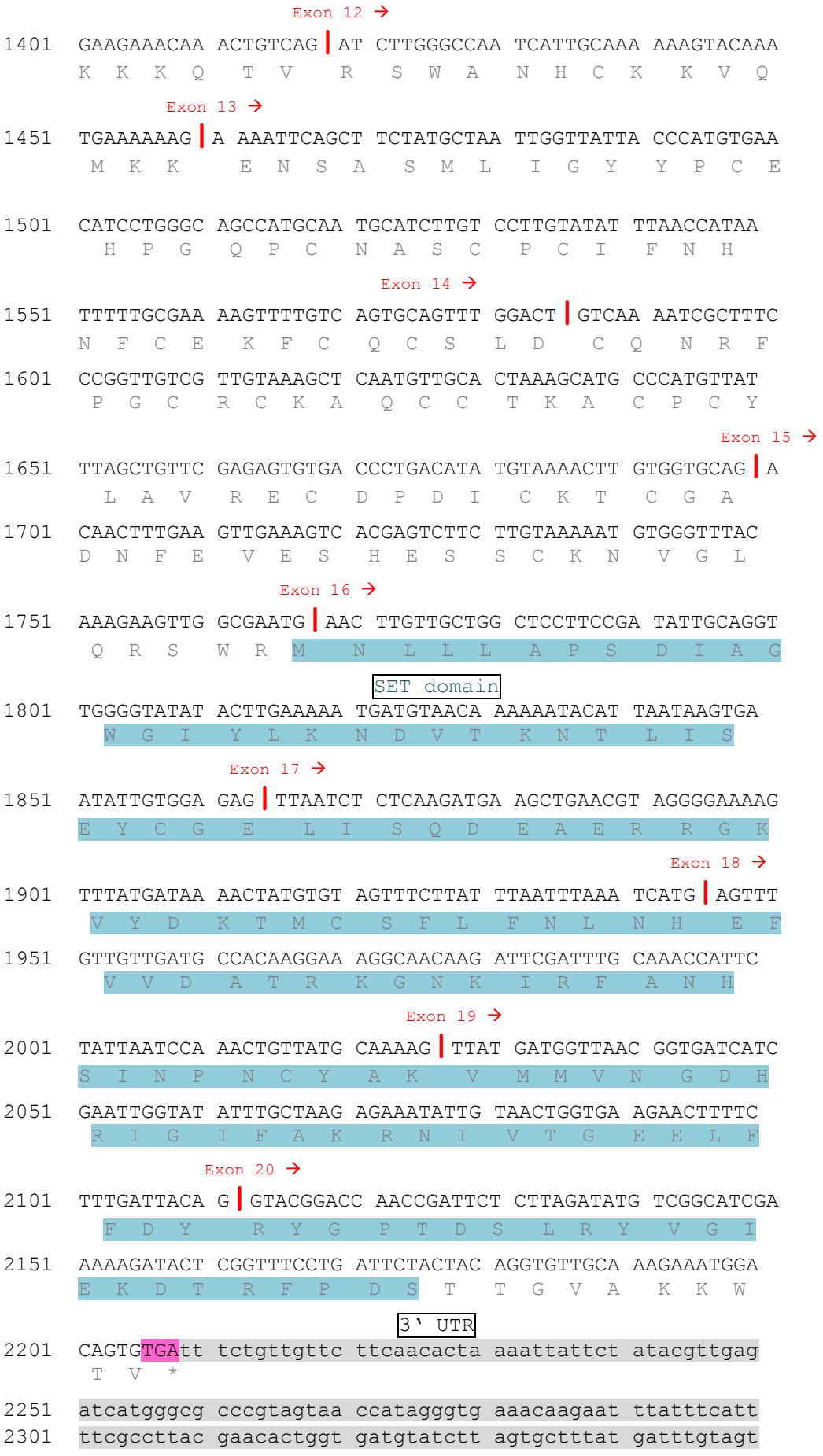


Sequence information:

Nucleotide and derived amino acid sequence of *HyEZH2* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The untranslated regions are highlighted in grey; the polyadenylation site is indicated by underlined nucleotides. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The SANT domain is highlighted in light red and the SET domain in light blue.

```
Exon 1 →                5' UTR
  1  aaaatatcct gtgtatagtt gtgagctaag aacATGCAAG AAGATATTGA
                                        M  Q   E  D  I

 51  AGCTAAAGAG AGACATTTAG ATCCTGCTTA CTTAGAAAAG ATTAATCATT
     E  A  K  E   R  H  L   D  P  A   Y  L  E  K   I  N  H

101  TTAAAAAAGT TATTCAGTTG GAATATGTTC GTATTTGTAG GAAGAGAAAA
      F  K  K   V  I  Q  L   E  Y  V   R  I  C   R  K  R  K

                          Exon 2
151  CAAAATAAAA CTGAAGAAGC AAAG|TCTATT TACAAAAGCA GCCGAGAAGA
       Q  N  K   T  E  E   A  K   S  I   Y  K  S   S  R  E

201  GTTGTTAAAT GTATTGAAAA AACGTCAGCA AATTTTAAAA GAGAAAAAGA
     E  L  L  N   V  L  K   K  R  Q   Q  I  L  K   E  K  K

251  GAGAGCCATT GACCACTATT CCTATTGCTA TTGTAACAGA GGCTTCAACA
      R  E  P   L  T  T  I   P  I  A   I  V  T   E  A  S  T

     Exon 3 →
301  AGAAAG|GCTC AAGTTCAAAG TGGTTTTAAC TATTTATCTC AAAATGTTAC
       R  K   A   Q  V  Q   S  G  F  N   Y  L  S   Q  N  V

351  ACTACAAACA ATTTACTCAC CTCCAGCATT GCCGACTATG ATCTCATGGG
     T  L  Q  T   I  Y  S   P  P  A   L  P  T  M   I  S  W

                          Exon 4 →
401  CTCCTCTACA ACAAAATTTT ATT|GTTGATG ATGAAACAGT TTTGCATAAT
      A  P  L   Q  Q  N  F   I   V  D   D  E  T   V  L  H  N
```

```
451  ATTCCTTACA TGGGAGAAGA TGTTATTGAC CAAGATGGGC AGTTTATTGA
       I  P  Y   M  G  E   D  V  I  D   Q  D  G   Q  F  I

501  AGAGCTTATT AAAAACTATG ATGGCAAAGT TCACACTTCT TCAAGTTTAG
     E  E  L  I   K  N  Y   D  G  K   V  H  T  S   S  S  L

  Exon 5 →
551  |ATAGTATTAT GGATGATGAA CTTTTTTTGG AATTAATAAA AGCTGGGATT
       D  S  I   M  D  D  E   L  F  L   E  L  I   K  A  G  I

601  GTTTACCAAG ATGAATATAG GAGCTCTGAG AATTATAAAG AACAGTTTTC
       V  Y  Q   D  E  Y   R  S  S  E   N  Y  K   E  Q  F

     Exon 6 →
651  AGACG|AATCT TTAACTAAAA AATTATCAAA CGATTTAAAA TTTCAAACTA
     S  D   E  S   L  T  K   K  L  S   N  D  L  K   F  Q  T

701  ATGATCCTAG CGAGTTACTT TTTGACACAA TAGCACATTA TTTCTCAGAG
       N  D  P   S  E  L  L   F  D  T   I  A  H   Y  F  S  E

                                          Exon 7 →
751  CAAGGTGTGA CTGCAAAAGA AGTAAAACAG AGGCAT|ATGC TATTAAAAGA
       Q  G  V   T  A  K   E  V  K  Q   R  H   M  L  L  K

                                                  Exon 8 →
801  AAAAGAAAGC AGTCAGCCGC CTGCTGAATG TACTCCAAAT ATTGATGG|CC
     E  K  E  S   S  Q  P   P  A  E   C  T  P  N   I  D   G

851  CAGATGTAAT AACAACGAAC AGGGAGAGAA CCATGCACTC TTACCATACA
       P  D  V   I  T  T  N   R  E  R   T  M  H   S  Y  H  T

901  TTGTTTTGTC GTAGATGTTA CAAGTATGAT TGCTTTCTTC ATGCTGTACG
       L  F  C   R  R  C   Y  K  Y  D   C  F  L   H  A  V

951  CCCTGGTCCA GTTAAAGTGC AAAGAAAGCA GTTAATGGAT ATACAAAATA
     R  P  G  P   V  K  V   Q  R  K   Q  L  M  D   I  Q  N

                                      'Exon 9' →
1001 TTGAGCCTTG TGGAGAAAAC TGCTATATGC ATTTG|GAAAG GTCTTTAACA
       I  E  P   C  G  E  N   C  Y  M   H  L   E   R  S  L  T

1051 ATTCAAGATA CAAAAGATTT ACCAAAAAGT TTACCAGAAA CAGATCGAAA
       I  Q  D   T  K  D   L  P  K  S   L  P  E   T  D  R

1101 TGAAGACTCA AAAATGAAGA AATTTAAGAG GAAAAAAAGT TTTAATAAAG
     N  E  D  S   K  M  K   K  F  K   R  K  K  S   F  N  K

                                  Exon 10 →
1151 CAGAAACATC AGTAGAAGTT TTAGCG|CCTG TTAAATCAAC ATACTTAATT
       A  E  T   S  V  E   V  L  A   P   V  K  S   T  Y  L  I

1201 GAAGATACAA ATGATGAACC ATGGACATCA TCTGATTTAT CAATGTTTCG
       E  D  T   N  D  E   P  W  T  S   S  D  L   S  M  F

                                SANT domain
1251 TGTGCTTATC AAGAATTTTC CAAATAATTA CTGCACAATA GCCCAACTAT
     R  V  L  I   K  N  F   P  N  N   Y  C  T  I   A  Q  L

                                  Exon 11 →
1301 TGAATTACTC CCATACATGT AAACAG|ATAT ATCGTCATGC TATGTTAGAA
     L  N  Y   S  H  T  C   K  Q   I   Y  R  H   A  M  L  E

1351 CCTCGTGATG AAAATCCTTC CAATGATGTT ATGACTCCTC CAACAAAAAA
       P  R  D   E  N  P   S  N  D  V   M  T  P   P  T  K
```

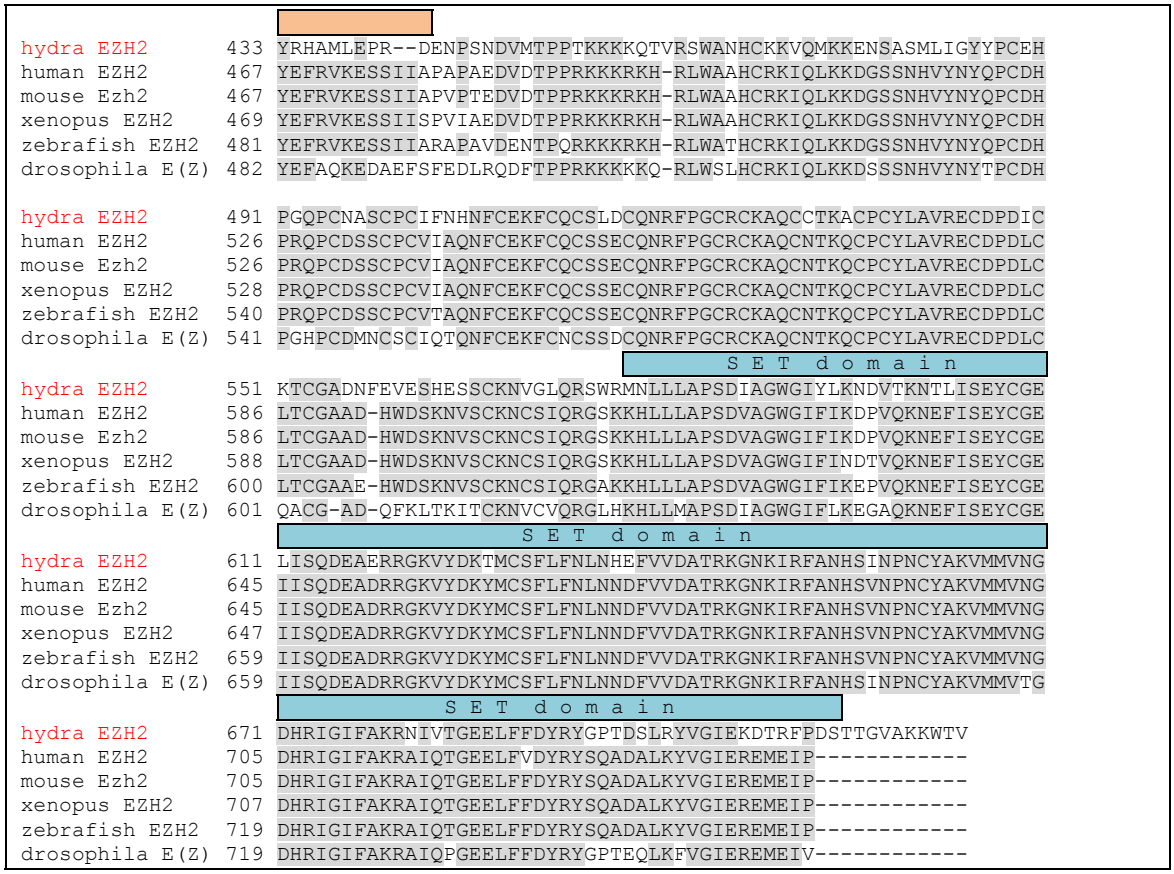

Exon 12 →
1401 GAAGAAACAA ACTGTCAG | AT CTTGGGCCAA TCATTGCAAA AAAGTACAAA
K K K Q T V R S W A N H C K K V Q
Exon 13 →
1451 TGAAAAAAG | A AAATTCAGCT TCTATGCTAA TTGGTTATTA CCCATGTGAA
M K K E N S A S M L I G Y Y P C E
1501 CATCCTGGGC AGCCATGCAA TGCATCTTGT CCTTGTATAT TTAACCATAA
H P G Q P C N A S C P C I F N H
Exon 14 →
1551 TTTTTGCGAA AAGTTTTGTC AGTGCAGTTT GGACT | GTCAA AATCGCTTTC
N F C E K F C Q C S L D C Q N R F
1601 CCGGTTGTCG TTGTAAAGCT CAATGTTGCA CTAAAGCATG CCCATGTTAT
P G C R C K A Q C C T K A C P C Y
Exon 15 →
1651 TTAGCTGTTC GAGAGTGTGA CCCTGACATA TGTAAAACTT GTGGTGCAG | A
L A V R E C D P D I C K T C G A
1701 CAACTTTGAA GTTGAAAGTC ACGAGTCTTC TTGTAAAAAT GTGGGTTTAC
D N F E V E S H E S S C K N V G L
Exon 16 →
1751 AAAGAAGTTG GCGAATG | AAC TTGTTGCTGG CTCCTTCCGA TATTGCAGGT
Q R S W R M N L L L A P S D I A G
SET domain
1801 TGGGGTATAT ACTTGAAAAA TGATGTAACA AAAAATACAT TAATAAGTGA
W G I Y L K N D V T K N T L I S
Exon 17 →
1851 ATATTGTGGA GAG | TTAATCT CTCAAGATGA AGCTGAACGT AGGGGAAAAG
E Y C G E L I S Q D E A E R R G K
Exon 18 →
1901 TTTATGATAA AACTATGTGT AGTTTCTTAT TTAATTTAAA TCATG | AGTTT
V Y D K T M C S F L F N L N H E F
1951 GTTGTTGATG CCACAAGGAA AGGCAACAAG ATTCGATTTG CAAACCATTC
V V D A T R K G N K I R F A N H
Exon 19 →
2001 TATTAATCCA AACTGTTATG CAAAAG | TTAT GATGGTTAAC GGTGATCATC
S I N P N C Y A K V M M V N G D H
2051 GAATTGGTAT ATTTGCTAAG AGAAATATTG TAACTGGTGA AGAACTTTTC
R I G I F A K R N I V T G E E L F
Exon 20 →
2101 TTTGATTACA G | GTACGGACC AACCGATTCT CTTAGATATG TCGGCATCGA
F D Y R Y G P T D S L R Y V G I
2151 AAAAGATACT CGGTTTCCTG ATTCTACTAC AGGTGTTGCA AAGAAATGGA
E K D T R F P D S T T G V A K K W
3' UTR
2201 CAGTGTGAtt tctgttgttc ttcaacacta aaattattct atacgttgag
T V *
2251 atcatgggcg cccgtagtaa ccatagggtg aaacaagaat ttatttcatt
2301 ttcgccttac gaacactggt gatgtatctt agtgctttat gatttgtagt

```
                                                    polyA site
2351  ttttatttag aatatgaaac gtacacacaa aaaaaaaaaa aaaaaaaa
```

