Functional analysis of *Drosophila melanogaster* linker histone dH1

Olivera Vujatovic

TESI DOCTORAL UPF / 2012

Thesis supervisor: Dr Ferran Azorín Marin

Institute for Research in Biomedicine Barcelona

The Molecular Biology Institute of Barcelona

Thesis tutor: Dr. José Francisco Aramburu Beltrán

Department of Experimental and Health Sciences, University Pompeu Fabra

The Department of Experimental and Health Sciences

Acknowledgements

Dragi roditelji, najsrdacnije vam hvala za stalnu podrsku. Toliko vaseg razumevanja za moju zelju da se bavim biologijom, svi uslovi i svestrana pomoc da se skolujem, ispunili su moje snove. Tokom godina izrade doktorata jedno su rezultati, projekat, eksperimenti i sva ta nauka. Postoji i mnogo vise od toga kao deo teze. Pridodala bih joj prilagodjavanje na novu, cudnu zemlju, nostalgiju, padove motivacije sa jedne strane i sa druge strane ushicenost zbog zanimljive konferencije, rezultata, nekog malog ili velikog uspeha. I vi ste za njih hteli uvek da znate. Znam da je svaka moja sreca, tuga ili briga bila makar duplo toliko i vasa i da nikada nisam bila zaista sama. Hrabrile su me vase reci ''plivaj Loli'' ili su me smirivale ''tiha voda breg roni''. Koliko sam samo puta pomislila na njih kao na magicne reci koje bi mi spasile dan i volju. Priveli smo kraju moju doktorsku tezu zajedno!

Vlado, hvala ti za paznju koju si mi davao kad sam morala sa nekim da porazgovaram, za sve savete kako da budem jaca, da se borim i cenim sebe. Jos od malena sam uvek igrala igre koje si ti igrao, trenirala karate, samo da budem kao ti. Klonem li nekad duhom zbog nostalgije, desavanja u nasoj tuznoj Srbiji ili zbog osecaja da sam poslednji stranac, setim se tebe, kako si i ti preko i kako si posao daleko neizvesnijim putem od mog. I tada, nema tu vise mesta samosazaljenju, znala sam da moram da se borim, kao ti.

I wish to thank to Dr. Ferran Azorín for giving me the opportunity to do PhD in his laboratory and for supervising my thesis. I am very grateful for all the time you dedicated me to talk about our project, to teach me about science in more general terms, motivate me when I was feeling down and advise me with all the kindness when it was necessary to work harder or do something better. Having a dedicated PhD supervisor with so much interest in science was a great inspiration for me and essential for my decision to continue with scientific career. Many thanks for all this.

Sincere gratitude for the members of defence tribunal of this thesis for accepting to be in the tribunal: Dr. Miguel Beato, Dr. Jordi Bernués, Dr. Alejandro Vaquero, Dr.

Albert Jordan and Dr. Travis Stracker. I am looking forward to hear your comments and questions about the thesis.

Dr. Jordi Casanova, Dr. José Ayté and, one more time, Dr. Alejandro Vaquero were following progression of my thesis during all years in the form of Thesis Advisory Committee. Dr. José Aramburu was also at these meetings as my thesis tutor. I appreciate all your suggestions, questions and discussions we had.

The Institute for Research in Biomedicine Barcelona (IRB Barcelona) where I did my PhD, was a great place to work. It created stimulating environment with seminars, conferences and other events that exposed me to a big scientific community. One could see efforts of the institute to develop and grow in their constant innovations and prizes the institute was winning. As IRB student I could count on help of the administration related to any issue. I am the most grateful to Margarita Navia, Silvia Aiguadé, Maria Rovira and Clara Caminal and all the rest of the people that make the institute a wonderful place. The Molecular Biology Institute of Barcelona (IBMB) is also a part of this atmosphere.

I was fortunate to have Alex as a person who introduced me to the laboratory work. Apart from his contribution to the project that was advancing very fast with him, his sense of humour, advices for career and translations of conversations in Spanish he did for me at the beginning were of great help to me.

Many thanks to Jordi for his help in experiments, paper and interest in the progression of my PhD. It was of big help to have somebody with knowledge needed to work with Drosophila melanogaster and broad knowledge about many other topics.

Thanks to Oscar Reina for the great contribution he had in our first paper, for efforts to explain me the analysis he was doing, for having patience to go back to our project in different moments.

I owe much gratitude to different people that were working in our laboratory during my PhD –to Esther and Estefania for so much help with flies; to Alicia for her help with experiments, her tidiness and reliability; to Salva for being an enthusiastic student with passion for experiments; to Marta Lloret for her readiness to help with any issue and for being extremely interesting lab mate; to Carles for scientific discussions and for

making me feel comfortable asking for any advice; to Katrin for help on the project in experiments and discussions and for positivism she was bringing to the lab in the mornings; to Roman for friendship, politeness and clear protocols; to Tomás for his picturesque explications of scientific problems and for willingness to help with ideas in the project; to Olga for help with cloning; to Joan Font for his critical point of view; to Eva for scientific discussions and for irradiating enthusiasm for science; to Aleix for sharing impressions about events in basketball; to Sergi for bringing leisure spirit to the lab; to Marcia Lami for making cheerful atmosphere; to Dr. Josep Portugal for interesting talks; to Sònia, Mònica, Lluisa, Anne, Lorena, Syvia Mansilla, people from Marian's laboratory, Marta y David, Rute, Silvia Perez, Marc Bataller, Bet, Gemma, Sani etc. There was not much time to get to know better all youngsters that spent few months with us or the ones that came short time before I left – George, Joan, Ujue, Lucia etc.

I am thankful to many people from the institute: to Dr. Herbert Auer for help with micro array analysis, protocols for real time PCR and for the interesting discussions; to the group of Dr. Albert Jordan for the meetings where I learned a lot about H1; to Dr. Julien Colombelli and Anna Llado from the microscope facility for development of the software we used to analyse images related to PTMs in this thesis; to Luciano and Angela from technical service for allowing me prepare special kind of food for flies that we needed in the study; to Alberto Adeva for chemical synthesis of the peptides and all questions I had about them.

Zelim da zahvalim i drustvu i rodbini koji nisu bili direktno ukljuceni u projekat, ali su mi itekako znacili: Isabella-i i Mariji, za nase druzenje, sale, nezaboravno putovanje, interesovanje za to kako sam, mejlove podrske, misljenje u vezi svega i svacega, caskanje o nauci i zivotu; Mirki Milanovic za nase razgovore na skajpu; tetki Rosi za lepe poruke iznenadjenja i zato sto je mislila jesam li dobro i kako mi ide; Nikoli Markovicu, Milosu Tatarskom i Saski Ivanovoj na druzenju; Ani Janic na prakticnim savetima i pomoci; teta Ljubici i Nadi jer mi je znacilo kada su pitale za mene i htele da me vide; Lidiji Andonovic, Jeleni Kostic, Surki i Jeleni Stanisavljevic za delenje utiska o doktoratu i nauci uopste

The last days of my thesis preparation I did in Dr. Guillaume Filion's laboratory. I want to thank him and Heng-Chang for making me these days interesting and for giving me the time I needed. I also met Roberto Blesa at the end of the thesis preparation. He made me these days cheerful and was a big support to me. Thank you for the help with the resume and more importantly, for being with me when I needed it.

Rekose mi da na kraju treba spomenuti onog kome sam najzahvalnija to su naravno moji roditelji, jos jednom mnogo vam hvala.

Contents

Abstract

We did functional characterisation of *Drosophila melanogaster* linker histone, dH1. In the mutant state for this protein, we observed structural changes in polytene chromosomes, chromocenter and nucleoli of mutant larvae. In addition, we performed a microarray analysis in H1 mutant background in order to determine contribution of dH1 to gene expression regulation. We determined effects of dH1 loss in different types of chromatin and we identified groups of differentially expressed (DE) genes, groups in sense of physical clusters of genes and genomic elements rather than groups of functionally related genes. We found that dH1 affects in greater extent expression of heterochromatin genes compared to its effect on euchromatin genes; that dH1 regulates transcription in a regional manner, since the genes physically nearest to the most DE genes tend to be upregulated as well; and that dH1 is negatively regulating expression of transposable elements and members of certain gene families. In addition, we found that dH1 is necessary for preserving genome stability. Among DE transposable elements we detected R1 and R2 retrotransposons, elements that are integrating specifically in rRNA locus. We showed that activation of their transcription is also upregulating expression of aberrant, transposon-inserted, rDNA units of the locus. In this regard we observed an accumulation of extra-chromosomal rDNA circles, increased γ-H2Av content, stop in cell proliferation and activation of apoptosis. Altogether, these results are revealing so far unknown role of histone H1 in preserving genome stability and its effects on cell proliferation.

Resumen

Este proyecto ha consistido en la caracterización funcional de la histona linker de *Drosophila melanogaster*, dH1. En el estado mutante de esta proteína observamos los cambios estructurales en los cromosomas politénicos, cromocentro y nucléolos de las larvas mutantes. Además, utilizamos la tecnología de microarrays en la dH1 mutante para determinar la contribución de la dH1 a la regulación de la expresión génica. Determinamos los efectos de la pérdida de dH1 que ocurren en diferentes tipos de la cromatina e identificamos los grupos de genes con la expresión cambiada (EC), los grupos en el sentido de las agrupaciones físicas de los genes y los elementos genómicos en vez de grupos de genes relacionados funcionalmente. Vimos que dH1 afecta en mayor medida la expresión de los genes de heterocromatina en comparación con su efecto sobre los genes de eucromatina y que dH1 regula la transcripción de una manera regional, ya que los genes físicamente más cercanos a la mayoría de los genes EC tienden a tener una expresión aumentada. También vimos que dH1 regula negativamente la expresión de los elementos transponibles y los miembros de las ciertas familias de genes. Además, encontramos que dH1 es necesario para preservar la estabilidad del genoma. Entre los elementos transponibles que cambian su expresión, detectamos retrotransposones R1 y R2, los elementos que se integran específicamente en el loci del rRNA. Demostramos que su activación transcripcional también aumenta la expresión de las unidades del loci de rDNA con los transposones insertados, las unidades aberrantes. A este respecto observamos una acumulación de los círculos extra-cromosómicos de rDNA, el aumento de contenido de γ-H2Av, la detención de la proliferación celular y la activación de la apoptosis. En conjunto, estos resultados revelan un papel nuevo de dH1 en la preservación de la estabilidad del genoma y sus efectos sobre la proliferación celular.

Prologue

Histone H1 is evolutionary well conserved, very abundant protein, found in cells of all eukaryotic species while part of prokaryotic species contain proteins homologues to H1. These three characteristics are suggesting that H1 is a protein with important roles in a cell. During the relatively long history of studies of H1, there were changes in ideas about its functions.

The existence of H1 was reported for the first time in 1951, when it was detected in various tissues of different organisms (Stedman, 1951). It was described as a basic nuclear protein similar to ″main histones″ (today termed as core histones) that however was different from them and based on its greater solubility could be partially separated from them. At that time and for a long time all histones were classified as inert proteins that organize DNA into more compact structures. A role in regulation of transcription was prescribed to histones after the experiments in which was demonstrated that they inhibit RNA synthesis in isolated thymus nuclei (Bonner & Huang, 1963) and that in their absence there is greater mRNA synthesis (Allfrey et al, 1963). When it comes to H1, it was shown that it occupies the most condensed regions in polytene chromosomes of *Drosophila melanogaster* and that these regions are not occupied by RNA polymerase II (Jamrich et al, 1977). Following studies of transcription *in vitro* defined linker histone as a general repressor of gene expression, since H1 addition to the transcriptional reaction inhibited the transcription (Hannon et al, 1984). This was a long-standing belief that was modified after more reliable *in vivo* analysis have challenged this belief and shown that H1 is actually a regulator of specific genes. The

number of genes whose expression H1 was affecting was limited in the majority of cases. There are couple of examples where even a molecular mechanism by which H1 is accomplishing it was determined (L.-jung Juan et al., 1997**,** Nishiyama et al., 2009). It was shown that those are infrequent examples so once again a role of H1 as a structural element is gaining importance against its role as a general regulator of transcription. Crystal structure of linker histone globular domain was determined (Ramakrishnan et al, 1993) and different models for the way H1 is binding to DNA to compact it are proposed. It is also becoming clear that posttranslational modifications of linker histone can determine its functions just like it's the case with core histones. However, much more results were obtained for core histones modifications.

We decided to address the question of H1 functions in *Drosophila melanogaster* because this is the only multicellular organism with a single H1 variant in the genome. We made a dH1 mutant condition and observed various phenotypes that led as to conclusions about different roles of dH1. We also decided to study a role of a particular modification of dH1 by raising antibodies that specifically recognize the modification.

To present and discuss the results, an introduction about the chromatin and how H1 is compacting it; about H1 gene and protein characteristics and finally about functions that have already been attributed to this protein will be presented first.

Introduction

Chromatin

Genetic and epigenetic material of cells is placed in nucleus in the form of chromatin. Chromatin contains DNA, histones and other non-histone proteins. While DNA is a carrier of protein and non-coding RNA sequences, protein portion of chromatin is necessary for structuring and providing proper functioning of DNA. Only together, DNA and chromatin proteins, in a cell allow cell growth, division, specific differentiation, cell death etc. To accomplish these activities, chromatin needs to answer to stimuli from cell interior and environment by constant modulation of chromatin and its functions - gene expression, DNA replication, DNA repair and others.

The most abundant chromatin proteins are histones. Histones are relatively small proteins that directly bind DNA and lead to the first level of DNA packing. Due to the positive charge of histones, these proteins have great binding affinity for negatively charged DNA. The positive charge comes from numerous basic Lysines (Lys) in their sequence. There are five main types of histones: linker histone H1 and core histones: H2A, H2B, H3 and H4. Protein structure and domains of core histones are similar among them and conserved through the evolution. The main part of core histones is central globular core with histone fold domain. The rest of the proteins are consisting of long N- and short C-terminal tails. Linker histones have winged helix domain in their globular domain and the rest of the protein is represented by N- and C terminus.

Chromatin packing

Long DNA molecules have multiple levels of folding by chromatin proteins that allow placing into micronic-size nucleus. Those are nucleosomes (beads on the string structure), 30 nm chromatin fiber and other higher order chromatin structures.

Packing of DNA starts with DNA wrapping around histone octamers. Histone octamers contain two tetramers of core histones, each of them carrying H2A, H2B, H3 and H4. Approximately 146bp long DNA wrapped around the octamer forms a nucleosome. Crystal structure of the nucleosome has been determined at 7Å (Richmond et al, 1984) resolution when it was showed that it has a disc-like shape and in 1997 at 2.8Å resolution more details were obtained (Luger et al, 1997). For simplifying a model all core histone components were expressed and purified from bacteria so they were missing any posttranslational modification and DNA was of defined sequence. The main part of the structure was determined although part of histone C-terminus regions is not included in the model. Finally in 2002 researchers made a closer look to nucleosome with 1,9Å resolution (Davey et al., 2002) (Figure 1).

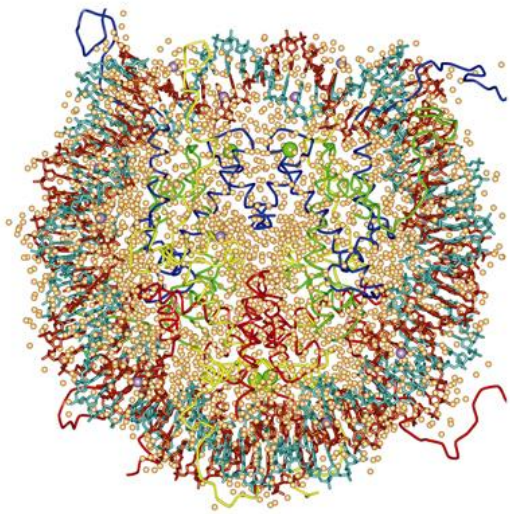

Figure 1: Nucleosome core at 1,9Å resolution

Array of nucleosomes forms 11nm wide, “beads on the string” structure (it is called like that because the array resembles it when looked under an electron microscope). Between adjacent nucleosomes extends linker DNA where linker histone H1/H5 can bind and further compact DNA into 30 nm structure. Binding of linker histones is increasing spacing between nucleosomes, termed as nucleosome repeat length (NRL). There is a linear relationship between H1 stoichiometry and NRL (Woodcoc et al, 2006). H1 stoichiometry is variable from one cell type to another and in many of them not all nucleosomes are occupied by linker histone (splenocytes contain 0,79, thymocytes 0,83 H1 molecule per nucleosome (Fan et al., 2003)) which is possibly facilitating chromatin modulating and access to chromatin by regulatory complexes. Using *in vitro* conditions with high concentration of salt, 30 nm fiber structure can be assembled even in the absence of H1. The exact arrangement of DNA and histones even at this level of chromatin structure is still incomplete (reviewed in Li and Reinberg, 2011). There are two models trying to explain the positioning of chromatin elements at this level: 1) one-start helix (solenoid) and 2) two-start helix (zig-zag) model. Both models are based on studies on *in vitro* reconstituted chromatin to avoid obstacles of native chromatin: various types of DNA sequences, posttranslational histone modifications, histone variants and irregular nucleosome spacing.

Bases for proposing one-start helix (solenoid) structure model were obtained by electron microscope of long nucleosomal arrays of different repeat length with or without the presence of linker histone and salt. In this model nucleosomes are proposed to be connected with incurved linker DNA so that it delineates a complete circle within about six to eight nucleosomes.

The second model, two-start helix model, has been proposed based on electron microscope and X-ray structure at 9Å resolution obtained with tetranuleosomal array without the presence of linker histones, providing a proper salt concentration (Schalch et al., 2005). A longer, continuous chromatin fiber model was built by successively stacking tetranucleosomes one on another. In this model, linker DNA is connecting nucleosomes that are distributed in zig-zag mode.

When a structure of long regular chromatin fibers in different solutions and in the presence of linker histone was analysed by EM and cryo-EM, results supported the

existence of one-start helix model (Robinson & Rhodes, 2006). In the same study, in the absence of linker histone, the resulting fiber structure was highly disordered, more similar to the two-start helix model. This is suggesting that a general and unique model probably does not exist and that different types of chromatin exist even in the same cell, depending on the conditions in the local environment.

Little is known about the chromatin structure beyond 30nm fiber level. There are intra- and inter-fiber interactions between nucleosomes which *in vitro* conditions are observed in the presence of Mg2+. It has been suggested that interactions between adjacent nucleosomes, between histone H4 tail domain with H2A and H2B, can be achieved not only in the same fiber, but between adjacent fibers too, leading to higher order chromatin packing (Gordon et al, 2005).

Chromatin organization

Traditionally, in chromatin there are two main types - open, transcriptionally active part, euchromatin, and closed and transcriptionally silence part, heterochromatin. These two main types of chromatin differ in DNA sequence, in the presence of certain histone variants, degree of DNA methylation, presence of histone posttranslational modifications (PTMs) etc.

Euchromatin is not stained or is lightly stained in G-banding staining. This is an open fraction of the chromatin, available to transcriptional machinery and rich in genes. Besides being hypo DNA-methylated, euchromatin is marked by numerous histone modifications: di-methylation of Lys at position 4 in histone H3 (H3K4me2), tri-methylation at Lys79 in H3 (H3K79me3) and hyper acetylated core histones (H3K14, H4K16). During S phase, euchromatin is replicated before and faster than heterochromatic regions (Lima-De-Faria, 1959).

Heterochromatic DNA sequence was originally described as densely stained part of chromatin by G-banding staining. It is characterised by the presence of repetitive sequences – (satellite DNA) mainly at telomeric and centromeric regions. Its high compaction makes it hardly accessible to transcriptional machinery. However, several

genes are placed in heterochromatin and are expressed (Smith et al, 2007). It is the part of the genome which is replicated later then euchromatin, at the end of S phase (Lima-De-Faria, 1959). In nuclei, heterochromatin is placed at the nuclear periphery and perinuclear area (Cremer & Cremer, 2001) (Towbin et al, 2009). For a long time it has been believed that this heterochromatic sequences have no role in cells and it has been unfairly named as "junk DNA". It turned out that this part of the genome has important roles (Weiler & Wakimoto, 1995) in regulation of transcription , mostly gene silencing of repetitive DNA elements, by providing non coding RNAs. In addition, it has been demonstrated that it contributes to normal centromere and telomere functioning as well as to chromosome pairing during meiosis. The main histone modifications associated with heterochromatin include methylations: H3K9me3 (Peters et al., 2003), H3K27me3 and H4K20me3 and general histone hypo-acetylation. Besides these histone PTMs, DNA methylation is also heterochromatin mark. The main non-histone protein present in heterochromatin is heterochromatin protein 1 (HP1).

HP1 protein has been first identified in *Drosophila melanogaster* (James & Elgin, 1986), but their homologues have been determined in many other species, including humans. In *D. melanogaster* there are three types of HP1 protein, but only HP1a and HP1b can be found in heterochromatin while HP1c is found in euchromatin. HP1 molecules contain chromo domain that recognizes methyl group of H3K9me histones. HP1 is a dominant suppressor of position-effect variegation (PEV) (Wustmann et al, 1989). PEV is a phenomenon of euchromatin gene silencing once they are placed into heterochromatin by genetic manipulation (Muller, 1930). When proteins that are involved in heterochromatin formation and functioning (like HP1) are mutated, heterochromatin cannot silence the gene so that PEV effect is not happening and these proteins are called suppressors of PEV.

Types of heterochromatin – classical types

The classical view on heterochromatin distinguishes two types of heterochromatin: facultative and constitutive.

Facultative heterochromatin is silenced chromatin which however can be converted to open chromatin state when necessarily. Here belong genes whose expression is needed in particular periods during cell cycle, differentiation, development or only under certain environmental conditions. In different cell types of an organism these heterochromatic regions do not necessarily correspond because of their different differentiation programs.

Constitutive heterochromatin is permanently silenced through the cell cycle and is common to different cell types in an organism. It contains highly repetitive sequences, such as clusters of satellite sequences or transposable elements that are usually found around telomeres and centromeres.

Types of chromatin – newly characterized types

The recent genome-wide studies (Consortium 2011 and Filion et al., 2010) addressing a question of histone modifications marks and other non-histone chromatin proteins distribution along genes in genomes of different organisms, in *D. melanogaster* as well, put a new light on chromatin types. In this way, today we can distinguish even five main types of chromatin in *Drosophila melanogaster* instead of only two main known before (Filion et al., 2010).

Each of them contains a specific combination of chromatin proteins. Among classical euchromatin portion of the genome, there are actually two chromatin types, named YELLOW and RED chromatin. They share characteristics of classical euchromatin in the sense that they are substantially transcribed, that they contain similar posttranslational modifications of histones (except for H3K36me3 which is more abundant in YELLOW chromatin) and that they share many non-histone chromatin proteins – histone deacetylases (HDAC) RPD3 and SIR2, DF31, ASH2 and MAX. There are also proteins that are not shared between them and there is also difference in their gene composition. YELLOW chromatin mainly contains genes that are expressed in many different tissues and during many embryonic stages while RED chromatin

contains genes that are determining cellular functions that are specific for certain cell types and that require more complex gene regulation.

This novel view on chromatin types distinguishes three distinct heterochromatin kinds: BLUE, GREEN and BLACK chromatin.

The most characteristic components of BLUE chromatin are Polycomb Group (PcG) proteins, proteins involved in regulation of developmental genes (Schuettengruber et al, 2007). For the first time they were discovered in *Drosophila melanogaster* as regulators of developmental, *Hox* genes, and their orthologs in mammals have been determined. *Hox* genes expression is initiated early in embryonic development and PcG proteins are necessary for maintaining their expression in later embryonic stages. These proteins bind Polycomb response elements (PREs) and act as part of three complexes – PRC1, PRC2 and PhoRC. PRC2 has methyltransferase activity, it trimethylates Lys at H3K27 position. H3K27met3 is the main histone modification mark of BLUE chromatin.

GREEN chromatin is present in pericentric regions and on chromosome 4. The components that distinguish this chromatin from other chromatin types are SU(VAR)3-9, H3K9me2, HP1 and HP1 interacting proteins, LHR and HP6.

BLACK chromatin covers long domains in the genome and presents 48% of the portion of the genome that was tested. It is characterised by very low number of genes and their low transcriptional activity. RNII18 protein, a subunit common to all three RNA polymerases, is poorly represented in BLACK chromatin, indicating a low transcriptional activity in this portion of the genome. When a transgene is placed in this chromatin type, it gets actively silenced. The prevalent proteins in this type of chromatin are linker histone H1, D1, IAL and SUUR, while LAM, SU(HW) and EFF are also frequently encounter in it. Presence of LAM is suggesting that nuclear lamina is playing a role in regulation transcription which is in accordance with the observation that peripheral chromatin in nuclei is silent (Towbin et al, 2009) (Cremer & Cremer, 2001).

Heterochromatin assembly

HP1-dependent and H3K9me-dependent heterochromatin assembly

As already mentioned, HP1 is recruited to heterochromatin by recognition of methyl-group in H3K9me histones by its chromodomain. In *Drosophila melanogaster* HP1 is then recruiting SU(VAR)3-9, a methyltransferase that introduces H3K9 methylation mark. In this way a mechanism of positive feedback in heterochromatin assembly is initiated and can be spread, since there is a successive binding of HP1 for methylated histones and as a consequence recruitment of methyltransferase that introduces the same mark to the adjacent nucleosome until the appearance of boundary elements that stop the spreading (Figure 2). Similarly, in *Saccharomyces pombe* there is methylation of H3K9 with Clr4 (a yeast homologue of SU(VAR)3-9) and recruitment of Swi6 (yeast homologue of HP1) which provokes spreading of H3K9 methylation mark. Additional proteins that will form heterochromatin can be then recruited, like histone deacetylases (HDAC) that need to erase acetylation before the introduction of methylation mark by histone methyltransferases (HMTs) on target histones.

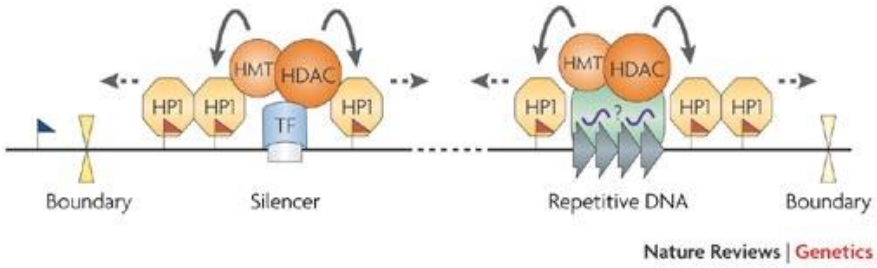

Figure 2: Heterochromatin assembly dependent on HP1 and H3K9me

RNAi-dependent heterochromatin assembly

RNAi-dependent assembly of heterochromatin (Figure 3) is mainly studied in *S. pombe*. One of the key players in this type of heterochromatin formation is RNA-induced transcriptional silencing (RITS) complex (Grewal, 2011).

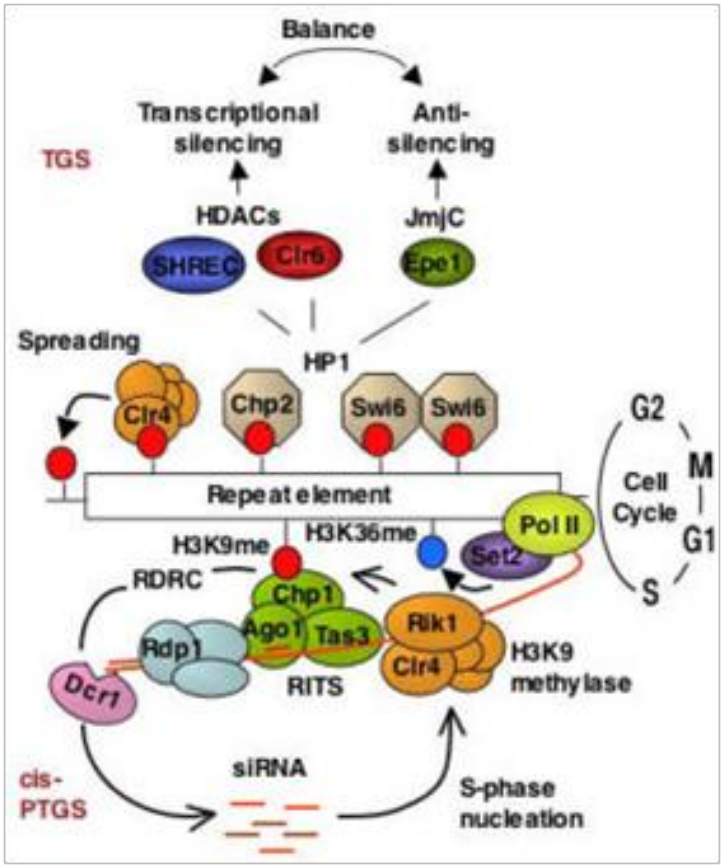

Figure 3: RNAi-dependent heterochromatin assembly

This complex contains Argonaute 1 (Ago1), heterochromatin-associated chromodomain protein (Chp1), Targeting complex subunit 3 (Tas3) and siRNA that correspond to *dg* and *dh* pericentromeric repeats (Grewal, 2011). RITS complex is targeted to heterochromatin by Chp1 subunit that recognizes H3K9me (Schalch et al., 2009).

Binding of RITS and Swi6 provides binding of RNA dependent polymerase 1 (Rdp1) that forms dsRNAs from nascent centromeric transcripts. These dsRNAs are then processed by Dicer1 (Dcr1) into siRNAs. It is not know how exactly siRNAs are promoting formation of heterochromatin, but it is possible that they recruit Clr4 that introduces H3K9me heterochromatin mark.

Heterochromatin sequence

Large portion of heterochromatin is presented by repetitive elements. Among them are short repeats (like satellite DNAs), middle repetitive elements (like transposable elements and ribosomal RNAs) and some single-copy DNA (Hoskins et al., 2007). Stability of repeated elements is essential for maintaining whole genome stability. The reason for instability of repetitive sequences comes from the difficulty in their replication that can provoke change in the number of the gene copies. Namely, the presence of tandemly repeated sequences can provoke slippage of replication fork along the sequences that have homologous or identical parts (Pearson et al, 2005). This can lead to sequence alterations, replication fork stalling and, as a final result, double stranded breaks (Figure 4). The other crucial point during cell cycle important for repetitive elements is the moment of homologous chromosomes pairing, when recombination can occur. Since repetitive sequences can accomplish recognition and pairing at different repetitive units, it is easy to imagine that these misalignments can lead to unequal recombination, generating duplications on one and deletions on other chromosome. Recognition can also occur between the homologues sequences located at the same chromosome, when recombination is resulting in excision of the part of DNA between the two sequences and extrachromosomal circular DNAs are formed. ecc DNAs are inherited instability and result in the loss of the genes placed in the circles in one of the daughter cells. Recognition between repetitive sequences can occur between units located at different chromosomes when recombination can result in greater chromosomal aberrations, like translocations, loss of parts of chromosome arms etc (Pearson et al, 2005).