Protein alignment:

Comparison of *Hy*EZH2 with EZH2 protein sequences from other species reveals significant conservation on the amino acid level (45%). Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

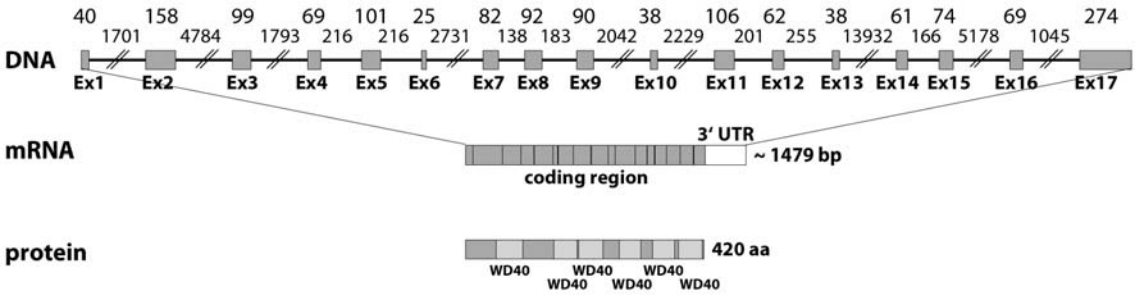
```
hydra EZH2          1 MQEDIEAKERHLDPAYLEKINHFKKVIQLEYVRICRKRKQNKTEEAKSIYKSSREELLN-
human EZH2          1 MGQ--TGKKSEKGPVC------WRKRVKSEYMRLRQLKRFRRADEVKSMFSSNRQKILER
mouse Ezh2          1 MGQ--TGKKSEKGPVC------WRKRVKSEYMRLRQLKRFRRADEVKTMFSSNRQKILER
xenopus EZH2        1 MGQ--TGKKSEKGPVC------WRKRVKSEYMRLRQLKRFRRADEVKSMFNTNRQKIMER
zebrafish EZH2      1 MGL--TGRKSEKGPVC------WRRRVKSEYMRLRQLKRFRRADEVKSMFSSNRQKILER
drosophila E(Z)     1 -------MNSTKVPPE------WKRRVKSEYIKIRQQKRYKRADEIKEAWIRNWDEHNHN

hydra EZH2         60 --VLKKRQQILKEKKREPLTTIPIAIVTEASTRKAQVQSG---FNYLSQNVTLQTIYSPP
human EZH2         53 TEILNQE---WKQRRIQPVH-ILTSVSSLRGTRECSVTSDL-DFPT--QVIPLKTLNAVA
mouse Ezh2         53 TETLNQE---WKQRRIQPVH-IMTSVSSLRGTRECSVTSDL-DFPA--QVIPLKTLNAVA
xenopus EZH2       53 TEILNQE---WKQRRIQPVH-IMTTVSSLRGTRECSVTSDL-DFPK--QVIPLKTLTAVA
zebrafish EZH2     53 TDILNQE---WKLRRIQPVH-IMTPVSSLRGTRECTVDSGFSEFSR--QVIPLKTLNAVA
drosophila E(Z)    48 VQDLYCESKVWQAKPYDPPH-VDC-------VKRAEVTS-YNGIPSGPQKVPICVINAVT

hydra EZH2        115 ALPTMISWAPLQQNFIVDDETVLHNIPYMGEDVIDQDGQFIEELIKNYDGKVHTSSSLDS
human EZH2        106 SVPIMYSWSPLQQNFMVEDETVLHNIPYMGDEVLDQDGTFIEELIKNYDGKVHGDREC-G
mouse Ezh2        106 SVPIMYSWSPLQQNFMVEDETVLHNIPYMGDEVLDQDGTFIEELIKNYDGKVHGDREC-G
xenopus EZH2      106 SVPIMYSWSPLQQNFMVEDETVLHNIPYMGDEVLDQDGTFIEELIKNYDGKVHGDREC-G
zebrafish EZH2    107 SVPVMYSWSPLQQNFMVEDETVLHNIPYMGDEILDQDGTFIEELIKNYDGKVHGDREC-G
drosophila E(Z)    99 PIPTMYTWAPTQQNFMVEDETVLHNIPYMGDEVLDKDGKFIEELIKNYDGKVHGDKDP-S

hydra EZH2        175 IMDDELFLELIKA---------------------------GIVYQDEYRSSENYKEQFSD
human EZH2        165 FINDEIFVELVNAL--------------------------GQYNDDDD--DDDGDD--PE
mouse Ezh2        165 FINDEIFVELVNAL--------------------------GQYNDDDD--DDDGDD--PD
xenopus EZH2      165 FINDEIFVELVNAL--------------------------AQYSDYED--DEDGDDNQDD
zebrafish EZH2    166 FINDEIFVELVNAL--------------------------NQYSDNEE--DDEEDDHH-D
drosophila E(Z)   158 FMDDAIFVELVHALMRSYSKELEEAAPSTSTAIKTEPLAKSKQGEDDGVVDVDADCESPM

hydra EZH2        208 ESLTKKLSNDLKFQTND---------------------------PSELLFDTIAHYFSEQ
human EZH2        195 --EREEKQKDLEDHRDDKESRPP------------RK------FPSDKILEAISSMFPDK
mouse Ezh2        195 --EREEKQKDLEDNRDDKETCPP------------RK------FPADKIFEAISSMFPDK
xenopus EZH2      197 --EQDDTAKDQDDNMEDKETQPL------------RK------FPSDKIFEAISSMFPDK
zebrafish EZH2    197 --YKFEKM-DLCDGKDDAEDHKEQLSSESHNNDGSKK------FPSDKIFEAISSMFPDK
drosophila E(Z)   218 KLEKTESKGDLTD-VEKKETEEPVETEDADVKPAVEEVKDKLPFPAPIIFQAISANFPDK

hydra EZH2        241 GVTAKEVKQRHMLLKEKE-SSQPPAECTPNIDGPDVITTNRERTMHSYHTLFCRRCYKYD
human EZH2        235 G-TAEELKEKYKELTEQQLPGALPPECTPNIDGPNAKSVQREQSLHSFHTLFCRRCFKYD
mouse Ezh2        235 G-TAEELKEKYKELTEQQLPGALPPECTPNIDGPNAKSVQREQSLHSFHTLFCRRCFKYD
xenopus EZH2      237 G-TSEELKEKYKELTEQQLPGALPPECTPNIDGPNAKSVQREQSLHSFHTLFCRRCFKYD
zebrafish EZH2    248 G-STEELKEKYKELTEQQLPGALPPECTPNIDGPNAKSVQREQSLHSFHTLFCRRCFKYD
drosophila E(Z)   277 G-TAQELKEKYIELTEHQDP-ERPQECTPNIDGIKAESVSRERTMHSFHTLFCRRCFKYD

hydra EZH2        300 CFLHAV--RPGPVKVQRKQLMDIQNIEPCGENCYMHLER---------SLTIQDTKDLPK
human EZH2        294 CFLHPF--HATPNTYKRKNTETALDNKPCGPQCYQHL--EG-AKEFAAALTAERIKTPPK
mouse Ezh2        294 CFLHPF--HATPNTYKRKNTETALDNKPCGPQCYQHL--EG-AKEFAAALTAERIKTPPK
xenopus EZH2      296 CFLHPF--HATPNTYKRKNNEAANDGKPCGPHCYQLL--EG-AREFAAALTAERIKTPPK
zebrafish EZH2    307 CFLHPF--QATPNTYKRKNMENLVDSKPCGIYCYMYMVQDGMVREYPAGVVPERAKTPSK
drosophila E(Z)   335 CFLHRLQGHAGPNLQKRRYPELKPFAEPCSNSCY--MLIDGMKEKLAA------------

hydra EZH2        349 --------------SLPET-------------DR---------NEDSKMKKFKRKKSFNK
human EZH2        349 RPGGRRRGRLPNNSSRPSTPTINVLESKDTDSDREAGTETGGENNDKEEEEKKDE-TSSS
mouse Ezh2        349 RPGGRRRGRLPNNSSRPSTPTISVLESKDTDSDREAGTETGGENNDKEEEEKKDE-TSSS
xenopus EZH2      351 RPSGRRRGRLPNNTSRPSTPTVNVSEAKDTDSDREAGTETGGESNDKEEEEKKDE-TSSS
zebrafish EZH2    365 RSTGRRRGRLPNSNSRPSTPTVN-SETKDTDSDREGGAD-GNDSNDKDDDKKDE-TTSS
drosophila E(Z)   381 ------------DSKTPPIDSCNEASSEDSND-----SNSQFSNKDFNHENSKDNGLTVN
                                                  SANT domain
hydra EZH2        373 AETSVEVLAPVKSTYLIEDTNDEPWTSSDLSMFRVLIKNFPNNYCTIAQLLNYSHTCKQI
human EZH2        408 SEANSRCQTPIKMKPNIEPPENVEWSGAEASMFRVLIGTYYDNFCAIARLIG-TKTCRQV
mouse Ezh2        408 SEANSRCQTPIKMKPNIEPPENVEWSGAEASMFRVLIGTYYDNFCAIARLIG-TKTCRQV
xenopus EZH2      410 SEANSRCQTPIKMKPNIEPPENVEWSGAEASLFRVLIGTYYDNFCAIARLIG-TKTCRQV
zebrafish EZH2    422 SEANSRCQTPVKLKLSSEPPENVDWSGAEASLFRVLIGTYYDNFCAIARLIG-TKTCRQV
drosophila E(Z)   424 SAAVAEINSIMAGMMNITSTQCV-WTGADQALYRVLHKVYLKNYCAIAHNML-TKTCRQV
```


SET domain
SET domain
SET domain
hydra EZH2 433 YRHAMLEPR--DENPSNDVMTPPTKKKKQTVRSWANHCKKVQMKKENSASMLIGYYPCEH
human EZH2 467 YEFRVKESSIIAPAPAEDVDTPPRKKKRKH-RLWAAHCRKIQLKKDGSSNHVYNYQPCDH
mouse Ezh2 467 YEFRVKESSIIAPVPTEDVDTPPRKKKRKH-RLWAAHCRKIQLKKDGSSNHVYNYQPCDH
xenopus EZH2 469 YEFRVKESSIISPVIAEDVDTPPRKKKRKH-RLWAAHCRKIQLKKDGSSNHVYNYQPCDH
zebrafish EZH2 481 YEFRVKESSIIARAPAVDENTPQRKKKRKH-RLWATHCRKIQLKKDGSSNHVYNYQPCDH
drosophila E(Z) 482 YEFAQKEDAEFSFEDLRQDFTPPRKKKKKQ-RLWSLHCRKIQLKKDSSSNHVYNYTPCDH
hydra EZH2 491 PGQPCNASCPCIFNHNFCEKFCQCSLDCQNRFPGCRCKAQCCTKACPCYLAVRECDPDIC
human EZH2 526 PRQPCDSSCPCVIAQNFCEKFCQCSSECQNRFPGCRCKAQCNTKQCPCYLAVRECDPDLC
mouse Ezh2 526 PRQPCDSSCPCVIAQNFCEKFCQCSSECQNRFPGCRCKAQCNTKQCPCYLAVRECDPDLC
xenopus EZH2 528 PRQPCDSSCPCVIAQNFCEKFCQCSSECQNRFPGCRCKAQCNTKQCPCYLAVRECDPDLC
zebrafish EZH2 540 PRQPCDSSCPCVTAQNFCEKFCQCSSECQNRFPGCRCKAQCNTKQCPCYLAVRECDPDLC
drosophila E(Z) 541 PGHPCDMNCSCIQTQNFCEKFCNCSSDCQNRFPGCRCKAQCNTKQCPCYLAVRECDPDLC
hydra EZH2 551 KTCGADNFEVESHESSCKNVGLQRSWRMNLLLAPSDIAGWGIYLKNDVTKNTLISEYCGE
human EZH2 586 LTCGAAD-HWDSKNVSCKNCSIQRGSKKHLLLAPSDVAGWGIFIKDPVQKNEFISEYCGE
mouse Ezh2 586 LTCGAAD-HWDSKNVSCKNCSIQRGSKKHLLLAPSDVAGWGIFIKDPVQKNEFISEYCGE
xenopus EZH2 588 LTCGAAD-HWDSKNVSCKNCSIQRGSKKHLLLAPSDVAGWGIFINDTVQKNEFISEYCGE
zebrafish EZH2 600 LTCGAAE-HWDSKNVSCKNCSIQRGAKKHLLLAPSDVAGWGIFIKEPVQKNEFISEYCGE
drosophila E(Z) 601 QACG-AD-QFKLTKITCKNVCVQRGLHKHLLMAPSDIAGWGIFLKEGAQKNEFISEYCGE
hydra EZH2 611 LISQDEAERRGKVYDKTMCSFLFNLNHEFVVDATRKGNKIRFANHSINPNCYAKVMMVNG
human EZH2 645 IISQDEADRRGKVYDKYMCSFLFNLNNDFVVDATRKGNKIRFANHSVNPNCYAKVMMVNG
mouse Ezh2 645 IISQDEADRRGKVYDKYMCSFLFNLNNDFVVDATRKGNKIRFANHSVNPNCYAKVMMVNG
xenopus EZH2 647 IISQDEADRRGKVYDKYMCSFLFNLNNDFVVDATRKGNKIRFANHSVNPNCYAKVMMVNG
zebrafish EZH2 659 IISQDEADRRGKVYDKYMCSFLFNLNNDFVVDATRKGNKIRFANHSVNPNCYAKVMMVNG
drosophila E(Z) 659 IISQDEADRRGKVYDKYMCSFLFNLNNDFVVDATRKGNKIRFANHSINPNCYAKVMMVTG
hydra EZH2 671 DHRIGIFAKRNIVTGEELFFDYRYGPTDSLRYVGIEKDTRFPDSTTGVAKKWTV
human EZH2 705 DHRIGIFAKRAIQTGEELFVDYRYSQADALKYVGIEREMEIP----------
mouse Ezh2 705 DHRIGIFAKRAIQTGEELFFDYRYSQADALKYVGIEREMEIP----------
xenopus EZH2 707 DHRIGIFAKRAIQTGEELFFDYRYSQADALKYVGIEREMEIP----------
zebrafish EZH2 719 DHRIGIFAKRAIQTGEELFFDYRYSQADALKYVGIEREMEIP----------
drosophila E(Z) 719 DHRIGIFAKRAIQPGEELFFDYRYGPTEQLKFVGIEREMEIV----------

HyEED

From gene to protein:

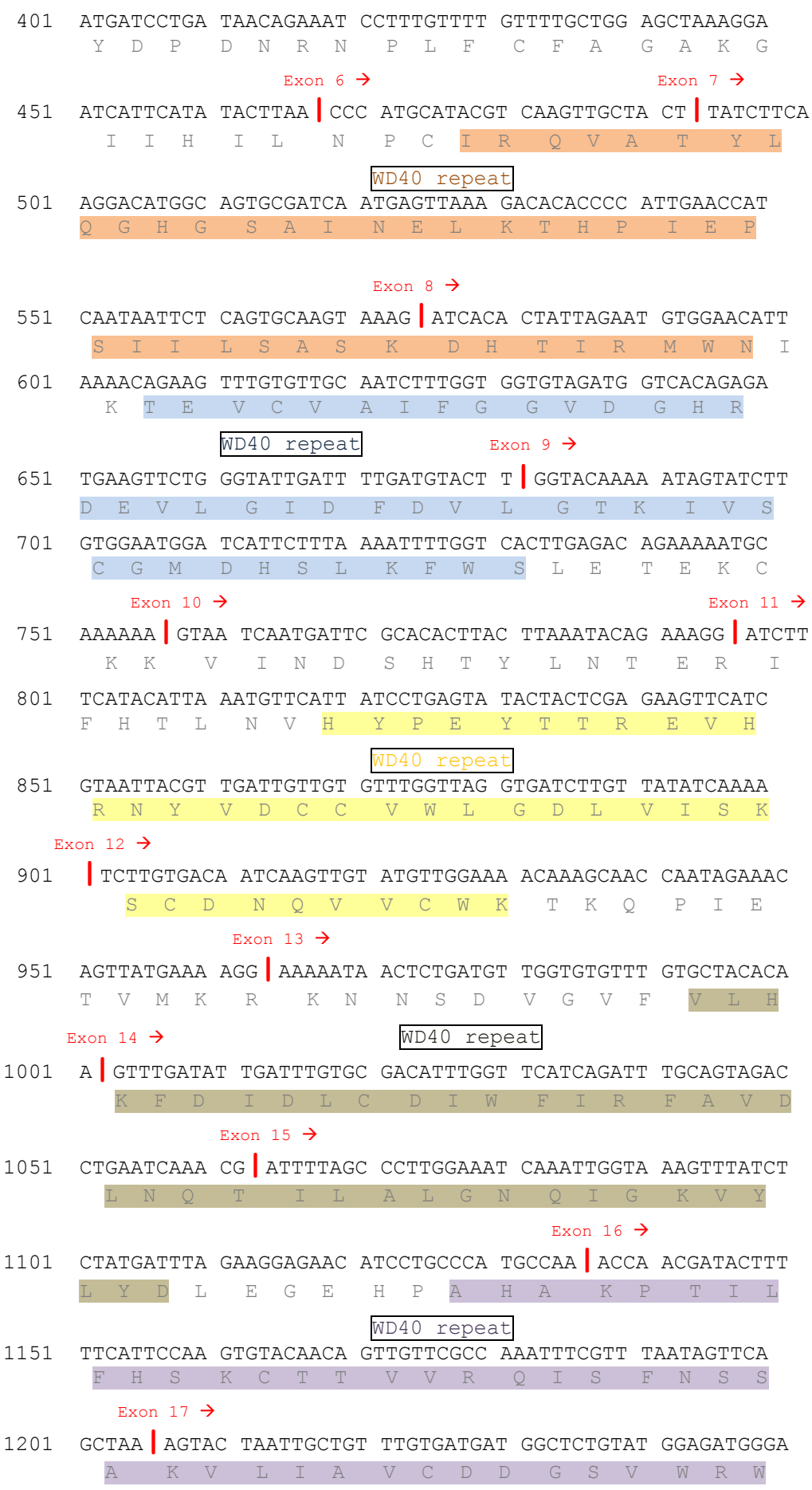
The *HyEED* gene is composed of 17 exons and spans a genomic region of > 38 kb. The respective exon and intron lengths are indicated. Undetermined lengths are indicated by a question mark. Based on a ~ 1.5 kb mRNA, a protein of 420 aa is translated.



Sequence information:

Nucleotide and derived amino acid sequence of *HyEED* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. No 5′ and 3′ untranslated region have been determined for this gene. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The six WD40 domains are highlighted in light green, red, blue, yellow, brown and blue lilac.

```
Exon 1 →                                                    Exon 2 →
  1  ATGGATACTG TTAATGATTT TATCGACGAT AGTTATCGTG | ATAATGATGA
       M  D  T   V  N  D   F  I  D  D   S  Y  R   D  N  D
 51  CGATACTGAC AGTAGTCAAA CTTATTCAAC ATCCAGTAAA ACAAAAGGAA
     D  D  T  D   S  S  Q   T  Y  S   T  S  S  K   T  K  G
101  GAAAGGTATA CAAAAAAAAA TGGAAAAAGC AAAATAAACA TGATGAGAAT
      R  K  V   Y  K  K  K   W  K  K   Q  N  K   H  D  E  N
                                                        Exon 3 →
151  TCAACTGGAA AATCAAAACT AAATTTTAAA TGCACTAATT TTATTAAG | GA
       S  T  G   K  S  K   L  N  F  K   C  T  N   F  I  K
                              WD40 repeat
201  GGATCACAAA CAACCGATTT TTGGTGTTCA ATTTTACCAA CAATGCTCCA
     E  D  H  K   Q  P  I   F  G  V   Q  F  Y  Q   Q  C  S
                                                    Exon 4 →
251  GTGGAGAAGA TGATCCATTA ATTTTTGGAA CTGTTGGAAG CAATAGG | GTT
      S  G  E   D  D  P  L   I  F  G   T  V  G   S  N  R   V
301  TCAATTTACA AATGTGCTGA AGATTCTGGA CAAATTATAT TGCTTCAAAG
       S  I  Y   K  C  A   E  D  S  G   Q  I  I   L  L  Q
                        Exon 5 →
351  CTATGCTGAT AGTGAT | CCAG AAGAAAGTTT CTATGCATGT TCATGGACAT
     S  Y  A  D   S  D   P   E  E  S   F  Y  A  C   S  W  T
```


401 ATGATCCTGA TAACAGAAAT CCTTTGTTTT GTTTTGCTGG AGCTAAAGGA
Y D P D N R N P L F C F A G A K G
Exon 6 →
Exon 7 →
451 ATCATTCATA TACTTAA | CCC ATGCATACGT CAAGTTGCTA CT | TATCTTCA
I I H I L N P C I R Q V A T Y L
WD40 repeat
501 AGGACATGGC AGTGCGATCA ATGAGTTAAA GACACACCCC ATTGAACCAT
Q G H G S A I N E L K T H P I E P
Exon 8 →
551 CAATAATTCT CAGTGCAAGT AAAG | ATCACA CTATTAGAAT GTGGAACATT
S I I L S A S K D H T I R M W N I
601 AAAACAGAAG TTTGTGTTGC AATCTTTGGT GGTGTAGATG GTCACAGAGA
K T E V C V A I F G G V D G H R
WD40 repeat
Exon 9 →
651 TGAAGTTCTG GGTATTGATT TTGATGTACT T | GGTACAAAA ATAGTATCTT
D E V L G I D F D V L G T K I V S
701 GTGGAATGGA TCATTCTTTA AAATTTTGGT CACTTGAGAC AGAAAAATGC
C G M D H S L K F W S L E T E K C
Exon 10 →
Exon 11 →
751 AAAAAA | GTAA TCAATGATTC GCACACTTAC TTAAATACAG AAAGG | ATCTT
K K V I N D S H T Y L N T E R I
801 TCATACATTA AATGTTCATT ATCCTGAGTA TACTACTCGA GAAGTTCATC
F H T L N V H Y P E Y T T R E V H
WD40 repeat
851 GTAATTACGT TGATTGTTGT GTTTGGTTAG GTGATCTTGT TATATCAAAA
R N Y V D C C V W L G D L V I S K
Exon 12 →
901 | TCTTGTGACA ATCAAGTTGT ATGTTGGAAA ACAAAGCAAC CAATAGAAAC
S C D N Q V V C W K T K Q P I E
Exon 13 →
951 AGTTATGAAA AGG | AAAAATA ACTCTGATGT TGGTGTGTTT GTGCTACACA
T V M K R K N N S D V G V F V L H
Exon 14 →
WD40 repeat
1001 A | GTTTGATAT TGATTTGTGC GACATTTGGT TCATCAGATT TGCAGTAGAC
K F D I D L C D I W F I R F A V D
Exon 15 →
1051 CTGAATCAAA CG | ATTTTAGC CCTTGGAAAT CAAATTGGTA AAGTTTATCT
L N Q T I L A L G N Q I G K V Y
Exon 16 →
1101 CTATGATTTA GAAGGAGAAC ATCCTGCCCA TGCCAA | ACCA ACGATACTTT
L Y D L E G E H P A H A K P T I L
WD40 repeat
1151 TTCATTCCAA GTGTACAACA GTTGTTCGCC AAATTTCGTT TAATAGTTCA
F H S K C T T V V R Q I S F N S S
Exon 17 →
1201 GCTAA | AGTAC TAATTGCTGT TTGTGATGAT GGCTCTGTAT GGAGATGGGA
A K V L I A V C D D G S V W R W

```
                                                3' UTR
1251  CAAGCAGACA TAAtgaaaaa atcttgtgta aatagtaaaa atgaagctat
      D  K  Q  T   *
1301  aatgacgaaa ctgctagacg ctgtgaaaat cgcaatattt tcagaggaga
1351  ataacaaaca attgattgtt agtatgaaaa taattaatgc gaaaatgttt
1401  caaaggaata aaatcagttt ttaataataa ctgtataaga taagaaaaat
1451  gtacattttt tagtaaataa aataaagtt
```

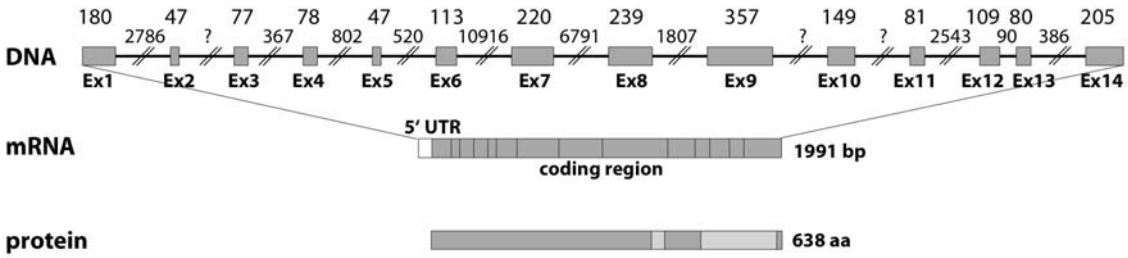
Protein alignment:

Comparison of *Hy*EED with EED protein sequences from other species reveals significant conservation on the amino acid level (47%). Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

```
hydra EED          1 -------------------------MDTVNDFIDDSYRDNDDD----------TDSSQTY
mouse Eed          1 MSEREVSTAPAGTDMPAAKKQKLSSDE--NSNPDLSGDEN--DDAVSIESGTNTERPDTP
human EED          1 MSEREVSTAPAGTDMPAAKKQKLSSDE--NSNPDLSGDEN--DDAVSIESGTNTERPDTP
xenopus EED        1 --------------MP-AKKQKLSSDE--NSNPELSGDEN--DDSVSIESGTNTERPDTP
zebrafish EED      1 --------------MP-NKKQKLSSDE--NSNPDLSGDDN--DDAVSVESGTHPERPDTP
drosophila ESC     1 ----------------------MSSDKVKNGNEPEESEESCGDESASYTTNSTTSRSKSP
                                                 W D 4 0  r e p e a t
hydra EED         26 STSSKTKGRKVYKK-KWKKQNKHDENSTGKSKLNFKCTNFIKEDHKQPIFGVQFYQQCSS
mouse Eed         57 TNTPNAPGRKSWGKGKWK--------SK-KCKYSFKCVNSLKEDHNQPLFGVQF--NWHS
human EED         57 TNTPNAPGRKSWGKGKWK--------SK-KCKYSFKCVNSLKEDHNQPLFGVQF--NWHS
xenopus EED       42 TNAANAPGRKGWGKGKWK--------SK-KCKYSFKCVNSLKEDHNQPLFGVQF--NWHS
zebrafish EED     42 TNTASAPGRKSWGKGKWK--------SK-KCKYSFKCVNSLKEDHGQPLFGVQF--NWHS
drosophila ESC    39 SSSTRSKRR---GRRSTK--------SKPKSRAAYKYDTHVKENHGANIFGVAF--NTLL
                     W D 4 0  r e p e a t
hydra EED         85 GEDDPLIFGTVGSNRVSIYKCAEDSGQIILLQSYADSDPEESFYACSWTYDPDNRNPLFC
mouse Eed        106 KEGDPLVFATVGSNRVTLYEC-HSQGEIRLLQSYVDADADENFYTCAWTYDSNTSHPLLA
human EED        106 KEGDPLVFATVGSNRVTLYEC-HSQGEIRLLQSYVDADADENFYTCAWTYDSNTSHPLLA
xenopus EED       91 KEGDPLVFATVGSNRVTLYEC-HPQGDIRLLQSYVDADADENFYTCAWTYDSNTSHPLLA
zebrafish EED     91 KEGDPLVFATVGSNRVTLYEC-HSQGEIRLLQSYVDADADENFYTCAWTFDCSSSHPLLA
drosophila ESC    86 GKDEPQVFATAGSNRVTVYEC-PRQGGMQLLHCYADPDPDEVFYTCAWSYDLKTSSPLLA
                                  W D 4 0  r e p e a t
hydra EED        145 FAGAKGIIHILNPCIRQVATYLQGHGSAINELKTHPIEPSIILSASKDHTIRMWNIKTEV
mouse Eed        165 VAGSRGIIRIINPITMQCIKHYVGHGNAINELKFHPRDPNLLLSVSKDHALRLWNIQTDT
human EED        165 VAGSRGIIRIINPITMQCIKHYVGHGNAINELKFHPRDPNLLLSVSKDHALRLWNIQTDT
xenopus EED      150 VAGSRGIIRIINPITMQCIKHYVGHGNAINELKFHPRDPNLLLSVSKDHALRLWNIQTDT
zebrafish EED    150 VAGSRGIIRIINHITMQCVKHYVGHGNAINELKFHPRDPNLLLSVSKDHALRLWNIQTDT
drosophila ESC   145 AAGYRGVIRVIDVEQNEAVGNYIGHGQAINELKFHPHKLQLLLSGSKDHAIRLWNIQSHV
                           W D 4 0  r e p e a t
hydra EED        205 CVAIFGGVDGHRDEVLGIDFDVLGTKIVSCGMDHSLKFWSLETEKCKKVINDSHTYL--N
mouse Eed        225 LVAIFGGVEGHRDEVLSADYDLLGEKIMSCGMDHSLKLWRINSKRMMNAIKESYDYNPNK
human EED        225 LVAIFGGVEGHRDEVLSADYDLLGEKIMSCGMDHSLKLWRINSKRMMNAIKESYDYNPNK
xenopus EED      210 LVAIFGGVEGHRDEVLSADYDLLGEKIMSCGMDHSLKLWRINSLRMKTAIKESYDYNPNK
zebrafish EED    210 LVAIFGGVEGHRDEVLSADFDLLGEKIMSCGMDHSLKLWRLDSERLQRAIRGSYEYNPSK
drosophila ESC   205 CIAILGGVEGHRDEVLSIDFNMRGDRIVSSGMDHSLKLWCLNTPEFHHKIELSNTFSQEK
                                    W D 4 0  r e p e a t
hydra EED        263 TERIFHTLNVHYPEYTTREVHRNYVDCCVWLGDLVISKSCDNQVVCWKTKQPIETVMKRK
mouse Eed        285 TNRPFISQKIHFPDFSTRDIHRNYVDCVRWLGDLILSKSCENAIVCWKPGKMEDDIDKIK
human EED        285 TNRPFISQKIHFPDFSTRDIHRNYVDCVRWLGDLILSKSCENAIVCWKPGKMEDDIDKIK
xenopus EED      270 TNRPFVSQKVHFPDFSTRDIHRNYVDCVRWLGDLILSKSCENAIVCWKPGKMEDDIEKIK
zebrafish EED    270 TNRPFVSQKIHFPDFSTRDIHRNYVDCVRWLGDLILSKSCENAIVCWKPGRMEDDIDRIK
drosophila ESC   265 STLPFPTVTKHFPDFSTRDIHRNYVDCVQWFGNFVLSKSCENAIVCWKPGQLHQSFEQVK
                                 W D 4 0  r e p e a t
hydra EED        323 NNSDVGVFVLHKFDIDLCDIWFIRFAVDLNQTILALGNQIGKVYLYDLEGEHPAHAKPTI
mouse Eed        345 P-SESNVTILGRFDYSQCDIWYMRFSMDFWQKMLALGNQVGKLYVWDLEVEDPHKAKCTT
human EED        345 P-SESNVTILGRFDYSQCDIWYMRFSMDFWQKMLALGNQVGKLYVWDLEVEDPHKAKCTT
xenopus EED      330 A-SESNVTILGRFDYSQCDIWYMRFSMDFWQKMLALGNQVGKLYVWDLEVEDPHKAKCTT
zebrafish EED    330 P-NESNVTILGRFDYSQCDIWYMRFSMDFWQKMLALGNQVGKLYVWDLEVEDPHKAKCTT
drosophila ESC   325 P-SDSSCTIIAEFEYDECEIWFVRFGFNPWQKVIALGNQQGKVYVWELDPSDPEGAHMTT
                           W D 4 0  r e p e a t
hydra EED        383 LFHSKCTTVVRQISFNSSAKVLIAVCDDGSVWRWDKQT----
mouse Eed        404 LTHHKCGAAIRQTSFSRDSSILIAVCDDASIWRWDRLR----
human EED        404 LTHHKCGAAIRQTSFSRDSSILIAVCDDASIWRWDRLR----
xenopus EED      389 LTYPKCASAIRQTSFSRDSSVLIAVCDDSTIWRWDRLR----
zebrafish EED    389 LTLPRCTSAIRQTSFSRDSSILIAVCDDASIWRWDRLR----
drosophila ESC   384 LHNSRSVATVRQIAFSRDASVLVYVCDDATVWRWNRRQTTSI
```

HySUZ12

From gene to protein:
The *HySUZ12* gene is composed of 14 exons and spans a genomic region of > 30 kb. The respective exon and intron lengths are indicated. 'Exon 10' cannot be found in any Contig. Undetermined lengths are indicated by a question mark. Based on a ~ 2 kb mRNA, a protein of 638 aa is translated.



Sequence information:
Nucleotide and derived amino acid sequence of *HySUZ12* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The 5' untranslated region is highlighted in grey. No 3' untranslated region has been determined for this gene. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The C2H2-zinc finger is highlighted in yellow, the VEFS-box is highlighted in light blue.

```
Exon 1 →                    5' UTR
    1  aggaataaat acaaaaaaaa aaaaaaaaga tattaacatc atacaagtat
   51  atatatatta tatgtataaa tataATGACT TCTTTAAATC AAAAGGTGGA
                                   M  T   S  L  N   Q  K  V
  101  ATCTTCACTT GGTACACCCA GAAAACTGAA GTTTGAACAA CTACGAGCAG
       E  S  S  L   G  T  P   R  K  L   K  F  E  Q   L  R  A
                                        Exon 2 →
  151  ACTATGAACT TTTTTTGCAT ACATATGAAA | AACCAACATG GATCTATAGA
        D  Y  E   L  F  L  H   T  Y  E   K  P  T   W  I  Y  R
                                   Exon 3 →
  201  TTTATACAAC GAAGACATCA ATTATCG | CCT TTTATGCTAC ACAGAAATTT
         F  I  Q   R  R  H   Q  L  S   P   F  M  L   H  R  N
  251  GTCTTATTTT AAAGATCGAA GAACAAAGAG ACCACAGCTG CAGCAAGTTA
       L  S  Y  F   K  D  R   R  T  K   R  P  Q  L   Q  Q  V
       Exon 4 →
  301  AAAG | GAAAAA TTTTCAATTG GACCTGTTAT TAAAAGACTT TAAATCAACT
        K    R  K  N  F  Q  L   D  L  L   L  K  D   F  K  S  T
                                        Exon 5 →
  351  GAATTGAATG TTAAAAGTGA AAGTAATGAG GA | AAATCATT ATCTGAATAT
         E  L  N   V  K  S   E  S  N  E    E  N  H   Y  L  N
```

Exon 6 →

401 ATCTTTTAAT GGATTCTTTC ATGGTCCAG|G TTTTGCTGAA GAAATATTTG
I S F N G F F H G P G F A E E I F

451 AAACAAATGA ACCGAAAGAA CAAGAACTGT TTAAAGACGA CTTTGTCAAA
E T N E P K E Q E L F K D D F V K

Exon 7 →

501 GTAGAAATAC ACTTATGTCA AGTATACCAT AAAAAGAAAA AG|TGCCAAGA
V E I H L C Q V Y H K K K K C Q

551 TTCTGAACAA ATGGATCTTA TTCCTCTTGG GTCTTGTAAA GCTCCTTGGA
D S E Q M D L I P L G S C K A P W

601 ATCCACGCAC TCAACTTCTT CTGCCATGTT CATCTGTTAA TCTCACTGTT
N P R T Q L L L P C S S V N L T V

651 CCTCCTAATG CTTTTGTTGA AAATGGTCGT GCTGTAAAAG CACACATATT
P P N A F V E N G R A V K A H I

701 ATCTTTCACA GTCATTATTA CAGTGCCTAA AAAAGGTTCA TCAAAAAACA
L S F T V I I T V P K K G S S K N

Exon 8 →

751 TAAAATCAGA CCGAACTGAA G|AAAATGGAT TACATCGCCC ATTAAAACGT
I K S D R T E E N G L H R P L K R

801 CAAAAGTGTG ACAATTCTGA TGCTGACAGT GAAACTAATA GTTTAACTTA
Q K C D N S D A D S E T N S L T

851 TCCAGCTGAT GTATGTGTAG CATCATTTTC GGCAGAAATG ATTCTCTATG
Y P A D V C V A S F S A E M I L Y

901 ACCGTCTTAA AAATTGCTTG TTGACAGATG GGGATTATGA GTTACTCTTA
D R L K N C L L T D G D Y E L L L

951 CATGAAAATT CACCGACAAT TCGAAAATCA AACAGTTGGG AAAGCAAGTT
H E N S P T I R K S N S W E S K

Exon 9 →

1001 CAAATCAAAG|CTGGGACCGT TTGCAGCATT TTCTTTTGGT CCGACTATAA
F K S K L G P F A A F S F G P T I

1051 AGTTTAACCT CACCTGGGGA ACAAAACCAA TAGCACCACT GGCAGCTGCC
K F N L T W G T K P I A P L A A A

1101 TATAAATGTC TTGCATCTCC TTATCAAGAC AATTCCAATA TAAAGATTGA
Y K C L A S P Y Q D N S N I K I

1151 GTTAGAAATG TCAAAAGAAA CAAAGTCTGT TGCAGTAACT TCTGCAGAAA
E L E M S K E T K S V A V T S A E

1201 AGAATATGCG AATGTTTTAT CGGTTTATTT ATAATGGATC CACTAGACAA
K N M R M F Y R F I Y N G S T R Q

1251 CAAACTGAAG CCAAAGAAGG TATGGCCTGT CCTTGGTGTG CATTGGACTG
Q T E A K E G M A C P W C A L D

C2H2 zinc finger

1301 CAAAAGATTA TATGTTTTGT TAAAGCATTT TAAAACTTGT CATCCTCGAT
C K R L Y V L L K H F K T C H P R

'Exon 10' →

1351 TTTCTTTTGT ATACACA|GCT ATAAAAAAAG GGCACCTAAT AGATGTACGT
F S F V Y T A I K K G H L I D V R

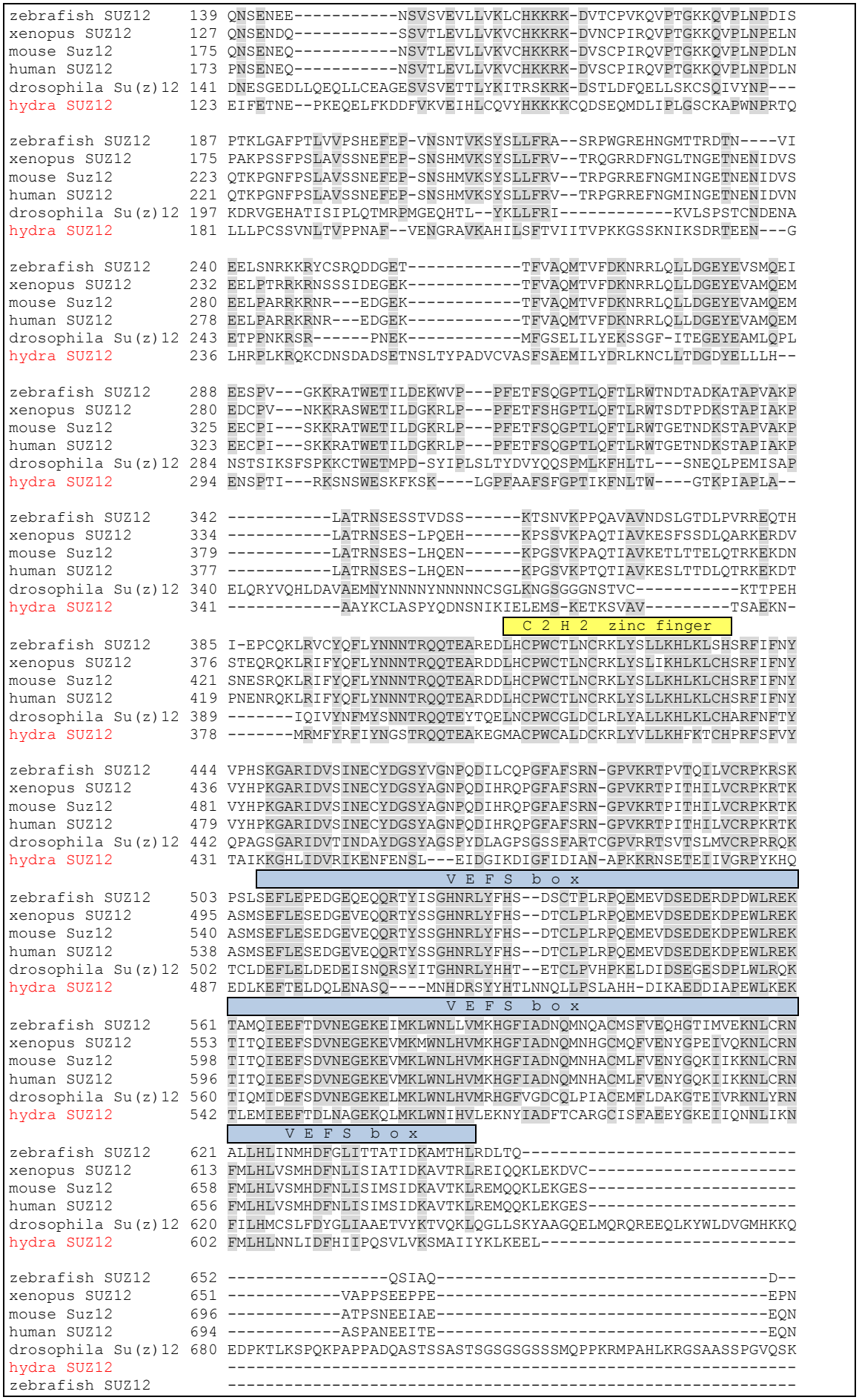
1401 ATCAAAGAGA ATTTTGAAAA TTCATTAGAG ATTGATGGTA TTAAAGATAT
I K E N F E N S L E I D G I K D

```
1451  TGGATTTATT GATATTGCAA ATGCTCCAAA GAAAAGGAAT TCAGAAACAG
        I  G  F  I   D  I  A   N  A  P   K  K  R  N   S  E  T

                    Exon 11 →
1501  AAATTATTGT TGGACG|GCCT TACAAACATC AAGAAGATTT AAAAGAGTTT
        E  I  I   V  G   R  P   Y  K  H   Q  E  D   L  K  E  F

                          VEFS-Box                  Exon 12 →
1551  ACAGAACTTG ATCAGTTAGA AAATGCATCA CAAATGAATC ATGACAG|GTC
        T  E  L   D  Q  L   E  N  A  S   Q  M  N   H  D   R

1601  TTATTACCAC ACGCTTAACA ACCAGCTATT ACCTAGTCTG GCACATCATG
        S  Y  Y  H   T  L  N   N  Q  L   L  P  S  L   A  H  H

1651  ATATAAAAGC TGAAGATGAT ATCGCTCCAG AGTGGTTAAA GGAGAAAACT
        D  I  K   A  E  D  D   I  A  P   E  W  L   K  E  K  T

          Exon 13 →
1701  TTAGAG|ATGA TTGAAGAATT TACTGATTTA AATGCTGGTG AGAAGCAACT
        L  E   M   I  E  E   F  T  D  L   N  A  G   E  K  Q

                                          Exon 14 →
1751  GATGAAATTG TGGAACATTC ATGTTTTAGA AAAAAA|TTAT ATTGCTGACT
        L  M  K  L   W  N  I   H  V  L   E  K   N  Y   I  A  D

1801  TCACTTGTGC TAGAGGGTGC ATATCTTTTG CCGAGGAATA TGGTAAAGAA
        F  T  C   A  R  G  C   I  S  F   A  E  E   Y  G  K  E

1851  ATTATTCAAA ACAATCTAAT CAAAAATTTT ATGCTTCATC TAAATAACTT
        I  I  Q   N  N  L   I  K  N  F   M  L  H   L  N  N

1901  AATTGATTTT CATATTATAC CACAATCTGT GCTGGTTAAG TCGATGGCTA
        L  I  D  F   H  I  I   P  Q  S   V  L  V  K   S  M  A

1951  TCATCTACAA GTTGAAAGAA GAACTTTTAC CAAAAATGTA G
        I  I  Y   K  L  K  E   E  L  L   P  K  M   *
```