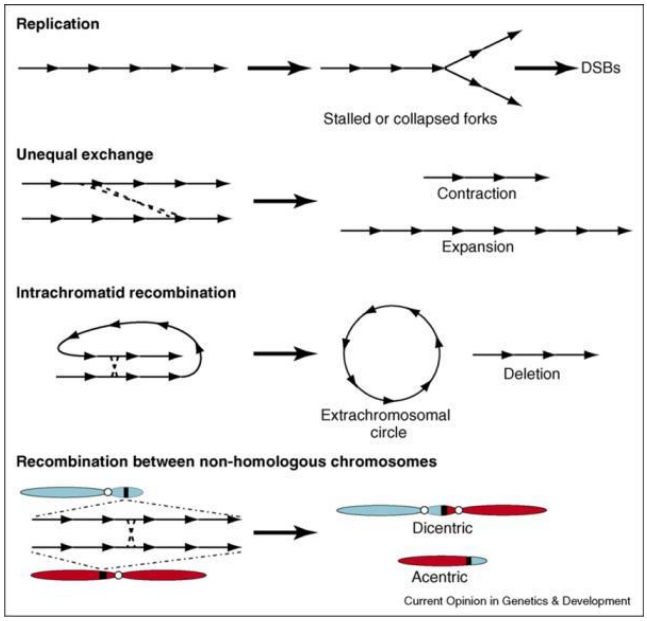

Figure 4 Tandemly repeated sequences can provoke genome instability

RNAi and SU(VAR)3-9 together with H3K9me mark are important in this sense. They have been shown to be essential for preserving stability of repetitive sequences and consequently, stability of the whole genome (Peng & Karpen, 2006). Loss of SU(VAR)3-9 or RNAi pathway causes decrease in H3K9me level and changes in organization of repetitive elements. A unique wild type nucleolus which contains repeats of rRNA sequence is split in many smaller ectopic nucleoli in the mutant. Satellite repeats in wild type are also localised together, in chromocenter, but in the mutant they also show mislocalization and presence in various sites in the nucleus. Both rRNA and satellite repeats are forming extrachromosomal circular (ecc) DNA in the mutants. This is showing that the general stability of repetitive sequences is dependent on H3K9 methylation and RNAi pathway, i.e. heterochromatin formation.

Linker histones

Structural features of linker histones

Linker histone of metazoans contains central globular domain and long C- and short N-terminal tails. Several features of H1 structure allow this protein dynamical, regulatory prone interaction with linker DNA, which is necessary for chromatin compaction. Among these features are presence of numerous Lysine (Lys) rich in protein sequence, presence of AKP amino acid motifs, existence of winged helix motif and characteristics of N and C terminus of the protein.

The presence of numerous Lys in H1 protein sequence gives a positive electrical charge that allows interaction with negatively charged phospho groups exposed on DNA molecule surface. Lys provides a great advantage of this interaction, dynamism, since H1 can bind DNA, but if necessarily, it can easily dissociate and permit access of other proteins to the DNA. Lys showed to be much more useful in this sense than Arginine (Arg) (Kasinsky et al, 2001.). Arg is also positively charged amino acid enriched in some other DNA binding proteins, like protamines. The interaction that Arg accomplishes with DNA is much tighter and leads to more rigid interaction with DNA. To dissociate from DNA and allow decondensation, protamines need a presence of another protein deriving from egg. Lys on the other hand establishes more loose interactions with DNA which makes them more dynamic.

Together with Lys, Alanines (Ala) and Prolines (Pro) play a critical role in providing H1 functions. Namely, line of these three amino acids (AKP) in H1 is repeated many times and this results in a special distribution of charge which makes H1 able binding DNA and it makes it amphipathic molecule (Subirana, 1990). Being amphipathic could

permit interaction with other H1 molecules and consequently their cooperation in chromatin folding.

Globular domain of linker histones has winged helix domain (WHD) (Figure 5) responsible for specific recognition of the four-way junction of DNA (4WJ) (Varga-Weisz et al., 1993). WHD contains three α-helices and a ß-hairpin at the carboxyl terminus, with a short ß-strand situated between helices I and II. The name ‘’wings’’ comes from the resemblance of large loops that connect structures of the proteins containing them to wings. It is believed that topology of DNA at the sites of entrance and exit into and from nucleosome resembles 4WJ structure (Lilley, 1992). In that way, linker histone recognizes DNA specifically at the sites where it enters and exits nucleosomes, at the sites of linker DNA.

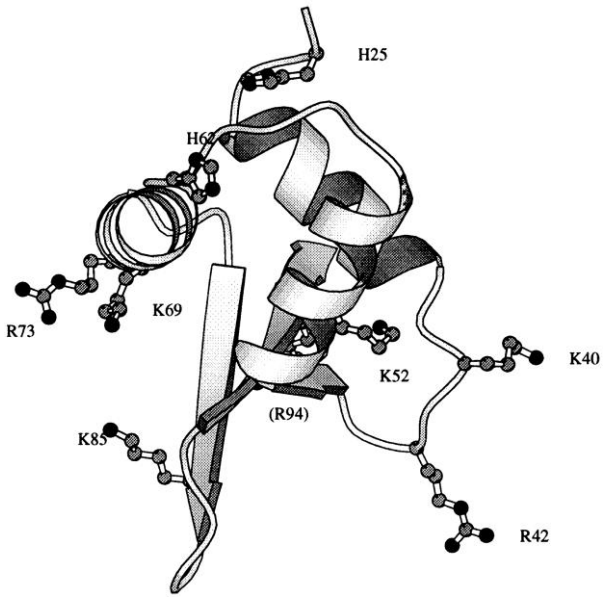

Figure 5: Crystal structure of globular domain in H1

C-terminus is evolutionary the most variable part of H1 protein (H E Kasinsky et al., 2001). Even so, it has a very important role in accomplishing H1 binding to chromatin and it is important that C-terminus of the same corresponding isoforms in different species of mammals are actually well conserved (Hendzel et al., 2004). C-teminus contains various positively charged amino acids and it is believed that this helps binding of H1 because it neutralises DNA charge (Dou et al., 1999). When targets for kinases in

C-terminus of human H1.1 isoform are mutated, this leads to a decrease in the binding affinity and residence time of the isoform on chromatin (Hendzel et al, 2004). Nonetheless, single amino acid mutant in human H1.1 isoform affects more binding of the histone to chromatin than deletion of the region that contains numerous positively charged amino acids (Hendzel et al, 2004). This is suggesting that not only a charge neutralisation, but most probably, also formation of specific structure in C-terminus affects H1 binding to chromatin. In relation to this, it is important to mention that C-terminus in mammals forms α-helical structure upon binding to chromatin (Vila, Ponte, Collado et al., 2001).

Little is known about posttranslational modifications that are decorating amino acids of N-terminus in linker histones. There are reports that phosphorylation of H1 is affecting its binding affinity to DNA (Dou et al., 1999), providing a fast way to change accessibility of DNA to various enzymes. Further text (section Posttranslational modifications) contains more information about PTMs in linker histones.

Evolution of linker histones

Appearance of the first proteins similar to linker histones in evolution happened in *Eubacteria*. Several of them contain in their genomes genes for basic proteins that have similarities to C-terminus of typical metazoan linker histones. These proteins have high Ala and low Lys content compared to canonical H1.

Certain groups of protists (like *Kinetoplastids*) also have proteins that correspond only to a portion of H1, again only to its C-terminal domain. In some other groups of protists (*Mycetozoans*) for the first time in the evolution emerged winged helix domain (WHD) as a part of linker histone protein (this domain has already existed in other proteins, in some transcription factors).

Apart from the emergence of WHD in H1 of *Mycetozoans* and its successors in evolution, WHD has emerged at least one more time independently in evolution, in ancestor of *Chlorophyta*.

As a possible event that led to the formation of more complete H1 protein in evolution, it has been suggested a fusion event between the carboxyl terminal domains of H1 related proteins and the proto-WHD of H1 related proteins.

Further evolution of H1 domains happened by different means. Globular domain has evolved by nucleotide substitutions while N- and C- terminus have evolved both by nucleotide substitutions and by accumulation of insertions and deletions events.

High incidence of insertion and deletions events are reflected in high variation in N- and C- terminus length through the evolution in organisms that contain all three H1 domains (40+-13 for N-terminus and 106+-17 for C-terminus compared to 79+-5 amino acids for globular domain). Significant simplicity of amino acids content and DNA sequence of the two termini of histone H1 (Ponte et al, 2003) can explain this high insertions and deletions events. Simple sequences (similarly to repetitive sequences) are prone to misalignement during occurrence of DNA functions – replication, repair and recombination (Pearson et al, 2005). The misalignment then can result in skipping or repeating replication of part of the sequence or in unequal crossing-over, all of them resulting in insertion or deletion events.

Evolution resulted in different outcomes in different species so today there are linker histones with two globular domains in *Saccharomyces cerevisiae*, linker histones with no globular domain at all in *Tetrahymena thermophila* etc. In majority of species it contains three domains and has numerous isoforms formed by process of gene duplication. Further text (section H1 isoforms) contains more information on this topic.

It is interesting that core histones have completely different origin from linker histones. In evolution, core histones appeared for the first time in *Archaebacteria*. There are suggestions that core and inker histones might unite in a single organism by lateral gene transfer, a common way of evolutional changes in *Eubacteria* and *Archaebacteria* (Doolittle, 1999).

H1 isoforms

Linker histones have numerous subtypes in different species. While there are very few species with a single H1 variant (*Saccharomyces cerevisiae*, *Tetrahymena thermophila, Drosophila melanogaster*), far most of the species have more H1 isoforms. Chicken for example contains six H1 isoforms while mice have five somatic in addition to two germ-line specific isoform, replacement linker histone $H1^0$. The variants can be specific for certain stage in development, for somatic or germ-line or for a tissue. Perhaps the most interesting example in this sense is spermatogenesis in mouse where there is a successive change of H1 isoforms all along maturation of gonad cells (Godde and Ura, 2009).

Humans have eleven different linker histone proteins. H1.X is ubiquitously expressed, five isoforms are present only in somatic cells (H1.1, H1.2, H1.3, H1.4 and H1.5) while the others are present only in certain tissues or organs (H1t, H1T2 and HILS1 present in testicle or H1oo present in oocytes) or in terminally differentiated cells of various tissues (replacement H1 subtype, H1.0). The subtypes differ among themselves at DNA and amino acid sequence level. The somatic type H1 genes (H1.1-H1.5) together with H1t gene are present in clusters with core histones genes, mainly on chromosome six. These genes are not simultaneously expressed. It has been proposed that this is accomplished by the specific chromosomal organization of the genes and by their different promoter structures (Doenecke et al, 1994).

Two possible explanations have been proposed for the existence of various H1 variants (D. T. Brown, 2001)**.** The existence of different H1 variants could reflect a need of a cell for H1 molecules with different properties (like providing a particular nucleosomal architecture). The other explanation would be that the cell initially contained numerous copies of the unique H1 gene as a way to answer a demand for great amount of the protein (gene dosage), particularly in S phase of a cell cycle. Consequently during evolution diversification of different gene copies occurred and they gained specific functions.

In any of these two cases, it is likely that H1 variants do play specific roles in cells although a common role in chromatin compaction can be attributed to all of them. However, little is known about particular roles that each of them might be playing in a cell. What has been demonstrated so far is that the variants show different characteristics in various terms – difference in turn-over rate (Cole, 1987), different moments of synthesis (Higurashi et al, 1987) and different phosphorylation level (Talasz et al, 1996). In addition, it has been shown (Orrego et al., 2007) that mammalian H1 variants have different affinity for DNA (scaffold-associated region) and chromatin (native chromatin isolated from cells). These experiments consisted of competition assays where was examined relative affinity of six rat H1 subtypes to bind DNA or chromatin. H1a showed the lowest binding affinity, H1b and H1c intermediate and H1e, $H1^{o}$ and H1d the highest binding affinity. There is a high, 19 fold-change, difference between the highest and the lowest binding affinity, suggesting a functional relevance of distinct binding abilities for H1 variants. Another study (Parseghian et al, 2000) has demonstrated that inactive and active chromatin are occupied by different ratio of human H1 variants. This group used the antibodies specific for four different H1 subtypes to immunoprecipitate genes that are known to be actively transcribed or silent. They have determined that the active chromatin contains less H1.2 and H1.4 and more H1.3; while inactive chromatin contains all four somatic H1 subtypes.

There are additional examples that encourage the assumption that H1 subtypes play specific roles in a cell. It has been determined that H1 subtypes differ in their ability to condense chromatin *in vitro* (Talasz et al, 1998). H1.3 could condense chromatin even when present in smaller quantities (1.5 times less then H1.2 or H1.4).

It was also found that H1t binding to chromatin has particularities. Namely, H1t binding to chromatin *in vitro* leaves it in more open state (which physiologically could have a significance in allowing recombination and facilitated replacement of H1t with protamines that should occur) (Talasz et al., 1998).

Further example that illustrates functional differences between two H1 subtypes is encountered in *Xenopus laevis* 5S rRNA gene. In the oocyte cell only oocyte specific 5S rRNA gens are expressed while early in embryo development both somatic and oocyte specific 5S rRNA genes are expressed. Later in embryogenesis, after mid-blastula

transition, there is a suppression only of oocyte specific genes. The suppression occurs at the same time when there is a change in H1 subtype in embryo, when germline-specific H1 (H1M or B4) is replaced by somatic H1 subtype. The experiments in which somatic H1 protein level was changed showed that this subtype is necessary for specific oocyte gene silencing (Kandolf, 1994) (Parseghian et al., 2000). It is also known that H1 in this case is blocking transcription by preventing nucleosome mobility and it seems that this can accomplish only a specific H1 subtype.

An obstacle that researches encounter in efforts to determine roles of specific H1 variants is that the variants are redundant. Isoforms play common roles in a cell and this as a consequence leads to an overexpression of the other subtypes and functional replacement of the missing variant (when it is eliminated by genetic manipulation). In this way, knock out experiments did not result in a real decrease in total H1 level so that the phenotypes could not be observed. For instance, knock out of H1c and H1e in mice did not give rise to any change in mice because of the compensatory effects of other H1 isoforms (Fan et al., 2003). Only with triple knock out problems could be observed. More successful try was in human breast cancer cells (Sancho et al, 2008). Namely, in the knock down of any particular H1 variant there was no compensatory increase in expression of other H1 subtypes. This study has pointed on the role of particular H1 subtypes. For example, H1.2 was shown to be the only variant to change a nucleosome repeat length. In addition, among all variants, H1.2 was the one causing the greatest downregulation in gene expression. H1.4 is the only variant essential for cell survival. It is noteworthy that depletion of specific H1 subtypes affected expression of mainly distinct group of genes, supporting the belief that H1 subtypes have distinct functions.

Nevertheless, presence of numerous H1 subtypes is a disadvantage of these model systems. It seems difficult to predict in which extent compensatory effects of the isoforms affect phenotypes one can detect and finally, if isoforms ever allow a sufficient H1 depletion to observe consequences of almost total linker histone loss. In this sense, *Drosophila melanogaster* is good choice for studying H1 since it is only multicellular organism that contains a single H1 variant.

Histone H1 in *Drosophila melanogaster*

In *Drosophila melanogaster*, dH1 gene is placed in a gene cluster together with four core histones genes. This quintet gene unit (one linker and four core histones genes) is repeated about one hundred of times in a haploid genome (Matsuo et al, 1989). dH1 gene is consisting of TATA-less promoter, 839bp long sequence without introns and at distal stem-and-loop structure (Marzluff, 2005), while polyadenylation signal is absent.

It has been demonstrated that TATA-less promoter of dH1 gene is transcriptionally regulated by TRF2 (TBP (TATA-box-binding protein)-related factor 2) (Isogai et al, 2007). On the other hand, core histones genes, placed together with dH1 gene in the common gene cluster, have a distinct transcriptional regulator, TPB (TATA-box-binding protein) / TFIID. Specific binding of TRF2 exclusively to dH1 gene promoter and TBP for core histones promoters have been proven by ChiP analysis. In addition, depletion of TRF2 affects negatively specifically level of dH1 mRNA (while not affecting mRNAs of core histones) and affects negatively expression of reported gene regulated by dH1 promoter. The existence of different regulators for linker and core histones is allowing a cell to provide distinct ratio of linker and core histones expression which is needed in different cells or tissues (Holmgren et al, 1985, Ner & Travers, 1994). It is also interesting that binding of PTB to core histone genes is increased in S phase of KC cells, while PTF2 is binding to dH1 promoter equally through the cell cycle (Isogai et al, 2007). Core histones also have a common negative regulator of transcription, Abnormal oocyte (Abo) protein. Again, this regulator of transcription is not shared between core and linker histones, since it was demonstrated that Abo is not controlling expression of dH1 gene, at least in early embryogenesis (Berloco et al, 2001).

It is interesting that dH1 gene sequence in *Drosophila melanogaster* lines deriving from different locations across the world (central Africa, North America, Australia and Japan) is identical (except for the three nucleotides preceding the start codon) (Nagel & Grossbach, 2000). Great physical distance of these *Drosophila melanogaster* lines

means a genetic separation for a time long even to accumulate gene polymorphisms. However, as mentioned, open reading frame of dH1 gene is completely monomorphic. This could be interpreted by a particular localisation that histone genes are occupying in *Drosophila melanogaster* genome. Namely, they are on chromosome II, locus 39D/E, very close to the centromere heterochromatin, where recombination rate is very low. To this day, poorness of recombination preserved the original dH1 gene sequence that was present in a common ancestor of all *Drosophila melanogaster* lines from any genetic change. In *Drosophila virilis*, on the other hand, histone genes are located in a region more prone to recombinations, which during evolution resulted in genetic polymorphism of histone genes, gene duplications and resulting existence of three linker histone subtypes. A particularity of almost all histone genes, dH1 gene among them, is that they do not contain introns. Only some germ line specific H1 isoforms contain them and they are present in rare examples, like in *Xenopus laevis* linker histone B4 (Cho, 1994).

A typical metazoan gene at the end contains a sequence which is a signal for polyadenylation. The signal is recognized by polyadenylate polymerase that introduces polyA tail (stretch of 20-250 adenosine monophosphates) at the end of mRNA transcribed from the gene. PolyA tail protects it from degradation and promotes mRNA export from the nucleus to cytoplasm and its translation.

In contrast to this, the majority of histone genes (all except from H3.3, H2a.Z, H1o, H3-cid and macroH2a genes) and linker histone genes among them contain at the end of the gene a stem-and-loop structure (Marzluff, 2005). The stem-and-loop contains a six-base stem, four-base loop and 25-26bp AC-rich sequence (histone downstream element, HDE). After synthesis of H1 mRNA stem-loop binding protein (SLBP) binds stem loop and helps binding of U7 snRNP to complementary HDE. U7 is them recruiting other necessary proteins (ZP100, Lsm10, Lsm11) that results in 3' cleavage of HDE part of histone mRNA. After fast transport of mRNA to cytoplasm, translation occurs.

There is a strong regulation of the level of histone mRNA level mediated by posttranscriptional changes at 3' end of the mRNA. If cell replication is blocked this will provoke degradation of the histone mRNA by activity of 3' Exo exonuclease. SLBP is protecting 3' end of the mRNA from degradation and is a necessary component

in histone mRNA processing. In this way level of SLBP is regulating the level of mature mRNA. SLBP is synthesised before entering S-phase and degraded very fast at the end of S-phase of cell cycle.

Histone H1 protein

In early *Drosophila melanogaster* embryo there are maternally inherited mRNA for dH1, but they are not translated before nuclear division seven (Sarbjit S Ner & Travers, 1994) so that the first stages of *Drosophila melanogaster* development are occurring in the absence of linker histones. In this period of development there are numerous replications and fast condensation and decondensation is needed. The presence of dH1 would be an obstacle to these cyclic changes of condensation level. On the other hand HMG-D proteins are more useful in this sense. HMG-D mRNAs are maternally inherited and HMG-D proteins can function as linker histones since they protect linker DNA from nucleases (Ner & Travers, 1994). It has been demonstrated that they alter nucleosome repeat length (NRL) when incorporated into reconstituted chromatin (Ner et al., 2001). The level of NRL extension by HMG-D is smaller than level of extension by histone H1 and this could determine the level of chromatin compaction that these two proteins can provide. Namely, it has been suggested that there is a correlation between nucleosome spacing and level of chromatin folding (Blank & Becker, 1996). Increased nucleosome spacing (longer nucleosome repeat length) is a characteristic of more compacted chromatin. That could mean that HMG-D containing chromatin is less compacted and could provide faster decondensation than dH1 containing chromatin.

By immunofluorescence dH1 can be detected in part of embryos earliest during nuclear division cycle 7 (Ner & Travers, 1994). During nuclear division cycles 8 and 9, dH1 staining becomes more abundant and stronger. At cycle 10 starts zygotic transcription so that after nuclear division cycle 14, there is great increase in dH1 amount. Estimates are that *Drosophila melanogaster* embryo at stage 8 has around 0.5-1 ng of dH1 protein ($>10^{10}$ molecules) (Ner & Travers, 1994). Amount of histone H1 is highly variable from one to another cell type (Fan et al., 2003).

First information about genomic localisation of H1 came from immunolocalization experiments in polytene chromosomes in *Drosophila melanogaster* (Jamrich et al., 1977). Staining of the chromosomes showed that histone H1 is present in gene poor, bands regions where RNA polymerase is absent.

Much more precise genomic position of linker histones in *Drosophila melanogaster* was determined much later, by usage of DamID (Braunschweig et al., 2009) and ChiP techniques. Project of Encyclopedia of DNA Elements (modENCODE) (Consortium et al., 2011) employed ChiP analysis to determine locations of many chromatin proteins, including dH1. Project aimed to determine binding sites of the proteins across a developmental time course and in multiple cell lines. The results are showing that dH1 is present along the great portion of the genome and that it is one of the main constituents in repressive chromatin. Scientific audience had to wait for a release of data about dH1 distribution, because of the doubts in quality of the results since dH1 binds at the majority of genomic locations and is not enriched at many regions. Rather, at least in the data obtained with DamID technique, regions where dH1 is depleted can be defined more easily (so called H1 dips).

H1 deposition to chromatin

Proper deposition and removal of linker histone is essential for providing its function in chromatin organization.

In vitro studies identified nucleophosmin 1 (NPM1) as a chaperon that interacts with H1 and deposits it onto dinucleosomal templates (Gadad et al., 2011).

ISWI remodelling factor has also been determined as a promoter of dH1 association with chromatin (Corona et al., 2007), but the exact mechanism of dH1 association is still unknown. It is speculated that ISWI-dependent assembly of dH1 on chromatin is happening independently from replication. This idea comes from the observation that ISWI mutants do not show change in NRL which is one of the phenotypes in H1 mutants. In H1 mutants since there is no sufficient H1 to be incorporated during DNA replication there is a change in NRL. Since ISWI mutant is not showing this phenotype,

most probably it is affecting assembly of H1 once the chromatin structure has already been established, after DNA replication.

Histone H1 functions

A great abundance of histone H1 molecules in nucleus suggests an important role of these molecules in a cell. Linker histones doubtless have structural role in packing DNA molecule into nucleus as described previously in the text. As mentioned, linker histones regulate spacing between nucleosomes and contribute to formation of 30 nm fiber and higher order chromatin structure. In addition to this, histone H1 plays many important roles in cells. Some of the roles are discovered for the moment only in specific specie, but even so, this information is contributing to the knowledge about all linker histones. The roles of H1 in many cases were determined by following phenotypes in various species with H1 mutant condition.

Saccharomyces cerevisiae mutant for H1 homologue, Hho1p, do not show any significant phenotype. The mutants are viable, without change in growth rate. At the level of chromatin neither could be observed significant changes – there are no changes in nucleosome repeat length neither in MNase accessibility of Hho1p depleted chromatin (Patterton et al, 1998). Compared to numerous phenotypes observed in higher eukaryotes (specified in the further text) the absence of any obvious phenotype is surprising fact in yeast mutants. However, Hho1p is only a putative H1 in yeast since it contains very different features, like structure, to H1 in higher eukaryotes. Instead of N- and C- terminus and globular domain, yeast H1 contains two globular domains separated by Lysine-, Alanine- and Proline-rich domain. Some more recent research identified a role for Hho1p in inhibition of recombination at rRNA locus (Li et al, 2008) and determined that H1 is playing this role independently from Sir2, another protein involved in suppression of rRNA recombination in yeast.

Tetrahymena thermophila also contains an atypical linker histone, without a globular domain and with only a sequence homologous to C-terminus of more canonical H1 histone proteins. The phenotypes in mutants of this protein are limited – there is no

effect on viability neither in growth-rate. However, an increase in nucleus volume has been reported as well as a change in transcription level of specific genes (Shen & Gorovsky, 1996).

In *Caenorhabditis elegans* H1.1 isoform is shown to be a specific regulator of gene silencing in germ-line cells (Jedrusik & Schulze, 2001). Depletion of only this specific variant results in reduced fertility of the mutants, formation of less germ nuclei and production of severely disordered gonads with lower number of differentiated oocytes. The same research group (Jedrusik & Schulze, 2007) in more recent report determined an increase in H3K4me and decrease in H3K9 level in the mutants, which could point on a role of H1 in establishing repressive chromatin.

There are many papers addressing the question of H1 role in regulation of transcription of rRNA genes in *Xenopus laevis* (described in section H1 isoforms). Apart from it, roles of embryonic histone H1 (B4) have been studied (Maresca et al, 2005). It has been determined that the loss of B4 provokes defects in mitotic chromosomes. They have elongated arms, they seem to be buckled and twisted and finally they cannot be properly aligned in the metaphase plate. Consequently, problems of chromosome segregation in anaphase occur. On the interphase chromatin on the other hand, no phenotype could be observed. These results are suggesting an essential role of B4 in metaphase chromosomes alignment and anaphase segregation.

Even an incomplete knock down of only 20% of total H1 level in *Drosophila melanogaster* causes late larval-pupal stage lethality (Lu et al., 2009). In the assay of position effect variegation (PEV) (Muller, 1930) dH1 acts as a suppressor since in dH1 mutant the active gene placed into heterochromatin location is not silenced and this is suggesting that dH1 is contributing to heterochromatin functioning. There are more proves that dH1 plays an important role in heterochromatin organization and functioning. For instance, a gene expression analysis of three genes (*concertina, light and rolled*) that are active only when present in heterochromatin showed that these genes are repressed in dH1 mutant. Chromocenter (aggregation of heterochromatin visible in polytene chromosomes) is split in parts in dH1 mutant. Finally, heterochromatin marks H3K9me2 and H4K20me2 and heterochromatin protein 1 (HP1) are increased in dH1 mutant background, although H3K9me2 is not incorporated in

chromatin. All these results indicate the involvement of dH1 in heterochromatin structuring (by contributing to unique chromocenter establishment) and functioning as a regulator of expression of heterochromatic genes or genes placed in heterochromatin by genetic manipulation (as proven to be a transcriptional activator of genes active inside of heterochromatin and as being suppressor of PEV, respectively). Apart from the effect on heterochromatin, histone dH1 is essential for the normal polytene chromosomes structure since in dH1 mutants these chromosomes are fragile, have problems of endoreplicated sister chromatids alignment and DNA replication.

Higher eukaryotes are characterised by the existence of genes for numerous H1 subtypes in their genome. These subtypes are redundant and as a consequence, an elimination of a single (or even more) variants is compensated by the increase in expression of other H1 subtypes. The real decrease in total H1 amount can be achieved only by elimination of more variants.

In the case of chicken DT40 B lymphocyte cell line, elimination of ten out of twelve in total alleles coding for six H1 subtypes, did not change total level of H1 in the cells (Hashimoto et al., 2010). Only elimination of eleventh allele decreased the total H1 mRNA level to 50%. These mutants show change in protein patterns in 2D-PAGE gels and the increase in HMG proteins, but they show no changes at the level of global chromatin structure, neither significant change in growth rate. Elimination of the last allele results in increased chromosomal aberration rate, increased nuclear volume, decrease in NRL and affected gene expression globally, mainly reduction in their expression. Growth rate is also affected, as proven not by stop in any part of the cell cycle, but most probably as a consequence of longer cell cycle. Another study in the same cell type determined a role of H1R isoform in DNA damage response (Hashimoto et al., 2007). Cells lacking this specific isoform are showing increased sensitivity to alkylating agent MMS and increased sensitivity to infra red radiation. The authors suggest involvement of H1R in Rad54-mediated homologous recombination.

In the case of mice (Fan et al., 2003), elimination of two out of five somatic H1 subtypes by homologous recombination did not change total H1 level, as mentioned, due to compensatory effects of the other subtypes. Only the elimination of three subtypes decreases the total H1 amount to 50%. The depletion is lethal for the mice as

they die by midgestation. At the chromatin level, occurs a reduction in NRL. Latter studies concerned with gene expression in these mice showed that only a small portion of the genes is changing expression and interestingly, among them are many genes regulated by DNA methylation (Fan et al., 2005). This is suggesting that H1 might be contributing to the formation or maintenance of DNA methylation patterns in mice.

In human breast cancer cell line, T47D, H1.2 depletion is provoking problems in cell growth, cell cycle arrest in G1 phase, decrease in NRL and repression of limited number of genes, many of which are involved in regulation of cell-cycle progression (which can explain G1 arrest). Loss of only H1.4 subtype is provoking cell lethality.

In HeLa cells linker histone was identified as a factor important in DNA repair. It was determined that H1 enhances non-homologous end joining (NHEJ) by DNA Ligase III and DNA Ligase IV (Rosidi et al., 2008).

H1 as a regulator of transcription

Gene expression analysis was done in some of the H1 mutants mentioned in the previous pages to check for the role of linker histones in regulation of transcription. Expression of genes was not drastically affected and H1 most probably has only a limited effect on gene expression. It is interesting that initial experiments that were examining this question got quite different conclusion.

First experiments addressing the question of linker histone role in chromatin were done using *in vitro* transcriptional system. At the beginning the system consisted of naked DNA used as a template for RNA polymerase activity. With time, naked DNA was changed for chromatin template that besides DNA contains core histones and can contain linker histones. Transcriptional models that were used very often were nucleosomal arrays of mouse mammary tumor virus (MMTV) and 5S rDNA chromatin reconstituted from *Xenopus* cells.

MMTV in the promoter region contains glucocorticoid hormone response element. Binding of the hormone activates transcription from the promoter and it has been

demonstrated that dissociation of histone H1 is necessary for the transcriptional activation to occur.

Expression of 5S rDNA is also regulated by linker histone. Histone H1 is preventing transcription from oocyte-type 5S rRNA genes by positioning nucleosomes at the site where key positive regulatory elements for transcription are found. In this way binding of transcription factors is prevented and there is no transcription of the genes (Pennings et al, 1994).

In these first experiments H1 has been identified as a general negative regulator of transcription and this epithet has been attributed to H1 for a long time. However, it has been demonstrated that H1 in some cases can act as a positive regulator of transcription since their overexpression leads to transcriptional induction of some genes (Brown et al, 1996). In addition, *in vivo* studies are giving less importance to histone H1 as a transcriptional regulator. Namely, in knockout models for histone H1 in different species the loss of H1 has only a minor effect on gene transcription.

In mice for example 50% decrease in total H1 content affected only 0.56% of examined genes with two-fold change in expression level (Fan et al., 2005). In chicken DT40 cells complete loss of H1 results in more than two-fold change in expression of only 4.2% of analyzed genes (Hashimoto et al., 2010).

In addition, ways by which H1 regulates transcription have been explained at the molecular level. Linker histones are regulating transcription by different means: by restricting nucleosome mobility, by regulating accessibility of different transcriptional factors to regulatory sequences, by forming a part of transcriptional complexes, by interacting with transcription factors and chromatin remodelling complexes, by regulating acetylation of core histones, in association with ribosomal proteins etc. I will make a brief description by making examples of these types of H1 transcriptional regulation.

H1 is repressing expression of oocyte 5S rRNA gene by restricting mobility of nucleosomes and positioning them to key regulatory elements of the gene (C-box region) so that binding of TFIIIA transcription factor which is necessary for transcription is not allowed. In this case, H1 acts by affecting nucleosome mobility. There are other evidences that mobile nucleosomes might facilitate transcription factor

access to DNA (Ura et al, 1995). In a sea urchin H1 is also limiting TFs access to rDNA (Pennings et al, 1994). Since H1 can easily associate and disassociate from chromatin and in this way prohibit or allow gene expression, H1 contributes to chromatin dynamics. The dynamics then depends on chromatin affinity for binding of H1 or possibility for redistribution of H1 on chromatin. In the context of H1 as a protein that restricts mobility of nucleosomes it's interesting to mention that *in vitro* has been demonstrated that H1 modulates remodelling of nucleosomes by SWI/SNF. On DNA molecule, SWI/SNF would slide nucleosomes to the end of the template while the presence of H1 changes the sliding to the central part of the template (Ramachandran et al, 2003).