Protein alignment:

Comparison of *Hy*SUZ12 with SUZ12 protein sequences from other species reveals a conservation of 28% on the amino acid level. Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

```
zebrafish SUZ12        1 MAPQKHG------------SGGNGFYGNSASKAA-------------NGGAHASS-----
xenopus SUZ12          1 MAPQKHGGS----SGPTSKSN-----------------------------------------
mouse Suz12            1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAAVAAAASGGKSGGGGCGGGGSYSASSSSAA
human SUZ12            1 MAPQKHGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCGGGGSYSASSSS--
drosophila Su(z)12     1 MAPAKK---REKDSNPD-GSAANGIIGLT-------------------HGAPDASNAGST
hydra SUZ12            1 ------------------------------------------------------MTSLNQ

zebrafish SUZ12       31 ------GVVMTSVKRLKMEQIQADHQLFLEAFEKPTQIYRFLRTRNLITPIFLHRSLTYM
xenopus SUZ12         18 -----ATASLLPVRKPKMEQIQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYM
mouse Suz12           61 AAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYM
human SUZ12           59 SAAAAAGAAVLPVKKPKMEHVQADHELFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYM
drosophila Su(z)12    38 VPPTAEG-------QVKLNGHQQEQELFLQAFEKPTQIYRYLRNRHETNPIFLNRTLSYM
hydra SUZ12            7 KVESSLGTP----RKLKFEQLRADYELFLHTYEKPTWIYRFIQRRHQLSPFMLHRNLSYF

zebrafish SUZ12      85 SHRNSRT----NSKRKSFSLNDLLFKVEKSKVDPESHNLAS-N-LQLTFTGFFHKTEKQL
xenopus SUZ12         73 SHRNSRS----NAKRKSFKVNDLLLKVEKMKGDQESHSVSA-H-LQLTFTGFFHKIDKPS
mouse Suz12          121 SHRNSRT----SIKRKTFKVDDMLSKVEKMKGEQESHSLSA-H-LQLTFTGFFHKNDKPS
human SUZ12          119 SHRNSRT----NIKRKTFKVDDMLSKVEKMKGEQESHSLSA-H-LQLTFTGFFHKNDKPS
drosophila Su(z)12    91 KERMSRN----NKKRISFQVNSMLESI-----TQKSEAVSQNY-LHVIYDSLHEKLPARL
hydra SUZ12           63 KDRRTKRPQLQQVKRKNFQLDLLLKDFKSTELNVKSESNEENHYLNISFNGFFHGPGFAE
```

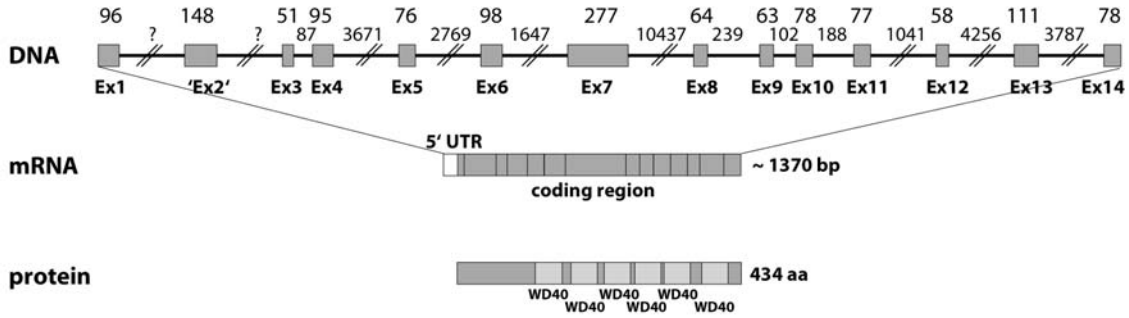

C 2 H 2 zinc finger
V E F S b o x
V E F S b o x
V E F S b o x

```
xenopus SUZ12       664 RTTNSFMESNGK------------DQALENDC--------------VSG---PPKHSKKQ
mouse Suz12         709 GTANGFSETNSK------------EKALETDG--------------VSG---VPKQSKKQ
human SUZ12         707 GTANGFSEINSK------------EKALETDS--------------VSG---VSKQSKKQ
drosophila Su(z)12  740 GTENGTNGSNSSSSNSKNVAKKSADQPLSTLANTRERRSEYGQKRNVSGSRLAATPASKR
hydra SUZ12         635 ------------------------------------------------------LPKM----

zebrafish SUZ12         ------------------------------------------------------------
xenopus SUZ12       695 KP----------------------------------------------------------
mouse Suz12         740 KL----------------------------------------------------------
human SUZ12         738 KL----------------------------------------------------------
drosophila Su(z)12  800 KLSSKDNTVLNKRQRYSDGSPGTGIGNGHGGGSGSGANRNKSNNHSLPATSNNASSSSSN
hydra SUZ12              ------------------------------------------------------------

zebrafish SUZ12          ----------------------------------------
xenopus SUZ12            ----------------------------------------
mouse Suz12              ----------------------------------------
human SUZ12              ----------------------------------------
drosophila Su(z)12  860  SKRAIARRRSTSERTKASGSTGGGAGGVRTRLSVPAKYERR
hydra SUZ12              ----------------------------------------
```

Hyp55

From gene to protein:

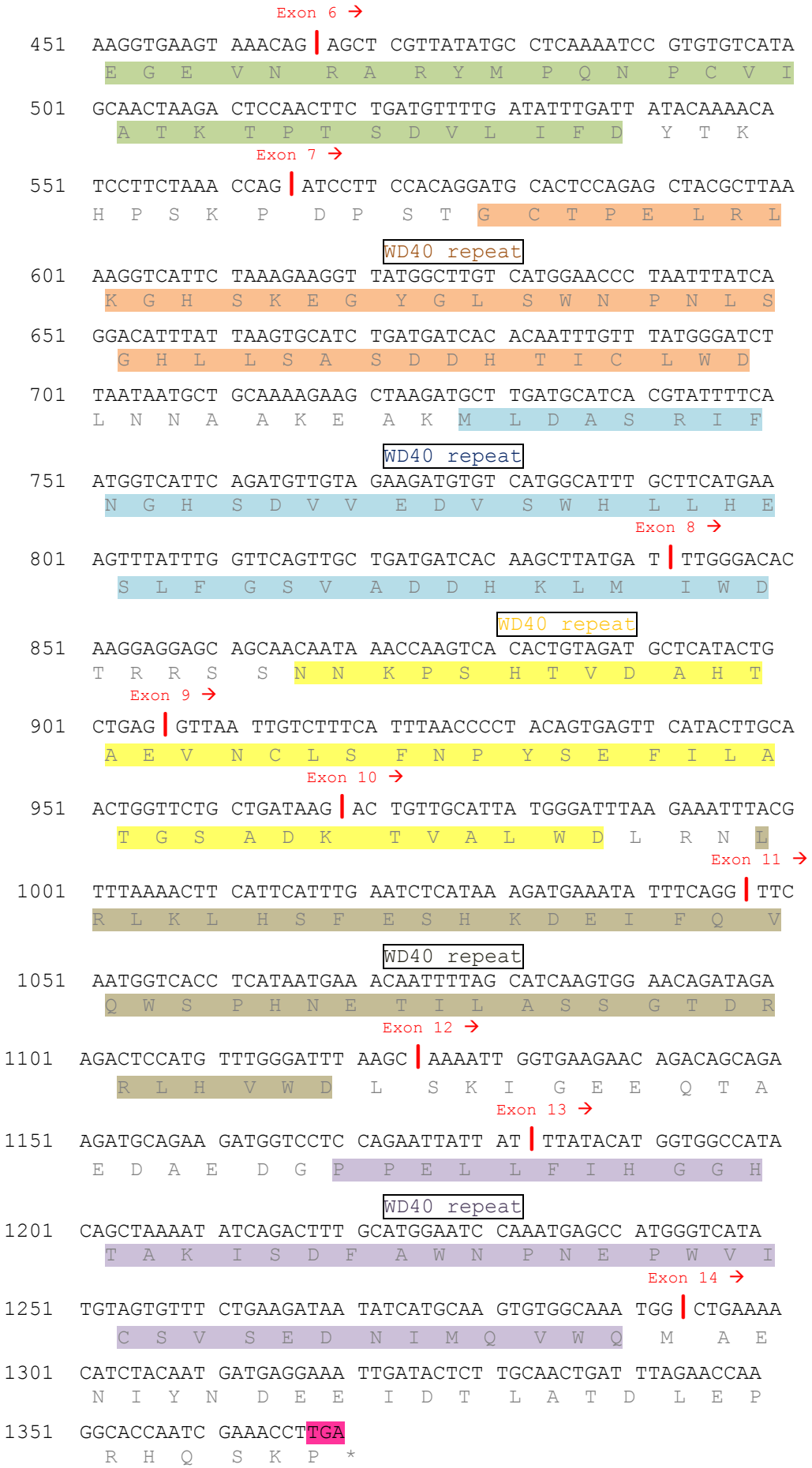
The *Hyp55* gene is composed of 14 exons and spans a genomic region of > 29 kb. The respective exon and intron lengths are indicated. 'Exon 2' cannot be found in any Contig. Undetermined lengths are indicated by a question mark. Based on a ~ 1.4 kb mRNA, a protein of 434 aa is translated.



Sequence information:

Nucleotide and derived amino acid sequence of *Hyp55* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The 5' untranslated region is highlighted in grey. No 3' untranslated region has been determined for this gene. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The six WD40 domains are highlighted in light green, red, blue, yellow, brown and blue lilac.

```
Exon 1 →                                  5' UTR
  1  ttttcccgcc atcaggaatt aggtgtcaac aaaaagtctc ataacctgca
                                                        Exon 2 →
 51  actaaataac cttatATGGC TGCTTTAAGT GCTCCAAACG TACCAG|ATGT
                         M   A   A   L   S   A   P   N   V   P   D
101  TTATGATGAT GCTGCTGAGG AGAGAATCAT TAATGAAGAG TACAAAATAT
     V   Y   D   D   A   A   E   E   R   I   I   N   E   E   Y   K   I
151  GGAAAAAAAA TACTCCTTTT TTATATGACC TTGTTATGAC GCATGCCTTA
      W   K   K   N   T   P   F   L   Y   D   L   V   M   T   H   A   L
                                                        Exon 3 →
201  GAATGGCCTA GTCTAACTGT ACAATGGCTT CCTGATGTCA CAAG|ACCTGA
       E   W   P   S   L   T   V   Q   W   L   P   D   V   T   R   P
                                                        Exon 4 →
251  AGGTAAAGAC TATTCTGTTC ATCGTTTGAT ATTAGGAACA CACAC|TTCTG
     E   G   K   D   Y   S   V   H   R   L   I   L   G   T   H   T   S
301  ATGAGCAAAA CCATCTTGTA ATAGCAACTG TACAATTACC AAATGATGAA
      D   E   Q   N   H   L   V   I   A   T   V   Q   L   P   N   D   E
                                                        Exon 5 →
351  ACTCATTTTG ATGCCAGTCA TTATGATCAT GATAGAGGAG|AGTTTGGTGG
       T   H   F   D   A   S   H   Y   D   H   D   R   G   E   F   G
                                                WD40 repeat
401  ATTTGGTACT GTAAGCGGAA AAATTGACGT AGAAATTAAA ATTAATCATG
     G   F   G   T   V   S   G   K   I   D   V   E   I   K   I   N   H
```


Exon 6
451 AAGGTGAAGT AAACAG | AGCT CGTTATATGC CTCAAAATCC GTGTGTCATA
E G E V N R A R Y M P Q N P C V I
501 GCAACTAAGA CTCCAACTTC TGATGTTTTG ATATTTGATT ATACAAAACA
A T K T P T S D V L I F D Y T K
Exon 7
551 TCCTTCTAAA CCAG | ATCCTT CCACAGGATG CACTCCAGAG CTACGCTTAA
H P S K P D P S T G C T P E L R L
WD40 repeat
601 AAGGTCATTC TAAAGAAGGT TATGGCTTGT CATGGAACCC TAATTTATCA
K G H S K E G Y G L S W N P N L S
651 GGACATTTAT TAAGTGCATC TGATGATCAC ACAATTTGTT TATGGGATCT
G H L L S A S D D H T I C L W D
701 TAATAATGCT GCAAAAGAAG CTAAGATGCT TGATGCATCA CGTATTTTCA
L N N A A K E A K M L D A S R I F
WD40 repeat
751 ATGGTCATTC AGATGTTGTA GAAGATGTGT CATGGCATTT GCTTCATGAA
N G H S D V V E D V S W H L L H E
Exon 8
801 AGTTTATTTG GTTCAGTTGC TGATGATCAC AAGCTTATGA T | TTGGGACAC
S L F G S V A D D H K L M I W D
WD40 repeat
851 AAGGAGGAGC AGCAACAATA AACCAAGTCA CACTGTAGAT GCTCATACTG
T R R S S N N K P S H T V D A H T
Exon 9
901 CTGAG | GTTAA TTGTCTTTCA TTTAACCCCT ACAGTGAGTT CATACTTGCA
A E V N C L S F N P Y S E F I L A
Exon 10
951 ACTGGTTCTG CTGATAAG | AC TGTTGCATTA TGGGATTTAA GAAATTTACG
T G S A D K T V A L W D L R N L
Exon 11
1001 TTTAAAACTT CATTCATTTG AATCTCATAA AGATGAAATA TTTCAGG | TTC
R L K L H S F E S H K D E I F Q V
WD40 repeat
1051 AATGGTCACC TCATAATGAA ACAATTTTAG CATCAAGTGG AACAGATAGA
Q W S P H N E T I L A S S G T D R
Exon 12
1101 AGACTCCATG TTTGGGATTT AAGC | AAAATT GGTGAAGAAC AGACAGCAGA
R L H V W D L S K I G E E Q T A
Exon 13
1151 AGATGCAGAA GATGGTCCTC CAGAATTATT AT | TTATACAT GGTGGCCATA
E D A E D G P P E L L F I H G G H
WD40 repeat
1201 CAGCTAAAAT ATCAGACTTT GCATGGAATC CAAATGAGCC ATGGGTCATA
T A K I S D F A W N P N E P W V I
Exon 14
1251 TGTAGTGTTT CTGAAGATAA TATCATGCAA GTGTGGCAAA TGG | CTGAAAA
C S V S E D N I M Q V W Q M A E
1301 CATCTACAAT GATGAGGAAA TTGATACTCT TGCAACTGAT TTAGAACCAA
N I Y N D E E I D T L A T D L E P
1351 GGCACCAATC GAAACCTTGA
R H Q S K P *

Protein alignment:

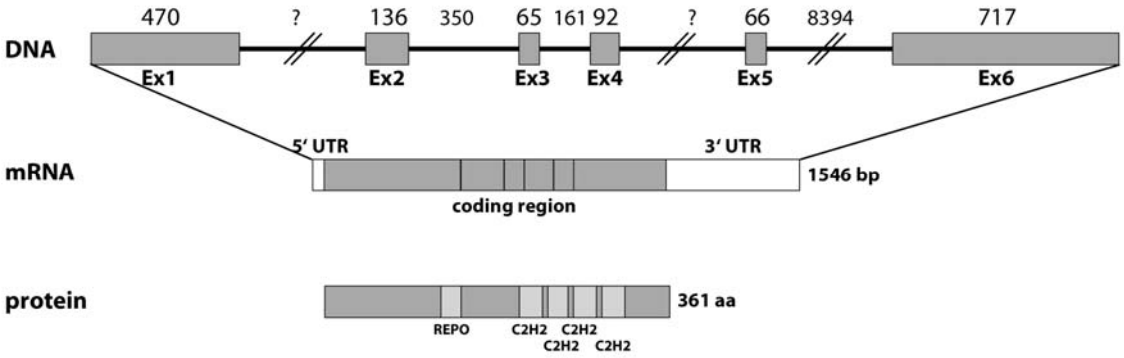
Comparison of *Hyp*55 with p55 protein sequences from other species reveals significant conservation on the amino acid level (82%). Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

```
hydra p55          1 MAALSAPNVPDVYDDAAEERIINEEYKIWKKNTPFLYDLVMTHALEWPSLTVQWLPDVTR
human RBBP4        1 -MADKE----AAFDDAVEERVINEEYKIWKKNTPFLYDLVMTHALEWPSLTAQWLPDVTR
mouse Rbbp4        1 -MADKE----AAFDDAVEERVINEEYKIWKKNTPFLYDLVMTHALEWPSLTAQWLPDVTR
xenopus RBBP4      1 -MADKE----AAFDDAVEERVINEEYKIWKKNTPFLYDLVMTHALEWPSLTAQWLSDVTR
zebrafish RBBP4    1 -MADKE----AAFDDAVEERVINEEYKIWKKNTPFLYDLVMTHALEWPSLTAQWLPDVTR
drosophila p55     1 -MVDRSDNAAESFDDAVEERVINEEYKIWKKNTPFLYDLVMTHALEWPSLTAQWLPDVTK

hydra p55         61 PEGKDYSVHRLILGTHTSDEQNHLVIATVQLPNDETHFDASHYDHDRGEFGGFGTVSGKI
human RBBP4       56 PEGKDFSIHRLVLGTHTSDEQNHLVIASVQLPNDDAQFDASHYDSEKGEFGGFGSVSGKI
mouse Rbbp4       56 PEGKDFSIHRLVLGTHTSDEQNHLVIASVQLPNDDAQFDASHYDSEKGEFGGFGSVSGKI
xenopus RBBP4     56 PDGKDFSIHRLVLGTHTSDEQNHLVIASVQLPNDDAQFDASHYDSEKGEFGGFGSVSGKI
zebrafish RBBP4   56 PEGKDFSVHRLVLGTHTSDEQNHLVIASVQLPNDDAQFDASHYDSEKGEFGGFGSVSGKI
drosophila p55    60 QDGKDYSVHRLILGTHTSDEQNHLLIASVQLPSEDAQFDGSHYDNEKGEFGGFGSVCGKI
                        W D 4 0  r e p e a t

hydra p55        121 DVEIKINHEGEVNRARYMPQNPCVIATKTPTSDVLIFDYTKHPSKPDPSTGCTPELRLKG
human RBBP4      116 EIEIKINHEGEVNRARYMPQNPCIIATKTPSSDVLVFDYTKHPSKPDPSGECNPDLRLRG
mouse Rbbp4      116 EIEIKINHEGEVNRARYMPQNPCIIATKTPSSDVLVFDYTKHPSKPDPSGECNPDLRLRG
xenopus RBBP4    116 EIEIKITHDGEVNRARYMPQNPCIIATKTPTSDVLVFDYTKHPSKPDPSGECNPNLRLRG
zebrafish RBBP4 116 EIEIKINHEGEVNRARYMPQNPCIIATKTPTSDVLVFDYTKHPSKPDPSGECTPDLRLRG
drosophila p55   120 EIEIKINHEGEVNRARYMPQNACVIATKTPSSDVLVFDYTKHPSKPEPSGECQPDLRLRG
                        W D 4 0  r e p e a t

hydra p55        181 HSKEGYGLSWNPNLSGHLLSASDDHTICLWDLNNAAKEAKMLDASRIFNGHSDVVEDVSW
human RBBP4      176 HQKEGYGLSWNPNLSGHLLSASDDHTICLWDISAVPKEGKVVDAKTIFTGHTAVVEDVSW
mouse Rbbp4      176 HQKEGYGLSWNPNLSGHLLSASDDHTICLWDISAVPKEGKVVDAKTIFTGHTAVVEDVSW
xenopus RBBP4    176 HQKEGYGLSWNPNLSGNLLSASDDHTICLWDISAVPKEGKVVDAKTIFTGHTAVVEDVSW
zebrafish RBBP4 176 HQKEGYGLSWNPNLSGCLLSASDDHTICLWDISTVPKEGKIVDAKTIFTGHTAVVEDVSW
drosophila p55   180 HQKEGYGLSWNPNLNGYLLSASDDHTICLWDINATPKEHRVIDAKNIFTGHTAVVEDVAW
                        W D 4 0  r e p e a t      W D 4 0  r e p e a t

hydra p55        241 HLLHESLFGSVADDHKLMIWDTRRSSNNKPSHTVDAHTAEVNCLSFNPYSEFILATGSAD
human RBBP4      236 HLLHESLFGSVADDQKLMIWDTRSNNTSKPSHSVDAHTAEVNCLSFNPYSEFILATGSAD
mouse Rbbp4      236 HLLHESLFGSVADDQKLMIWDTRSNNTSKPSHSVDAHTAEVNCLSFNPYSEFILATGSAD
xenopus RBBP4    236 HLLHESLFGSVADDQKLMIWDTRSNNTSKPSHSVDAHTAEVNCLSFNPYSEFILATGSAD
zebrafish RBBP4 236 HLLHESLFGSVADDQKLMIWDTRSNNTSKPSQAVDAHTAEVNCLSFNPYSEFILATGSAD
drosophila p55   240 HLLHESLFGSVADDQKLMIWDTRNNNTSKPSHTVDAHTAEVNCLSFNPYSEFILATGSAD
                        W D 4 0  r e p e a t

hydra p55        301 KTVALWDLRNLRLKLHSFESHKDEIFQVQWSPHNETILASSGTDRRLHVWDLSKIGEEQT
human RBBP4      296 KTVALWDLRNLKLKLHSFESHKDEIFQVQWSPHNETILASSGTDRRLNVWDLSKIGEEQS
mouse Rbbp4      296 KTVALWDLRNLKLKLHSFESHKDEIFQVQWSPHNETILASSGTDRRLNVWDLSKIGEEQS
xenopus RBBP4    296 KTVALWDLRNLKLKLHSFESHKDEIFQVQWSPHNETILASSGTDRRLNVWDLSKIGEEQS
zebrafish RBBP4 296 KTVALWDLRNLKLKLHSFESHKDEIFQVQWSPHNETILASSGTDRRLNVWDLSKIGEEQS
drosophila p55   300 KTVALWDLRNLKLKLHSFESHKDEIFQVQWSPHNETILASSGTDRRLHVWDLSKIGEEQS
                        W D 4 0  r e p e a t

hydra p55        361 AEDAEDGPPELLFIHGGHTAKISDFAWNPNEPWVICSVSEDNIMQVWQMAENIYNDEEID
human RBBP4      356 PEDAEDGPPELLFIHGGHTAKISDFSWNPNEPWVICSVSEDNIMQVWQMAENIYNDEDPE
mouse Rbbp4      356 PEDAEDGPPELLFIHGGHTAKISDFSWNPNEPWVICSVSEDNIMQVWQMAENIYNDEDPE
xenopus RBBP4    356 PEDAEDGPPELLFIHGGHTAKISDFSWNPNEPWVICSVSEDNIMQVWQMAENIYNDEDTE
zebrafish RBBP4 356 PEDAEDGPPELLFIHGGHTAKISDFSWTPNEPWVICSVSEDNIMQVWQMAENIYNDEDPE
drosophila p55   360 TEDAEDGPPELLFIHGGHTAKISDFSWNPNEPWIICSVSEDNIMQVWQMAENVYNDEEPE

hydra p55        421 TLATDLEPRHQSKP
human RBBP4      416 ---GSVDPEGQGS-
mouse Rbbp4      416 ---GSVDPEGQGS-
xenopus RBBP4    416 ---GGVDPEGQGS-
zebrafish RBBP4 416 ---GAADTEVQG--
drosophila p55   420 ---IPASELETNTA
```

HyYY1

From gene to protein:

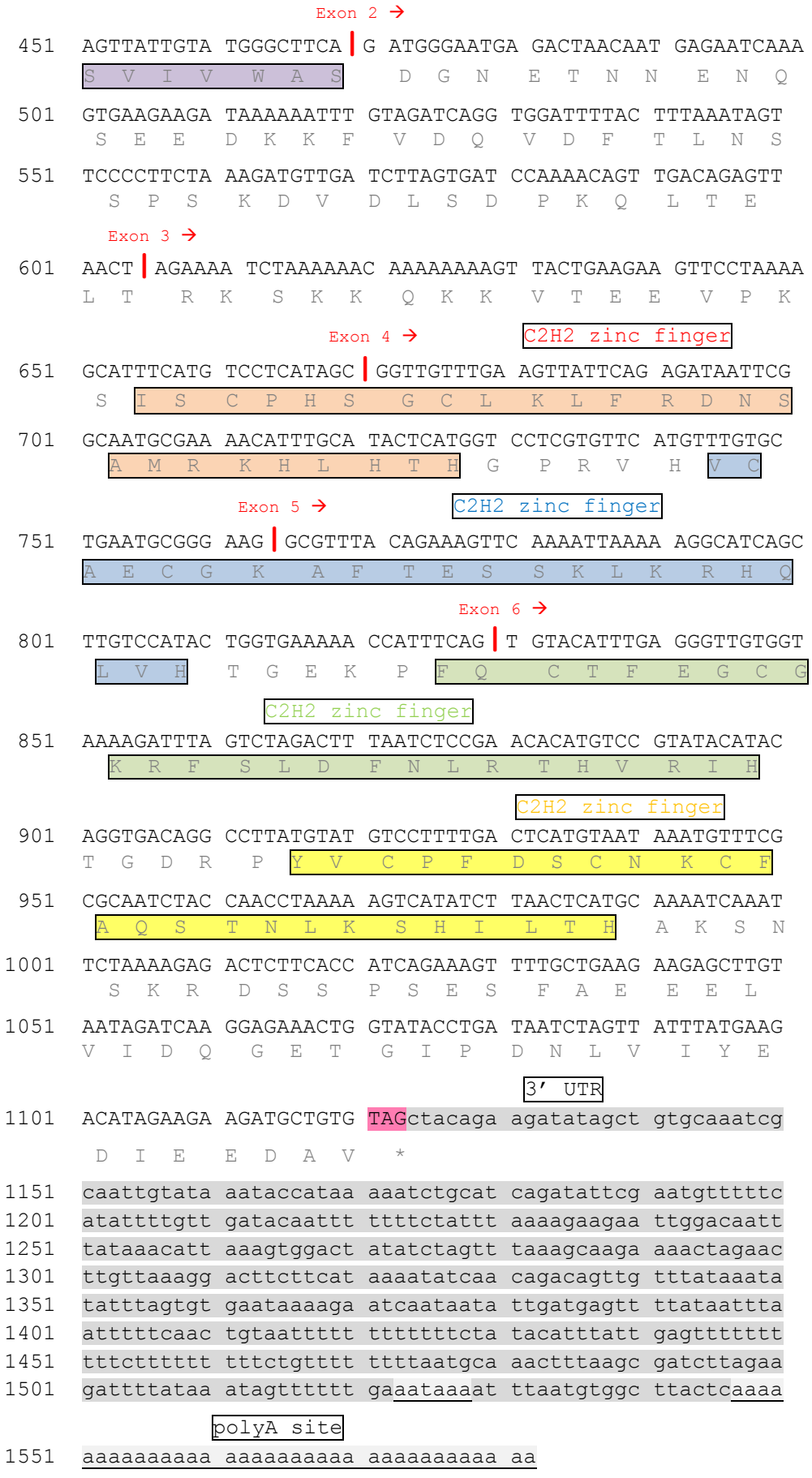
The *HyYY1* gene is composed of 6 exons and spans a genomic region of ~ 10.5 kb. The respective exon and intron lengths are indicated. Undetermined lengths are indicated by a question mark. Based on a ~ 1.5 kb mRNA, a protein of 361 aa is translated.