Another example where linker histone regulates binding of TFs is in nucleosomes from HeLa cells where H1 prevents binding of USF (Juan et al, 1994). The same authors do not observe repression in the same extend for Gal4-AH TF binding and propose existence of differential repression rate by which H1 prevents binding of different TFs. It has been show that H1 in this case is actually stabilizing interaction of core histones with DNA and leaving less exposure time of DNA for binding of TF (Juan et al., 1997). Acetylation of core histones negatively affects H1 binding and consequently abolishes H1 repression effect.

Next, H1 can act on transcription by interaction with chromatin remodelling complexes. For instance, H1 is interacting with chromodomain-helicase- DNA-binding 8 (CHD 8), an ATP dependent chromatin remodelling factor. In this way occurs repression of transcription of targets of two TFs, p53 and β–catenin (Nishiyama et al., 2009). In these experiments, done on mice by the same group for both p53 and β–catenin, the authors have proven that the presence of H1 is necessary in trimeric complexes (CHD8, H1 and p53 or CHD8, H1 and β-catenin) for transcriptional repression of specific target genes since either triple knock-out of three H1 isoforms or expression of dominant negative mutant of H1 abolishes the transcriptional inhibition.

It was reported that H1 interacts with FoxP3, a member of P subfamily of the forehead (FKH) box TFs (Mackey-Cushman et al., 2011). This TF represses transcription of interleukin-2 (IL2) gene by recruiting histone deacetylase-containing complex. In addition, FoxP3 inhibits transcription by interacting with H1 in humans, with H1.5

isoform, when it modulates affinity of H1.5 binding to IL2 gene promoter. Depletion of H1.5 abolishes IL2 transcriptional inhibition, demonstrating a necessity of linker histone for the inhibition. However, FoxP3 – H1 interaction is not always negatively contributing to transcription of the target genes. For example, in the case of CTLA 4 gene promoter, the interaction of FoxP3 with H1 enhances its expression when H1 binding is weaker.

Joint binding of H1 and ribosomal protein L22 to chromatin is regulating expression of specific genes (Ni et al, 2006). These two proteins have been shown to interact and depletion or overexpression of one of them results in changes in expression in the same direction for common set of genes, suggesting interplay of H1 with ribosomal proteins in regulation of transcription.

Linker histone has been implicated in regulation of transcription by interplay with core histone acetylation. Data suggest that linker histones affect core histones acetylation and *vice versa*. Acetylation level of core histones is tightly regulated in a cell by two main groups of enzymes: histone acetyltransferases (HAT) and histone deacetylases (HDAC) that introduce and erase the mark, respectively. This strong regulation of core histone acetylation is suggesting that the modification plays an important role in chromatin functioning. Indeed, there are many evidences that correlate H3 lysine acetylation with transcriptional activation (Parekh & Maniatis, 1999). By affecting core histone acetylation, linker histones are affecting gene expression. The interplay between H1 and core histone acetylation is mutual.

On one hand, *in vitro* analysis have shown that acetylation of core histones is altering the capacity of H1 to form higher order structures (Ridsdale et al, 1990) and that it decreases H1 repression of transcription factor binding on reconstituted nucleosomes (Juan et al, 1994). However, there are contrary *in vitro* results arguing that repressive effect of H1 on transcription and reduction effect of H1 on nucleosome mobility depends on core histones acetylation (Ura et al, 1997).

On the other hand, it has been demonstrated *in vitro* that H1 influences level of core histone acetylation (Gunjan et al, 2001). In this work, overexpression of H1 as a consequence has a decreased acetylation of core histones. It was demonstrated that the activity of HATs and HDAC is not changed in these extracts leading to a conclusion

that the effect is due to the changes in chromatin structure. In another study (Herrera et al, 2000) where H1 is affecting acetylation of core histones *in vitro*, the effect in this example is accomplished by inhibition of HAT p300/CBP-associated factor (PCAF), but again due to changes in chromatin, by sterically preventing approach to H3 with its protein tail. When present in a complex, p300/CBP-associated factor (PCAF), can overcome the inhibition, which is possibly because of remodelling complex that acts on H1 and change its interaction with chromatin.

Posttranslational modifications

Proteins can change their properties and functions by introduction of simple chemical groups on their amino acids. Some of the groups for instance contain a lot of electrical charge that affects structure of a protein and consequently, its function. Good example for this is greatly negatively charged phospho-group that in numerous enzymes is a regulatory modification for their activity. For a cell these, so called, posttranslational modifications (PTMs) are very practical since they can change the functionality of the proteins without a need to invert energy and time to synthesise a completely new protein.

In histones, groups that are introduced on amino acids of N-terminal tails and effect of these PTMs on histones' properties are studied the most. Majority of them are reversible due to the existence of, on one hand, enzymes that introduce the modifications (methyltransferases, acetyltransferases, kinases, ubiquitin ligases etc) and, on the other hand, enzymes that erase them (demethylases, histone deacetylases, phosphatases, deiminases etc) (Suganuma and Workman, 2011). Introduction of different PTMs makes histone molecules methylated, acetylated, phosphorylated, ubiquitinated, ADP-ribosylated or sumoylated. Presence of a certain PTM for a histone can mean a change in binding affinity for DNA and different interactions with other constituents of chromatin. Many chromatin proteins contain domains that recognize specific histone modification. Chromodomain is recognizing methylated histones; some PHD domain containing proteins bind H3K4me3; tudor and MBT domain containing proteins in protein- domain microarray technique bind methylated peptides in H3 and H4 (H3K4, H3K9 and K4K20, predominantly its dimethylated states) (Kim et al., 2006); WD40

domain of WDR5 recognizes H3R2 and H3K4me2 etc. Determination and studying PTMs of histones is then a very interesting task since these marks are regulating some of the most basal functions of chromatin and most crucial functions of a cell. As a good example of PTMs studies importance stand new medical approaches in which enzymes that regulate histone PTMs are targeted. By imitating their role in a cell gene expression can be regulated and this can have many consequences.

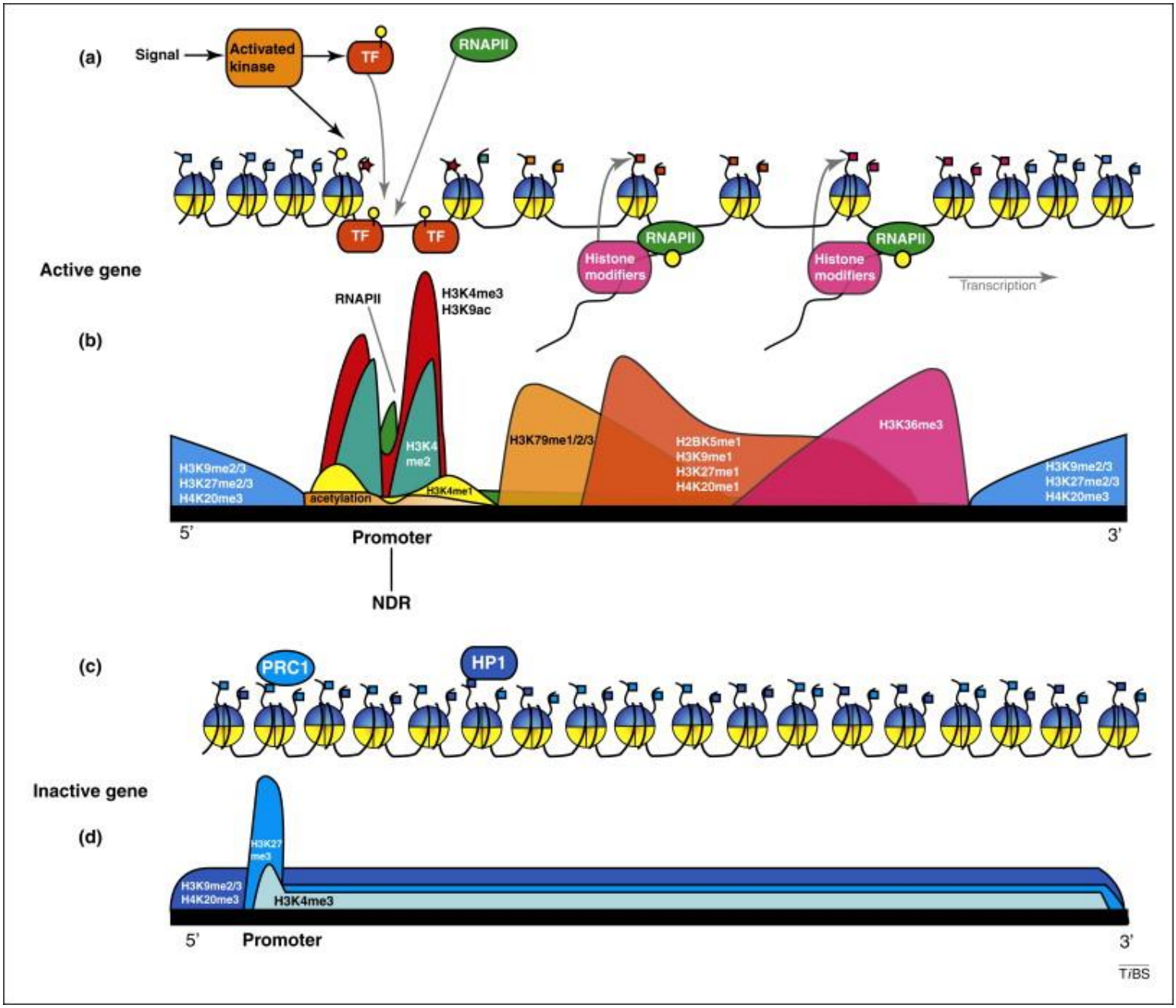

Figure 6: Posttranslational modifications along active and silent genes

From recently, there are genome-wide data about chromatin proteins distribution and genome sequences in different species. Their availability is unrevealing that there is regularity in the distribution of histone with particular modifications in different chromatin types and even along single genes. It is known that euchromatin is decorated with H3K4me, H3K36me, that heterochromatin contains H3K9me, H4K20me etc. In addition, active and silenced genes have a special pattern of histone modifications along their sequence (Figure 6). Sequences upstream from active gene promoters have H3K9me2/3, H3K27me2/3 and H4K20me3. Promoters of these genes have nucleosomes with H3K4me2, H3K4me3, H3K9ac and in less extent H3K4me1 and general histone acetylation. Downstream from TSS, active genes contain H3K79me1/2/3 followed by H2BK5me1, H3Kme1, H3K27me1 and H4K20me1 and further, closer to the gene end are H3K36me3 and completely distally are H3K9me2/3, H3K27me2/3 and H3K20me3. On the other hand, silenced genes have H3K9me2/3 and H4K20me3 upstream their promoters, H3K27me3 and H3K4me 3 around TSS and in less extent along the whole gene as well.

Mass spectrometry techniques have contributed much in detection of histone posttranslational modifications. Histone modification maps for chicken, mouse and human (Wood et al., 2009) and other species are becoming more complete and functions for some of these new observed modifications are being established.

Posttranslational modifications of dH1

dH1 modification map is very poorly characterized. It is known that Ser10 is phosphorylated (Villar-Garea et al., 2007) and that the protein is ubiquitinated. Phosphorilation level changes with cell cycle progression and it is known that S-phase or G2-like state cells have this modification mark. (Talmage et al, 1987; Villar-Garea et al., 2007). It is established that in *Drosophila* folicle cells dH1 is highly phosphorylated during gene amplification (Hartl et al, 2007). The enzyme that regulates H1

ubiquitination is TAFII250 (Pham et al., 2000), but the exact position of the modification is not determined yet.

It is interesting that there are several potential targets for methylation in dH1 (e.g. Lysines at positions 22, 23, 27 and 28 in N-terminus) and that in human H1 residues with methyl groups have been determined (Garcia et al, 2004). It is known that H1K26me is included in heterochromatin formation and HP1 protein in vitro recognizes this modification (Daujat et al., 2005) Methyltransferase and demethylase that regulate the mark have been determined (Trojer et al., 2009). Even so, methylation in dH1 has not been reported yet. Mass-spectrometry has limitations in detection of small methyl groups and particularly in the case of *Drosophila melanogaster* there is a case where an amino acid polymorphism produces spectrometric peaks that can be ambiguous since the same peak can be corresponding to an amino acid which is methylated or, on the other hand, the same peak might correspond to another, unmodified amino acid (Villar et al, 2007). Therefore analysis of peaks that contain methyl-groups can be equivocal and lead to wrong interpretation. Additional approaches besides classical mass-spectrometry are needed for confident determination of methylated residues.

Rising antibodies highly specific for modified histones is another approach that can be taken. An advantage of this approach is that it allows in vivo studies. A good-quality antibody, specific for modified form of the protein is the crucial in this approach.

Drosophila melanogaster as a model system

Drosophila melanogaster is widely used model system in biology. Apart from having a short life cycle and being easy to cultivate, there is a great pool of genetic tools that are developed for this specie, due to a very long history of laboratory work on this fly. There are only four chromosomal pairs in this specie and for each of them there are chromosomal markers that can be followed phenotypically and in this way inheritance of the chromosomes can be easily followed from generation to generation. There is a great collection of mutants available for this fly. Recently, *D. melanogaster* genome has been sequenced and genomic maps for numerous chromatin proteins have been determined as well. There are three cell lines of the specie. Gal4 system has been described (Brand & Perrimon, 1993) and widely used to induce RNAi response in *Drosophila melanogaster*. In this system, Gal4 gene, which is a yeast gene, not present in wild type fruit fly, is introduced into *D. melanogaster* genome (or in another genome) and expressed by control of different promoters. Gal4 protein has an affinity for binding upper activating sequence (UAS) and in this way activates expression of the downstream gene (Figure 7).

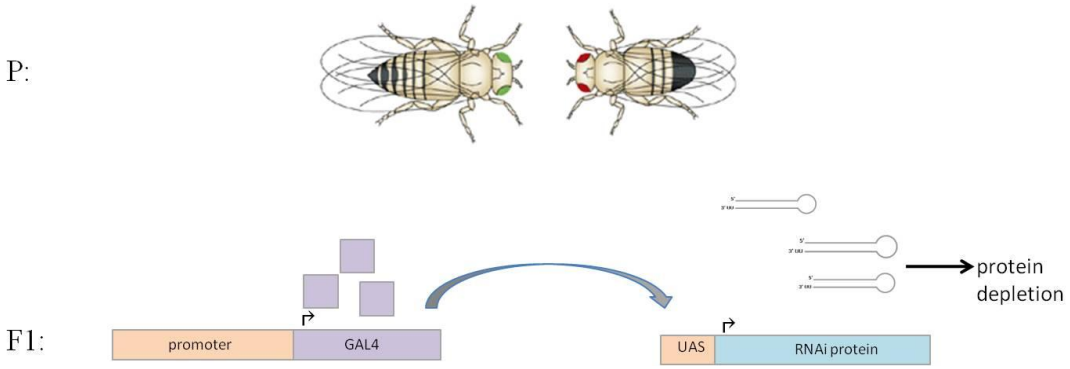

Figure 7: Gal4 system induction of RNAi machinery pathway; P-parental fly stock, F1-offspring

It means that the expression of the hairpin can be controlled by usage of a specific promoter controlling expression of Gal4. Some promoters are active in all cell types (like actin or tubulin promoter), some others only in particular cell types or tissue compartments. Some promoters are expressed early in embryogenesis, some others later or only during specific developmental stages. In other words, by making careful choice one can induce expression of Gal4 and the hairpin in specific part of fly life stage and part of the body. The hairpin is cut by Dicer in short pieces that will activate other elements in RNAi machinery pathway. This will result in degradation of mRNA corresponding to the hairpin sequence and consequently, decrease in the protein level. To accomplish stronger mRNA depletion, one can overexpress Dicer2 to stimulate RNAi machinery pathway.

Wing imaginal discs are *Drosophila* larval organs that after metamorphosis give rise to adult wings. They are very often used in studies. They can be easily isolated from larvae and there are optimized protocols for different techniques applied on them. This organ contains only a membrane and two layers of cells, which is very useful for immunostaining experiments because it is not necessary to make thinner cuts. Many signalling pathways are studied in discs and lot is known about them and about phenotypes that are produced as a consequence in of impairment in some of these pathways. In this way, by following phenotypes in wing's shape, vein morphology, state of sensory organs on the border of wings etc, one can get information about signalling pathways that are affected in the mutant that is under observation.

Objectives

The principal objective of this project was to analyze the contribution of *Drosophila melanogaster* histone H1, dH1, to structural and functional properties of chromatin.

We aimed to make a mutant state for dH1 protein by using RNAi machinery and to look for the phenotypes that would help us learn more about the protein functions.

We also aimed investigating posttranslational modifications in N-terminus of dH1 and how are they contributing to chromatin properties. In particular, we wanted to learn more about methylation modification marks since di-methylation was already reported in human H1 N-terminus region, which is conserved from *Drosophila melanogaster* to humans.

Papers and additional results

Article 1

In the first paper we are describing a strategy we used to make a mutant state for dH1. We are presenting phenotypes that occur in the mutants and we are looking for changes in gene expression.

We determined that only a small portion of genes are changing their expression and that transposable elements (TE) get activated. Among upregulated TE are R1 and R2 elements that are inserted in rRNA locus of a cell. We are showing that this is further changing expression of the whole rRNA locus and that it leads to production of extra chromosomal circular DNA. We also detected DNA damage, problems in cell proliferation and ultimately cell death.

In this way, here we are showing that linker histone plays a minor role in gene expression regulation, but that it is necessary for silencing of transposable elements and preserving genome stability.

This paper contains the majority of the experimental work of my PhD. Rescue experiment with human isoforms did postdoc in our laboratory, Katrin Zaragoza; Alejandro Vaquero started the project concerning H1 and supervised me during first year of my work; statistical analysis of the data did Oscar Reina from biostatistics/bioinformatics facility at IRB Barcelona; Jordi Bernués did some of the genetic crosses of the flies and contributed in proposing experiments, knowing well *Drosophila melanogaster* as a model system. Ferran Azorín as a principal investigator made a project and guided it. I did the rest of the experiments.

Bibliographic citation of the article 1

Vujatovic O, Zaragoza K, Vaquero A, Reina O, Bernués J, Azorín F. *Drosophila melanogaster* linker histone dH1 is required for transposon silencing and to preserve genome integrity. Nucleic Acids Res. 2012 Jul;40(12):5402-14.

Vujatovic O, Zaragoza K, Vaquero A, Reina O, Bernués J, Azorín F. Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity.Supplementary data. Nucleic Acids Res. 2012 Jul;40(12):5402-14.

Article 2

The second paper in this thesis is dealing with linker histone PTMs in SL2 cells of *Drosophila melanogaster*. Bottom up and top down mass-spectometry analysis was performed to detect PTMs. My contribution to the second paper is only partial. I did a part of the experiments concerning biological significance of 2mK27dH1 modification mark. Here we are presenting that the mark is enriched in centromeric regions in metaphase chromosomes of SL2 cells and the rest of the experiments on this subject I will present apart, after the second paper.

JOURNAL OF PROTEOMICS XX (2012) XXX–XXX

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of *Drosophila melanogaster* linker histone H1

Carles Bonet-Costa[a,b,1], Marta Vilaseca[c,1], Claudio Diema[c], Olivera Vujatovic[a,b], Alejandro Vaquero[a,b,2], Núria Omeñaca[c], Lucía Castejón[a,b], Jordi Bernués[a,b], Ernest Giralt[b,d,*], Fernando Azorín[a,b,*]

[a]Institute of Molecular Biology of Barcelona, CSIC, 08028 Barcelona, Spain
[b]Institute for Research in Biomedicine, IRB Barcelona, 08028 Barcelona, Spain
[c]Mass Spectrometry Core Facility, Institute for Research in Biomedicine, IRB Barcelona, 08028 Barcelona, Spain
[d]Department of Organic Chemistry, University of Barcelona, 08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received 30 April 2012
Accepted 20 May 2012

Keywords:
Chromatin
Histone H1
Post-translational modifications
Mass spectroscopy
Top-down MS
Drosophila

ABSTRACT

Linker histone H1 is a major chromatin component that binds internucleosomal DNA and mediates the folding of nucleosomes into a higher-order structure, namely the 30-nm chromatin fiber. Multiple post-translational modifications (PTMs) of core histones H2A, H2B, H3 and H4 have been identified and their important contribution to the regulation of chromatin structure and function is firmly established. In contrast, little is known about histone H1 modifications and their function. Here we address this question in *Drosophila melanogaster*, which, in contrast to most eukaryotic species, contains a single histone H1 variant, dH1. For this purpose, we combined bottom-up and top-down mass-spectrometry strategies. Our results indicated that dH1 is extensively modified by phosphorylation, methylation, acetylation and ubiquitination, with most PTMs falling in the N-terminal domain. Interestingly, several dH1 N-terminal modifications have also been reported in specific human and/or mouse H1 variants, suggesting that they have conserved functions. In this regard, we also provide evidence for the contribution of one of such conserved PTMs, dimethylation of K27, to heterochromatin organization during mitosis. Furthermore, our results also identified multiple dH1 isoforms carrying several phosphorylations and/or methylations, illustrating the high structural heterogeneity of dH1. In particular, we identified several non-CDK sites at the N-terminal domain that appear to be hierarchically phosphorylated. This study provides the most comprehensive PTM characterization of any histone H1 variant to date.

* *Corresponding authors at:* Institute of Molecular Biology of Barcelona, CSIC, Institute for Research in Biomedicine, IRB Barcelona, Barcelona Science Park, Baldiri Rexac, 10. 08028 Barcelona, Spain. Tel.: +34 93 4034958; fax: +34 93 4034979.
E-mail address: fambmc@ibmb.csic.es (F. Azorín).
[1] These authors contributed equally to this work.
[2] Present address: Chromatin Biology Laboratory, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Spain.

1874-3919/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2012.05.034

1. Introduction

Eukaryotic chromatin is primarily organized as nucleosomes, a nucleoprotein complex composed by 146 bp of DNA wrapped around a histone octamer formed by two copies of each core histone H2A, H2B, H3 and H4. Chromatin, however, exists mainly as a 30-nm fiber that results from the folding of nucleosomes into a higher-order structure (reviewed in [1–5]). The formation and stability of this higher-order chromatin structure depends on a fifth protein, histone H1, which binds the linker DNA connecting adjacent nucleosomes. In contrast to core histones that are highly conserved through evolution, histone H1 shows much higher variability [6], with most species containing several variants that appear to play both specific and redundant functions [7,8].

Core histones are extensively modified at multiple residues. Known post-translational modifications (PTMs) of core histones include: K-acetylation, K-methylation (mono-, di- and tri-), R-methylation (mono-, di-symmetric and di-asymmetric), S/T-phosphorylation, K-ubiquitination, K-sumoylation, E-ADP ribosylation, R-deimination and P-isomerization (reviewed in [9,10]). In contrast, information about the PTM pattern of histone H1 is limited, as it has been analyzed only in few species and variants. In addition, although recent studies have identified multiple PTMs in human, mouse and chicken H1 variants [11,12], most other published work focuses mainly on the mapping of phosphorylation sites [13–17]. Like core histone modifications, which contribute to the regulation of multiple genomic processes (reviewed in [9,10]), modifications on histone H1 are likely to play important regulatory functions. Little is known, however, about their functional significance. In this regard, given its remarkable lack of evolutionary conservation, describing histone H1 PTMs in different species appears essential.

Here we analyzed the PTM pattern of histone H1 in the fruit-fly *Drosophila melanogaster* that, in contrast to most eukaryotic species, contains a single histone H1 variant, dH1, which is encoded by the multicopy *His1* gene [18]. For this purpose, we used a combined bottom-up and top-down mass-spectrometry (MS) approach. Bottom-up and top-down strategies for MS-based protein characterization are complementary. In bottom-up strategies, proteomic measurements at the peptide level offer a basis for unambiguous PTM identification. On the other hand, top-down proteomics, which does not require the proteolysis of proteins to generate peptides [19–22], can provide an integrated view of PTM occupancy at the intact protein level. Top-down MS has been used to characterize several purified proteins, including some core histones [23–29]. However, this approach is currently envisaged as a complement of the classical bottom-up strategy since, despite recent progress in protein separation methods [30,31], the analysis of complex mixtures of proteins by top-down MS remains a challenge. In fact, histone H1 PTMs have been analyzed mainly by bottom-up MS [12,13,15–17] and, only in *Tetrahymena thermophila*, histone H1 has been partially analyzed by top-down MS [14].

Here, we report on the identification of multiple PTMs in dH1, including: seven phosphorylated S/T; three mono- and eight dimethylated K; three acetylated K, and four ubiquitinated K. Interestingly, most of these modifications are located at the N-terminal domain, some being conserved in specific human and/or mouse variants, suggesting that they play conserved functions. Our results also showed that dimethylation of K27, a main dH1 PTM that is conserved in vertebrates, accumulates at pericentromeric heterochromatin in metaphase, suggesting a functional contribution to heterochromatin organization and function during mitosis. In addition, we also present evidence of the co-existence of different PTMs in the same dH1 molecule. This study, which illustrates the usefulness of combining bottom-up and top-down analytical approaches, highlights the structural heterogeneity of individual dH1 molecules. It must also be noted that, previous to this study, only a single major phosphorylation site was mapped in dH1 [17].

2. Materials and methods

2.1. Obtaining dH1

dH1 was extracted from purified nuclei from *D. melanogaster* S2 cultured cells by treatment with 10% perchloric acid (PCA) for 1 h [32] (Fig. 1A). After extraction, dH1 was precipitated by addition of 50% trichloroacetic acid (TCA) to a final concentration of 20%, washed with acetone, and air-dried under vacuum.

2.2. Chemical oxidation of dH1

40 μl (8 μg) of dH1 was dissolved to a final volume of 150 μl in a freshly prepared mixture of 3% aqueous H_2O_2 and 3% formic acid (FA) following previously described conditions for mild performic acid treatment (MPA) [33].

2.3. Reverse phase high performance liquid chromatography (RP-HPLC) purification of dH1

For RP-HPLC purification, 8–10 μg of dH1, with and without prior oxidation, was subjected to chromatography on a BioSuite pPhenyl 1000 column (Waters) (10 μm RPC, 4.6×75 mm), using a linear gradient of 5% to 80% B in 60 min (A=0.1% FA in H_2O, B=0.1% FA in CH_3CN) at a flow rate of 1 ml/min (Fig. 1B). An Acquity UPLC chromatographic system (Waters) was used. The LC eluent was coupled after a 1/10 split post-column to a LCT-Premier XE mass spectrometer (Waters-Micromess) provided with an ESI source. A BioRad model 2110 fraction collector was used to collect the fractions of interest. Fractions were lyophilized and stored at −20 °C for further analysis.

2.4. CNBr treatment of dH1 and RP-HPLC purification of dNtH1 peptide

Non-oxidized RP-HPLC-purified dH1 (8 μg) was resuspended in 95 μl 0.1% aqueous FA. The solution was heated at 80 °C for 1 h and 10 μl of 5 M CNBr (Sigma-Aldrich) in acetonitrile was added. Digestion was allowed to proceed overnight at room temperature. dNtH1 peptide was purified using a BioSuite pPhenyl 1000 column (Waters) (10 μm RPC, 2.0×75 mm) and a linear gradient of 5% to 80% B in 60 min (A=0.1% FA in H_2O, B=0.1% FA in CH_3CN) at a flow rate of 100 μl/min. A Finnigan Micro AS autosampler and a Surveyor Finnigan MS quaternary

Please cite this article as: Bonet-Costa C, et al, Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of..., J Prot (2012), doi:10.1016/j.jprot.2012.05.034

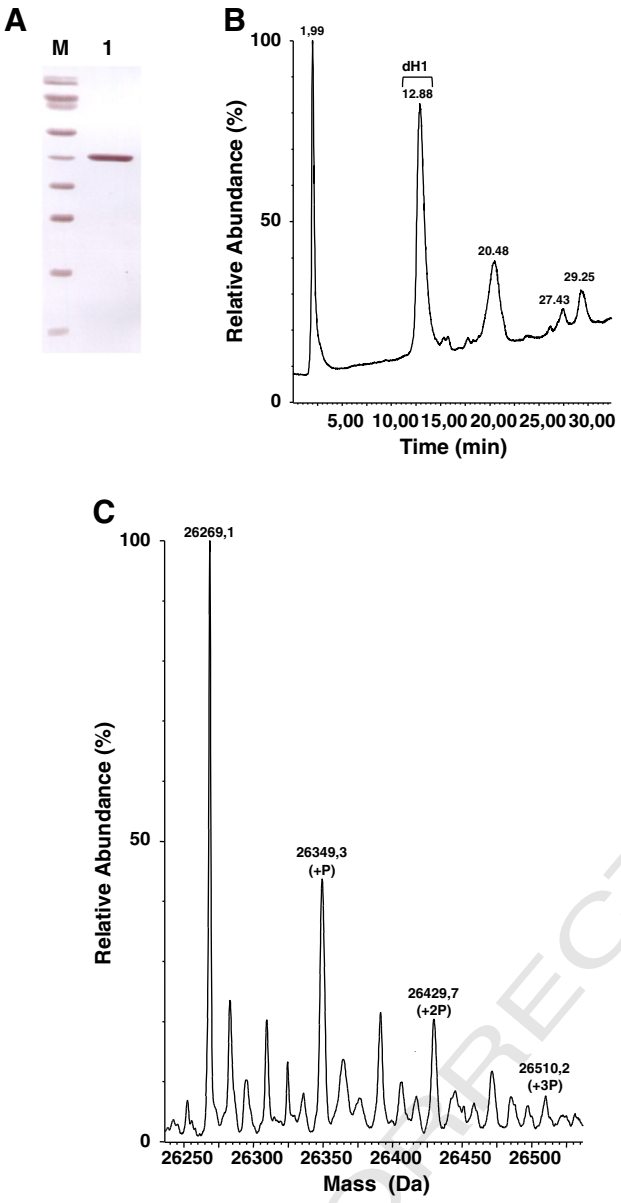

Fig. 1 – Purification of dH1. (A) 15% SDS-PAGE analysis of dH1 obtained from purified S2 nuclei by extraction with 10% perchloric acid (lane 1). Lane M corresponds to molecular weight markers. (B) RP-HPLC-purification of dH1 on an Acquity UPLC system using a BioSuite pPhenyl 1000 column (Waters) attached to an LCT-Premier mass spectrometer (Waters) (see Materials and methods section for details). The peak corresponding to dH1, which elutes at t_R 12.88 min, is indicated. (C) Deconvoluted on-line LC–MS spectrum of purified dH1. Average mass (Da) of selected peaks corresponding to the unmodified N^{α}-terminally acetylated, and mono-, di- and triphosphorylated dH1 forms is indicated.

pump (Thermo Electron Corporation) were used. Online LC-nanoESI-MS coupling on a 7 T LTQ-FT Ultra mass spectrometer (Thermo Scientific) with simultaneous fraction collection was performed using a TriVersa NanoMate (Advion BioSciences). MS/MS analyses of dNtH1 fractions of interest were performed off-line, using the NanoMate in infusion mode, by applying CID fragmentation on selected ions (see Top-down and middle-down MS section for details).

2.5. *Bottom-up MS*

Dried RP-HPLC-purified dH1 pellets were reconstituted in 100 mM ammonium bicarbonate and digested with trypsin overnight at 37 °C. The resulting peptide mixtures were desalted with a Zip-Tip C_{18} (Millipore), vacuum dried and reconstituted in 0.1% FA prior to analysis.

The nano-LC–MS/MS set up was as follows. Samples were injected through a Finnigan Micro AS autosampler and loaded to a NanoEase trap column Symmetry 300™, C_{18}, 5 μm (Waters) at a flow rate of 15 μl/min using the Micro AS-Surveyor MS chromatographic system described above. Peptides were separated using a C_{18} PepMap 100, 3 μm capillary column (75 μm, 15 cm) (Dionex, LC Packings) with a 160 min run, comprising a 10 min isocratic elution at 0% B, three consecutive steps with linear gradients from 0% to 15% B in 10 min, from 15% to 60% B in 70 min, and from 60% to 100% B in 20 min, followed by isocratic elution at 100% B in 10 min (A=0.1% FA in H_2O, B=0.1% FA in CH_3CN). The 110 nl/min flow rate used for peptide separation was provided by an in-house splitter system. The column outlet was connected to the LC coupler of the TriVersa NanoMate, which was coupled to a 7 T LTQ-FT Ultra. The mass spectrometer was operated in a data-dependent (DD) mode using three acquisition methods: DDMS10, neutral loss (DDNLMS3) and multistage activation (MSA). For DDMS10, survey MS scans were acquired in with the resolution set to 100,000 ($m/\Delta m_{50\%}$ at 400 m/z). Up to ten of the most intense ions per scan were fragmented and detected in the linear ion trap. Ion transmission into the FTICR cell and the linear trap was automatically controlled for optimal performance of the analyzers by setting the charge capacity to 1 million counts for the survey full scan and to a 50,000 counts for the MS/MS experiments. Target ions already selected for MS/MS were dynamically excluded for 60 s. The DDNLMS3 experiment was completed for the three most intense precursor peptide ions and the top three neutral loss masses. A threshold of 50 counts was used for triggering the MS/MS and MS/MS/MS scan events. The MSA acquisition method was completed for the five most intense precursor ions with additional activation upon phosphate neutral loss detection [34], and a 50 count minimum signal threshold.

Database search was performed with Proteome Discoverer software v1.2 (Thermo) using Sequest and Mascot engines. The databases used were a subset (*D. melanogaster*) of SWISSPROT and NCBInr. Search parameters included trypsin enzyme specificity, allowing for three missed cleavage sites and several variable modifications including N^{α}-terminal acetylation, M-oxidation, S/T-phosphorylation, K-acetylation, K-ubiquitination, and K mono-, di- and trimethylation. Peptide mass tolerance was 10 ppm and the MS/MS tolerance was 0.8 Da. Increasing mass tolerance to 12.5 ppm resulted in the identification of few additional peptides corresponding to low abundant forms with poor ion statistics.

2.6. *Top-down and middle-down MS*

Data were acquired on a 12 T and a 7 T LTQ-FT Ultra, as indicated in each case. RP-HPLC-purified dH1 (oxidized and non-oxidized)

and collected dNtH1 fractions were diluted or reconstituted with ESI solution (MeOH, 1% FA (1:1, v/v)) and infused by automated nanoelectrospray using the TriVersa NanoMate as the interface. Full MS spectra (m/z 200–2000) were acquired either at a 100,000 or 200,000 resolution ($m/\Delta m_{50\%}$ at 400 m/z). After full scan analysis, individual charge states of the multiply protonated protein/peptide molecular ions were selected for isolation in the LTQ using isolation widths of 5–10–100 m/z. Isolated ions were then fragmented by CID in the trap or guided to the FTICR cell and further fragmented by ECD. For CID experiments the precursor ions were activated using 30% to 50% normalized collision energy at the default activation *q*-value of 0.25. All ECD experiments were performed in the 12 T LTQ-FT-Ultra instrument with the following settings: 3–5 energy (arbitrary units) corresponding to a cathode voltage of 1.5 V to 3.5 V, 127 ms delay (with 0 ms additional delay) and 15–30 ms duration. In both types of fragmentation the efficiency of fragmentation was optimized to maximize product ion signal intensity. Fragment detection was done in the FTICR cell for both types of fragmentation at 100,000 resolution ($m/\Delta m_{50\%}$ at 400 m/z) and averaging 200–1000 scans. The analyzer charge capacity was set to a target value of 500,000 counts for both CID and ECD MS/MS experiments. Protein masses and zero-charged fragment masses from CID or ECD were determined by deconvolution

Table 1 – List of identified tryptic peptides with posttranslational modifications (PTMs). Nta, N-terminal acetylation; pS, phosphoserine; pT, phosphothreonine; mK, monomethylated lysine; 2m, dimethylated lysine; aK, acetylated lysine; uK, ubiquitinated lysine; oM, oxidized methionine. See Supplementary data SI for MS/MS spectra of indicated tryptic peptides.

PTM	Peptide	Sequence1 [a]	Score	Ion	Mass M	Mass C	Δmass	ppm
	2–2	SDSAVATSASPVAAPPATVEK	74	978,4997	1954,9848	1954,9847	0.0001	0.1
Nta	2–22	**aS**DSAVATSASPVAAPPATVEK	105	999,5079	1997,0013	1996,9953	0.0060	3.0
Nta, pS1l	2–22	**aS**DSAVATSA**pS**PVAAPPATVEK	66	1039,4880	2076,9631	2076,9616	0.0015	0.7
Nta, pS9	2–22	**aS**DSAVAT**pS**ASPVAAPPATVEK	49	693,3283	2076,9631	2076,9616	0.0015	0.7
Nta, pTl9	2–22	aSDSAVATSASPVAAPPApTVEK	23	693,3277	20,769,704	2076,9616	0.0088	4.2
pS2	2–22	pSDSAVATSASPVAAPPATVEK	18	1018,4837	2034,9530	2034,9511	0.0019	0.9
pS1l	2–22	SDSAVATSApSPVAAPPATVEK	70	1018,4902	2034,9660	2034,9511	0.0149	7.3
Nta	2–23	aSDSAVATSASPVAAPPATVEKK	123	1063,5540	2125,0948	2125,0903	0.0045	2.1
Nta, 2mK23	2–23	aSDSAVATSASPVAAPPATVEK2mK	21	1077,5563	2153,0982	2153,1216	−0.0234	10.91 [b]
Nta, mK22, mK23	2–23	aSDSAVATSASPVAAPPATVEmKmK	21	1077,5555	2153,0965	2153,1216	−0.0251	11.7 [b]
Nta, 2mK22	2–23	aSDSAVATSASPVAAPPATVE2mKK	28	1077,5564	2153,0982	2153,1216	−0.0234	10.9 [b]
Nta, pS11	2–23	aSDSAVAT3ApSPVAAPPA1V[KK]	94	1103,5410	2205,0677	2205,0566	0.0111	5.0
Nta, pS9	2–23	aSDSAVATpSASPVAAPPATVEKK	69	1103,5390	2205,0496	2205,0566	−0.0070	3.2
Nta, pS4	2–23	aSDpSAVATSASPVAAPPATVEKK	73	1103,5390	2205,0497	2205,0566	−0.0069	3.1
p(S2 or S4)	2–23	p(SDS)AVATSASPVAAPPATVEKK	29	541,7661	21,630,353	2163,0460	−0.0107	4.9
pSll	2–23	SDSAVATSApSPVAAPPATVEKK	44	541,7661	2163,0353	2163,0460	−0.0107	4.9
pS9	2–23	SDSAVATpSASPVAAPPATVEKK	40	541,7660	2163,0453	2163,0460	−0.0007	0.3
pS2, pSil	2–23	pSDSAVATSApSPVAAPPATVEKK	33	748,6748	2243,0044	2243,0123	−0.0079	3.5
pS2, pS4	2–23	pSDpSAVATSASPVAAPPATVEKK	28	748,6754	2243,0044	2243,0123	−0.0079	3.5
pS4, pS1l	2–23	SDpSAVATSApSPVAAPPATVEKK	33	748,6748	2243,0044	2243,0123	−0.0079	3.5
pT8, pS1l	2–23	SDSAVApTSApSPVAAPPATVEKK	34	748,6748	2243,0044	2243,0123	−0.0079	3.5
pT8, pS9	2–23	SDSAVApTpSASPVMPPATVEKK	25	748,6754	2243,0044	2243,0123	−0.0079	3.5
pS9, pS1l	2–23	SDSAVATpSApSPVAAPPATVEKK	31	748,6750	2243,0044	2243,0123	−0.0079	3.5
Nta	2–27	aSDSAVATSASPVAAPPATVEKKVVQK	75	860,8008	2579,3805	2579,3806	−0.0001	0.0
Nta, 2mK27	2–27	aSDSAVATSASPVMPPATVEKKWQ2mK	13	870,1430	2607,4092	2607,4119	−0.0027	1.0
Nta, mK23, uK27	2–27	aSDSAVATSASPVAAPPATVEKmKWQuK	18	677,8740	2707,4706	2707,4392	0.0314	11.6 [b]
Nta, pS1l	2–27	aSDSAVATSApSPVAAPPATVEKKVVQK	64	887,4570	2659,3509	2659,3470	0.0039	1.5
Nta, pS9	2–27	aSDSAVATpSASPVAAPPATVEKKVVQK	64	930,1542	2659,3445	2659,3470	−0.0025	0.9
Nta, pTl9	2–27	aSDSAVATSASPVAAPPApTVEKKVVQK	52	930,1542	2659,3445	2659,3470	−0.0025	0.9
Nta, pS1l, mK23, uK27	2–27	aSDSAVATSApSPVAAPPATVEKmKVVQuK	25	697,8672	2787,4399	2787,4055	0.0344	12.3 [b]
pT8	2–27	SDSAVApTSASPVAAPPATVIKKVVQK	30	655,3389	2617,3269	2617,3364	−0.0095	3.6
pS4, pS1l	2–27	SDpSAVATSApSPVAAPPATVEKKWQK	33	675,3309	2697,2904	2697,3027	−0.0123	4.6
pS9, pS1l	2–27	SDSAVAT**pS**A**pS**PVAAPPATVEKKVVQK	30	675,3299	2697,3001	2697,3027	−0.0026	1.0
pT8, pS1l	2–27	SDSAVA**pT**SA**pS**PVAAPPATVEKKVVQK	34	675,3299	2697,3001	2697,3027	−0.0026	1.0
pT8, pS9, mK22, uK23	2–27	SDSAVA**pTpS**ASPVAAPPATVE**mKuK**VVQK	25	707,3542	2825,3879	2825,3613	0.0266	9.4
oxM52	39–58	KASATPSHPPTQQ**oM**VDASIK	60	704,0250	2109,0561	2109,0524	0.0037	1.8
2mK39, oxM52	39–58	**2mK**ASATPSHPPTQQ**oM**VDASIK	31	713,3622	2137,0649	2137,0837	−0.0188	8.8
oxM52, 2mK58	39–58	KASATPSHPPTQQ**oM**VDASI**2mK**	12	713,3641	2137,0706	2137,0837	−0.0131	6.1
2mK39, oxM52, 2mK58	39–58	**2mK**ASATPSHPPTQQ**oM**VDASI**2mK**	15	722,7046	2165,0922	2165,1150	−0.0228	10.5 [b]
p(S66 or S67), 2mK72	62–72	ERGG**p(SS)**LLAI**2mK**	32	413,5515	1237,6327	1237,6431	−0.0104	8.4
uK8O	74–85	YITAT**YuK**CDAQK	39	759,8750	15,177,364	1517,7184	0.0180	11.9 [b]
2mK122	108–122	GKGASGSFKLSASA**2mK**	26	475,2644	1422,7716	1422,7831	−0.0115	8.1
2mK132	131–143	S**2mK**VLSAEKKVQSK	27	487,3015	14,588,826	1458,8769	0.0057	3.9
aK164, uK165	157–168	TAVGAAD**aKuK**PK	21	621,3281	1240,6697	1240,6775	−0.0078	6.3
aK250, aK255	249–256	A**aK**TTAA**aK**K	17	451,7685	901,5224	901,5232	−0.0008	0.9

[a] Determined from the DNA sequence considering N-terminal cleaved off M as residue 1.
[b] Peptides identified at a peptide mass tolerance of 12.5 ppm.

Please cite this article as: Bonet-Costa C, et al, Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of..., J Prot (2012), doi:10.1016/j.jprot.2012.05.034

using the integrated Xcalibur Xtract software (Thermo Scientific). All data validation was done using ProSight PC 2.0 software (Thermo Scientific, San Jose, CA) in single protein mode using sequence gazer option. To reduce the noise arising from low-abundance non-specific peaks, an in-house algorithm (kindly provided by Prof. Neil Kelleher) was used to trim the fragment mass list before introducing it into ProSight software. For each .puf file, fragments were sorted into 50 to 100 Da mass bins and only the three or five most intense fragment ions within each bin were retained [35].

2.7. Antibodies

Rabbit αdH1 antibody was kindly provided by Dr. Kadonaga. Rabbit polyclonal α2meK27dH1 antibodies were generated by conventional methods using a peptide spanning dH1 residues 22 to 35, containing dimethylated K27, purified by affinity-chromatography and their specificity determined by dot-blot and Western-blot assays (see Fig. 7A and B).

2.8. FACS sorting and immunostaining experiments

Drosophila S2 cells were grown in Schneider medium containing 10% FCS to a density of ~1×10^6 cells/ml. Cells were harvested, washed in cold PBS and fixed in 70% ethanol overnight. Propidium Iodide-RNAse treated cells [36] were sorted using an ARIA Sorp (Becton Dickinson) machine and were recovered in PBS. Cells were recovered by centrifugation and resuspended in SDS-PAGE protein loading buffer for Western-blot analysis. Immunostaining experiments in S2 cells were performed as previously described [37] using α2meK27dH1 antibodies.

3. Results and discussion

3.1. dH1 shows a complex PTM pattern

dH1 was obtained from cultured S2 cells by extraction with 10% $HClO_4$ (see Materials and methods section for details) (Fig. 1A). Extracts were then subjected to LC–MS purification on a reverse phase BioSuite pPhenyl column, where dH1 eluted as a single peak at t_R 12.88 min (Fig. 1B). The corresponding on-line LC–MS spectrum, once deconvoluted, showed evidence of a complex pattern of potential PTMs (Fig. 1C). Some of the observed peaks fit by mass with varying degrees of phosphorylation. Also, we detected peaks that could be associated with concomitant acetylation or methylation. This initial TOF (time of flight)-resolution analysis revealed that dH1 is extensively modified post-translationally.

3.2. Mapping dH1 PTMs by bottom-up MS

Conventional bottom-up MS analysis was used to map PTM sites. For this purpose, we employed a hybrid 7 T LTQ-FT Ultra mass spectrometer (Thermo Scientific), capable of high resolution and high accuracy analysis in the ICR (ion cyclotron resonance cell) part of the instrument. To identify modifications, we used various MS methodologies: standard MS/MS to search for all kind of PTMs, and multistage activation (MSA) and neutral loss (NL) to identify phosphorylation sites (see Materials and methods section for details). All mapped PTMs are listed in Table 1 and summarized in Fig. 2. Altogether, we have identified: seven phosphorylation sites, two T and five S; two mono- and eight dimethylated K; three acetylated K; and four ubiquitinated K. In addition, a third monomethylated K was identified by top-down MS (see Table 5). Oxidation of the single M52 was also detected. Notice that most of these modifications were located within the N-terminal domain and, although it is the largest domain in dH1, only five modifications were mapped to the C-terminal domain. This low number is probably due to the high K-content of this domain, which makes it particularly refractory to analysis.

The PTM pattern of dH1 shows several relevant features. On one hand, dH1 has cleaved off N-terminal methionine (M1) (Table 1 and Fig. 2). Cleavage of the N-terminal methionine is most frequently associated with N^{α}-terminal acetylation, which occurs co-translationally in the same polypeptide (reviewed in [38]). However, in the case of dH1, the resulting N-terminal residue (S2) is present in both acetylated and non-acetylated forms (Table 1 and Fig. 2). Analysis of the intact protein by top-down MS confirmed these results (see Tables 2 and 3). This type of heterogeneity, where N^{α}-terminally acetylated and non-acetylated forms co-exist, appears to be a general phenomenon in histone H1, as it has also been

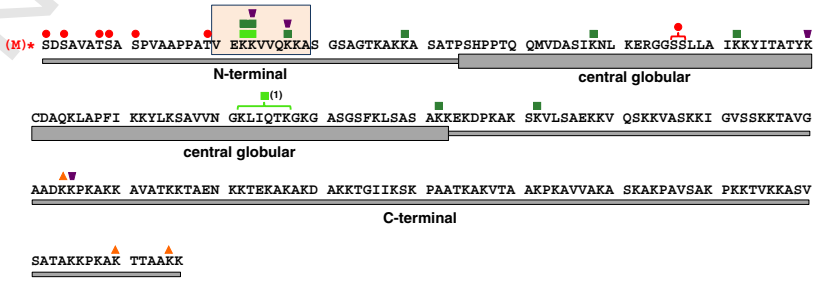

Fig. 2 – Summary of identified PTMs of dH1. The amino acid sequence of dH1 is presented. PTMs are indicated: phosphorylation (red circles), monomethylation (light green squares), dimethylation (dark green squares), acetylation (orange triangles) and ubiquitination (purple trapezoids). Cleavage of the N-terminal M (red in parenthesis) and N^{α}-terminal acetylation (red asterisk) is indicated. Domain organization of dH1 is shown underneath the sequence. Within the N-terminal domain, the position of a conserved region, spanning residues 19 to 29, is indicated (pale orange box). (1)Monomethylation mapped by top-down MS analysis of dH1 (see Table 5).

Table 2 – Isoforms detected by top-down MS of chemically oxidized dH1. The unmodified form is compatible with N^{α}-terminal acetylation. –Nta, non-N-terminal acetylation. 1p, 2p and 3p, mono-, di- and triphosphorylation. 1m, 2m and 3m, mono-, di- and trimethylation. ac, K-acetylation. (1)Low abundant. (2) and (3) isoforms for which –1 Da and +1 Da species are also detected, respectively.

Monoisotopic obsM	Monoisotopic theoM	ppm	PTM
26290.0086	26291.2069	45.57	–Nta(1)
26334.2132	26334.2016	0.44	Unmodified(2)
26348.2127	26348.2172	0.16	1m(3)
26362.1289	26362.2329	3.90	2m(2)(3)
26372.1755	26372.1573	0.69	–Nta,1p
26376.2307	26376.2485	0.67	3m(3)
	26376.2121	0.71	ac(3)
26386.2317	26386.1729	2.20	–Nta,1p, 1m
26414.2124	26414.1679	1.68	1p(2)
26428.1769	26428.1835	2.49	1p, 1m(2)(3)
26442.1993	26442.1992	1.73	1p, 2m(2)(3)
26456.1509	26456.2147	2.41	1p, 3m(2)(3)
	26456.1783	1.04	1p, ac(2)(3)
26466.2230	26466.1392	3.17	–Nta, 2p, 1m
26494.1604	26494.1342	0.98	2p(2)
26546.1170	26546.1055	0.43	–Nta, 3p, 1m
26574.1380	26574.1005	1.41	3p
26588.0526	26588.1161	2.38	3p, 1m(3)

observed in all human, mouse and chicken variants, where, in addition, the canonical N^{α}-termini has been detected [11,12]. In this study, we did not identify non-cleaved dH1 forms by either bottom-up or top-down MS analysis.

Phosphorylation is the most frequent PTM of histone H1. Abundant evidence indicates that, in mammals, histone H1 is phosphorylated at multiple sites in a cell cycle-dependent manner, as many phosphorylations map to consensus CDK-sites, and, furthermore, the abundance of polyphosphorylated forms increases during cell cycle progression [12,13,15,16,39–41]. The single histone H1 variant of *Tetrahymena* also shows multiple phosphorylations [14], some occurring at CDK-sites, and previous work in *D. melanogaster* identified a single major phosphorylation site at S11 [17], providing also evidence for the presence of a second phosphorylation site. Our bottom-up analysis confirmed phosphorylation of S11 and, in addition, we identified six other phosphorylation sites (Table 1 and Fig. 2). None of these sites correspond to consensus CDK-sites, as dH1 contains no CDK sites at all. Histone H1 phosphorylation at non-CDK sites has also been reported in *Tetrahymena* [14], and in several human, mouse and chicken variants [11–13,40]. Six out of the seven mapped phosphorylation sites of dH1 locate at the N-terminal domain (S2, S4, T8, S9, S11 and T19), and the seventh occurs within the central globular domain at S66 or S67. Interestingly, most non-CDK sites mapped in other histone H1 variants are also located within the N-terminal tail [11,12,14,27,40], and a non-CDK site has been mapped in the central globular domain of mouse H1.1 and H1.4 [12].

Apart from phosphorylation, K-methylation, K-acetylation and K-ubiquitination have been reported in vertebrate histone H1 variants [11–13]. K-acetylation has also been detected in *Tetrahymena* [14] and, in addition, previous work by others showed that dH1 is ubiquitinated, although the precise site(s) were not identified [42]. Our bottom-up analysis identified several K-methylation, K-acetylation and K-ubiquitination sites in dH1 (Table 1 and Fig. 2). Interestingly, three methylated K-residues mapped at a conserved region of the N-terminal tail (K22, K23 and K27), being also methylated in some vertebrate variants (see Fig. 7E under Biological significance section) [11,12]. These three residues are detected dimethylated and, in addition, K22 and K23 are also detected monomethylated. Fig. 3 shows the MS/MS spectrum of peptide ^{2}aSDSAVATSASPVAAPPATVEKK^{23}, where K22 and K23 are detected both mono- and dimethylated. As determined from the intensity of the discriminating ions b_{21}^{2+}, the dimethylated 2mK22 form is significantly more abundant than the double monomethylated mK22mK23 or the dimethylated 2mK23 forms. Note that K23 and K27 are also detected ubiquitinated. It must also be noticed that we did not identify any trimethylated K-residues or methylation at R-residues. In fact, no K-trimethylation has been reported for histone H1 and evidence of R-methylation is limited to one residue in human H1.4 [12]. Finally, in comparison to vertebrate variants, K-acetylation appears to be less frequent in dH1, as we mapped only three acetylated K-residues at the C-terminal domain (K164, K250 and K255).

Table 3 – Isoforms detected by top-down analysis of dNtH1 peptide. The unmodified form is compatible with N^{α}-terminal acetylation and homoserine lactone derivatization at C-terminus. –Nta, non-N-terminal acetylation. 1p, 2p, 4p, and 5p, mono-, di-, tetra- and pentaphosphorylation. 2m, dimethylation.

Monoisotopic obsM	Monoisotopic theoM	ppm	PTM
4979.6358	4979.6316	0.84	Unmodified
5059.6011	5059.5979	0.63	1p
5087.6694	5087.6292	7.90	1p, 2m
5097.5497	5097.5537	0.78	–Nta, 2p
5139.5826	5139.5642	3.58	2p
5257.4823	5257.4863	0.76	–Nta, 4p
5337.4777 [a]	5337.4527	4.68	–Nta, 5p

[a] Manually deconvoluted.

3.3. *Analysis of the structural heterogeneity of dH1 forms*

The identification of multiple PTMs by bottom-up MS suggests that individual dH1 molecules might contain different combinations of modifications. Initial on-line LC–MS analysis of the intact protein supports this hypothesis (Fig. 1C). In addition, our bottom-up analyses identified a number of peptides carrying several PTMs (Table 1). To further address the structural heterogeneity of dH1 forms, we performed top-down MS of intact dH1. For this purpose, RP-HPLC-fractions corresponding to purified dH1 (Fig. 1A) were analyzed on a 7 T LTQ-FT Ultra mass spectrometer. Individual species corresponding to charge +26 (m/z 1015) were selected (isolation width 10 m/z) and FT-MS data was derived from an average of 500-to-1000 transient signals to enhance the detection of the less represented forms with masses within approximately ± 130 Da. Note that, under these experimental conditions, ubiquitinated forms are not detectable. The profile obtained after Xtract deconvolution shows a peak corresponding to the N^{α}-terminally acetylated form, plus peaks corresponding to

Please cite this article as: Bonet-Costa C, et al, Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of..., J Prot (2012), doi:10.1016/j.jprot.2012.05.034

b ions **aS**$_1^{21}$**D**$_2^{20}$**S**$_3^{19}$**A**$_4^{18}$**V**$_5^{17}$**A**$_6^{16}$**T**$_7^{15}$**S**$_8^{14}$**A**$_9^{13}$**S**$_{10}^{12}$**P**$_{11}^{11}$**V**$_{12}^{10}$**A**$_{13}^{9}$**A**$_{14}^{8}$**P**$_{15}^{7}$**P**$_{16}^{6}$**A**$_{17}^{5}$**T**$_{18}^{4}$**V**$_{19}^{3}$**E**$_{20}^{2}$**2mK**$_{21}^{1}$**K** y ions

b ions **aS**$_1^{21}$**D**$_2^{20}$**S**$_3^{19}$**A**$_4^{18}$**V**$_5^{17}$**A**$_6^{16}$**T**$_7^{15}$**S**$_8^{14}$**A**$_9^{13}$**S**$_{10}^{12}$**P**$_{11}^{11}$**V**$_{12}^{10}$**A**$_{13}^{9}$**A**$_{14}^{8}$**P**$_{15}^{7}$**P**$_{16}^{6}$**A**$_{17}^{5}$**T**$_{18}^{4}$**V**$_{19}^{3}$**E**$_{20}^{2}$**mK**$_{21}^{1}$**mK** y ions

b ions **aS**$_1^{21}$**D**$_2^{20}$**S**$_3^{19}$**A**$_4^{18}$**V**$_5^{17}$**A**$_6^{16}$**T**$_7^{15}$**S**$_8^{14}$**A**$_9^{13}$**S**$_{10}^{12}$**P**$_{11}^{11}$**V**$_{12}^{10}$**A**$_{13}^{9}$**A**$_{14}^{8}$**P**$_{15}^{7}$**P**$_{16}^{6}$**A**$_{17}^{5}$**T**$_{18}^{4}$**V**$_{19}^{3}$**E**$_{20}^{2}$**K**$_{21}^{1}$ **2mK** y ions

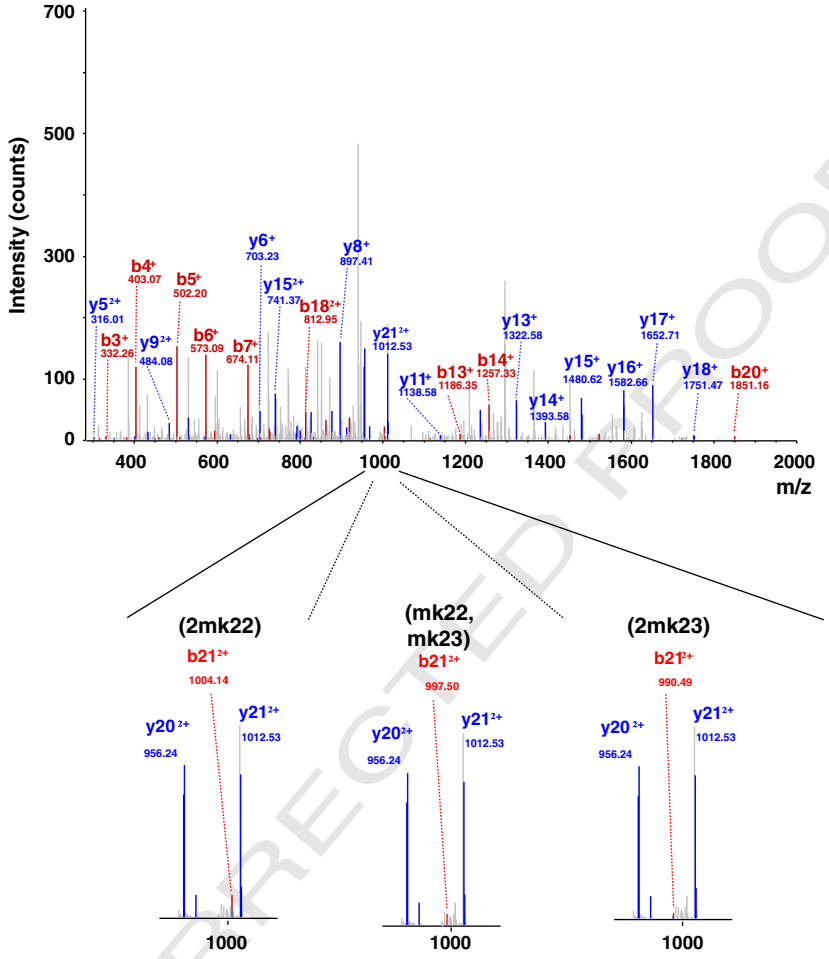

Fig. 3 – K22 and K23 are found both mono- and dimethylated. MS/MS spectrum corresponding to tryptic peptides showing dimethylation at K22, dimethylation at K23 and double monomethylation at K22 and K23 (see Table 1 for details). The inset shows a blow-up of the region corresponding to the discriminating b_{21}^{2+} ions. 2m, indicates dimethylation. m, indicates monomethylation. aS, indicates acetylated S2.

mono- and diphosphorylated forms (Fig. 4A and Table SI). Associated with these three major forms, intermediate species that could fit with other added modifications were also detected (Table SI). However, they could not be unambiguously assigned due to the difficulty of deconvolution when overlapping isotopes, coming from neighboring modified forms of similar weight, occur simultaneously with different abundances. Actually, we detected partial oxidation of the sample during analysis, which causes spreading of the signal across multiple oxidized masses, potentially masking biological relevant PTMs. More specifically, partial oxidation of M and C residues interferes with PTM identification by top-down MS, as it results in mass shifts similar to those of some specific modifications [33,43]. For instance, oxidation of M-residues to methionine sulfoxide results in a mass shift similar to methylation, with overlapping isotopic distributions that originate mass errors when deconvoluting to zero or single charged masses. dH1 contains one M-residue (M52) and one C-residue (C81). Partial oxidation of M52 was detected by bottom-up analysis (Table 1), and, in addition, intact protein analysis showed prominent peaks fitting by mass with oxidation of both the unmodified and monophosphorylated forms (Table SI).