Sequence information:

Nucleotide and derived amino acid sequence of *HyYY1* from *Hydra magnipapillata 105*. The beginning of the respective exons is indicated. The untranslated regions are highlighted in grey; the Kozak sequence and the polyadenylation site are indicated by underlined nucleotides. The translation initiation start and stop codons are highlighted in red. The coding sequence and the derived amino acid sequence are indicated by capital letters. The protein domains are indicated. The REPO domain is highlighted in blue lilac, the four C2H2-zinc fingers are highlighted in light red, blue, green and yellow.

```
Exon 1 →          5' UTR                    Kozak sequence
  1  ccggtacaaa ttggacctaa gttaatgccg taaaaaaATG GCTGACGATT
                                              M   A  D  D

 51  TAACGATAGA TGGAGTTGAA AGTAATTCTT TTCTTTCTGT CGATGAAGAC
      L  T  I   D  G  V  E   S  N  S   F  L  S   V  D  E  D

101  AATCTTAGCA TCATCGATGA GAATTTTTAT AATCAACCTT TGATAGCACT
      N  L  S   I  I  D   E  N  F  Y   N  Q  P   L  I  A

151  ACAACCTAAT TCTCATTTAG GTGGCAATGA AACAGTTTCA GATGGGTTCT
      L  Q  P  N   S  H  L   G  G  N   E  T  V  S   D  G  F

201  TTGAAGATTC TTTAGTTGAT ATCCCTTTAG TTGGGGAAGG GCAAGAAATT
      F  E  D   S  L  V  D   I  P  L   V  G  E   G  Q  E  I

251  TCAACTGAAG CAGCGATACA ACAAACCCCT GAAAAATCAG GTATTAAGCG
      S  T  E   A  A  I   Q  Q  T  P   E  K  S   G  I  K

301  AAAAAATAAT AAAAAATCAA AACCAAAACT CAAGGTAAAA AATAAATTGA
      R  K  N  N   K  K  S   K  P  K   L  K  V  K   N  K  L

351  CAACATCCAA TCTTGATAAC GAGGACAGTA ACGATATCGA TGTTCCTCGT
      T  T  S   N  L  D  N   E  D  S   N  D  I   D  V  P  R

                                  REPO domain
401  AAATGGTCAA GAAAAAAGGT TCAGATTAAA ACATTTGAAG GAGAGTTCTC
      K  W  S   R  K  K   V  Q  I  K   T  F  E   G  E  F
```


Exon 2 →
451 AGTTATTGTA TGGGCTTCA | G ATGGGAATGA GACTAACAAT GAGAATCAAA
S V I V W A S D G N E T N N E N Q
501 GTGAAGAAGA TAAAAAATTT GTAGATCAGG TGGATTTTAC TTTAAATAGT
S E E D K K F V D Q V D F T L N S
551 TCCCCTTCTA AAGATGTTGA TCTTAGTGAT CCAAAACAGT TGACAGAGTT
S P S K D V D L S D P K Q L T E
Exon 3 →
601 AACT | AGAAAA TCTAAAAAAC AAAAAAAAGT TACTGAAGAA GTTCCTAAAA
L T R K S K K Q K K V T E E V P K
Exon 4 →
C2H2 zinc finger
651 GCATTTCATG TCCTCATAGC | GGTTGTTTGA AGTTATTCAG AGATAATTCG
S I S C P H S G C L K L F R D N S
701 GCAATGCGAA AACATTTGCA TACTCATGGT CCTCGTGTTC ATGTTTGTGC
A M R K H L H T H G P R V H V C
Exon 5 →
C2H2 zinc finger
751 TGAATGCGGG AAG | GCGTTTA CAGAAAGTTC AAAATTAAAA AGGCATCAGC
A E C G K A F T E S S K L K R H Q
Exon 6 →
801 TTGTCCATAC TGGTGAAAAA CCATTTCAG | T GTACATTTGA GGGTTGTGGT
L V H T G E K P F Q C T F E G C G
C2H2 zinc finger
851 AAAAGATTTA GTCTAGACTT TAATCTCCGA ACACATGTCC GTATACATAC
K R F S L D F N L R T H V R I H
C2H2 zinc finger
901 AGGTGACAGG CCTTATGTAT GTCCTTTTGA CTCATGTAAT AAATGTTTCG
T G D R P Y V C P F D S C N K C F
951 CGCAATCTAC CAACCTAAAA AGTCATATCT TAACTCATGC AAAATCAAAT
A Q S T N L K S H I L T H A K S N
1001 TCTAAAAGAG ACTCTTCACC ATCAGAAAGT TTTGCTGAAG AAGAGCTTGT
S K R D S S P S E S F A E E E L
1051 AATAGATCAA GGAGAAACTG GTATACCTGA TAATCTAGTT ATTTATGAAG
V I D Q G E T G I P D N L V I Y E
3′ UTR
1101 ACATAGAAGA AGATGCTGTG TAGctacaga agatatagct gtgcaaatcg
D I E E D A V *
1151 caattgtata aataccataa aaatctgcat cagatattcg aatgtttttc
1201 atattttgtt gatacaattt ttttctattt aaaagaagaa ttggacaatt
1251 tataaacatt aaagtggact atatctagtt taaagcaaga aaactagaac
1301 ttgttaaagg acttcttcat aaaatatcaa cagacagttg tttataaata
1351 tatttagtgt gaataaaaga atcaataata ttgatgagtt ttataattta
1401 atttttcaac tgtaattttt ttttttttcta tacatttatt gagtttttt
1451 tttctttttt tttctgtttt ttttaatgca aactttaagc gatcttagaa
1501 gattttataa atagtttttt gaaataaaat ttaatgtggc ttactcaaaa
polyA site
1551 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa

Protein alignment:

Comparison of *Hy*YY1 with YY1 protein sequences from other species reveals significant conservation on the amino acid level (39%). Areas of high matches are highlighted in grey. The BLOSUM62 matrix was used for the multi-way protein sequence alignment.

```
hydra YY1           1 MADDLTI----DGVESNS----------FLSVDEDNLSIIDE-----------------
human YY1           1 MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDDEDGGGGDHG
mouse Yy1           1 MASGDTLYIATDGSEMPAEIVELHEIEVETIPVETIETTVVGEEEEEDDDDEDGGGGDHG
xenopus YY1         1 MASGDTLYIASDGSEMPAEIVELHEIEVECIPVETIETTVVG---DDEDDDDDQG-----
zebrafish YY1       1 MASGETLYIEADGSEMPAEIVELHEIE-----VETIETTVVG---GDDDE----------
drosophila Pho      1 MA-----------YERFGIILQSEQYD------EDIGNTKVNQKMNEGNHYD--------

hydra YY1          29 ------------------NFYNQPLIALQP---------NSH------------LGGNE-
human YY1          61 GGGG-HGHAGHHHHHHHHHH-PPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDD-
mouse Yy1          61 GGGGGHGHAGHHHHHHHHHHPPMIALQPLVTDDPTQVHHHQEVILVQTREEVVGGDD-
xenopus YY1        53 --------------------HHQQPMIALQPLDSDDLVHP-HHQEVILVQTREEVVGGDD-
zebrafish YY1      43 ----------------------HQPMIALQPLVTDDPNHV-NHQEVILVQTREEVVGCDD-
drosophila Pho     36 ---------------LHRKNAFDRIIHSESKKGDNVINYNIHE-------NDKIKAADNI

hydra YY1          49 -T----VSDGFFEDSLVDIPLVGEG------QEISTEAAIQQTP-----EKSG--IKRKN
human YY1         118 -SDGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGGSSSSGGGRVKKGG
mouse Yy1         120 -SDGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGG--ASSGGGRVKKGG
xenopus YY1        92 -SD-LRADDGYEDQILIPVPVPA-GEDEYIEQTLVTV--AGKSSSSGGRMKKGGG----GS
zebrafish YY1      80 -SD-LHADDSFEDQILIPVPVPV-AEEEYIEQTLVTV--SGKNPSG-RMKKG-G----GS
drosophila Pho     74 FSSKLKMNPNMSYEMNINCFKNIG----YGENQETSKVLTNSLSNNDINTEESGVVDK--
                                                  R E P O   d o m a i n
hydra YY1          91 NKKSKPKLKVKNKLTTSNLDNEDSNDIDVPRKWSRKKVQIKTFEGEFSVIVWASDG----
human YY1         177 GKKSGKKSYLSGGAGAAGGGGADPGN----KKWEQKQVQIKTLEGEFSVTMWSSDEKKDI
mouse Yy1         177 GKKSGKKSYLGGGAGAAGGGGADPGN----KKWEQKQVQIKTLEGEFSVTMWSSDEKKDI
xenopus YY1       143 GKKSSKKSYL---------SGTEPSG----RKWEQKQVQIKTLEGEFSVTMWASDDKKDI
zebrafish YY1     129 GKRVVKKSFL---------NSAEASG----RKWEQKQVQIKTLEGEFSVTMWASDDKKDV
drosophila Pho    128 -----NSPFLTLGTTILNSNGKS-------RRWEQKLVHIKTMEGEFSVTMWASGIS---

hydra YY1         147 -NETNNENQSEEDKK-----------------FVDQ----VDFT----LN-SSPSKDV--
human YY1         233 DHETVVEEQIIGENS-----------------PPDY----SEYMTGKKLP-PGGIPGI--
mouse Yy1         233 DHETVVEEQIIGENS-----------------PPDY----SEYMTGKKLP-PGGIPGI--
xenopus YY1       190 DHETVVEEQIIGENS-----------------PPDY----SEYMTGKKLP-PGGIPGI--
zebrafish YY1     176 DHETVVEEQIIGENS-----------------PPDY----SEYMTGKKLP-PGGIPGI--
drosophila Pho    173 DDEYSGSDQIVGASDLLKGKEEFGIDGFTSQQNKEYQKMESKFTNAQTLEMPHPISSVQI

hydra YY1         178 ---------DLSDPKQLTE----------------------------LTR----------
human YY1         269 ---------DLSDPKQLAE----------------------------FA-----------
mouse Yy1         269 ---------DLSDPKQLAE----------------------------FA-----------
xenopus YY1       226 ---------DLSDPKQLAE----------------------------FASETRCTNNING
zebrafish YY1     212 ---------DLSDPKQLAE----------------------------FA-----------
drosophila Pho    233 MDHLIKERGNLSQENNISERILSKTTLSFEEPILLPDSSSIELVNETAAMTINNHRTLSN

hydra YY1         191 --------KSKKQKKV----------------------TEEVPKS--------------
human YY1         281 --------RMKPRKIK----------------------EDDAPRT--------------
mouse Yy1         281 --------RMKPRKIK----------------------EDDAPRT--------------
xenopus YY1       249 RKKQAPSWRMKPRKIK----------------------EDDAPRT--------------
zebrafish YY1     224 --------RMKPRKIK----------------------EDDSPRT--------------
drosophila Pho    293 HTGNTGDLHALPSSVPFRIGLHEGQVNDCLSTISQSTHQDNTDSTGCGEMNLSEVTVSYT
                          C 2 H 2   zinc   finger      C 2 H 2  zinc  finger
hydra YY1         206 ----ISCPHSGCLKLFRDNSAMRKHLHTHGPRVHVCAECGKAFTESSKLKRHQLVHTGEK
human YY1         296 ----IACPHKGCTKMFRDNSAMRKHLHTHGPRVHVCAECGKAFVESSKLKRHQLVHTGEK
mouse Yy1         296 ----IACPHKGCTKMFRDNSAMRKHLHTHGPRVHVCAECGKAFVESSKLKRHQLVHTGEK
xenopus YY1       272 ----IACPHKGCTKMFRDNSAMRKHLHTHGPRVHVCAECGKAFVESSKLKRHQLVHTGEK
zebrafish YY1     239 ----IACPHKGCTKMFRDNSAMRKHLHTHGPRVHVCAECGKAFVESSKLKRHQLVHTGEK
drosophila Pho    353 NDKKIACPHKGCNKHFRDSSAMRKHLHTHGPRVHVCAECGKAFVESSKLKRHQLVHTGEK
                        C 2 H 2   zinc   finger      C 2 H 2   zinc   finger
hydra YY1         262 PFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPYVCPFDSCNKCFAQSTNLKSHILTHAKSN
human YY1         352 PFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPYVCPFDGCNKKFAQSTNLKSHILTHAKAK
mouse Yy1         352 PFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPYVCPFDGCNKKFAQSTNLKSHILTHAKAK
xenopus YY1       328 PFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPYVCPFDGCNKKFAQSTNLKSHILTHAKAK
zebrafish YY1     295 PFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPYVCPFDGCNKKFAQSTNLKSHILTHAKAK
drosophila Pho    413 PFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPFVCPFDACNKKFAQSTNLKSHILTHAKAK

hydra YY1         322 SKRDSSPSESFAEEELVIDQGETGIPDNLVIYEDIEEDAV--------
human YY1         412 NNQ-----------------------------------------------
mouse Yy1         412 NNQ-----------------------------------------------
xenopus YY1       388 NNQ-----------------------------------------------
zebrafish YY1     355 NNQ-----------------------------------------------
drosophila Pho    473 RNTSISGKSGCSNAESNSQSEDTSANYVKVELQDSVTENHVPFVVYAD
```

B. Histone modifications and *Hydra* histones

B.1. Histone posttranslational modifications

Histone are subjected to many posttranslational modifications. These modifications can alter the chromatin structure, disrupt the binding of associated proteins or provide a binding platform for effector proteins. The recruitment of proteins or protein complexes can activate or repress transcription. Whereas certain modifications are associated with transcriptional activation, others are linked to transcriptional repression. Histone modifications include the acetylation, phosphorylation, methylation, sumoylation and ubiquitylation of histone residues. Jenuwein and Allis (2001) proposed that modifications of histone tails form a kind of language, which became known as the 'histone code' hypothesis. An overview of known histone modifications is depicted in figure 1.14 on page 17. Table **??** gives an overview of histone posttranslational modifications and their role in transcription. In more recent studies, the existence of a histone 'cross-talk' is proposed where one histone modification promotes the generation of another (for a review see Suganuma and Workman (2008)). For example, the PcG complexes catalyze the modification of H3K27 followed by the modification of H2AK119.