To overcome this problem, dH1 was oxidized by mild performic acid (MPA) treatment [33] to fully convert M52 to methionine sulfone and C81 to cysteic acid. After purification, oxidized dH1 was analyzed in a 12 T LTQ-FT Ultra mass spectrometer to reach higher resolution and clear mass shift elucidation of the intermediate species between the various phosphorylated forms. The MS profile obtained was similar to that of the non-oxidized form (Fig. 4B and Table 2). The chemically oxidized unmodified form fits an N^{α}-terminal acetylation, and mono-, di- and triphosphorylated forms are clearly distinguishable. In addition, several peaks show mass

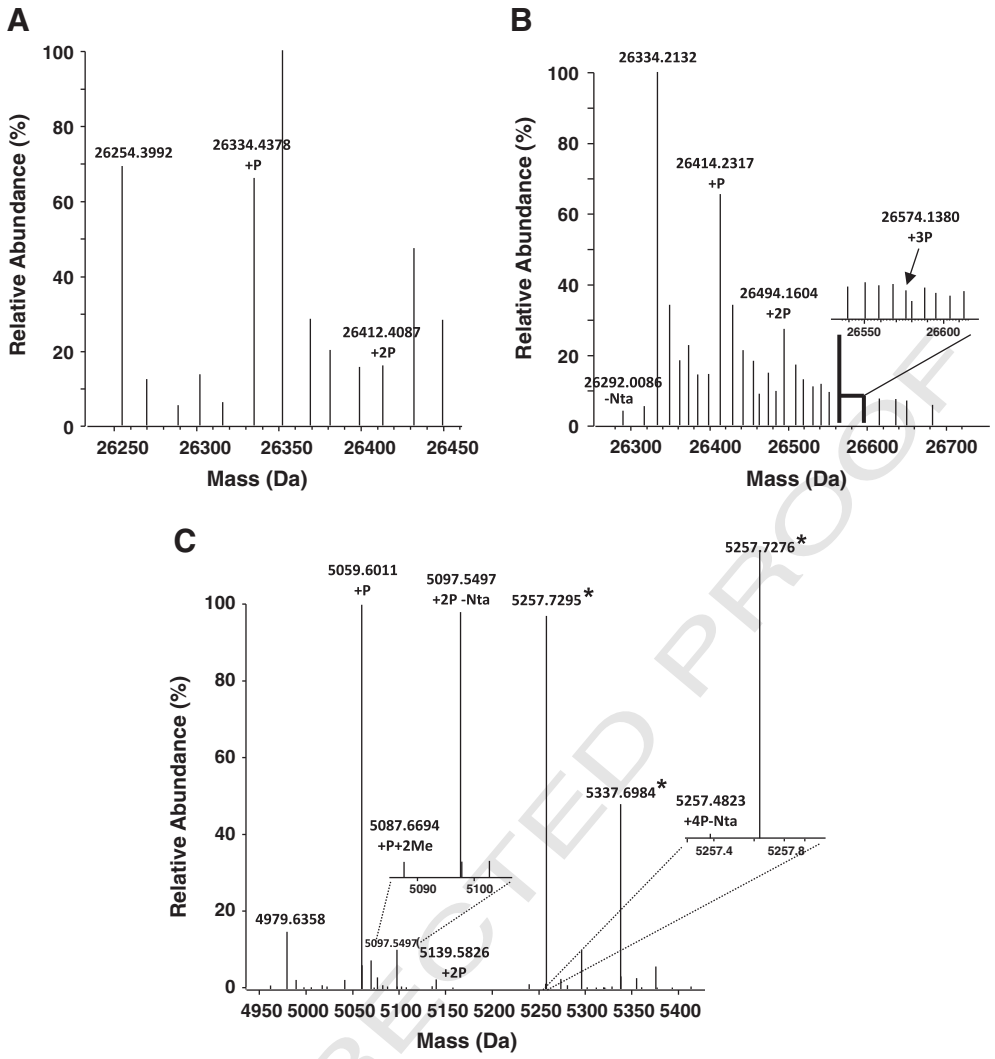

Fig. 4 – Multiple PTMs co-exist in dH1. (A) Xtract deconvolution of the MS spectrum of RP-HPLC-purified full length dH1 acquired on a 7 T LTQ-FT Ultra after ion isolation of species of charge z = 26 (isolation width 10 m/z). Monoisotopic mass (Da) of selected peaks corresponding to the unmodified N^{α}-terminally acetylated, and mono- and diphosphorylated dH1 forms is indicated. See also Table SI. (B) As in A but corresponding to chemically oxidized dH1. In this case MS spectrum was acquired on a 12 T LTQ-FT Ultra (Thermo Scientific). Monoisotopic mass (Da) of selected peaks corresponding to the unmodified N^{α}-terminally acetylated and non-acetylated, and mono-, di- and triphosphorylated dH1 forms are indicated. The inset is a blow-up of the region showing triphosphorylated dH1 forms. See also Table 2. (C) Xtract deconvolution of the MS spectrum of RP-HPLC-purified dNtH1 peptide acquired on a 7 T LTQ-FT Ultra after isolation of ion m/z 865 (isolation width 100 m/z) and allowing 1000 scans averaging. Monoisotopic mass (Da) of selected peaks corresponding to the indicated dNtH1 forms is shown. Prominent peaks that could not be assigned from the observed mass shifts are indicated with an asterisk. The two insets correspond to blow-ups of the regions showing monophospho-dimethylated and tetraphosphorylated dNtH1 species, respectively. See also Table 3.

increments of 14 Da and 28 Da, corresponding to mono- and dimethylations, respectively. Notice that phosphorylated forms are also found mono- or dimethylated. Increments of 42 Da, which could correspond to either trimethylated or acetylated forms, are also observed. Repetitive increments of 37 Da (79–42 Da) are also detected, indicating the presence of phosphorylated forms that are not N^{α}-terminally acetylated, thereby confirming bottom-up results showing N^{α}-terminal heterogeneity of dH1 (Table 1). As a matter of fact, the unmodified non-acetylated form is also detected as a minor species although, in this case, the monoisotopic mass could not be accurately measured, probably due to bad ion statistics. These results are summarized in Table 2. For various isoforms, we also detected the presence of species differing in −1 Da (Table 2), suggesting the presence of amidated forms (OH to NH2: −0.984016 Da). Top-down analyses mapped this modification to ^{124}EKD^{126} (Figure S1), suggesting the existence of forms carrying an E124Q or D126N substitution that could arise from single nucleotide polymorphisms (SNPs) [44]. This polymorphism could not be detected in bottom-up analyses, as we did not obtain coverage of this region. In this study, we have also encountered some misassignments of monoisotopic peaks

Please cite this article as: Bonet-Costa C, et al, Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of..., J Prot (2012), doi:10.1016/j.jprot.2012.05.034

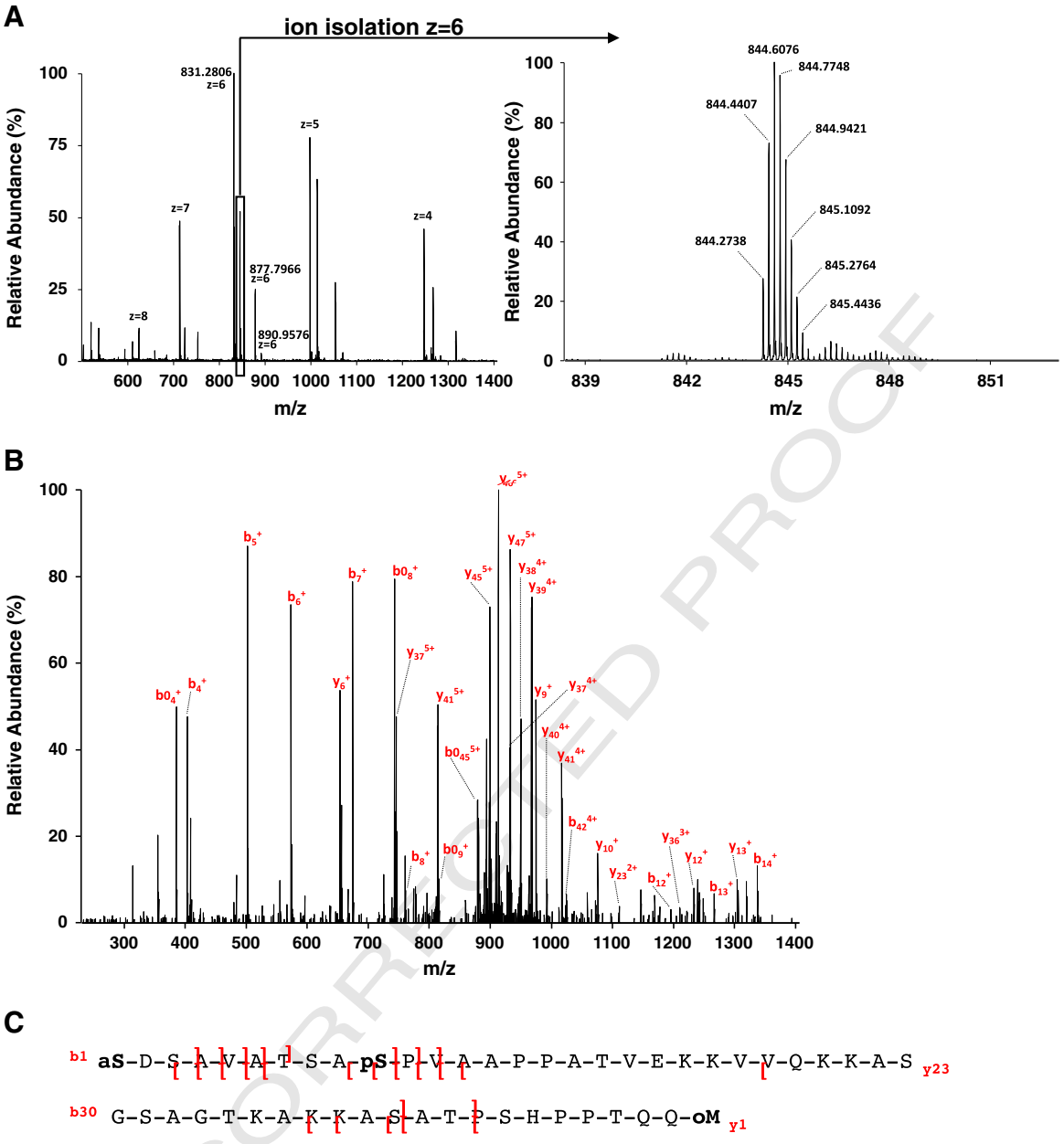

Fig. 5 – Mapping PTM sites in dNtH1 by top-down MS. (A) On the left, the broadband MS spectrum of RP-HPLC-purified dNtH1 acquired on a 7 T LTQ-FT-Ultra is presented. On the right, MS spectrum after isolation of ion m/z 844, corresponding to charge z = 6 species of the monophosphorylated dNtH1 form. (B) CID fragmentation spectrum of the isolated z = 6 charge state species. The most relevant b- and y-ions are indicated. (C) CID fragmentation map showing phosphorylation at S11 and N$^{\alpha}$-terminal acetylation in bold. The C-terminal M derivatized as homoserine lactone is also indicated in bold. Brackets facing N-terminal and C-terminal indicate b- and y-ions coverage, respectively.

with masses off by +1 Da with respect to some isoforms (Table 2). As suggested earlier by others [44–46], variations in the natural $^{12}C/^{13}C$ abundance of laboratory samples with slightly altered isotopomer distributions, together with deconvolution software imperfections, can shift by 1 Da the assignment of the calculated monoisotopic mass, which is based on an averagine isotope envelope model. Moreover, the presence of deamidated species arising from SNPs, could also account for these results (46). It must be noted, however, that no deamination could be mapped to any specific site by bottom-up or top-down analyses.

To further characterize structural heterogeneity of dH1 isoforms, we also explored a middle-down approach [47] since proteins of large molecular weight, such as dH1 (26 kDa), are in general difficult to analyze by top-down MS. For this purpose, we used CNBr to cleave dH1 at the single M52 residue, which results in two peptides spanning residues 2–51 (dNtH1) and residues 52–255 (dCtH1), the former derivatized as homoserine lactone at the C-terminus. In this study, we focused on the analysis of the dNtH1 peptide (t_R 14.88 min) that was purified by on-line LC-nanoESI-MS using an Advion TriVersa NanoMate as interface

source, splitter and fraction collector. Purified fractions were reanalyzed by automated nanoESI ionization. Fig. 4C shows deconvolution of the spectrum obtained by infusion using a window isolation of 100 centered at m/z 865 and allowing an averaging of 1000 transient signals. Together with mono- and diphosphorylated forms carrying N^{α}-terminal acetylation, species corresponding to a monophospho-dimethyl form and a diphosphorylated form missing N^{α}-terminal acetylation are also detected. In addition, two prominent peaks are observed (m/z 5257.7295 and 5337.6984), which could not be assigned from the observed mass shifts. Accurate inspection of the region corresponding to peak m/z 5257.7295 reveals the presence of a minor species fitting by mass with a tetraphosphorylated form missing N^{α}-terminal acetylation, which coionizes with a major species differing only in 0.24 Da (Fig. 4C). A similar situation is observed at the region corresponding to m/z 5337.6984, where manual deconvolution shows the presence of a pentaphosphorylated form of very low abundance. CID analysis of the major species coionizing together with the tetra- and pentaphosphorylated forms suggests that it arises from an uncharacterized side reaction of the CNBr cleavage occurring at C-terminus (Figure S2). Table 3 summarizes these results.

3.4. *Mapping co-existing PTMs by top-down MS*

Results reported above indicate the presence of dH1 isoforms carrying several phosphorylations and methylations. Next, to map PTMs co-existing in individual dH1 forms, we performed MS/MS analyses in the dNtH1 peptide and the full-length dH1 protein. For this purpose, individual ions were selected and subjected to CID or ECD. Fragmentation maps were then interrogated for the presence of PTMs using ProSightPC™ (see Materials and methods section for details).

In the case of the dNtH1 peptide, ion m/z 844 (z=6), corresponding to the monophosphorylated and N^{α}-terminally acetylated form, was selected with an isolation width of 10 m/z (Fig. 5A). Subsequent fragmentation by CID (Fig. 5B), revealed four major monophosphorylation sites (S11, S9, T19 and S4), based on pScores of the corresponding fragmentation maps (Table 4). Fig. 5C shows the full CID fragmentation map corresponding to phosphorylation at S11. The rest of modified species were analyzed together, as the use of narrow isolation widths for isolation of their corresponding charged ions, and their low abundances, limited the quality of the MS/MS spectra. Results obtained after isolation of ion m/z 865 (z=6, isolation width=100 m/z) and CID fragmentation are summarized in Table 4. In the case of the monophospho-dimethyl species, two forms are detected showing phosphorylation at S11 or T19, and a dimethylation, or two monomethylations, at K22, K23 or K27. In the case of the diphosphorylated species, a major form phosphorylated at S9 and S11 is detected for the N^{α}-terminally acetylated form. For the non-acetylated form, though the exact positions could not be assigned, the two phosphorylation sites map between S2 and S11. In the case of the tetra- and pentaphosphorylated forms, the phosphorylated sites could not be mapped due to the interference of the unassigned species described above (Fig. 4C). Notice that all detected PTMs are consistent with those identified in bottom-up analyses (Table 1 and Fig. 2).

When full-length dH1 was analyzed, the oxidized protein was purified and infused through the Advion TriVersa NanoMate. Individual ions were then isolated and subjected to fragmentation by CID. It must be noted that, even when narrow isolation widths were used, selected ions corresponded to a mixture of modified forms of similar mass, as the large molecular weight of dH1, the complexity of the modifications and their low abundance prevented isolation of homogeneous species. For the interrogation of monophosphorylated species, ion m/z 1102 (z=24) was isolated with an isolation width 8 m/z (Fig. 6A). Subsequent CID (Fig. 6B) confirmed the four major phosphorylation sites identified in the dNtH1 peptide. In this case, however, data could not discriminate between phosphorylation at S4 and S2 (Table 5). In addition, fragment ions were also consistent with phosphorylation at T8 and S66 or S67 (Table 5). The interrogation of the same monophosphorylated species (same ion isolation) by ECD (Fig. 6C) showed a pattern of fragmentation with predominant C-terminal ions. In this case S66 or S67 were also confirmed as consistent phosphorylated residues (Table 5). Fig. 6D shows the full CID and ECD maps showing phosphorylation of S11. Fragmentation maps also contained information about other modified species. In particular, data were consistent with a monophospho-monomethyl species showing phosphorylation at S11 or T19, and monomethylation at K22, K23 or K27. Isolation of a second ion m/z 1098 (z=24) with an isolation width=50 provided evidence of monomethylation at K102 or K107 (Table 5), a modification that was not

Table 4 – PTMs identified by top-down MS of dNtH1 peptide. The indicated ions were isolated and subjected to CID fragmentation. –Nta, non-N-terminal acetylation. 1p and 2p, mono- and diphosphorylation. 2m, dimethylation. pS, phosphoserine. pT, phosphothreonine. pScore and PDE score (ProSightPC™) are indicated. The number of ion fragments detected (b- and y-ions) and the corresponding percentage (%) of total possible ions are indicated. See Supplementary data S2 for CID fragmentation spectra corresponding to the indicated PTMs.

PTM	Ion	obsM	theoM	Δm	ppm	Residue(s) [a]	pScore	PDE	Fragments	%	b-Ions	y-Ions
1p	844	5059.6011	5059.5979	0.0032	0.6325	pS11	$5.4\ e^{-41}$	144	27	26	10	17
						pS9	$4.1\ e^{-39}$	144	26	25	9	17
						pT19	$9.8\ e^{-25}$	76.3	18	17	7	11
						pS4	$4.6\ e^{-18}$	116	17	16	5	12
1p, 2m	865	5087.6694	5087.6292	0.0402	7.9015	pS11, 2m(K22/23/27)	$3.4\ e^{-19}$	80.7	16	23	5	18
						pT19, 2m(K22/23/27)	$3.4\ e^{-19}$	87.4	16	23	5	18
–Nta, 2p	865	5097.5497	5097.5537	–0.004	0.7847	2p(pS2–pS11)	$1.1\ e^{-17}$	106	15	21	1	14
2p	865	5139.5826	5139.5642	0.0184	3.5801	pS9,pS11	$9.7\ e^{-21}$	124	17	24	3	14

[a] Determined from the DNA sequence considering N-terminal cleaved off M as residue 1.

Please cite this article as: Bonet-Costa C, et al, Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of..., J Prot (2012), doi:10.1016/j.jprot.2012.05.034

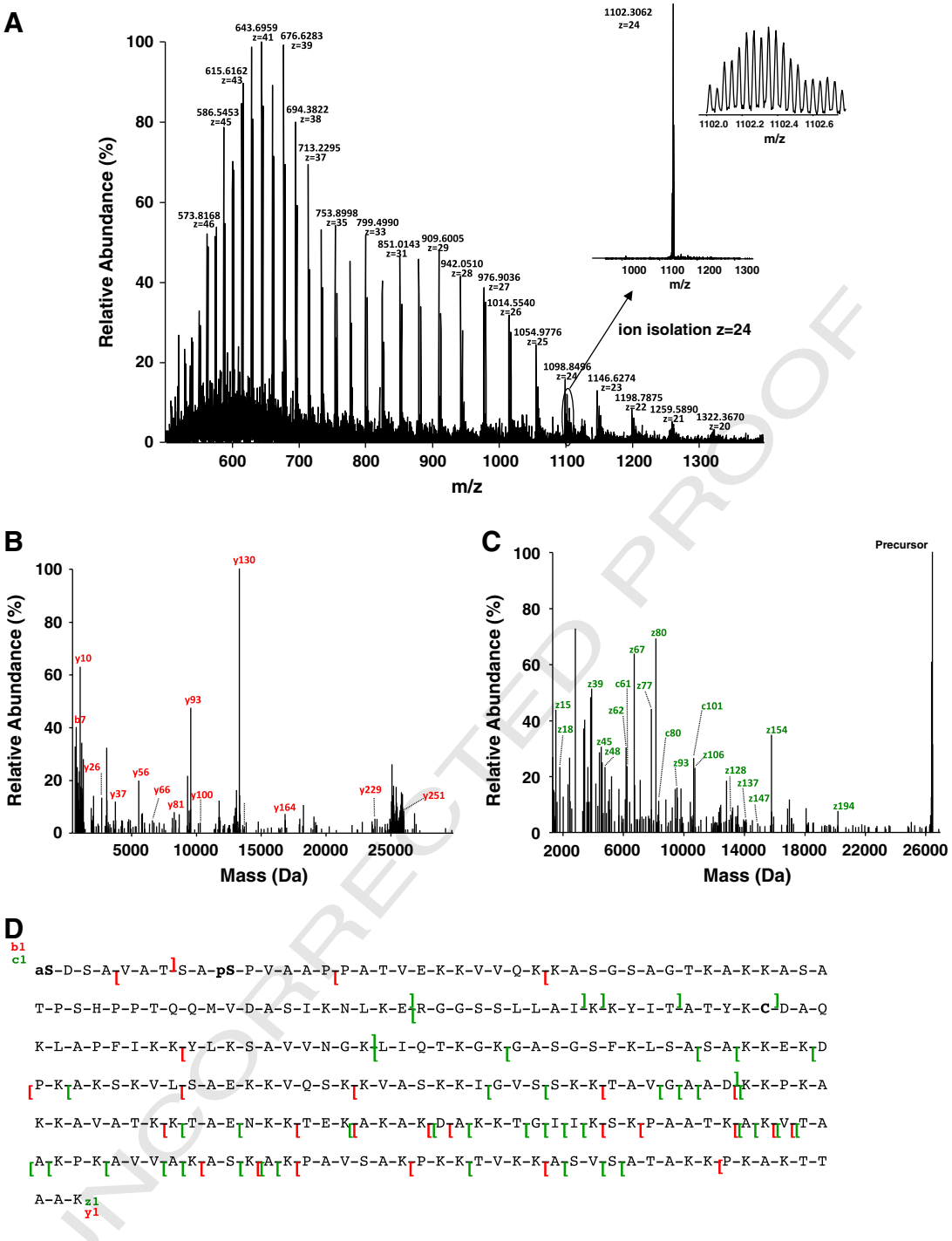

Fig. 6 – Mapping PTM sites in dH1 by top-down MS. (A). The broadband MS spectrum of RP-HPLC-purified oxidized dH1 acquired on a 12 T LTQ-FT-Ultra is presented. The inset shows the MS spectrum after isolation of ion m/z 1102 corresponding to charge z=24 species with a molecular mass consistent with a monophosphorylated dH1 form showing N^{α}-terminal acetylation and oxidation of M52 to methionine sulfone and C81 to cysteic acid resulting from MPA treatment. Panels (B) and (C) show CID and ECD fragmentation spectra of the isolated z=24 charge state species, respectively. The most relevant b- and y-ions (CID), and c- and z-ions (ECD), are indicated. (D) Full CID and ECD fragmentation map showing phosphorylation at S11 and N^{α}-terminal acetylation in bold. Oxidized C81 is also indicated in bold. Red brackets facing N-terminal and C-terminal indicate b- and y-ions coverage, respectively. Green brackets facing N-terminal and C-terminal indicate c- and z-ions coverage, respectively.

Table 5 – PTMs identified by top-down MS of intact dH1. The indicated ions were isolated and subjected to CID or ECD fragmentation. –Nta, non-N-terminal acetylation. 1p and 2p, mono- and diphosphorylation. 1m, 2m and 5m, mono-, di- and pentamethylation. pS, phosphoserine. pT, phosphothreonine. pScore and PDE score (ProSightPC™) are indicated. The number of ion fragments detected (b- and y-ions, or c-(1) and z-ions(2)) and the corresponding percentage (%) of total possible ions are also indicated. See Supplementary data S3 for CID/ECD fragmentation spectra corresponding to the indicated PTMs. The asterisk indicates a form carrying a putative amidation. (3)Determined from the DNA sequence considering N-terminal cleaved off M as residue 1.

PTM	Ion	obsM	theoM	Δm	ppm	Residue(s) (3)	pScore	PDE	Fragments	%	b-Ions	y-Ions
1p	1102	26414.2385	26414.1679	0.0706	2.67	pS11	$9.4\ e^{-13}$	163	26	11	1	25
						pS9	$9.2\ e^{-13}$	163	26	11	1	25
						pT8	$5.7\ e^{-12}$	161	25	11	0	25
						pT19	$5.7\ e^{-12}$	162	25	11	1	24
						1p(S2/S4)	$5.7\ e^{-12}$	161	25	11	0	25
						1p(S66/S67)	$3.4\ e^{-11}$	161	24	10	1	23
						ECD						
						1p(S2–S55)	$3.5\ e^{-26}$	44.5	49	20	9(1)	40(2)
						1p(S66/S67)	$3.5\ e^{-26}$	43	48	19	7(1)	41(2)
1p, 1m	1102	26428.1769	26428.1835	−0.0066	0.25	pS11, m(K22/23/27)	$2.1\ e^{-22}$	170	37	16	2	35
						pT19, m(K22/23/27)	$2.1\ e^{-22}$	166	37	16	2	35
–Nta, 2p *	1102	26451.1512	26451.1396	0.0116	0.44	pS2. pS4	$3.1\ e^{-15}$	174	29	13	0	29
						2p(T8–S11)	$4.5\ e^{-16}$	175	30	13	0	30
						1p(S2/S4), 1pS11	$6.1\ e^{-17}$	175	31	14	0	31
1m	1098	26348.2016	26348.2172	−0.0156	0.59	mK(22–27)	$6\ e^{-7}$	85	22	7	1	21
1p	1098	26414.2385	26414.1679	0.0706	2.67	1p(T8/S9/S11)	$8.8\ e^{-11}$	88.9	28	9	4	24
						pT19	$4.2\ e^{-10}$	89.6	27	9	4	23
–Nta, 1p, 5m	1098	26442.1627	26442.1992	−0.0365	1.38	1p(T8–T19), 4m(K23–K39), m(K102–K107)	$8.7\ e^{-9}$	84	25	8	2	23
1p, 2m	1098	26442.1627	26442.1995	−0.0368	1.39	1p(T8–T19), 2m (K23–K39)	$3.7\ e^{-8}$	85.3	24	8	2	22

detected by bottom-up analysis. Additional modifications identified by top-down MS are summarized in Tables 4 and 5.

3.5. *Biological significance*

Results reported above identified multiple PTMs in dH1. Their functional significance is, however, unknown. As a first step to address this question, we raised antibodies against some of the identified PTMs. In particular, we generated α2mK27dH1 antibodies that specifically recognize 2mK27, which is a main dH1 PTM. These antibodies were raised against a peptide spanning residues 22 to 35, carrying dimethylated K27 (see Materials and methods section for details). As shown in Fig. 7A, α2mK27dH1 antibodies specifically recognize 2mK27, since they do not bind the same peptide unmethylated or dimethylated at K28. Similarly, α2mK27dH1 antibodies recognize dH1 purified from S2 cells, which contains endogenous PTMs, but not bacterially expressed recombinant dH1, which is not modified (Fig. 7B). Next, to determine the pattern of chromosomal distribution of 2mK27dH1, we performed immunostaining experiments in S2 cells using α2mK27dH1 antibodies (Fig. 7C). These experiments show that 2mK27dH1 occurs predominantly in metaphase cells, localizing preferentially at pericentromeric heterochromatin. To confirm these results, S2 cells were sorted at different cell-cycle phases (G1, S and G2/M) by FACS and, then, analyzed by Western-blot using α2mK27dH1. As shown in Fig. 7D, G2/M-cells have higher 2mK27dH1 content than either G1- or S-cells. Altogether these results indicate that 2mK27dH1 is regulated during cell-cycle progression, accumulating at pericentromeric heterochromatin, suggesting a contribution to heterochromatin organization and function during mitosis. Most interestingly, K27 occurs at a short conserved region within the N-terminal domain (residues 19 to 29), being also methylated in some human/mouse variants [12] (Fig. 7E), indicating that the potential contribution of this modification to heterochromatin organization and function might be evolutionarily conserved. Two additional K-residues (K22 and K23) located within this conserved region are also found methylated in dH1 (Figs. 2 and 7E). Methylation of K22 has also been detected in some chicken H1 variants [11]. In addition, K22 is also acetylated in some human/mouse/chicken H1 variants [11,12] and, although with a low score, this PTM is also detected in dH1 (not shown). Altogether, these observations suggest that methylation/acetylation at these sites is evolutionarily conserved from *Drosophila* to vertebrates, suggesting that these modifications play conserved functions.

Phosphorylation is a main PTM of histone H1 [11–17,39–41]. Most H1 variants contain multiple consensus CDK-sites that preferentially locate at the C-terminal domain and appear to be hierarchically phosphorylated during cell cycle progression. Here we mapped multiple phosphorylations in dH1, all occurring at non-CDK sites mostly located within the N-terminal domain. Interestingly, H1 phosphorylation at non-CDK sites of the N-terminal domain has also been reported in vertebrates and *Tetrahymena* [11,12,14,27,40]. Phosphorylation at the N-terminal domain of dH1 also appears to be hierarchical, as S11 constitutes a major phosphorylation site (Tables 1, 4 and 5) [17], which occurs in most polyphosphorylated forms identified by either top-down or bottom-up analyses (Tables 1, 4 and 5). As a matter of fact, *in silico* prediction using NetPhos 2.0 Server Software (Technical University of Denmark) identifies S11 as the main phosphorylation site within the N-terminal domain (Table SII), and, most remarkably, S11 phosphorylation strongly favors phosphorylation at S9 and T8 (Table SII).

Please cite this article as: Bonet-Costa C, et al, Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of..., J Prot (2012), doi:10.1016/j.jprot.2012.05.034

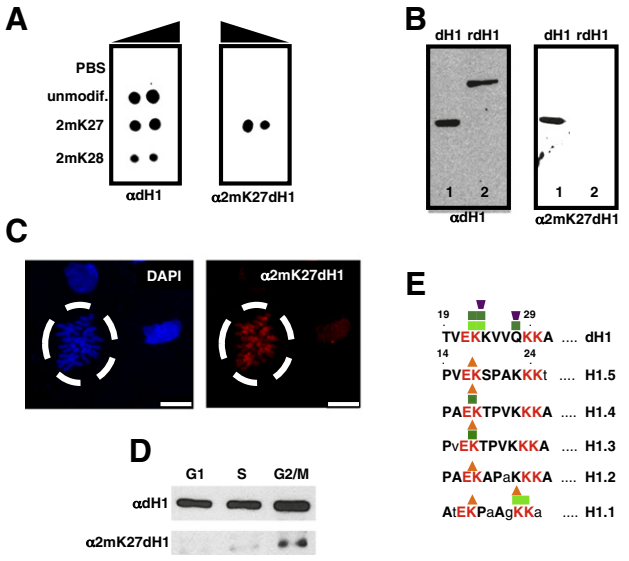

Fig. 7 – Dimethylation of K27 accumulates at peri-centromeric heterochromatin during mitosis. (A) Specificity of α2mK27dH1 antibodies was determined by dot-blot using increasing amounts of the dH1-peptide spanning residues 22 to 35 carrying dimethylated K27 (row 2mK27), dimethylated K28 (row 2mK28), unmethylated (row unmodif.) or no peptide (row PBS). Membranes were assayed for binding with α2mK27dH1 (right) and, as a control, αdH1 (left). (B) Specificity of α2mK27dH1 antibodies was determined by Western-blot using endogenous dH1 purified from S2 cells (dH1, lanes 1) and bacterially expressed recombinant dH1 (rdH1, lanes 2). Membranes were assayed for binding with α2mK27dH1 (right) and, as a control, αdH1 (left). (C) Immunostaining of S2 cells with α2mK27dH1 antibodies (in red). The circle shows a metaphase cell showing intense α2mK27dH1 in peri-centromeric heterochromatin. DNA was stained with DAPI (in blue). (D) Analysis of the abundance of 2mK27dH1 during cell-cycle progression. S2-cells were sorted by FACS at G1, S and G2/M-phases and analyzed by Western-blot with α2mK27dH1 antibodies (bottom) and, as a control, αdH1 antibodies (top). (E) Several dH1PTMs, including 2mK27dH1, map at a conserved region within the N-terminal domain of dH1. The amino acid sequence of dH1 (residues 19 to 29) and mouse/human H1.1, H1.2, H1.3, H1.4 and H1.5 variants (residues 14 to 24) is shown. For mouse/human variants, the mouse sequences are shown and residues that are not identical in the corresponding human variant are indicated in lower case. Identical residues are shown in red. PTMs are indicated: monomethylation (light green squares), dimethylation (dark green squares), acetylation (orange triangles) and ubiquitination (purple trapezoids).

4. Conclusions

Here we addressed determination of the PTM pattern of the *D. melanogaster* linker histone H1, dH1. For this purpose, we combined bottom-up and top-down MS strategies. Our results show that dH1 is extensively modified post-translationally by phosphorylation, methylation, acetylation and ubiquitination, with some residues showing different PTMs and several PTMs co-existing in the same molecule. Most previous studies on H1 modifications focused on the identification of phosphorylation sites and evidence of the existence of other PTMs was limited to a few vertebrate variants [11–17]. In particular, only a single phosphorylation site was previously mapped in dH1 [17]. To date, our study provides the most comprehensive characterization of the PTM pattern of any histone H1 variant. At least in part, this is due to the low H1 complexity of *D. melanogaster* that holds a single dH1 variant, which is in contrast to most metazoans that contain multiple closely related H1 variants.

Except monomethylation of K102 or K107, all dH1 PTMs were detected by bottom-up MS, reflecting the high sensitivity of this approach to unambiguously identify PTM sites. On the other hand, multiple dH1 isoforms carrying several phosphorylations and/or methylations were identified by top/middle-down MS, which illustrates the usefulness of this methodology to analyze the structural heterogeneity of complex mixtures of proteins and determine the PTM pattern of individual isoforms. However, it must be noticed that, in several cases, fragmentation efficiency by CID or ECD was insufficient to unambiguously map PTM sites. This is likely due to the relatively low charge state that had to be used as the precursor ion in order to select the desired isoform with a feasible isolation width. In fact, it has been described that higher charge states facilitate ECD fragmentation [48] and, in general, they are more unfolded species with higher fragmentation yields [49]. However, in our case, the isolation of higher charge states required very narrow isolation width and, therefore, sensitivity decreased dramatically.

Linker histone H1 PTMs likely constitute important epigenetic instructions that regulate chromatin structure and function. Results reported here and elsewhere [11,12] indicate that the PTM pattern of linker histone H1 is as complex as in core histones, pointing to a relevant functional contribution. Little is known, however, about the functional significance of histone H1 PTMs. Given the powerful genetic, molecular and cellular tools provided by *D. melanogaster*, and its low H1 complexity, we anticipate that identification of the PTM pattern of histone H1 in this model organism would strongly facilitate this analysis. Actually, we also provide evidence for the regulation during cell-cycle progression, and its potential contribution to heterochromatin function, of dimethylation of K27, a main dH1 PTM that appears conserved in vertebrates. It must be mentioned that, to date, no other histone H1 PTM was mapped at the chromosomal level.

Acknowledgments

The authors thank Dr. J. Kadonaga for αdH1 antibodies and Prof. Neil Kelleher (Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University) and the members of his research group for technical support and advice. Thanks also go to Prof. Julian Whitelegge for his helpful advice. We also acknowledge Dr. Michaela Scigelova and Dr. Vlad Zabrouskov from Thermo Scientific for instrumental training and advice. We are also grateful to Tanya Yates for correcting the manuscript. This work was supported by grants from MICINN

(CSD2006-49, BIO2008-00799 and BFU2009-07111), the CSIC (200420E391) and the "Generalitat de Catalunya" (SGR2009-1023 and SGR2009-1005). This work was carried out within the framework of the "Xarxa de Referència en Biotecnologia" of the "Generalitat de Catalunya".

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2012.05.034.

REFERENCES

[1] Bassett A, Cooper S, Wu C, Travers A. The folding and unfolding of eukaryotic chromatin. Curr Opin Genet Dev 2009;19:159–65.
[2] Kasinky HE, Lewis JD, Dacks JB, Ausio J. Origin of H1 linker histones. FASEB J 2001;15:34–42.
[3] Ramakrishnan V. Histone H1 and chromatin higher-order structure. Crit Rev Eukaryotic Gene Expr 1997;7:215–30.
[4] Robinson PJJ, Rhodes D. Structure of the "30 nm" chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol 2006;16:1–8.
[5] van Holde K, Zlatanova J. What determines the folding of the chromatin fiber. Proc Natl Acad Sci U S A 1996;93:10548–55.
[6] Mariño-Ramírez L, Hsu B, Baxenavis AD, Landsman D. The histone database: a comprehensive resource for histones and histone fold-containing proteins. Proteins 2006;62:838–42.
[7] Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, et al. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell 2005;123:1199–212.
[8] Sancho M, Diani E, Beato M, Jordan A. Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. PLoS Genet 2008;4:e1000227.
[9] Kouzarides T. Chromatin modifications and their function. Cell 2007;128:693–705.
[10] Ruthenburg AJ, Li H, Patel DJ, Allis DC. Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol 2007;8:983–94.
[11] Snijders APL, Pongdam S, Lambert SJ, Wood CM, Baldwin JP, Dickman MJ. Characterization of post-translational modifications of the linker histones H1 and H5 from chicken erythrocytes using mass spectrometry. J Proteome Res 2008;7:4326–35.
[12] Wisniewski JR, Zougman A, Krüger S, Mann M. Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylations as well as differences between cell culture and tissue. Mol Cell Proteomics 2007;6:72–87.
[13] Garcia BA, Busby SA, Barber CM, Shabanowitz J, Allis CD, Hunt DF. Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. J Proteome Res 2004;3:1219–27.
[14] Garcia BA, Joshi S, Thomas CE, Chitta RK, Diaz RL, Busby SA, et al. Comprehensive phosphoprotein analysis of linker histone H1 from *Tetrahymena thermophila*. Mol Cell Proteomics 2006;5:1593–609.
[15] Sarg B, Helliger W, Talasz H, Forg B, Lidner HH. Histone H1 phosphorylation occurs site-specifically during interphase and mitosis. Identification of a novel phosphorylation site on histone H1. J Biol Chem 2006;281:6573–80.
[16] Talasz H, Helliger W, Puschendorf B, Lidner HH. In vivo phosphorylation of histone H1 variants during the cell cycle. Biochemistry 1996;35:1761–7.
[17] Villar-Garea A, Imhof A. Fine mapping of posttranslational modifications of the linker histone H1 from *Drosophila melanogaster*. PLoS One 2008;3:e1553.
[18] Nagel S, Grossbach U. Histone H1 genes and histone clusters in the genus *Drosophila*. J Mol Evol 2000;51:286–98.
[19] Casado-Vela J, Cebrian A, Gomez T, Sanchez E, Vilaseca M, Menchen L, et al. Lights and shadows of proteomic technologies for the study of protein species including isoforms, splicing variants and protein post-translational modifications. Proteomics 2011;11:590–603.
[20] Hernández H, Dziembowski A, Taverner T, Séraphin B, Robinson CV. Subunit architecture of multimeric complexes isolated directly from cells. EMBO Rep 2006;7:605–10.
[21] Whitelegge JP, Zabrouskov V, Halgand F, Souda P, Bassilian S, Yan W, et al. Protein-sequence polymorphisms and post-translational modifications in proteins from human saliva using top-down Fourier transform ion cyclotron resonance mass spectrometry. Int J Mass Spectrom 2007;268:190–7.
[22] Zabrouskov V, Whitelegge JP. Increase coverage in the transmembrane domain with activated-ion electron capture dissociation for top-down Fourier-transform mass spectrometry of integral membrane proteins. J Proteome Res 2007;6:2205–10.
[23] Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. Mass Spectrom Rev 2005;24:168–200.
[24] Chait BT. Chemistry. Mass spectrometry: bottom-up or top-down? Science 2006;314:65–6.
[25] Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. Top down *versus* bottom up protein characterization by tandem high-resolution mass spectrometry. J Am Chem Soc 1999;121:806–12.
[26] Boyne II MT, Pesavento JJ, Mizzen CA, Kelleher NL. Precise characterization of human histones in the H2A gene family by top-down mass spectrometry. J Proteome Res 2006;5:248–53.
[27] Garcia BA, Pesavento JJ, Mizzen CA, Kelleher NL. Pervasive combinatorial modification of histone H3 in human cells. Nat Methods 2007;4:487–9.
[28] Pesavento JJ, Bullock C, LeDuc RD, Mizzen CA, Kelleher NL. Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top-down mass spectrometry. J Biol Chem 2008;282:14927–37.
[29] Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. Nat Methods 2007;4:817–21.
[30] Tran JC, Doucette AA. Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: a proteome prefractionation tool. J Proteome Res 2008;7:1761–6.
[31] Tran JC, Doucette AA. Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. Anal Chem 2008;80:1568–73.
[32] Johns EW. Studies on histones. 7. Preparative methods for histone fractions from calf thymus. Biochem J 1964;92:55.
[33] Pesavento JJ, Garcia BA, Streeky JA, Kelleher NL, Mizzen CA. Mild performic acid oxidation enhances chromatographic and top-down mass spectrometric analyses of histones. Mol Cell Proteomics 2007;6:1510–26.
[34] Ulintz PJ, Yocum AK, Bodenmiller B, Aebersold R, Andrews PC, Nesvizhskii A. Comparison of MS2-only, MSA, and MS2/MS3 methodologies for phosphopeptide identification. J Proteome Res 2009;8:887–99.
[35] Vellaichamy A, Tran JC, Catherman AD, Lee JE, Kellie JF, Sweet SMM, et al. Size-sorting combined with improved nanocapillary liquid chromatography–mass spectrometry for

Please cite this article as: Bonet-Costa C, et al, Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of..., J Prot (2012), doi:10.1016/j.jprot.2012.05.034

identification of intact proteins up to 80 kDa. Anal Chem 2010;82:1234–44.
[36] Darzynkiewicz Z, Juan G. Current protocols in cytometry. John Wiley & Sons, Inc.; 1997. 7.5.1–7.5.24.
[37] Font-Burgada J, Rossell D, Auer H, Azorín F. *Drosophila* HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC (ROW) to regulate gene expression. Genes Dev 2008;22:3007–23.
[38] Polevoda B, Sherman F. N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J Mol Biol 2003;325:595–622.
[39] Gurley LR, D'Anna JA, Barham SS, Deaven LL, Tobey RA. Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. Eur J Biochem 1978;84:1–15.
[40] Gurley LR, Valdez JG, Buchanan JS. Characterization of the mitotic specific phosphorylation site of histone H1. Absence of a consensus sequence for the p34cdc2/cyclin B kinase. J Biol Chem 1995;246:27653–60.
[41] Hohman P, Tobey RA, Gurley LR. Phosphorylation of distinct regions of f1 histone. Relationship to the cell cycle. J Biol Chem 1976;251:3685–92.
[42] Pham AD, Sauer F. Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*. Science 2000;289:2357–60.
[43] Medzihradszky KF, Zhang X, Chalkley RJ, Guan S, McFarland MA, Chalmers MJ, et al. Characterization of *Tetrahymena* histone H2B variants and posttranslational populations by electron capture dissociation (ECD) Fourier transform ion cyclotron mass spectrometry (FT-ICR MS). Mol Cell Proteomics 2004;3:872–86.
[44] Ryan CM, Souda P, Bassiliant S, Ujwal R, Zhang J, Abramson J, et al. Post-translational modifications of integral membrane proteins resolved by top-down Fourier transform mass spectrometry with collisionally activated dissociation. Mol Cell Proteomics 2010;9:791–803.
[45] Beavis RC. Chemical mass of carbon in proteins. Anal Chem 1993;65:496–7.
[46] Halgand F, Zabrouskov V, Bassilian S, Souda P, Wong DT, Loo JA, et al. Micro-heterogeneity of human saliva peptide P–C characterized by high-resolution top-down Fourier-transform mass spectrometry. J Am Soc Mass Spectrom 2010;21:868–77.
[47] Xu P, Peng J. Characterization of polyubiquitin chain structure by middle-down mass spectrometry. Anal Chem 2008;80:3438–44.
[48] Breuker K, Oh H, Lin C, Carpenter BK, McLafferty FW. Nonergodic and conformational control of the electron capture dissociation of protein cations. Proc Natl Acad Sci U S A 2004;101:14011–6.
[49] Mikhailov VA, Cooper HJ. Activated Ion Electron Capture Dissociation (AI ECD) of proteins: synchronization of infrared and electron irradiation with ion magnetron motion. J Am Soc Mass Spectrom 2009;20:763–71.

Additional results

The second paper presented in this thesis contains only a small part of the results that I did, part of biological significance of 2mK27dH1 modification. On the other hand, more experiments that are concerning dH1 PTMs are part of my thesis and are not presented in the paper. Due to this, I will make a description of these experiments and results apart, as well as materials and methods that we used.

Our laboratory made bottom up and top down mass-spectometry analysis of histone dH1 isolated from SL2 cells. In this way numerous PTMs were identified, methylation, acetylation, phosphorylation and ubiquitination. Before this work it was known only about the existence of dH1 phosphorylation and ubiquitination in *Drosophila melanogaster*. Aim of my PhD project was to biologically characterize some of these modifications situated in N-terminus of dH1 (Figure 9 shows detected N-terminus modifications). For this purpose we raised antibodies that recognize these modifications.

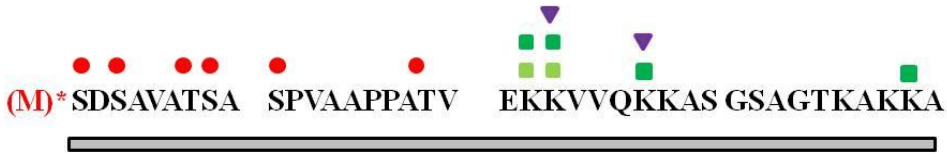

Figure 9: Postranslational modifications in N-terminus of dH1 that were detected in our laboratory. (M)* - N-terminal acetylation, red circle – phosphorylation, light green square – dimethylation, dark green square – monomethylation and purple triangle - ubiquitination

The majority of the experiments concern methylation of Lys at position 27 (K27) while for Ser10 phosphorylation and K22 and K23 methylation, there was time only to purify and characterize antibodies.

Raising and purification of the antibodies

To raise antibodies against the modifications rabbits were immunized for chemically synthesized peptides that correspond to the modified part of dH1. The animal facility was collecting samples of blood from the immunized animals in three time points: 1) before the immunization, 2) approximately a month after the immunization and 3) two months after the immunization. In this way, we could follow the synthesis of the antibodies against the epitope and choose the serum fraction with the best ability to recognize the peptides. Antibodies both for unmodified and modified versions of the peptides were used in two different animals (four in total for each modification).

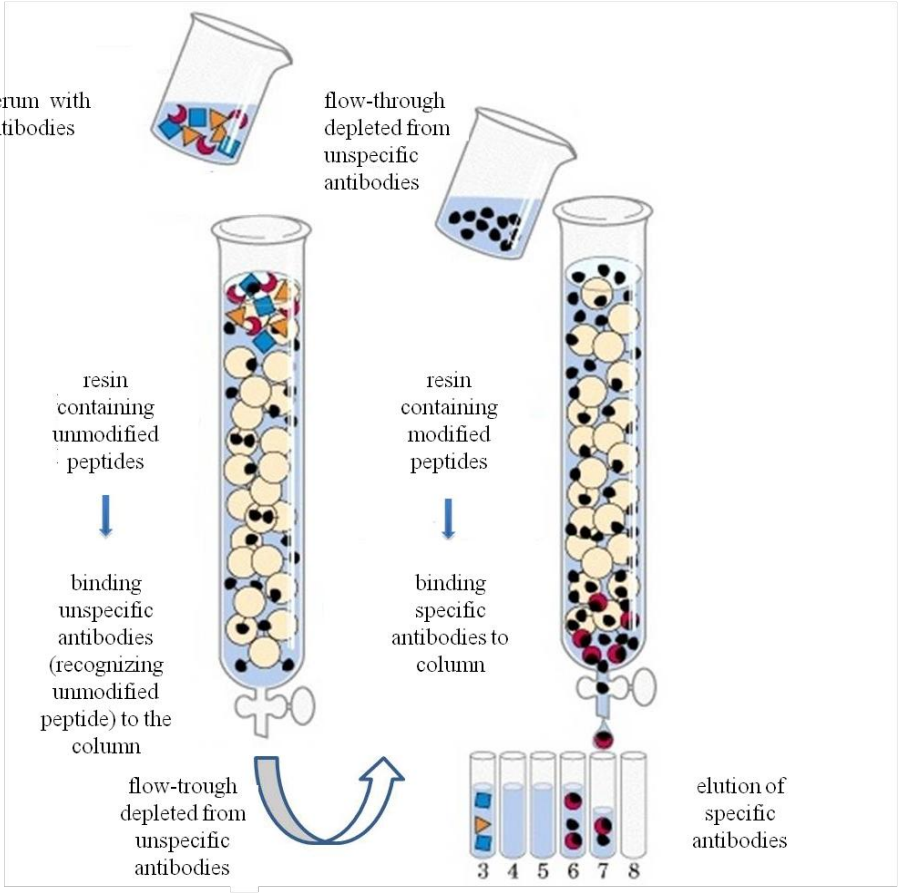

Figure 10: Process of purification of specific antibodies from rabbit serum. The serum first goes through the column with unmodified peptides to bind unspecific antibodies; Afterwards, flow-through then goes through column with modified peptides where specific antibodies will bind; After washes antibodies specific for modified peptides are eluted

Unmodified and modified peptides differ only in the presence of two methyl or one phospho groups so we had to perform extensive purification procedures to finally accomplish great degree of distinction of antibodies for the corresponding modified peptides. Serums from two animals immunized with modified peptide were first run through the affinity purification column that had unmodified peptide bound to its SulfoLink coupling resin. In this way are bound to the column and excluded from the serum all the antibodies that have affinity for recognition of unmodified peptide. Flow-through of this column was containing serum free of antibodies specific for unmodified peptide and we could further purify it by using column coupled with modified peptide. Antibodies that recognize modified peptide bound the peptides coupled to the column and after washes were eluted in few aliquots (Figure 10).

Characterization of α2mK23dH1 and αpSer10dH1

We managed in purifying α2mK23dH1 (Figure11A), which specifically recognizes only 2mK23 peptide, while K23 and 2mK22 peptide are not recognized by the antibody. In the case of αpSer10dH1 we have the signal specific for pSer10 peptide, but rather weak signal (Figure 11B). When it comes to α2mK22, we did not manage to purify antibody specific for 2mK22 peptide (data not shown).

Characterization of α2mK27dH1

In dot blot analysis (Figure 11C) antibodies for K27 and 2mK27 peptide specifically recognized both unmodified and modified peptide and only lysine di-methylated peptide, respectively.

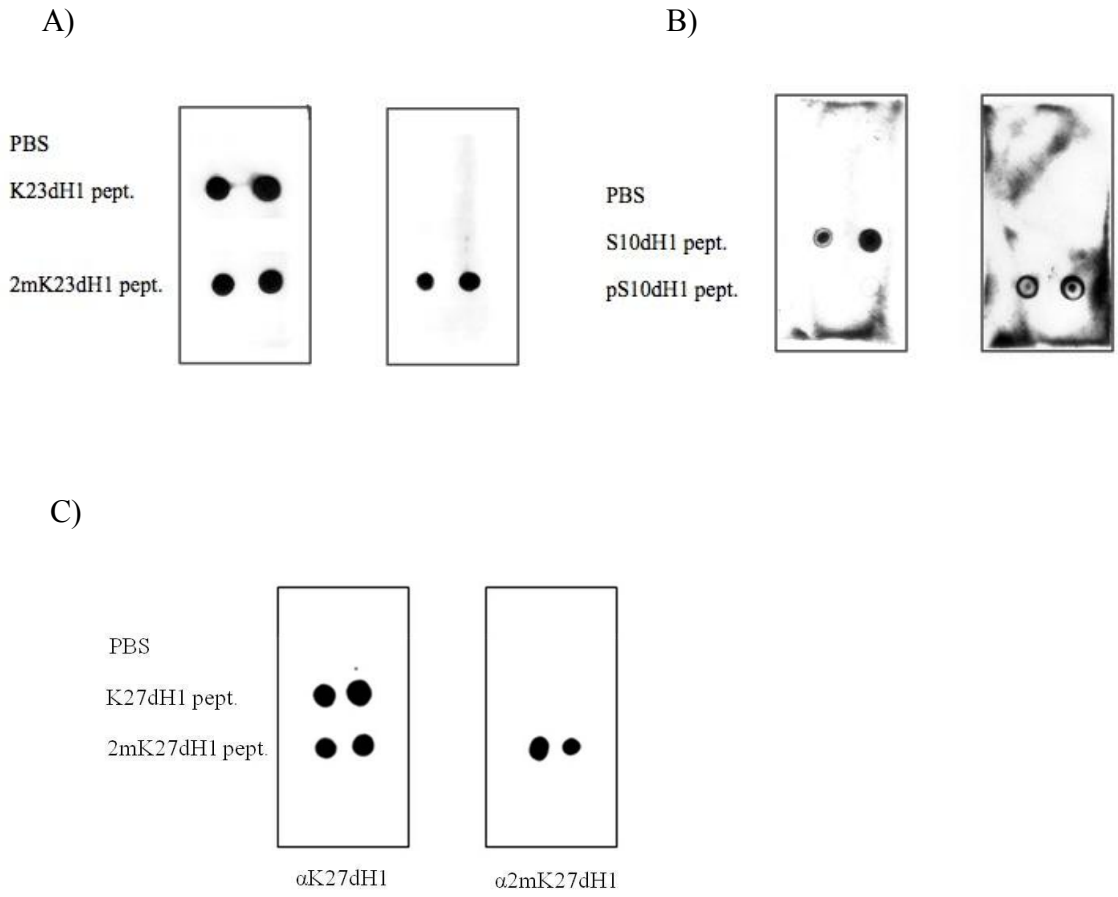

Figure 11: **A)** αK23dH1 recognizes both K23 and 2mK23 peptide (left), while α2mK23dH1 recognises exclusively 2mK23 peptide (right). **B)** αS10dH1 recognizes S10 peptide (left), while αpS10dH1 recognises exclusively pS10 peptide (right). **C)** αdH1 recognizes both K27 and 2mK27 peptide, while α2mK27dH1 recognises exclusively 2mK27 peptide. PBS is buffer used to dissolve the peptides

Western blot analysis (Figure 7B in paper 2) also showed that α2mK27dH1 is specific for 2mK27 group since the antibody recognizes better endogenous dH1 that histone H1 overexpressed in bacteria. One is comparing signal of the antibody in endogenous protein with the modification and a recombinant protein expressed in bacteria which is not modified. In this way, the difference in the signal is indicating a specificity of the antibody for the modification. Some cross-reactivity directed to unmodified form of the protein was also present, but there was a clear preference in recognition of the modified protein. After the extensive processes of purification through purification columns

α2mK27dH1 indeed recognized histone dH1 isolated from SL2 cells better than bacterially overexpressed H1. Degree of difference in recognition is significant, especially if we recall that percentage of the dH1 molecules containing the modification is low.

We included an additional experiment in SL2 cells to prove the specificity of the antibody for the modification. We performed immunostaining experiments with the antibody and antibody incubated with K27 or 2mK27 peptides. During the incubation with the antibody, the peptides are binding them if there is compatibility between them. In this way, immunostaining with the antibodies is not going to give a signal, since all recognition sites of the antibody will be occupied by the peptide and no binding to epitope used in immunostaining can occur. As expected for specific antibody, α2mK27dH1 in these competition experiments gave equal signals when it was used alone and when it was incubated with K27 or 2mK28 peptide or recombinant dH1 protein..

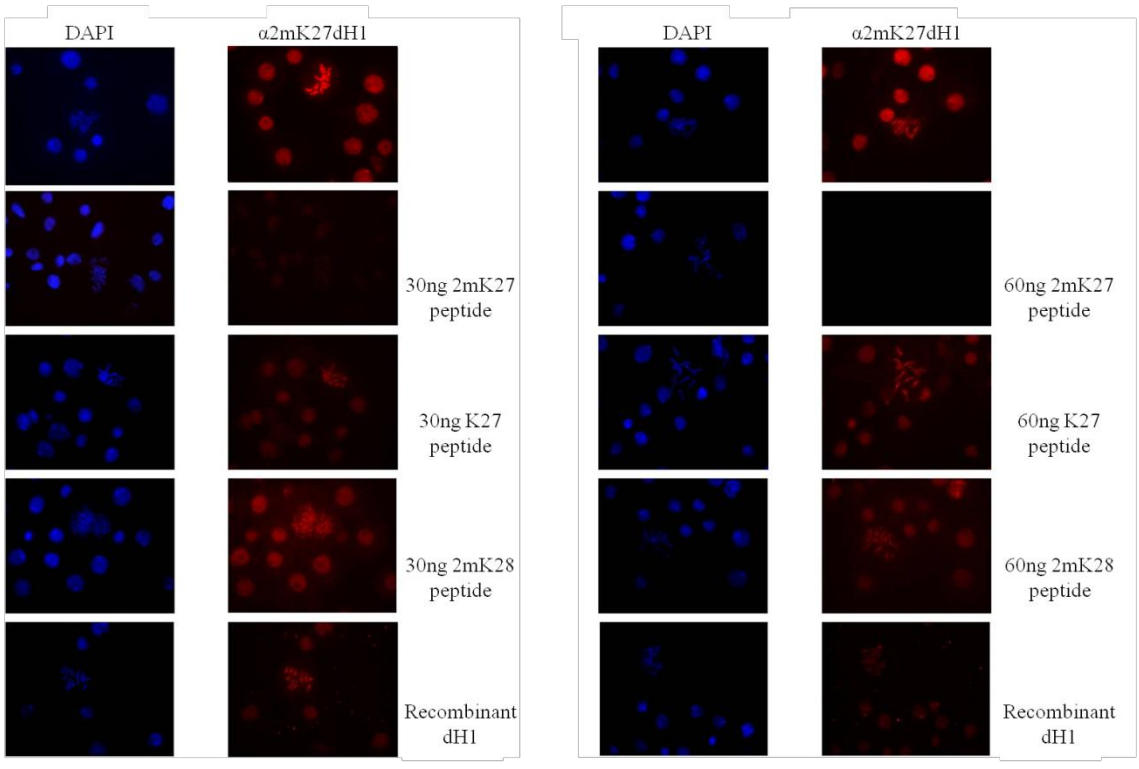

Figure 12: α2mK27dH1 binding to SL2 cells is competed with 2mK27 peptide, but not with K27, K28 peptide or recombinant dH1 protein

This is saying that the antibody does not recognize unmodified epitope or other examined epitopes. On the other hand, incubation of the antibody with 2mK27 peptide completely abolished antibody binding (Figure 12), saying that the antibody recognizes 2mK27 epitope.

These results together show that the antibody has required specificity and that it distinguishes 2mK27 from K27 epitope.

Functional analysis of α2mK27dH1

We performed immunostaining of polytene chromosomes with α2mK27dH1 and observed banding pattern and colocalization with HP1a protein at the chromocenter. Comparing the banding staining pattern of the antibody with DAPI staining pattern one can see that they overlap in a great manner, like was already reported (Jamrich et al., 1977). Band regions in polytene chromosomes are well stained both with DAPI and the antibodies while interband regions are stained poorly. However, we could not observe any particularity of α2mK27dH1 signal (Figure 13) compared to αK27dH1 signal.

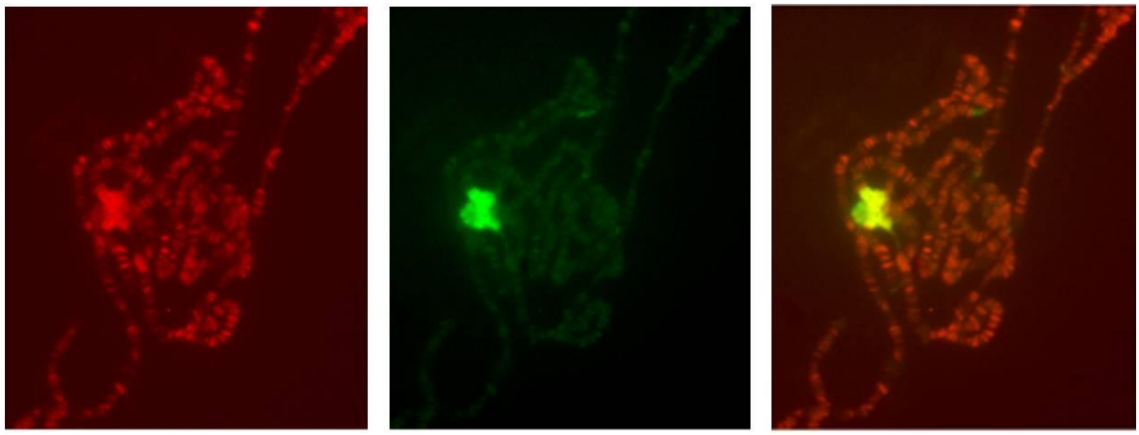

Figure 13: α2mK27dH1 gives banding pattern in polytene chromosomes and colocalizes with HP1a in chromocenter; α2mK27dH1-red, HP1a-green

Nevertheless, in SL2 cells, α2mK27dH1 shows distinctive staining pattern from αK27dH1. While antibody for unmodified peptide can bind uniformly all over the nucleus region (Figure 14 left) and all along metaphase chromosomes, α2mK27dH1 preferentially binds to centromeric regions of metaphase chromosomes (Figure 14 right).

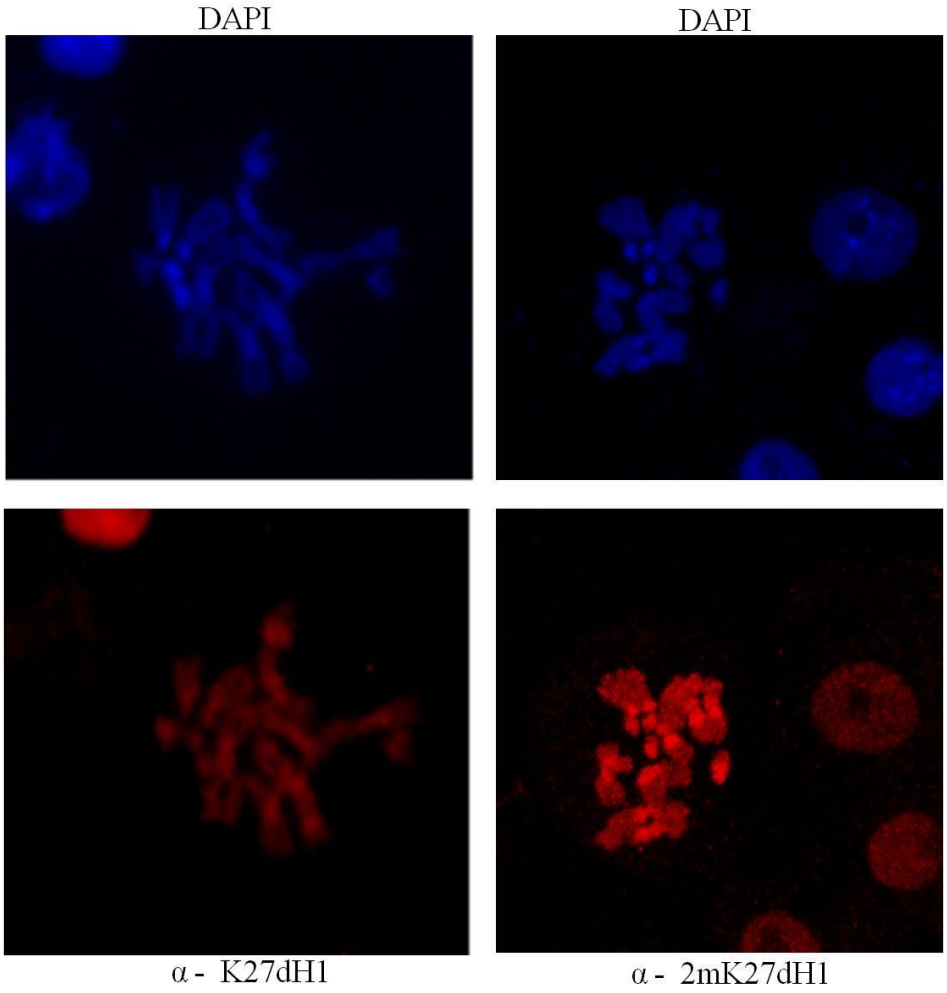

Figure 14: staining patterns of αK27dH1 and α2mK27dH1 in SL2 cell are different; DAPI -blue, αK27dH1 and α2mK27dH1-red

In order to learn *in vivo* about the physiological significance of the 2meK27dH1 modification mark, we constructed transgenic flies that express dH1 by its own, dH1, promoter and transgenic flies that overexpress histone dH1 directed by UAS sequence. At the C-terminus end we added a GFP-tag for the easier detection, distinction from the

endogenous wild type protein and for any further manipulation with the protein. We constructed wt dH1 and three types of dH1 with mutated lysine residue/s. Instead of Lys27 and/or Lys28, mutant histones have a similar amino acid, Ala, that cannot be methylated and which is often used as a substitution in similar studies. We had to include Lys28 in the analysis even though we did not detect any modification of this residue, because there is a possibility that the absence of methylation site at Lys27 can re-direct methylation activity of the enzyme that could introduce the mark towards adjacent Lys28.