TABLE B.1 *List of histone posttranslational modifications. The table was adapted from Allis et al. (2006).*

modification	role in transcription	modified residue
Acetylation	activation	H3 (K9, K14, K18, K56) H4 (K5, K8, K12, K16) H2B (K6, K7, K16, K17)
Phosphorylation	activation	H3 (S10)
Methylation	activation repression	H3 (K4, K36, K79) H3 (K9, K27) H4 (K20)
Ubiquitylation	activation repression	H2B (K123) H2A (K119)
Sumoylation	repression	H4 (K5, K8, K12, K16) H2A (K126) H2B (K6, K7, K16, K17)

B.2. *Hydra* histones

Histone proteins are highly conserved. By genome searches with histone proteins from human and by the comparison of *Hydra* gene models, homologs could be identified in the *Hydra* genome. All putative *Hydra* histone proteins differ only in a few residues from the human histone proteins. In figure B.1, the sequences of the *Hydra* histone proteins are aligned with human histone variants. For histone H3, the tail region is one hundred per cent identical to its human counterpart - including lysine 27, the site trimethylated by the PRC2 complex.

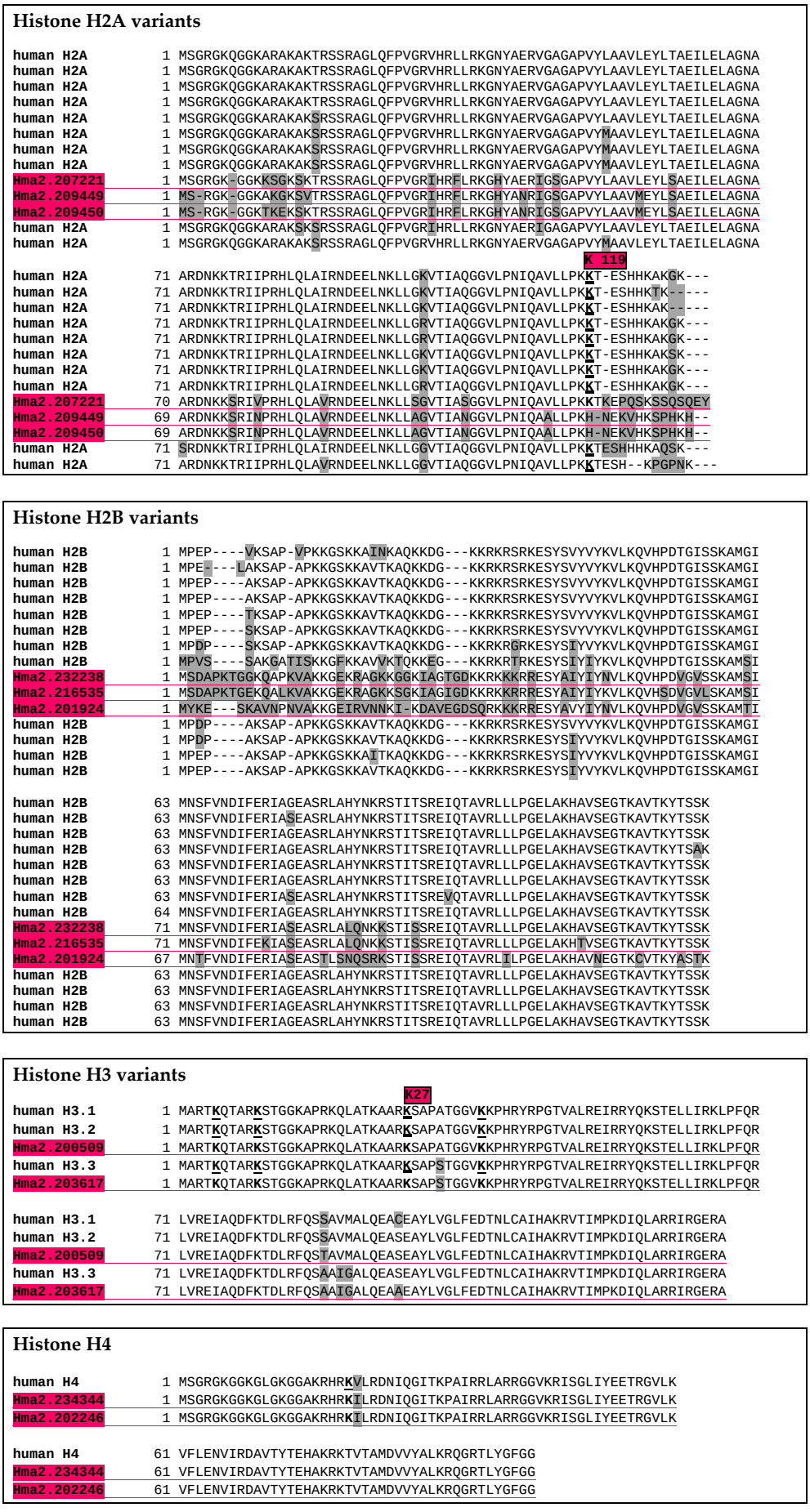


FIGURE B.1 *Hydra histones. Alignment of human histone genes with gene models of putative Hydra histones. Sites of epigenetic modifications involved in PcG repression are marked.*

C. Chromatin remodeling complexes

Chromatin remodeling complexes disrupt DNA-histone interactions and thus alter the chromatin structure. The topic is reviewed by (Clapier and Cairns, 2009). Whereas some of the remodelers promote the dense packaging of chromatin, others allow the access to the packaged DNA. Some of the complexes act as transcriptional activators and have TrxG-like functions. To alter histone-DNA contacts, the chromatin remodeling complexes use the energy of ATP-hydrolysis. Figure C.1 gives an overview of the mechanisms for the ATP-dependent chromatin remodeling. Dependent on their ATPase subunit, the chromatin remodelers are classified into four different families: SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation SWI), CHD (chromodomain, helicase, DNA binding) and INO80 (inositol requiring 80). Like PcG mediated gene regulation, chromatin remodeling is accompanied by certain histone modifications, which are recognized by conserved domains like the Chromodomain or Bromodomain.
Subunits of the PcG complexes have been found to be associated with chromatin remodeling complexes. For example, p55 is can be found in an ISWI-family complex (NURF) and a CHD-family complex (NuRD), the *Drosophila* YY1 homolog Pho is found in a complex of the INO80-family called Pho-dINO80. The table on the next page gives an overview of the chromatin remodeling complexes. Members, which are also found in the PcG group complexes are highlighted.

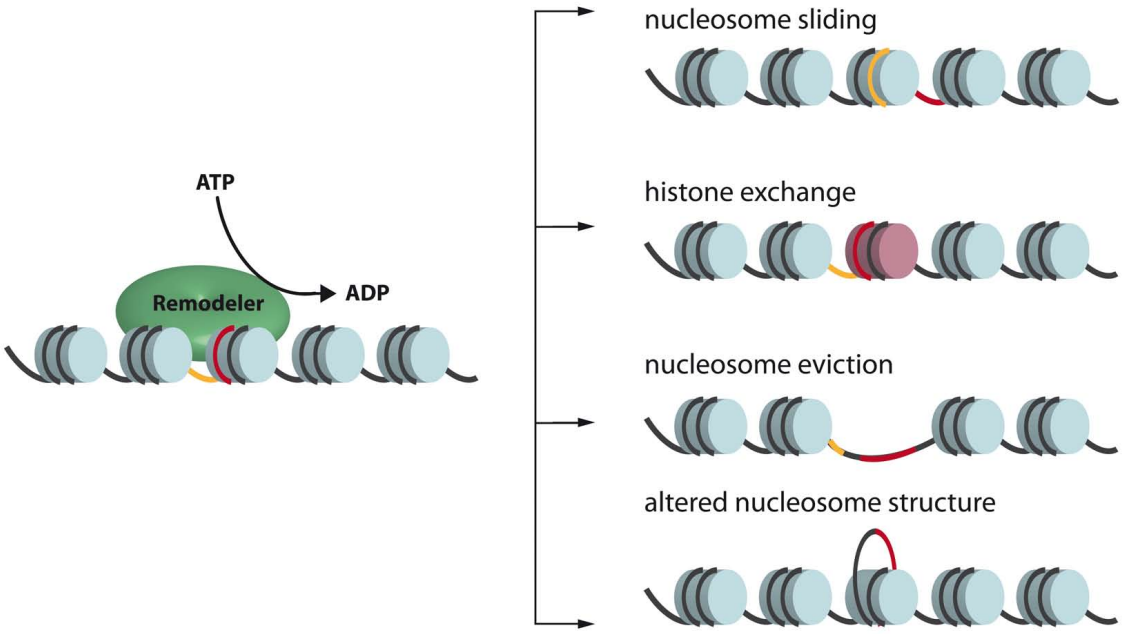


FIGURE C.1 *Chromatin remodeling by ATP-dependent chromatin remodeling complexes.* ***A*** *Nucleosome 'sliding' allows the exposure of a region (marked in red) that was occluded.* ***B*** *A variant nucleosome is created by the exchange of a variant histone for a standard histone.* ***C*** *A large region of DNA is opened by nucleosome eviction.* ***D*** *By creating a loop on the surface of the nucleosome, the nucleosome structure gets altered and sites get available (marked in red). The figure was adapted from Allis et al. (2006)*

TABLE C.1 *Chromatin remodeling complexes. Subunits, which also belong to the PcG complexes are highlighted. The table was taken from Clapier and Cairns (2009) and modified.*

Family		**Drosophila**			**Human**		
SWI/SNIF		**BAP**	**PBAP**		**BAF**	**PBAF**	
	ATPase	Brahma	Brahma		hBRM or BRG1	BRG1	
	non-catalytic	Osa	Osa		BAF250/hOSA1		
	subunits	Moira/BAP155	Moira/BAP155		BAF155, BAF170	BAF180/BAF200	
		BAP60	BAP60		BAF60a, b or c		
		SNR1/BAP45	SNR1/BAP45		hSNF5/BAF47/INI1		
		BAP111	BAP111		BAF57		
		BAP55 or BAP47	BAP55 or BAP47		BAF53a or b		
		Actin	Actin		β-actin		
			Polybromo/ BAP170				
ISWI		**NURF**	**CHRAC**	**ACF**	**NURF**	**CHRAC**	**ACF**
	ATPase	ISWI	ISWI	ISWI	SNF2L	SNF2H	SNF2H
	non-catalytic subunits	NURF301	ACF1	ACF1	BPTF	hACF1/ WCRF180	hACF1/WCRF180
		NURF55/p55	CHRAC14		RbAP46 or 48	hCHRAC17	
		NURF38	CHRAC16			hCHAC15	
CHD		**CHD1**	**Mi-2/NuRD**		**CHD1**	**NuRD**	
	ATPase	dCHD1	dMi-2		CHD1	Mi-2α/CHD3, Mi-2β/CHD4	
	non-catalytic		dMBD2/3			MBD3	
	subunits		dMTA			MTA1, 2, 3	
			dRPD3			HDAC1, 2	
			p55			RbAP46 or 48	
			p66/68			p66α, β	
						DOC-1?	
INO80		**Pho-dN80**	**Tip60**		**INO80**	**SRCAP**	**TRAP/Tip60**
	ATPase	dIno80	Domino		hINO80	SCRAP	p400
	non-catalytic subunits	Reptin, Pontin	Reptin, Pontin		RUVBL1, 2/ TIP49a, b	RUVBL1, 2/ TIP49a, b	RUVBL1, 2/ TIP49a, b
		dArp5,8/dActin1	BAP55/Actin87E		BAF53a	BAF53a	BAF53a
		Pho	dGAS41		Arp5, 8	Arp6	Actin
			dYL-1		hIes2, 6	GAS41	GAS41
			dBrd8			DMAP1	DMAP1
			H2Av, H2B			YL-1	YL-1
			dTra1			H2AZ, H2B	Brd8/TRC/p120
			dTip60			ZnF-HIT1	TRRAP
			dMRG15				Tip60
			dEaf6				MRG15/MRGX
			dMRGBP				FLJ11730
			E(Pc)				MRGBP
			dING3				EPC1, EPC-like
							ING3

Acknowledgments

I am deeply grateful to my supervisor and first referee Prof. Dr. Gabriele Petersen for the opportunity to accomplish this research. I want to thank her for the guidance throughout this thesis and for sharing her scientific knowledge with me. She allowed me great latitude in my research. I consider myself very fortunate for being one of her students.

I want to show my sincere gratitude to PD Dr. Renate Voit for being my second referee and for spending her time to evaluate my thesis.

I am much obliged to Prof. Dr. Thomas W. Holstein for the opportunity to work in his lab and the funding.

I would like to thank my colleagues Dr. Anne Kuhn, Christoph Herbel and Volker Lauschke for their helpful suggestions and their encouragement. All their kind advice and cheerfulness made the lab routine enjoyable. They initially set up the vast field of Polycombs and epigenetics in our lab and they made me feel I was not isolated with my research. I have especially appreciated the inspiring discussions, our 'Epi.com' meetings and the 'conspirative' meetings at *Mamma Leone*. I have to thank for the cloning of the *Hydra* PRC1 and PRC2 genes and for the antibody production.

I would like to thank my colleague Graciela Delgado del Moissl who always believes in the goodness of mankind and who tried to make our lab a better place. She always has some encouraging words for everybody!

I want to thank Brigitta Martin who took care of all the administration stuff. Without her, everything would end up in chaos.

I want to thank all present and former members of the Holstein lab who contributed to making the time in the lab likable for me.

I am deeply grateful to my family, my mom, my dad, my brother and my grandparents who always believed in me. Especially, I want to thank my mom who was always willing to listen to my minor and major problems and who saved me from starving during the exhausting time of writing. I want to thank my brother who helped me to understand my computer a bit better. I want to thank my grandparents for their financial support and particularly my grandma who provided me with some gorgeous food packages.