First, we confirmed that there is expression of dH1-GFP transgene in these constructs by detecting direct GFP fluorescence under GFP exciting light (Figure 15). We could detect dH1-GFP expression only in the constructs carrying UAS promoter, while no tagged dH1 protein could be detected with the endogenous dH1 promoter. For this reason, in the further analysis we used only UAS-dH1-GFP flies.

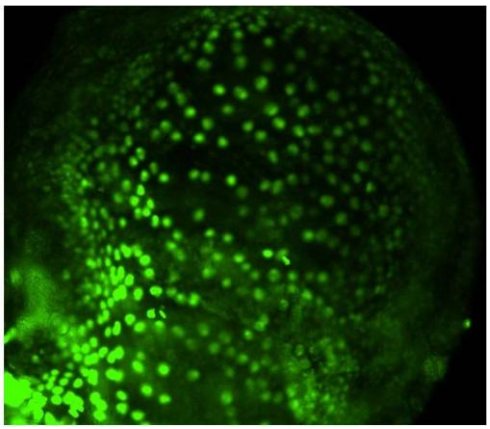

Figure 15: UAS-dH1-GFP gene is expressed in transgenic flies: direct GFP signal detected in cells of salivary glands; GFP - green

Next, we tried to observe a phenotype in these flies due to the competition for dH1 binding sites on chromatin between wild type and mutated form of the protein. We could not observe any lethality or any kind of milder phenotype.

We tried answering question of enzymes that regulate the modification mark. JMJD2A/KDM4 is a mammalian HMT that demethylates K26 in H1.4 (Trojer et al., 2009). We decided to test JMJD2A homologue in *D. melanogaster* for demethylase

activity towards 2mK27dH1. In our laboratory was available inducible cell line that overexpresses GFP-tagged version of JMJD2A (Lloret-Linares et al., 2008). Addition of copper sulphate induced part of the cells to express the demethylase and these cells could be detected by immunostaining with the antibodies that recognize the protein tag. We detected the level of dimethylated K27dH1 in these cells and compared it to the level of the signal in the surrounding cells, cells that had only endogenous level of the demethylase. If the enzyme is indeed demthylating 2mK27dH1, one would expect to see a decrease of the α2mK27dH1 signal in cells with the increased level of the JMJD2A enzyme. A master student in our laboratory, Sanela Mrkonjic, with my monitoring did these experiments. We used Leica SPE confocal microscope (High-Throughput Image acquisition) and quantified the intensity of the signal in nuclei. It was necessary to make images of whole microscope slide in order to have as many as possible cells that overexpress the enzyme, since small portion of them are induced to overexpress the enzyme.

The cells overexpressing JMJD2A show a decrease in α2mK27dH1 signal (Figure 16).

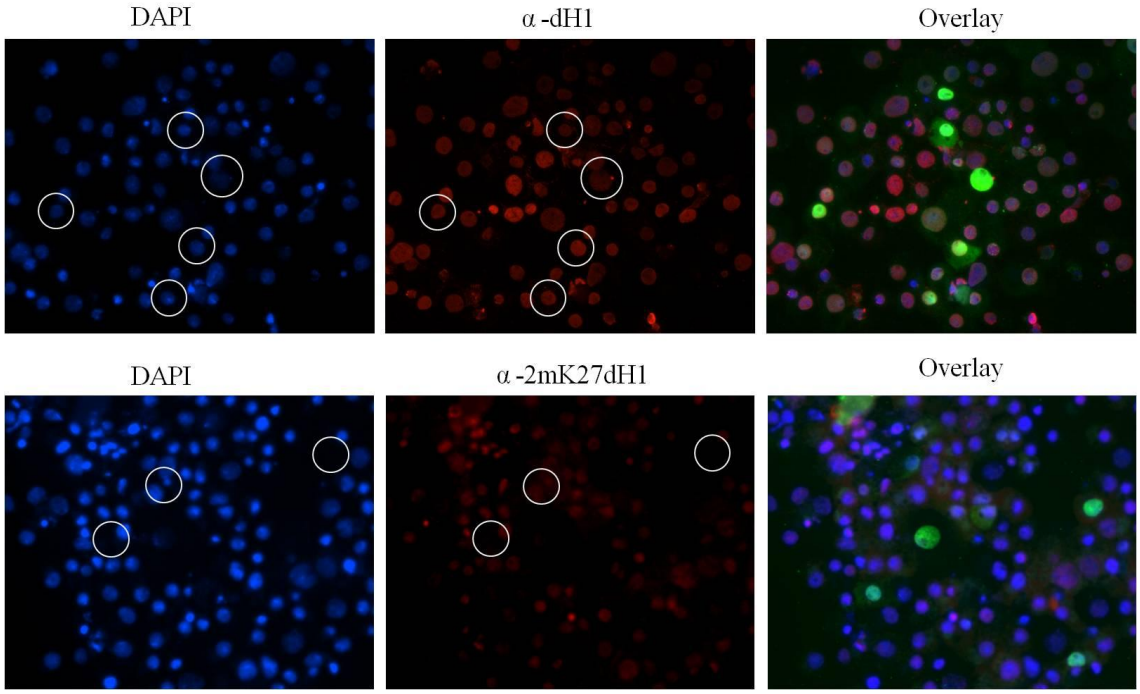

Figure 16: Down: Cells overexpressing JMJD2A (green signal) show a decrease in α-2mK27dH1 signal compared to the surrounding cells with control level of JMJD2A; **Up**: level of total H1 remains unchanged in cells overexpressing JMJD2A; DAPI – blue, dH1 – red (up), 2mK27dH1 – red (down), JMJD2A-GFP - green

However, quantification of the decrease was rather complex. The main problem we encountered was that the signal intensity of α2mK27dH1 was variable already in the cells that had wild type level of the enzyme (control cells) just like in cells that are overexpressing JMJD2A. We had to compare the mean value of signals in cells that overexpress and do not overexpress the demethylase. Advanced digital microscopy facility developed a program that was calculating a mean value of the signal in all pixels present in the region of the nuclei for control cells and cells that overexpress the enzyme. The mean values of both groups of cells were compared. As Figure 17 shows, there is a great decrease in α2mK27dH1 signal in cells that overexpress JMJD2A when compared to the surrounding cells that do not overexpress it. On the other hand, αdH1 signal is very similar in these two groups of cells (right). Our results are showing that there is decrease in α2mK27dH1 signal in JMJD2A overexpressing cells, going in favour of possibility that this enzyme is demethylating our modification of interest.

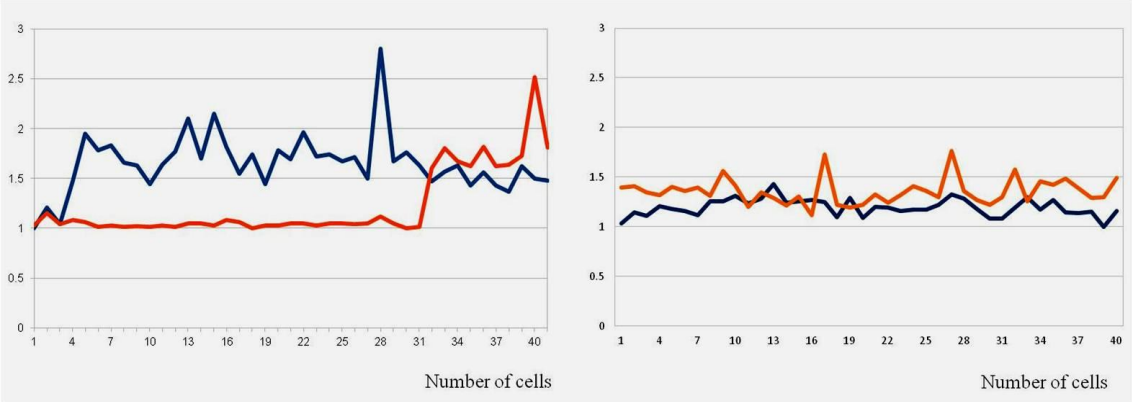

Figure 17: Left: α2mK27dH1 signal (red) and αK27dH1 signal (blue) in cells overexpressing JMJD2A; **Right**: α2mK27dH1 signal (blue) and αK27dH1 signal (red) in wild type cells. Y axis: antibodies signal value normalized by signal value in untransfected cells

Materials and methods

Antibodies purification

We purified antibodies from serums of rabbits immunized against the peptides using SulfoLink Coupling Gel (Pierce) column. The column binds sulfur element in Cysteine residues of the peptide that later in the purification process interact with the antibodies from the serum. Seven milliliters of the serums were loaded on the columns and incubated at room temperature for one hour and over-nigh at 4°C while rocking to allow binding of the antibodies to the resin. We left flow-through to leak and washed the remaining serum bound to the column with washing buffer (PBS) to clean the sample from molecules that do not recognize and bind the peptide. Only the specific antibodies stayed bound to the column. Elution with 2ml of glicyne solution (0.2 M glycine•HCl at pH 2.5) released the antibodies from the column. Low pH of the elution buffer is neutralized right away by 100 µl neutralization buffer (1 M Tris•HCl at pH 8.5) in the collection tubes.

Dot blot analysis

We spotted 1µl of peptides of two concentrations (0,5 and 1 µg/µl) onto the nitrocellulose membrane and let the membrane to dry. We blocked non-specific sites by soaking the membrane in 5% milk in PBS TritonX-100 for one hour at room

temperature. We added the corresponding primary antibodies (1:1000 dilution) dissolved in PBS TritonX-100 for one hour at room temperature. We wash the membrane three times with PBS TritonX-100 for 5 minutes. We added rabbit secondary antibody and incubated it for one hour at room temperature. We washed the membrane three times with PBS TritonX-100 for five minutes. We added luminol and enchancer (oxidizing agent) to induce the fluorescence of the secondary antibody. We exposed X-ray film in the dark room to save the antibodies' signal.

Peptide sequences

Chemically synthesized peptides were used for immunization of rabbits, purification procedures, dot blot analysis and peptide competition experiments. Sequence of the peptides are the following ones: K22dH1: PATVEKKVVQKC, 2mK22dH1: PATVE(2mK)KVVQKC, K23dH1:PATVEKKVVQKC, 2mK23dH1: PATVEK(2mK)VVQKC, K27dH1: H-CKKWQKKASGSAGT-NH2, 2mK27dH1: H-CKKWQKK(Me)2ASGSAGT-NH2, Ser10dH1:DSAVATSA(pS)PVAC, P-Ser10H1: DSAVATSASPVAC

Immunostainig of SL 2 cells

We used SL2 for immunostaining experiments when they reached concentration of 3 million per milliliter. The cells were placed into hypotonic solution (50mM glycerol, 5mM KCl, 10mM NaCl, 0.8mM CaCl2, 10mM sucrose) for five minutes. Around 200µl of the cells for each slide were spinned down at 500rpm during 10 minutes (ThermoShandon Cytospin 4) and fixed with 4% paraformaldehyde for 10 minutes. The slides were washed for 15 minutes in PBS after which they could be stored at 4°C or used directly for immunostaining. Immunostaining was started by permeabilisation in PBS, 0,1% TritonX-100 two times for 10 minutes. Blocking of unspecific binding sites was done two times during ten minutes in PBS, 0,1% TritonX-100, 1% BSA. Rabbit α2mK27dH1, αK27dH1 and αGFP antibodies were used at 1/250 1/250 and 1/1000

dilution respectively. Incubation was one hour at room temperature followed by over-night incubation at 4°C in a humid chamber. We washed the antibody that did not bind with PBS, Tween 0,1% TritonX-100 for ten minutes. We used Cy3-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch) at 1/200 dilution for 45 minutes on a room temperature, protected from light. The antibody that did not bind was eliminated by three times ten minutes washes in PBS, 0,1% TritonX-100, 1% BSA. The last wash was in PBS without the detergent. 15µl Mowiol (Calbiochem-Novabiochem) containing 10% DAPI (2ng/µl). Preparations were mounted in Mowiol containing DAPI. The pictures were taken using a Leica SPE confocal microscope (High-Throughput image acquisition with 60x water objective) and analyzed using ImageJ macro software.

Peptide competition assay used in immunostaining experiments

We prepared SL2 cells as in the regular immunostaining experiment. Different form the standard procedure, in peptide competition assay, we pre-incubated α-2mK27dH1 with 2mK27dH1 or K27dH1 or 2mK28dH1 peptides or recombinant dH1 protein prior to use in immunostaining assays. As a control antibody staining we used the antibody not pre-incubated with peptide. All other parameters of the immunostainig experiment were constant for all immunostaining experiments.

Immunostainig of polytene chromosomes

Polytene chromosomes were obtained from third-instar larvae raised at 25°C. Dissection of salivary glands and immunostaining were performed in Cohens buffer (10mM MgCl2, 25mM sodium glycerol 3P, 3mM CaCl2, 10mM KH2PO2, 0.5% NP-

40, 30mM KCl, 160mM sacarosa). Fixation was done in PBS supplemented with 18.5% formaldehyde for 2 minutes and later in 50% acetic acid, 18.5% formaldehyde for 3 minutes. After the fixation, salivary glands were squashed on the microscope glass to make chromosome spreads and were soaked into liquid nitrogen. The microscope slides were washed three times for five minutes in PBS 0.05% Tween®-20. We used 1/30 dilution of α2mK27dH1 antibodies. The antibody was incubated for one hour on room temperature, following over-night incubation at 4°C in a humid box. We washed the antibody that did not bind with PBS, Tween 0,05% Tween-®20 three times for five minutes. We used Cy3-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch) at 1/200 dilution during an hour on a room temperature, protected from light. The antibody that did not bind was eliminated by three times five minutes washes in PBS, 0,05%Tween®-20. The last wash was in PBS without the detergent. Preparations were then mounted in Vectashield mounting medium (Vector Laboratories) with 1.5 µg/ml DAPI (4,6-diamidino-2-phenylindole) and visualized in an Eclipse E-1000 (Nikon) fluorescence microscope equipped with a CoolSnapfx camera (Photometrics) and Metamorph software (version 6.3r1).

Analysis of cells overexpressing JMJD2A

JMJD2A with Flag tag at C-terminus was in the *Drosophila* expression vector, pActPPA. Expression of the transgene was accomplished by addition of calcium-phosphate. 48 hours after the induction of the transgene expression, cells were collected and prepared for immunostaining experiments, like explained above. The pictures of the slides were taken with Leica 19 SPE confocal microscope (High-Throughput image acquisition with 60x water objective) and analyzed using ImageJ macro software developed by Microscope facility at the IRB Barcelona institute. The software was analysing images of DAPI to select for cells in a good shape that would be included in the analysis. After the selection of these cells, a signal of Cy2 and Cy3 that were marking JMJD2A and α-2mK27dH (respectively) were detected. The mean intensity of all pixels in the cell field was calculated and used as a final value.

Discussion

Extent of dH1 depletion in different tissues

In *Drosophila melanogaster* it is practically inconceivable to have a complete knock out of dH1 protein since there are more than hundred copies of dH1 gene and one should eliminate all of them by homologous recombination approach. Only by targeting mRNA by using RNAi machinery, one can deplete the protein level in great extent. In *Drosophila melanogaster* there is only a single variant with polymorphism in two amino acids, so that a unique dH1 hairpin activates RNAi machinery against all dH1 transcripts. Level of small hairpin against dH1 ($shRNA^{H1}$) depletion that we could provoke in different *Drosophila* organs is not the same. The depletion level depends on initial level of dH1, on presence of regulatory mechanisms that counteract the manipulation in dH1 amount and effectiveness of promoter that drives $shRNA^{H1}$ expression in a specific tissue. Initial levels of dH1 are not the same in different tissues and organs. It is known that dH1 level is greater in differentiated than undifferentiated cells. It also seems reasonable that particular tissues might have regulatory mechanisms that would counteract depletion of dH1 we are causing. In accordance to this, we have observed that overexpression of dH1 (data not presented) is not provoking any obvious phenotype in *Drosophila* which is surprising, but could be explained by the existence of mechanisms that tend to maintain a constant level of dH1 in a cell. Our optimization of conditions for RNAi that included a search for the optimal temperature conditions, driver that would lead expression of $shRNA^{H1}$ and target tissue where the sufficient dH1 depletion can be achieved, have teach us that salivary glands are the organs that show the best depletion of dH1 when a general driver, Actin5C driver, is used. Imaginal discs have a good level of depletion too, but brains are organs where little H1 depletion could

be observed. Only simultaneous overexpression of Dicer2, a component of RNAi machinery pathway, is leading to a decrease in dH1 level in brain, detectable by western blot analysis. This is how we decided to use imaginal discs in majority of further analysis.

Effect of dH1 on gene expression

Specificity of observed phenotypes to dH1 loss

The substantial portion (>97%) of genes we detected as DE are occupied by dH1 in wild type conditions. We used only ModEncode data about the distribution of dH1 in SL2 and BG3 cells, since there is no available information about dH1 distribution in wing imaginal discs. We had to presuppose that the distribution is similar in these systems. Occupancy of nearly all affected genes by dH1 is suggesting that the majority of the effects we are seeing in dH1 mutant, including effects on gene expression, are direct consequences of dH1 loss.

dH1 depletion affects small portion of genes

In *Drosophila melanogaster* dH1 is acting as suppressor of position effect variegation (Lu et al., 2009) which suggests that dH1 can regulate transcription of genes. In our *D. melanogaster* mutants, depletion of 90% of dH1 in wing imaginal discs of larvae is provoking rather mild effect on gene expression. Only 4.35% of genes present in Affymetrix analysis are changing their expression with 1.5 fold change and 1.35% with 2.5 fold change. This result goes in line with *in vivo* findings of limited effect of H1 on gene expression in other species. For instance, knock out of all chicken linker histones provoked two fold change in expression of only 4.2% examined genes (Hashimoto et al., 2010) and different extent in depletion of various H1 isoforms in human breast

cancer cells had an effect in transcription of 1-2% of genes, depending on the H1 isoform (Sancho, et al., 2008).

This finding is suggesting that dH1 has predominant role as a structural element of chromatin, rather than as a regulator of transcription. dH1 has an effect on transcription of genes that are physically close to each other, suggesting that it is controlling transcription by affecting chromatin structure. It is possible that dH1 makes the whole chromosome region more open and more accessible to the transcriptional machinery. The other possibility is that dH1 contributes to the formation and organization of nuclear compartments with transposable elements and heterochromatin portion of the genome (it was reported that transposable elements can make these cluster). dH1 depletion would then disturb these structures and perturb expression of all genes located in them.

dH1 mainly acts as a gene repressor

When it comes to direction in which gene expression is changed when dH1 amounts are decreased, our results showed that the genes can be both up- and downregulated. Genes that are inactive in wild type are mainly upregulated in dH1 mutant. In conditions of smaller concentration of dH1 these genes are probably more accessible to transcriptional machinery. Genes that are active in wild type tend to be downregulated in dH1 mutant. In normal conditions, genes that are active are depleted of dH1 at least around TSS. At first glance, downregulation of these genes seems difficult to understand. However, we believe that dH1 protein distributed along these genes (downstream from TSS) also play an important role in regulation of transcription by organizing folding of chromatin fiber. Proper fiber folding during transcription might facilitate the whole process and prevent interactions between DNA and nascent RNA polymerase transcripts when R loops can be formed.

We have found roughly the same number of upregulated and downregulated genes with more than 1.5 fold change in their expression compared to the control expression level. However, among genes that are changing their expression in greater extend, higher than

2.5 fold, the vast majority are upregulated, revealing that dH1 preferentially functions as a repressor of gene expression in *D. melanogaster*. In mice, H1 mutant effects on gene transcription are both positive and negative, but predominantly H1 is acting as a repressor as well, since its loss highly upregulates expression of several genes (Fan et al., 2005). On the contrary, chicken linker histones and human isoform H1.2 act as activators of gene expression since their knock out is provoking dowregulation of certain genes (Hashimoto et al., 2010 and Sancho et al., 2008, respectively).

dH1 affects expression of greater portion of heterochromatic than euchromatic genes

Loss of dH1 changed expression in greater portion of heterochromatin than euchromatin genes included in Affymetrix analysis. This observation is consistent with the already observed extensive effect of dH1 on heterochromatin (Lu et al., 2009). Namely, linker histone in *Drosophila melanogaster* is essential for the establishment of chromocenter since dH1 loss causes disruption of the structure. In our work, we showed that this effect is not purely morphological, but that it changes levels of expression of the genes embedded in heterochromatin.

dH1 silences TE expression

Among upregulated genes we identified a specific group of genomic elements, transposable elements (TE). Even before our results, there was an indication that dH1 might be involved in the regulation of TE silencing. Namely, dH1 has been mapped by DamID method along transposons, starting from transcription start site of the elements (Braunschweig et al., 2009), but no regulatory meaning could be attributed to dH1 at that time. With gene expression analysis we could see that the presence of dH1 is having a role in repression of their expression. TEs are located both in euchromatin and

heterochromatin (Bartolome et al., 2002), saying that their overexpression is not due simply to the location of transposons in heterochromatin. More abundant transcripts that we detected could originate from the increase in their expression from euchromatin or/and heterochromatin. There is no convenient way to determine this, but the fact that the vast majority of full-length TEs are located in euchromatin (Kaminker et al., 2002) is actually suggesting that detected transcripts are probably deriving mostly from euchromatin. Linker histone is a novel player in regulation of TE silencing and probably it acts on transcriptional level, but we don't know the molecular mechanism by which dH1 could act. It is possible that it prevents the access of transcriptional machinery by compacting chromatin. There also might be interplay between dH1 and already known elements that are regulating TE expression. The majority of studies dealing with the subject of transposons silencing are done in germ-line cells and little is known about their regulation in somatic cells. In somatic cells, at the transcriptional level, HP1 is an important protein in this sense (Minervini et al., 2007). There are also indications that DNA methylation could be involved in this process (Phalke et al., 2009) although existence of DNA methylation in *Drosophila melanogaster* is still not certain. However it is not likely that dH1 cooperates with HP1 to repress TE since in dH1 mutants there is actually an increase in HP1 level (Lu et al., 2009). One would expect to observe a positive correlation in the amount of dH1 and HP1a in the case of their putative cooperation. Interpretation of increase in HP1a level could be that the cell is activating a response on TE activation by synthesizing proteins that are involved in their silencing. As mentioned above, there is HP1 independent mechanism of transposons silencing (Phalke et al., 2009).

dH1 is necessary to maintain genome stability

Our results are showing that loss of dH1 is provoking genomic instability which is reflected in production or extrachromosomal circular (ecc) DNA deriving from rDNA locus and the presence of DNA double--strand breaks (DSB). It is possible that the production of ecc is limited only to rDNA locus and that it can be explained by activation of TE encountered in the locus, R1 and R2. In wild type conditions these

transposons are not transcribed or are transcribed at very low level compared to transcription of un-inserted rDNA units (Dawid & Rebbert, 1981). ChiP analysis on inserted and un-inserted units of rDNA showed that H3K9me3 mark is highly present in both inserted and un-inserted rDNA units, suggesting that majority of rDNA units are transcriptionally inactive (Ye and Eickbush, 2006). In our dH1 mutant we see upregulation of R1 and R2 expression, but upregulation of rDNA units as well. It is possible that silencing of R1 and R2 is actually necessary for silencing of the whole rDNA locus. There are reports about the involvement of transposons in regulation of transcription of genes in their close vicinity (Lippman et al, 2004). It is possible that silencing of R1 and R2 is contributing to regulation of a number of transcriptionally active rDNA units which is very important for proper cell functioning. In this way, H1 by controlling expression of R1 and R2 TEs is regulating proper expression of rDNA locus. There are additional data suggesting involvement of H1 is regulation rDNA locus. It is known that *Xenopus laevis* H1 is repressing rDNA locus expression. In order to activate the locus expression, upstream binding factor (UBF) needs to displace H1 (Kermekchiev & Workman, 1997). In *Saccharomyces cerevisiae* linker histones are preventing recombination at rDNA locus (Li et al., 2008).

ecc DNA can contain sequences of various genes that are repeated in the genome, like rDNA, histone genes, stellate and satellite sequences (Cohen et al., 2003). We detected rDNA in ecc in dH1 mutant, but we were not able to detect stellate genes in ecc. However, SU(VAR)3-9 mutant neither contain dramatically increased stellate ecc DNA (Peng and Karpen, 2006) compared to wild type larvae, possibly saying that there is a specific susceptibility of particular repeated sequences for ecc formation. It would be informative to determine if any other group of repetitive sequences or genes are found in ecc of dH1 mutants. It is interesting that in the microarray data we produced, we detected members of various gene families to be upregulated (data not presented in results). Among them are CPR (cytochrome P450 reductase proteins), LCP (larval cuticular proteins), GST (gluthathion S-transferase proteins), MST (monosaccharide transporter proteins), PGRP (peptidoglycan recognition proteins) etc and their gene members are enriched for upregulated genes. This is suggesting that there might be problems in stability of number of these genes that in some extent resemble repetitive genes because there are regions of homology between members of a certain gene family

(Cornman et al., 2009). More general effect on stability of different repetitive sequences, more than the effect on rDNA locus, would add linker histones into group of regulators that are in charge of preserving stability of repetitive sequences. Some heterochromatin proteins are in this group and SU(VAR)3-9 is one of them. This methyltransferase is introducing H3K9 methylation mark which is necessary for prevention of recombination events between rDNA units and recombination between satellite sequences, that otherwise result in splitting of nucleoli and satellite aggregation in few ectopic parts, respectively (Lu et al., 2009). From this perspective it is interesting that linker histone in *Saccharomyces cerevisiae* is necessary for prevention of recombination at the rDNA locus (Li et al., 2008).

Apart from SU(VAR)3-9 and its binding partner HP1, components of RNAi pathway are also preserving stability of these sequences and are important for their silencing. All these proteins important for stability of repetitive sequences are involved in heterochromatin formation, suggesting that chromatin environment of heterochromatin is much more favourable for controlling behaviour of repetitive sequences. We and others (Lu et al., 2009) showed that linker histone is also important for heterochromatin functioning (it is establishing chromocenter and gene expression in heterochromatin) and that it is important for stability of rDNA repeats. It is possible that dH1 through heterochromatin formation maintains stability of repetitive sequence. It is also possible that its binding to rDNA locus directly prevents recombination between units or rDNA, just like it is reported in *Saccharomyces cerevisiae* (Li et al., 2008).

dH1 loss causes DNA damage, stops cell proliferation and induces apoptosis

In response to DNA damage, a cell marks the affected DNA region and recruits proteins necessary for the DNA repair. In *Drosophila melanogaster*, right after the DNA damaging, mei-41 kinase (ATM/ATR homologue in *Drosophila*) is phosphorylating H2AV around the lesion site, in both directions along the DNA molecule. Phosphorylated form of H2AV, γ-H2AV, is often used as a marker of DNA damage.

We detected preapoptotic γ-H2AV immunofluorescence signal as another indication of genome instability in H1 mutants. It is possible that great activation of TE is provoking these numerous DSB in DNA as a consequence of their displacement in the genome. The accurate way to check for this would be to do a ChiP experiment with γ-H2AV antibodies and sequence DNA where this mark occurs. γH2AV signal is spread in both directions from the DNA damage site and as a result of sequencing we would have long DNA regions, that would reduce the resolution by which we can localize the precise DNA damage site. However, an eventual enrichment of the sequences in TEs would confirm that TEs displacement is an important cause of DNA damage in dH1 mutant. Another way by which activation of TE can provoke DNA damage and genome instability is by causing mutations in genes crucial for preserving integrity of the genome since mobilization of TE can cause disruption of these genes, disruption of their regulatory sequences etc. In addition, when there are many DSB (caused by TE displacement or by some other means), in some cases DNA repair mechanisms are not re-establishing connections between the corresponding DNA parts. In this way inefficient DNA repair can produce great chromosomal rearrangements and genome instability.

Another possible cause of phenotypes (among others, DNA breaks) in dH1 mutant could be a problem in progression of DNA replication fork, especially at the sites of repetitive sequences that are difficult to replicate. It was reported that problems in progression of replication fork may cause some of the problems we detected in dH1 mutant. Namely, obstacles in DNA replication fork have implications on hyper-recombination rate (in dH1 mutant we have ecc that are formed most probably as a result of intrachromosomal recombination), DNA damage and genome instability. In addition, DNA replication is more difficult exactly at sites of repetitive elements, transposable elements and heterochromatin sequences (Rozenzhak et al., 2010). DNA replication fork blockage is provoked when there is a collision of the fork with transcriptional machinery. Since dH1 loss causes increase in transcription in rDNA, heterochromatic genes and transposable elements, it is possible that increase engagement of transcription at these already peculiar sequences leads to many collisions of replication forks and transcriptional machinery, resulting in DNA damage. ChiP analysis that would determine the approximate location of DNA damage would help

understanding this better. As mentioned previously, γ-H2AV signal is located upstream and downstream from the exact damage site, so that we could know just rough site of the DNA damage. It is possible that many sequences containing γ-H2AV signal would be close to the sites of problematic replication progression, which then might confirm our predictions.

Extent of DSB in dH1 mutants is high, judged by immunofluorescence images staining for γ-H2AV. When encountered in the situation of DNA damage, a cell can choose between few strategies. The cell can stop cell cycle until the damage is repaired. Alternatively, the cell can decide to enter apoptosis, especially when there is an excess of DNA damage and all cellular elements responsible for DNA repair are occupied. Our immune staining with cleaved-caspese3 antibodies and analysis of cell cycle progression (data not shown) are going in favour of the second strategy in the presence of DNA damage caused by dH1 loss. Range of DNA damage is so high that the cell is not stopping cell cycle in tries to repair the damage, but decides to enter apoptosis and save the rest of the body of possibly deleterious effects DNA damage and fixation of mutations.

These effects we actually observe when we block apoptosis. There is a hyperproliferation resulting in tissue overgrowth, an effect that most probably comes as a result of inefficient DNA repair when numerous gene mutations and great chromosomal rearrangements are fixed and transmitted to all cells deriving from the mother cell with the initial mutation. Accumulation of more mutations, particularly the once that regulate genes involved in cell cycle progression, finally results in uncontrolled cell growth and tissue transformation (Figure 18).

Learning more about the proteins that detect DNA damage and about the way the cells die upon dH1 depletion, could tells us which problems occur in the conditions of decreased amounts of dH1. To do this, we would have to test whether some of the mutant states for the key players in DNA damage response could rescue the phenotypes we are observing in dH1 mutants. Mei-41 kinase phosporylates H2AV and spreads this mark along DNA molecule. Afterwards, kinase activity of checkpoint kinase 1 (Chk1) and checkpoint kinase 2 (Chk2) are responsible for the arrest in cell cycle and, if necessary, for the activation of the cell death. Our laboratory has preliminary results of

the MNK (*Drosophila* homolougue of Chk2), p53 and Grapes (*Drosophila* homologue of Chk1) mutant experiments.

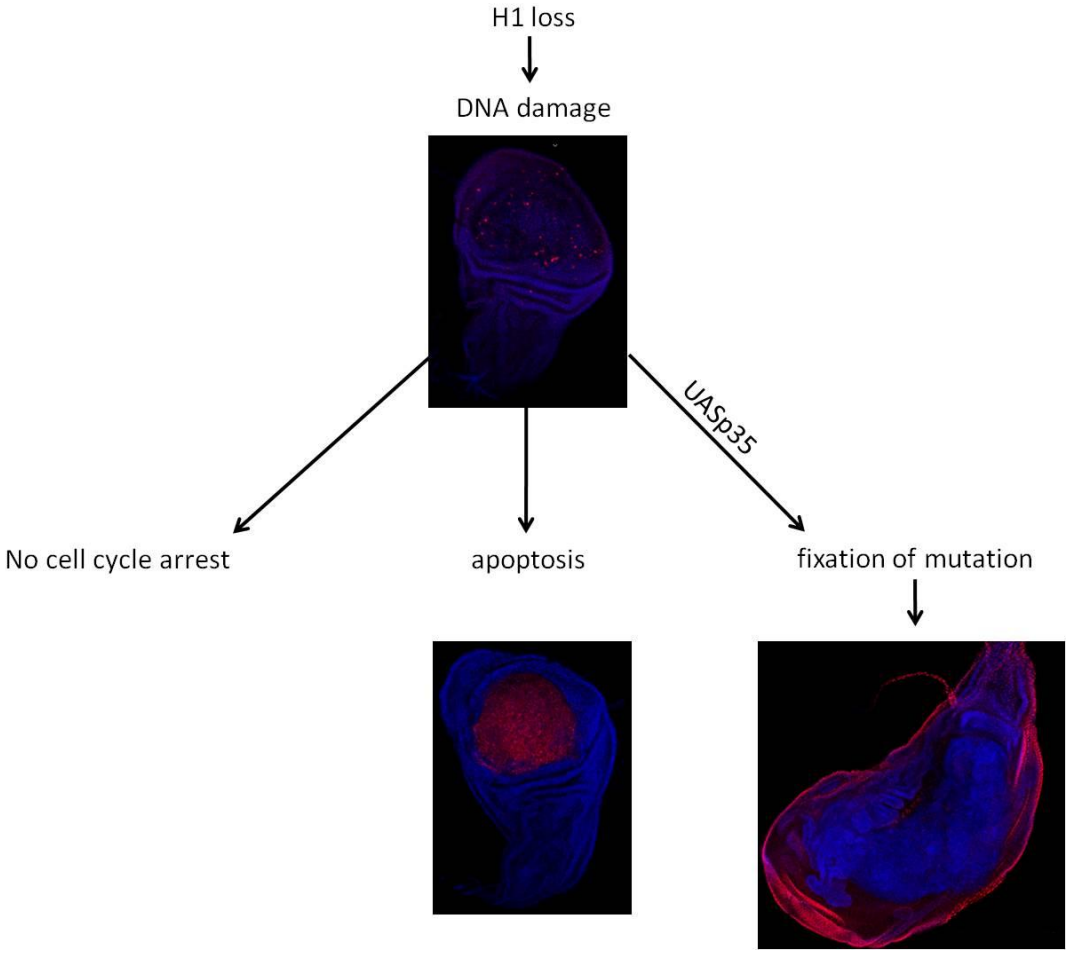

Figure 18: Consequences of H1 loss and decisions that a cell takes as a response

PTMs of dH1

Our laboratory detected various PTMs in H1 isolated from *Drosophila melanogaster* cells in culture, SL2 cells. We decided to raise antibodies against 2mK27dH1 modification since no methylation has been described so far in dH1 and since human H1.4 linker histone has di-methylation on N-terminus as well, on residue Lys26 (Wisniewski et al.,2007).

Crucial property of the antibody we made was its specificity towards di-methylated form of H1K27 and no reactivity with unmodified form of the protein. We reached this level of specificity by affinity purification of the antibody on columns. Dot blot analysis proved selectivity of the purified α2mK27dH1 for solely H1K27 when it is methylated. In addition, western blot analysis confirmed it, since α2mK27dH1 recognizes endogenous dH1 protein better than bacterially overexpressed, unmodified protein. In some extent recombinant protein is also recognized by the antibody, but much less then endogenous one. When discussing this it is important to bear in mind that the level of dH1 protein that is modified is very low and that in the western blot analysis a large amount of recombinant protein is loaded that surely influences antibody's specificity. Peptide competition experiments gave us a reason more to believe in antibody specificity since only modified peptide is competing with endogenous epitop binding, while no such an effect is detected for unmodified version of the peptide.

After proving antibody's specificity, we were examining the presence of the modification in polytene chromosomes and SL2 cells.

We have observed distinctive staining patterns for αK27dH1 and α2mK27dH1 in SL2 cells. In polytene chromosomes we could not see any difference in staining with two antibodies and this could mean that the modification is uniformly distributed along the chromosomes. In SL2 cells αK27dH1 that recognizes all dH1 molecules, regardless of their modification state, is uniformly bound in the whole nucleus and metaphase chromosome surface. On the other hand, α2mK27dH1 stains nucleus with less affinity

than for metaphase chromosomes and in the chromosomes it preferentially binds to centromeric regions. Centromeric chromatin has many particularities in comparison to the rest of the chromatin. This specialized heterochromatic region enables chromosome attachment to the spindle and segregation of chromosomes into daughter cells. Many proteins involved in kinetochore assembly and centromere function are highly conserved and it even contains distinctive nucleosome different than in canonical ones (Palmer et al, 1991). Centromeric nucleosome consists of histone H3 variant - CenH3 (variant is called CID in *Drosophila*) together with H4, H2A, and H2B. Centromere chromatin is also genome region where numerous transposons are inserted. It is possible that 2mK27dH1 also plays a role in specification of centromeric chromatin.

Only UAS promoter is capable to induce detectable levels of the dH1-GFP constructs, while endogenous dH1 protein is not activating the protein expression enough to be detected by western blot or by GFP direct fluorescence. The explanation for this most probably comes from the existence of more than a hundred copies of dH1 gene in the fly genome, when addition of only one more transgene is not changing significantly and detectably the total pool of linker histone protein in a cell. On the other hand, UAS promoter is a great activator of downstream sequences so that even one copy of it is enough to produce great amounts of the protein.

We didn't manage to determine physiological significance of 2mK27dH1 modification since we could not observe any phenotype in flies overexpressing mutated form of dH1. Absence of Lys residue on positions 27 and/or 28 did not make any obvious effect in flies carrying the mutant protein. In this experiment we are only overexpressing the mutant forms of the protein, while endogenous protein is still present. We expected that the presence of mutant protein could act as a dominant negative form since it competes with endogenous protein for binding sites in chromatin. It is possible that the 2mK27dH1 is not an essential modification, not even when both Lys27 and Lys28 are changed for Ala or that manipulated form of the protein is actually not capable of binding to the chromatin and that that's why we don't see any effect. There is a report saying that in human H1, tagging of the protein is diminishing ability of the protein to bind DNA (Hendzel et al., 2004). The signal we detect in salivary glands is nuclear (Figure 16) which is a good indication that it might bind DNA, but more reliable would be ChiP confirmation of the protein binding.

We addressed the question about demethylase that erases the modification mark. We had a candidate approach based on homology with the human enzyme that erases H1K26me2 mark in humans (Trojer et al., 2009). Sequence analysis is showing that this part of the N-terminus is well conserved from flies to humans and we reasoned that there might be the conservation of the enzymes that regulate PTMs in this region of H1 protein. Our results suggest that the homologue, JMJ2DA protein, indeed is a potential HDM for 2mK27dH1. The main concern we have in the interpretation of the results is that initially there is a great variability in endogenous 2mK27dH1 levels which complicates comparison with the level of 2mK27dH1 in cells that overexpress the candidate HDM. Additional experiments are needed to confirm this result. Making a mutant state for the putative demethylase and an increase in the level of the modification signal could be one of the approaches to be taken, but in this way we would still encounter the problem of the signal variability and its quantification. The main problem with cell line that overexpresses JMJD2A is that only a small portion of the cells are induced to produce the protein. The better solution would be to study the mutant state for the enzyme and to search for more sensitive ways to quantify the exact levels of 2mK27dH1 mark, for instance, by making western blot analysis in the mutant and wild type state.

Further questions that are of interest and that would help understanding the mark role in a cell are the following ones: what are the features of chromatin enriched in the mark, which proteins recognize this mark and what is the final result of their interaction, does it change during development, why does it change during the cell cycle, what does it mean for the global chromatin structure – heterochromatin, centromeric chromatin etc.

Conclusions

- Histone H 1 of *Drosophila melanogaster* (dH1) is mainly gene repressor and affects greater portion of genes in heterochromatin than genes in the rest of the genome

- dH1 represses transposable elements in *Drosophlia meanogaster*

- dH1 contributes to regulation of proper expression of rRNA locus

- dH1 is necessary for genome stability

- dH1 is essential for cell viability and proliferation

- Antibodies that specificity recognize 2mK27dH1, 2mK23dH1 and pSer10dH1 have been raised and purified

- α2mK27dH1 signal is concentrated in pericentromeric regions suggesting that the modification might be contributing to properties of heterochromatin;

Bibliography

Allfrey, V. G, V. C. Littau, Mirsky A. E. (1963). On the role of histones in regulating ribonucleic acid synthesis in the cell nucleus. *Methods*, 1238: 414-421.

Bartolome C., Maside X. and Charlesworth B. (2002). On the abundance of transposable elements in the genome of *Drosophila meanogaster*. *Mol*. *Bol*. *Evol*. 19(6): 926-937

Berloco, M., Fanti, L., Breiling, a, Orlando, V., & Pimpinelli, S. (2001). The maternal effect gene, abnormal oocyte (abo), of *Drosophila melanogaster* encodes a specific negative regulator of histones. *PNAS*, 98(21): 12126-31.

Blank, T. A., & Becker, P. B. (1996). The effect of nucleosome phasing sequences and DNA topology on nucleosome spacing. *J*. *Mol*. *Bio*., 260(1): 1-8.

Bonner, J., Huang, R.-chih C. (1963). Properties of chromosomal nucleohistone. . *J*. *Mol*. *Bio*, 6(3): 169-174.

Brand, H., Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, 118(2): 401-15.

Braunschweig, U., Hogan, G. J., Pagie, L., van Steensel, B. (2009). Histone H1 binding is inhibited by histone variant H3.3. *The EMBO journal*, 28(23): 3635-45.

Brown, D. T. (2001). Histone variants: are they functionally heterogeneous? *Genome Biol*., 2(7): reviews0006.1-reviews6.6

Brown, D. T., Alexander, B. T., Sittman, D. B. (1996). Differential effect of H1 variant overexpression on cell cycle progression and gene expression. *Nucl. Acids Res.* 24(3): 486-93.

Li C., Mueller, J. E., Elfline, M., & Bryk, M. (2008). Linker histone H1 represses recombination at the ribosomal DNA locus in the budding yeast *Saccharomyces cerevisiae*. *Mol. Microbiol*., 67(4): 906-919.

Cells, L. L. Y. (1987). Synthesis and Degradation of H1 histone subtypes in mouse lymphoma L5178Y cells. *J. Biol. Chem*., (27): 13075-13080.

Cho, H. Wolffe A.P. (1994). *Xenopus laevis* B4, an intron-containing oocyte-specific linker histone-encoding gene, *Gene*, 143: 233-238.

Cohen S., Yacobi K. and Segal D. (2003).Extracomosomal circular DNA of tandemly repeated genomic sequences in *Drosophila*. *Genome Res*.; 13(6a): 1133-1145

Cole, R. D. (1987). Microheterogeneity in H1 histones and its consequences. *Int. J. Pept. Protein Res*., 30(4): 433-449.

Consortium, T., Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., et al. (2011). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. Oct, 330(6012): 1787-1797.

Cornman R. S. (2009). Molecular evolution of *Drosophila* cuticular protein genes.PLoS One. Dec 17; 4(12): e8345

Corona, D. F. V., Siriaco, G., Armstrong, J. a, Snarskaya, N., McClymont, S. a, Scott, M. P., & Tamkun, J. W. (2007). ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo. *PLoS Biol*. Sep; 5(9): e232.

Cremer, T., & Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet*, Apr; 2(4): 292-301.

Daujat S., Zeissler U., Waldmann T., Happel N., Schneider R. (2005). HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J. Biol. Chem*.Nov 11; 280(45): 38090–38095.

Davey, C. A, Sargent, D. F., Luger, K., Maeder, A. W., Richmond, T. J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol.*, Jun 21; 319(5): 1097-113.

Dawid, I. B., Rebbert, M. L. (1981). Nucleotide sequences at the boundaries between gene and inserton regions in *Drosohila melanogaster*. *Nucl. Acids Res*. 9(19): 5011-5020.

Doenecke, D., Albig, W., Bouterfa, H., Drabent, B. (1994). Organization and expression of H1 histone and H1 replacement histone genes. *J. Cell. Biochem.*, Apr; 54(4): 423-431.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science*, Jun 25; 284(5423): 2124-2128.

Dou, Y., Mizzen, C. A., Abrams, M., Allis, C. D., Gorovsky, M. A., York, N. (1999). Phosphorylation of Linker Histone H1 regulates gene expression in vivo by Mimicking H1 Removal. *Mol. Cell*, Oct; 4(4): 641-647.

Fan, Y., Nikitina, T., Morin-kensicki, E. M., Zhao, J., Magnuson, T. R., Woodcock, C. L., Skoultchi, A. I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo H1. *Mol. Cell Biol.*, 23(13): 4559-72

Fan, Y., Nikitina, T., Zhao, J., Fleury, T. J., Bhattacharyya, R., Bouhassira, E. E., Stein, A., et al. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, 123(7): 1199-212.

Fanti L. and Pimpinelli S. (2004). Analysis of mitosis in squash preparations of larval brains orcein, giemsa, hoechst 33258, DAPI, quinacrine, and N-banding. *Methods Mol. Biol*, 247: 325-332

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., et al. (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, 143(2): 212-24.

Garcia B.A., Busby S.A., Barber C.M. Shabanowitz J., Allis C.D. and Hunt D.F. (2004). Characterization of Phosphorylation Sites on Histone H1 Isoforms by Tandem Mass Spectrometry. *J. Proteome Res.*,3 (6): 1219–1227

Gadad, S. S., Senapati, P., Syed, S. H., Rajan, R. E., Shandilya, J., Swaminathan, V., Chatterjee, S., et al. (2011). The multifunctional protein nucleophosmin (NPM1) is a human linker histone H1 chaperone. *Biochemistry*, 50(14): 2780-9.

Godde, J. S., Ura, K. (2009). Dynamic alterations of linker histone variants during development. *Int. J. Dev. Biol.*, 53(2-3): 215-24.

Gordon, F., Luger, K., Hansen, J. C. (2005). The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. *J. Biol. Chem.*, 280(40): 33701-6.

Grewal, S. I. S. (2011). RNAi-dependent formation of heterochromatin and its diverse functions. *Curr. Opin. Genet. Dev*, 20(2): 134-141.

Gunjan, a, Sittman, D. B., Brown, D. T. (2001). Core histone acetylation is regulated by linker histone stoichiometry in vivo. *J. Biol. Chem.*, 276(5): 3635-40.

Guohong Li and Danny Reinberg. (2011). Chromatin higher-order structures and gene regulation. *Chromatin* ,Oct, 21(2): 175-186.

Hannon, R., Bateman, E., Allan, J., Harborne, N., & Gould, H. (1984). Control of RNA polymerase binding to chromatin by variations in linker histone composition. *J. Mol. Biol.*, 180(1): 131-149.

Hartl T, Boswell C, Orr-Weaver TL, Bosco G. (2007). Developmentally regulated histone modifications in *Drosophila* follicle cells: initiation of gene amplification is associated with histone H3 and H4 hyperacetylation and H1 phosphorylation. *Chromosoma*; 116(2): 197-214.

Hashimoto, H., Sonoda, E., Takami, Y., Kimura, H., Nakayama, T., Tachibana, M., Takeda, S., et al. (2007). Histone H1 variant, H1R is involved in DNA damage response. *DNA Repair*, *6*(11): 1584-1595.

Hashimoto, H., Takami, Y., Sonoda, E., Iwasaki, T., Iwano, H., Tachibana, M., Takeda, S., et al. (2010). Histone H1 null vertebrate cells exhibit altered nucleosome architecture. *Nucl. Acids Res*, *38*(11): 3533-45.

Hendzel, Michael J, Lever, M. a, Crawford, E., Th'ng, J. P. H. (2004). The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J. Biol. Chem.*, *279*(19): 20028-34.

Herrera, J. E., West, K. L., Schiltz, R. L., Nakatani, Y., Bustin, M. (2000). Histone H1 is a specific repressor of core histone acetylation in chromatin. *Mol Cell. Biol.*, *20*(2): 523-9.

Holmgren, P., Johansson, T., Lambertsson, A., Rasmuson, B. (1985). Content of histone H1 and histone phosphorylation in relation to the higher order structures of chromatin in *Drosophila*. *Chromosoma*, *93*(2): 123-131.

Hoskins, R. A., Carlson, J. W., Kennedy, C., Acevedo, D., Evans-Holm, M., Frise, E., Wan, K. H., et al. (2007). Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science*, *316*(5831): 1625-1628.

Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., & Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.*, *21*(22): 2936-49.

James, T. C., & Elgin, S. C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. *Mol. Cell. Biol.*, *6*(11): 3862-72.

Jamrich, M., Greenleaf, A. L., & Bautz, E. K. F. (1977). Localization of RNA polymerase in polytene chromosomes of *Drosophila melanogaster*. *Genetics*, *74*(5): 2079-2083.

Jedrusik, M. A., & Schulze, E. (2001). A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in *Caenorhabditis elegans*. *Development, 128*(7): 1069-80.

Jedrusik, M. A., Schulze, E. (2007). Linker histone HIS-24 (H1.1) cytoplasmic retention promotes germ line development and influences histone H3 methylation in Caenorhabditis elegans. *Mol. Cell. Biol.*, *27*(6): 2229-39.

Juan, L. J., Utley, R. T., Adams, C. C., Vettese-Dadey, M., Workman, J. L. (1994). Differential repression of transcription factor binding by histone H1 is regulated by the core histone amino termini. *EMBOJ.*, *13*(24): 6031-40.

Juan, L. J., Utley, R. T., Vignali, M., Bohm, L., Workman, J. L. (1997). H1-mediated Repression of Transcription Factor Binding to a Stably Positioned Nucleosome *EMBO J*. *272*(6): 3635-3640.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R.,Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M. et al. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomic perspective. *Genome Biol*., 3, (12): research0084.0081-0084.0020.

Kandolf, H. (1994). The H1A histone variant is an in vivo repressor of oocyte-type 5S gene transcription in Xenopus laevis embryos. *Proc. Nat. Acad. Sci. U. S. A.*, *91*(15): 7257-61.

Kasinsky, H E, Lewis, J. D., Dacks, J. B., Ausió, J. (2001). Origin of H1 linker histones. *FASEB J.*, *15*(1): 34-42.

Kermekchiev, M., Workman, J. L. (1997). Nucleosome binding by the polymerase I transactivator upstream binding factor displaces linker histone H1. *Mol. Cell. Biol.*, *17*(10): 5833-42

Kim J., Jeremy Daniel, Alexsandra Espejo, Aimee Lake, Murli Krishna, Li Xia, Yi Zhang, Mark T Bedford (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMB Rep*, 7(4): 397-403

Lilley, D. M. (1992). DNA--protein interactions. HMG has DNA wrapped up. *Nature*, *357*(6376), 282-3.

Lima-De-Faria, a. (1959). Differential uptake of tritiated thymidine into hetero- and euchromatin in Melanoplus and Secale. *J. Biophys. Biochem. Cytol.*, *6*: 457-66.

Lippman Z., Gendrel A., Black M., Vaughn M. W., Dedhia N., McCombie W. R., Lavine K., Mittal V., May B., Kasschau K. D., Carrington J. C., Doerge R. W., Colot V. & Martienssen R. (2004). Role of transposable elements in heterochromatin and epigenetic control . *Nature* 430(6998): 471-476

Lloret-Llinares M., Carre´ C., Vaquero A., de Olano N.and Azorı´n F. (2008). Characterization of *Drosophila melanogaster* JmjC+N histone demethylases. *Nucl. Acids Res*,36(9): 2852-2863

Lu, X., Wontakal, S. N., Emelyanov, A. V., Morcillo, P., Konev, A. Y., Fyodorov, D. V., Skoultchi, A. I. (2009). Linker histone H1 is essential for *Drosophila* development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes Dev.*, *23*(4): 452-65.

Luger, K, M., A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648): 251-60.

Mackey-Cushman, S. L., Gao, J., Holmes, D. A., Nunoya, J.I., Wang, R., Unutmaz, D., & Su, L. (2011). FoxP3 interacts with linker histone H1.5 to modulate gene expression and program Treg cell activity. *Genes Immun.*, *12*(7): 559-567.

Maresca, T. J., Freedman, B. S., & Heald, R. (2005). Histone H1 is essential for mitotic chromosome architecture and segregation in *Xenopus laevis* egg extracts. *J. Cell Biol.*, *169*(6): 859-69.

Marzluff, W. F. (2005). Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr. Opin. Cell Biol.*, *17*(3): 274-280.

Minervini, C.F., Marsano, R.M., Casieri, P., Fanti, L., Caizzi, R.,Pimpinelli, S., Rocchi, M. and Viggiano, L. (2007) Heterochromatin protein 1 interacts with 5'UTR of transposable element ZAM in a sequence-specific fashion. *Gene:* 393: 1-10.

Muller, H. J. (1930). Types of visible variations induced by X-rays in *Drosophila*. *J. Genet.*, *22*(3): 299-334.

Nagel, S., Grossbach, U. (2000). Histone H1 genes and histone gene clusters in the genus *Drosophila*. *J. Mol. Evol.*, *51*(3): 286-298.

Ner, S S, Blank, T., Pérez-Paralle, M. L., Grigliatti, T. a, Becker, P. B., Travers, A. A. (2001). HMG-D and histone H1 interplay during chromatin assembly and early embryogenesis. *J. Bio.l Chem.*, *276*(40): 37569-76.

Ner, Sarbjit S, Travers, A. A. (1994). HMG-D, *Drosophila* homologue. *EMBO J.*, *13*(8): 1817-1822.

Ni, J.-quan, Liu, L.-ping, Hess, D., Rietdorf, J., Sun, F.-lin. (2006). *Drosophila* ribosomal proteins are associated with linker histone H1 and suppress gene transcription. *Genes Dev.*, 20(14): 1959-1973.

Nishiyama, M., Oshikawa, K., Tsukada, Y.-ichi, Nakagawa, T., Iemura, S.-ichiro, Natsume, T., Fan, Y., et al. (2009). CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. *Nature Cell. Biol.*, *11*(2): 172-182.

Orrego, M., Ponte, I., Roque, A., Buschati, N., Mora, X., Suau, P. (2007). Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biology*, *5*: 22.

Palmer, D.K., K. O'Day, H.L. Trong, H. Charbonneau, R.L. Margolis (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl. Acad. Sci. U. S. A.* 88:3734–3738.

Parekh, B. S., & Maniatis, T. (1999). Virus infection leads to localized hyperacetylation of histones H3 and H4 at the IFN-beta promoter, *Mol. Cell. Biol. 3*: 125-129.

Parseghian, M. H., Newcomb, R. L., Winokur, S. T., Hamkalo, B. A. (2000). The distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin: distribution in human fetal fibroblasts. *J. Cell. Biochem.*, *83*(5): 643-659.

Patterton, H. G., Landel, C. C., Landsman, D., Peterson, C. L., Simpson, R. T. (1998). The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, *273*(13): 7268-76.

Pearson, C. E., Nichol Edamura, K., Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nature reviews. Genetics*, *6*(10): 729-42.

Peng, J. C., Karpen, G. H. (2006). H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nature Cell. Biol.* 9: 25 - 35

Peng, J. C., Karpen, G. H. (2009). Heterochromatic genome stability requires regulators of histone H3K9 methylation. *PLoS genetics*, *5*(3): e1000435.

Pennings, S., Meersseman, G., Bradbury, E. M. (1994). Linker histones H1 and H5 prevent the mobility of positioned nucleosomes. *Proc. Nat. Acad. Sci. U. S. A.*, *91*(22), 10275-10279.

Peters, A. H. F. M., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. a H. a, Perez-Burgos, L., Kohlmaier, A., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular cell*, *12*(6): 1577-1589.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M. C., Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila*

depends on the cytosine-5 methyltransferase DNMT2. *Nature Gen.*, *41*(6): 696-702.

Pham A.D. and Sauer F. (2000). Ubiquitin-Activating/Conjugating Activity of TAFII250, a Mediator of Activation of Gene Expression in *Drosophila*. Science 29: 289(5488): 2357-2360.

Ponte, I. (2003). Sequence Complexity of Histone H1 Subtypes. *Mol. Biol. Evol.*, *20*(3): 371-380.

Ramachandran, A., Omar, M., Cheslock, P., Schnitzler, G. R. (2003). Linker histone H1 modulates nucleosome remodelling by human SWI/SNF. *J. Biol. Chem.*, *278*(49): 48590-601.

Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, *362*(6417): 219-23.

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. *Nature*, *311*(5986): 532-537.

Ridsdale, J. a, Hendzel, M. J., Delcuve, G. P., Davie, J. R. (1990). Histone acetylation alters the capacity of the H1 histones to condense transcriptionally active/competent chromatin. *J. Biol. Chem.*, *265*(9): 5150-6.

Robinson, P. J. J., & Rhodes, D. (2006). Structure of the “30 nm” chromatin fibre: a key role for the linker histone. *Curr. Opin. Struct. Biol. 16*(3): 336-43.

Rosidi, B., Wang, M., Wu, W., Sharma, A., Wang, H., Iliakis, G. (2008). Histone H1 functions as a stimulatory factor in backup pathways of NHEJ. *Nucl. Acids Res.*, *36*(5): 1610-23.

Rozenzhak S, Mejía-Ramírez E, Williams JS, Schaffer L, Hammond JA, Head SR, Russell P. (2010). Rad3 decorates critical chromosomal domains with gammaH2A to protect genome integrity during S-Phase in fission yeast. *PLoS Genet.* 22;6(7): e1001032.

Sancho, M., Diani, E., Beato, M., Jordan, A. (2008). Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet.*, *4*(10): e1000227.

Schalch, T., Duda, S., Sargent, D. F., Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, *436*(7047): 138-41.

Schalch, T., Job, G., Noffsinger, V. J., Shanker, S., Joshuator, L., Partridge, J. F. (2009). High-affinity of Chp1 chromodomain to K9 methylated H3 is required to establish centromeric heterochromatin.*Mol*. *Cell*, *34*(1): 36-46.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell*, *128*(4): 735-45.

Shen, X., Gorovsky, M. A. (1996). Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell*, *86*(3): 475-83.

Smith, C. D., Shu, S., Mungall, C. J., Karpen, G. H. (2007). The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science*, *316*(5831): 1586-91.

Stedman, E. (1951). The Basic Proteins of Cell Nuclei. *Philos. Trans..R.Soc. London, Series B, Biological Sciences*, *235*(630): 565-595.

Subirana, J. A. (1990). Analysis of the charge distribution in the C-terminal region of histone H1 as related to its interaction with DNA. *Biopolymers*, *29*(10-11): 1351-7.

Suganuma T, Workman JL. (2011). Signals and combinatorial functions of histone modifications. *Annu Rev Biochem*. 7(80):473-99.

Talasz, H., Helliger, W., Puschendorf, B., Lindner, H. (1996). In vivo phosphorylation of histone H1 variants during the cell cycle. *Biochemistry*, *35*(6): 1761-1767.

Talasz, H., Sapojnikova, N., Helliger, W., Lindner, H., Puschendorf, B. (1998). In vitro binding of H1 histone subtypes to nucleosomal organized mouse mammary tumor virus long terminal repeat promotor. *J. Biol. Chem.*, *273*(48): 32236-43.

Talmage D.A. and M Blumenfeld (1987) *Drosophila melanogaster* H1 histone is phosphorylated stably. *Mol Cell Biol.;* 7(11): 4118–4121

Towbin, B. D., Meister, P., & Gasser, S. M. (2009). The nuclear envelope--a scaffold for silencing? *Curr. Opin. Genet. Dev.*, *19*(2): 180-6.

Trojer P., Zhang J., Yonezawa M., Schmidt A., Zheng H., Jenuwein T., Reinberg D. (2009). Dynamic histone H1 isotype 4 methylation and demethylation by histone

lysine methyltransferase G9a/KMT1C and the Jumonji domain-containing JMJD2/KDM4 proteins. *J. Biol. Chem.* 284(13): 8395–8405.

Ura, K, Hayes, J. J., Wolffe, a P. (1995). A positive role for nucleosome mobility in the transcriptional activity of chromatin templates: restriction by linker histones. *EMBO J.*, *14*(15): 3752-65.

Ura, K, Kurumizaka, H., Dimitrov, S., Almouzni, G., Wolffe, a P. (1997). Histone acetylation: influence on transcription, nucleosome mobility and positioning, and linker histone-dependent transcriptional repression. *EMBO J.*, *16*(8): 2096-107.

Varga-Weisz, P., van Holde, K., Zlatanova, J. (1993). Preferential binding of histone H1 to four-way helical junction DNA. *J. Biol. Chem.*, *268*(28): 20699-700.

Vila, R., Ponte, I., Collado, M., Arrondo, J. L., Suau, P. (2001). Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA: an infrared spectroscopy study. *J. Biol. Chem.*, *276*(33): 30898-903.

Villar-Garea A., Imhof A. (2008). Fine Mapping of Posttranslational Modifications of the Linker Histone H1 from *Drosophila melanogaster*. *PLoS* 3(2): e1553

Weiler, K. S., Wakimoto, B. T. (1995). Heterochromatin and gene expression in *Drosophila*. *Ann. Rev. Genet.*, *29*, 577-605.

Weiss T., Hergeth S., Zeissler U., Izzo A., Tropberger P., Zee B.M, Dundr M., Garcia B.A., Daujat S. and Schneider R. (2010). Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. *Epigenetics Chromatin* 3(1):7

Wisniewski J. R., Zougman A., Kruger S. and Mann M. (2007). Mass Spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol. Cell. Proteomics* 6(1): 72-87

Wood C., Snijders A., Williamson J., Reynolds C., Baldwin J, Dickman M. (2009). Post-translational modifications of the linker histone variants and their association with cell mechanisms. *FEBS J*. 276(14): 3685–3697

Woodcock, C. L., Skoultchi, A. I., Fan, Y. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res., 14*(1): 17-25.

Wustmann, G., Szidonya, J., Taubert, H., Reuter, G. (1989). The genetics of position-effect variegation modifying loci in *Drosophila melanogaster*. *Mol*. *Gen*. *Genet*.: 217(2-3): 520-7.

Ye J. and Eickbush T. H. (2006). Chromatin Structure and Transcription of the R1- and R2-Inserted rRNA Genes of *Drosophila melanogaster*. *Mol*. *Cell*. *Biol*. 26(23): 8781-8790

Figure 1 was taken from the web site:
http://www.mol.biol.ethz.ch/groups/richmond/Achievements

Figure 2 was taken from the publication:
Grewal S. I., Jia S. Heterochromatn revised. (2007) *Nat Rev Genet*. Jan;8(1): 35-46.

Figure 3 was taken from the publication:
Grewal S. I. RNAi-dependent formation of heterochromatin and its diverse functions. (2010) *Curr Opin Genet Dev*. Apr;20(2): 134-41.

Figure 4 was taken from the publication:
Peng J. C., Karpen G. H. Epigenetic regulation of heterochromatic DNA stability. (2008)*Curr Opin Genet Dev*. Apr;18(2): 204- 11.

Figure 5 was taken from the publication:
Zlatanova J., van Holde K. Binding to four-way junction DNA: a common property of architectural proteins? (1998) *FASEB J*. Apr;12(6): 421-31.

Figure 6 was taken from the publication:
Barth T. K., Imhof A. (2010) Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci*. Nov;35(11): 618-26